COMMUN. MATH. SCI.
Vol. 21, No. 2, pp. 351–378

ON ANISOTROPIC NON-LIPSCHITZ RESTORATION MODEL: LOWER BOUND THEORY AND ITERATIVE ALGORITHM*

CHUNLIN WU†, XUAN LIN‡, AND YUFEI ZHAO§

Abstract. For nonconvex and nonsmooth restoration models, the lower bound theory reveals their good edge recovery ability, and related analysis can help to design convergent algorithms. Existing such discussions are focused on isotropic regularization models, or only the lower bound theory of anisotropic model with a quadratic fidelity. In this paper, we consider a general image recovery model with a non-Lipschitz anisotropic composite regularization term and an ℓ_q norm ($1\leq q\leq+\infty$) data fidelity term. We establish the lower bound theory for the anisotropic model with an ℓ_1 fidelity or an ℓ_∞ fidelity, which applies to impulsive noise or uniform noise (quantization error) removal problems. For the general case with $1\leq q\leq+\infty$, a support inclusion analysis is provided. To solve this non-Lipschitz composite minimization model, we are then naturally motivated to extend previous works to introduce a support shrinking strategy in the iterative algorithm and relax the support constraint to a τ-support (a thresholded version) constraint, which is more consistent with practical computation. The objective function at each iteration is also linearized to construct a strongly convex subproblem. To make the algorithm more implementable, we compute an approximation solution to this subproblem at each iteration, but not an exact one. The global convergence result of the proposed inexact iterative thresholding and support shrinking algorithm with proximal linearization is established. The experiments on image restoration and two-stage image segmentation demonstrate the effectiveness of the proposed algorithm.

Keywords. Image restoration; anisotropic model; non-Lipschitz optimization; lower bound theory; thresholding; support shrinking.

AMS subject classifications. 49K30; 49N45; 90C26; 94A08.

1. Introduction

We consider the problem of image recovery, i.e., trying to find the unknown ground truth $\underline{x}\in\mathbb{R}^N$ from the degraded observation $b\in\mathbb{R}^M$ as follows

$$b=A\underline{x}+n.$$

Herein, M and N are positive integers, $n\in\mathbb{R}^M$ represents the noise, and the matrix $A\in\mathbb{R}^{M\times N}$ is related to the information acquisition process, e.g., an identity matrix for the image denoising problem, a convolution matrix generated by blur kernel for image deblurring, and a measurement matrix in the problem of compressed sensing. This linear inverse problem is usually ill posed. To solve such a problem, we consider the following minimization model

$$\min_{x\in\mathbb{R}^N}\sum_{i\in\mathrm{J}}\phi(|G_i^\top x|)+\mathcal{Q}_q(x), \tag{1.1}$$

where

$$\mathcal{Q}_q(x)=\begin{cases}\frac{\beta}{q}\|Ax-b\|_q^q, & q\in[1,+\infty),\\ \beta\|Ax-b\|_\infty, & q=+\infty.\end{cases} \tag{1.2}$$

*Received: March 02, 2021; Accepted (in revised form): May 29, 2022. Communicated by Wotao Yin.

†School of Mathematical Sciences, Nankai University, Tianjin 300071, China (wucl@nankai.edu.cn).

‡School of Mathematical Sciences, Nankai University, Tianjin 300071, China (1120200026@mail.nankai.edu.cn).

§Corresponding author. School of Mathematical Sciences, Nankai University, Tianjin 300071, China (matzyf@nankai.edu.cn).

Herein, ϕ is a potential function satisfying certain properties, J is an index set, $\{G_i\}_{i\in \mathrm{J}} \subset \mathbb{R}^N$ are the columns of matrix $G \in \mathbb{R}^{N\times \sharp \mathrm{J}}$, which can be a set of sparsifying operators (e.g., the discrete gradient operators or the atoms in a wavelet transform), and $G_i^\top x$ denotes the transform coefficient. As one can see, the ℓ_q norm $\|\cdot\|_q$ $(1 \le q \le +\infty)$ is adopted in the data fidelity term of model (1.1) and it can be used to handle different additive noises by choosing different values of q from a maximum likelihood (ML) estimation. For example, people use the squared ℓ_2 norm to deal with the white Gaussian noise [11, 15, 22], the ℓ_1 norm for impulse noise removal [35, 40, 55, 60], and the ℓ_∞ norm for uniform noise (or quantization error) removal [13, 26, 66]. It is also known that the ℓ_q norm minimization is the maximum likelihood estimate of the generalized Gaussian distribution with shape parameter q $(q > 0)$ [21, 44, 59], which has served as the statistical distribution of image wavelet coefficients in high frequency subbands [53], or used to model the prediction errors in deep learning based speech enhancement [17]. If $\{G_i\}_{i\in \mathrm{J}}$ is the identity operator, (1.1) with $q \in [1, +\infty]$ is reduced to the model in [63]. Meanwhile, the energy of the transform coefficients (like discrete derivatives) of x is penalized in the regularization term and the potential functions ϕ considered in this paper are general nonconvex and non-Lipschitz ones.

Nonconvex and nonsmooth regularizers have the advantages in finding sparse solutions than convex ones and they can help to preserve or generate neat edges in restored images [19, 34, 56, 58]. This very useful edge property of nonconvex and nonsmooth regularization in image restoration is due to the lower bound theory, which provides a uniform lower bound of nonzero entries or image differences of local minimizers. One question we are interested in about model (1.1) is whether it satisfies the lower bound theory. The other crucial question we are concerned about is how to design efficient and convergent algorithms for solving this model. Compared to convex minimization models, the difficulty for solving model (1.1) arises mainly from the possible nonsmoothness of the fidelity term, the nonconvex and nonsmooth property of the regularization term, as well as the composition of ϕ and G_i's, which makes the model a composite optimization problem. In the following two paragraphs, we review some existing works related to these two aspects.

The lower bound theory has been studied in the literature; see, e.g., [23, 25, 31, 32, 36, 54, 56, 57, 70–72]. In general, the key to derive the lower bound theory is to apply the inequality derived from the optimality condition to some carefully constructed testing vectors. For the model with an anisotropic regularizer as given in (1.1), only the case of a squared ℓ_2 norm fidelity term has been considered; see the pioneering work [56], or [23] where box constraints were added to the minimization model. The technique in [56] and [23] uses the second order optimality condition, which cannot be applied to model (1.1) with other fidelity terms. On the other hand, [71] established the lower bound theory of image restoration model with non-Lipschitz isotropic regularization and quadratic fidelity, which was extended to other restoration or decomposition models [36, 67, 70] with non-Lipschitz isotropic regularization and different fidelities in the single- or multiple-channel case. This theory for nonconvex and nonsmooth isotropic models with box constraints was also established in [72] and [31], the latter of which focuses on the case of ℓ_0 "norm" potential function. The derivation in [31] is based on the gradient based image representation and the property of ℓ_0 "norm", while [36, 70–72] mainly used the conservation of image gradient fields. The sparsifying system $\{G_i\}_{i\in \mathrm{J}}$ in all these isotropic regularization cases [31, 36, 70–72] is limited to be the discrete gradient operators (first order differences). At the same time, the discussion is restricted to the case that the regularization quantity corresponding to each pixel is only one.

These limitations make such techniques difficult to be generalized to an anisotropic regularization with an arbitrary sparsifying system. Also, the existing construction of testing vectors is either only suitable for the second order optimality condition, or restricted to the case of the isotropic first order regularization. Indeed, the lower bound theory of the anisotropic model (1.1) with non-quadratic fidelity terms has not been studied in existing works. We mention that, still using the second optimality condition, a very recent work [54] reported the lower bound theory for nonlocal nonconvex models with quadratic fidelity solved by split Bregman iteration [39].

For nonconvex and nonsmooth composite minimization models, the other crucial problem is how to design effective and convergent algorithms to solve them. One straightforward approach is to utilize variable splitting and alternating direction method of multipliers (ADMM) [37] or equivalently, split Bregman iteration [39], but their convergence for nonconvex optimization problems is only guaranteed under specific assumptions like surjectiveness on the linear operator [43, 49, 65], which do not hold in 2D image restoration models with gradient or higher order sparsifying systems. In the existing studies, one popular class of approaches are smoothing approximation methods, e.g. [7,18,20,24,42], in which smoothing functions with auxiliary parameters are utilized to approximate the original objective function. These include the smoothing gradient method [20], the smoothing quadratic regularization (SQR) method [7], the smoothing trust region Newton method [24] and the R-regularized Newton scheme [42] both for twice differentiable fidelity terms, the half-quadratic technique based method [18] for specific potential function $\phi(t)=t^p$ with $0<p,q\leq 2$. In such methods, it was usually only proved that the iterative sequence converges to a stationary point of the smoothed variational model, or satisfies the subsequence convergence result by letting the smoothing parameter tend to zero. Another class of approaches is the iteratively reweighted methods, which concentrate on the special case of $\phi(t)=t^p$. Under the assumption that $0<p,q\leq 2$, the iterative reweighted norm (IRN) algorithm [61] and the generalized Krylov subspace method for ℓ_p-ℓ_q minimization (GKSpq, [48]) are proposed, where the ℓ_p (or ℓ_q) term is approximated by a weighted ℓ_2 norm with iteratively updated weighting matrices. However, the convergence analysis of iterate sequences in [61] and [48] only focuses on the case of $1\leq p,q\leq 2$, which is not applicable to the more interesting case with nonconvex regularization terms. The last approach we mention here is a support shrinking strategy derived respectively for various signal recovery problems [32,50,63] and different image restoration models with isotropic first order regularization [36,69,70,73], yielding some (two-loop) iterative algorithms with either exact or inexact inner loops, when incorporated with the iteratively reweighted ℓ_1 [16,34] or least squares [30, 47]. During a very recent study [75] on nonconvex wavelet image restoration, we got aware that a similar support shrinking strategy was also induced from a majorization-minimization framework for non-Lipschitz synthesis wavelet model [33], but without convergence result given in [33]. In such works with convergence results, the proof techniques for non-Lipschitz composite optimization problems all use the conservation of image gradient fields, which cannot be used for composite models with the anisotropic regularization by a general sparsifying system. To our best knowledge, existing algorithms for anisotropic composite minimization models have no proved global sequence convergence to a stationary point of the original objective function.

As can be seen, the research on the lower bound theory and related algorithm study for nonconvex anisotropic regularization models is very limited. Also, existing related derivation techniques cannot be applied to analyze anisotropic regularization models with non-quadratic fidelities, or iterative algorithms solving a general anisotropic

model. In this paper, we aim to analyze and solve the general non-Lipschitz anisotropic regularization model (1.1). The key technique we introduce is a new construction of the testing vectors, based on which, the first order optimality condition can be applied to prove a lower bound theory and to analyze the algorithm in the situation of anisotropic regularization. The main contributions of this paper can be summarized as follows.

(1) We show a lower bound theory for the composite minimization model (1.1) with an ℓ_1 or ℓ_∞ data fitting term, which is useful for impulse noise or uniform noise (quantization error) removal. For the general case with $1 \le q \le \infty$, a support inclusion analysis is provided. Such theoretical results apply to anisotropic regularization by a general sparsifying system. They help us not only understand more about the model, but also to design iterative algorithms.

(2) Motivated by the support inclusion analysis, we extend previous works to propose an inexact version of an iterative thresholding and support shrinking algorithm with proximal linearization to solve model (1.1). Being more consistent with real computation, the algorithm thresholds at each iteration the transform coefficients to determine the support, and constructs a strongly convex subproblem. This makes the algorithm more practical and computationally efficient. The sequence convergence of the generated iterates is also established, provided that the subgradient condition is satisfied.

The remainder of this paper is organized as follows. In Section 2, we investigate the lower bound theory and support inclusion property of the non-Lipschitz anisotropic model (1.1). Based on these conclusions, in Section 3, we introduce an iterative algorithm with the strategy of thresholding and support shrinking. Then we consider an inexact version of the algorithm and establish its convergence result. We present the details of algorithm implementation in Section 4. In Section 5, the experiments on image deconvolution and two-stage image segmentation are carried out to exhibit the usefulness of the proposed algorithm. We conclude this paper in Section 6.

2. Lower bound theory and support inclusion analysis

We now give some notations, which will be used throughout this paper. For a set S, let $\sharp$S or |S| denote its cardinality. For a matrix B, we use B_i to denote the i'th column of B, use $B^\top$ and $B_i^\top$ to denote the transposes of B and B_i, respectively. For a vector $x \in \mathbb{R}^N$, x_l is the l'th entry of x, $1 \le l \le N$. Given $x \in \mathbb{R}^N$, the ℓ_p (quasi-)norm $\|x\|_p$ with $p \in (0, +\infty)$ is defined by $\|x\|_p = \left(\sum_{1\le l\le N} |x_l|^p\right)^{1/p}$ and the infinity norm is defined by $\|x\|_\infty = \max\limits_{1\le l\le N}\{|x_l|\}$. Sometimes, we abbreviate $\|\cdot\|_2$ as $\|\cdot\|$ as usual. We use I_N to denote the $N \times N$ identity matrix. We denote the set of indices corresponding to nonzero sparsifying coefficients $G_i^\top x$ of x as the coefficient support (or, abbreviated as support) S(x), i.e., $\mathrm{S}(x) = \{i \in \mathrm{J} : G_i^\top x \ne 0\}$. The set of indices corresponding to differences $G_i^\top x$ whose absolute values are larger than τ ($\tau \ge 0$), is called the coefficient τ-support (or, abbreviated as τ-support) of x and written as $\mathrm{T}(x) = \{i \in \mathrm{J} : |G_i^\top x| > \tau\}$. Specifically, if taking $\tau = 0$, T(x) degenerates to S(x).

For the convenience of description, we denote

$$\mathcal{F}(x) := \sum_{i\in \mathrm{J}} \phi(|G_i^\top x|) + \mathcal{Q}_q(x) \tag{2.1}$$

for our considered model (1.1), and rewrite it as

$$(\mathfrak{F}) \quad \min_{x\in\mathbb{R}^N} \mathcal{F}(x). \tag{2.2}$$

The potential function ϕ is supposed to satisfy the following assumptions.

ASSUMPTION 2.1.

(a) $\phi:[0,+\infty)\to[0,+\infty)$ *is a continuous concave coercive function, and* $\phi(0)=0$.

(b) ϕ *is* C^1 *on* $(0,+\infty)$, $\phi'(t)|_{(0,+\infty)}>0$ *and* $\phi'(0+)=+\infty$.

(c) *For any* $\alpha>0$, ϕ' *is* L_α*-Lipschitz continuous on* $[\alpha,+\infty)$, *i.e., there exists a constant* L_α *determined by* α, *such that for any* $t,s\in[\alpha,+\infty)$, $|\phi'(t)-\phi'(s)|\le L_\alpha|t-s|$.

EXAMPLE 2.1. *Two examples of potential functions satisfying Assumption 2.1 are* $\phi_1(t)=t^p$ $(0<p<1)$ *and* $\phi_2(t)=\log(1+t^p)$ $(0<p<1)$ *[69, 73].*

By Assumption 2.1, the subdifferential of $\phi(|t|)$ at t is

$$\partial\phi\circ|\cdot|(t)=\begin{cases}(-\infty,+\infty), & \text{if } t=0,\\ \left\{\frac{t}{|t|}\phi'(|t|)\right\}, & \text{otherwise,}\end{cases}$$

and $\phi(|t|)$ is subdifferentially regular [62] at any $t\in\mathbb{R}$ [50].

For the fidelity term, $\mathcal{Q}_q(x)$ is convex, subdifferentially regular at any $x\in\mathbb{R}^N$ and its subdifferential is given by

$$\partial\mathcal{Q}_q(x)=\begin{cases}\frac{\beta}{q}A^\top\partial\|\cdot\|_q^q(Ax-b) & q\in[1,+\infty)\\ \beta A^\top\partial\|\cdot\|_\infty(Ax-b) & q=+\infty\end{cases}, \tag{2.3}$$

where the subdifferential of the infinity norm is as follows [6]

$$\partial\|h\|_\infty=\{d\in\mathbb{R}^M\,|\,\|d\|_1\le 1, h^\top d=\|h\|_\infty\}.$$

Throughout this paper, A and G are assumed to satisfy the following basic property, which is trivial in image restoration problems [64, 68].

ASSUMPTION 2.2. $\ker A\cap\ker G^\top=\{0\}$.

Following the arguments in existing works, e.g., [73] and [52], one can show the coercive property of the objective function $\mathcal{F}(x)$.

THEOREM 2.1. *Suppose that Assumption 2.1 and Assumption 2.2 hold true. Then the function* $\mathcal{F}(x)$ *in* (2.1) *is coercive and thus* (2.2) *has at least one solution.*

In this section, we focus on discussing the lower bound theory and related support inclusion analysis for the model (1.1), i.e., the minimization problem (2.2). In [56], the authors studied the lower bound theory for a special case of (2.2) with $q=2$, i.e.,

$$\min_{x\in\mathbb{R}^N}\sum_{i\in\mathrm{J}}\phi(|G_i^\top x|)+\frac{\beta}{2}\|Ax-b\|_2^2. \tag{2.4}$$

To establish such a conclusion, the second order derivative of ϕ is required. We revise the conclusion in [56] to Theorem 2.2 below, in which the requirement on ϕ is modified to be consistent with Assumption 2.1.

THEOREM 2.2. *Let* ϕ *be the potential function satisfying Assumption 2.1 and the following:* ϕ *is* C^2 *on* $(0,+\infty)$, ϕ'' *is increasing on* $(0,+\infty)$ *with* $\phi''(t)\le 0$ *for any* $t>0$, $\phi''(0^+)=-\infty$ *and* $\lim_{t\to+\infty}\phi''(t)=0$. *Then there exists a constant* $\theta_2>0$, *such that for any local minimizer* x^* *of model* (2.4) *(i.e., the problem* (2.2) *with* $q=2$*), the difference* $G_i^\top x^*$ *is either zero or satisfies* $|G_i^\top x^*|\ge\theta_2$.

Proof. The proof is similar to that in [56]. ∎

We here mention that a related theorem to Theorem 2.2 was reported very recently in [54], which is a nonlocal version of the lower bound theory in [56].

Theorem 2.2 is for the case of Gaussian measurement noise. Another common and important situation is the impulse noise removal, and the corresponding model is as follows,

$$\min_{x\in\mathbb{R}^N}\sum_{i\in \mathrm{J}}\phi(|G_i^\top x|)+\beta\|Ax-b\|_1, \tag{2.5}$$

which is the $q=1$ case of (2.2). The key technique for proving Theorem 2.2 is the second order necessary condition on local minimizers with the corresponding testing vector construction [56], which cannot be applied to deal with the model (2.5). For the model (2.5), we will instead establish a lower bound theory on the stationary points (including local minimizers) by using the first order optimality condition with a new construction of testing vector.

THEOREM 2.3. *There exists a constant $\theta_1>0$, such that for any stationary point x^* of (2.5) (i.e., the problem (2.2) with $q=1$), it satisfies*

$$\text{either } G_i^\top x^*=0 \text{ or } |G_i^\top x^*|\ge\theta_1,\quad \forall i\in \mathrm{J}.$$

Proof. Suppose that x^* is a stationary point of (2.5). Define

$$\mathrm{S}^*=\mathrm{S}(x^*)=\left\{i\in \mathrm{J}: G_i^\top x^*\neq 0\right\} \text{ and } C(\mathrm{S}^*)=\left\{x\in\mathbb{R}^N:\ G_i^\top x=0\ \forall i\in \mathrm{J}\backslash \mathrm{S}^*\right\}.$$

Without loss of generality, we consider the case $\mathrm{S}^*\neq\emptyset$. By a similar calculation to [73, Theorem 2.3], we have $\partial\left(\sum_{i\in \mathrm{J}\backslash \mathrm{S}^*}\phi(|G_i^\top x^*|)\right)=\sum_{i\in \mathrm{J}\backslash \mathrm{S}^*}(\ker G_i^\top)^\perp$ and that $\sum_{i\in \mathrm{J}\backslash \mathrm{S}^*}\phi(|G_i^\top x|)$ is subdifferentially regular at x^*. By Corollary 10.9 of [62], we derive

$$\begin{aligned}\partial\mathcal{F}(x^*)=&\partial(\beta\|Ax^*-b\|_1)+\sum_{i\in \mathrm{S}^*}\phi'(|G_i^\top x^*|)\frac{G_i^\top x^*}{|G_i^\top x^*|}G_i+\partial\left(\sum_{i\in \mathrm{J}\backslash \mathrm{S}^*}\phi(|G_i^\top x^*|)\right)\\ =&\beta A^\top\partial\|\cdot\|_1(Ax^*-b)+\sum_{i\in \mathrm{S}^*}\phi'(|G_i^\top x^*|)\frac{G_i^\top x^*}{|G_i^\top x^*|}G_i+\sum_{i\in \mathrm{J}\backslash \mathrm{S}^*}(\ker G_i^\top)^\perp.\end{aligned}$$

Since x^* is a stationary point of $\mathcal{F}(x)$, $0\in\partial\mathcal{F}(x^*)$ and there exists $d\in\partial\|\cdot\|_1(Ax^*-b)$, such that for any $v\in C(\mathrm{S}^*)$,

$$\sum_{i\in \mathrm{S}^*}\phi'(|G_i^\top x^*|)\frac{G_i^\top x^*}{|G_i^\top x^*|}G_i^\top v=-\beta d^\top Av\le\beta\|d\|_2\|A\|_2\|v\|_2\le\alpha\|v\|_2, \tag{2.6}$$

where $\alpha=\beta\sqrt{M}\|A\|_2$.

Define $\mathrm{S}^*_{++}=\left\{i\in \mathrm{S}^*: G_i^\top x^*>0\right\}\subset \mathrm{S}^*$. We consider a fixed $j\in \mathrm{S}^*$, and define a closed set

$$V(\mathrm{S}^*,j,\mathrm{S}^*_{++})=C(\mathrm{S}^*)\cap\left\{v: G_i^\top v\ge 0\ \forall i\in \mathrm{S}^*_{++},\quad G_i^\top v\le 0\ \forall i\in \mathrm{S}^*\backslash \mathrm{S}^*_{++}\right\}\cap\left\{v: |G_j^\top v|=1\right\}$$

which is nonempty, since $v=\frac{x^*}{|G_j^\top x^*|}\in V(\mathrm{S}^*,j,\mathrm{S}^*_{++})$. Let $\tilde v=\tilde v(\mathrm{S}^*,j,\mathrm{S}^*_{++})$ be the solution to the following problem

$$\min\|v\|_2\quad \text{subject to } v\in V(\mathrm{S}^*,j,\mathrm{S}^*_{++}).$$

Then it implies that: $\tilde{v}\in C(\mathrm{S}^*)$, and for any $i\in\mathrm{S}^*$, $G_i^\top x^*\cdot G_i^\top\tilde{v}\geq 0$, and $\frac{G_j^\top x^*}{|G_j^\top x^*|}G_j^\top\tilde{v}=1$.

Next, we derive a uniform upper bound of $\|\tilde{v}\|_2$ for any stationary point x^*. Similar to the definition of $C(\mathrm{S}^*)$ and $V(\mathrm{S}^*,j,\mathrm{S}^*_{++})$, for any set $\mathrm{I}\subset\mathrm{J}$, define

$$C(\mathrm{I})=\left\{u\in\mathbb{R}^N:\ G_i^\top u=0\ \forall i\in\mathrm{J}\backslash\mathrm{I}\right\},$$

and for any subset $\mathrm{I}_{++}\subset\mathrm{I}$ and index $i'\in\mathrm{I}$, define

$$V(\mathrm{I},i',\mathrm{I}_{++})=C(\mathrm{I})\cap\{v:G_i^\top v\geq 0\ \forall i\in\mathrm{I}_{++},\quad G_i^\top v\leq 0\ \forall i\in\mathrm{I}\backslash\mathrm{I}_{++}\}\cap\{v:|G_{i'}^\top v|=1\}.$$

For those nonempty $V(\mathrm{I},i',\mathrm{I}_{++})$, let $\tilde{v}(\mathrm{I},i',\mathrm{I}_{++})$ be the solution to the following constrained optimization problem

$$\min\|v\|_2\quad\text{ subject to }v\in V(\mathrm{I},i',\mathrm{I}_{++}).$$

Denote $\mu(\mathrm{I})=\max\limits_{i'\in\mathrm{I},\mathrm{I}_{++}\subset\mathrm{I}}\|\tilde{v}(\mathrm{I},i',\mathrm{I}_{++})\|_2$ and let

$$\mu=\max_{\mathrm{I}\subset\mathrm{J}}\{\mu(\mathrm{I})\},\tag{2.7}$$

where μ is well-defined and positive. Then μ is a uniform upper bound of $\|\tilde{v}\|_2$, i.e., $\|\tilde{v}\|_2=\|\tilde{v}(\mathrm{S}^*,j,\mathrm{S}^*_{++})\|_2\leq\mu$ for any stationary point x^*.

By the assumption on function $\phi(\cdot)$ and the construction of $\tilde{v}$, we derive from (2.6) that

$$\phi'(|G_j^\top x^*|)=\phi'(|G_j^\top x^*|)\frac{G_j^\top x^*}{|G_j^\top x^*|}G_j^\top\tilde{v}\leq\sum_{i\in\mathrm{S}^*}\phi'(|G_i^\top x^*|)\frac{G_i^\top x^*}{|G_i^\top x^*|}G_i^\top\tilde{v}\leq\alpha\|\tilde{v}\|_2\leq\alpha\mu.$$

Since $\phi'(0+)=+\infty$, the constant $\theta_1=\inf\{t>0:\phi'(t)\leq\alpha\mu\}>0$ is well defined and independent of x^*. Thus, we obtain that

$$|G_j^\top x^*|\geq\theta_1,$$

which holds for any $j\in\mathrm{S}^*$ and any stationary point x^*. ∎

For the removal of uniformly distributed noise (or quantization error), the corresponding model is as follows,

$$\min_{x\in\mathbb{R}^N}\sum_{i\in\mathrm{J}}\phi(|G_i^\top x|)+\beta\|Ax-b\|_\infty,\tag{2.8}$$

which is the $q=+\infty$ case of (2.2). For the model (2.8), a lower bound theory similar to Theorem 2.3 holds.

THEOREM 2.4. *There exists a constant $\theta_\infty>0$, such that for any stationary point x^* of (2.8) (i.e., the problem (2.2) with $q=+\infty$), it satisfies*

$$\textit{either } G_i^\top x^*=0\ \textit{ or }\ |G_i^\top x^*|\geq\theta_\infty,\quad\forall i\in\mathrm{J}.$$

Proof. By (2.3), $\partial\mathcal{Q}_q$ is uniformly bounded when $q=+\infty$. Then the remaining proof is similar to Theorem 2.3. ∎

REMARK 2.1. By Theorem 2.2, the lower bound theory holds for the local minimizers of the objective function $\mathcal{F}(x)$ in (2.1), in the case of squared ℓ_2 norm fidelity, i.e.,

$q=2$. By Theorem 2.3 and Theorem 2.4, in the cases of ℓ_1 norm and ℓ_∞ norm fidelity terms, i.e., $q=1$ and $q=+\infty$, the lower bound theory can hold for all stationary points (including the local minimizers) of $\mathcal{F}(x)$.

By Theorem 2.2 (or Theorem 2.3, Theorem 2.4), there exists some constant θ, such that for any local minimizer x^* of $\mathcal{F}(x)$ with $q=2$ (or any stationary point x^* of $\mathcal{F}(x)$ with $q=1,+\infty$), it satisfies

$$\text{either } G_i^\top x^*=0 \text{ or } |G_i^\top x^*|\geq\theta.$$

Suppose that x^* is very near to a given point $\tilde{x}$, such that $|G_i^\top x^*-G_i^\top\tilde{x}|\leq\|G_i\|_2\|x^*-\tilde{x}\|_2<\theta,\ \forall i\in\mathrm{J}$. Then $|G_i^\top x^*|\leq|G_i^\top x^*-G_i^\top\tilde{x}|+|G_i^\top\tilde{x}|<\theta+|G_i^\top\tilde{x}|$, which implies *the support inclusion property*, i.e., for $i\in\mathrm{J}$,

$$G_i^\top x^*=0,\quad \text{when } G_i^\top\tilde{x}=0.$$

In fact, although there is so far no lower bound theory for model (2.2) with $q\in[1,+\infty]$ and $q\neq1,2,+\infty$, the support inclusion property can be shown to hold for the stationary points of model (2.2) with any $q\in[1,+\infty]$ as stated in Theorem 2.5.

THEOREM 2.5. *Consider* (2.2) *with a fixed* $q\in[1,+\infty]$. *Given* $\tilde{x}\in\mathbb{R}^N$, *assume that a stationary point* $\hat{x}$ *of* (2.2) *is sufficiently close to* $\tilde{x}$. *Then for* $i\in\mathrm{J}$,

$$G_i^\top\hat{x}=0,\quad \textit{when } G_i^\top\tilde{x}=0.$$

Proof. Suppose that for some $j\in\mathrm{J}$, $G_j^\top\tilde{x}=0$ but $G_j^\top\hat{x}\neq0$. Define nonempty sets $\hat{\mathrm{S}}=\mathrm{S}(\hat{x})=\{i\in\mathrm{J}:G_i^\top\hat{x}\neq0\}$ and $C(\hat{\mathrm{S}})=\left\{x\in\mathbb{R}^N:\ G_i^\top x=0\ \forall i\in\mathrm{J}\backslash\hat{\mathrm{S}}\right\}$. By Corollary 10.9 of [62], we derive

$$\begin{aligned}\partial\mathcal{F}(\hat{x})&=\partial\mathcal{Q}_q(\hat{x})+\sum_{i\in\hat{\mathrm{S}}}\phi'(|G_i^\top\hat{x}|)\frac{G_i^\top\hat{x}}{|G_i^\top\hat{x}|}G_i+\partial\left(\sum_{i\in\mathrm{J}\backslash\hat{\mathrm{S}}}\phi(|G_i^\top\hat{x}|)\right)\\&=\begin{cases}\frac{\beta}{q}A^\top\partial\|\cdot\|_q^q(A\hat{x}-b)+\sum_{i\in\hat{\mathrm{S}}}\phi'(|G_i^\top\hat{x}|)\frac{G_i^\top\hat{x}}{|G_i^\top\hat{x}|}G_i+\sum_{i\in\mathrm{J}\backslash\hat{\mathrm{S}}}(\ker G_i^\top)^\perp, & q\in[1,+\infty),\\ \beta A^\top\partial\|\cdot\|_\infty(A\hat{x}-b)+\sum_{i\in\hat{\mathrm{S}}}\phi'(|G_i^\top\hat{x}|)\frac{G_i^\top\hat{x}}{|G_i^\top\hat{x}|}G_i+\sum_{i\in\mathrm{J}\backslash\hat{\mathrm{S}}}(\ker G_i^\top)^\perp, & q=+\infty.\end{cases}\end{aligned}$$

Since $\hat{x}$ is a stationary point of (2.2), $0\in\partial\mathcal{F}(\hat{x})$ and for any $v\in C(\hat{\mathrm{S}})$,

$$\sum_{i\in\hat{\mathrm{S}}}\phi'(|G_i^\top\hat{x}|)\frac{G_i^\top\hat{x}}{|G_i^\top\hat{x}|}G_i^\top v\leq\begin{cases}\beta\|A\hat{x}-b\|_{2q-2}^{q-1}\|A\|_2\|v\|_2, & q\in(1,+\infty)\\ \beta\sqrt{M}\|A\|_2\|v\|_2, & q=1\\ \beta\|A\|_2\|v\|_2, & q=+\infty\end{cases}.$$

Define

$$\alpha_1(\tilde{x})=\begin{cases}\|A\tilde{x}-b\|_{2q-2}^{q-1}+1, & q\in(1,+\infty),\\ \sqrt{M}, & q=1,\\ 1, & q=+\infty.\end{cases}$$

Then there exists $\delta>0$ such that if $\|\hat{x}-\tilde{x}\|_2<\delta$,

$$\sum_{i\in\hat{\mathrm{S}}}\phi'(|G_i^\top\hat{x}|)\frac{G_i^\top\hat{x}}{|G_i^\top\hat{x}|}G_i^\top v\leq\alpha_1(\tilde{x})\beta\|A\|_2\|v\|_2.\tag{2.9}$$

By using similar notations in Theorem 2.3, we denote

$$\hat{\mathrm{S}}_{++}=\{i\in\hat{\mathrm{S}}:G_i^\top\hat{x}>0\}\subset\hat{\mathrm{S}}$$

and

$$V(\hat{\mathrm{S}},j,\hat{\mathrm{S}}_{++})=C(\hat{\mathrm{S}})\cap\{v:G_i^\top v\geq 0\ \forall i\in\hat{\mathrm{S}}_{++},\quad G_i^\top v\leq 0\ \forall i\in\hat{\mathrm{S}}\backslash\hat{\mathrm{S}}_{++}\}\cap\{v:|G_j^\top v|=1\}.$$

Since $v=\frac{\hat{x}}{|G_j^\top\hat{x}|}\in V(\hat{\mathrm{S}},j,\hat{\mathrm{S}}_{++})$, $V(\hat{\mathrm{S}},j,\hat{\mathrm{S}}_{++})\neq\emptyset$. Let $\hat{v}=\hat{v}(\hat{\mathrm{S}},j,\hat{\mathrm{S}}_{++})$ denote the solution to the following problem

$$\min\|v\|_2\quad\text{subject to } v\in V(\hat{\mathrm{S}},j,\hat{\mathrm{S}}_{++}),$$

which satisfies $\|\hat{v}\|_2=\|\hat{v}(\hat{\mathrm{S}},j,\hat{\mathrm{S}}_{++})\|_2\leq\mu$ with μ as defined in (2.7). Then, by (2.9),

$$\begin{aligned}\phi'(|G_j^\top\hat{x}|)=\phi'(|G_j^\top\hat{x}|)\frac{G_j^\top\hat{x}}{|G_j^\top\hat{x}|}G_j^\top\hat{v}\leq\sum_{i\in\hat{\mathrm{S}}}\phi'(|G_i^\top\hat{x}|)\frac{G_i^\top\hat{x}}{|G_i^\top\hat{x}|}G_i^\top\hat{v}\leq\alpha_1(\tilde{x})\beta\|A\|_2\|v\|_2\\ \leq\alpha_2(\tilde{x}),\end{aligned}\tag{2.10}$$

where $\alpha_2(\tilde{x})=\alpha_1(\tilde{x})\beta\mu\|A\|_2$. Since $\phi'(0+)=+\infty$, the constant $\tilde{\theta}=\inf\{t>0:\phi'(t)\leq\alpha_2(\tilde{x})\}$ is well defined. If $\hat{x}$ is sufficiently close to $\tilde{x}$ such that $\|\hat{x}-\tilde{x}\|<\min\{\frac{\tilde{\theta}}{\max_{i\in\mathrm{J}}\|G_i\|},\delta\}$, we have $|G_j^\top\hat{x}|\leq|G_j^\top\hat{x}-G_j^\top\tilde{x}|+|G_j^\top\tilde{x}|\leq\|G_j\|\|\hat{x}-\tilde{x}\|<\tilde{\theta}$ and $\phi'(|G_j^\top\hat{x}|)>\alpha_2(\tilde{x})$, which contradicts with (2.10). Thus, for any $j\in\mathrm{J}$ with $G_j^\top\tilde{x}=0$, we have $G_j^\top\hat{x}=0$. ∎

Note that similar support inclusion analysis to Theorem 2.5 was derived for different sparse signal recovery problems ([50, 63]) and different isotropic image restoration or decomposition models ([36, 67, 69, 71, 73]), etc.

3. Algorithms and convergence analysis

The theoretical results in the previous section not only exhibit some interesting model properties, but also help us to design iterative algorithms and perform convergence analysis for the non-Lipschitz composite minimization model (2.2).

3.1. Algorithms. We now derive an iterative algorithm to solve the composite minimization model (2.2). In order to find a local minimizer or stationary point near to a given point, the support inclusion analysis in Theorem 2.5 naturally motivates a so-called support shrinking strategy at each iteration, like [50, 63, 69, 70, 73] for various sparse signal reconstruction and image restoration models with isotropic regularization. Given x^k, we denote $\mathrm{S}^k=\mathrm{S}(x^k)=\{i\in\mathrm{J}:G_i^\top x^k\neq 0\}$, and define $C_k=\{x:G_i^\top x=0,\ \forall i\in\mathrm{J}\backslash\mathrm{S}^k\}$. Then for $k=0,1,2,\ldots$, the following

$$\begin{cases}\min\limits_{x\in\mathbb{R}^N}\left\{\mathcal{F}_k(x):=\sum\limits_{i\in\mathrm{S}^k}\phi(|G_i^\top x|)+\mathcal{Q}_q(x)\right\},\\ \text{s.t. } x\in C_k,\end{cases}$$

is considered to compute x^{k+1}.

However, considering the finite word length of real-world computers and to avoid extremely large linearization weights described later, we do not track the indices just in S^k or use the constraint $x\in C_k$. Instead, we relax these at each iteration by a thresholding operation, and propose the strategy of iterative thresholding and support shrinking

(ITSS). Indeed, this relaxation is actually usually adopted in computer codes for support shrinking based algorithms, but so far, has been formulated explicitly only in [32] with incorporation into the iteratively reweighted least square (IRLS) algorithm for group sparse signal recovery (where G=Identity). In particular, let us denote the τ-support $\mathrm{T}^k = \mathrm{T}(x^k) = \{i \in \mathrm{J} : |G_i^\top x^k| > \tau\}$ and the set $C_k^\tau = \{x : G_i^\top x = 0,\ \forall i \in \mathrm{J}\backslash\mathrm{T}^k\}$ accordingly, for some nonnegative τ. At the k'th iteration, the coefficients $G_i^\top x^{k+1}$ with $i \in \mathrm{J}\backslash\mathrm{T}^k$, instead of $i \in \mathrm{J}\backslash\mathrm{S}^k$, are constrained to be zero. That is, we construct the following

$$(\mathfrak{F}_k^\tau) \quad \begin{cases} \min\limits_{x\in\mathbb{R}^N} \left\{ \mathcal{F}_k^\tau(x) := \sum\limits_{i\in\mathrm{T}^k} \phi(|G_i^\top x|) + \mathcal{Q}_q(x) \right\}, \\ \text{s.t. } x \in C_k^\tau. \end{cases}$$

This problem is still nonconvex and difficult. We however can linearize the ϕ at $|G_i^\top x^k|$ with $i \in \mathrm{T}^k$. With a proximal term, we define a strongly convex function

$$\mathcal{H}_k^\tau(x) := \sum_{i\in\mathrm{T}^k} \left\{ \phi(|G_i^\top x^k|) + \phi'(|G_i^\top x^k|) \left(|G_i^\top x| - |G_i^\top x^k| \right) \right\} + \mathcal{Q}_q(x) + \frac{\rho}{2}\|x - x^k\|^2,$$

and consider the following optimization problem

$$(\mathfrak{H}_k^\tau) \begin{cases} \min\limits_{x\in\mathbb{R}^N} \mathcal{H}_k^\tau(x), \\ \text{s.t. } x \in C_k^\tau, \end{cases} \tag{3.1}$$

for computing x^{k+1}. Obviously $\mathcal{H}_k^\tau(x)$ in (3.1) has a unique optimal solution. Thus, we have derived the *Iterative thresholding and support shrinking algorithm with proximal linearization (ITSS-PL)* as stated in Algorithm 1, which is with iteratively reweighted ℓ_1 flavor.

Algorithm 1 Iterative thresholding and support shrinking algorithm with proximal linearization (ITSS-PL)

Require: A, b, $\beta > 0$, $\rho > 0$, $\tau \geq 0$, MaxIter, $x^0 \in \mathbb{R}^N$.
  while $k \leq$ MaxIter **do**
    Compute x^{k+1} by solving $(\mathfrak{H}_k^\tau)$;
  end while

Precisely solving the subproblem $(\mathfrak{H}_k^\tau)$ still needs infinite iteration steps, which is not practical in real computation. Therefore, instead of finding the exact minimizer, we solve the subproblem $(\mathfrak{H}_k^\tau)$ in Algorithm 1 to some given accuracy, like [4, 50, 63, 73, 74]. For any set C, recall the indicator function $\mathcal{I}_C$ as

$$\mathcal{I}_C(x) = \begin{cases} 0 & x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

Then we come to the *inexact iterative thresholding and support shrinking algorithm with proximal linearization (Inexact ITSS-PL)*, as given in Algorithm 2. We mention that *Inexact ITSS-PL* degenerates to *ITSS-PL*, if $\epsilon = 0$.

3.2. Convergence analysis. In what follows, we prove that the sequence generated by Algorithm 2 with the inexact inner loop does converge to a stationary point of the original minimization model $(\mathfrak{F})$ in (2.2), provided that the stopping condition in (3.2) is satisfied. The proof is based on the Kurdyka-Łojasiewicz (KL) property [46,51], which has been extensively applied to the analysis of various optimization methods [1–4,8,32,50,69,73,74], etc.

Algorithm 2 Inexact iterative thresholding and support shrinking algorithm with proximal linearization (Inexact ITSS-PL)

Require: $A\in\mathbb{R}^{M\times N}$, $b\in\mathbb{R}^M$, $\beta>0$, $\rho>0$, $\tau\geq 0$, $0<\epsilon<1$, MaxIter, $x^0\in\mathbb{R}^N$.

while $k\leq$ MaxIter **do**

Compute x^{k+1} by

$$\begin{cases} x^{k+1}\approx\arg\min\limits_{x\in\mathbb{R}^N}\mathcal{H}_k^\tau(x)+\mathcal{I}_{C_k^\tau}(x) \text{ with } h^{k+1}\in\partial\left(\mathcal{H}_k^\tau(x^{k+1})+\mathcal{I}_{C_k^\tau}(x^{k+1})\right),\\ \text{s.t. } \|h^{k+1}\|_2\leq\dfrac{\rho}{2}\epsilon\|x^{k+1}-x^k\|_2; \end{cases} \tag{3.2}$$

end while

Note that

$$\partial\mathcal{I}_{C_k^\tau}(x)=\begin{cases}\sum\limits_{i\in\mathrm{J}\setminus\mathrm{T}^k}(\ker G_i^\top)^\perp, & x\in C_k^\tau,\\ \emptyset, & \text{otherwise.}\end{cases}$$

By (3.2), we then have

$$G_i^\top x^{k+1}=0,\quad i\in\mathrm{J}\setminus\mathrm{T}^k,$$

which implies $\mathrm{J}\setminus\mathrm{S}^k\subset\mathrm{J}\setminus\mathrm{T}^k\subset\mathrm{J}\setminus\mathrm{S}^{k+1}\subset\mathrm{J}\setminus\mathrm{T}^{k+1}$ and $\mathrm{T}^{k+1}\subset\mathrm{S}^{k+1}\subset\mathrm{T}^k\subset\mathrm{S}^k\subset\cdots\subset\mathrm{J}$. Due to the finiteness of the set J, it is also straightforward to see the finite convergence of both support and τ-support sequences S^k and T^k, like the finite convergence of the support sequence mentioned in the literature (e.g. [32,50,69,73]). That is to say, there exists an integer $K>0$, such that

$$\mathrm{S}^k=\mathrm{T}^k=\mathrm{S}^K\quad\text{for any } k\geq K. \tag{3.3}$$

Therefore, the sets S^k and T^k will be unchanged when $k\geq K$. In the following, to prove the convergence, we only need to focus on the iterates with $k\geq K$. We denote

$$\bar{\mathrm{S}}=\bar{\mathrm{T}}=\mathrm{S}^k=\mathrm{T}^k=\{i\in\mathrm{J}:G_i^\top x^k\neq 0\}=\{i\in\mathrm{J}:|G_i^\top x^k|>\tau\}\quad\text{for } k\geq K, \tag{3.4}$$

and

$$\bar{\mathcal{H}}_k^\tau(x):=\sum_{i\in\bar{\mathrm{T}}}\left\{\phi(|G_i^\top x^k|)+\phi'(|G_i^\top x^k|)\left(|G_i^\top x|-|G_i^\top x^k|\right)\right\}+\mathcal{Q}_q(x)+\frac{\rho}{2}\|x-x^k\|^2.$$

Then for $k\geq K$, the subproblem at the k'th step in Algorithm 2 becomes

$$\begin{cases} x^{k+1}\approx\arg\min\limits_{x\in\mathbb{R}^N}\bar{\mathcal{H}}_k^\tau(x)+\mathcal{I}_{\bar{C}}(x) \text{ with } h^{k+1}\in\partial\left(\bar{\mathcal{H}}_k^\tau(x^{k+1})+\mathcal{I}_{\bar{C}}(x^{k+1})\right)\\ \text{s.t. } \|h^{k+1}\|_2\leq\frac{\rho}{2}\epsilon\|x^{k+1}-x^k\|_2 \end{cases}, \tag{3.5}$$

where $\bar{C}=\{x: G_i^\top x=0,\ \forall i\in \mathrm{J}\backslash\bar{\mathrm{T}}\}$.

The next lemma is the decreasing property of the objective function.

LEMMA 3.1 (Sufficient decrease). *The sequence $\{\mathcal{F}(x^k)\}_{k\geq K}$ is nonincreasing, and more precisely, for any $k\geq K$,*

$$(1-\epsilon)\frac{\rho}{2}\|x^{k+1}-x^k\|_2^2\leq \mathcal{F}(x^k)-\mathcal{F}(x^{k+1}).$$

Proof. For any $k\geq K$, by the convexity of function $\bar{\mathcal{H}}_k^\tau(x)+\mathcal{I}_{\bar{C}}(x)$ in (3.5), we have

$$\begin{aligned}
&\bar{\mathcal{H}}_k^\tau(x^{k+1})\\
=&\sum_{i\in\bar{\mathrm{T}}}\left\{\phi(|G_i^\top x^k|)+\phi'(|G_i^\top x^k|)\left(|G_i^\top x^{k+1}|-|G_i^\top x^k|\right)\right\}+\mathcal{Q}_q(x^{k+1})+\frac{\rho}{2}\|x^{k+1}-x^k\|_2^2\\
\leq&\bar{\mathcal{H}}_k^\tau(x^k)-\langle h^{k+1},x^k-x^{k+1}\rangle\\
\leq&\sum_{i\in\bar{\mathrm{T}}}\phi(|G_i^\top x^k|)+\mathcal{Q}_q(x^k)+\frac{\rho}{2}\epsilon\|x^{k+1}-x^k\|_2^2\\
=&\mathcal{F}(x^k)+\frac{\rho}{2}\epsilon\|x^{k+1}-x^k\|_2^2.
\end{aligned}$$

By Assumption 2.1,

$$\phi(t)\leq\phi(\bar{t})+\phi'(\bar{t})(t-\bar{t}),\quad \forall t\geq 0 \text{ and } \bar{t}>0. \tag{3.6}$$

Then by (3.6),

$$\bar{\mathcal{H}}_k^\tau(x^{k+1})\geq\sum_{i\in\bar{\mathrm{T}}}\phi(|G_i^\top x^{k+1}|)+\mathcal{Q}_q(x^{k+1})+\frac{\rho}{2}\|x^{k+1}-x^k\|_2^2=\mathcal{F}(x^{k+1})+\frac{\rho}{2}\|x^{k+1}-x^k\|_2^2.$$

Therefore, we obtain

$$\mathcal{F}(x^k)-\mathcal{F}(x^{k+1})\geq(1-\epsilon)\frac{\rho}{2}\|x^{k+1}-x^k\|_2^2\quad\text{for any } k\geq K.$$

∎

From Lemma 3.1, we see the boundness of the sequence $\{x^k\}$.

LEMMA 3.2 (Square summable and asymptotic convergence). *The sequence $\{x^k\}_{k\geq 0}$ is bounded and satisfies*

$$\sum_{k=0}^{\infty}\|x^{k+1}-x^k\|^2<\infty.$$

Hence $\lim_{k\to\infty}\|x^{k+1}-x^k\|=0$.

Proof. By Lemma 3.1, $\{\mathcal{F}(x^k)\}_{k\geq 0}$ is bounded and convergent. Since $\mathcal{F}(x)$ is coercive (Theorem 2.1), $\{x^k\}_{k\geq 0}$ is bounded. Again by Lemma 3.1, for any $K_1\geq K$, we get

$$\begin{aligned}
(1-\epsilon)\frac{\rho}{2}\sum_{k=0}^{K_1}\|x^{k+1}-x^k\|^2&\leq(1-\epsilon)\frac{\rho}{2}\sum_{k=0}^{K-1}\|x^{k+1}-x^k\|^2+\mathcal{F}(x^K)-\mathcal{F}(x^{K_1+1})\\
&\leq(1-\epsilon)\frac{\rho}{2}\sum_{k=0}^{K-1}\|x^{k+1}-x^k\|^2+\mathcal{F}(x^K).
\end{aligned}$$

Letting $K_1 \to \infty$ completes the proof. ∎

A very useful lemma is the following uniform lower bound result of nonzero transform coefficients of the iterate sequence.

LEMMA 3.3 (Lower bound of the sequence). *There exists a constant $\bar{\theta} > 0$ such that*

$$|G_i^\top x^k| \geq \bar{\theta} \quad \forall k \geq K, \forall i \in \bar{\mathrm{T}}.$$

Proof. Without loss of generality, suppose that $\bar{\mathrm{T}} \neq \emptyset$. By (3.4), a natural and obvious consequence is

$$|G_i^\top x^k| > \tau \geq 0 \quad \forall k \geq K, \forall i \in \bar{\mathrm{T}}.$$

Next, we derive a positive lower bound of nonzero transform coefficients of the iterate sequence, even if $\tau = 0$.

Since $\mathcal{I}_{\bar{C}}(x)$ is subdifferentially regular at $x \in \bar{C}$, we derive

$$\begin{aligned} h^{k+1} &\in \partial\left(\bar{\mathcal{H}}_k^\tau(x^{k+1}) + \mathcal{I}_{\bar{C}}(x^{k+1})\right) \\ &= \partial \mathcal{Q}_q(x^{k+1}) + \sum_{i \in \bar{\mathrm{T}}} \phi'(|G_i^\top x^k|) \frac{G_i^\top x^{k+1}}{|G_i^\top x^{k+1}|} G_i + \rho(x^{k+1} - x^k) + \partial \mathcal{I}_{\bar{C}}(x^{k+1}), \end{aligned} \tag{3.7}$$

where $\partial \mathcal{I}_{\bar{C}}(x^{k+1}) = \sum_{i \in \mathrm{J} \setminus \bar{\mathrm{T}}} (\ker G_i^\top)^\perp$.

Denote $\bar{\mathrm{T}}_{++}^{k+1} = \{i \in \bar{\mathrm{T}} : G_i^\top x^{k+1} > 0\}$. Consider a fixed index $j \in \bar{\mathrm{T}}$ and define the set

$$V(\bar{\mathrm{T}}, j, \bar{\mathrm{T}}_{++}^{k+1}) = \bar{C} \cap \{v : G_i^\top v \geq 0 \ \forall i \in \bar{\mathrm{T}}_{++}^{k+1}, \quad G_i^\top v \leq 0 \ \forall i \in \bar{\mathrm{T}} \setminus \bar{\mathrm{T}}_{++}^{k+1}\} \cap \{v : |G_j^\top v| = 1\}.$$

Note that $V(\bar{\mathrm{T}}, j, \bar{\mathrm{T}}_{++}^{k+1}) \neq \emptyset$, since $v = \frac{x^{k+1}}{|G_j^\top x^{k+1}|} \in V(\bar{\mathrm{T}}, j, \bar{\mathrm{T}}_{++}^{k+1})$. Take $v = \bar{v}^{k+1}$ to be the solution to the following problem

$$\min \|v\|_2 \quad \text{s.t. } v \in V(\bar{\mathrm{T}}, j, \bar{\mathrm{T}}_{++}^{k+1})$$

implying that

$$\bar{v}^{k+1} \in \bar{C}; \text{ for any } i \in \bar{\mathrm{T}}, \ G_i^\top x^{k+1} \cdot G_i^\top \bar{v}^{k+1} \geq 0; \ \frac{G_j^\top x^{k+1}}{|G_j^\top x^{k+1}|} G_j^\top \bar{v}^{k+1} = 1,$$

and $\|\bar{v}^{k+1}\|_2 \leq \mu$ with μ as defined in (2.7). Then by (3.7), there exists $z^{k+1} \in \partial \mathcal{Q}_q(x^{k+1})$ and $\eta_i^{k+1} \in (-\infty, +\infty)$ $(i \in \mathrm{J} \setminus \bar{\mathrm{T}})$ such that

$$\begin{aligned} \langle h^{k+1}, \bar{v}^{k+1} \rangle = \langle z^{k+1}, \bar{v}^{k+1} \rangle &+ \sum_{i \in \bar{\mathrm{T}}} \phi'(|G_i^\top x^k|) \frac{G_i^\top x^{k+1}}{|G_i^\top x^{k+1}|} G_i^\top \bar{v}^{k+1} \\ &+ \rho \langle x^{k+1} - x^k, \bar{v}^{k+1} \rangle + \sum_{i \in \mathrm{J} \setminus \bar{\mathrm{T}}} \eta_i^{k+1} G_i^\top \bar{v}^{k+1}, \end{aligned}$$

indicating

$$\begin{aligned} &\langle h^{k+1}, \bar{v}^{k+1} \rangle - \langle z^{k+1}, \bar{v}^{k+1} \rangle - \rho \langle x^{k+1} - x^k, \bar{v}^{k+1} \rangle \\ &= \sum_{i \in \bar{\mathrm{T}}} \phi'(|G_i^\top x^k|) \frac{G_i^\top x^{k+1}}{|G_i^\top x^{k+1}|} G_i^\top \bar{v}^{k+1} \geq \phi'(|G_j^\top x^k|). \end{aligned}$$

By the boundedness of $\{x^k\}$, the stopping condition (3.2) in Algorithm 2, the fact that $\|\bar{v}^{k+1}\|_2 \le \mu$, and (2.3), there exists some constant $\bar{\alpha}$ (independent of k) such that

$$\phi'(|G_j^\top x^k|) \le \langle h^{k+1}, \bar{v}^{k+1}\rangle - \langle z^{k+1}, \bar{v}^{k+1}\rangle - \rho\langle x^{k+1} - x^k, \bar{v}^{k+1}\rangle \le \bar{\alpha}.$$

Since $\phi'(0+) = +\infty$, the constant $\theta' = \inf\{t > 0 : \phi'(t) \le \bar{\alpha}\} > 0$ is well defined and consequently

$$|G_j^\top x^k| \ge \theta',$$

holds $\forall j \in \bar{\mathrm{T}}$ and $\forall k \ge K$. Let $\bar{\theta} = \max\{\tau, \theta'\} > 0$, which completes the proof. ∎

In Lemma 3.3, we have proved the existence of a uniform lower bound for nonzero transform coefficients generated by the iterative sequence. Meanwhile, as stated in Assumption 2.1 (c), ϕ' is L_α-Lipschitz continuous on $[\alpha, +\infty)$ for any $\alpha > 0$. We can then overcome the difficulties in the convergence analysis brought by the non-Lipschitz property of ϕ' at the zero point, and a subgradient lower bound for the iteration gap is given in the following lemma.

LEMMA 3.4 (A subgradient lower bound for the iteration gap). *For each $k \ge K$, there exists $s^{k+1} \in \partial\mathcal{F}(x^{k+1})$ such that*

$$\|s^{k+1}\|_2 \le c\|x^{k+1} - x^k\|_2,$$

for some constant c.

Proof. We calculate

$$\begin{aligned}\partial\mathcal{F}(x^{k+1}) =& \partial\mathcal{Q}_q(x^{k+1}) + \sum_{i\in\bar{\mathrm{T}}} \phi'(|G_i^\top x^{k+1}|)\frac{G_i^\top x^{k+1}}{|G_i^\top x^{k+1}|}G_i + \partial\left(\sum_{i\in \mathrm{J}\setminus\bar{\mathrm{T}}} \phi(|G_i^\top x^{k+1}|)\right)\\ =& \partial\mathcal{Q}_q(x^{k+1}) + \sum_{i\in\bar{\mathrm{T}}} \phi'(|G_i^\top x^{k+1}|)\frac{G_i^\top x^{k+1}}{|G_i^\top x^{k+1}|}G_i + \sum_{i\in \mathrm{J}\setminus\bar{\mathrm{T}}} (\ker G_i^\top)^\perp.\end{aligned}$$

On the other hand, by Algorithm 2, $\|h^{k+1}\|_2 \le \frac{\rho}{2}\epsilon\|x^{k+1} - x^k\|_2$ and

$$h^{k+1} \in \partial\mathcal{Q}_q(x^{k+1}) + \sum_{i\in\bar{\mathrm{T}}} \phi'(|G_i^\top x^k|)\frac{G_i^\top x^{k+1}}{|G_i^\top x^{k+1}|}G_i + \rho(x^{k+1} - x^k) + \partial\mathcal{I}_{\bar{C}}(x^{k+1}),$$

where $\partial\mathcal{I}_{\bar{C}}(x^{k+1}) = \sum_{i\in \mathrm{J}\setminus\bar{\mathrm{T}}} (\ker G_i^\top)^\perp$.

It is then easy to see

$$s^{k+1} = h^{k+1} + \sum_{i\in\bar{\mathrm{T}}} \left(\phi'(|G_i^\top x^{k+1}|) - \phi'(|G_i^\top x^k|)\right)\frac{G_i^\top x^{(k+1)}}{|G_i^\top x^{(k+1)}|}G_i - \rho(x^{k+1} - x^k),$$

is in $\partial\mathcal{F}(x^{k+1})$. By Assumption 2.1 (c) and Lemma 3.3, we further derive

$$\begin{aligned}\|s^{k+1}\|_2 &\le \|h^{k+1}\|_2 + \sum_{i\in\bar{\mathrm{T}}} |\phi'(|G_i^\top x^{k+1}|) - \phi'(|G_i^\top x^k|)|\|G_i\|_2 + \rho\|x^{k+1} - x^k\|_2\\ &\le \sum_{i\in\bar{\mathrm{T}}} L_{\bar{\theta}}\big||G_i^\top x^{k+1}| - |G_i^\top x^k|\big| \cdot \|G_i\|_2 + \rho(1 + \frac{\epsilon}{2})\|x^{k+1} - x^k\|_2\end{aligned}$$

$$\leq \left(L_{\bar{\theta}} \sum_{i \in \mathrm{J}} \|G_i\|^2 + \rho(1+\frac{\epsilon}{2}) \right) \|x^{k+1} - x^k\|_2.$$

Taking $c = L_{\bar{\theta}} \sum_{i \in \mathrm{J}} \|G_i\|^2 + \rho(1+\frac{\epsilon}{2})$ completes the proof. ∎

We can now establish the global convergence of the sequence generated by Inexact ITSS-PL (Algorithm 2).

THEOREM 3.1 (Global convergence). *Suppose that $\mathcal{F}(x)$ is a KL function. The sequence $\{x^k\}$ generated by Algorithm 2 converges to a stationary point of $(\mathfrak{F})$.*

Proof. By Lemma 3.2, $\{x^k\}$ is bounded. Then there exists a subsequence $\{x^{k_j}\}$ and a point x^* such that

$$x^{k_j} \to x^* \text{ and } \mathcal{F}(x^{k_j}) \to \mathcal{F}(x^*) \quad \text{as } j \to \infty.$$

Since $\mathcal{F}(x)$ is a KL function, by Lemma 3.1, Lemma 3.4 and [4, Theorem 2.9], the sequence $\{x^k\}$ converges globally to x^*, which is a stationary point of problem $(\mathfrak{F})$. ∎

REMARK 3.1. A proper lower semicontinous function satisfying the KL property at all points in its domain is called a KL function. The objective functions $\mathcal{F}(x)$ in our examples are KL functions. Indeed, it is known that any proper lower semicontinuous function that is definable on an o-minimal structure is a KL function. See [9] and [3, Theorem 4.1]. A class of o-minimal structure is the log-exp structure ([27, Example 2.5]). By this structure, the examples of potential function $\phi(t) = t^p$ and $\phi(t) = \log(1+t^p)$ are definable, and $|G_i^\top x|$, as a semi-algebraic function, is also definable. Then the composition $\phi(|G_i^\top x|)$ and the finite sum $\sum_{i \in \mathrm{J}} \phi(|G_i^\top x|)$ are definable functions. Similarly, the fidelity $\mathcal{Q}_q(x)$ is a definable function. As the sum of regularization and fidelity terms, the objective function $\mathcal{F}(x)$ is definable, therefore it is a KL function. See, e.g., [2,3,63,69,73,74] for discussions. For more details about the KL function, one may refer to the appendix.

4. Algorithm implementation

The subproblem $(\mathfrak{H}_k^\tau)$ at each iteration of Inexact ITSS-PL can be solved by many effective convex optimization algorithms. We consider to use the alternating direction method of multipliers (ADMM [12,37,38,41,68]) and the following are the implementation details.

In the description of ADMM, the variable x in the subproblem $(\mathfrak{H}_k^\tau)$ is replaced by u to avoid confusion with the outer iterations. We first consider the general form of $(\mathfrak{H}_k^\tau)$ with $q \in [1, +\infty] \backslash \{2\}$ in the fidelity. To solve such a subproblem, we introduce the variables $v \in \mathbb{R}^M$ and $\{w_i\}_{i \in \mathrm{T}^k} \in \mathbb{R}^{|\mathrm{T}^k|}$, and reformulate the subproblem $(\mathfrak{H}_k^\tau)$ as the following equivalent form,

$$\begin{aligned} \min_{u,v,\{w_i\}_{i\in\mathrm{T}^k}} & \sum_{i \in \mathrm{T}^k} \phi'(|G_i^\top x^k|)|w_i| + \psi(v) + \frac{1}{2}\rho \left\|u - x^k\right\|_2^2 \\ \text{s.t.} & \begin{cases} G_i^\top u = 0, & \forall i \in \mathrm{J} \backslash \mathrm{T}^k \\ G_i^\top u = w_i, & \forall i \in \mathrm{T}^k \\ Au - b = v, \end{cases} \end{aligned} \tag{4.1}$$

where

$$\psi(v)=\begin{cases}\frac{\beta}{q}\|v\|_q^q, & q\in[1,+\infty)\backslash\{2\},\\ \beta\|v\|_\infty, & q=+\infty.\end{cases}$$

The augmented Lagrangian function of the above constrained problem reads

$$\begin{aligned}&\mathcal{L}(u,v,\{w_i\}_{i\in\mathrm{T}^k},\lambda_v,\lambda_w;r_v,r_w)\\ &=\psi(v)+\sum_{i\in\mathrm{T}^k}\phi'(|G_i^\top x^k|)|w_i|+\frac{1}{2}\rho\left\|u-x^k\right\|_2^2+\langle\lambda_v,Au-b-v\rangle+\frac{1}{2}r_v\|Au-b-v\|_2^2\\ &\quad+\sum_{i\in\mathrm{T}^k}\langle(\lambda_w)_i,G_i^\top u-w_i\rangle+\frac{1}{2}r_w\sum_{i\in\mathrm{T}^k}|G_i^\top u-w_i|^2+\sum_{i\in\mathrm{J}\backslash\mathrm{T}^k}\langle(\lambda_w)_i,G_i^\top u\rangle\\ &\quad+\frac{1}{2}r_w\sum_{i\in\mathrm{J}\backslash\mathrm{T}^k}|G_i^\top u|^2,\end{aligned}\tag{4.2}$$

with Lagrangian multipliers $\lambda_v\in\mathbb{R}^M$, $\lambda_w\in\mathbb{R}^{\sharp\mathrm{J}}$, and penalty parameters $r_v,r_w>0$. Then the ADMM to solve $(\mathfrak{H}_k^\tau)$ is given in Algorithm 3.

Algorithm 3 ADMM to solve the subproblem $(\mathfrak{H}_k^\tau)$ with $q\in[1,+\infty]\backslash\{2\}$

Require: $r_v>0$, $r_w>0$, $\mathrm{MaxIter}_{\mathrm{in}}$, $\epsilon_{\mathrm{in}}>0$, $u^0=x^k$, $\lambda_v^0=0\in\mathbb{R}^M$, $\lambda_w^0=0\in\mathbb{R}^{\sharp\mathrm{J}}$

while $l\le\mathrm{MaxIter}_{\mathrm{in}}$ and $\frac{\|u^{l+1}-u^l\|_2}{\|u^{l+1}\|_2}>\epsilon_{\mathrm{in}}$ **do**

1. Compute $(v^{l+1},\{w_i^{l+1}\}_{i\in\mathrm{T}^k})$ by $\min_{v,\{w_i\}_{i\in\mathrm{T}^k}}\{\mathcal{L}(u^l,v,\{w_i\}_{i\in\mathrm{T}^k},\lambda_v^l,\lambda_w^l;r_v,r_w)\}$;
2. Compute u^{l+1} by $\min_u\{\mathcal{L}(u,v^{l+1},\{w_i^{l+1}\}_{i\in\mathrm{T}^k},\lambda_v^l,\lambda_w^l;r_v,r_w)\}$;
3. Update λ_v^{l+1} and λ_w^{l+1} by $\lambda_v^{l+1}=\lambda_v^l+r_v(Au^{l+1}-b-v^{l+1})$ and

$$(\lambda_w^{l+1})_i=\begin{cases}(\lambda_w^l)_i+r_w(G_i^\top u^{l+1}-w_i^{l+1}), & \text{if}\quad i\in\mathrm{T}^k,\\ (\lambda_w^l)_i+r_wG_i^\top u^{l+1}, & \text{if}\quad i\in\mathrm{J}\backslash\mathrm{T}^k.\end{cases}$$

end while

For the $(v,\{w_i\}_{i\in\mathrm{T}^k})$-subproblem in Algorithm 3, v and $\{w_i\}_{i\in\mathrm{T}^k}$ are separable variables and we can compute them in parallel. Herein, $\{w_i^{l+1}\}_{i\in\mathrm{T}^k}$ has the explicit form solution given by the soft-thresholding [28,29]:

$$w_i^{l+1}=\max\left\{|G_i^\top u^l+\frac{1}{r_w}(\lambda_w^l)_i|-\frac{1}{r_w}\phi'(|G_i^\top x^k|),0\right\}\cdot\mathrm{sign}\left(G_i^\top u^l+\frac{1}{r_w}(\lambda_w^l)_i\right),\forall i\in\mathrm{T}^k.$$

v^{l+1} can be solved by numerical algorithms from the following convex optimization problem

$$\min_v\left\{\psi(v)+\langle\lambda_v^l,Au^l-b-v\rangle+\frac{1}{2}r_v\left\|Au^l-b-v\right\|_2^2\right\}.\tag{4.3}$$

If $q\in(1,+\infty)\backslash\{2\}$, the objective function is convex and continuously differentiable, and the problem (4.3) can be solved by differentiable convex optimization approaches, e.g. the gradient descent algorithm. If $q=1$, v^{l+1} also has the explicit form solution:

$$v_j^{l+1}=\max\left\{|(Au^l)_j-b_j+\frac{1}{r_v}(\lambda_v^l)_j|-\frac{\beta}{r_v},0\right\}\cdot\mathrm{sign}\left((Au^l)_j-b_j+\frac{1}{r_v}(\lambda_v^l)_j\right),\forall 1\le j\le M,$$

where $(Au^l)_j$ is the j'th element of vector Au^l. If $q=+\infty$, the v-subproblem (4.3) is equivalent to the following problem

$$\min_v \|v\|_\infty + \frac{r_v}{2\beta}\|v-y\|_2^2,$$

where $y=Au^l-b+\frac{1}{r_v}\lambda_v^l$. There are several methods to compute the solution to this problem [5,63]. Here, we use the explicit form solution in [63].

For solving the u-subproblem, we augment $\{w_i^{l+1}\}_{i\in \mathrm{T}^k}$ by $w_i^{l+1}=0$, $\forall i\in \mathrm{J}\backslash \mathrm{T}^k$, so that we can make full use of the structure of the operators $\{G_i\}_{i\in \mathrm{J}}$, like [69,73]. Then the u-subproblem can be equivalent to

$$\min_u \left\{ \frac{1}{2}u^\top \left(\rho I_N + r_v A^\top A + r_w \sum_{i\in \mathrm{J}} G_i G_i^\top\right) u - \left(\rho x^k + A^\top (r_v b + r_v v^{l+1} - \lambda_v^l) + \sum_{i\in \mathrm{J}} \left(r_w w_i^{l+1} - (\lambda_w^l)_i\right) G_i\right)^\top u \right\}, \tag{4.4}$$

whose solution can be obtained by solving the corresponding normal equation. In the case of convolutional operators A,G with periodic boundary condition in image-deblur like applications, $\rho I_N + r_v A^\top A + r_w \sum_{i\in \mathrm{J}} G_i G_i^\top$ is a block circulant matrix and it can be diagonalized by the two-dimensional discrete Fourier transforms. The u-subproblem (4.4) can then be efficiently solved by utilizing the fast Fourier transform, like [64,68].

We are then left with the special case of $(\mathfrak{H}_k^\tau)$ with $q=2$ in the fidelity, i.e., the case of quadratic fidelity term. This is an easier case and the subproblem $(\mathfrak{H}_k^\tau)$ can be reformulated into

$$\begin{aligned} \min_{u,\{w_i\}_{i\in \mathrm{T}^k}} & \sum_{i\in \mathrm{T}^k} \phi'(|G_i^\top x^k|)|w_i| + \frac{\beta}{2}\|Au-b\|_2^2 + \frac{1}{2}\rho\left\|u-x^k\right\|_2^2 \\ \text{s.t.} & \begin{cases} G_i^\top u = 0, & \forall i\in \mathrm{J}\backslash \mathrm{T}^k \\ G_i^\top u = w_i, & \forall i\in \mathrm{T}^k. \end{cases} \end{aligned} \tag{4.5}$$

The corresponding augmented Lagrangian function is

$$\begin{aligned} &\mathcal{L}(u,\{w_i\}_{i\in \mathrm{T}^k},\lambda_w;r_w) \\ =&\frac{\beta}{2}\|Au-b\|_2^2 + \sum_{i\in \mathrm{T}^k} \phi'(|G_i^\top x^k|)|w_i| + \frac{\rho}{2}\left\|u-x^k\right\|_2^2 + \sum_{i\in \mathrm{T}^k} \langle (\lambda_w)_i, G_i^\top u - w_i\rangle \\ &+\frac{1}{2}r_w \sum_{i\in \mathrm{T}^k} |G_i^\top u - w_i|^2 + \sum_{i\in \mathrm{J}\backslash \mathrm{T}^k} \langle (\lambda_w)_i, G_i^\top u\rangle + \frac{1}{2}r_w \sum_{i\in \mathrm{J}\backslash \mathrm{T}^k} |G_i^\top u|^2, \end{aligned} \tag{4.6}$$

where $\{w_i\}_{i\in \mathrm{T}^k}\in \mathbb{R}^{\sharp \mathrm{T}^k}$ is the auxiliary variable, $\lambda_w\in \mathbb{R}^{\sharp \mathrm{J}}$ is the Lagrangian multiplier and $r_w>0$. The ADMM to solve $(\mathfrak{H}_k^\tau)$ with $q=2$ is shown in Algorithm 4.

REMARK 4.1. The ADMM converges to the unique minimizer (with zero subgradient) of the strongly convex $(\mathfrak{H}_k^\tau)$. After a large enough number of iterations of ADMM and a projection to the feasible set, one can get a point close enough to the unique minimizer of $(\mathfrak{H}_k^\tau)$ and check condition (3.2). However, the projection step and checking condition (3.2) at each iteration of the subsolver are time consuming. Therefore, in practical computation, we propose a simple stopping criterion, that is to check whether

Algorithm 4 ADMM to solve the subproblem $(\mathfrak{H}_k^\tau)$ with $q=2$

Require: $r_w>0$, $\mathrm{MaxIter}_{\mathrm{in}}$, $\epsilon_{\mathrm{in}}>0$, $u^0=x^k$, $\lambda_w^0=0\in\mathbb{R}^{\sharp \mathrm{J}}$

while $l\le \mathrm{MaxIter}_{\mathrm{in}}$ and $\frac{\|u^{l+1}-u^l\|_2}{\|u^{l+1}\|_2}>\epsilon_{\mathrm{in}}$ **do**

1. Compute $\{w_i^{l+1}\}_{i\in \mathrm{T}^k}$ by $\min_{\{w_i\}_{i\in \mathrm{T}^k}}\{\mathcal{L}(u^l,\{w_i\}_{i\in \mathrm{T}^k},\lambda_w^l;r_w)\}$;
2. Compute u^{l+1} by $\min_u\{\mathcal{L}(u,\{w_i^{l+1}\}_{i\in \mathrm{T}^k},\lambda_w^l;r_w)\}$;
3. Update λ_w^{l+1} by

$$(\lambda_w^{l+1})_i=\begin{cases}(\lambda_w^l)_i+r_w(G_i^\top u^{l+1}-w_i^{l+1}), & \text{if} \quad i\in \mathrm{T}^k,\\ (\lambda_w^l)_i+r_w G_i^\top u^{l+1}, & \text{if} \quad i\in \mathrm{J}\backslash \mathrm{T}^k.\end{cases}$$

end while

it satisfies $\frac{\|u^{l+1}-u^l\|_2}{\|u^{l+1}\|_2}<\epsilon_{\mathrm{in}}$ and whether the iteration number l exceeds the predefined maximum iteration number $\mathrm{MaxIter}_{\mathrm{in}}$. Such an approach can save running time and achieve fairly good restoration results.

5. Experiments

In this section, we test the performance of the proposed algorithm in image deconvolution, and in particular, we consider three types of i.i.d. noises: (1) the salt and pepper impulse noise, (2) the Gaussian noise and (3) the uniform noise. Therefore, we recall here the model (2.5), (2.4) and (2.8), all of which can be solved by the Inexact ITSS-PL given in Algorithm 2. For convenience, we rename the Inexact ITSS-PL for model (2.5) as *Inexact iterative thresholding and support shrinking algorithm with proximal linearization for ℓ_1 fidelity (Inexact ITSS-PL-ℓ_1)*, the ITSS-PL for model (2.4) as *Inexact iterative thresholding and support shrinking algorithm with proximal linearization for ℓ_2 fidelity (Inexact ITSS-PL-ℓ_2)*, and the ITSS-PL for model (2.8) as *Inexact iterative thresholding and support shrinking algorithm with proximal linearization for ℓ_∞ fidelity (Inexact ITSS-PL-ℓ_∞)*.

In the experiments, we take the potential function ϕ as $\phi(t):=t^p$, $\forall t\in[0,+\infty)$, with $p\in(0,1)$, and take the sparsifying system $\{G_i\}_{i\in \mathrm{J}}$ as the horizontal and vertical discrete derivative operators. In this case, model (2.5), (2.4) and (2.8) become the anisotropic $\ell_1\mathrm{TV}^p$, $\ell_2\mathrm{TV}^p$ and $\ell_\infty\mathrm{TV}^p$ models (namely the $\ell_1\mathrm{aTV}^p$, $\ell_2\mathrm{aTV}^p$, and $\ell_\infty\mathrm{aTV}^p$ models), and they are respectively solved by the Inexact ITSS-PL-ℓ_1, Inexact ITSS-PL-ℓ_2, and Inexact ITSS-PL-ℓ_∞. The subproblems $(\mathfrak{H}_k^\tau)$ in Inexact ITSS-PL-ℓ_1 and Inexact ITSS-PL-ℓ_∞ are solved by the ADMM given in Algorithm 3, and those in Inexact ITSS-PL-ℓ_2 are solved by the ADMM given in Algorithm 4, respectively.

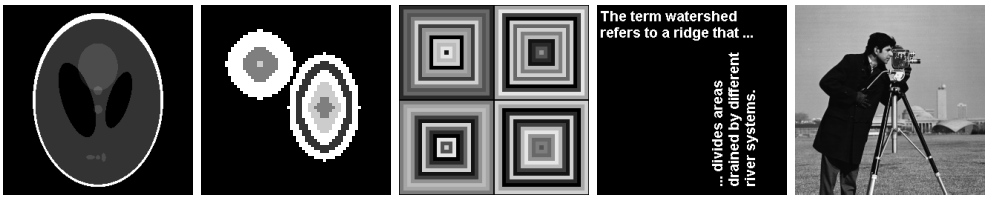
5.1. Test platform and parameter choices. The experiments are performed under Windows 8 and MATLAB R2018a running on a desktop equipped with an Intel Core i7-6700 CPU @ 3.40GHz and 8.00G RAM memory.

For the experiments of image deconvolution, the test images are first degraded by a blur kernel and then some noise is added. Three types of blurring kernels are used, which include (1) average blur (fspecial ('average',5)); (2) Gaussian blur (fspecial ('gaussian', [23,23], 12)); (3) disk blur (fspecial ('disk',6)). Three types of i.i.d. noises are considered: (1) salt-and-pepper impulse noise with noise level 30%; (2) Gaussian noise with mean 0 and variance 10^{-6}; (3) uniform noise with amount 10^{-4}. The quality of the restored images is measured by the peak signal-to-noise ratio (PSNR), which is

defined as follows:

$$\mathrm{PSNR} = 10\log_{10}\frac{N}{\|x_{\mathrm{Alg}} - \underline{x}\|_2^2},$$

where N is the number of image pixels, x_{Alg} is the restored image, and $\underline{x}$ is the ground truth. The test images are shown in Figure 5.1.



(a) (b) (c) (d) (e)

FIG. 5.1. *Test images for restoration experiments. (a): Phantom* (256×256)*; (b): Twocircles* (64×64) *(c): Squares* (256×256)*; (d): Text* (256×256)*; (e): Cameraman* (256×256)*.*

For the model parameters, we use $p=0.5$ in the $\ell_q\mathrm{aTV}^p$ $(q=1,2,+\infty)$ model, and the parameter β is tuned up to achieve the best performance of each method, which will be specified in the subsequent experiments. For the algorithm parameters, they will be set as follows and remain unchanged throughout the experiments, unless otherwise specified. For the Inexact ITSS-PL-ℓ_q $(q=1,2,+\infty)$, the parameter ρ in the proximal term is set as $\rho=10^{-10}$, the parameter τ for defining the τ-support T^k is taken as $\tau=10^{-7}$. In this experiment, the tolerance of the outer loop in the Inexact ITSS-PL-ℓ_q $(q=1,2,+\infty)$ is set to be $\epsilon=10^{-3}$, and the maximum outer iteration number is set to be $\mathrm{MaxIter}=25$. We use $r_v=3\times10^5$ and $r_w=200$ in the ADMM. The tolerance of the inner loop is set to be $\epsilon_{\mathrm{in}}=10^{-5}$, and the maximum inner iteration number is set to be $\mathrm{MaxIter}_{\mathrm{in}}=500$.

The initial value of the Inexact ITSS-PL-ℓ_q $(q=1,2,+\infty)$ is taken as the solution to the ℓ_qaTV model (i.e., the anisotropic $\ell_q\mathrm{TV}^p$ model with $p=1$) solved by the ADMM described in Algorithm 3 and Algorithm 4 with $p=1$ and $\rho=0$. For solving the ℓ_qaTV model, we set $r_v=3\times10^3$, $r_w=200$, $\epsilon_{\mathrm{in}}=10^{-5}$ and $\mathrm{MaxIter}_{\mathrm{in}}=500$ in the ADMM.

5.2. The convergence and τ-support shrinkage properties. As Lemma 3.1 stated, the objective function value $\mathcal{F}(x^k)$ of the Inexact ITSS-PL-ℓ_q is guaranteed to be nonincreasing if $k\geq K$. However, in practice, the exact value of K is usually unknown, and such a property can only be observed numerically. In this subsection, we test the numerical evolution behavior of objective function values, support sizes and relative change when applying the Inexact ITSS-PL-ℓ_q in image restoration. We test on three sample images corrupted by different blur kernels. To reveal the behavior, the maximum outer iteration number is set to be $\mathrm{MaxIter}=30$ in all the algorithms. We show the evolution behavior of $\mathcal{F}(x^k)$, the percentage $\sharp\mathrm{T}^k/\sharp\mathrm{J}$, and the logarithm of relative change $\log_{10}\frac{\|x^k-x^{k-1}\|}{\|x^k\|}$ in Figure 5.2, for $q=1,2,+\infty$ respectively. It is worth mentioning that ADMM is an infeasible optimization method. If the subproblem $(\mathfrak{H}_k^\tau)$ is not solved accurately enough, the objective function value and support size may have minor fluctuations.

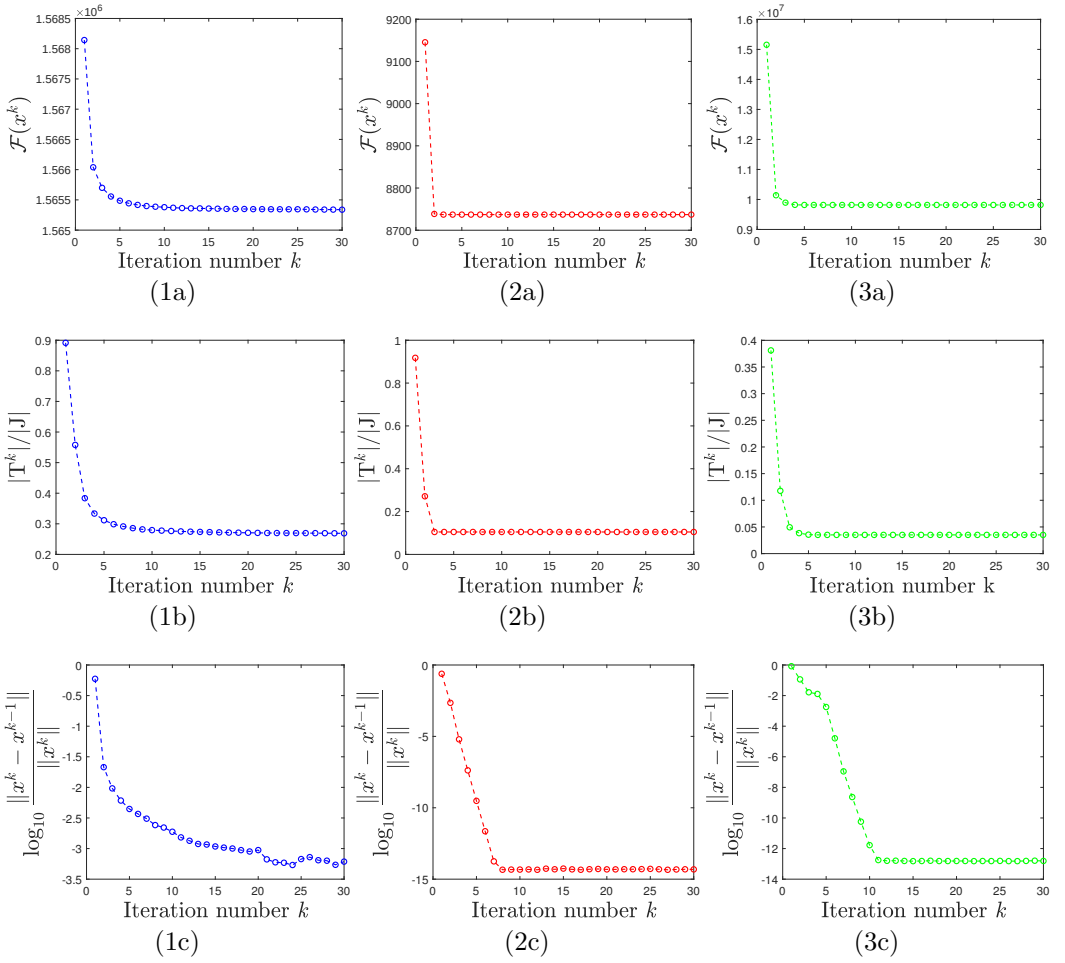
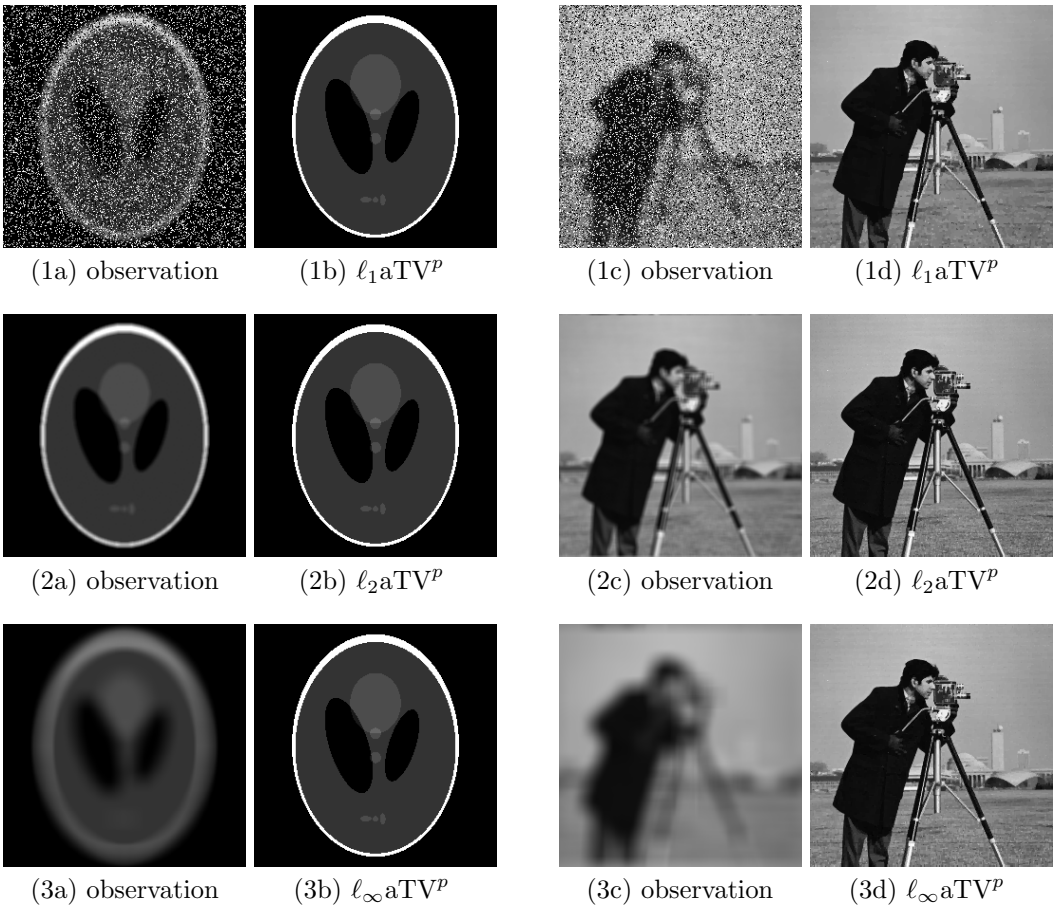


FIG. 5.2. *The behavior of objective function value $\mathcal{F}(x^k)$ (the first row), support ratio $|\mathrm{T}^k|\backslash|\mathrm{J}|$ (the second row), and logarithm of the relative change between succesive iterates $\|x^k-x^{k-1}\|/\|x^k\|$ (the third row), versus the outer iteration number k. (1a)(1b)(1c): results of Inexact ITSS-PL-ℓ_1 for the "Cameraman" degraded by the disk blur and impulse noise; (2a)(2b)(2c): results of Inexact ITSS-PL-ℓ_2 for the "Squares" degraded by the average blur and Gaussian noise; (3a)(3b)(3c): results of Inexact ITSS-PL-ℓ_∞ for the "Text" degraded by the Gaussian blur and uniform noise. The blur kernel parameters and noise levels are given in Section 5.1.*

5.3. Restoration performance. In this subsection, we test the performance of ℓ_qaTV^p $(q=1,2,+\infty)$ model and the Inexact ITSS-PL in image restoration. Figure 5.3 shows some restoration results by our methods. The performance of the ℓ_1aTV^p, ℓ_2aTV^p and ℓ_∞aTV^p are then compared to the ℓ_1aTV, ℓ_2aTV and ℓ_∞aTV, respectively. The ℓ_qaTV$(q=1,2,+\infty)$ model for comparison is also solved by the ADMM with the same algorithm parameters used in the ℓ_qaTV model for initializing the ℓ_qaTV$^p(q=1,2,+\infty)$ model. Note that the model parameter β for ℓ_qaTV is also fine tuned. The PSNR values of the results recovered by different methods are illustrated in Table 5.1. It can be seen that in all the cases of three types of noises, the ℓ_qaTV^p $(q=1,2,+\infty)$ model performs much better (with much higher PSNR values) than the ℓ_qaTV $(q=1,2,+\infty)$ model on piecewise constant images, while the ℓ_qaTV model sometimes outperforms the ℓ_qaTV^p model on the natural image.

Image		disk blur impulse noise		average blur Gaussian noise		Gaussian blur uniform noise	
		ℓ_1aTV	ℓ_1aTV^p	ℓ_2aTV	ℓ_2aTV^p	ℓ_∞aTV	ℓ_∞aTV^p
Phantom	PSNR(dB)	99.56	**156.96**	56.89	**73.76**	46.44	**89.98**
	β/q (or β if $q=\infty$)	100	25	7×10^3	2×10^4	5×10^6	4×10^6
Twocircles	PSNR(dB)	63.51	**216.01**	46.79	**69.72**	33.01	**80.88**
	β/q (or β if $q=\infty$)	160	75	2×10^4	2×10^4	8×10^5	2×10^6
SS3	PSNR(dB)	91.89	**263.52**	57.67	**71.83**	82.89	**88.43**
	β/q (or β if $q=\infty$)	45	60	8×10^3	7×10^3	2×10^6	8×10^5
Text	PSNR(dB)	40.98	**188.52**	49.00	**70.60**	29.93	**88.95**
	β/q (or β if $q=\infty$)	140	85	2×10^4	2×10^4	9×10^6	6×10^6
Cameraman	PSNR(dB)	**30.04**	29.11	**35.48**	34.80	29.93	**32.13**
	β/q (or β if $q=\infty$)	130	160	8×10^4	2×10^5	9×10^6	6×10^8

TABLE 5.1. *PSNR values of restoration results and the model parameters. The blur kernel parameters and noise levels are given in Section 5.1.*

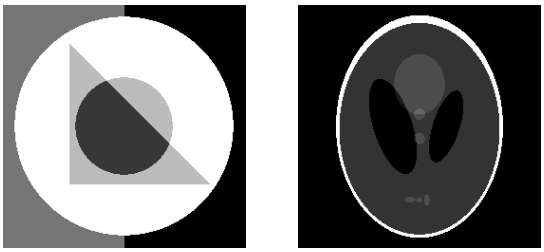


(1a) observation (1b) ℓ_1aTV^p (1c) observation (1d) ℓ_1aTV^p

(2a) observation (2b) ℓ_2aTV^p (2c) observation (2d) ℓ_2aTV^p

(3a) observation (3b) ℓ_∞aTV^p (3c) observation (3d) ℓ_∞aTV^p

FIG. 5.3. *Image restoration by the ℓ_qaTV^p model solved by the Inexact ITSS-PL-ℓ_q ($q=1,2,+\infty$). (1a)(1c): images degraded by disk blur kernel and impulse noise; (2a)(2c): images degraded by average blur kernel and Gaussian noise; (3a)(3c): images degraded by Gaussian blur kernel and uniform noise. The blur kernel parameters and noise levels are given in Section 5.1. (1b), (1d), (2b), (2d), and (3b), (3d): restoration results of different methods.*



(a) Geometry (b) Phantom

FIG. 5.4. *Test images for segmentation experiments. (a): 5-phase Geometry (*256×256*); (b): 6-phase Phantom (*256×256*).*

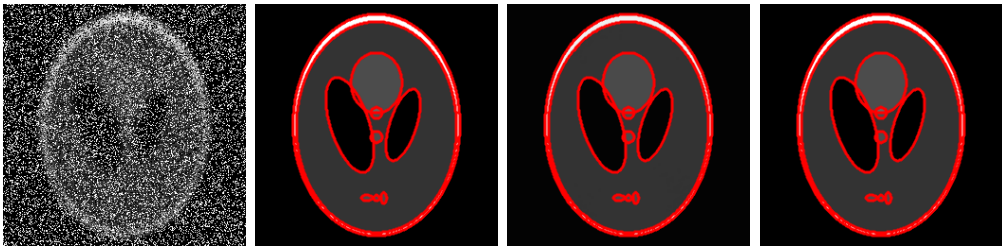
5.4. Applications in image segmentation. Our proposed algorithm can also be applied to the two-stage image segmentation method. Given a blurry and noisy image, the segmentation problem is to partition the image into several regions based on the image intensity. The two-stage image segmentation method [14] includes two major steps: the first stage is to get a piecewise constant approximation of the observation; the second stage is a simple thresholding operation applied to the approximation to obtain the segmentation result. We apply the ℓ_qaTV^p model, and the ℓ_qaTV model as a comparison, at the first stage of finding the approximations. At the second stage, we use the same segmentation method in [14].

In the experiments of image segmentation, we use two piecewise constant test images shown in Figure 5.4, whose ground truth of segmentation can be obtained by MATLAB function *tabulate*, and thus convenient for the quantitive comparision defined later. We use three different blurring kernels: (1) average blur (fspecial('average',9)); (2) Gaussian blur (fspecial('gaussian',[15,15],10)); (3) disk blur (fspecial('disk',6)). Three types of noises are considered: (1) 40% salt-and-pepper impulse noise; (2) Gaussian noise with mean 0 and variance 10^{-6}; (3) uniform noise with amount 10^{-4}. At the first stage of image restoration, the ℓ_1aTV^p model with $p=0.5$ and the ℓ_1aTV model, as a comparison, are used to obtain a piecewise constant approximation of the observation degraded by the impulse noise; the ℓ_2aTV^p model with $p=0.5$ and the ℓ_2aTV model are used for the case of Gaussian noise; and the ℓ_∞aTV^p model with $p=0.5$ and the ℓ_∞aTV model are used for the case of uniform noise. All the algorithm parameters for solving the ℓ_qaTV^p model and the ℓ_qaTV model ($q=1,2,+\infty$) are the same as those used in the experiments of image deconvolution in Section 5.1. In this experiment, parameter β in each method is tuned up to achieve an overall good segmentation result for all the phases. The *Jaccard Similarity* (denoted as JS, [45]) value is used to evaluate the segmentation results, and is defined as follows:

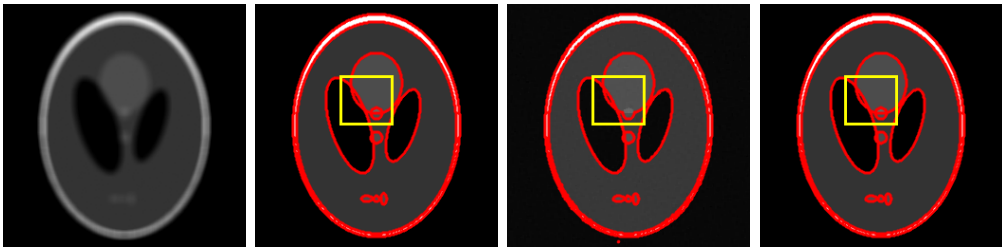
$$\mathrm{JS}(S_{\mathrm{GT}}^{\mathrm{phase}},S_{\mathrm{Alg}}^{\mathrm{phase}}):=\frac{|S_{\mathrm{GT}}^{\mathrm{phase}}\cap S_{\mathrm{Alg}}^{\mathrm{phase}}|}{|S_{\mathrm{GT}}^{\mathrm{phase}}\cup S_{\mathrm{Alg}}^{\mathrm{phase}}|}, \tag{5.1}$$

where $S_{\mathrm{GT}}^{\mathrm{phase}}$ and $S_{\mathrm{Alg}}^{\mathrm{phase}}$ denote the region of one certain phase in the ground truth and the region in the segmentation result, and $|\cdot|$ denotes the area of a region. Clearly, the higher the JS value is, the better the corresponding segmentation result is obtained.

Table 5.2 illustrates the JS values of the experiments. The performance of ℓ_qaTV and ℓ_qaTV^p models are almost the same or comparable. Therefore, the images of "Geometry" are not shown. For "Phantom", our methods have a little advantage in the case of Gaussian noise and average blur; see also Figure 5.5.



(1a) observation (1b) ground truth (1c) ℓ_1aTV (1d) ℓ_1aTV^p



(2a) observation (2b) ground truth (2c) ℓ_2aTV (2d) ℓ_2aTV^p

(3a) observation (3b) ground truth (3c) ℓ_∞aTV (3d) ℓ_∞aTV^p

FIG. 5.5. *Image segmentation based on restoration results by the $\ell_q aTV$ and $\ell_q aTV^p$ models ($q=1,2,+\infty$). (1a): Phantom degraded by disk blur and impulse noise; (2a): Phantom degraded by average blur and Gaussian noise; (3a): Phantom degraded by Gaussian blur and uniform noise. The blur kernel parameters and noise levels are given in Section 5.4. (1b) to (1d), (2b) to (2d), and (3b) to (3d): segmentation results of ground truth and different methods.*

Image			disk blur impulse noise		average blur Gaussian noise		Gaussian blur uniform noise	
			ℓ_1aTV	ℓ_1aTV^p	ℓ_2aTV	ℓ_2aTV^p	ℓ_∞aTV	ℓ_∞aTV^p
Geometry	JS	phase 1(43.6661%)	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**
		phase 2(18.3701%)	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**
		phase 3(18.0176%)	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**
		phase 4(10.4874%)	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**
		phase 5(9.4589%)	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**	**1.000000**
	β/q (or β if $q=\infty$)		110	45	3.6×10^4	4.4×10^4	6.0×10^5	2.0×10^5
Phantom	JS	phase 1(58.1772%)	0.999895	**1.000000**	0.998348	**0.999948**	**1.000000**	**1.000000**
		phase 2(32.9269%)	0.999444	**1.000000**	0.997730	**0.999676**	**1.000000**	**1.000000**
		phase 3(4.3427%)	**1.000000**	**1.000000**	0.930077	**1.000000**	**1.000000**	**1.000000**
		phase 4(4.3350%)	0.996492	**1.000000**	0.967885	**0.997891**	**1.000000**	**1.000000**
		phase 5(0.1389%)	0.978495	**1.000000**	0.551282	**0.967033**	**1.000000**	**1.000000**
		phase 6(0.0793%)	**1.000000**	**1.000000**	0.000000	**1.000000**	**1.000000**	**1.000000**
	β/q (or β if $q=\infty$)		110	40	5.8×10^4	2.3×10^4	3.0×10^6	5.0×10^5

TABLE 5.2. *JS values of segmentation results and the model parameters for "Geometry" and "Phantom". The blur kernel parameters and noise levels are given in Section 5.4. The percentages in the brackets are the ratios of the phase areas to the whole image in the ground truth.*

6. Conclusions

In this paper, we studied a non-Lipschitz restoration model with anisotropic regularization and the $\ell_q, q \in [1, +\infty]$ fidelity. We proved the lower bound theory for the cases of $q = 1$ and $+\infty$, and for a general $q \in [1, +\infty]$, the support inclusion property was derived. For solving such a nonconvex and non-Lipschitz model, we proposed the inexact iterative thresholding and support shrinking algorithm with proximal linearization, which is shown to globally converge to a stationary point of the objective function, provided that the subgradient condition in the subproblem is reached. The proof techniques used in such analysis can be applied to first order related analysis of other image models with anisotropic regularizations. Numerical experiments on image deconvolution and two-stage image segmentation also illustrated the potential advantages of the proposed algorithm in applications.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant No. 11871035 and 11901312) and a Key Program (No. 21JCZDJC00220) of Natural Science Foundation of Tianjin, China.

Appendix.

A.1. Subdifferential.

DEFINITION A.1 ([62, Definition 8.3]). *Let $\omega : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. The domain of ω is defined by dom $\omega := \{y \in \mathbb{R}^n : \omega(y) < +\infty\}$.*

(1) *For each $y \in$ dom ω, the regular subdifferential of ω at y is defined as:*

$$\widehat{\partial}\omega(y) := \{y^\star \in \mathbb{R}^n : \liminf_{\substack{z \to y \\ z \neq y}} \frac{\omega(z) - \omega(y) - \langle y^\star, z - y \rangle}{\|z - y\|} \geq 0\}.$$

If $y \notin$ dom ω, then $\widehat{\partial}\omega(y) = \emptyset$.

(2) *The limiting subdifferential of ω at $y \in$ dom ω is defined as:*

$$\partial\omega(y) := \{y^\star \in \mathbb{R}^n : \exists y^k \to y, \omega(y^k) \to \omega(y), s^k \in \widehat{\partial}\omega(y^k), s^k \to y^\star \quad as \quad k \to +\infty\}.$$

(3) *The horizon subdifferential of ω at $y \in$ dom ω is defined as:*

$$\partial^\infty\omega(y) := \{y^\star \in \mathbb{R}^n : \exists y^k \to y, \omega(y^k) \to \omega(y), s^k \in \widehat{\partial}\omega(y^k), \nu^k s^k \to y^\star \text{ for some sequence } \nu^k \searrow 0 \quad as \quad k \to +\infty\}.$$

REMARK A.1. A point y is said to be a stationary point of ω, if $0 \in \partial\omega(y)$.

A.2. Kurdyka-Łojasiewicz function. The definition for a proper lower semicontinuous function f to have the Kurdyka-Łojasiewicz (KL) property at $\bar{x} \in \text{dom}\partial f$ can be found in [3, Definition 3.1]. A proper lower semicontinuous function f satisfying the KL property at all points in dom∂f is called a *KL function*. A large class of KL functions widely used in applications are given by functions definable in an o-minimal structure introduced in [27]. See also [69] for a summary and discussion.

DEFINITION A.2 ([3, Definition 4.1]). *Let $\mathcal{O} = \{\mathcal{O}_n\}_{n \in \mathbb{N}}$ be such that each $\mathcal{O}_n$ is a collection of subsets of $\mathbb{R}^n$. The family $\mathcal{O}$ is an o-minimal structure over $\mathbb{R}$, if it satisfies the following axioms:*

(1) *Each $\mathcal{O}_n$ is a Boolean algebra. Namely $\emptyset \in \mathcal{O}_n$ and for each $A, B \in \mathcal{O}_n, A \cup B, A \cap B$, and $\mathbb{R}^n \setminus A$ belong to $\mathcal{O}_n$.*

(2) *For all* $A\in\mathcal{O}_n, A\times\mathbb{R}$ *and* $\mathbb{R}\times A$ *belong to* $\mathcal{O}_{n+1}$.

(3) *For all* $A\in\mathcal{O}_{n+1}$, $\prod(A):=\{(x_1,...,x_n)\in\mathbb{R}^n\,|\,(x_1,...,x_n,x_{n+1})\in A\}$ *belongs to* $\mathcal{O}_n$.

(4) *For all* $i\neq j$ *in* $\{1,2,...,n\}$, $\{(x_1,...,x_n)\in\mathbb{R}^n\,|\,x_i=x_j\}\in\mathcal{O}_n$.

(5) *The set* $\{(x_1,x_2)\in\mathbb{R}^2\,|\,x_1<x_2\}$ *belongs to* $\mathcal{O}_2$.

(6) *The elements of* $\mathcal{O}_1$ *are exactly finite unions of intervals.*

We say that a set $A\subseteq\mathbb{R}^n$ belongs to $\mathcal{O}$ if $A\in\mathcal{O}_n$. A map $\Psi:A\to\mathbb{R}^m$ with $A\subseteq\mathbb{R}^n$ is said to belong to $\mathcal{O}$ if its graph $\{(x,\Psi(x))|x\in\mathrm{dom}\Psi\}\subseteq\mathbb{R}^{n+m}$ belongs to $\mathcal{O}$. We say that sets and maps are definable in $\mathcal{O}$ if they belong to $\mathcal{O}$. Definable functions are defined like definable maps. By [3], the o-minimal structure has the properties that (1) the composition of definable functions is definable; (2) the finite sum of definable functions is definable.

A class of o-minimal structure is the log-exp structure given in [27, Example 2.5], by which the following functions are definable:

(1) semi-algebraic functions [10, Definition 5], including real polynomial functions.

(2) $x^r:\mathbb{R}\to\mathbb{R}$ with $r\in\mathbb{R}$, which is given by

$$a\mapsto\begin{cases}a^r, & a>0\\ 0, & a\leq 0.\end{cases}$$

(3) The exponential function: $\mathbb{R}\to\mathbb{R}$ defined by $x\mapsto e^x$ and the logarithm function: $(0,\infty)\to\mathbb{R}$ defined by $x\mapsto\log(x)$.

It has been shown that any proper lower semicontinuous function that is definable in an o-minimal structure is a KL function ([9] and [3, Theorem 4.1]). Then by the aforementioned properties and examples of definable functions, the objective functions $\mathcal{F}(x)$ in the examples of this paper are KL functions.

REFERENCES

[1] P.A. Absil, R. Mahony, and B. Andrews, *Convergence of the iterates of descent methods for analytic cost functions*, SIAM J. Optim., 16(2):531–547, 2005. 3.2

[2] H. Attouch and J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Math. Program., 116:5–16, 2009. 3.2, 3.1

[3] 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*, Math. Oper. Res., 35(2):438–457, 2010. 3.2, 3.1, A.2, A.2

[4] 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*, Math. Program., 137(1-2):91–129, 2013. 3.1, 3.2, 3.2

[5] E. van den Berg and M.P. Friedlander, *Probing the pareto frontier for basis pursuit solutions*, SIAM J. Sci. Comput., 31(2):890–912, 2009. 4

[6] D.P. Bertsekas, *Control of uncertain systems with a set-membership description of the uncertainty*, PhD thesis, Massachusetts Institute of Technology, 1971. 2

[7] W. Bian and X. Chen, *Linearly constrained non-Lipschitz optimization for image restoration*, SIAM J. Imaging Sci., 8(4):2294–2322, 2015. 1

[8] J. Bolte, A. Daniilidis, and A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., 17(4):1205–1223, 01 2007. 3.2

[9] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM J. Optim., 18(2):556–572, 2007. 3.1, A.2

[10] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146(1-2):459–494, 2014. A.2

[11] C. Bouman and K. Sauer, *A generalized Gaussian image model for edge-preserving MAP estimation*, IEEE Trans. Image Process., 2(3):296–310, 1993. 1
[12] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3(1):1-122, 2011. 4
[13] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. 1
[14] X. Cai, R. Chan, and T. Zeng, *A two-stage image segmentation method using a convex variant of the Mumford-Shah model and thresholding*, SIAM J. Imaging Sci., 6(1):368–390, 2013. 5.4
[15] E.J. Candés, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Commun. Pure Appl. Math., 59(8):1207–1223, 2006. 1
[16] E.J. Candés, M.B. Wakin, and S.P. Boyd, *Enhancing sparsity by reweighted ℓ_1 minimization*, J. Fourier Anal. Appl., 14(5-6):877–905, 2008. 1
[17] L. Chai, J. Du, Q.-F. Liu, and C.-H. Lee, *Using generalized Gaussian distributions to improve regression error modeling for deep learning-based speech enhancement*, IEEE/ACM Trans. Audio Speech Lang. Process., 27(12):1919–1931, 2019. 1
[18] R.H. Chan and H.-X. Liang, *Half-quadratic algorithm for ℓ_p-ℓ_q problems with applications to TV-ℓ_1 image restoration and compressed sensing*, in A. Bruhn, T. Pock, and X.C. Tai (eds.), Efficient Algorithms for Global Optimization Methods in Computer Vision, Springer, 8293:78–103, 2014. 1
[19] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett., 14(10):707–710, 2007. 1
[20] X. Chen, *Smoothing methods for nonsmooth, nonconvex minimization*, Math. Program., 134(1):71–99, 2012. 1
[21] J.-M. Chen and B.-S. Chen, *System parameter estimation with input/output noisy data and missing measurements*, IEEE Trans. Signal Process., 48(6):1548–1558, 2000. 1
[22] S.S. Chen, D.L. Donoho, and M.A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Rev., 43(1):129–159, 2001. 1
[23] X. Chen, M. Ng, and C. Zhang, *Non-Lipschitz ℓ_p-regularization and box constrained model for image restoration*, IEEE Trans. Image Process., 21(12):4709–4721, 2012. 1
[24] X. Chen, L. Niu, and Y. Yuan, *Optimality conditions and a smoothing trust region Newton method for non-Lipschitz optimization*, SIAM J. Optim., 23(3):1528–1552, 2013. 1
[25] X. Chen, F. Xu, and Y. Ye, *Lower bound theory of nonzero entries in solutions of l_2-l_p minimization*, SIAM J. Sci. Comput., 32(5):2832–2852, 2010. 1
[26] C. Clason, *L^∞ fitting for inverse problems with uniform noise*, Inverse Probl., 28(10):104007, 2012. 1
[27] L. van den Dries and C. Miller, *Geometric categories and o-minimal structures*, Duke Math. J., 84(2):497–540, 1996. 3.1, A.2, A.2
[28] D.L. Donoho and I.M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81(3):425–455, 1994. 4
[29] D. Donoho, *De-noising by soft-thresholding*, IEEE Trans. Inform. Theory, 41:613–627, 1995. 4
[30] I. Daubechies, R. Devore, M. Fornasier, and C.S. Gunturk, *Iteratively reweighted least squares minimization for sparse recovery*, Commun. Pure Appl. Math., 63(1):1–38, 2010. 1
[31] X. Feng, C. Wu, and C. Zeng, *On the local and global minimizers of ℓ_0 gradient regularized model with box constraints for image restoration*, Inverse Probl., 34(9):095007, 2018. 1
[32] X. Feng, S. Yan, and C. Wu, *The $\ell_{2,q}$ regularized group sparse optimization: Lower bound theory, recovery bound and algorithms*, Appl. Comput. Harmon. Anal., 49(2):381–414, 2020. 1, 3.1, 3.2, 3.2
[33] M.A.T. Figueiredo, J.M. Bioucas-Dias, and R.D. Nowak, *Majorization-minimization algorithms for wavelet-based image restoration*, IEEE Trans. Image Process., 16(12):2980–2991, 2007. 1
[34] S. Foucart and M.-J. Lai, *Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0<q\leq 1$*, Appl. Comput. Harmon. Anal., 26(3):395–407, 2009. 1
[35] J.-J. Fuchs, *Fast implementation of a ℓ_1-ℓ_1 regularized sparse representations algorithm*, Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 3329–3332, 2009. 1
[36] Y. Gao and C. Wu, *A general non-Lipschitz joint regularized model for multi-channel/modality image reconstruction*, CSIAM Trans. Appl. Math., 2(3):395–430, 2021. 1, 2
[37] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-splitting Methods in Nonlinear Mechanics*, SIAM, 9, 1989. 1, 4
[38] R. Glowinski, S. Osher, and W. Yin, *Splitting Methods in Communication, Imaging, Science, and Engineering*, Springer International Publishing, 2016. 4
[39] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci., 2(2):323–343, 2009. 1
[40] L. Granai and P. Vandergheynst, *Sparse approximation by linear programming using an L_1 data-*

fidelity term, Proceedings of Workshop on Signal Processing with Adaptative Sparse Structured Representations, 2005. 1

[41] B. He and X. Yuan, *On the O(1/n) convergence rate of the Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 50(2):700–709, 2012. 4

[42] M. Hintermuller and T. Wu, *A superlinearly convergent R-regularized Newton scheme for variational models with concave sparsity-promoting priors*, Comput. Optim. Appl., 57(1):1–25, 2014. 1

[43] M. Hong, Z. Luo, and M. Razaviyayn, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*, SIAM J. Optim., 26(1):337–364, 2016. 1

[44] W. Ho Pun and B.D. Jeffs, *Adaptive image restoration using a generalized Gaussian model for unknown noise*, IEEE Trans. Image Process., 4(10):1451–1456, 1995. 1

[45] P. Jaccard, *The distribution of the flora in the alpine zone*, New Phytol., 11(2):37–50, 1912. 5.4

[46] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Ann. Inst. Fourier (Grenoble), 48(3):769–783, 1998. 3.2

[47] M.-J. Lai, Y. Xu, and W. Yin, *Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization*, SIAM J. Numer. Anal., 51(2):927–957, 2013. 1

[48] A. Lanza, S. Morigi, L. Reichel, and F. Sgallari, *A generalized Krylov subspace method for ℓ_p-ℓ_q minimization*, SIAM J. Sci. Comput., 37(5):S30–S50, 2015. 1

[49] G. Li and T.-K. Pong, *Global convergence of splitting methods for nonconvex composite optimization*, SIAM J. Optim., 25(4):2434–2460, 2015. 1

[50] Z. Liu, C. Wu, and Y. Zhao, *A new globally convergent algorithm for non-Lipschitz ℓ_p-ℓ_q minimization*, Adv. Comput. Math., 45(3):1369–1399, 2019. 1, 2, 2, 3.1, 3.1, 3.2, 3.2

[51] S. Łojasiewicz, *Une propriété topologique des sous-ensembles analytiques réels*, Les Équations aux Dérivées Partielles (Paris, 1962), 87–89, 1963. 3.2

[52] Y. Lou, T. Zeng, S. Osher, and J. Xin, *A weighted difference of anisotropic and isotropic total variation model for image processing*, SIAM J. Imaging Sci., 8(3):1798–1823, 2015. 2

[53] S.G. Mallat, *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE Trans. Pattern Anal. Mach. Intell., 11(7):674–693, 1989. 1

[54] J. Meng, F. Wang, L. Cui, and J. Liu, *The lower bound of nonlocal gradient for non-convex and non-smooth image patches based regularization*, Inverse Probl., 38(3):035010, 2022. 1, 2

[55] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging Vis., 20(1):99–120, 2004. 1

[56] M. Nikolova, *Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares*, Multiscale Model. Simul., 4(3):960–991, 2005. 1, 2, 2, 2, 2

[57] M. Nikolova, M. Ng, and C.P. Tam, *Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction*, IEEE Trans. Image Process., 19(12):3073–3088, 2010. 1

[58] M. Nikolova, M. Ng, S. Zhang, and W.K. Ching, *Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization*, SIAM J. Imaging Sci., 1(1):2–25, 2008. 1

[59] T.T. Pham and R.J.P. deFigueiredo, *Maximum likelihood estimation of a class of non-Gaussian densities with application to l_p deconvolution*, IEEE Trans. Acoust. Speech Signal Process., 37(1):73–82, 1989. 1

[60] P. Rodriguez and B. Wohlberg, *An efficient algorithm for sparse representations with ℓ^p data fidelity term*, Proceedings of 4th IEEE Andean Technical Conference (ANDESCON), 2008. 1

[61] P. Rodriguez and B. Wohlberg, *Efficient minimization method for a generalized total variation functional*, IEEE Trans. Image Process., 18(2):322–332, 2009. 1

[62] R. Tyrrell Rockafellar and R.J.-B. Wets, *Variational Analysis*, Springer, 317, 2009. 2, 2, 2, A.1

[63] Y. Xue, Y. Feng, and C. Wu, *An efficient and globally convergent algorithm for $\ell_{p,q}-\ell_r$ model in group sparse optimization*, Commun. Math. Sci., 18:227–258, 2020. 1, 2, 3.1, 3.1, 3.1, 4

[64] Y. Wang, J. Yang, W. Yin, and Y. Zhang, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imaging Sci., 1(3):248–272, 2008. 2, 4

[65] Y. Wang, W. Yin, and J. Zeng, *Global convergence of ADMM in nonconvex nonsmooth optimization*, J. Sci. Comput., 78(1):29–63, 2019. 1

[66] Y.-W. Wen, W.-K. Ching, and M. Ng, *A semi-smooth Newton method for inverse problem with uniform noise*, J. Sci. Comput., 75:713–732, 2018. 1

[67] C. Wu, X. Guo, Y. Gao, and Y. Xue, *A general non-Lipschitz infimal convolution regularized model: lower bound theory and algorithm*, SIAM J. Imaging Sci. 15(3):1499–1538, 2022. 1, 2

[68] C. Wu and X.-C. Tai, *Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models*, SIAM J. Imaging Sci., 3(3):300–339, 2010. 2, 4, 4

[69] C. Zeng, R. Jia, and C. Wu, *An iterative support shrinking algorithm for non-Lipschitz optimization in image restoration*, J. Math. Imaging Vis., 61(1):122–139, 2019. 1, 2.1, 2, 3.1, 3.2, 3.2, 3.1, 4, A.2

[70] C. Zeng, C. Wu, and R. Jia, *Non-Lipschitz models for image restoration with impulse noise removal*, SIAM J. Imaging Sci., 12(1):420–458, 2019. 1, 3.1

[71] C. Zeng and C. Wu, *On the edge recovery property of noncovex nonsmooth regularization in image restoration*, SIAM J. Numer. Anal., 56(2):1168–1182, 2018. 1, 2

[72] C. Zeng and C. Wu, *On the discontinuity of images recovered by noncovex nonsmooth regularized isotropic models with box constraints*, Adv. Comput. Math, 45:589–610, 2019. 1

[73] Z. Zheng, M. Ng, and C. Wu, *A globally convergent algorithm for a class of gradient compounded non-Lipschitz models applied to non-additive noise removal*, Inverse Probl., 36(12):125017, 2020. 1, 2.1, 2, 2, 2, 3.1, 3.1, 3.2, 3.2, 3.1, 4

[74] X. Zhang, M. Bai, and M. Ng, *Nonconvex-TV based image restoration with impulse noise removal*, SIAM J. Imaging Sci., 10(3):1627–1667, 2017. 3.1, 3.2, 3.1

[75] Y. Zhao, C. Wu, Q. Dong, and Y. Zhao, *An accelerated majorization-minimization algorithm with convergence guarantee for non-Lipschitz wavelet synthesis model*, Inverse Probl., 38(1):015001, 2022. 1