LINEARLY CONSTRAINED NON-LIPSCHITZ OPTIMIZATION FOR IMAGE RESTORATION

WEI BIAN* AND XIAOJUN CHEN†

18 July 2015

Abstract. Nonsmooth nonconvex optimization models have been widely used in the restoration and reconstruction of real images. In this paper, we consider a linearly constrained optimization problem with a non-Lipschitz regularization term in the objective function which includes the l_p norm ($0 < p < 1$) of the gradient of the underlying image in the l_2-l_p problem as a special case. We prove that any cluster point of ϵ scaled first order stationary points satisfies a first order necessary condition for a local minimizer of the optimization problem as ϵ goes to 0. We propose a smoothing quadratic regularization (SQR) method for solving the problem. At each iteration of the SQR algorithm, a new iterate is generated by solving a strongly convex quadratic problem with linear constraints. Moreover, we show that the SQR algorithm can find an ϵ scaled first order stationary point in at most $O(\epsilon^{-2})$ iterations from any starting point. Numerical examples are given to show good performance of the SQR algorithm for image restoration.

Key words. Image restoration, total-variation regularization, non-Lipschitz optimization, smoothing quadratic regularization method, worst-case complexity,

AMS subject classifications. 90C30, 90C26, 65K05, 49M37

1. Introduction. In this paper, we consider the following constrained minimization problem

$$
\begin{aligned}
\min \quad & f(x) := \Theta(x) + \sum_{i=1}^{m} \varphi(|d_i^T x|^p) \\
\text{s.t.} \quad & x \in \Omega := \{x : Ax \le b\},
\end{aligned}
\tag{1.1}
$$

where $\Theta : \mathbb{R}^n \to \mathbb{R}_+$ is continuously differentiable, $0 < p \le 1$, $D = (d_1, d_2, \ldots, d_m)^T \in \mathbb{R}^{m \times n}$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous with $\varphi(0) = 0$, $A \in \mathbb{R}^{r \times n}$ and $b \in \mathbb{R}^r$. Problem (1.1) has many important applications in medical and astronomical image restoration [6, 12, 14, 23, 24, 36, 37, 38] and film restoration [29]. In (1.1), the first term measures how well the restored image is fitting the observed data under the imaging system, the second term induces special properties of the restored image, and the constraints can improve the restored image using a priori information.

Using a nonconvex nonsmooth non-Lipschitz regularization function in the second term of the objective function of (1.1) has remarkable advantages for the restoration of piecewise constant images [12, 23, 37]. Typical choices of D for the potential function include the identity matrix, first order difference operator, second order difference operator or some overcomplete dictionary [32]. Constrained optimization has been used in various applications with substantial improvements in the image restoration [1, 7, 31, 39]. For example, all gray level images with intensity values are ranging from 0 to 1. Based on the constrained TV-l_2 models, the numerical tests in [1] indicate that the peak signal-to-noise ratio (PSNR) can be improved more than 2 dB for some

special images by imposing the box constraint. The numerical experiments in [7] revel that the improvement can even be as big as 10.28bB for a special image by adding constraints. Moreover, most original images are comprised of entries in a nonnegative region. The linear constraint in (1.1) includes the box constraint and the nonnegative constraint as special cases.

The success of model (1.1) with $0 < p < 1$ in sparse optimization is related to the non-Lipschitz property of the objective function. Local minimizers of (1.1) with $0 < p < 1$ have various nice properties over the minimizers of (1.1) with $p = 1$. In image restoration, (i) it is shown in [8] that model (1.1) with $0 < p < 1$ promotes a better gradient sparsity than model (1.1) with $p = 1$; (ii) model (1.1) with $0 < p < 1$ is also robust with respect to noise; (iii) theoretical and numerical results show that local minimizers of (1.1) with $0 < p < 1$ have advantages in distinguishing zero and nonzero entries of coefficients in sparse high-dimensional approximation [2, 8, 9, 26] and bring the restored image closed contours and neat edges [12, 23]. Recently, Hintermuüller et. al [22] present new results of the non-Lipschitz TV^p with $(0 < 1 < p)$ in the function space setting for efficient restoration of real images. Moreover, in variable selection, the l_p potential function with $0 < p < 1$ owns the oracle property [18, 30] in statistics, while l_1 does not; the problem (1.1) with $0 < p < 1$ can be used for variable selection at the group and individual variable levels simultaneously, while (1.1) with $p = 1$ can only work for individual variable selection [27]. Thus, we focus on the minimization problem (1.1) with $0 < p < 1$ in this paper.

Theory and algorithms for some special cases of problem (1.1) with $0 < p < 1$ have been developed in the last few years. The lower bound theory [12, 13, 36, 38] ensures that each component of $|Dx|$ at any local minimizer of problem (1.1) is either zero or not less than a positive constant, which implies that the restored image has closed contours and neat edges in the imaging system. Moreover, various optimality conditions for local minimizers of non-Lipschitz optimization have been established [11, 13]. In particular, all local minimizers are the scaled first and second order stationary points which satisfy the first and second order necessary optimality conditions of non-Lipschitz optimization, respectively. For the unconstrained version of problem (1.1) with $(0 < p < 1)$, the reweighted algorithms [33, 34] are globally convergent to a stationary point. Moreover, the generalized Newton method using the R-regularization of the Hessian is superlinearly convergent to a stationary point, where the R-regularization is motivated by reweighting to handle the possible non-positive definiteness of the Hessian [24]. The trust region Newton method with subspace techniques converges to a scaled second order stationary point [11]. An effective majorize-minimize subspace algorithm is proposed in [15] for solving a class of unconstrained nonconvex regularization optimization in image computing problems, where the asymptotic connection to l_0-penalized problems and its convergence rate are discussed. For problem (1.1) with D being the identity matrix, the smoothing quadratic regularization (SQR) algorithm for (1.1) with $\Omega = \mathbb{R}^n$ and the interior point algorithm for (1.1) with the nonnegative constraint converge to an ϵ scaled first order stationary point with worst-case iteration complexity $O(\epsilon^{-2})$ [3, 4]. Moreover, when $\epsilon \to 0$, any cluster point of ϵ scaled first order stationary points is a scaled first order stationary point which satisfies a first order necessary condition for a local minimizer of the optimization problem. However, to the best of our knowledge, for problem (1.1) with an arbitrary matrix D in the regularization term and arbitrary linear constraint, there is no algorithm which can always find an ϵ scaled first order stationary point in no more than $O(\epsilon^{-2})$ iterations from any staring point.

The matrix D and linear constraints in (1.1) bring problem (1.1) many advantages in image restoration and reconstruction [11, 12, 28, 36, 37, 38, 39]. However, because the non-Lipschitz potential function in (1.1) is neither separable with respect to components of x nor concave in the feasible set, (1.1) is harder to solve than the problem with D being the identity matrix and with the nonnegative constraints. Algorithms in [3, 4, 11, 24, 33, 34, 38] cannot be directly extended to solve (1.1), and the lower bound theory in [12, 13, 36, 38] and the definitions of the ϵ scaled first order stationary point in [3, 4] are invalid. Thus, (1.1) gives many challenging problems in developing effective algorithms with desired convergence theorems and computational complexity bounds.

In this paper, we generalize the subspace idea for unconstrained optimization in [11] and define the first order necessary optimality condition and an ϵ scaled first order stationary point of problem (1.1). We prove that any cluster point of ϵ scaled first order stationary points satisfies the first order necessary optimality condition as ϵ tending to 0. Moreover, a new SQR algorithm with the worst-case complexity $O(\epsilon^{-2})$ is given for solving (1.1), where the updating techniques in the algorithm for approximating the Lipschitz constant of the gradient of Θ is adopted from [5].

The rest of this paper is organized as follows. In section 2, we derive a first order necessary condition for a local minimizer of (1.1), and define an ϵ scaled first order stationary point of (1.1). We prove that any cluster point of ϵ scaled first order stationary points of (1.1) satisfies the first order necessary condition as ϵ tends to 0. In section 3, using a smoothing function of the objective function f in (1.1), we present an SQR algorithm for solving (1.1). In section 4, we show that the SQR algorithm can find an ϵ scaled first order stationary point of (1.1) in at most $O(\epsilon^{-2})$ iterations. In section 5, we report numerical experiments with one randomly generated test example and four image restoration problems to validate the theoretical results and show good performance of the proposed SQR algorithm for image problems.

Notations: Denote $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and $e = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$. For $x, y \in \mathbb{R}^n$, let $\|x\| := \|x\|_2$, $x \leq y$ means $x_i \leq y_i$, $i = 1, 2, \ldots, n$. For a matrix $M \in \mathbb{R}^{r \times n}$ and an index set $J \subseteq \{1, \ldots, r\}$, M_J denotes the submatrix of M whose rows are indexed by J. For $\Pi \subseteq \mathbb{R}^n$ consisted by a class of column vectors of $\mathbb{R}^n$, spanΠ indicates the subspace of $\mathbb{R}^n$ spanned by the elements in Π, and $(\text{span}\Pi)^\perp$ is its orthogonal complement space. For a subspace $\mathbb{S} \subseteq \mathbb{R}^n$, orthon$(\mathbb{S})$ is a subset of $\mathbb{R}^{n \times \dim(S)}$, in which the columns of each matrix form an orthonormal basis of $\mathbb{S}$ if $\dim(\mathbb{S}) \geq 1$ and orthon$(\mathbb{S}) = 0_{n \times 1}$ if $\mathbb{S} = \{0\}$.

2. First order necessary condition. The scaled first order and second order necessary conditions for unconstrained non-Lipschitz optimization have been studied in [11, 13]. For the constrained non-Lipschitz optimization (1.1) with $D = I_n$, the scaled first and second order stationary points are defined in [4, 20]. Inspired by the subspace idea, we first derive a first order necessary condition for local minimizers of constrained non-Lipschitz optimization (1.1), whereafter the scaled and ϵ scaled first order stationary points of (1.1) are defined. Note that the results established in this paper have no assumption on the matrices D and A.

First, we give some notations used in the following. For fixed $x \in \mathbb{R}^n$ and $\epsilon > 0$, denote

$$\mathcal{D}_x^\epsilon = \{d_i : i \in \{1, 2, \ldots, m\}, \ |d_i^T x| \leq \epsilon\}.$$

For simplicity, we denote $\mathcal{D}_x := \mathcal{D}_x^\epsilon$ when $\epsilon = 0$. Obviously, $\dim(\text{span}\mathcal{D}_{\bar{x}}) = n$ implies $\bar{x} = 0$, and f is non-Lipschitz at $\bar{x}$ if $\mathcal{D}_{\bar{x}}$ is nonempty.

Now, we derive a first order necessary condition for local minimizers of (1.1) using two matrices whose columns form orthonormal basis of $\text{span}\mathcal{D}_{\bar{x}}$ and $(\text{span}\mathcal{D}_{\bar{x}})^{\perp}$, namely,

$$Y_{\bar{x}} \in \text{orthon}(\text{span}\mathcal{D}_{\bar{x}}) \quad \text{and} \quad Z_{\bar{x}} \in \text{orthon}((\text{span}\mathcal{D}_{\bar{x}})^{\perp}).$$

From the definitions of $\mathcal{D}_{\bar{x}}$ and $Z_{\bar{x}}$, we have $Z_{\bar{x}}^T d_{\bar{x}} = 0 \; \forall d_{\bar{x}} \in \text{span}\mathcal{D}_{\bar{x}}$ and there is a unique vector $\bar{z}$ such that

$$\bar{x} = Z_{\bar{x}}\bar{z} \quad \text{and} \quad \bar{z} = Z_{\bar{x}}^T \bar{x}. \tag{2.1}$$

In what follows, we denote

$$\Phi(t) = \begin{cases} p\varphi'(s)_{s=|t|^p}|t|^{p-1}\text{sign}(t) & \text{if } t \neq 0 \\ 0 & \text{if } t = 0, \end{cases}$$

which is the first order derivative of $\varphi(|t|^p)$ with respect to t when $t \neq 0$.

LEMMA 2.1. *If $\bar{x}$ is a local minimizer of (1.1), there exists a nonnegative vector $\gamma \in \mathbb{R}^r$ such that $\bar{x}$ satisfies*

$$Z_{\bar{x}}^T(\nabla\Theta(\bar{x}) + \sum_{i=1}^m \Phi(d_i^T\bar{x})d_i + A^T\gamma) = 0, \tag{2.2a}$$

$$A\bar{x} - b \leq 0, \; (A\bar{x} - b)^T\gamma = 0. \tag{2.2b}$$

Proof. If $\dim(\text{span}\mathcal{D}_{\bar{x}}) = n$, then $\bar{x} = 0$ and $Z_{\bar{x}} = 0_{n\times 1}$, which means the conditions in this lemma naturally holds for $\bar{x}$ with $\gamma = 0_{r\times 1}$.

Now, we suppose $\dim(\text{span}\mathcal{D}_{\bar{x}}) = n - t < n$, then $Z_{\bar{x}} \in \mathbb{R}^{n\times t}$ is a nonzero matrix and there is an $\eta_{\bar{x}} > 0$ such that

$$f(\bar{x}) = \min_x \{f(x) : Ax - b \leq 0, \|x - \bar{x}\| \leq \eta_{\bar{x}}\}$$

$$= \min_{y,z} \{\Theta(Y_{\bar{x}}y + Z_{\bar{x}}z) + \sum_{i=1}^m \varphi(|d_i^T(Y_{\bar{x}}y + Z_{\bar{x}}z)|^p) :$$

$$A(Y_{\bar{x}}y + Z_{\bar{x}}z) - b \leq 0, \|Y_{\bar{x}}y + Z_{\bar{x}}z - Z_{\bar{x}}\bar{z}\| \leq \eta_{\bar{x}}\}$$

$$\leq \min_z \{\Theta(Y_{\bar{x}}0 + Z_{\bar{x}}z) + \sum_{i=1}^m \varphi(|d_i^T(Y_{\bar{x}}0 + Z_{\bar{x}}z)|^p) :$$

$$A(Y_{\bar{x}}0 + Z_{\bar{x}}z) - b \leq 0, \|Y_{\bar{x}}0 + Z_{\bar{x}}z - Z_{\bar{x}}\bar{z}\| \leq \eta_{\bar{x}}\}$$

$$= \min_z \{\Theta(Z_{\bar{x}}z) + \sum_{i=1}^m \varphi(|d_i^T Z_{\bar{x}}z|^p) : AZ_{\bar{x}}z - b \leq 0, \|Z_{\bar{x}}z - Z_{\bar{x}}\bar{z}\| \leq \eta_{\bar{x}}\}$$

$$= \min_z \{\Theta(Z_{\bar{x}}z) + \sum_{d_i^T\bar{x}\neq 0} \varphi(|d_i^T Z_{\bar{x}}z|^p) : AZ_{\bar{x}}z - b \leq 0, \|Z_{\bar{x}}z - Z_{\bar{x}}\bar{z}\| \leq \eta_{\bar{x}}\},$$

where the last equality uses $d_i^T Z_{\bar{x}}z = 0, \; \forall z \in \mathbb{R}^t, \; d_i \in \mathcal{D}_{\bar{x}}$.

In what follows, we will find the first order necessary condition for local minimizers of (1.1) from the reduced optimization problem in $\mathbb{R}^t$:

$$\min \quad v(z) = \Theta(Z_{\bar{x}}z) + \sum_{d_i^T\bar{x}\neq 0} \varphi(|d_i^T Z_{\bar{x}}z|^p), \quad \text{s.t.} \quad AZ_{\bar{x}}z - b \leq 0, \tag{2.3}$$

4

where $v(z)$ is continuously differentiable and its gradient is locally Lipschitz continuous around $\bar{z}$.

By (2.1) and (2.3),

$$v(\bar{z}) = \Theta(Z_{\bar{x}}\bar{z}) + \sum_{d_i^T \bar{x} \neq 0} \varphi(|d_i^T Z_{\bar{x}}\bar{z}|^p) = f(\bar{x}).$$

Therefore, $v(\bar{z}) \leq \min_z \{v(z) : AZ_{\bar{x}}z - b \leq 0, \|Z_{\bar{x}}(z - \bar{z})\| \leq \eta_{\bar{x}}\}$.

Since $Z_{\bar{x}} \in \mathbb{R}^{n \times t}$ is of full column rank, $\bar{z}$ is a local minimizer of (2.3). By the KKT condition for a local minimizer of (2.3), there exists a nonnegative vector $\gamma \in \mathbb{R}^r$ such that $\bar{z}$ satisfies

$$\nabla v(\bar{z}) + Z_{\bar{x}}^T A^T \gamma = 0, \tag{2.4a}$$

$$AZ_{\bar{x}}\bar{z} - b \leq 0, \ (AZ_{\bar{x}}\bar{z} - b)^T \gamma = 0. \tag{2.4b}$$

By (2.1), (2.2b) can be obtained from (2.4b), and

$$\begin{aligned}
\nabla v(\bar{z}) &= Z_{\bar{x}}^T (\nabla\Theta(y)_{y=Z_{\bar{x}}\bar{z}} + \sum_{d_i^T \bar{x} \neq 0} \Phi(d_i^T Z_{\bar{x}}\bar{z})d_i) \\
&= Z_{\bar{x}}^T (\nabla\Theta(\bar{x}) + \sum_{d_i^T \bar{x} \neq 0} \Phi(d_i^T \bar{x})d_i),
\end{aligned} \tag{2.5}$$

which together with (2.4a) and the definition on $\Phi(d_i^T \bar{x})$ gives (2.2a). $\square$

In view of the first order necessary condition for local minimizers of (1.1) given in Lemma 2.1, we define the scaled and ϵ scaled first order stationary points of (1.1).

DEFINITION 2.2. *We call $\bar{x}$ a scaled first order stationary point of (1.1), if there exists a nonnegative vector $\gamma \in \mathbb{R}^r$ such that $\bar{x}$ satisfies (2.2) in Lemma 2.1.*

DEFINITION 2.3. *For $\epsilon > 0$, we call x^ϵ an ϵ scaled first order stationary point of (1.1), if there exists a nonnegative vector $\gamma^\epsilon \in \mathbb{R}^r$ such that x^ϵ satisfies*

$$\left\| (Z_{x^\epsilon}^\epsilon)^T (\nabla\Theta(x^\epsilon) + \sum_{i=1}^m \Phi(d_i^T x^\epsilon)d_i + A^T \gamma^\epsilon) \right\|_\infty \leq \epsilon, \tag{2.6a}$$

$$Ax^\epsilon - b \leq 0, \ -\epsilon \leq (Ax^\epsilon - b)^T \gamma^\epsilon \leq 0, \tag{2.6b}$$

where $Z_{x^\epsilon}^\epsilon \in \text{orthon}((\text{span}\mathcal{D}_{x^\epsilon}^\epsilon)^\perp)$.

Definitions 2.2 and 2.3 are consistent at $\epsilon = 0$. The next proposition validates this consistence for ϵ tending to 0, which gives some hints on how to find a scaled first order stationary point of (1.1).

PROPOSITION 2.4. *Let x^ϵ be an ϵ ($\epsilon > 0$) scaled first order stationary point of (1.1). Then any cluster point of x^ϵ is a scaled first order stationary point of (1.1) as $\epsilon \to 0$.*

Proof. Suppose $\bar{x}$ is a limit point of $\{x^k\}$ as k tending to ∞, where x^k is an ϵ_k scaled first order stationary point of (1.1) and $\lim_{k\to\infty} \epsilon_k = 0$.

If $\dim(\text{span}\mathcal{D}_{\bar{x}}) = n$, then $\bar{x} = 0$ and $Z_{\bar{x}} = 0_{n\times 1}$, which implies that $\bar{x}$ is a scaled first order stationary point. In what follows, we suppose that $\dim(\text{span}\mathcal{D}_{\bar{x}}) < n$.

First, we prove that there is $k_{\bar{x}} \in \mathbb{N}_0$ such that $\mathcal{D}_k \subseteq \mathcal{D}_{\bar{x}} \ \forall k \geq k_{\bar{x}}$, where $\mathcal{D}_k := \mathcal{D}_{x^k}^{\epsilon_k}$. If not, there is a subsequence $\{x^{k_j}\} \subseteq \{x^k\}$ such that $\lim_{j\to\infty} \epsilon_{k_j} = 0$ and $\mathcal{D}_{k_j} \not\subseteq \mathcal{D}_{\bar{x}}$ for all j, by $\mathcal{D}_{k_j} \subseteq \{d_1, d_2, \ldots, d_m\}$, there is an element $d \in \mathbb{R}^n$ and a subsequence of $\{k_j\}$ (also denoted as $\{k_j\}$) such that $d \in \mathcal{D}_{k_j}$ but $d \notin \mathcal{D}_{\bar{x}}$. Then,

$|d^T x^{k_j}| \leq \epsilon_{k_j}$, letting j tend to ∞, we have $|d^T \bar{x}| = 0$, which leads a contradiction with $d \notin \mathcal{D}_{\bar{x}}$.

Denote $Z_k \in \operatorname{orthon}((\operatorname{span}\mathcal{D}_k)^{\perp})$. Then, we can find matrices Z_k and $Z_{\bar{x}}$ such that Z_k contains all columns of $Z_{\bar{x}}$ for all $k \geq k_{\bar{x}}$.

By (2.6a), there is a nonnegative vector $\gamma^k \in \mathbb{R}^r$ such that

$$\left\| Z_k^T \left(\nabla \Theta(x^k) + \sum_{i=1}^m \Phi(d_i^T x^k) d_i + A^T \gamma^k \right) \right\|_\infty \leq \epsilon_k,$$

from the inclusion property between $Z_{\bar{x}}$ and Z_k, which gives

$$\left\| Z_{\bar{x}}^T \left(\nabla \Theta(x^k) + \sum_{i=1}^m \Phi(d_i^T x^k) d_i + A^T \gamma^k \right) \right\|_\infty \leq \epsilon_k.$$

By the definition of $Z_{\bar{x}}$, we have

$$\left\| Z_{\bar{x}}^T \left(\nabla \Theta(x^k) + \sum_{d_i^T \bar{x} \neq 0} \Phi(d_i^T x^k) d_i + A^T \gamma^k \right) \right\|_\infty \leq \epsilon_k. \tag{2.7}$$

For $i \in \{i : d_i^T \bar{x} \neq 0\}$, we obtain

$$\lim_{k \to \infty} \Phi(d_i^T x^k) = \Phi(d_i^T \bar{x}). \tag{2.8}$$

Moreover, from (2.6b), we have

$$\lim_{k \to \infty} (Ax^k - b)^T \gamma^k = 0. \tag{2.9}$$

Then, $\lim_{k \to \infty} \gamma_j^k = 0$, $\forall j \notin J_{\bar{x}} = \{j \in \{1, 2, \ldots, r\} : [A\bar{x}]_j - b_j = 0\}$. If $J_{\bar{x}} = \emptyset$, letting k tend to ∞ in (2.7), (2.8) and (2.9) imply that $\bar{x}$ satisfies (2.2) with $\gamma = 0$.

We suppose $J_{\bar{x}} \neq \emptyset$ and let $J_{\bar{x}} = \{q+1, \ldots, r\}$ without loss of generality. Letting $k \to \infty$ in (2.7), by (2.8), we have

$$-\lim_{k \to \infty} Z_{\bar{x}}^T A^T \gamma^k = Z_{\bar{x}}^T \left(\nabla \Theta(\bar{x}) + \sum_{i=1}^m \Phi(d_i^T \bar{x}) d_i \right),$$

which follows

$$-\lim_{k \to \infty} [AZ_{\bar{x}}]_{J_{\bar{x}}}^T \gamma_{J_{\bar{x}}}^k = Z_{\bar{x}}^T \left(\nabla \Theta(\bar{x}) + \sum_{i=1}^m \Phi(d_i^T \bar{x}) d_i \right). \tag{2.10}$$

Consider the quadratic programming

$$\min_{y \geq 0} \left\| [AZ_{\bar{x}}]_{J_{\bar{x}}}^T y + Z_{\bar{x}}^T \left(\nabla \Theta(\bar{x}) + \sum_{i=1}^m \Phi(d_i^T \bar{x}) d_i \right) \right\|^2. \tag{2.11}$$

By the Frank-Wolfe theorem [19], a global minimizer y^* of (2.11) exists, and thus

$$\| [AZ_{\bar{x}}]_{J_{\bar{x}}}^T y^* + Z_{\bar{x}}^T (\nabla \Theta(\bar{x}) + \sum_{i=1}^m \Phi(d_i^T \bar{x}) d_i) \|^2 \leq \| [AZ_{\bar{x}}]_{J_{\bar{x}}}^T \gamma_{J_{\bar{x}}}^k + Z_{\bar{x}}^T (\nabla \Theta(\bar{x}) + \sum_{i=1}^m \Phi(d_i^T \bar{x}) d_i) \|^2.$$

6

Passing k to ∞ in the above equation and by (2.10), we have

$$\|[AZ_{\bar{x}}]^T_{J_{\bar{x}}} y^* + Z^T_{\bar{x}}(\nabla\Theta(\bar{x}) + \sum_{i=1}^{m} \Phi(d_i^T\bar{x})d_i)\| = 0,$$

which implies that there exists a nonnegative vector $y^* \in \mathbb{R}^{r-q}$ such that

$$-[AZ_{\bar{x}}]^T_{J_{\bar{x}}} y^* = Z^T_{\bar{x}}(\nabla\Theta(\bar{x}) + \sum_{i=1}^{m} \Phi(d_i^T\bar{x})d_i).$$

Hence $\bar{x}$ satisfies (2.2) with the nonnegative vector $\gamma \in \mathbb{R}^r$, where $\gamma_{J_{\bar{x}}} = y^*$ and the other elements of γ are 0. This implies that $\bar{x}$ is a scaled first order stationary point of (1.1). $\square$

3. Smoothing quadratic regularization method. In this section, we present an SQR method for solving (1.1). The objective function is not Lipschitz continuous at points in $\{x : d_i^T x = 0, \text{ for some } i \in \{1, \ldots, m\}\}$. Throughout this paper, we need the following assumptions on Θ and φ:

- $\Theta : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and its gradient $\nabla\Theta$ is globally Lipschitz with a Lipschitz constant $\hat{\beta}$;
- φ is differentiable and concave in $(0, \infty)$, φ' is globally Lipschitz continuous and there is a positive constant α such that for all $t \in (0, \infty)$,

$$0 \leq \varphi'(t) \leq \alpha \quad \text{and} \quad |\xi| \leq \alpha \quad \forall \xi \in \partial(\varphi'(t)), \qquad (3.1)$$

where

$$\partial(\varphi'(t)) = \mathrm{con}\{v \,|\, \varphi''(s) \to v, \, \varphi \text{ is twice differentiable at } s, \; s \to t\}, \qquad (3.2)$$

and "con" means the convex hull [16].

Many data fitting functions and penalty functions in sparse image restoration and reconstruction satisfy these conditions [12, 14, 23, 36, 37, 38].

3.1. Smoothing approximation. One of difficulties for solving (1.1) comes from the nonsmoothness and nonconvexity of the objective function f, which let the usual gradient-based methods be inappropriate. The approximation idea is often used for image restoration problems [12, 37, 38]. The approximate energy minimization in [37] can be split into a total variational-type regularization term and a smooth nonconvex term. The approximate energy function in [38] is with an additional variable, which can be separated to a twice continuously differentiable problem and a TV denoising problem.

In this paper, we approximate f by a smoothing function $\tilde{f} : \mathbb{R}^n \times (0, \infty) \to \mathbb{R}$, which is continuously differentiable with respect to x for any fixed parameter $\mu > 0$ and $\lim_{z\to x, \mu\downarrow 0} \tilde{f}(z, \mu) = f(x)$ holds for any $x \in \mathbb{R}^n$. We can construct a quadratic approximation function of f by using the gradient of the smoothing function $\tilde{f}(x, \mu)$, where the continuous differentiability of φ on $(0, \infty)$ is sufficient to support the algorithm proposed in this paper.

According to the assumptions on Θ and φ, we define $\tilde{f}(x, \mu)$ by using a smoothing function of the absolute value function as follows

$$\tilde{f}(x, \mu) = \Theta(x) + \sum_{i=1}^{m} \varphi(\theta^p(d_i^T x, \mu)) \quad \text{with} \quad \theta(t, \mu) = \begin{cases} |t| & \text{if } |t| > \mu \\ \dfrac{t^2}{2\mu} + \dfrac{\mu}{2} & \text{if } |t| \leq \mu. \end{cases}$$

7

The function $\theta(t, \mu)$ is continuously differentiable with respect to t for any fixed $\mu > 0$, nondecreasing with respect to μ, and

$$0 = \arg\min_{t \in \mathbb{R}} |t| = \arg\min_{t \in \mathbb{R}} \theta(t, \mu) = \arg\min_{t \in \mathbb{R}} \varphi(\theta^p(t, \mu)), \quad \forall \mu \in (0, \infty).$$

Moreover, we have

$$|\nabla_t \theta^p(t, \mu)| \leq p\theta^{p-1}(t, \mu), \ |\nabla_t^2 \theta^p(t, \mu)| \leq p\theta^{p-2}(t, \mu) \text{ when } |t| \neq \mu. \quad (3.3)$$

Specially, when $|t| < \mu$, $\nabla_t^2 \theta^p(t, \mu) > 0$, which means that $\theta^p(t, \mu)$ is a convex smoothing function of $|t|^p$ in $(-\mu, \mu)$.

Since $\varphi'(t) \geq 0$ for all $t \in (0, \infty)$, $\tilde{f}(x, \mu)$ is nondecreasing with respect to μ for any fixed $x \in \mathbb{R}^n$. Denote

$$g(x, \mu) := \nabla_x \tilde{f}(x, \mu) = \nabla\Theta(x) + \sum_{i=1}^{m} \nabla_t \varphi(\theta^p(t, \mu))_{t=d_i^T x} d_i. \quad (3.4)$$

From $0 \leq \theta^p(t, \mu) - |t|^p \leq \theta^p(0, \mu) = (\frac{\mu}{2})^p$ and (3.1), we have

$$0 \leq \varphi(\theta^p(t, \mu)) - \varphi(|t|^p) \leq \alpha(\frac{\mu}{2})^p,$$

which gives

$$0 \leq \tilde{f}(x, \mu) - f(x) \leq \sum_{|d_i^T x| < \mu} \alpha(\frac{\mu}{2})^p \quad \forall x \in \mathbb{R}^n, \ \mu \in (0, \infty). \quad (3.5)$$

Denote

$$\mathcal{O} = \{t : \varphi \text{ is not twice differentiable at } \theta^p(t, \mu) \text{ or } |t| = \mu\}.$$

Then, when $t \notin \mathcal{O}$, $\varphi(\theta^p(t, \mu))$ is twice differentiable at t and

$$\nabla_t^2 \varphi(\theta^p(t, \mu)) = \varphi''(s)_{s=\theta^p(t, \mu)}(\nabla_t \theta^p(t, \mu))^2 + \varphi'(s)_{s=\theta^p(t, \mu)} \nabla_t^2 \theta^p(t, \mu). \quad (3.6)$$

For any fixed $\mu \in (0, \infty)$, $\theta^p(t, \mu)$ is twice differentiable when $|t| \neq \mu$ and strictly increasing with respect to t in $\mathbb{R}_+$ and $\mathbb{R}_-$, respectively. Combining this with the Lipschitz property of φ', the measure of $\mathcal{O}$ is 0. Thus, by (3.2), (3.3), (3.6) and the assumptions on φ, the following estimation on the elements in $\partial_t(\nabla_t \varphi(\theta^p(t, \mu)))$ holds for all $t \in \mathbb{R}$ and $\mu \in (0, 1]$

$$\max\{|\xi| : \xi \in \partial_t(\nabla_t \varphi(\theta^p(t, \mu)))\} \leq \alpha p^2 \theta^{2p-2}(t, \mu) + \alpha p\theta^{p-2}(t, \mu)$$
$$\leq \alpha p^2 (\frac{\mu}{2})^{2p-2} + \alpha p (\frac{\mu}{2})^{p-2} \leq 8\alpha p \mu^{p-2}.$$

Inspired by Taylor's expansion, for any $x^+, x \in \mathbb{R}^n$,

$$\varphi(\theta^p(d_i^T x^+, \mu)) - \varphi(\theta^p(d_i^T x, \mu))$$
$$\leq \nabla_t \varphi(\theta^p(t, \mu))_{t=d_i^T x} d_i^T(x^+ - x) + 4\alpha p\mu^{p-2}(d_i^T(x^+ - x))^2. \quad (3.7)$$

Notice that $\varphi(|t|^p)$ is concave on $\mathbb{R}_+$ and $\mathbb{R}_-$, respectively. We have

$$\varphi(|\hat{t}|^p) \leq \varphi(|t|^p) + \nabla\varphi(|t|^p)(\hat{t} - t)$$

8

for any $\hat{t}, t \in \mathbb{R}$ such that $\hat{t}t > 0$. Hence, for any $x, x^+ \in \mathbb{R}^n$ satisfying

$$(d_i^T x^+)(d_i^T x) > 0, \ |d_i^T x^+| \geq \mu \text{ and } |d_i^T x| \geq \mu,$$

by $\theta(s, \mu) = |s|$ when $|s| \geq \mu$, we have

$$\varphi(\theta^p(d_i^T x^+, \mu)) \leq \varphi(\theta^p(d_i^T x, \mu)) + \nabla_t \varphi(\theta^p(t, \mu))_{t=d_i^T x} d_i^T (x^+ - x). \quad (3.8)$$

LEMMA 3.1. *For $x \in \Omega$, $s \in \mathbb{R}^n$ and $\mu \in (0, 1]$, if $-\|D\|_\infty^{-1} \mu^p e \leq s \leq \|D\|_\infty^{-1} \mu^p e$ and the following inequality holds*

$$\Theta(x + s) \leq \Theta(x) + \langle \nabla\Theta(x), s \rangle + \frac{\beta}{2} \|s\|^2 \quad (3.9)$$

with $\beta > 0$, then

$$\tilde{f}(x + s, \mu) - \tilde{f}(x, \mu) \leq \langle g(x, \mu), s \rangle + \frac{\beta}{2} \|s\|^2 + 4\alpha p \mu^{p-2} \sum_{|d_i^T x| \leq 2\mu^p} (d_i^T s)^2. \quad (3.10)$$

Proof. From $-\|D\|_\infty^{-1} \mu^p e \leq s \leq \|D\|_\infty^{-1} \mu^p e$, we have

$$|d_i^T s| \leq \|Ds\|_\infty \leq \|D\|_\infty \|s\|_\infty \leq \mu^p, \quad i = 1, 2, \ldots, m.$$

Then, $|d_i^T x| > 2\mu^p$ implies $|d_i^T (x + s)| > \mu^p$ and $(d_i^T x)(d_i^T (x + s)) > 0$ for $i = 1, 2, \ldots, m$, which together with (3.7) and (3.8) gives

$$\sum_{i=1}^m \varphi(\theta^p(d_i^T (x + s), \mu)) - \sum_{i=1}^m \varphi(\theta^p(d_i^T x, \mu))$$

$$\leq \langle \sum_{i=1}^m \nabla_t \varphi(\theta^p(t, \mu))_{t=d_i^T x}, d_i^T s \rangle + 4\alpha p \mu^{p-2} \sum_{|d_i^T x| \leq 2\mu^p} (d_i^T s)^2.$$

Thus, we obtain (3.10) based on (3.9). $\square$

From the assumption on Θ, (3.9) holds with $\beta = \hat{\beta}$. However, in general, the Lipschitz constant $\hat{\beta}$ of $\nabla\Theta$ is difficult to evaluate. We use β as an approximation of $\hat{\beta}$ and update β in the algorithm.

For $x \in \Omega$ and $\mu \in (0, 1]$, to achieve a potential reduction, we solve the following strongly convex quadratic program in $\mathbb{R}^n$ with linear constraints:

$$\begin{aligned} \min \quad & \langle g(x, \mu), s \rangle + \frac{\beta}{2} \|s\|^2 + 4\alpha p \mu^{p-2} \sum_{|d_i^T x| \leq 2\mu^p} (d_i^T s)^2 \\ \text{s.t.} \quad & -\delta \mu^p e \leq s \leq \delta \mu^p e, \ A(x + s) - b \leq 0, \end{aligned} \quad (3.11)$$

where $\delta = \|D\|_\infty^{-1}$ and $g(x, \mu)$ is defined in (3.4).

9

SQR Algorithm.
Step 0: Initialization: Choose $x^0 \in \Omega$, $0 < \mu_0 \leq 1$, $\beta_0 \geq 1$, $0 < \sigma < 1$ and $\eta > 1$. Set $k = 0$.
Step 1: New point calculation: Solve (3.11) with $x = x^k$, $\mu = \mu_k$ and $\beta = \beta_k$ for s^k, and let $y^k = x^k + s^k$.
Step 2: Updating the regularization weight: If

$$\Theta(y^k) - \Theta(x^k) > \langle \nabla\Theta(x^k), y^k - x^k \rangle + \frac{1}{2}\beta_k \|x^k - y^k\|^2,$$

let

$$\beta_{k+1} = \eta\beta_k, \ x^{k+1} = x^k, \ \mu_{k+1} = \mu_k$$

and return to Step 1; otherwise, let

$$\beta_{k+1} = \beta_k, \ x^{k+1} = y^k$$

and go to Step 3.
Step 3: Updating the smoothing parameter: Let

$$\mu_{k+1} = \begin{cases} \mu_k & \text{if } \tilde{f}(x^{k+1}, \mu_k) - \tilde{f}(x^k, \mu_k) < -\mu_k^{2p} \\ \sigma\mu_k & \text{otherwise.} \end{cases} \qquad (3.12)$$

Step 4: Constructing convergence sequence: Let

$$z^{k+1} = \begin{cases} x^k & \text{if } \mu_{k+1} = \sigma\mu_k \\ z^k & \text{otherwise.} \end{cases} \qquad (3.13)$$

Increment k by one and return to Step 1.

The proposed SQR algorithm is for the linearly constrained optimization problem (1.1) with an arbitrary D in the potential function φ, while the SQR method in [3] can only be applied to the unconstrained problem with $\varphi(|d_i^T x|^p) = \varphi(|x_i|^p)$. Hence, the two SQR algorithms in this paper and [3] are different through the framework of the SQR algorithm in this paper is adopted from [3]. The construction of the quadratic subproblem and the updating scheme of the smoothing parameter are entirely different in the two algorithms. In [3], the quadratic program is an unconstrained problem and can be split into n one dimensional problems to get a simple closed form solution. In this paper, the quadratic program is a constrained problem and cannot have a closed form solution. Moreover, the convergence and complexity analysis in [3] uses the separability of variables in the term $\sum_{i=1}^{n} \varphi(|x_i|^p)$ without considering the feasibility of iterates. The convergence and complexity analysis for the SQR algorithm in this paper is more comprehensive due to the existence of linear constraints and the arbitrary D in the non-Lipschitz potential function.

From Lemma 3.1, for $x^k \in \Omega$ and $\mu_k \in (0, 1]$, by $A(x^k + s^k) \leq b$ and $\sigma \in (0, 1)$, we have $x^{k+1} \in \Omega$ and $\mu_{k+1} \in (0, 1]$. Then, the proposed SQR algorithm is well defined, and $x^k \in \Omega$, $\mu_k \in (0, 1]$ for all $k \in \mathbb{N}_0$.

Let $\{x^k\}$, $\{y^k\}$, $\{z^k\}$, $\{\mu_k\}$ and $\{\beta_k\}$ be the sequences generated by the SQR

10

algorithm. Denote

$$N_s = \{k \in \mathbb{N}_0 : \beta_{k+1} = \beta_k\}, \quad \mathcal{T} = \{k \in \mathbb{N}_0 : \ \mu_{k+1} = \sigma\mu_k\}. \tag{3.14}$$

We call the kth iteration is successful if $k \in N_s$. Note that $\mathcal{T} \subseteq N_s$.

LEMMA 3.2. *The sequence $\{\tilde{f}(x^k, \mu_k)\}$ is non-increasing. Moreover, when $k \in N_s$, there are nonnegative vectors ν^k and γ^k such that*

$$\tilde{f}(x^{k+1}, \mu^{k+1}) - \tilde{f}(x^k, \mu^k) \leq -\frac{\beta_k}{2}\|s^k\|^2 - 4\alpha p\mu_k^{p-2} \sum_{|d_i^T x^k| \leq 2\mu_k^p} (d_i^T s^k)^2$$
$$- \mu_k^p e^T \nu^k + (\gamma^k)^T(Ax^k - b).$$

Proof. From the KKT condition of (3.11), the solution s^k of the strongly convex quadratic program (3.11) satisfies

$$g(x^k, \mu_k) + \beta_k s^k + 8\alpha p\mu_k^{p-2} \sum_{|d_i^T x^k| \leq 2\mu_k^p} (d_i^T s^k)d_i - \varrho_1^k + \varrho_2^k + A^T\gamma^k = 0, \tag{3.15a}$$

$$A(x^k + s^k) \leq b, \ (As^k + Ax^k - b)^T\gamma^k = 0, \tag{3.15b}$$

$$-\delta\mu_k^p e \leq s^k \leq \delta\mu_k^p e, \ (s^k + \delta\mu_k^p e)^T\varrho_1^k = 0, \ (s^k - \delta\mu_k^p e)^T\varrho_2^k = 0, \tag{3.15c}$$

with $\varrho_1^k, \varrho_2^k, \gamma^k \geq 0$.

On the one hand, when $k \in N_s$, $x^{k+1} = y^k$, from Lemma 3.1 and (3.15), we have

$$\tilde{f}(x^{k+1}, \mu_k) - \tilde{f}(x^k, \mu_k)$$

$$\leq \langle g(x^k, \mu_k), s^k \rangle + \frac{\beta_k}{2}\|s^k\|^2 + 4\alpha p\mu^{p-2} \sum_{|d_i^T x^k| \leq 2\mu_k^p} (d_i^T s^k)^2$$

$$= -\frac{\beta_k}{2}\|s^k\|^2 - 4\alpha p\mu_k^{p-2} \sum_{|d_i^T x^k| \leq 2\mu_k^p} (d_i^T s^k)^2 + (\varrho_1^k)^T s^k - (\varrho_2^k)^T s^k - (\gamma^k)^T As^k$$

$$= -\frac{\beta_k}{2}\|s^k\|^2 - 4\alpha p\mu_k^{p-2} \sum_{|d_i^T x^k| \leq 2\mu_k^p} (d_i^T s^k)^2 - \delta\mu_k^p e^T(\varrho_1^k + \varrho_2^k) + (\gamma^k)^T(Ax^k - b) \leq 0. \tag{3.16}$$

Since $\mu_{k+1} \leq \mu_k$ and $\tilde{f}(x, \mu)$ is non-decreasing about μ for any fixed $x \in \mathbb{R}^n$, the results in this lemma holds for $k \in N_s$, where $\nu^k = \delta(\varrho_1^k + \varrho_2^k)$ is a nonnegative vector.

On the other hand, when $k \notin N_s$, $x^{k+1} = x^k$ and $\mu_{k+1} = \mu_k$, which implies $\tilde{f}(x^{k+1}, \mu_{k+1}) = \tilde{f}(x^k, \mu_k)$. Thus, $\tilde{f}(x^k, \mu_k)$ is nonincreasing for all $k \in \mathbb{N}_0$. □

LEMMA 3.3. *For all $k \in N_s$, if*

$$\tilde{f}(x^{k+1}, \mu_k) - \tilde{f}(x^k, \mu_k) > -\mu_k^{2p}, \tag{3.17}$$

then there exists a nonnegative vector $\gamma^k \in \mathbb{R}^r$ such that x^k satisfies

$$\left\|Z_k^T\left(\nabla\Theta(x^k) + \sum_{i=1}^m \Phi(d_i^T x^k)d_i + A^T\gamma^k\right)\right\|_\infty \leq (\sqrt{2\beta_k} + \delta^{-1})\mu_k^p,$$
$$Ax^k - b \leq 0, \quad -\mu_k^{2p} \leq (Ax^k - b)^T\gamma^k \leq 0, \tag{3.18}$$

where $Z_k \in$ orthon$((\text{span}\{d_i : |d_i^T x^k| \leq 2\mu_k^p\})^\perp)$.

11

Proof. If one of the following four inequalities fails

$$\frac{\beta_k}{2}\|s^k\|^2 < \mu_k^{2p}, \quad 4\alpha p \mu_k^{p-2} \sum_{|d_i^T x^k| \le 2\mu_k^p} (d_i^T s^k)^2 < \mu_k^{2p}, \qquad (3.19a)$$

$$\mu_k^p e^T \nu^k < \mu_k^{2p}, \quad (b - Ax^k)^T \gamma^k < \mu_k^{2p}, \qquad (3.19b)$$

by (3.16), we obtain $\tilde{f}(x^{k+1}, \mu_k) - \tilde{f}(x^k, \mu_k) \le -\mu_k^{2p}$. Hence, (3.17) implies all inequalities in (3.19) hold and then we only need to prove the estimation in (3.18) under the conditions in (3.19).

First, from $Ax^k - b \le 0$, $\gamma^k \ge 0$, (3.19b) gives

$$-\mu_k^{2p} \le (Ax^k - b)^T \gamma^k \le 0. \qquad (3.20)$$

By ϱ_1^k, $\varrho_2^k \ge 0$, $\nu^k = \delta(\varrho_1^k + \varrho_1^k)$ and the first inequality in (3.19b), we have

$$\|\varrho_1^k - \varrho_2^k\| \le \|\varrho_1^k + \varrho_2^k\| \le \delta^{-1}\|\nu^k\|_1 \le \delta^{-1}\mu_k^p. \qquad (3.21)$$

Moreover, (3.15a) can be rewritten as

$$g(x^k, \mu_k) + A^T \gamma^k = -\beta_k s^k - 8\alpha p \mu_k^{p-2} \sum_{|d_i^T x^k| \le 2\mu_k^p} (d_i^T s^k)d_i + \varrho_1^k - \varrho_2^k. \qquad (3.22)$$

From the definition of Z_k, we have $Z_k^T d_i = 0$ for all d_i such that $|d_i^T x^k| \le 2\mu_k^p$. Combining this with $\varphi(\theta^p(d_i^T x^k, \mu_k)) = \varphi(|d_i^T x^k|^p)$ for all d_i such that $|d_i^T x^k| > 2\mu_k^p$, we obtain

$$Z_k^T \sum_{i=1}^m \nabla_t \varphi(\theta^p(t, \mu_k))_{t=d_i^T x^k} d_i = Z_k^T \sum_{i=1}^m \Phi(d_i^T x^k) d_i. \qquad (3.23)$$

Thus, multiplying Z_k^T to the both sides of (3.22) gives

$$Z_k^T(g(x^k, \mu_k) + A^T \gamma^k) = Z_k^T\left(\nabla\Theta(x^k) + \sum_{i=1}^m \Phi(d_i^T x^k)d_i + A^T \gamma^k\right), \qquad (3.24a)$$

$$Z_k^T\left(-\beta_k s^k - 8\alpha p \mu_k^{p-2} \sum_{|d_i^T x^k| \le 2\mu_k^p} (d_i^T s^k)d_i + \varrho_1^k - \varrho_2^k\right) = -\beta_k Z_k^T s^k + Z_k^T(\varrho_1^k - \varrho_2^k). \qquad (3.24b)$$

Then, from (3.19a), (3.21), (3.22) and (3.24), we obtain

$$
\begin{aligned}
&\left\|Z_k^T\left(\nabla\Theta(x^k) + \sum_{i=1}^m \Phi(d_i^T x^k)d_i + A^T \gamma^k\right)\right\|_\infty \\
&= \left\|\beta_k Z_k^T s^k - Z_k^T(\varrho_1^k - \varrho_2^k)\right\|_\infty \\
&\le \beta_k\|Z_k^T s^k\|_\infty + \|Z_k^T(\varrho_1^k - \varrho_2^k)\|_\infty \\
&\le \beta_k\|s^k\| + \|\varrho_1^k - \varrho_2^k\| \le (\sqrt{2\beta_k} + \delta^{-1})\mu_k^p,
\end{aligned}
\qquad (3.25)
$$

where we use that the columns of Z_k are orthonormal. By (3.20) and (3.25), we obtain the results in this lemma. $\square$

The following lemma presents some properties of the sequences $\{\beta_k\}$, $\{\mu_k\}$ and $\{f(x^k)\}$.

LEMMA 3.4. *The following statements hold.*

(i) $\beta_k \leq \bar{\beta} := \max\{\beta_0, \eta\hat{\beta}\}$ *for all* $k \in \mathbb{N}_0$;

(ii) $\lim_{k\to\infty} \mu_k = 0$;

(iii) $\lim_{k\to\infty} f(x^k)$ *exists.*

Proof. By Step 2 in the SQR algorithm, β_k is updated when $\beta_k \leq \hat{\beta}$, then statement (i) can be easily proved by the assumption on Θ.

From (3.14), we have

$$\sum_{k\in\mathcal{T}} \mu_k^{2p} < \sum_{k=1}^{\infty} \mu_0^{2p} \sigma^{2p(k-1)} = \frac{\mu_0^{2p}}{1-\sigma^{2p}}. \tag{3.26}$$

Note that when $k \in N_s \backslash \mathcal{T}$, from (3.12), we have

$$\mu_k^{2p} < \tilde{f}(x^k, \mu_k) - \tilde{f}(x^{k+1}, \mu_{k+1}) \quad \text{and} \quad \mu_{k+1} = \mu_k.$$

This, together with the nonincreasing property of $\tilde{f}(x^k, \mu_k)$ and (3.5), gives

$$\sum_{k\in N_s\backslash\mathcal{T}} \mu_k^{2p} < \sum_{k\in N_s\backslash\mathcal{T}} (\tilde{f}(x^k, \mu_k) - \tilde{f}(x^{k+1}, \mu_{k+1})) \leq \tilde{f}(x^0, \mu_0) - \min_{x\in\Omega} f(x). \tag{3.27}$$

Adding (3.26) and (3.27), we have

$$\sum_{k\in N_s} \mu_k^{2p} < \tilde{f}(x^0, \mu_0) - \min_{x\in\Omega} f(x) + \frac{\mu_0^{2p}}{1-\sigma^{2p}}. \tag{3.28}$$

If there are finite elements in N_s, then there is $\bar{k} \in \mathbb{N}_0$ such that $k \notin N_s \, \forall k \geq \bar{k}$, which implies that $\beta_k \geq \beta_0 \eta^{k-\bar{k}} \, \forall k \geq \bar{k}$. By $\eta > 1$, $\lim_{k\to\infty} \beta_k = \infty$, which leads a contradiction with the boundedness of $\{\beta_k\}$ given in (i). Thus, there are infinite elements in N_s, which together with (3.28) gives $\lim_{k\to\infty} \mu_k = 0$.

By (3.5) and $x^k \in \Omega$, we have

$$\tilde{f}(x^k, \mu_k) \geq f(x^k) \geq \min_{x\in\Omega} f(x),$$

which shows that $\{\tilde{f}(x^k, \mu_k)\}$ is bounded from below. Combining this with the nonincreasing property of $\{\tilde{f}(x^k, \mu_k)\}$, we obtain $\lim_{k\to\infty} \tilde{f}(x^k, \mu_k)$ exists. By virtue of $\lim_{k\to\infty} \mu_k = 0$ and (3.5), we have

$$\lim_{k\to\infty} \tilde{f}(x^k, \mu_k) = \lim_{k\to\infty} f(x^k) = \lim_{k\to\infty} f(z^k).$$

☐

From the proof of Lemma 3.4, we can guarantee the existence of the limit $f(x^k)$ for (1.1) with the unbounded linear constraints when the minimum of f in the feasible region exists.

4. Worst-case complexity analysis. We are now ready to present the worst-case complexity of the SQR algorithm for finding an ϵ scaled first order stationary point of (1.1).

THEOREM 4.1. *Given any* $\epsilon \in (0,1]$, *the proposed SQR algorithm obtains an* ϵ *scaled first order stationary point of (1.1) defined in Definition 2.3 in no more than* $O(\epsilon^{-2})$ *iterations.*

Proof. Without loss of generality, we suppose $\mu_0 = 1$. Fix $\epsilon > 0$ and let j be the smallest positive integer such that

$$C\sigma^{p(j-1)} \leq \epsilon \quad \text{and} \quad C\sigma^{p(j-2)} > \epsilon, \tag{4.1}$$

where $C = \max\{\sqrt{2\bar{\beta}} + \delta^{-1}, 2\}$ with $\bar{\beta}$ given in Lemma 3.4.

Denote t_j be the jth element of $\mathcal{T}$ defined in (3.14). Then, we will prove that x^{t_j} is an ϵ scaled first order stationary point of (1.1).

Note that

$$\mu_k = \sigma^{j-1} \quad \forall \, t_{j-1} + 1 \leq k \leq t_j. \tag{4.2}$$

Using $2\mu_{t_j}^p = 2\sigma^{p(j-1)} < \epsilon$, $\mathcal{D}_{t_j} \subseteq \mathcal{D}_{t_j}^\epsilon$, where

$$\mathcal{D}_{t_j} = \{d_i : |d_i^T x^{t_j}| \leq 2\mu_{t_j}^p\}, \quad \mathcal{D}_{t_j}^\epsilon = \{d_i : |d_i^T x^{t_j}| \leq \epsilon\}.$$

Then, we can find Z_{t_j} and $Z_{t_j}^\epsilon$ such that $Z_{t_j} \in \text{orthon}((\text{span}\mathcal{D}_{t_j})^\perp)$, $Z_{t_j}^\epsilon \in \text{orthon}((\text{span}\mathcal{D}_{t_j}^\epsilon)^\perp)$ and Z_{t_j} contains all columns of $Z_{t_j}^\epsilon$.

From Lemma 3.3, (3.12) and (4.2), there is a nonnegative vector $\gamma^{t_j} \in \mathbb{R}^r$ such that x^{t_j} satisfies

$$\left\| Z_{t_j}^T \left(\nabla\Theta(x^{t_j}) + \sum_{i=1}^m \Phi(d_i^T x^{t_j}) d_i + A^T \gamma^{t_j} \right) \right\|_\infty \leq \epsilon, \tag{4.3a}$$

$$Ax^{t_j} - b \leq 0, \quad -\epsilon^2 \leq (Ax^{t_j} - b)^T \gamma^{t_j} \leq 0. \tag{4.3b}$$

(4.3a) implies that

$$\left\| (Z_{t_j}^\epsilon)^T \left(\nabla\Theta(x^{t_j}) + \sum_{i=1}^m \Phi(d_i^T x^{t_j}) d_i + A^T \gamma^{t_j} \right) \right\|_\infty \leq \epsilon. \tag{4.4}$$

Thus, we conclude that x^{t_j} is an ϵ scaled first order stationary point of (1.1) and we need at most t_j iterations to find it.

Suppose there are s_j successful iterations up to the t_jth iteration. From Step 2 in the SQR algorithm and Lemma 3.4 (i), $\bar{\beta} \geq \beta_{t_j} \geq \beta_0 \eta^{t_j - s_j}$, which implies that $\eta^{t_j - s_j} \leq \bar{\beta}/\beta_0$. Then,

$$t_j \leq s_j + \log_\eta \bar{\beta} - \log_\eta \beta_0. \tag{4.5}$$

Thus, in order to evaluate t_j, we only need to evaluate s_j.

From (3.12), when $k \in N_s \backslash \mathcal{T}$,

$$\tilde{f}(x^{k+1}, \mu_{k+1}) - \tilde{f}(x^k, \mu_k) \leq -\mu_k^{2p}. \tag{4.6}$$

Since there are at least $s_j - j + 1$ successful iterations before the t_jth iteration such that (4.6) holds, from the nonincreasing of $\tilde{f}(x^k, \mu_k)$, we have

$$\tilde{f}(x^{t_j}, \mu_{t_j}) - \tilde{f}(x^0, \mu_0) \leq -(s_j - j + 1)\sigma^{2p(j-1)}. \tag{4.7}$$

By the second inequality in (4.1), we have

$$j \leq \frac{1}{p}\log_\sigma \frac{\epsilon}{C} + 2, \quad \sigma^{2p(j-1)} \geq \sigma^{2p}C^{-2}\epsilon^2. \tag{4.8}$$

14

(4.7) and (4.8) give

$$s_j \leq \frac{C^2(\tilde{f}(x^0, \mu_0) - \min_{x \in \Omega} f(x))}{\sigma^{2p} \epsilon^2} + \frac{1}{p} \log_\sigma \frac{\epsilon}{C} + 1. \qquad (4.9)$$

From (4.5) and (4.9), we have

$$t_j \leq \frac{C^2(\tilde{f}(x^0, \mu_0) - \min_{x \in \Omega} f(x))}{\sigma^{2p} \epsilon^2} + \frac{1}{p} \log_\sigma \frac{\epsilon}{C} + \log_\eta \bar{\beta} - \log_\eta \beta_0 + 1. \qquad (4.10)$$

Thus, the worst-case complexity of the proposed SQR algorithm for obtaining an ϵ scaled first order stationary point of (1.1) is $O(\epsilon^{-2})$. □

COROLLARY 4.2. *Suppose f is level bounded in Ω, i.e. $\{x \in \Omega : f(x) \leq \Gamma\}$ is bounded for any $\Gamma \in \mathbb{R}$. Then, any accumulation point of $\{z^k\}$ is a scaled first order stationary point of (1.1).*

Proof. By $x^k \in \Omega$, $f(x^k) \leq \tilde{f}(x^k, \mu_k) \leq \tilde{f}(x^0, \mu_0)$, $\forall k \in \mathbb{N}_0$, and the level boundedness of f in Ω, the iterates $\{x^k\}$ is bounded. By (3.13), for any fixed $j = 1, 2, \ldots$, $z^k = z^{t_j+1} = x^{t_j}$, $\forall t_j + 1 \leq k \leq t_{j+1}$. From Lemma 3.4 (ii), (4.3b) and (4.4), any accumulation point of $\{z^k\}$ is a scaled first order stationary point of (1.1). □

When Ω is bounded, the supposition in Corollary 4.2 holds naturally.

REMARK 4.1. *Whether the complexity and convergence of the SQR algorithm are dependent on the dimension of the problem are interesting problems. When initialized properly, the semismooth Newton process enjoys the mesh independence convergence[21]. Due to the computational cost of the preconditioning conjugate gradient technique, the overall computational cost of the proposed method in [37] is $O(cn \log n)$, where c depends on the number of approximate value ε. Since the right hand of (4.10) is independent on n, the worst-case iteration complexity of the proposed SQR algorithm for obtaining an ϵ scaled first order stationary point of (1.1) is independent on the dimension n. Moreover, if the inequality in (2.6a) is defined by the Euclidean 2-norm, then this independence result also holds. However, the total computational cost for reaching an ϵ scaled first order stationary point must increase with n and the estimation of it is up to the solution method for the subproblem (3.11).*

REMARK 4.2. *Let $\mu_0 = 1$. From the proof of Theorem 4.1, if j satisfies (4.1), then $z^{t_j+1} = x^{t_j}$ is an ϵ scaled first order stationary point of (1.1). By (4.1) and (4.2), if $\bar{k} \in \mathbb{N}_0$ satisfies*

$$\max\{\sqrt{2\beta_0} + \delta^{-1}, \sqrt{2\eta\hat{\beta}} + \delta^{-1}, 2\}\mu_{\bar{k}}^p \leq \epsilon \quad and \quad \mu_{\bar{k}+1} = \sigma\mu_{\bar{k}}, \qquad (4.11)$$

then z^k is an ϵ scaled first order stationary point of (1.1) for all $k \geq \bar{k} + 1$.

Though it is difficult to judge which iterate is an ϵ scaled first order stationary point of (1.1) from Definition 2.3, we can use (4.11) to find an ϵ scaled first order stationary point of (1.1), which is a sufficient condition for the conditions in Definition 2.3.

5. Numerical Experiments. In this section, we report numerical results of five examples to validate the theoretical results and show the good performance of the proposed SQR algorithm. The numerical testing is performed using MATLAB 2009a on a Lenov PC (3.00GHz, 2.00GB of RAM). The strongly convex quadratic subproblem (3.11) is solved by the projected alternating Barzilai-Borwein method in [17] with the zero vector as the initial iterate, and we stop when we find s^* such that $\|s^* - P_{S^k}[s^* - \nabla q(s^*, x^k, \mu_k, \beta_k)]\| \leq 10^{-5}$, where S^k and $q(s, x^k, \mu_k, \beta_k)$ are the

constraint set and objective function in (3.11) with $x = x^k$, $\mu = \mu_k$ and $\beta = \beta_k$. Without special clarification, we let $\mu_0 = 1$, $\beta_0 = 2$ and $\eta = 1.1$ in the SQR algorithm throughout the numerical experiments. Denote the first order difference operator by

$$D_1 = \begin{pmatrix} L_1 \otimes I \\ I \otimes L_1 \end{pmatrix} \quad \text{with} \quad L_1 = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{pmatrix}.$$

Example 5.1 is a randomly generated test problem to support the complexity theory of the SQR algorithm for finding a scaled first order stationary point of (1.1) given in Theorem 4.1 and the independence of parameter $\sigma \in (0, 1)$ for the complexity of SQR algorithm. Meantime, the computational performance of the SQR algorithm with respect to the dimension of problem (1.1) is also discussed.

Example 5.2- Example 5.5 are four often used gray level image restoration problems with intensity values ranging from 0 to 1 with size 64×64, 256×256 and 512×512. Two classes of observed images are considered. One is the observed image with blurring and Gaussian noise, and another is the observed image only with Gaussian noise. Numerical results show that the proposed SQR algorithm is robust and efficient for these image restoration problems. Let x_o and x_b be the original and observed images with the dimension $n_l \times n_w$. We use the peak signal-to-noise ratio (PSNR) to evaluate the quality of the restored image, i.e.

$$\text{PSNR}(x^k) = -10 \log_{10} \frac{\|x^k - x_o\|}{n_l \times n_w}.$$

In Example 5.2 and Example 5.3, we will show the importance of the constraint in (1.1) for image restoration, where we test the following three constraints:

$$\Omega^1 = \{x : 0 \leq x \leq e\}, \quad \Omega^2 = \{x : x \geq 0\}, \quad \Omega^3 = \mathbb{R}^n. \tag{5.1}$$

Moreover, the independence on initial iterate for the SQR algorithm is also discussed, where we test three different initial iterates: the zero vector denoted by 0, the observed data projected on Ω^1 denoted by $P_{\Omega^1}(x_b)$, and a randomly generated vector in Ω^1 denoted by x_r. For the regularization term, we use two different potential functions

$$\varphi_1(s) = \lambda s, \quad \varphi_2(s) = \lambda \frac{0.5s}{1 + 0.5s}, \quad \text{where} \quad \lambda > 0. \tag{5.2}$$

EXAMPLE 5.1. In this example, we solve (1.1) with $\Theta(x) = \|x - c\|^2$, $D = D_1$, $\Omega = \{x : -2e \leq x \leq 2e\}$ and $\varphi(s) := \varphi_1(s)$ with $\lambda = 0.2$. For a positive integer n_l, we use the following codes to generate $c \in \mathbb{R}^n$.
$n_w = n_l$; n=n_w*n_l; s=randn(n,1);
s'=median([-2*ones(1,n);s';2*ones(1,n)]); c=s+normrnd(0,0.05,[n,1]);

The algorithms in recent literatures [23, 25, 37, 38] also focus on solving this problem without constraint, where the approximation ideas are also used. The algorithms and convergence proof are based on a fixed approximation parameter in [23, 37, 38]. The smoothing descent method in [25] forces the smoothing parameter converging to zero and has the provable convergence to a stationary point of the non-smooth and non-convex TV^q model. However, there is no computational complexity analysis in

all of these literatures, while the convergence and computational complexity bound
are proved theoretically with the updating parameter for the SQR algorithm in this
paper. Here, we use this example to verify these results.

Independence of σ for the complexity of the SQR algorithm. The Parameter σ
in the SQR algorithm is to control the decreasing scale of smoothing parameter μ_k.
From the proof of Theorem 4.1, we find that the worst-case complexity of the SQR
algorithm is stable with respect to $\sigma \in (0, 1)$.

Let $p = 0.5$ in (1.1) and generate 20 samples of c by the above codes with $n_l = 2$.
For these 20 randomly generated vectors c and some different values of σ in $(0, 1)$,
the mean average number of iterations for obtaining an ϵ scaled first order stationary
point of (1.1) by the SQR algorithm with $x^0 = 0_{4\times 1}$ is illustrated in Figure 5.1(a),
where we can find that the worst-case complexity is stable for different values of
σ. For a certain random generated c and three different values of $\sigma \in (0, 1)$, the
convergence of the SQR algorithm is plotted in Figure 5.1(b), where we can find that
the obtained solutions are slightly different for $\sigma = 0.3, 0.5, 0.9$, but the first order
difference at those solutions x^* has the same zero elements, $(Dx^*)_1 = (Dx^*)_4 = 0$, as
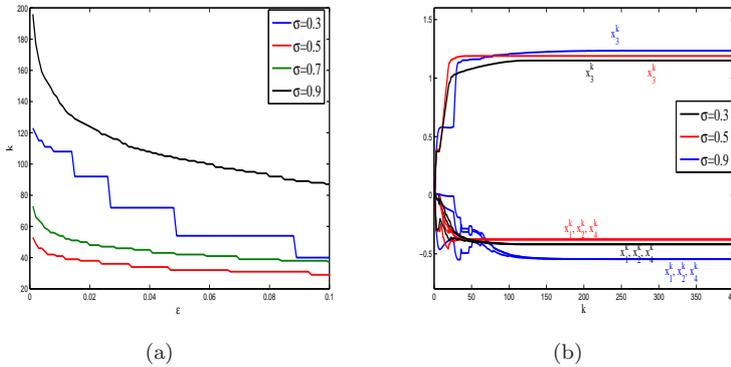$x_1^* = x_2^* = x_4^*$.

(a)　　　　　　　　　　　　(b)

Fig. 5.1: For different values of σ (a) Iteration complexity; (b) convergence of x^k

Computational complexity bound and convergence tests. Let $n_l = 2$ in the codes.
Choose $\sigma = 0.5$. With the above 20 randomly generated vectors c , the average
number of iterations for obtaining an ϵ scaled first order stationary point of (1.1)
with three different values of p is illustrated in Figure 5.2(a), where we use the stop
criterion in Remark 4.2. Moreover, in order to show the total computational effort of
the SQR algorithm, which includes the effort for calculating the quadratic subproblem
(3.11) with BB method, we show the computational complexity of the SQR algorithm
with respect to the CPU time for obtaining an ϵ scaled first order stationary point of
(1.1) in Figure 5.2(b), where the stop criterion is same as in Figure 5.2(a).

Computational performance with respect to dimension n. In this part, we test the
computational complexity of the SQR algorithm with respect to n for finding an ϵ
($\epsilon = 10^{-3}$) scaled first order stationary point of (1.1). In the codes, we let n_l=2:2:50
and generate 20 random vectors $c \in \mathbb{R}^n$ for each n_l. For $n = k^2$, $k = 2, 4, \ldots, 50$, the
average number of iterations for obtaining an ϵ ($\epsilon = 10^{-3}$) scaled first order stationary
point of (1.1) with the 20 random generated c is shown in Figure 5.3(a), where we
use the stop criterion in (4.11). From Figure 5.3(a), we find that the total iteration
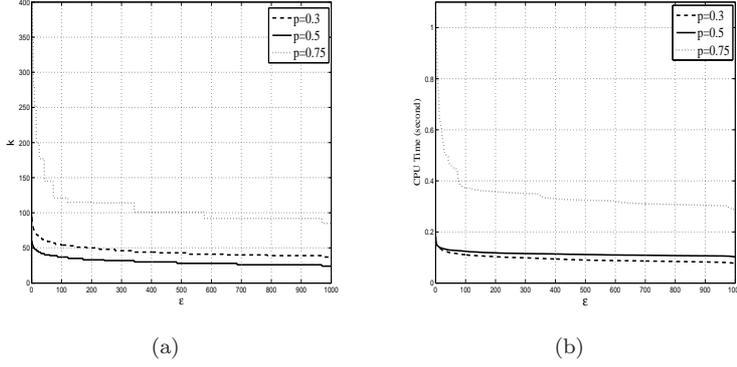
(a)　　　　　　　　　　　　　　　(b)

Fig. 5.2: (a) Iteration complexity; (b) CPU time complexity

for finding an ϵ scaled first order stationary point of (1.1) is not increase with n. The total computational cost with respect to the CPU time is given in Figure 5.3(b), which indicates the increasing effort with respect to the increasing of n.
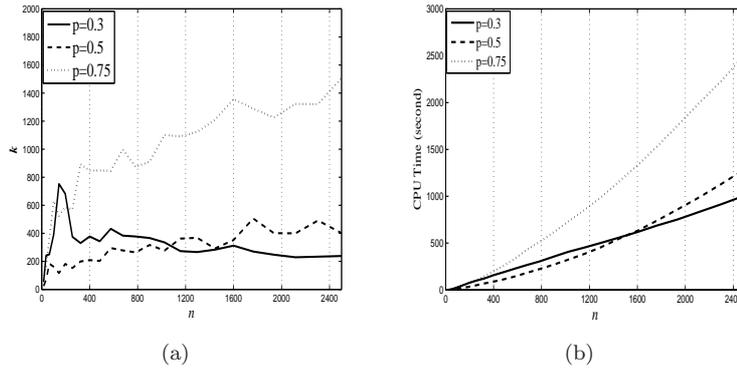

(a)　　　　　　　　　　　　　　　(b)

Fig. 5.3: For different values of n, (a) Iteration complexity; (b) CPU time complexity

EXAMPLE 5.2. ***Circles image with size*** 64×64***.*** In this example, we test the proposed SQR algorithm using the 64×64 Circles image [12, 23, 37]. We discuss the restoration of the circles image in two parts according to the class of observed images.

A. Observed image with blurring and noise. In this part, the observed image x_b is that all the pixels are blurred by a two dimensional Gaussian function, and then added a Gaussian noise. The blurring function is chosen to be

$$h(i,j) = \exp^{-2(i/3)^2 - 2(j/3)^2},$$

truncated such that the function has a support of 7×7, and the Gaussian noise is with the mean of 0 and the standard deviation of 0.05. The original image and the observed image are shown in Figure 5.4.

Dependence on parameter σ in the SQR algorithm. First, we test the influence of σ in the SQR algorithm for this image. Let $\varphi := \varphi_2$ with $\lambda = 0.019$, $\Omega := \Omega^1$, $D := D_1$
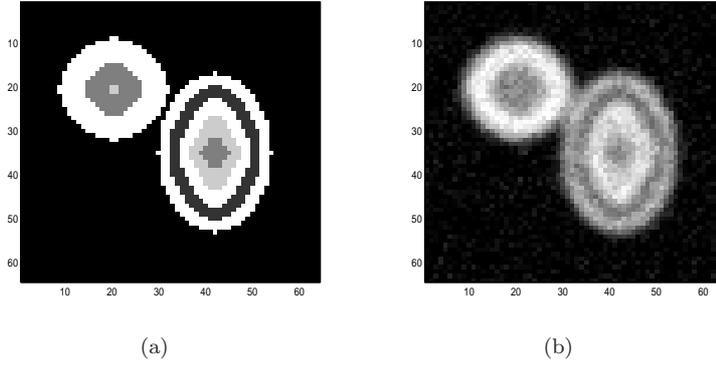
(a)   (b)

Fig. 5.4: (a) original image; (b) observed image(PSNR=15.50)

σ	0.99	0.7	0.5	0.3	0.1
PSNR(x^{500})	21.15	19.91	19.92	19.76	19.37
$f(x^{500})$	15.51	15.54	15.56	15.63	15.60

Table 5.1: The SQR algorithm for the circles image with different values of σ

and $p = 0.5$ in (1.1), and $x^0 = P_{\Omega^1}(x_b)$ in the SQR algorithm. Then, $f(x_o) = 13.71$, $f(x_b) = 32.62$ and $f(x^0) = 30.99$. For different values of σ in $(0, 1)$, the PSNR and objective function values at the 500th iteration are listed in Table 5.1, which shows that the SQR algorithm with a larger σ can find an image with higher PSNR value. The convergence of the PSNR and objective function values are figured in Figure 5.5 for $\sigma = 0.1, 0.5, 0.99$. In the sequence of this example, we let $\sigma = 0.99$.
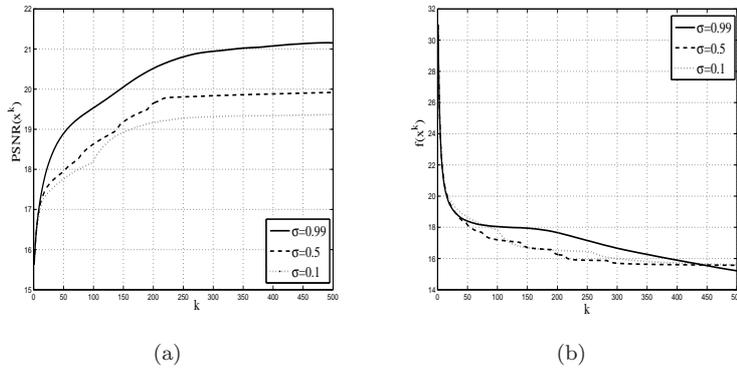

(a)   (b)

Fig. 5.5: For different values of σ: (a) PSNR values; (b) objective function values

Dependence on D in (1.1). In order to show the importance of the difference operators in image restoration, we choose $\varphi := \varphi_1$, $\Omega := \Omega^1$ and $p = 0.5$ in (1.1) and we test the SQR algorithm with $D := I$ and $D := D_1$ to restore the circles image

19

with blurring and noise. Let $x^0 = P_{\Omega^1}(x_b)$. Figure 5.6(a) shows the convergence of PSNR(x^k) with $D := I$ and $\lambda = 0.081$, which is the best choice of λ among $0.0001 : 0.0002 : 0.1$ to let the SQR algorithm find an x^k with the highest PSNR before 500 iterations. Also with $D := I$ and $\lambda = 0.081$, $f(x^{500}) = 80.82$ and $f(x_o) = 85.22$. When
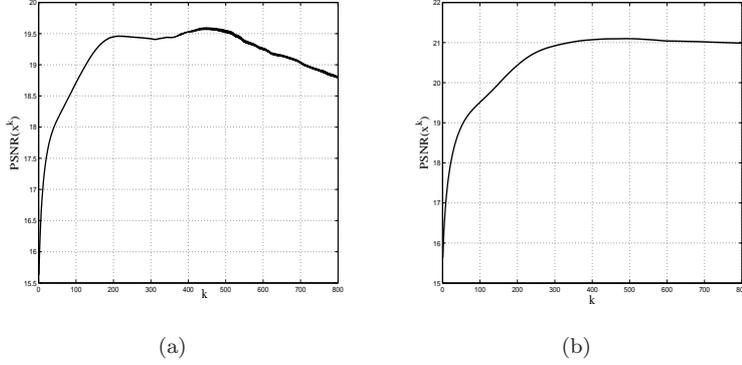

(a)   (b)

Fig. 5.6: (a) PSNR with $D := I$; (b) PSNR with $D := D_1$

$D := I$, from our numerical experiments, for almost all λ among $0.0001 : 0.0002 : 0.1$, $\tilde{f}(x^k, \mu_k)$ is monotone decreasing, whereas, the PSNR is not monotonely increasing and $f(x^k)$ can decrease below $f(x_o)$. Thus, the original image is not the optimal solution of (1.1) with $D := I$ and λ among $0.0001 : 0.0002 : 0.1$. However, when $D := D_1$ and $\lambda = 0.006$, PSNR(x^k) is monotonely increasing as shown in Figure 5.6(b).

The first or second order difference operators are also considered in the models in [1, 2, 7, 11, 12, 14, 15, 23, 24, 36, 37, 38, 39]. From this numerical experiment, we find that problem (1.1) with $D := I$ seems not suitable for the restoration of this image, but using $D := D_1$ performs well. This shows the importance of the linear operator D in the image restoration problems.

In the sequel parts of this example, we shall choose $D := D_1$ in (1.1).

Dependence on the constraints in (1.1). In this part, we test the importance of the constraints in (1.1) for obtaining the image with high PSNR values. To let the observed image in the usual dynamical range of the intensity values, we set the observed image be the projection of the above observed image on $[0, e]$ in this part. Then, the PSNR of the observed image is 15.63dB. Let $\varphi := \varphi_2$ and $x^0 = 0_{n \times 1}$. For different values of p and different constraints given in (5.1), the PSNR values at the 500th iteration are given in Table 5.2, where the parameter λ is also manually chosen in order to obtain the best PSNR value. From this table, we can find that the box constraints Ω^1 in (5.1) can improve the performance of the restored image clearly for different values of p. The average improvement of PSNR values is about 2.10dB. Moreover, when $p = 0.5$, the convergence of PSNR(x^k) by the SQR algorithm for the three constraints in (5.1) is shown in Figure 5.7(a).

Dependence on p in (1.1). In this part, we consider the influence of the value of p in (1.1) to restore the circles image with blurring and noise. Let $\varphi := \varphi_2$, $\Omega := \Omega^1$ and $x^0 = 0_{n \times 1}$. From Table 5.2, $0 < p < 1$ in (1.1) brings the restored image with better performance. With the same λ as in Table 5.2 for different values of p, the PSNR values at the 200th iteration are listed in Table 5.3, where the convergence of the

20

p	1	0.8	0.7	0.6	0.5	0.4	0.3	0.1
λ	0.017	0.018	0.018	0.019	0.019	0.020	0.020	0.040
Ω^1	20.95	21.06	21.17	21.24	21.27	21.20	21.20	21.20
Ω^2	19.89	19.96	20.08	20.17	20.24	20.20	20.22	20.20
Ω^3	18.81	18.93	19.09	19.16	19.18	19.04	19.10	19.15

Table 5.2: PSNR values of the SQR algorithm for (1.1) with different constraints and values of p
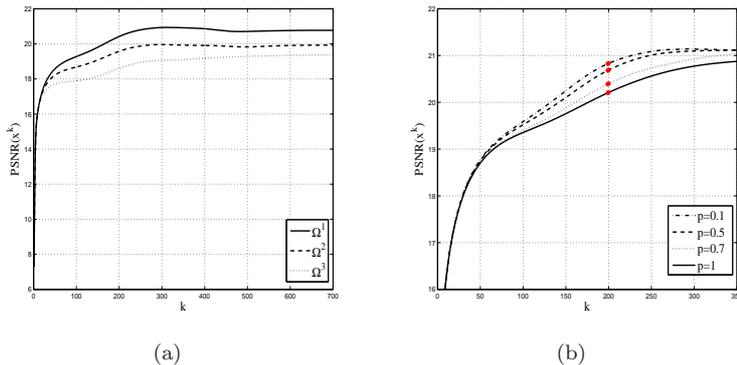
Fig. 5.7: Convergence of PSNR(x^k) for the circles image: (a) with different constraints; (b) with different values of p

PSNR values for some cases are plotted in Figure 5.7(b). From Table 5.3 and Figure 5.7(b), we find that solving (1.1) with a smaller value of p by the SQR algorithm can find an image with higher PSNR value in 200 iterations.

Independence on initial iterate in the SQR algorithm. In this part, we let $p = 0.5$, $\Omega := \Omega^1$ and $\varphi := \varphi_1$ with $\lambda = 0.006$. The objective value with the original image is $f(x_o) = 13.44$. For the three different initial iterates given in the beginning of this section, the corresponding results at the 500th iteration are given in Table 5.4, where we observe that the SQR algorithm is stable with respect to the choice of initial iterates, in terms of the PSNR values and objective values. Moreover, the mean CPU time of the SQR algorithm with 50 different iterates in Ω^1 to calculate x^{500} is 24.62 seconds, and the variance of it is 0.31.

At the end of this part, we should state that the PSNR of the restored image by the SQR algorithm for (1.1) with $D := D_1$ is better than the resorted images in [12, 37] (PSNR=19.03 reported in [37] and PSNR=19.97 reported in [12] by using different stop rules).

B. Observed image with Gaussian noise. In this part, we generate the observed image x_b without blurring that all the pixels are contaminated by Gaussian noise with mean of 0 and standard deviation of 0.1. Define $p = 0.5$, $\varphi := \varphi_2$ with $\lambda = 0.15$ and $D := D_1$ in (1.1). Then $f(x_o) = 64.27$, $f(x_b) = 172.43$ and PSNR$(x_b) = 20.07$.

Let $\Omega := \Omega^1$. With the three initial iterates used in Table 5.4, the numerical results of the SQR algorithm for solving (1.1) to restore the circles image with Gaussian noise

p	1	0.8	0.7	0.6	0.5	0.4	0.3	0.1
PSNR(x^{200})	20.16	20.26	20.34	20.39	20.50	20.59	20.72	20.85

Table 5.3: The SQR algorithm for the circles image with different values of p

x^0	PSNR(x^0)	PSNR(x^{500})	$f(x^0)$	$f(x^{500})$
0	7.30	21.13	585.41	14.32
$P_{\Omega^1}(x_b)$	15.63	21.10	27.34	14.29
x_r	5.04	20.83	818.19	14.37

Table 5.4: The SQR algorithm for the circles image with different initial iterates

are given in Table 5.5, which shows that the SQR algorithm is stable with respect to the initial iterates. With 50 different random initial iterates in Ω^1, the mean CPU time of the SQR algorithm to calculate x^{500} is 40.88 seconds and its variance is 0.49. Moreover, with $x^0 = P_{\Omega^1}(x_b)$, the convergence of μ_k and PSNR(x^k) are illustrated in Figure 5.8. From the results in Table 5.5 and Figure 5.8, the PSNR of the restored image by the SQR algorithm is also higher than the restored images in [23, 37] (PSNR=31.03 reported in [23] and PSNR=31.28 reported in [37] by using different stop rules) for the circles image with the same Gaussian noise.

For the three constraints in (5.1), the restored images by the SQR algorithm with $x^0 = 0$ at the 700th iteration are shown in Figure 5.9. We see that the quality of the restored image with box constraint Ω^1 is the best and more accurate of the feasible region in (1.1) brings better restoration image, which further shows the importance of the constraints in the image models. Moreover, let the allowable error be 10^{-3}, the original image has 3027 pixels with value 0 and 660 pixels with value 1, while the recovered image in Figure 5.9(a) has 1073 pixels with value 0 and 25 pixels with value 1.
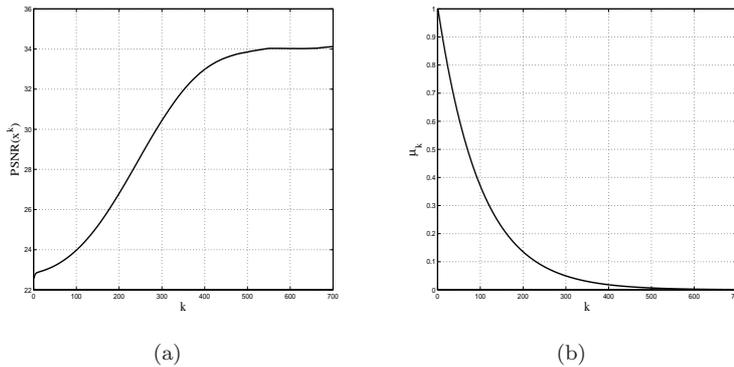

(a)   (b)

Fig. 5.8: Convergence of PSNR(x^k) and μ_k for the circles image with Gaussian noise: (a) PSNR(x^k); (b) μ_k

EXAMPLE 5.3. ***Phantom image with size*** 256×256***.*** First, we test the

x^0	PSNR(x^0)	PSNR(x^{500})	$f(x^0)$	$f(x^{500})$
0	7.30	33.83	794.42	79.65
$P_{\Omega^1}(x_b)$	22.57	33.85	139.82	79.12
x_r	5.04	33.83	1.56e+3	79.74

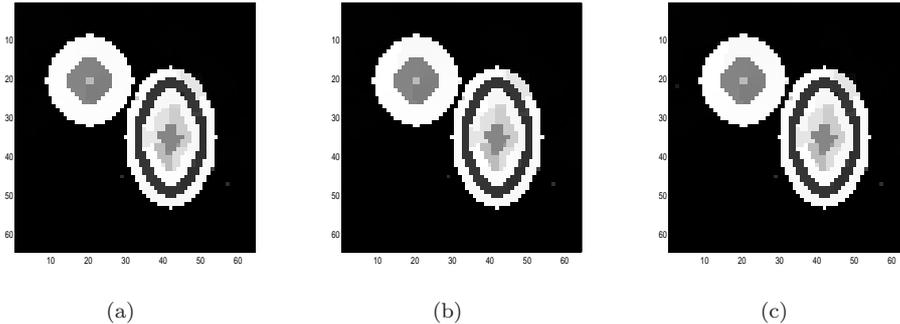Table 5.5: Stability of the SQR algorithm with different initial iterates for the circles image without blurring

(a) (b) (c)

Fig. 5.9: Restored images with different constraints: (a) Ω^1(PSNR(x^{700})=34.11); (b) Ω^2 (PSNR(x^{700})=33.66); (c) Ω^3(PSNR(x^{700})=32.76)

proposed SQR algorithm using the 256×256 phantom image with blurring and Gaussian noise as in Example 5.2-(A), where the original and observed images are given in Figure 5.10.

Define $\varphi := \varphi_1$ with $\lambda = 0.009$, $p = 0.5$ and $D := D_1$ in (1.1), and let $\sigma = 0.99$ in the SQR algorithm. With the three different constraints in (5.1), we show the restored images by the SQR algorithm with $x^0 = 0$ in Figure 5.11. And we draw the PSNRs and images of the 20 slices from 180 to 200 with the three constraints in Figure 5.12. Similar as the performance in Example 5.2, the box constraint Ω^1 can provide a better image restoration with higher PSNR value. Also let the allowable error be 10^{-3}, the original image has 38127 pixels with value 0 and 2846 pixels with value 1, while the recovered image in Figure 5.11(a) has 29866 pixels with value 0 and 1740 pixels with value 1.

Let $\Omega := \Omega^1$. The convergence of PSNR(x^k) with $x^0 = P_{\Omega^1}(x_b)$ is plotted in Figure 5.13(a). Figure 5.13(b) shows convergence of $f(x^k)$ and $\tilde{f}(x^k, \mu_k)$ generated by the SQR algorithm with $x^0 = P_{\Omega^1}(x_b)$. In order to see the edge-preserving property by the SQR algorithm, we display the 176th line of the original, observed and restored images in Figure 5.14.

Next, we test this phantom image only with Gaussian noise as in Example 5.2-(B). The PSNR of the observed image is 26.05dB. Choose $\varphi := \varphi_2$, $D := D_1$, $\Omega := \Omega^1$ in (1.1), let $x^0 = P_{\Omega^1}(x_b)$, $\sigma = 0.9$ in the SQR algorithm and we stop when $\mu_k < 0.01$. For $p = 0.25, 0.5, 0.75$, $f(x_o)$, $f(x^0)$, K, $f(x^K)$, PSNR(x^K) and PSNR[23] are reported in Table 5.6, where K denotes the number of terminate iterates and PSNR[23] is the PSNR values reported in [23] for this image with the same power on the regularization term. From this table, we would find that the proposed SQR algorithm performs well
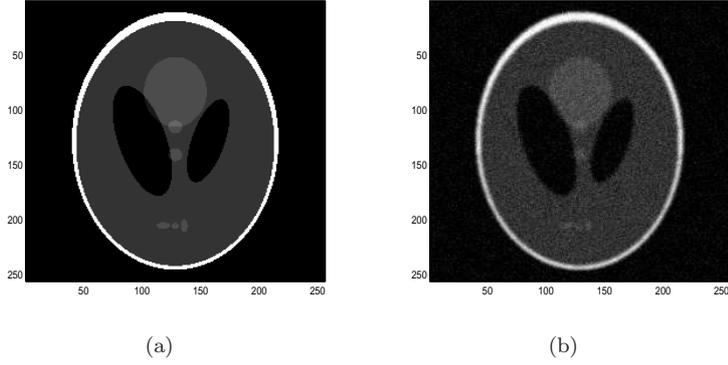
(a)   (b)

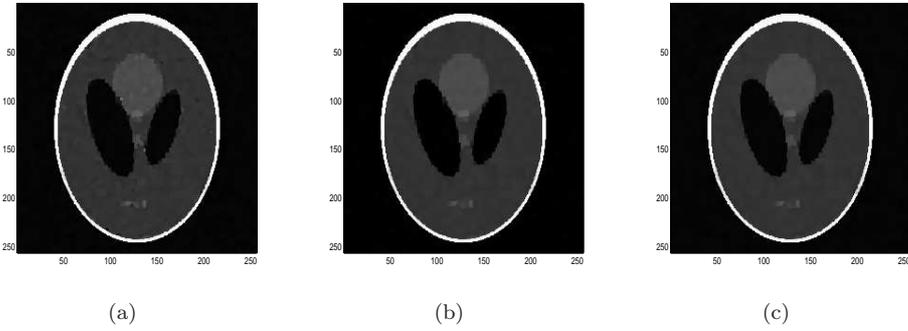Fig. 5.10: Phantom image: (a) original image; (b) observed image(PSNR=21.23)

(a)   (b)   (c)

Fig. 5.11: Restored images for the phantom image with different constraints: (a) Ω_1 (PSNR(x^{500})=29.60); (b) Ω_2 (PSNR(x^{500})=28.90); (c) Ω_3 (PSNR(x^{500})=28.12)

(a)   (b)
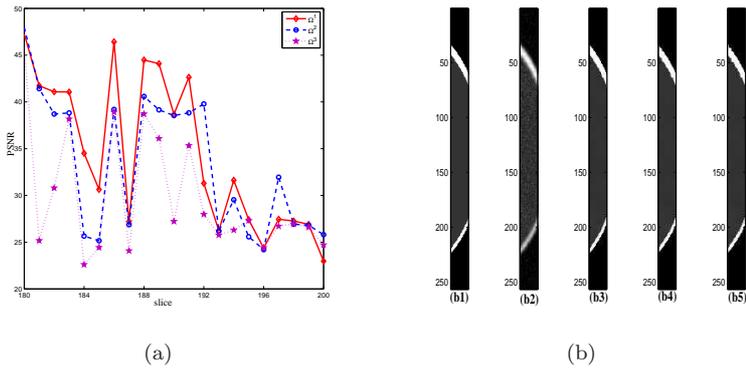
Fig. 5.12: (a) PSNRs of the restored 20 slices; (b) image of 20 slices: (b1) original (b2) observed (b3) with Ω^1 (b4) with Ω^2 (b5) with Ω^3

24
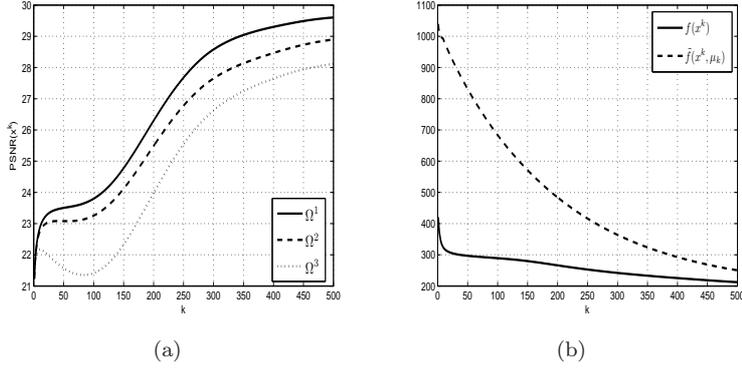
Fig. 5.13: Convergence for the phantom image: (a) PSNR(x^k) with different constraints; (b) $f(x^k)$ and $\tilde{f}(x^k, \mu_k)$ with $\Omega = \Omega^1$.
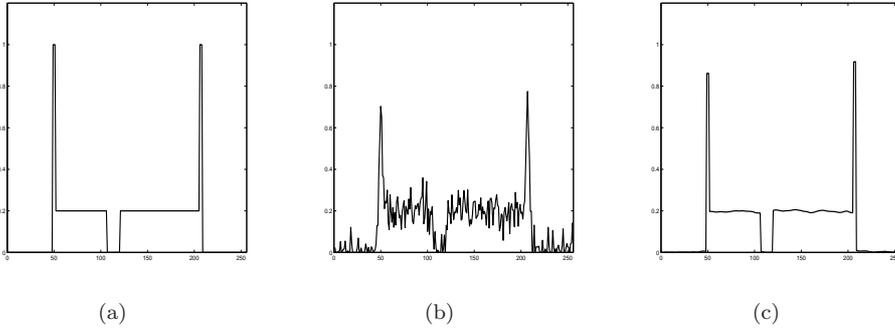

Fig. 5.14: 176th line of phantom image: (a) original image; (b) observed image; (c) restored image

for this image with Gaussian noise and the smaller p brings the restored image with higher PSNR values and fewer iterations.

EXAMPLE 5.4. **Phantom image with size** 512×512. In this example, we use another phantom image with size 512×512 to test the performance of the SQR algorithm in image restoration. The observed image is generated with the blurring and Gaussian noise as in Example 5.2. Figures 5.15(a)-5.15(b) show the original and observed images. Define $\Omega := \Omega^1$, $\varphi := \varphi_2$ with $\lambda = 0.07$, $p = 0.5$ and $D := D_1$ in (1.1). Let $x^0 = P_{\Omega^1}(x_b)$ and $\sigma = 0.9$ in the SQR algorithm. The restored image is given in Figure 5.15(c), where the restored image is generated by the 400th iteration of the SQR algorithm with the smoothing parameter $\mu_{400} = 0.01$. The convergence of the PSNR(x^k), $f(x^k)$ and $\tilde{f}(x^k, \mu_k)$ are plotted in Figure 5.16. Similar as the simulation results in Examples 5.2 and 5.3, (1.1) with box constraint Ω^1 can provide a good restoration for this image.

EXAMPLE 5.5. **Rice image with size** 256×256. Our fifth experiment corresponds to the rice image with size 256×256. We generate the observed image with the same blurring as in Example 5.2-(A) and Gaussian noise with mean 0 and stand

p	λ	$f(x_o)$	$f(x^0)$	K	$f(x^K)$	PSNR(x^K)	PSNR[23]
0.75	0.088	313.46	1.08×10^3	307	381.35	40.97	41.00
0.5	0.054	264.60	1.28×10^3	192	465.46	42.66	41.02
0.25	0.046	261.45	2.28×10^3	113	1.19×10^3	43.37	39.93

Table 5.6: Performance of the SQR algorithm for phantom image with Gaussian noise

(a)   (b)   (c)
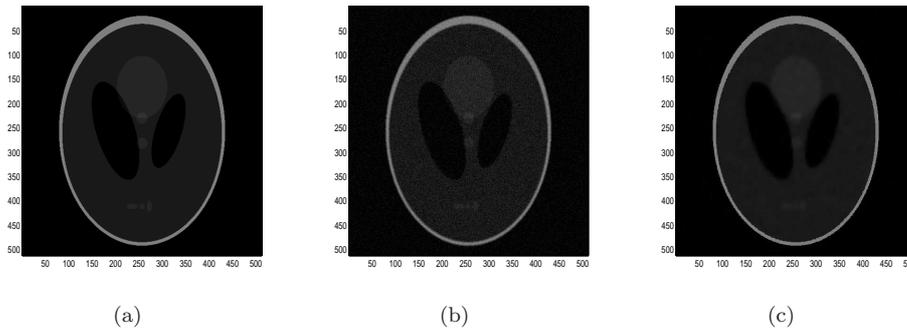
Fig. 5.15: Phantom image with size 512×512: (a) original image; (b) observed image(PSNR=25.01); (c)restored image (PSNR(x^{400})=34.88)

deviation of 0.1. The original and observed images are shown in Figures 5.17(a)-5.17(b). Define $\Omega := \Omega^1$, $\varphi := \varphi_1$ with $\lambda = 0.0003$, $p = 0.5$ and $D := D_1$ in (1.1). Let $x^0 = P_{\Omega^1}(x_b)$ and $\sigma = 0.9$ in the SQR algorithm. The restored image and the convergence of PSNR values are given in Figures 5.17(c)-5.17(d).

REFERENCES

[1] A. BECK AND M. TEBOULLE *A fast gradient-based algorithms for constrained total variation image denosing and deblurring problems*, IEEE Trans. Image Process., 18(2009), pp. 2419-2434.

[2] W. BIAN AND X. CHEN, *Smoothing neural network for constrained non-Lipschitz optimization with applications*, IEEE Trans. Neural Netw. Learn. Syst., 23(2012), pp. 399-411.

[3] W. BIAN AND X. CHEN, *Worst-case complexity of smoothing quadratic regularization methods for non-Lipschitzian optimization*, SIAM J. Optim., 23(2013), pp. 1718-1741.

[4] W. BIAN, X. CHEN AND Y. YE, *Complexity analysis of interior point algorithms for non-Lipschitz and nonconvex minimization*, Math. Program., 149(2015), pp. 301-327.

[5] C. CARTIS, N.I.M. GOULD AND PH.L. TOINT, *On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming*, SIAM J. Optim., 21(2011), pp. 1721-1739.

[6] R.H. CHAN AND K. CHEN, *A multilevel algorithm for simultaneously denoising and deblurring images*, SIAM J. Sci. Comput., 32(2010), pp. 1043-1063.

[7] R.H. CHAN, M. TAO AND X.M. YUAN, *Constrained total variation deblurring models abd fast algorithms based on alternating direction method of multipliers*, SIAM J. Imaging Sci., 6(2013), pp. 680-697.

[8] R. CHARTRAND, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett., 14(2007), pp. 707-710.

[9] R. CHARTRAND AND V. STANEVA, *Restricted isometry properties and nonconvex compressive sensing*, Inverse Prob., 24(2008), pp. 1-14.

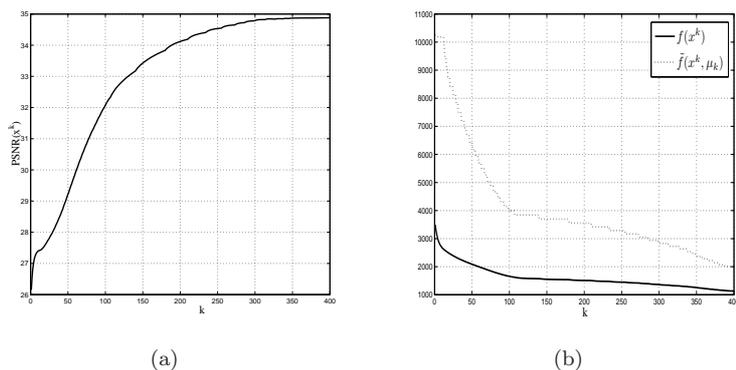[10] X. CHEN, *Smoothing methods for nonsmooth, nonconvex minimization*, Math. Program., 134(2012), pp. 71-99.

(a)   (b)

Fig. 5.16: Phantom image with size 512×512: (a) PSNR values; (c)objective and its smoothing values

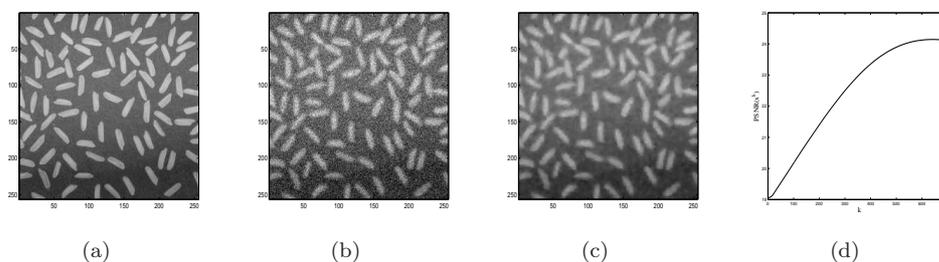(a)   (b)   (c)   (d)

Fig. 5.17: Rice image with size 256×256: (a) original image; (b) observed image(PSNR=19.04); (c) restored images (PSNR(x^{700})=24.11); (d) convergence of PSNR values

[11] X. CHEN, L. NIU AND Y. YUAN, *Optimality conditions and smoothing trust region Newton method for non-Lipschitz optimization*, SIAM J. Optim., 23(2013), pp. 1528-1552.

[12] X. CHEN, M. NG AND C. ZHANG, *Nonconvex l_p regularization and box constrained model for image restoration*, IEEE Trans. Image Processing, 21(2012), pp. 4709-4721.

[13] 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(2010), pp. 2832-2852.

[14] X. CHEN AND W. ZHOU, *Smoothing nonlinear conjugate gradient method for image restoration using nonsmooth nonconvex minimization*, SIAM J. Imaging Sci., 3(2010), pp. 765-790.

[15] E. CHOUZENOUS, A. JEZIERSKA, J. PESQUET AND H. TALBOT, *A majorize-minimize subspace approach for l_2-l_0 image regularization*, SIAM J. Imaging Sci., 6(2013), pp. 563-591.

[16] F.H. CLARKE, *Optimization and Nonsmooth Analysis*, New York: Wiley, 1983.

[17] Y.H. DAI AND R. FLETCHER, *Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programing*, Numer. Math., 100(2005), pp. 21-47.

[18] J. FAN AND R. LI, *Variable selection via nonconcave penalized likelihood and its oracle properties*, J. Amer. Stat. Soc., 96(2001), pp. 1348-1360.

[19] M. FRANK AND P. WOLFE, *An algorithm for quadratic programming*, Nav. Res. Logist., 3(1956), pp. 95-110.

[20] D. GE, X. JIANG AND Y. YE, *A note on the complexity of L_p minimization*, Math. Program., 21(2011), pp. 1721-1739.

[21] M. HINTERMÜLLAER AND M. ULBRICH, *A mesh-independence result for semismooth Newton methods*, Math. Profram., 101(2004), pp. 151-184.

[22] M. Hintermüllaer, T. Valkonen and T. Wu, *Limiting aspects of non-convex TV^φ models* Preprint, (2014).

[23] M. Hintermüller and T. Wu, *Nonconvex TV^q-models in image restoration: analysis and a trust-region regularization based superlinearly convergent solver*, SIAM J. Imaging Sci., 6(2013), pp. 1385-1415.

[24] M. Hintermüller and T. Wu, *A superlinearly convergent R-regularized Newton scheme for variational models with concave sparsity-promoting priors*, Computat. Optim. Appl., 57(2014), pp. 1-25.

[25] M. Hintermüller and T. Wu, *A smoothing descent method for nonconvex TV^q -models*, In: A. Bruhn, T. Pock, X.-C. Tai (eds.) Efficient Algorithms for Global Optimization Methods in Computer Vision, LNCS, Springer, 8293(2014), pp. 119-133.

[26] J. Huang, J. L. Horowitz and S. Ma, *Asymptotic properties of bridge estimators in sparse high-dimensional regression models*, Ann. Stat., 36(2008), pp. 587-613.

[27] J. Huang, S. Ma, H. Xue and C. Zhang, *A group bridge approach for variable selection*, Biometrika, 96(2009), pp. 339-355.

[28] P. Huber, Robust Estimation, New York: Wiley, 1981.

[29] A. Jain, *Fundamentals of Digital Image Processing*, Upper Saddle River, NJ: Prentice-Hall, 1989.

[30] K. Knight and W.J. Fu, *Asymptotics for lasso-type estimators*, Ann. Stat., 28(2000), pp. 1356-1378.

[31] D. Krishnan, L. Ping, and A.M. Yip, *A primal-dual active-set method for non-negativity constrained total variation deblurring problems*, IEEE Trans. Image Process., 16(2007), pp. 2766-2777.

[32] K. Kunisch and T. Pock, *A bilevel optimization approach for parameter learning in variational models*, SIAM J. Imaging Sci., 6(2013), pp. 938-983.

[33] M. Lai and Y. Wang, *An unconstrained l_q minimization with $0 < q < 1$ for sparse solution of under-determined linear systems*, SIAM J. Optim., 21(2011), pp. 82-101.

[34] Z. Lu, *Iterative reweighted minimization methods for l_p regularized unconstrained nonlinear programming*, Math. Program., 147(2014), pp. 277-307.

[35] Yu. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program., 103(2005), pp. 127-152.

[36] M. Nikolova, *Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares*, Multiscale Model. Simul., 44(2005), pp. 960-991.

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

[38] M. Nikolova, M.K. Ng and C.P. Tam, *Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction*, IEEE Trans. Image Process., 19(2010), pp. 3073-3088.

[39] M. Ng, P. Weiss and X. Yuan, *Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods*, SIAM J. Sci. Comput., 32(2010), pp. 2710-2736.

[40] Y. Ouyang, Y. Chen, G. Lan and E.JR. Pasiliao, *A new semiblind deconvolution approach for Fourier-based image restoration: an application in astronomy*, SIAM J. Imaging Sci., 6(2013), pp. 1736-1757.

[41] R.T. Rockafellar and R.J-B Wets, *Variational Analysis*, Berlin: Springer, 1998.