Iteratively Linearized Reweighted Alternating Direction Method of Multipliers for a Class of Nonconvex Problems

Tao Sun[*] Hao Jiang[†] Lizhi Cheng[‡]

September 7, 2017

Abstract

In this paper, we consider solving a class of nonconvex and nonsmooth problems frequently appearing in signal processing and machine learning research. The traditional alternating direction method of multipliers encounters troubles in both mathematics and computations in solving the nonconvex and nonsmooth subproblem. In view of this, we propose a reweighted alternating direction method of multipliers. In this algorithm, all subproblems are convex and easy to calculate. We also provide several guarantees for the convergence and prove that the algorithm globally converges to a critical point of an auxiliary function with the help of the Kurdyka-Łojasiewicz property. Several numerical results are presented to demonstrate the efficiency of the proposed algorithm.

Keywords: Alternating direction method of multipliers, iteratively reweighted algorithm, nonconvex and nonsmooth minimization, Kurdyka-Łojasiewicz property

Mathematical Subject Classification 90C30, 90C26, 47N10

1 Introduction

Minimization of composite functions with linear constrains finds various applications in signal and image processing, statics, machine learning, to name a few. Mathematically, such a problem can be presented as

$$\min_{x,y}\{f(x)+g(y) \ \text{ s.t. } \ Ax+By=c\}, \tag{1.1}$$

where $A\in\mathbb{R}^{r\times M}$, $B\in\mathbb{R}^{r\times N}$, and g is usually the regularization function and f is usually the loss function.

The well-known alternating direction method of multipliers (ADMM) method [12, 15] is a powerful tool for the problem mentioned above. The ADMM actually aims to focusing on the augmented Lagrangian problem of (1.1) which reads as

$$\widetilde{\mathcal{L}}_\alpha(x,y,p):=f(x)+g(y)+\langle p,Ax+By-c\rangle+\frac{\alpha}{2}\|Ax+By-c\|_2^2, \tag{1.2}$$

where $\alpha>0$ is a parameter. The ADMM minimizes only one variable and fixing others in each iteration; the variable p is updated by a feedback strategy. Mathematically, the standard ADMM method can be presented as

$$\left\{\begin{array}{lll} y^{k+1} & = & \arg\min_y \widetilde{\mathcal{L}}_\alpha(x^k,y,p^k)\\ x^{k+1} & = & \arg\min_x \widetilde{\mathcal{L}}_\alpha(x,y^{k+1},p^k)\\ p^{k+1} & = & p^k+\alpha(Ax^{k+1}+By^{k+1}-c)\end{array}\right. \tag{1.3}$$

[*] College of Science, National University of Defense Technology, Changsha, 410073, Hunan, China. Email: nudtsuntao@163.com

[†] College of Computer, National University of Defense Technology, Changsha, 410073, Hunan, China. Email: haojiang@nudt.edu.cn

[‡] College of Science & The State Key Laboratory for High Performance Computation, National University of Defense Technology, Changsha, 410073, Hunan, China. Email: clzcheng@nudt.edu.cn

The ADMM algorithm attracts increasing attentions for its efficiency in dealing with sparse-related problems [40, 37, 39, 26, 35]. Obviously, the ADMM has a self-explanatory assumption; all the subproblems shall be solved efficiently. In fact, if the proximal maps of the f and g are easy to calculate, the linearized ADMM [8] proposes the linearized technique to solve the subproblem efficiently; the subproblems all need to compute once proximal map of f or g. The core part of the linearized ADMM lies on linearizing the quadratical terms $\frac{\alpha}{2}\|Ax+By^k-c\|_2^2$ and $\frac{\alpha}{2}\|Ax^{k+1}+By-c\|_2^2$ in each iteration. The linearized ADMM is also called as preconditioned ADMM in [11]; in fact, it is also a special case when $\theta=1$ in Chambolle-Pock primal dual algorithm [5]. In the latter paper, the linearized ADMM is more generalized as the Bregman ADMM [34].

The convergence of the ADMM in convex case is also well studied; numerous excellent works have made contributions to this field [16, 17, 19, 10]. Recently, the ADMM algorithm is even developed for the infeasible problems [24]. The earlier analyses focus on the convex case, i.e., both f and g are all convex. But as the nonconvex penalty functions perform efficiently in applications, nonconvex ADMM is developed and studied: in paper [6], Chartrand and Brendt directly used the ADMM to the group sparse problems. They replace the nonconvex subproblems as a class of proximal maps. Latter Ames and Hong consider applying ADMM for certain non-convex quadratic problems [1]. The convergence is also presented. A class of nonconvex problems are solved by Hong et al by a provable ADMM [20]. They also allow the subproblems to be solved inexactly by taking gradient steps which can be regarded as a linearized way. Recently, with weaker assumptions, [36] present new analysis for nonconvex ADMM by novel mathematical techniques. With the Kurdyka-Łojasiewicz property, [22, 23] consider the convergence of the generated iterative points. [32] consider a structure constrained problem and proposed the ADMM-DC algorithm. In nonconvex ADMM literature, either the proximal maps of f and g or the subproblems are assumed to be easily calculated.

1.1 Motivated example and problem formulation

This subsection contains two parts: the first one presents an example and discusses the problems in directly using the ADMM; the second one describes the problem considered in this paper.

1.1.1 A motivated example: the problems in directly using ADMM

The methods mentioned above are feasibly applicable provided the subproblems are relatively solvable, i.e., either the proximal maps of f and g or the subproblems are assumed to be easily calculated. However, the nonconvex cases may not always promise such a convention. We recall the $\mathrm{TV}_{\varepsilon}^{q}$ problem [18] which arises in imaging science

$$\min_{u}\{\frac{1}{2}\|f-\Psi u\|_2^2+\sigma\|Tu\|_{q,\varepsilon}^q\}, \tag{1.4}$$

where T is the total variation operator and $\|y\|_{q,\varepsilon}^q:=\sum_i(|y_i|+\varepsilon)^q$. By denoting $v=Tu$, the problem then turns to being

$$\min_{u,v}\{\frac{1}{2}\|f-\Psi u\|_2^2+\sigma\|v\|_{q,\varepsilon}^q, \quad \text{s.t.} \ \ Tu-v=\mathbf{0}\}. \tag{1.5}$$

The direct ADMM for this problem can be presented as

$$\left\{\begin{array}{lll} v^{k+1} & = & \arg\min_v\{\sigma\|v\|_{q,\varepsilon}^q+\langle p^k, v-Tu^k\rangle+\frac{\alpha}{2}\|v-Tu^k\|_2^2\}, \\ x^{k+1} & = & \arg\min_x\{\frac{1}{2}\|f-\Psi u\|_2^2+\langle p^k, v^{k+1}-Tu\rangle+\frac{\alpha}{2}\|v^{k+1}-Tu\|_2^2\}, \\ p^{k+1} & = & p^k+\alpha(v^{k+1}-Tu^{k+1}). \end{array}\right. \tag{1.6}$$

The first subproblem in the algorithm needs to minimize a nonconvex and nonsmooth problem. If $q=\frac{1}{2},\frac{2}{3}$, the point v^k can be explicitly calculated. This is because the proximal map of $\|\cdot\|_{q,\varepsilon}^q$ can be easily obtained. But for other q, the proximal map cannot be easily derived. Thus, we may must employ iterative algorithms to compute v^{k+1}. That indicates three drawbacks which cannot be ignored:

1. The stopping criterion is hard to set for the nonconvexity[1].
2. The error may be accumulating in the iterations due to inexact solution of the subproblem.
3. Even the subproblem can be solved without any error, the nonconvexity always promises a critical point for the subproblem, which is not "really" argmin.

In fact, the other penalty functions like Logistic function [38], Exponential-Type Penalty (ETP) [13], Geman [14], Laplace [33] also encounter such a problem.

1.1.2 Optimization problem and basic assumptions

In this paper, we consider the following problem

$$\min_{x,y} f(x) + \sum_{i=1}^{N} g[h(y_i)] \quad \text{s.t.} \quad Ax + By = c, \tag{1.7}$$

where $A \in \mathbb{R}^{r\times M}$, $B \in \mathbb{R}^{r\times N}$, and f, g and h satisfy the following assumptions:

A.1 $f : \mathbb{R}^N \to \mathbb{R}$ is a differentiable convex function with a Lipschitz continuous gradient, i.e.,

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L_f \|x - y\|_2. \tag{1.8}$$

A.2 $h : \mathbb{R} \to \mathbb{R}$ is convex and proximable.

A.3 $g : \mathrm{Im}(h) \to \mathbb{R}$ is a differentiable concave function with a Lipschitz continuous gradient whose Lipschitz continuity modulus is bounded by $L_g > 0$; that is

$$\mid g'(s) - g'(t) \mid \leq L_g \mid s - t \mid, \tag{1.9}$$

and $g'(t) > 0$ when $t \in \mathrm{Im}(h)$.

It is easy to see that the TV^q problem can be regarded as a special one of (1.7) if we set $g(s) = (s+\varepsilon)^q$ and $h(t) = |t|$. The augmented lagrange dual function of model (1.7) is

$$\mathcal{L}_\alpha(x, y, p) = f(x) + \sum_{i=1}^{N} g[h(y_i)] + \langle p, Ax + By - c\rangle + \frac{\alpha}{2}\|Ax + By - c\|_2^2, \tag{1.10}$$

where $\alpha > 0$ is a parameter.

1.2 Linearized ADMM meets the iteratively reweighted strategy: convexifying the subproblems

In this part, we present the algorithm for solving problem (1.7). The term $\sum_{i=1}^N g[h(y_i)]$ has a deep relationship with several iteratively reweighted style algorithms [7, 9, 31, 41, 30]. Although the function $\sum_{i=1}^N g[h(y_i)]$ may be nondifferentiable itself, the reweighted style methods still propose an elegant way: linearization of outside function g. Precisely, in $(k+1)$-th iteration of the iteratively reweighted style algorithms, the term $\sum_{i=1}^N g[h(y_i)]$ is usually replaced by $\sum_{i=1}^N g'[h(y_i^k)]\cdot[h(y_i) - h(y_i^k)] + \sum_{i=1}^N g[h(y_i^k)]$, where y^k is obtained in the k-th iteration. Motivated by the iteratively reweighted strategy, we propose the following scheme for solving (1.7) which reads as

$$\left\{\begin{array}{lcl} y^{k+1} & = & \arg\min_y\{\sum_{i=1}^N g'[h(y_i^k)]h(y_i) + \langle By, \alpha(Ax^k + By^k - c) + p^k\rangle + \frac{r_1}{2}\|y - y^k\|_2^2\}, \\ x^{k+1} & = & \arg\min_x\{f(x) + \langle Ax, \alpha(Ax^k + By^{k+1} - c) + p^k\rangle + \frac{(x-x^k)^\top[D(\vec{r_2}) - \alpha A^\top A](x-x^k)}{2}\}, \\ p^{k+1} & = & p^k + \alpha(Ax^{k+1} + By^{k+1} - c), \end{array}\right. \tag{1.11}$$

[1] The convex methods usually enjoy a convergence rate.

where $D(\vec{r_2}) = U^\top \mathrm{diag}(r_{2,1}, r_{2,2}, \ldots, r_{2,N})U$ and $r_{2,i} > 0$ for $i \in [1, 2, \ldots, N]$, and U is the SVD matrix of $A^\top A$. We combined both linearized ADMM and reweighted algorithm in the new scheme: for the nonconvex part $\sum_{i=1}^N g[h(y_i)]$, we linearize the outside function g and keep h, which aims to derive the convexity of the subproblem; for the quadratical parts $\frac{\alpha}{2}\|Ax + By^k - c\|_2^2$ and $\frac{\alpha}{2}\|Ax^{k+1} + By - c\|_2^2$, linearizations are for the use of the proximal map of f and h. We call this new algorithm as Iteratively Linearized Reweighted Alternating Direction Method of Multipliers (ILR-ADMM). It is easy to see that each subproblem just needs to solve a convex problem in this scheme. With the expression of proximal maps, scheme (1.11) can be equivalently presented as the following forms

$$\left\{\begin{array}{lll} y_i^{k+1} & = & \mathbf{prox}_{\frac{g'[h(y_i^k)]}{r_1}h}(y_i^k - \frac{B_i^\top(\alpha(Ax^k+By^k-c)+p^k)}{r_1}), i \in [1, 2, \ldots, N] \\ x^{k+1} & = & \arg\min_x\{f(x) + \langle Ax, \alpha(Ax^k + By^{k+1} - c) + p^k\rangle + \frac{(x-x^k)^\top[D(\vec{r_2})-\alpha A^\top A](x-x^k)}{2}\}, \\ p^{k+1} & = & p^k + \alpha(Ax^{k+1} + By^{k+1} - c), \end{array}\right. \tag{1.12}$$

where B_i denotes the i-th column of the matrix B. In many applications, f is the quadratical function, and then solving x^{k+1} is also very easy. With this form, the algorithm can be programmed with the absence of inner loops. In fact, if A, B and c all vanish, ILR-ADMM immediately reduces to the proximal reweighted algorithm [**?**].

Algorithm 1 Iteratively Linearized Reweighted Alternating Direction Method of Multipliers (ILR-ADMM)

```
Require: parameters $\alpha > 0, r_1, r_{2,1}, r_{2,1}, \ldots, r_{2,N}$
  Initialization: $x^0, y^0, p^0$
  for $k = 0, 1, 2, \ldots$
    $y_i^{k+1} = \mathbf{prox}_{\frac{g'[h(y_i^k)]}{r_1}h}(y_i^k - \frac{B_i^\top(\alpha(Ax^k+By^k-c)+p^k)}{r_1}), i \in [1, 2, \ldots, N]$
    $x^{k+1} = \arg\min_x\{f(x) + \langle Ax, \alpha(Ax^k + By^{k+1} - c) + p^k\rangle + \frac{(x-x^k)^\top[D(\vec{r_2})-\alpha A^\top A](x-x^k)}{2}\}$
    $p^{k+1} = p^k + \alpha(Ax^{k+1} + By^{k+1} - c)$
  end for
```

1.3 Contribution and Organization

In this paper, we consider a class of nonconvex and nonsmooth problems which are ubiquitous in applications. Direct use of ADMM algorithms will lead troubles in both computations and mathematics for the nonconvexity of the subproblem. In view of this, we propose the iteratively linearized reweighted alternating direction method of multipliers for these problems. The new algorithm is an organic combination of iteratively reweighted strategy and the linearized ADMM. All the subproblems in the proposed algorithm are convex and easy to be solved if the proximal maps of h is easy to calculate and f is quadratical. Compared with the direct application of ADMM to problem (1.7), we now list the advantages of the new algorithm:

1. Computational perspective: each subproblem just needs to compute once proximal map of g and minimize a quadratical problem, the computational cost is low in each iteration.

2. Practical perspective: without any inner loop, the programming is very easy.

3. Mathematical perspective: all the subproblems is convex and exactly solved. Thus, we get "really" argmin everywhere, which makes the mathematical convergence analysis solid and meaningful.

With the help of the Kurdyka-Łojasiewicz property, we provide the convergence results of the algorithm with proper selections of the parameters. The applications of the new algorithm to the signal and image processing are presented. The numerical results demonstrate the efficiency of the proposed algorithm.

The rest of this paper is organized as follows. Section 2 introduces the preliminaries including the definitions of subdifferential and the Kurdyka-Łojasiewicz property. Section 3 provides the convergence analysis. The core part is using an auxiliary Lyapunov function and bounding the generated sequence. Section 4 applies the proposed algorithm to signal and image processing. And several comparisons are reported. Finally, section 5 concludes the paper.

2 Preliminaries

We introduce the basic tools in the analysis: the subdifferential and Kurdyka-Łojasiewicz property. These two definitions play important roles in the variational analysis.

2.1 Subdifferential

Given a lower semicontinuous function $J : \mathbb{R}^N \to (-\infty, +\infty]$, its domain is defined by

$$\mathrm{dom}(J) := \{x \in \mathbb{R}^N : J(x) < +\infty\}.$$

The graph of a real extended valued function $J : \mathbb{R}^N \to (-\infty, +\infty]$ is defined by

$$\mathrm{graph}(J) := \{(x, v) \in \mathbb{R}^N \times \mathbb{R} : v = J(x)\}.$$

Now, we are prepared to present the definition of subdifferential. More details can be found in [27].

Definition 1. *Let $J : \mathbb{R}^N \to (-\infty, +\infty]$ be a proper and lower semicontinuous function.*

1. *For a given $x \in dom(J)$, the Fréchet subdifferential of J at x, written as $\hat{\partial} J(x)$, is the set of all vectors $u \in \mathbb{R}^N$ satisfying*
$$\lim_{y \neq x} \inf_{y \to x} \frac{J(y) - J(x) - \langle u, y - x \rangle}{\|y - x\|_2} \geq 0.$$
When $x \notin dom(J)$, we set $\hat{\partial} J(x) = \emptyset$.

2. *The (limiting) subdifferential, or simply the subdifferential, of J at $x \in \mathbb{R}^N$, written as $\partial J(x)$, is defined through the following closure process*
$$\partial J(x) := \{u \in \mathbb{R}^N : \exists x^k \to x, J(x^k) \to J(x) \text{ and } u^k \in \hat{\partial} J(x^k) \to u \text{ as } k \to \infty\}.$$

When J is convex, the definition agrees with the subgradient in convex analysis [28] which is defined as

$$\partial J(x) := \{v \in \mathbb{R}^N : J(y) \geq J(x) + \langle v, y - x \rangle \ \text{ for any } \ y \in \mathbb{R}^N\}.$$

It is easy to verify that the Fréchet subdifferential is convex and closed while the subdifferential is closed. Denote that

$$\mathrm{graph}(\partial J) := \{(x, v) \in \mathbb{R}^N \times \mathbb{R}^N : v \in \partial J(x)\},$$

thus, graph(∂J) is a closed set. Let $\{(x^k, v^k)\}_{k \in \mathbb{N}}$ be a sequence in $\mathbb{R}^N \times \mathbb{R}$ such that $(x^k, v^k) \in$ graph (∂J). If (x^k, v^k) converges to (x, v) as $k \to +\infty$ and $J(x^k)$ converges to v as $k \to +\infty$, then $(x, v) \in$ graph (∂J). This indicates the following simple proposition.

Proposition 1. *If $v^k \in \partial J(x^k)$, $\lim_k v^k = v$ and $\lim_k x^k = x$. Then, we have that*

$$v \in \partial J(x). \tag{2.1}$$

A necessary condition for $x \in \mathbb{R}^N$ to be a minimizer of $J(x)$ is

$$\mathbf{0} \in \partial J(x). \tag{2.2}$$

When J is convex, (2.2) is also sufficient.

Definition 2. *A point that satisfies (2.2) is called (limiting) critical point. The set of critical points of* $J(x)$ *is denoted by* $\mathrm{crit}(J)$.

Proposition 2. *If* (x^*,y^*,p^*) *is a critical point of* $\mathcal{L}_\alpha(x,y,p)$ *with any* $\alpha>0$*, it must hold that*

$$\begin{array}{rcl} -B^\top p^* & \in & W^*\partial h(y^*), \\ -A^\top p^* & = & \nabla f(x^*), \\ Ax^*+By^*-c & = & \boldsymbol{0}, \end{array}$$

where $\mathcal{L}_\alpha(x,y,p)$ *is defined in (1.10) and* $W^*=\mathrm{Diag}\{g'[h(y_i^*)]\}_{1\leq i\leq N}$.

2.2 Kurdyka-Łojasiewicz function

The domain of a subdifferential is given as

$$\mathrm{dom}(\partial J):=\{x\in\mathbb{R}^N:\partial J(x)\neq\emptyset\}.$$

Definition 3. *(a) The function* $J:\mathbb{R}^N\to(-\infty,+\infty]$ *is said to have the Kurdyka-Łojasiewicz property at* $\overline{x}\in\mathrm{dom}(\partial J)$ *if there exist* $\eta\in(0,+\infty)$*, a neighborhood* U *of* $\overline{x}$ *and a continuous function* $\varphi:[0,\eta)\to\mathbb{R}^+$ *such that*

1. $\varphi(0)=0$.
2. φ *is* C^1 *on* $(0,\eta)$.
3. *for all* $s\in(0,\eta)$, $\varphi^{'}(s)>0$.
4. *for all* x *in* $U\bigcap\{x|J(\overline{x})<J(x)<J(\overline{x})+\eta\}$*, it holds*

$$\varphi^{'}(J(x)-J(\overline{x}))\cdot\mathrm{dist}(\boldsymbol{0},\partial J(x))\geq 1. \tag{2.3}$$

(b) Proper lower semicontinuous functions which satisfy the Kurdyka-Łojasiewicz property at each point of dom(∂J*) are called KL functions.*

The readers can find [25, 21, 3] for more details. In the following part of the paper, we use KL for Kurdyka-Łojasiewicz for short. Direct checking whether a function is KL or not is hard, but the semi-algebraic functions [3] do much help.

Definition 4. *(a) A subset* S *of* $\mathbb{R}^N$ *is a real semi-algebraic set if there exists a finite number of real polynomial functions* $g_{ij},h_{ij}:\mathbb{R}^N\to\mathbb{R}$ *such that*

$$S=\bigcup_{j=1}^{p}\bigcap_{i=1}^{q}\{u\in\mathbb{R}^N:g_{ij}(u)=0 \textit{ and } h_{ij}(u)<0\}.$$

(b) A function $h:\mathbb{R}^N\to(-\infty,+\infty]$ *is called semi-algebraic if its graph*

$$\{(u,t)\in\mathbb{R}^{N+1}:h(u)=t\}$$

is a semi-algebraic subset of $\mathbb{R}^{N+1}$.

Better yet, the semi-algebraicity enjoys many quite nice properties and various kinds of functions are KL [2]. We just put a few of them here:

- Real polynomial functions.
- Indicator functions of semi-algebraic sets.

- Finite sums and product of semi-algebraic functions.
- Composition of semi-algebraic functions.
- Sup/Inf type function, e.g., $\sup\{g(u,v) : v \in C\}$ is semi-algebraic when g is a semi-algebraic function and C a semi-algebraic set.
- Cone of PSD matrices, Stiefel manifolds and constant rank matrices.

Lemma 1 ([2])**.** *Let $J : \mathbb{R}^N \to \mathbb{R}$ be a proper and lower semicontinuous function. If J is semi-algebraic then it satisfies the KL property at any point of dom(J). In particular, if J is semi-algebraic and* $\mathrm{dom}(J) = \mathrm{dom}(\partial J)$*, then it is a KL function.*

The previous definition and property of KL is about a certain point in $\mathrm{dom}(J)$. In fact, the property has been extended to a certain closed set [4]. And this property makes previous convergence proofs related to KL property much easier.

Lemma 2. *Let $J : \mathbb{R}^N \to \mathbb{R}$ be a proper lower semi-continuous function and Ω be a compact set. If J is a constant on Ω and J satisfies the KL property at each point on Ω, then there exists function φ and $\eta, \varepsilon > 0$ such that for any $\overline{x} \in \Omega$ and any x satisfying that dist$(x, \Omega) < \varepsilon$ and $f(\overline{x}) < f(x) < f(\overline{x}) + \eta$, it holds that*

$$\varphi^{'}(J(x) - J(\overline{x})) \cdot \mathrm{dist}(\boldsymbol{0}, \partial J(x)) \geq 1. \tag{2.4}$$

3 Convergence analysis

In this part, the function $\mathcal{L}_\alpha(x, y, p)$ is defined in (1.10). We provide the convergence guarantee and the convergence analysis of ILR-ADMM (Algorithm 1). We first present a sketch of the proofs, which is also a big picture for the purpose of each lemma and theorem, :

- In the first step, we bound the dual variables by the primal points (Lemma 3).
- In the second step, the sufficient descent condition is derived for a new Lyapunov function (Lemma 4).
- In the third step, we provide several conditions to bound the points (Lemma 5).
- In the fourth step, the relative error condition is proved (Lemma 6).
- In the last step, we prove the convergence under semi-algebraic assumption (Thoerem 1).

Lemma 3. *If*

$$\mathrm{Im}(B) \bigcup \{c\} \subseteq \mathrm{Im}(A) \tag{3.1}$$

and

$$r_{2,i} = \alpha\sigma_i^2(A) + r_2, \tag{3.2}$$

where $r_2 > 0$. Then, we have

$$\|p^k - p^{k+1}\|_2^2 \leq \eta_1 \|x^{k+1} - x^k\|_2^2 + \eta_2 \|x^k - x^{k-1}\|_2^2, \tag{3.3}$$

where $\eta_1 = 2(\frac{r_2 + L_f}{\theta})^2$, $\eta_2 = 2(\frac{r_2}{\theta})^2$, and θ is the smallest strictly-positive eigenvalue of $(A^\top A)^{1/2}$.

Proof. The second step in each iteration actually gives

$$D(\vec{r_2})(x^{k+1} - x^k) + \nabla f(x^{k+1}) = -A^\top(\alpha(Ax^k + By^{k+1} - c) + p^k). \tag{3.4}$$

With the expression of p^{k+1},

$$D(\vec{r_2})(x^{k+1} - x^k) + \nabla f(x^{k+1}) = -A^\top p^{k+1} + \alpha A^\top A x^{k+1} - \alpha A^\top A x^k. \tag{3.5}$$

Replacing $k+1$ with k, we can obtain

$$D(\vec{r_2})(x^k - x^{k-1}) + \nabla f(x^k) = -A^\top p^k + \alpha A^\top A x^k - \alpha A^\top A x^{k-1}. \tag{3.6}$$

Under condition (3.1), $p^{k+1} - p^k \in \mathrm{Im}(A)$; and substraction of the two equations above gives

$$\begin{aligned}
&\|p^k - p^{k+1}\|_2 \leq \frac{1}{\theta}\|A^\top(p^k - p^{k+1})\|_2 \\
\leq\ & \frac{\|(D(\vec{r_2}) - \alpha A^\top A)(x^{k+1} - x^k)\|_2}{\theta} + \frac{\|(D(\vec{r_2}) - \alpha A^\top A)(x^k - x^{k-1})\|_2}{\theta} + \frac{\|\nabla f(x^{k+1}) - \nabla f(x^k)\|_2}{\theta} \\
\leq\ & \frac{r_2 + L_f}{\theta}\|x^{k+1} - x^k\|_2 + \frac{r_2}{\theta}\|x^k - x^{k-1}\|_2. \qquad (3.7)
\end{aligned}$$

In the third inequality, we used the fact

$$D(\vec{r_2}) - \alpha A^\top A = r_2 \mathbb{I}. \tag{3.8}$$

□

Remark 1. *If $p^0 \in Im(A)$, we have $p^k \in Im(A)$. Then, from (3.6), we have that*

$$\|p^k\|_2^2 \leq \frac{2}{\theta^2}\|\nabla f(x^k)\|_2^2 + \eta_2\|x^k - x^{k-1}\|_2^2. \tag{3.9}$$

We will use this inequality in bounding the sequence.

Remark 2. *In many applications, the matrix A is usually full row-rank.*

The condition (3.13) is satisfied if A is full row-rank. We return the example presented in Section 1.1, and we can see that T is full row-rank. In applications, the condition (3.13) can always hold. Now, we introduce several notations to present the following lemma. Denote the variable d and the sequence d^k as

$$d := (x, y, p, \tilde{x}), d^k := (x^k, y^k, p^k, x^{k-1}), z^k := (x^k, y^k). \tag{3.10}$$

The Lyapunov function is given as

$$\xi(d) := \mathcal{L}_\alpha(x, y, p) + \frac{\eta_2}{\alpha}\|x - \tilde{x}\|_2^2. \tag{3.11}$$

An auxiliary function is always used in the proof

$$\mathcal{L}_\alpha^k(x, y, p) := f(x) + \sum_{i=1}^{N} g'[h(y_i^k)]h(y_i) + \langle p, Ax + By - c\rangle + \frac{\alpha}{2}\|Ax + By - c\|_2^2. \tag{3.12}$$

Lemma 4 (Descent)**.** *Let the sequence $\{(x^k, y^k, p^k)\}_{k=0,1,2,\ldots}$ be generated by ILR-ADMM. If condition (3.1) and the following condition*

$$\min\{r_2 - 4\frac{(r_2 + L_f)^2}{\alpha\theta^2} - \frac{4r_2^2}{\alpha\theta^2}, r_1 - \alpha\|B\|_2^2\} > 0 \tag{3.13}$$

hold, then there exists $\nu > 0$ such that

$$\xi(d^k) - \xi(d^{k+1}) \geq \nu\|z^{k+1} - z^k\|_2^2. \tag{3.14}$$

Proof. Direct calculation shows that the first step is actually minimizing the function $\mathcal{L}_\alpha^k(x^k, y, p^k) + \frac{(y-y^k)(r_1\mathbb{I}-\alpha B^\top B)(y-y^k)}{2}$ with respect to y. Thus, we have

$$\begin{aligned}
\mathcal{L}_\alpha^k(x^k, y^{k+1}, p^k) \quad + &\quad \frac{r_1 - \alpha\|B\|_2^2}{2}\|y^{k+1} - y^k\|_2^2 \\
\leq &\quad \mathcal{L}_\alpha^k(x^k, y^{k+1}, p^k) + \frac{(y - y^k)(r_1\mathbb{I} - \alpha B^\top B)(y^{k+1} - y^k)}{2} \\
\leq &\quad \mathcal{L}_\alpha^k(x^k, y^k, p^k).
\end{aligned}$$

Similarly,

$$\mathcal{L}_\alpha^k(x^{k+1},y^{k+1},p^k)+\frac{r_2}{2}\|x^{k+1}-x^k\|_2^2\leq\mathcal{L}_\alpha^k(x^k,y^{k+1},p^k). \tag{3.15}$$

Direct calculation yields

$$\begin{aligned}\mathcal{L}_\alpha^k(x^{k+1},y^{k+1},p^{k+1}) &= \mathcal{L}_\alpha^k(x^{k+1},y^{k+1},p^k)+\langle p^{k+1}-p^k,Ax^{k+1}+By^{k+1}-c\rangle\\ &= \mathcal{L}_\alpha^k(x^{k+1},y^{k+1},p^k)+\frac{1}{\alpha}\|p^{k+1}-p^k\|_2^2. \end{aligned}\tag{3.16}$$

Combining the equations above, we can have

$$\begin{aligned}\mathcal{L}_\alpha^k(x^k,y^k,p^k) &\geq \mathcal{L}_\alpha^k(x^{k+1},y^{k+1},p^{k+1})+\frac{r_2}{2}\|x^{k+1}-x^k\|_2^2\\ &+ \frac{r_1-\alpha\|B\|_2^2}{2}\|y^{k+1}-y^k\|_2^2-\frac{1}{\alpha}\|p^{k+1}-p^k\|_2^2. \end{aligned}\tag{3.17}$$

Noting g is concave, we have

$$\begin{aligned}\sum_{i=1}^N g[h(y_i^k)]-\sum_{i=1}^N g[h(y_i^{k+1})] &= \sum_{i=1}^N\{g[h(y_i^k)]-g[h(y_i^{k+1})]\}\\ &\geq \sum_{i=1}^N g'[h(y_i^k)][h(y_i^k)-h(y_i^{k+1})]\\ &= \sum_{i=1}^N g'[h(y_i^k)]h(y_i^k)-\sum_{i=1}^N g'[h(y_i^k)]h(y_i^{k+1}). \end{aligned}\tag{3.18}$$

Then, we can derive

$$\begin{aligned}&\mathcal{L}_\alpha(x^k,y^k,p^k)-\mathcal{L}_\alpha(x^{k+1},y^{k+1},p^{k+1})\\ =\ &\sum_{i=1}^N g[h(y_i^k)]-\sum_{i=1}^N g[h(y_i^{k+1})]+f(x^k)+\langle p^k,Ax^k+By^k-c\rangle\\ +\ &\frac{\alpha}{2}\|Ax^k+By^k-c\|_2^2-\{f(x^{k+1})+\langle p^{k+1},Ax^{k+1}+By^{k+1}-c\rangle\\ +\ &\frac{\alpha}{2}\|Ax^{k+1}+By^{k+1}-c\|_2^2\}\\ \geq\ &\sum_{i=1}^N g'[h(y_i^k)]h(y_i^k)-\sum_{i=1}^N g'[h(y_i^k)]h(y_i^{k+1})+f(x^k)+\langle p^k,Ax^k+By^k-c\rangle\\ +\ &\frac{\alpha}{2}\|Ax^k+By^k-c\|_2^2-\{f(x^{k+1})+\langle p^{k+1},Ax^{k+1}+By^{k+1}-c\rangle\\ +\ &\frac{\alpha}{2}\|Ax^{k+1}+By^{k+1}-c\|_2^2\}\\ =\ &\mathcal{L}_\alpha^k(x^k,y^k,p^k)-\mathcal{L}_\alpha^k(x^{k+1},y^{k+1},p^{k+1})\\ \geq\ &\frac{r_2}{2}\|x^{k+1}-x^k\|_2^2+\frac{r_1-\alpha\|B\|_2^2}{2}\|y^{k+1}-y^k\|_2^2-\frac{1}{\alpha}\|p^{k+1}-p^k\|_2^2.\end{aligned}$$

With Lemma 3, we then have

$$\begin{aligned}\mathcal{L}_\alpha(x^k,y^k,p^k) &- \mathcal{L}_\alpha(x^{k+1},y^{k+1},p^{k+1})\geq\frac{r_2}{2}\|x^{k+1}-x^k\|_2^2+\frac{r_1-\alpha\|B\|_2^2}{2}\|y^{k+1}-y^k\|_2^2\\ &- \frac{\eta_1}{\alpha}\|x^{k+1}-x^k\|_2^2-\frac{\eta_2}{\alpha}\|x^k-x^{k-1}\|_2^2.\end{aligned}$$

Then, we can further obtain

$$
\begin{aligned}
& \mathcal{L}_\alpha(x^k, y^k, p^k) + \frac{\eta_2}{\alpha}\|x^k - x^{k-1}\|_2^2 \\
- \quad & (\mathcal{L}_\alpha(x^{k+1}, y^{k+1}, p^{k+1}) + \frac{\eta_2}{\alpha}\|x^{k+1} - x^k\|_2^2) \\
\geq \quad & (\frac{r_2}{2} - \frac{\eta_1 + \eta_2}{\alpha})\|x^{k+1} - x^k\|_2^2 + \frac{r_1 - \alpha\|B\|_2^2}{2}\|y^{k+1} - y^k\|_2^2.
\end{aligned}
$$

Letting $\nu := \min\{\frac{r_2}{2} - \frac{\eta_1+\eta_2}{\alpha}, \frac{r_1 - \alpha\|B\|_2^2}{2}\}$, we then prove the result. □

In fact, condition (3.13) can be always satisfied in applications because the parameters r_1, r_2 and α are all selected by the user. For example, we can set the parameters as

$$
r_1 = \alpha\|B\|_2^2 + 1, r_{2,i} = 1 + \alpha\sigma_i^2(A), i \in [1, 2, \dots, N], \alpha > \frac{4L_f^2 + 8L_f + 8}{\theta^2}. \tag{3.19}
$$

Different with the ADMMs in convex setting, the parameter α is nonarbitrary. After fixing r_1 and r_2, the α should be sufficiently large.

Lemma 5 (Boundedness)**.** *If $p^0 \in \mathrm{Im}(A)$ and conditions (3.1) and (3.13) hold, and there exists $\sigma_0 > 0$ such that*

$$
\inf\{f(x) - \sigma_0\|\nabla f(x)\|_2^2\} > -\infty, \tag{3.20}
$$

and

$$
0 < \sigma \leq \sigma_0, \frac{1}{\sigma\theta^2} \leq \alpha \leq \frac{2}{\sigma\theta^2}. \tag{3.21}
$$

The sequence $\{d^k\}_{k=0,1,2,\dots}$ is bounded, if one of the following conditions hold:
***B1**. $g(y)$ is coercive, and $f(x) - \sigma_0\|\nabla f(x)\|_2^2$ is coercive.*
***B2**. $g(y)$ is coercive, and A is invertible.*
***B3**. $\inf\{g(y)\} > -\infty$, $f(x) - \sigma_0\|\nabla f(x)\|_2^2$ is coercive, and A is invertible.*

Proof. We have

$$
\begin{aligned}
\mathcal{L}_\alpha(d^k) &= f(x^k) + g(y^k) + \langle p^k, Ax^k + By^k - c\rangle + \frac{\alpha}{2}\|Ax^k + By^k - c\|_2^2 + \frac{\eta_2}{\alpha}\|x^k - x^{k-1}\|_2^2 \\
&= f(x^k) + g(y^k) - \frac{\|p^k\|_2^2}{2\alpha} + \frac{\alpha}{2}\|Ax^k + By^k - c + \frac{p^k}{\alpha}\|_2^2 + \frac{\eta_2}{\alpha}\|x^k - x^{k-1}\|_2^2 \\
&= f(x^k) + g(y^k) - \frac{\sigma\theta^2}{2}\|p^k\|_2^2 + (\frac{\sigma\theta^2}{2} - \frac{1}{2\alpha})\|p^k\|_2^2 \\
&+ \frac{\alpha}{2}\|Ax^k + By^k - c + \frac{p^k}{\alpha}\|_2^2 + \frac{\eta_2}{\alpha}\|x^k - x^{k-1}\|_2^2 \\
\overset{(3.9)}{}&\geq f(x^k) + g(y^k) - \sigma_0\|\nabla f(x^k)\|_2^2 + \eta_2(\frac{1}{\alpha} - \frac{\sigma\theta^2}{2})\|x^k - x^{k-1}\|_2^2 + (\sigma_0 - \sigma)\|\nabla f(x^k)\|_2^2 \\
&+ (\frac{\sigma\theta^2}{2} - \frac{1}{2\alpha})\|p^k\|_2^2 + \frac{\alpha}{2}\|Ax^k + By^k - c + \frac{p^k}{\alpha}\|_2^2.
\end{aligned} \tag{3.22}
$$

We then can see $\{g(y^k)\}_{k=0,1,2,\dots}$, $\{p^k\}_{k=0,1,2,\dots}$, $\{Ax^k + By^k - c + \frac{p^k}{\alpha}\}_{k=0,1,2,\dots}$ are all bounded. It is easy to see that one of the three conditions holds, $\{d^k\}_{k=0,1,2,\dots}$ will be bounded. □

Remark 3. *The condition (3.20) holds for many quadratical functions [22, 29]. This condition also implies the function f is similar to quadratical function and its property is "good".*

Remark 4. *Both conditions **B2** and **B3** actually overlap condition (3.9). If using **B2** and **B3**, condition (3.9) could be absent.*

Remark 5. *The intersection between conditions (3.21) and (3.13) can be always nonempty. This is because we can always choose small σ. For example, we still use the setting (3.19). In this case, we just set that*

$$\sigma \leq \min\{\frac{1}{2L_f^2 + 4L_f + 4}, \sigma_0\}. \tag{3.23}$$

Lemma 6 (Relative error)**.** *If conditions (3.1) and (3.13) hold, for any $k \in \mathbb{Z}_+$, there exists $\tau > 0$ such that*

$$\mathrm{dist}(\boldsymbol{0}, \partial\xi(d^{k+1})) \leq \tau(\|z^k - z^{k+1}\|_2 + \|z^{k+1} - z^{k+2}\|_2). \tag{3.24}$$

Proof. Due to that h is convex, h is Lipschitz continuous with some constant if being restricted to some bounded set. Thus, there exists $L_h > 0$ such that

$$|h(y_i^{k+1}) - h(y_i^k)| \leq L_h |y_i^{k+1} - y_i^k|.$$

Updating y^{k+1} in each iteration certainly yields

$$r_1(y^k - y^{k+1}) - B^\top(\alpha(Ax^k + By^k - c) + p^k) \in W^k \partial h(y^{k+1}), \tag{3.25}$$

where $h(y) := \sum_{i=1}^N h(y_i)$ and $W^k := \mathrm{Diag}\{g'[h(y_1^k)], g'[h(y_2^k)], \ldots, g'[h(y_N^k)]\}$. Noting the boundedness of the sequence and $h(y_i^k)$, the continuity of g' indicates there exist $\delta_1, \delta_2 > 0$ such that

$$\delta_1 \leq g'[h(y_i^k)] \leq \delta_2, i \in [1, 2, \ldots, N], k \in \mathbb{Z}_+. \tag{3.26}$$

Easy computation gives

$$\begin{aligned} W^{k+1}(W^k)^{-1}[r_1(y^k - y^{k+1}) &- \alpha B^\top(Ax^k + By^k - c) - B^\top p^k] \\ &+ B^\top p^{k+1} + \alpha B^\top(Ax^{k+1} + By^{k+1} - c) \in \partial_y \xi(d^{k+1}). \end{aligned} \tag{3.27}$$

The left side of (3.27) can be rewritten as

$$\begin{aligned} &(W^{k+1} - W^k)(W^k)^{-1}[r_1(y^k - y^{k+1}) - \alpha B^\top(Ax^k + By^k - c) - B^\top p^k] \\ + \; &r_1(y^k - y^{k+1}) + B^\top(p^{k+1} - p^k) + \alpha B^\top[A(x^{k+1} - x^k) + B(y^{k+1} - y^k)] \in \partial_y \xi(d^{k+1}). \end{aligned} \tag{3.28}$$

With the boundedness of the generated points, there exist $R_1 > 0$ such that

$$\|r_1(y^k - y^{k+1}) - \alpha B^\top(Ax^k + By^k - c) - B^\top p^k\|_2 \leq R_1. \tag{3.29}$$

Thus, we have

$$\begin{aligned} \mathrm{dist}(\mathbf{0}, \partial_y \xi(d^{k+1})) &\leq \frac{R_1}{\delta_1}\|W^{k+1} - W^k\|_2 + \|B^\top p^{k+1} - B^\top p^k\|_2 \\ &+ \|\alpha B^\top A(x^{k+1} - x^k)\|_2 + \|\alpha B^\top B(y^{k+1} - y^k)\|_2 + r_1\|y^{k+1} - y^k\|_2 \\ &\leq \frac{R_1}{\delta_1}\|W^{k+1} - W^k\|_2 + \|B\|_2 \cdot \|p^{k+1} - p^k\|_2 \\ &+ \alpha\|B\|_2 \cdot \|A\|_2 \cdot \|x^{k+1} - x^k\|_2 + (\alpha\|B\|_2^2 + r_1) \cdot \|y^{k+1} - y^k\|_2. \end{aligned} \tag{3.30}$$

Obviously, it holds that

$$\begin{aligned} \|W^{k+1} - W^k\|_2 &\leq \max_i |g'[h(y_i^{k+1})] - g'[h(y_i^k)]| \\ &\leq L_g \max_i |h(y_i^{k+1}) - h(y_i^k)| \\ &\leq L_g L_h \|y^{k+1} - y^k\|_\infty \leq L_g L_h \|y^{k+1} - y^k\|_2. \end{aligned}$$

Thus, with Lemma 3, we derive that

$$\mathrm{dist}(\mathbf{0}, \partial_y \xi(d^{k+1})) \leq \tau_y(\|z^{k+1} - z^k\|_2 + \|z^k - z^{k-1}\|_2), \tag{3.31}$$

for $\tau_y = \max\{\frac{L_g L_h R_1}{\delta_1} + \alpha\|B\|_2^2 + r_1, \alpha\|B\|_2\|A\|_2 + \|B\|_2\sqrt{\eta_1}, \|B\|_2\sqrt{\eta_2}\}$.

From the second step in each iteration,

$$\nabla f(x^{k+1}) = r_2(x^k - x^{k+1}) - A^\top p^k - \alpha A^\top(Ax^k + By^{k+1} - c). \tag{3.32}$$

Direct calculation gives

$$\nabla f(x^{k+1}) + A^\top p^{k+1} + \alpha A^\top(Ax^{k+1} + By^{k+1} - c) + \frac{2\eta_2}{\alpha}(x^{k+1} - x^k) \in \partial_x \xi(d^{k+1}). \tag{3.33}$$

With Lemma 3, we have

$$\begin{aligned}
\mathrm{dist}(\mathbf{0}, \partial_x \xi(d^{k+1})) &\leq \|\alpha A^\top A(x^{k+1} - x^k)\|_2 + \|A^\top p^{k+1} - A^\top p^k\|_2 + \frac{2\eta_2}{\alpha}\|x^{k+1} - x^{k+2}\|_2 \\
&\leq \|A\|_2 \cdot \|p^{k+1} - p^k\|_2 + (\frac{2\eta_2}{\alpha} + \alpha\|A\|_2^2 + r_2)\|x^{k+1} - x^k\|_2 \\
&\leq \tau_x(\|z^{k+1} - z^k\|_2 + \|z^k - z^{k-1}\|_2),
\end{aligned}$$

where $\tau_x = \max\{\frac{2\eta_2}{\alpha} + \alpha\|A\|_2^2 + r_2 + \|A\|\sqrt{\eta_1}, \|A\|\sqrt{\eta_2}\}$.

It is easy to see that

$$Ax^{k+1} + By^{k+1} - c \in \partial_p \xi(d^{k+1}). \tag{3.34}$$

And we have

$$\mathrm{dist}(\mathbf{0}, \partial_p \xi(d^{k+1})) \leq \tau_p(\|z^{k+1} - z^k\|_2 + \|z^k - z^{k-1}\|_2), \tag{3.35}$$

for $\tau_p = \max\{\frac{\sqrt{\eta_1}}{\alpha}, \frac{\sqrt{\eta_2}}{\alpha}\}$.

Noting that

$$\frac{2\eta_2}{\alpha}(x^{k-1} - x^k) \in \partial_{\tilde{x}} \xi(d^{k+1}), \tag{3.36}$$

we then have that

$$\mathrm{dist}(\mathbf{0}, \partial_{\tilde{x}} \xi(d^{k+1})) \leq \frac{2\eta_2}{\alpha}(\|z^{k+1} - z^k\|_2 + \|z^k - z^{k-1}\|_2). \tag{3.37}$$

With the deductions above,

$$\mathrm{dist}(\mathbf{0}, \partial \xi(d^{k+1})) \quad \leq \quad (\tau_x + \tau_y + \tau_p + \frac{2\eta_2}{\alpha})(\|z^{k+1} - z^k\|_2 + \|z^k - z^{k-1}\|_2). \tag{3.38}$$

Denote that $\tau = \tau_x + \tau_y + \tau_p + \frac{2\eta_2}{\alpha}$, we then finish the proof. □

Lemma 7. *If the sequence* $\{(x^k, y^k, p^k)\}_{k=0,1,2,\ldots}$ *is bounded and conditions (3.1) and (3.13) hold, then we have*

$$\lim_k \|z^{k+1} - z^k\|_2 = 0. \tag{3.39}$$

For any cluster point (x^*, y^*, p^*)*, it is also a critical point of* $\mathcal{L}_\alpha(x, y, p)$.

Proof. We can easily see that $\{d^k\}_{k=0,1,2,\ldots}$ is also bounded. The continuity of ξ indicates that $\{\xi(d^k)\}_{k=0,1,2,\ldots}$ is bounded. From Lemma 4, $\xi(d^k)$ is decreasing. Thus, the sequence $\{\xi(d^k)\}_{k=0,1,2,\ldots}$ is convergent, i.e., $\lim_k[\xi(d^k) - \xi(d^{k+1})] = 0$. With Lemma 4, we have

$$\lim_k \|z^{k+1} - z^k\|_2 \leq \lim_k \sqrt{\frac{\xi(d^k) - \xi(d^{k+1})}{\nu}} = 0. \tag{3.40}$$

From the scheme of the ILR-ADMM, we also have

$$\lim_k \|p^{k+1} - p^k\|_2 = 0. \tag{3.41}$$

For any cluster point (x^*, y^*, p^*), there exists $\{k_j\}_{j=0,1,2,\ldots}$ such that $\lim_j (x^{k_j}, y^{k_j}, p^{k_j}) = (x^*, y^*, z^*)$. Then, we further have $\lim_j z^{k_j+1} = (x^*, y^*)$. From Lemma 3, we also have $\lim_j p^{k_j+1} = p^*$. That also means

$$\lim_j W^{k_j} = W^*. \tag{3.42}$$

From the scheme, we have the following conditions

$$\begin{aligned} (W^{k_j})^{-1}[r_1(y^{k_j} - y^{k_j+1}) &- B^\top(\alpha(Ax^{k_j} + By^{k_j} - c) + p^{k_j})] \in \partial h(y^{k_j+1}), \\ r_2(x^{k_j} - x^{k_j+1}) - A^\top p^{k_j} &- \alpha A^\top(Ax^{k_j} + By^{k_j+1} - c) = \nabla f(x^{k_j+1}), \\ p^{k_j+1} = p^{k_j} &+ \alpha(Ax^{k_j+1} + By^{k_j+1} - c). \end{aligned}$$

Letting $j \to +\infty$, with Proposition 1, we have

$$\begin{aligned} (W^*)^{-1}[-B^\top p^*] &\in \partial h(y^*), \\ -A^\top p^* &= \nabla f(x^*), \\ Ax^* + By^* - c &= \mathbf{0}. \end{aligned}$$

The first relation above is actually $-B^\top p^* \in W^* \partial h(y^*)$. From Proposition 2, z^* is a critical point of L. □

In the following, to prove the convergence result, we first establish some results about the limit points of the sequence generated by ILR-ADMM. There results are presented for the use of Lemma 2. We recall a definition about the limit point which is introduced in [4].

Definition 5. *Define that*

$$\mathcal{M}(d^0) := \{u \in \mathbb{R}^N : \exists \text{ an increasing sequence of integers } \{k_j\}_{j\in\mathbb{N}} \text{ such that } d^{k_j} \to u \text{ as } j \to \infty\},$$

where $d^0 \in \mathbb{R}^n$ is an arbitrary starting point.

Lemma 8. *Suppose that $\{d^k\}_{k=0,1,2,\ldots}$ is generated by scheme (1.11). Then, we have the following results.*

(1) $\mathcal{M}(d^0)$ is nonempty and $\mathcal{M}(d^0) \subseteq \mathrm{cri}(\xi)$.

(2) $\lim_k \mathrm{dist}(d^k, \mathcal{M}(d^0)) = 0$.

(3) The objective function is finite and constant on $\mathcal{M}(d^0)$.

Proof. (1) Due to that $\{d^k\}_{k=0,1,2,\ldots}$ is bounded, $\mathcal{M}(d^0)$ is nonempty. Assume that $d^* \in \mathcal{M}(d^0)$, from the definition, there exists a subsequence $d^{k_i} \to d^*$. From Lemma 4, we have $d^{k_i-1} \to d^*$. From Lemma 6, we have $\omega^{k_i} \in \partial\xi(d^{k_i})$ and $\omega^{k_i} \to \mathbf{0}$. Proposition 1 indicates that $\mathbf{0} \in \partial\xi(d^*)$, i.e. $d^* \in \mathrm{cri}(\xi)$.

(2) This item follows as a consequence of the definition of the limit point.

(3) The continuity of ξ directly yields this result. □

Theorem 1 (Convergence result)**.** *Suppose that f and g are all semi-algebraic functions and* $\mathrm{dom}(f) = \mathrm{dom}(\partial f)$, *and* $\mathrm{dom}(g) = \mathrm{dom}(\partial g)$. *Assume that conditions (3.1), (3.2), (3.13), (3.20), (3.21) and one of **B1, B2, B3** hold. Let the sequence $\{(x^k, y^k, p^k)\}_{k=1,2,3,\ldots}$ generated by ILR-ADMM be bounded. Then, the sequence $\{z^k = (x^k, y^k)\}_{k=1,2,3,\ldots}$ has finite length, i.e.*

$$\sum_{k=0}^{+\infty} \|z^{k+1} - z^k\|_2 < +\infty. \tag{3.43}$$

And $\{(x^k, y^k)\}_{k=1,2,3,\ldots}$ converges to (x^*, y^*), where (x^*, y^*, p^*) is a critical point of $\mathcal{L}_\alpha(x, y, p)$.

Proof. Obviously, ξ is also semi-algebraic. And with Lemma 1, ξ is KL. From Lemma 8, ξ is constant on $\mathcal{M}(d^0)$. Let d^* be a stationary point of $\{d^k\}_{k=0,1,2,\dots}$. Also from Lemma 8, we have $\mathrm{dist}(d^k,\mathcal{M}(d^0))<\varepsilon$ and $\xi(d^k)<\xi(d^*)+\eta$ if any $k>K$ for some K. Hence, from Lemma 2, we have

$$\mathrm{dist}(\mathbf{0},\partial\xi(d^k))\cdot\varphi'(\xi(d^k)-\xi(d^*))\geq 1, \tag{3.44}$$

which together with Lemma 6 gives

$$\frac{1}{\varphi'(\xi(d^k)-\xi(d^*))}\leq \mathrm{dist}(\mathbf{0},\partial\xi(d^k))\leq\tau(\|z^k-z^{k-1}\|_2+\|z^{k+1}-z^k\|_2). \tag{3.45}$$

Then, the concavity of φ yields

$$\begin{aligned}
\xi(d^k)-\xi(d^{k+1}) &= \xi(d^k)-\xi(d^*)-[\xi(d^{k+1})-\xi(d^*)]\\
&\leq \frac{\varphi[\xi(d^k)-\xi(d^*)]-\varphi[\xi(d^{k+1})-\xi(d^*)]}{\varphi'[\xi(d^k)-\xi(d^*)]}\\
&\leq \{\varphi[\xi(d^k)-\xi(d^*)]-\varphi[\xi(d^{k+1})-\xi(d^*)]\}\times\tau(\|z^k-z^{k-1}\|_2+\|z^{k+1}-z^k\|_2).
\end{aligned}$$

With Lemma 4, we have

$$\nu\|z^{k+1}-z^k\|_2^2\leq\{\varphi[\xi(d^k)-\xi(d^*)]-\varphi[\xi(d^{k+1})-\xi(d^*)]\}\times\tau(\|z^k-z^{k-1}\|_2+\|z^{k+1}-z^k\|_2),$$

which is equivalent to

$$\begin{aligned}
3\frac{\nu}{\tau}\|z^{k+1}-z^k\|_2 &\leq 2\times\frac{3}{2}\sqrt{\varphi[\xi(d^k)-\xi(d^*)]-\varphi[\xi(d^{k+1})-\xi(d^*)]}\\
&\times\sqrt{\frac{\nu}{\tau}}\sqrt{\|z^k-z^{k-1}\|_2+\|z^{k+1}-z^k\|_2}.
\end{aligned}\tag{3.46}$$

Using the Schwartz's inequality, we then derive that

$$\begin{aligned}
3\frac{\nu}{\tau}\|z^{k+1}-z^k\|_2 &\leq \frac{9}{4}\{\varphi[\xi(d^k)-\xi(d^*)]-\varphi[\xi(d^{k+1})-\xi(d^*)]\}\\
&+\frac{\nu}{\tau}(\|z^k-z^{k-1}\|_2+\|z^{k+1}-z^k\|_2).
\end{aligned}\tag{3.47}$$

Summing (3.47) from K to $K+j$ yields that

$$\begin{aligned}
\frac{\nu}{\tau}\sum_{k=K}^{K+j-1}\|z^{k+1}-z^k\|_2 &+ \frac{2\nu}{\tau}\|z^{K+j+1}-z^{K+j}\|_2\\
&\leq \frac{9}{4}\varphi[\xi(d^K)-\xi(d^*)]-\frac{9}{4}\varphi[\xi(d^{K+j+1})-\xi(d^*)].
\end{aligned}\tag{3.48}$$

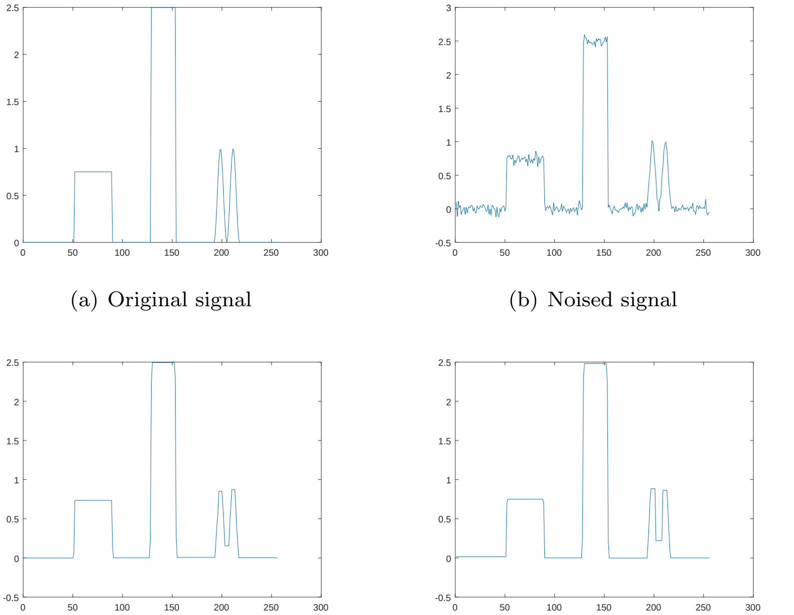
Letting $j\to+\infty$, with Lemma 6, we have

$$\frac{\nu}{\tau}\sum_{k=K}^{+\infty}\|z^{k+1}-z^k\|_2\leq\frac{9}{4}\varphi[\xi(d^K)-\xi(d^*)]<+\infty. \tag{3.49}$$

From Lemma 8, there exists a critical point (x^*,y^*,p^*) of $\mathcal{L}_\alpha(x,y,p)$. Then, $\{z^k\}_{k=0,1,2,\dots}$ is convergent and (x^*,y^*) is a stationary point of $\{z^k\}_{k=0,1,2,\dots}$. That is to say $\{z^k\}_{k=0,1,2,\dots}$ converges to (x^*,y^*). □

4 Applications and numerical results

In this part, we consider using (1.4) for signal and image denoising. Considering the proximal map of $\|\cdot\|_{q,\varepsilon}^q$ is easy to derive if $q=\frac{1}{2},\frac{2}{3}$, and hard when $q\neq\frac{1}{2},\frac{2}{3}$. The numerical results in this section consist of two parts: the first one is about the case $q=\frac{1}{2}$; and the second one is about $q\in(0,\frac{1}{2})\bigcup(\frac{1}{2},1)$.



(a) Original signal

(b) Noised signal

(c) Recovery by ILR-ADMM, SNR=76.3dB

(d) Recovery by $L - 1/2$ ADMM, SNR=74.6dB

Figure 1: Recovered signal and original signal

4.1 The parameter $q = \frac{1}{2}$

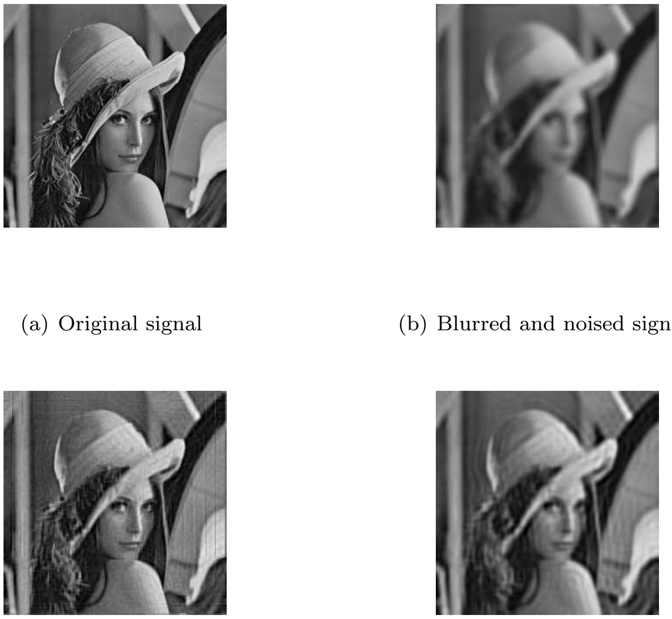
In fact, the proximal map of $\|\cdot\|_{q,\varepsilon}^q$ also exists when $q = \frac{2}{3}$. But it is a little harder than the case $q = \frac{1}{2}$. This part just wants to demonstrate the performance of the proposed algorithm when the direct use of the ADMM without any computational hindrance for solving the subproblems. Thus, we just present the case $q = \frac{1}{2}$. Two problems are considered in this part. In the first one, we consider the case $\Psi = I$. In the iteration, the parameters are chose as $q = 1/2$, $\varepsilon = 0.001$. The observed vector is $b = x^* + e$. We compare Algorithm 2 with the direct use of ADMM for the TV-$\frac{1}{2}$ denoising. We can directly use the algorithm because the proximal map of $\|\cdot\|_{\frac{1}{2},\varepsilon}^{\frac{1}{2}}$ is easy to obtain. We choose the parameter $\sigma = 0.1$. The noise $e \in \mathbb{R}^{256}$ is generated by a random variable. Figure 1 presents the numerical results of a 1-dimensional signal denoising. From the results, we can see that ILR-ADMM performs better.

In the second test, we consider the deblurring problem. The blurring operator Ψ is generated by the Matlab order `fspecial('gaussian',.,.)`. The parameter is set as $\sigma = 10^{-4}$, e is generated by the Gaussian noise $\mathcal{N}(0, 0.01)$ and $b = \Psi x^* + e$. Figure 2 presents the numerical results of a 2-dimensional signal deblurring and denoising. And we can see that the ILR-ADMM outperforms the direct nonconvex ADMM.

4.2 The parameter $q \neq \frac{1}{2}, \frac{2}{3}$

This subsection contains three parts. In the first one, we present the performance of the algorithm for different q; the second one is the comparison with the nonconvex ADMM. Because the proximal map of $\|\cdot\|_{q,\varepsilon}^q$ does not enjoy explicit format, an inner loop is used for the subproblem. Precisely, the inner loop is constructed by the proximal reweighted algorithm.

In the first test, the blurring operator Ψ is generated by the Matlab order `fspecial('motion',.,.)`. The parameter is set as $\sigma = 10^{-4}$, e is generated by the Gaussian noise $\mathcal{N}(0, 0.01)$ and $b = \Psi x^* + e$. For $q = 0.2, 0.4, 0.6, 0.8$, Figure 3 presents the deblurred images.



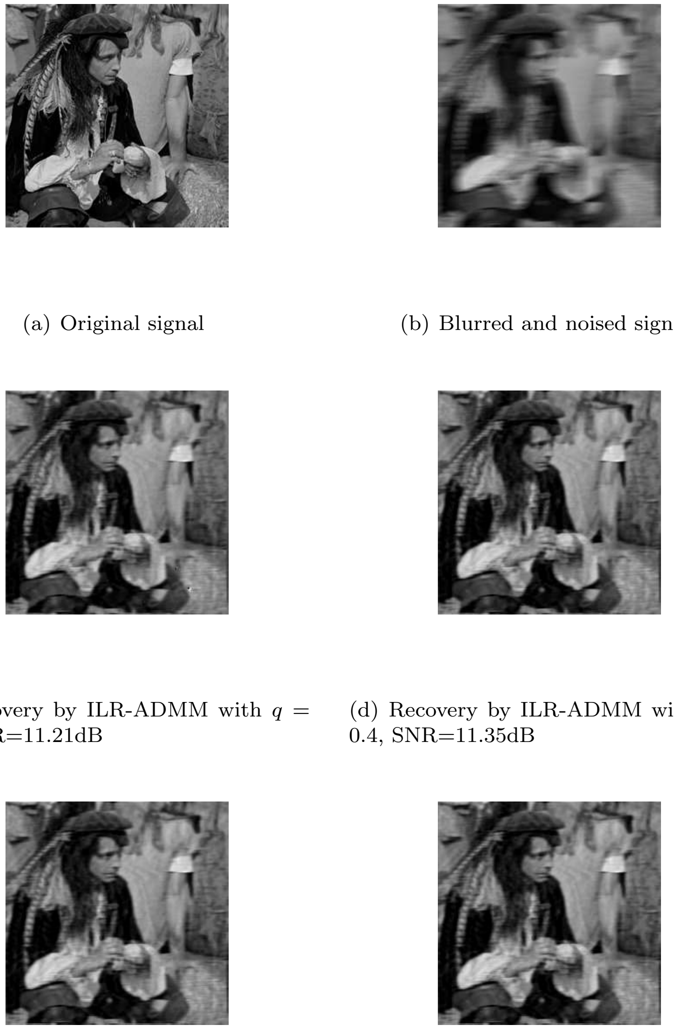
(a) Original signal

(b) Blurred and noised signal

(c) Recovery by ILR-ADMM, SNR=13.09dB

(d) Recovery by $\frac{1}{2}$-ADMM, SNR=11.55dB

Figure 2: Recovered signal and original signal



(a) Original signal

(b) Blurred and noised signal

(c) Recovery by ILR-ADMM with $q = 0.2$, SNR=11.21dB

(d) Recovery by ILR-ADMM with $q = 0.4$, SNR=11.35dB

(e) Recovery by ILR-ADMM with $q = 0.6$, SNR=11.42dB

(f) Recovery by ILR-ADMM with $q = 0.8$, SNR=11.37dB

Figure 3: Recovered images with different q

In the second test, we compare our algorithm with the nonconvex ADMM. We focus on the case $q = 0.4$. The blurring operator Ψ is generated by the Matlab orders `fspecial('gaussian',.,.)`. The parameter is set as $\sigma = 10^{-4}$, e is generated by the Gaussian noise $\mathcal{N}(0, 0.01)$. The result reconstructed by the convex method (i.e., $q = 1$) is also reported. Figure 4 presents the results with different algorithms. And we can see that ILR-ADMM performs best in the perspective of SNR. ILR-ADMM is faster than the nonconvex ADMM due to that the nonconvex ADMM employs inner loops. And LIR-ADMM is just a little slower than the convex ADMM but with a better effect.

5 Conclusion

In this paper, we consider a class of nonconvex and nonsmooth minimizations with linear constrains which have applications in signal processing and machine learning research. The classical ADMM method for these problems always encounter both computational and mathematical barriers in solving the subproblem. We organically combined the reweighted algorithm and linearized techniques, and then designed a new ADMM. In the proposed algorithm, each subproblem just needs to calculate the proximal maps. The convergence is proved under several assumptions on the parameters and functions. And numerical results demonstrate the efficiency of our algorithm.

Acknowledgments

We are grateful for the support from the National Science Foundation of China (No.61402495).

References

[1] Brendan PW Ames and Mingyi Hong. Alternating direction method of multipliers for penalized zero-variance discriminant analysis. *Computational Optimization and Applications*, 64(3):725–754, 2016.

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

[3] Jérôme Bolte, Aris Daniilidis, and Adrian Lewis. The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17(4):1205–1223, 2007.

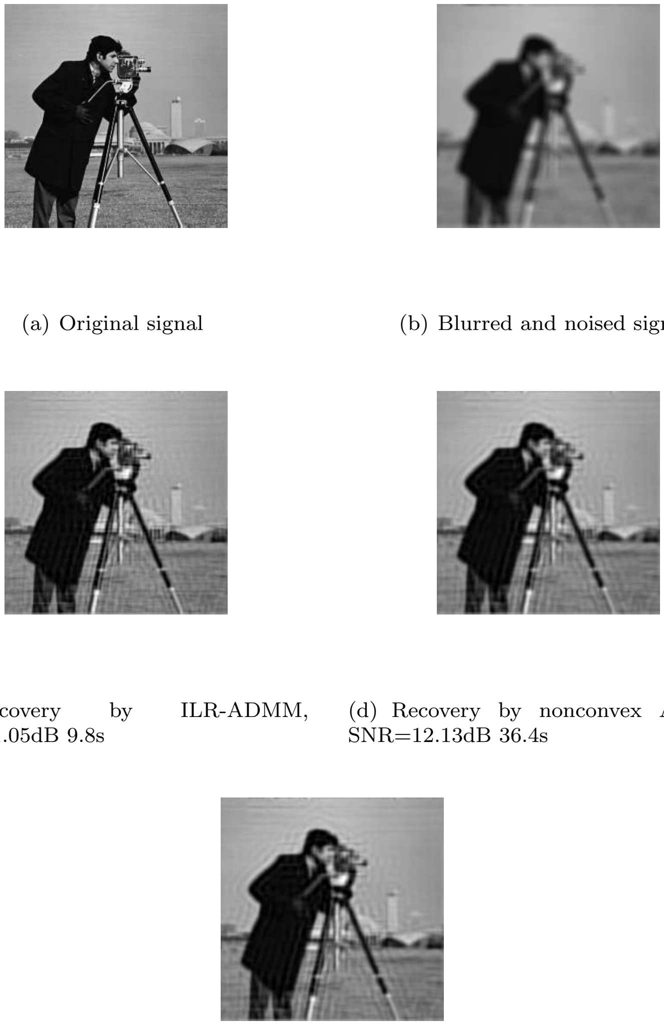
[4] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.

[5] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

[6] Rick Chartrand and Brendt Wohlberg. A nonconvex admm algorithm for group sparsity with sparse groups. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 6009–6013. IEEE, 2013.

[7] Rick Chartrand and Wotao Yin. Iteratively reweighted algorithms for compressive sensing. In *Acoustics, speech and signal processing, 2008. ICASSP 2008. IEEE international conference on*, pages 3869–3872. IEEE, 2008.

[8] Gong Chen and Marc Teboulle. A proximal-based decomposition method for convex minimization problems. *Mathematical Programming*, 64(1-3):81–101, 1994.



(a) Original signal

(b) Blurred and noised signal

(c) Recovery by ILR-ADMM, SNR=11.05dB 9.8s

(d) Recovery by nonconvex ADMM, SNR=12.13dB 36.4s

(e) Recovery by convex ADMM, SNR=10.84dB 7.7s

Figure 4: Reconstructed images and time cost by different methods

[9] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C Sinan Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63(1):1–38, 2010.

[10] Wei Deng and Wotao Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. *Journal of Scientific Computing*, 66(3):889–916, 2016.

[11] Ernie Esser, Xiaoqun Zhang, and Tony F Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.

[12] Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976.

[13] Cuixia Gao, Naiyan Wang, Qi Yu, and Zhihua Zhang. A feasible nonconvex relaxation approach to feature selection. In *Aaai*, pages 356–361, 2011.

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

[15] Roland Glowinski and A Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, 9(R2):41–76, 1975.

[16] Bingsheng He and Xiaoming Yuan. On the o(1/n) convergence rate of the douglas–rachford alternating direction method. *SIAM Journal on Numerical Analysis*, 50(2):700–709, 2012.

[17] Bingsheng He and Xiaoming Yuan. On non-ergodic convergence rate of douglas–rachford alternating direction method of multipliers. *Numerische Mathematik*, 130(3):567–577, 2015.

[18] Michael Hintermüler and Tao Wu. Nonconvex tv^q-models in image restoration: Analysis and a trust-region regularization–based superlinearly convergent solver. *SIAM Journal on Imaging Sciences*, 6(3):1385–1415, 2013.

[19] Mingyi Hong and Zhi-Quan Luo. On the linear convergence of the alternating direction method of multipliers. *Mathematical Programming*, 162(1-2):165–199, 2017.

[20] Mingyi Hong, Zhi-Quan Luo, and Meisam Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM Journal on Optimization*, 26(1):337–364, 2016.

[21] Krzysztof Kurdyka. On gradients of functions definable in o-minimal structures. In *Annales de l'institut Fourier*, volume 48, pages 769–784. Chartres: L'Institut, 1950-, 1998.

[22] Guoyin Li and Ting Kei Pong. Global convergence of splitting methods for nonconvex composite optimization. *SIAM Journal on Optimization*, 25(4):2434–2460, 2015.

[23] Guoyin Li and Ting Kei Pong. Douglas–rachford splitting for nonconvex optimization with application to nonconvex feasibility problems. *Mathematical programming*, 159(1-2):371–401, 2016.

[24] Yanli Liu, Ernest K Ryu, and Wotao Yin. A new use of douglas-rachford splitting and admm for identifying infeasible, unbounded, and pathological conic programs. *arXiv preprint arXiv:1706.02374*, 2017.

[25] Stanislas Łojasiewicz. Sur la géométrie semi-et sous-analytique. *Ann. Inst. Fourier*, 43(5):1575–1595, 1993.

[26] Michael K Ng, Pierre Weiss, and Xiaoming Yuan. Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods. *SIAM journal on Scientific Computing*, 32(5):2710–2736, 2010.

[27] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[28] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

[29] Tao Sun, Roberto Barrio, Lizhi Cheng, and Hao Jiang. Precompact convergence of the nonconvex primal-dual hybrid gradient algorithm. *Journal of Computational and Applied Mathematics*, 2017.

[30] Tao Sun, Hao Jiang, and Lizhi Cheng. Convergence of proximal iteratively reweighted nuclear norm algorithm for image processing. *IEEE Transactions on Image Processing*, 2017.

[31] Tao Sun, Hao Jiang, and Lizhi Cheng. Global convergence of proximal iteratively reweighted algorithm. *Journal of Global Optimization*, pages 1–12, 2017.

[32] Tao Sun, Penghang Yin, Lizhi Cheng, and Hao Jiang. Alternating direction method of multipliers with difference of convex functions. *Advances in Computational Mathematics*, pages 1–22, 2017.

[33] Joshua Trzasko and Armando Manduca. Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization. *IEEE Transactions on Medical imaging*, 28(1):106–121, 2009.

[34] Huahua Wang and Arindam Banerjee. Bregman alternating direction method of multipliers. In *Advances in Neural Information Processing Systems*, pages 2816–2824, 2014.

[35] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[36] Yu Wang, Wotao Yin, and Jinshan Zeng. Global convergence of admm in nonconvex nonsmooth optimization. *arXiv preprint arXiv:1511.06324*, 2015.

[37] Zaiwen Wen, Donald Goldfarb, and Wotao Yin. Alternating direction augmented lagrangian methods for semidefinite programming. *Mathematical Programming Computation*, 2(3):203–230, 2010.

[38] Jason Weston, André Elisseeff, Bernhard Schölkopf, and Mike Tipping. Use of the zero-norm with linear models and kernel methods. *Journal of machine learning research*, 3(Mar):1439–1461, 2003.

[39] Yangyang Xu, Wotao Yin, Zaiwen Wen, and Yin Zhang. An alternating direction algorithm for matrix completion with nonnegative factors. *Frontiers of Mathematics in China*, 7(2):365–384, 2012.

[40] Junfeng Yang, Yin Zhang, and Wotao Yin. A fast alternating direction method for tvl1-l2 signal reconstruction from partial fourier data. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):288–297, 2010.

[41] Tong Zhang. Analysis of multi-stage convex relaxation for sparse regularization. *Journal of Machine Learning Research*, 11(Mar):1081–1107, 2010.