A convergence frame for inexact nonconvex and nonsmooth algorithms and its applications to several iterations

Tao Sun* Hao Jiang† Lizhi Cheng‡ Wei Zhu§

September 12, 2017

Abstract

In this paper, we consider the convergence of an abstract inexact nonconvex and nonsmooth algorithm. We promise a *pseudo sufficient descent condition* and a *pseudo relative error condition*, which both are related to an auxiliary sequence, for the algorithm; and a continuity condition is assumed to hold. In fact, a wide of classical inexact nonconvex and nonsmooth algorithms allow these three conditions. Under the finite energy assumption on the auxiliary sequence, we prove the sequence generated by the general algorithm converges to a critical point of the objective function if being assumed Kurdyka-Łojasiewicz property. The core of the proofs lies on building a new Lyapunov function, whose successive difference provides a bound for the successive difference of the points generated by the algorithm. And then, we apply our findings to several classical nonconvex iterative algorithms and derive corresponding convergence results.

Keywords: Nonconvex minimization; Inexact algorithms; Semi-algebraic functions; Kurdyka-Łojasiewicz property; Convergence analysis

Mathematical Subject Classification 90C30, 90C26, 47N10

1 Introduction

Minimization of a nonconvex and nonsmooth function

$$\min_x F(x) \tag{1.1}$$

is a core part of nonlinear programming and applied mathematics. In most practical cases, the objective functions enjoy the Kurdyka-Łojasiewicz property (see definitions in Sec. 2). Thus, the convergence analysis of the nonconvex algorithms discussed by the literature is always under the Kurdyka-Łojasiewicz property assumption on the objective function. But different with convergence routine results of the global minimizers in the convex community, the convergence of the nonconvex algorithm basically just promises that the sequence falls into a critical point.

A very general result is presented by paper [4]: for the sequence $(x^k)_{k\geq 0}$ generated a very general scheme for problem (1.1), the authors assume three conditions, *sufficient descent condition*, *relative error condition* and *continuity condition*, for some $a>0, c>0$

$$\left\{\begin{array}{c} F(x^k)-F(x^{k+1})\geq a\|x^{k+1}-x^k\|^2, \\ \mathrm{dist}(\mathbf{0},\partial F(x^{k+1}))\leq c\|x^{k+1}-x^k\|, \\ \text{for any stationary point } x^* \text{ there exists subsequence } (x^{k_j})_{j\geq 0}\to x^* \text{ satisfying } F(x^{k_j})\to F(x^*) \end{array}\right. \tag{1.2}$$

*Department of Mathematics and Systems 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

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

§Hunan Key Laboratory for Computation and Simulation in Science and Engineering, School of Mathematics and Computational Science, Xiangtan University, Xiangtan, Hunan, 411105, China. Email: zhuwei@xtu.edu.cn

where ∂F means the limiting subdifferential of F (see definition in Sec. 2). Actually, various algorithms satisfy these three conditions. The third condition is usually derived by the minimization in each iteration, which can give an upper semi-continuity for the generated sequence. The proofs in [4] use a *local area analysis*; the authors prove that the sequence falling into an oracle of some point after enough iterations and then employ the Kurdyka-Łojasiewicz property on the point. In latter paper [8], the authors prove a uniformed Kurdyka-Łojasiewicz lemma for a closed set and much simplify the proofs.

In this paper, we consider the convergence for inexact nonconvex and nonsmooth algorithms. *We stress that the inexact algorithms discussed in our paper are different from the paper [4].* In their paper, they an assumption is posed for the noise: the noise should be bounded by the successive difference of the iteration. The "inexact algorithm" in [4] is much more closed to "proximal algorithm". For example, if F is differentiable (may be nonconvex), the gradient descent algorithm performs as

$$x^{k+1} = x^k - h \cdot \nabla F(x^k). \tag{1.3}$$

If the gradient of f is Lipschitz with L and $0 < h < \frac{1}{L}$, the sequence $(x^k)_{k\geq 0}$ generated by (1.3) satisfies condition (1.2). However, if the iteration is corrupted by some noise e^k in each step, i.e.,

$$x^{k+1} = x^k - h \cdot \nabla F(x^k) + e^k. \tag{1.4}$$

Unfortunately, the sequence $(x^k)_{k\geq 0}$ generated by (1.4) does not satisfy any one of conditions (1.2) if $e^k \neq \mathbf{0}$. The existing analysis cannot be directly used for the algorithm (1.4). The authors in [4] propose the assumption for the noise as

$$\|e^k\| \leq \ell \cdot \|x^{k+1} - x^k\|, \tag{1.5}$$

where $\ell > 0$. Under this assumption, they can continue using the sufficient descent condition and relative error condition. Besides this, the sequence is also assumed to satisfy the sufficient descent condition. In this paper, we get rid of the dependent assumption like (1.5) and the sufficient descent condition.

This paper is devoted to the convergence analysis for inexact algorithms like (1.4). The inexact algorithms like (1.4) are deeply studied in the convex community [30, 11, 19, 22]. However, for the nonconvex problems, to our best knowledge, it is the first time to discuss the convergence. Although the inexact algorithms always fail to obey the first two of the core condition (1.2), we find that many of them satisfy an alternative condition:

$$\begin{cases} F(x^k) - F(x^{k+1}) \geq a\|\omega^{k+1} - \omega^k\|^2 - b\eta_k^2 \\ \mathrm{dist}(\mathbf{0}, \partial F(x^{k+1})) \leq c\sum_{j=k-\tau}^{k} \|\omega^{j+1} - \omega^j\| + d\eta_k \\ \text{for any stationary point } x^* \text{ there exist subsequence } (x^{k_j})_{j\geq 0} \to x^* \text{ satisfying } F(x^{k_j}) \to F(x^*) \end{cases} \tag{1.6}$$

where $a, b, c, d > 0$ are constants, and $(\eta_k)_{k\geq 0}$ is a nonnegative sequence, and $\tau \in \mathbb{Z}_0^+$ and $(\omega^k)_{k\geq 0}$ is a sequence satisfying

$$\|x^k - x^{k+1}\| \leq \mathfrak{e}\|\omega^k - \omega^{k+1}\| \tag{1.7}$$

for some $\mathfrak{e} > 0$. The continuity condition is kept here. Obviously, if $\eta_k \equiv 0$, $\omega^k = x^k$ and $\tau = 0$, the condition will reduce to (1.2). Thus, our work can also be regarded as a generation of paper [4]. Our approach is first proving convergence for a general inexact algorithm whose sequence $(x^k)_{k\geq 0}$ satisfying the condition (1.6) under energy assumption on $(\eta_k)_{k\geq 0}$. Then, we then prove several classical inexact algorithms satisfying condition (1.6).

The core of the proof lies on using an auxiliary function whose successive difference gives a bound to the successive difference of the sequence $\|\omega^{k+1} - \omega^k\|^2$. If F is semi-algebraic, the new function is then Kurdyka-Łojasiewicz. And then, we build sufficient descent involving with the new function and $\|\omega^{k+1} - \omega^k\|^2$. Under assuming the finite the energy of η_k, we denote t_k in (3.3). In the $(k+1)$-th iteration, the distance between subdifferential of the new function and the origin is bounded by the composition of $\|\omega^{k+1} - \omega^k\|$, t_k and t_{k+1}. And then, we prove the finite length of $(x^k)_{k\geq 0}$ provided $(t_k)_{k\geq 0}$ is also finite. In proving the finite length, the key part is using the KL property of the new Lyapunov function. The proof techniques are motivated by the methodology proposed in [4].

The rest of the paper is organized as follows. In section 2, we list needed preliminaries. Section 3 contains the main results. In section 4, we provide the applications. Section 5 concludes the paper.

2 Preliminaries

This section presents the mathematical tools which will be used in our proofs, and contains two parts: in the first one, we introduce the basic definitions and properties for subdifferentials; in the second one, the KL property is introduced.

2.1 Subdifferential

More details about the definition of subdifferential can be found in the textbooks [28, 29]. Given an lower semicontinuous function $J : \mathbb{R}^N \to (-\infty, +\infty]$, its domain is defined by

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

The notation of subdifferential plays a central role in variational analysis.

Definition 1 (subdifferential)**.** *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 $\hat{\partial}J(x)$, is the set of all vectors $u \in \mathbb{R}^N$ which satisfy*
$$\lim_{y \neq x} \inf_{y \to x} \frac{J(y) - J(x) - \langle u, y - x\rangle}{\|y - x\|} \geq 0.$$
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 $\partial J(x)$, is defined through the following closure process*
$$\partial J(x) := \{u \in \mathbb{R}^N : \exists x^k \to x, J(x^k) \to J(x) \text{ and } u^k \in \hat{\partial}J(x^k) \to u \text{ as } k \to \infty\}.$$

It is easy to verify that the Fréchet subdifferential is convex and closed while the subdifferential is closed. When J is convex, the definition agrees with the subgradient in convex analysis as

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

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

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

And the domain of the subdifferential of ∂J is given as

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

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

$$\mathbf{0} \in \partial J(x). \tag{2.1}$$

When J is convex, (2.1) is also sufficient. A point that satisfies (2.1) is called (limiting) critical point. The set of critical points of $J(x)$ is denoted by crit(J).

2.2 Kurdyka-Łojasiewicz function

With the definition of subdifferential, we now are prepared to introduce the Kurdyka-Łojasiewicz property and function.

Definition 2. *[23, 17, 6] (a) The function $J : \mathbb{R}^N \to (-\infty, +\infty]$ is said to have the Kurdyka-Łojasiewicz property at $\overline{x} \in dom(\partial J)$ 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* $U \bigcap \{x | J(\overline{x}) < J(x) < J(\overline{x}) + \eta\}$, *the Kurdyka-Łojasiewicz inequality holds*

$$\varphi^{'}(J(x) - J(\overline{x})) dist(\boldsymbol{0}, \partial J(x)) \geq 1. \tag{2.2}$$

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

It is hard to judge whether a function is Kurdyka-Łojasiewicz or not. Fortunately, the concept of semi-algebraicity can help to find and check a very rich class of Kurdyka-Łojasiewicz functions.

Definition 3 (Semi-algebraic sets and functions [6, 7])**.** *(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*

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

(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 [4]. 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 ([4])**.** *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* $\text{dom}(J)$. *In particular, if* J *is semi-algebraic and* $\text{dom}(J) = \text{dom}(\partial J)$, *then it is a KL function.*

Now we present a lemma for the uniformized KL property. With this lemma, we can make the proofs much more concise.

Lemma 2 ([8])**.** *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* $\delta, \varepsilon > 0$ *such that for any* $\overline{x} \in \Omega$ *and any* x *satisfying that* $dist(x, \Omega) < \varepsilon$ *and* $J(\overline{x}) < J(x) < J(\overline{x}) + \delta$, *it holds that*

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

3 Convergence analysis

The sequence $(\eta_k)_{k\geq 0}$ is assumed to satisfy

$$\sum_l \eta_l^2 < +\infty. \tag{3.1}$$

We denote a new function, which plays an important role in the analysis, as

$$\xi(z) := F(x) + \frac{t^2}{2}, z := (x,t) \in \mathbb{R}^{N+1}. \tag{3.2}$$

We also need to define the new sequences as

$$t_k := \sqrt{2a \sum_{l=k}^{+\infty} \eta_l^2}, z^k := (x^k, t_k). \tag{3.3}$$

The aim in this part is proving that $\{z^k\}$ generated by the algorithm converges to a critical point of ξ. The proof contains two main steps:

1. Find a positive constant ρ_1 such that

$$\rho_1 \|\omega^{k+1} - \omega^k\|_2^2 \leq \xi(z^k) - \xi(z^{k+1}), \forall k = 0, 1, \cdots.$$

2. Find another positive constants ρ_2, ρ_3, ρ_4 such that

$$\mathrm{dist}(\mathbf{0}, \partial\xi(z^{k+1})) \leq \rho_2 \sum_{j=k-\tau}^{k} \|\omega^{j+1} - \omega^j\| + \rho_3 t_k + \rho_4 t_{k+1},\ \forall k = 0, 1, \cdots.$$

Lemma 3. *Assume that $\{x^k\}_{k=0,1,2,\ldots}$ is generated by the general inexact algorithm satisfying condition (1.6), and condition (3.1) holds. Then, we have the following results.*

(1) It holds that

$$\xi(z^k) - \xi(z^{k+1}) \geq a\|\omega^k - \omega^{k+1}\|^2. \tag{3.4}$$

And then, $(z^k)_{k\geq 0}$ is bounded if F is coercive.

(2) $\sum_k \|x^{k+1} - x^k\|^2 < +\infty$, which implies that

$$\lim_k \|x^{k+1} - x^k\| = 0. \tag{3.5}$$

Proof. (1) From the direct algebra computations, we can easily obtain

$$\begin{aligned} \xi(z^k) - \xi(z^{k+1}) &= F(x^k) - F(x^{k+1}) + \frac{t_k^2 - t_{k+1}^2}{2} \\ &= F(x^k) - F(x^{k+1}) + b\eta_k^2 \\ &\geq a\|\omega^k - \omega^{k+1}\|^2. \end{aligned} \tag{3.6}$$

(2) From (3.4), $\{\xi(z^k)\}_{k=0,1,2,\ldots}$ is descending. Note that $\xi > -\infty$, $\{\xi(z^k)\}_{k=0,1,2,\ldots}$ is convergent. Hence, we can easily have that

$$\sum_{n=0}^{k} \|\omega^{n+1} - \omega^n\|^2 \leq \frac{\xi(z^0) - \xi(z^{k+1})}{a} < +\infty.$$

With (1.7), we then prove the result. □

Lemma 4. *Let the sequence generated by the general inexact algorithm,*

$$\mathrm{dist}(\boldsymbol{0}, \partial\xi(z^{k+1})) \leq c \sum_{j=k-\tau}^{k} \|\omega^{j+1} - \omega^{j}\| + dt_k + t_{k+1}. \tag{3.7}$$

Proof. Direct calculation yields

$$\partial\xi(z^{k+1}) = \begin{pmatrix} \partial F(x^{k+1}) \\ t_{k+1} \end{pmatrix}. \tag{3.8}$$

Thus, we have

$$\begin{aligned} \mathrm{dist}(\mathbf{0}, \partial\xi(z^{k+1})) &\leq \mathrm{dist}(\mathbf{0}, \partial F(x^{k+1})) + t_{k+1} \\ &\leq c \sum_{j=k-\tau}^{k} \|\omega^{j+1} - \omega^{j}\| + d\eta_k + t_{k+1} \\ &\leq c \sum_{j=k-\tau}^{k} \|\omega^{j+1} - \omega^{j}\| + dt_k + t_{k+1}. \end{aligned} \tag{3.9}$$

□

In the following, we establish some results about the limit points of the sequence generated by the general algorithm. We need a definition about the limit point which is introduced in [4].

Definition 4. *For a sequence* $\{d^k\}_{k=0,1,2,...}$*, define that*

$$\mathcal{M}(d^0) := \{d \in \mathbb{R}^{N+1} : \exists \text{ an increasing sequence of integers } \{k_j\}_{j\in\mathbb{N}} \text{ such that } d^{k_j} \to d \text{ as } j \to \infty\},$$

where $d^0 \in \mathbb{R}^{N+1}$ *is the starting point.*

Lemma 5. *Suppose that* $\{z^k = (x^k, t_k)\}_{k=0,1,2,...}$ *is generated by general algorithm and* F *is coercive. And it is assumed that* $t_k \to 0$*. Then, we have the following results.*

(1) For any $z^* = (x^*, t^*) \in \mathcal{M}(z^0)$*, we have* $t^* = 0$ *and* $\xi(z^*) = F(x^*)$*.*
(2) $\mathcal{M}(z^0)$ *is nonempty and* $\mathcal{M}(z^0) \subseteq crit(\xi)$*.*
(2') $\mathcal{M}(x^0)$ *is nonempty and* $\mathcal{M}(x^0) \subseteq crit(F)$
(3) $\lim_k dist(z^k, \mathcal{M}(z^0)) = 0$*.*
(3') $\lim_k dist(x^k, \mathcal{M}(x^0)) = 0$*.*
(4) The function ξ *is finite and constant on* $\mathcal{M}(z^0)$*.*
(4') The function F *is finite and constant on* $\mathcal{M}(x^0)$*.*

Proof. (1) Noting $(t_k)_{k\geq 0} \to 0$, $t^* = 0$ and

$$\xi(z^*) = \xi(x^*, 0) = F(x^*).$$

(2) It is easy to see the coercivity of ξ. With Lemma 3 and the coercivity of ξ, $\{z^k\}_{k=0,1,2,...}$ is bounded. Thus, $\mathcal{M}(z^0)$ is nonempty. Assume that $z^* \in \mathcal{M}(z^0)$, from the definition, there exists a subsequence $z^{k_i} \to z^*$. From Lemmas 3 and 4, we have $\mathrm{dist}(\mathbf{0}, \partial\xi(z^{k_i})) \to \mathbf{0}$. The closedness of $\partial\xi$ indicates that $\mathbf{0} \in \partial\xi(z^*)$, i.e. $z^* \in crit(\xi)$.

(2') With the facts $z = (x, t)$ and $\xi(z) = F(x) + \frac{t^2}{2}$, we can easily derive the results.

(3)(3') This item follows as a consequence of the definition of the limit point.

(4) Let l be the limitation of $(\xi(x^k))_{k\geq 0}$. For one stationary point x^*, from the continuity condition, there exists $x^{k_j} \to x^*$ satisfying $F(x^k_j) \to F(x^*)$. We denote that $z^{k_j} = (x^{k_j}, t_{k_j})$. Thus, the subsequence $(z^{k_j}) \to z^* \in \mathrm{crit}(\xi)$ and $(\xi(z^{k_j}))_{j\geq 0} \to l$. And we have

$$\xi(\overline{z}) = \lim_j \xi(z^{k_j}) = l.$$

(4') The proof is similar to (4).

□

Theorem 1 (Convergence result)**.** *Suppose that F is a semi-algebraic function and coercive. Let the sequence $\{x^k\}_{k=0,1,2,3,...}$ be generated by general scheme and the conditions (1.6) hold. If the sequence $(\eta_k)_{k\geq 0}$ satisfies*

$$\sum_k \sqrt{\sum_{l=k}^{+\infty} \eta_l^2} < +\infty. \tag{3.10}$$

Then, the sequence $\{x^k\}_{k=0,1,2,3,...}$ has finite length, i.e.

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|_2 < +\infty. \tag{3.11}$$

And $\{x^k\}_{k=0,1,2,3,...}$ converges to a critical point x^* of F.

Proof. Obviously, ξ is semi-algebraic, and then KL. Let x^* be a cluster point of $\{x^k\}_{k=0,1,2,...}$, then, $z^* = (x^*, 0)$ is also a cluster point of $\{z^k\}_{k=0,1,2,...}$. From Lemmas 2 and 5, there exist $\delta, \varepsilon > 0$ such that for any $\overline{x} \in \mathcal{M}(z^0)$ and any x satisfying that $\mathrm{dist}(z, \mathcal{M}(z^0)) < \varepsilon$ and $\xi(z^*) < \xi(z) < \xi(z^*) + \delta$. From Lemma 5, as k is large enough,

$$z^k \in \{z \mid \mathrm{dist}(z, \mathcal{M}(z^0)) < \varepsilon\} \bigcap \{z \mid \xi(z^*) < \xi(z) < \xi(z^*) + \delta\}.$$

Thus, there exist φ such that

$$\varphi'(\xi(z^{k+1}) - \xi(z^*))\mathrm{dist}(\mathbf{0}, \partial\xi(z^{k+1})) \geq 1. \tag{3.12}$$

Therefore, we have

$$\begin{aligned}
\varphi(\xi(z^{k+1}) - \xi(z^*)) - \varphi(\xi(z^{k+2}) - \xi(z^*)) &\geq \varphi'(\xi(z^{k+1}) - \xi(z^*))(\xi(z^{k+1}) - \xi(z^{k+2})) \\
&\geq a\varphi'(f(x^{k+1}) - \xi(z^*))\|\omega^{k+2} - \omega^{k+1}\|_2^2 \\
&\geq a\frac{1}{\mathrm{dist}(\mathbf{0}, \partial\xi(z^{k+1}))}\|\omega^{k+2} - \omega^{k+1}\|_2^2 \\
&\geq \frac{a}{c\|x^{k+1} - x^k\| + dt_k + t_{k+1}}\|\omega^{k+2} - \omega^{k+1}\|_2^2.
\end{aligned}$$

That is also

$$\begin{aligned}
2\|\omega^{k+2} - \omega^{k+1}\|_2 &\leq \frac{2}{a}\left\{[\varphi(\xi(z^{k+1}) - \xi(z^*)) - \varphi(\xi(z^{k+2}) - \xi(z^*))] \cdot [c\sum_{j=k-\tau}^{k}\|\omega^{j+1} - \omega^j\| + d\eta_k + \eta_{k+1}]\right\}^{\frac{1}{2}} \\
&\leq \frac{c\tau}{a^2}[\varphi(\xi(z^{k+1}) - \xi(z^*)) - \varphi(\xi(z^{k+2}) - \xi(z^*))] \\
&+ \frac{\sum_{j=k-\tau}^{k}\|\omega^{j+1} - \omega^j\|}{\tau} + \frac{ad}{c\tau}t_k + \frac{a}{c\tau}t_{k+1}.
\end{aligned} \tag{3.13}$$

After simplifications, we have

$$\begin{aligned}
2\tau\|\omega^{k+2} - \omega^{k+1}\|_2 &\leq \frac{c\tau^2}{a^2}[\varphi(\xi(z^{k+1}) - \xi(z^*)) - \varphi(\xi(z^{k+2}) - \xi(z^*))] \\
&+ \sum_{j=k-\tau}^{k}\|\omega^{j+1} - \omega^j\| + \frac{ad}{c}t_k + \frac{a}{c}t_{k+1}.
\end{aligned} \tag{3.14}$$

Summing both sides from k to K, and with simplifications,

$$\begin{aligned}
\sum_{l=k+1}^{K+1}\|\omega^{l+1} - \omega^l\| + \sum_{j=K+1-\tau}^{K+1}\|x^{j+1} - x^j\| &\leq \frac{c}{a^2}[\varphi(\xi(z^{k+1}) - \xi(z^*)) - \varphi(\xi(z^{K+2}) - \xi(z^*))] \\
&+ \sum_{j=k-\tau}^{k}\|\omega^{j+1} - \omega^j\| + \frac{ad}{c}\sum_{l=k}^{K}t_l + \frac{a}{c}\sum_{l=k+1}^{K+1}t_l < +\infty.
\end{aligned} \tag{3.15}$$

Letting $K \to +\infty$ and using $\sum_{j=K+1-\tau}^{K+1} \|\omega^{j+1} - \omega^j\| \to 0$, we then derive the result by using (1.7). $\square$

Remark 1. *If the sequence $(\eta_k)_{k\geq 0}$ performs as $\eta_k \sim O(\frac{1}{k^\alpha})$ and $\alpha > \frac{3}{2}$, the condition (3.10) holds.*

4 Applications to several nonconvex algorithms

In this part, several classical nonconvex inexact algorithms are considered. We apply our theoretical findings to these algorithms and derive corresponding convergence results for the algorithms. As presented before, we just need to check whether the algorithm satisfies the three conditions in (1.6). For a closed function (may be nonconvex) J, we denote

$$\mathbf{prox}_J(x) \in \arg\min_y \{J(y) + \frac{\|y-x\|^2}{2}\}. \tag{4.1}$$

Different with convex cases, the $\mathbf{prox}_J$ is a point-to-set operator and may have more than one solution. We present a useful lemma which plays a very important role in the analysis.

Lemma 6. *For any x and y, if $z \in \mathbf{prox}_J(x)$,*

$$J(z) + \frac{\|z-x\|^2}{2} \leq J(y) + \frac{\|y-x\|^2}{2}. \tag{4.2}$$

Of course, we also have

$$x - z \in \partial J(z). \tag{4.3}$$

In subsections 4.1-4.4, the point ω^k is x^k itself, i.e., $\omega^k \equiv x^k$.

4.1 Inexact nonconvex gradient and proximal algorithm

The nonconvex proximal gradient algorithm is developed for the nonconvex composite optimization

$$\min_x \{F(x) = f(x) + g(x)\}, \tag{4.4}$$

where f is differentiable and ∇f is Lipschitz with L, and g is closed. And both f and g may be nonconvex. The nonconvex inexact proximal gradient algorithm can be described as

$$x^{k+1} = \mathbf{prox}_{hg}(x^k - h\nabla f(x^k) + e^k), \tag{4.5}$$

where h is the stepsize, **prox** is the proximal operator and e^k is the noise. In the convex case, this algorithm is discussed in [38, 30].

Lemma 7. *Let $0 < h < \frac{1}{L}$ and the sequence $(x^k)_{k\geq 0}$ be generated by algorithm (4.5), we have*

$$F(x^k) - F(x^{k+1}) \geq \frac{1}{4}(\frac{1}{h} - L)\|x^{k+1} - x^k\|^2 - \frac{1}{h(1-hL)}\|e^k\|^2. \tag{4.6}$$

Proof. The L-Lipschitz of ∇f gives

$$f(x^{k+1}) - f(x^k) \leq \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2. \tag{4.7}$$

On the other hand, with Lemma 6, we have

$$hg(x^{k+1}) + \frac{\|x^k - h\nabla f(x^k) + e^k - x^{k+1}\|^2}{2} \leq hg(x^k) + \frac{\|-h\nabla f(x^k) + e^k\|^2}{2}. \tag{4.8}$$

This is also

$$g(x^{k+1})-g(x^k)\le -\langle\nabla f(x^k),x^{k+1}-x^k\rangle-\frac{\|x^k-x^{k+1}\|^2}{2h}+\frac{\langle e^k,x^{k+1}-x^k\rangle}{h}. \tag{4.9}$$

Summing (4.7) and (4.9),

$$F(x^{k+1})-F(x^k)\le\frac{1}{2}(L-\frac{1}{h})\|x^{k+1}-x^k\|^2+\frac{\langle e^k,x^{k+1}-x^k\rangle}{h}. \tag{4.10}$$

With the Cauchy-Schwarz inequality, we have

$$\frac{\langle e^k,x^{k+1}-x^k\rangle}{h}\le\frac{1}{4}(\frac{1}{h}-L)\|x^{k+1}-x^k\|^2+\frac{1}{h(1-hL)}\|e^k\|^2 \tag{4.11}$$

Combining (4.11) and (4.10), we then prove the result. □

Lemma 8. *Let the sequence be generated by algorithm (4.5), we have*

$$\mathrm{dist}(\boldsymbol{0},\partial F(x^{k+1}))\le(\frac{1}{h}+L)\|x^k-x^{k+1}\|+\|e^k\|. \tag{4.12}$$

Proof. We have

$$\frac{x^k-x^{k+1}}{h}-\nabla f(x^k)+\frac{e^k}{h}\in\partial g(x^{k+1}). \tag{4.13}$$

Therefore,

$$\frac{x^k-x^{k+1}}{h}+\nabla f(x^{k+1})-\nabla f(x^k)+\frac{e^k}{h}\in\nabla f(x^{k+1})+\partial g(x^{k+1})=\partial F(x^{k+1}). \tag{4.14}$$

Thus, we have

$$\begin{aligned}\mathrm{dist}(\mathbf{0},\partial F(x^{k+1})) &\le \|\frac{x^k-x^{k+1}}{h}+\nabla f(x^{k+1})-\nabla f(x^k)+\frac{e^k}{h}\|\\ &\le \frac{1}{h}\|x^k-x^{k+1}\|+L\|x^k-x^{k+1}\|+\frac{\|e^k\|}{h}.\end{aligned} \tag{4.15}$$

□

Lemma 9. *Let $0<h<\frac{1}{L}$ and the sequence $(x^k)_{k\ge0}$ be generated by algorithm (4.5), and F be coercive. We also assume that $e^k\to\boldsymbol{0}$. Then, for any $x^*\in\mathrm{crit}(F)$, there exists a subsequence $(x^{k_j})_{j\ge0}$ converges to x^* satisfying $F(x^{k_j})\to F(x^*)$.*

Proof. With Lemma 5, $(x^k)_{k\ge0}$ is bounded. For any $x^*\in\mathrm{crit}(F)$, there exists a subsequence $(x^{k_j})_{j\ge0}$ converges to x^*. With Lemmas 5 and 7, we also have

$$x^{k_j-1}\to x^*. \tag{4.16}$$

And in each iteration, with Lemma 6, we have

$$hg(x^{k_j})+\frac{\|x^{k_j-1}-h\nabla f(x^{k_j-1})+e^{k_j}-x^{k_j}\|^2}{2}\le hg(x^*)+\frac{\|x^{k_j-1}-h\nabla f(x^{k_j-1})+e^{k_j-1}-x^*\|^2}{2}. \tag{4.17}$$

Taking $j\to+\infty$, we have

$$\limsup_{j\to+\infty}\ g(x^{k_j})\le g(x^*). \tag{4.18}$$

And recalling the lower semi-continuity of g,

$$g(x^*)\le\liminf_{j\to+\infty}\ g(x^{k_j}). \tag{4.19}$$

That means $\lim g(x^{k_j})=g(x^*)$; and combining the continuity of f, we then prove the result. □

And then, we then prove the following result.

Theorem 2. *Suppose that f and g are all semi-algebraic,* $\mathrm{dom}(f) = \mathrm{dom}(\partial f)$ *and* $\mathrm{dom}(g) = \mathrm{dom}(\partial g)$*, F is coercive, and $0 < h < \frac{1}{L}$. Let the sequence $(x^k)_{k\geq 0}$ be generated by scheme (4.5). If the sequence $(\|e^k\|)_{k\geq 0}$ satisfies*

$$\sum_k \sqrt{\sum_{l=k}^{+\infty} \|e^l\|^2} < +\infty. \tag{4.20}$$

Then, the sequence $(x^k)_{k\geq 0}$ has finite length, i.e.

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\| < +\infty. \tag{4.21}$$

And $\{x^k\}_{k=0,1,2,3,\dots}$ converges to a critical point x^* of F.

Proof. From (4.20), we have $\|e^k\| \to 0$. From Lemma 1, F is a semi-algebraic function. With lemmas proved before in this subsection and Theorem 1, we then obtain the result. □

Remark 2. *That means for the nonconvex inexact proximal gradient algorithm, the iteration converges provided $\|e^k\| \sim O(\frac{1}{k^\alpha})$ and $\alpha > \frac{3}{2}$. While in the convex setting, $\alpha > 1$ is sufficient for the convergence [22].*

4.2 Inexact proximal linearized alternating minimization algorithm

In this part, we use the convention

$$x = (y, z), x^k = (y^k, z^k), e^k = (\alpha^k, \beta^k)$$

The following problem is considered

$$\min_{y,z}\{\Phi(y, z) := f(y) + H(y, z) + g(z)\}, \tag{4.22}$$

where the function H is assumed to be differentiable and satisfy

$$\|\nabla_y H(y^1, z) - \nabla_y H(y^2, z)\| \leq M(z)\|y^1 - y^2\|, 0 < \inf M(z) \leq \sup M(z) < +\infty, \tag{4.23a}$$
$$\|\nabla_z H(y, z^1) - \nabla_z H(y, z^2)\| \leq N(y)\|z^1 - z^2\|, 0 < \inf N(y) \leq \sup N(y) < +\infty, \tag{4.23b}$$
$$\|\nabla_y H(x^1) - \nabla_y H(x^2)\| \leq L(x^1, x^2)\|x^1 - x^2\|, 0 < \inf L(x^1, x^2) \leq \sup L(x^1, x^2) < +\infty. \tag{4.23c}$$

An intuitive algorithm for solving problem (4.22) is the alternating minimization scheme, i.e., fixing one of y and z in each iteration and then minimizing the other one [27]; and the convex rate is proved in [5] in the convex case. In the nonconvex case, the alternating minimization scheme can barely derive the descent property, thus the authors propose the proximal alternating minimization [3]. However, both alternating minimization and proximal alternating minimization have an obvious drawback: both algorithms need to solve a minimization problem in each iteration, the stopping criterion is hard to determine, and error accumulates. Therefore, several variants are developed [8, 33, 31], and the Proximal Linearized Alternating Minimization (PLAM) algorithm [8] is one of them. The inexact PLAM can be described as

$$y^{k+1} = \mathbf{prox}_{\gamma_k f}(y^k - \gamma_k \nabla_y H(y^k, z^k) + \alpha^k), \tag{4.24a}$$
$$z^{k+1} = \mathbf{prox}_{\lambda_k g}(z^k - \lambda_k \nabla_z H(y^{k+1}, z^k) + \beta^k). \tag{4.24b}$$

Lemma 10. *Let the sequence $(x^k)_{k\geq 0}$ be generated by algorithm (4.24). If*

$$\inf_k\{1-\gamma_k M(z^k), 1-\lambda_k N(y^{k+1})\} > 0, \tag{4.25}$$

we have

$$\Phi(x^k) - \Phi(x^{k+1}) \geq \nu\|x^{k+1}-x^k\|^2 - \sigma\|e^k\|^2, \tag{4.26}$$

where $\nu = \inf_k\{\frac{1}{4}(M(z^k)-\frac{1}{\gamma_k}), \frac{1}{4}(N(y^{k+1})-\frac{1}{\lambda_k})\}$ *and* $\sigma = \sup_k\{\frac{1}{\gamma_k(1-\gamma_k M(z^k))}, \frac{1}{\lambda_k(1-\lambda_k N(y^{k+1}))}\}$

Proof. The $L(z^k)$-Lipschitz of $\nabla_y H(y, z^k)$ gives

$$H(y^{k+1}, z^k) - H(y^k, z^k) \leq \langle \nabla_y H(y^k, z^k), y^{k+1}-y^k\rangle + \frac{M(z^k)}{2}\|y^{k+1}-y^k\|^2. \tag{4.27}$$

From Lemma 6, we have

$$\gamma_k f(y^{k+1}) + \frac{\|y^k - \gamma_k\nabla_y H(y^k, z^k) + \alpha^k - y^{k+1}\|^2}{2} \leq \gamma_k f(y^k) + \frac{\|-\gamma_k\nabla_y H(y^k, z^k) + \alpha^k\|^2}{2}. \tag{4.28}$$

This is also

$$f(y^{k+1}) - f(y^k) \leq -\langle \nabla_y H(y^k, z^k), y^{k+1}-y^k\rangle - \frac{\|y^k-y^{k+1}\|^2}{2\gamma_k} + \frac{\langle \alpha^k, y^{k+1}-y^k\rangle}{\gamma_k}. \tag{4.29}$$

Summing (4.27) and (4.29), with the Cauchy-Schwarz inequality

$$\frac{\langle \alpha^k, y^{k+1}-y^k\rangle}{\gamma_k} \leq \frac{1}{4}(\frac{1}{\gamma_k} - M(z^k))\|y^{k+1}-y^k\|^2 + \frac{1}{\gamma_k(1-\gamma_k M(z^k))}\|\alpha^k\|^2, \tag{4.30}$$

we then have

$$[f(y^{k+1}) + H(y^{k+1}, z^k)] - [f(y^k) + H(y^k, z^k)] \leq \frac{1}{4}(M(z^k) - \frac{1}{\gamma_k})\|y^{k+1}-y^k\|^2 + \frac{\|\alpha^k\|^2}{\gamma_k(1-\gamma_k M(z^k))}. \tag{4.31}$$

Similarly, we can prove

$$[g(z^{k+1}) + H(y^{k+1}, z^{k+1})] - [g(z^k) + H(y^{k+1}, z^k)] \leq \frac{1}{4}(N(y^{k+1}) - \frac{1}{\lambda_k})\|z^{k+1}-z^k\|^2 + \frac{\|\beta^k\|^2}{\lambda_k(1-\lambda_k N(y^{k+1}))}. \tag{4.32}$$

Combining (4.11) and (4.10), we then prove the result. □

Lemma 11. *Let the sequence be generated by algorithm (4.24) and condition (1.6) be satisfied, we have*

$$\mathrm{dist}(\boldsymbol{0}, \partial\Phi(x^{k+1})) \leq S\|x^k - x^{k+1}\| + D\|e^k\|, \tag{4.33}$$

where $S = \sup\{\frac{1}{\lambda_k} + \frac{1}{\gamma_k} + L(x^k, x^{k+1}) + L(y^{k+1})\}$ *and* $D = \sup_k\{\sqrt{\frac{1}{\gamma_k^2} + \frac{1}{\lambda_k^2}}\}$.

Proof. In updating y^{k+1}, we have

$$\frac{y^k - y^{k+1}}{\gamma_k} - \nabla_y H(y^k, z^k) + \frac{\alpha^k}{\gamma_k} \in \partial f(y^{k+1}). \tag{4.34}$$

Therefore,

$$\frac{y^k - y^{k+1}}{\gamma_k} + \nabla_y H(y^{k+1}, z^{k+1}) - \nabla_y H(y^k, z^k) + \frac{\alpha^k}{\gamma_k} \in \nabla_y H(y^{k+1}, z^{k+1}) + g(x^{k+1}) = \partial_y \Phi(x^{k+1}). \tag{4.35}$$

Thus, we have

$$\begin{aligned}\mathrm{dist}(\mathbf{0}, \partial_y \Phi(x^{k+1})) &\leq \|\frac{y^k - y^{k+1}}{\gamma_k} + \nabla_y H(y^{k+1}, z^{k+1}) - \nabla_y H(y^k, z^k) + \frac{\alpha^k}{\gamma_k}\| \\ &\leq \frac{\|y^k - y^{k+1}\|}{\gamma_k} + L(x^{k+1}, x^k)\|x^{k+1} - x^k\| + \frac{\|\alpha^k\|}{\gamma_k}. \end{aligned} \tag{4.36}$$

In updating z^{k+1}, we have

$$\mathrm{dist}(\mathbf{0}, \partial_z \Phi(x^{k+1})) \leq \frac{\|z^k - z^{k+1}\|}{\lambda_k} + L(y^{k+1})\|z^{k+1} - z^k\| + \frac{\|\beta^k\|}{\lambda_k}. \tag{4.37}$$

If condition (1.6) is satisfied, we can easy see that

$$\sup\{\frac{1}{\lambda_k}, \frac{1}{\gamma_k}\} < +\infty. \tag{4.38}$$

Combining (4.36) and (4.37), we then prove the result. □

Lemma 12. *Let the sequence $(x^k)_{k\geq 0}$ be generated by algorithm (4.5), and Φ be coercive, and condition (4.25) hold, $\sum_k \sqrt{\sum_{l=k}^{+\infty}(\|\alpha^l\|^2 + \|\beta^l\|^2)} < +\infty$. Then, for any $x^* \in \mathrm{crit}(\Phi)$, there exists a subsequence $(x^{k_j})_{j\geq 0}$ converges to x^* satisfying $\Phi(x^{k_j}) \to \Phi(x^*)$.*

Proof. Obviously, we have $e^k \to \mathbf{0}$. With Lemma 5, $(x^k)_{k\geq 0}$ is bounded. For any $x^*\mathrm{crit}(F)$, there exists a subsequence $(x^{k_j})_{j\geq 0}$ converges to x^*. With Lemmas 5 and 10, we also have

$$x^{k_j-1} = (y^{k_j-1}, z^{k_j-1}) \to x^* = (y^*, z^*). \tag{4.39}$$

And in each iteration of updating y^{k_j}, with Lemma 6, we have

$$\begin{aligned}\gamma_k f(y^{k_j}) &+ \frac{\|y^{k_j-1} - \gamma_{k_j-1}\nabla_y H(y^{k_j-1}, z^{k_j-1}) + \alpha^{k_j-1} - y^{k_j}\|^2}{2} \\ &\leq \gamma_{k_j-1} f(y^{k_j-1}) + \frac{\| - \gamma_{k_j-1}\nabla_y H(y^{k_j-1}, z^{k_j-1}) + \alpha^{k_j-1}\|^2}{2}. \end{aligned} \tag{4.40}$$

Taking $j \to +\infty$, we have

$$\limsup_{j\to+\infty} \ f(y^{k_j}) \leq f(y^*). \tag{4.41}$$

And recalling the lower semi-continuity of f,

$$f(y^*) \leq \liminf_{j\to+\infty} \ f(y^{k_j}). \tag{4.42}$$

That means $\lim f(y^{k_j}) = f(x^*)$; and similarly, $\lim g(z^{k_j}) = g(z^*)$; combining the continuity of H, we then prove the result. □

And then, we then prove the following result.

Theorem 3. *Suppose that Φ is and coercive, and condition (4.25) holds. Functions f, g and H are all semi-algebraic, and their domains satisfy $\mathrm{dom}(\partial f) = \mathrm{dom}(f)$, $\mathrm{dom}(\partial g) = \mathrm{dom}(g)$, $\mathrm{dom}(\nabla H) = \mathrm{dom}(H)$. Let the sequence $(x^k)_{k\geq 0}$ be generated by scheme (4.24). If the sequence $(\|\alpha^k\|, \|\beta^k\|)_{k\geq 0}$ satisfies*

$$\sum_k \sqrt{\sum_{l=k}^{+\infty}(\|\alpha^l\|^2 + \|\beta^l\|^2)} < +\infty. \tag{4.43}$$

Then, the sequence $(x^k)_{k\geq 0}$ has finite length, i.e.

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\| < +\infty. \tag{4.44}$$

And $\{x^k\}_{k=0,1,2,3,\ldots}$ converges to a critical point x^* of Φ.

4.3 Inexact proximal reweighted algorithm

This part considers an iteratively reweighted algorithm for a broad class of nonconvex and nonsmooth problems with the following form

$$\min_x \{\Psi(x) = f(x) + \sum_{i=1}^{N} h(g(x_i))\}, \tag{4.45}$$

where $x \in \mathbb{R}^N$, and functions f has a Lipschitz gradient with constant L_f, and $g : \mathbb{R} \to \mathbb{R}$ is a lower-semicontinuous convex function, and $h : \mathrm{Im}(g) \to \mathbb{R}$ is a differentiable concave function with a Lipschitz continuous gradient with constant L_h, i.e.,

$$\mid h'(s) - h'(t) \mid \leq L_h \mid s - t \mid, \tag{4.46}$$

and $h'(t) > 0$ for any $t \in \mathrm{Im}(g)$. This model generalizes various problems in the machine learning and signal processing satisfy. The reweighted style algorithms [10, 9, 13, 18, 34, 25] (or also called multi-stage algorithm [40]) are popular in solving this problem. To make each subproblem easy to be solved. The Proximal Iteratively REweighted (PIRE) algorithm is proposed in [24]. The convergence of PIRE under KL property is proved by [35]. We consider the inexact version of PIRE as

$$x_i^{k+1} = \mathbf{prox}_{\mu w_i^k g}(x_i^k - \mu \nabla_i f(x^k) + e_i^k), i \in [1, 2, \ldots, N] \tag{4.47}$$

where $w_i^k := h'(g(x_i^k))$ and $\mu > 0$ is the stepsize, e^k is the noise vector. If $e^k \equiv 0$, the algorithm then reduces to PIRE.

Lemma 13. *Let $(x^k)_{k\geq 0}$ is generated by scheme (4.47) and $\frac{1}{\mu} > \frac{L_f}{2}$. Then, we will have*

$$\Psi(x^k) - \Psi(x^{k+1}) \geq (\frac{1}{\mu} - \frac{L_f}{2})\|x^k - x^{k+1}\|^2 - \frac{\|e^k\|^2}{\mu(2 - \mu L_f)}. \tag{4.48}$$

Proof. We can easily obtain that

$$\begin{aligned}
\Psi(x^k) - \Psi(x^{k+1}) &= f(x^k) - f(x^{k+1}) + \sum_{i=1}^{N} h(g(x_i^k)) - h(g(x_i^{k+1})) \\
&\geq \langle \nabla f(x^k), x^k - x^{k+1} \rangle - \frac{L_f}{2}\|x^k - x^{k+1}\|_2^2 + \sum_{i=1}^{N} h(g(x_i^k)) - h(g(x_i^{k+1})) \\
&\geq \sum_{i=1}^{N} \langle \nabla_i f(x^k), x_i^k - x_i^{k+1} \rangle - \frac{L_f}{2}\|x^k - x^{k+1}\|^2 \\
&+ \sum_{i=1}^{N} h'(g(x_i^k))(g(x_i^k) - g(x_i^{k+1})).
\end{aligned} \tag{4.49}$$

Note that x_i^{k+1} is obtained by (4.47); the K.K.T condition gives

$$\nabla_i f(x^k) + w_i^k v_i^{k+1} + \frac{(x_i^{k+1} - x_i^k)}{\mu} - \frac{e_i^k}{\mu} = \mathbf{0}, \tag{4.50}$$

where $v_i^{k+1} \in \partial g(x_i^{k+1})$. Note that g is convex, we have that

$$\sum_{i=1}^{N} h'(g(x_i^k))(g(x_i^k) - g(x_i^{k+1})) \geq \sum_{i=1}^{N} \langle w_i^k v_i^{k+1}, x_i^k - x_i^{k+1} \rangle. \tag{4.51}$$

Substituting (4.50) and (4.51) into (4.49), we derive that

$$\begin{aligned}\Psi(x^k)-\Psi(x^{k+1}) &\geq (\frac{1}{\mu}-\frac{L_f}{2})\|x^k-x^{k+1}\|^2+\frac{\langle e^k, x^k-x^{k+1}\rangle}{\mu}\\ &\geq \frac{1}{2}(\frac{1}{\mu}-\frac{L_f}{2})\|x^k-x^{k+1}\|^2-\frac{\|e^k\|_2^2}{\mu(2-\mu L_f)},\end{aligned} \tag{4.52}$$

where we use the inequality $\langle e^k, x^k-x^{k+1}\rangle \geq -\frac{1}{2}(1-\frac{\mu L_f}{2})\|x^k-x^{k+1}\|^2-\frac{\|e^k\|_2^2}{2-\mu L_f}$.

□

Lemma 14. *Let $(x^k)_{k\geq 0}$ is generated by scheme (4.47) and $\frac{1}{\mu}>\frac{L_f}{2}$, and $\sum_k\sqrt{\sum_{l=k}^{+\infty}\|e^l\|^2}<+\infty$, and function Φ be coercive. Then, there exist $S, D>0$ such that*

$$\operatorname{dist}(\boldsymbol{0}, \partial\Psi(x^{k+1}))\leq S\|x^{k+1}-x^k\|+D\|e^k\|. \tag{4.53}$$

Proof. We can easily have that

$$\nabla f(x^{k+1})+W^{k+1}v^{k+1}\in\partial\Psi(x^{k+1}), \tag{4.54}$$

where $v_i^{k+1}\in\partial g(x_i^{k+1})$ and $W^{k+1}=\operatorname{diag}(h'(g(x_1^k)), h'(g(x_2^k)), \ldots, h'(g(x_N^k)))$. We employ relation (4.50). Then, we have that

$$v_i^{k+1}=[\frac{e_i^k}{\mu}+\frac{(x_i^k-x_i^{k+1})}{\mu}-\nabla_i f(x^k)]/w_i^k, i\in[1,2,\ldots,N]. \tag{4.55}$$

Combining (4.54) and (4.55), we have

$$\frac{w_i^{k+1}-w_i^k}{w_i^k}\nabla_i f(x^{k+1})+\frac{w_i^{k+1}}{w_i^k}[\frac{e_i^k}{\mu}+\frac{(x_i^k-x_i^{k+1})}{\mu}]\in\partial_i\Psi(x^{k+1}), i\in[1,2,\ldots,N]. \tag{4.56}$$

In view of that ∇f is continuous, so is $\nabla_i f(x)$; and from Lemmas 13 and 5, $\{x^k\}_{k=0,1,2,\ldots}$ is bounded. Hence, there exist $\widetilde{L}>0$ such that

$$\max_{1\leq i\leq N}\|\nabla_i f(x)\|_2\leq\widetilde{L}. \tag{4.57}$$

Considering that h' is nonzero and continuous, and $\{g(x_i^k)\}_{k=0,1,2,\ldots}$ is bounded ($i\in[1,2,\ldots,N]$). With [] and the convexity of g, there exists d_g

$$|g(x_i^k)-g(x_i^{k+1})|\leq d_g|x_i^k-x_i^{k+1}|. \tag{4.58}$$

Therefore, for any k and $i\in\{1,2,\ldots,N\}$, there exists $\delta,\pi>0$ such that

$$\delta\leq h'(g(x_i^k))=w_i^k\leq\pi.$$

Hence, we derive that

$$\max_{1\leq i\leq N}|\frac{w_i^{k+1}}{w_i^k}|\leq\frac{\pi}{\delta}, \max_{1\leq i\leq N}|\frac{1}{w_i^k}|\leq\frac{1}{\delta}. \tag{4.59}$$

From (4.56), with (4.59), we have

$$\begin{aligned}\operatorname{dist}(\mathbf{0}, \partial\Psi(x^{k+1})) &\leq \sum_{i=1}^N\left|\frac{w_i^{k+1}-w_i^k}{w_i^k}\nabla_i f(x^{k+1})+\frac{w_i^{k+1}}{w_i^k}[\frac{e_i^k}{\mu}+\frac{(x_i^k-x_i^{k+1})}{\mu}]\right|\\ &\leq \frac{\tilde{L}}{\delta}\sum_{i=1}^N|w_i^{k+1}-w_i^k|+\frac{\pi\sqrt{N}}{\mu\delta}\|e^k\|+\frac{\pi\sqrt{N}}{\mu\delta}\|x^{k+1}-x^k\|\end{aligned} \tag{4.60}$$

The problem also turns to estimating $|w_i^{k+1} - w_i^k|$. For any $i \in [1, 2, \ldots, N]$,

$$\begin{aligned} w_i^k - w_i^{k+1} &= h'(g(x_i^k)) - h'(g(x_i^{k+1})) \\ &\leq L_h|g(x_i^k) - g(x_i^{k+1})| = L_h d_g |x_i^{k+1} - x_i^k|. \end{aligned} \tag{4.61}$$

Combining (4.60) and (4.61), we obtain

$$\text{dist}(\mathbf{0}, \partial\Psi(x^{k+1})) \leq (\frac{\tilde{L}L_h d_g\sqrt{N}}{\delta} + \frac{\pi\sqrt{N}}{\mu\delta})\|x^{k+1} - x^k\| + \frac{\pi\sqrt{N}}{\mu\delta}\|e^k\| \tag{4.62}$$

□

Lemma 15. *Let $(x^k)_{k\geq 0}$ is generated by scheme (4.47) and $\frac{1}{\mu} > \frac{L_f}{2}$, and $\sum_k \sqrt{\sum_{l=k}^{+\infty} \|e^l\|^2} < +\infty$, and function Φ be coercive. Then, for any $x^* \in \text{crit}(\Phi)$, there exists a subsequence $(x^{k_j})_{j\geq 0}$ converges to x^* satisfying $\Phi(x^{k_j}) \to \Phi(x^*)$.*

Proof. The continuity of the function Ψ directly gives the result. □

Theorem 4. *Suppose that f, g, h are all semi-algebraic, and $\text{dom}(f) = \text{dom}(\nabla f)$, $\text{dom}(g) = \text{dom}(\partial g)$; and Ψ is coercive, and $\frac{1}{\mu} > \frac{L_f}{2}$. Let the sequence $(x^k)_{k\geq 0}$ be generated by scheme (4.47). If the sequence $(\|e^k\|)_{k\geq 0}$ satisfies*

$$\sum_k \sqrt{\sum_{l=k}^{+\infty} \|e^l\|^2} < +\infty. \tag{4.63}$$

Then, the sequence $(x^k)_{k\geq 0}$ has finite length, i.e.

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\| < +\infty. \tag{4.64}$$

And $\{x^k\}_{k=0,1,2,3,\ldots}$ converges to a critical point x^* of Ψ.

4.4 Inexact DC algorithm

In this part, we consider nonconvex optimization problems of the following type

$$\min\{\Phi(x) = f(x) + g(x) - h(x)\}, \tag{4.65}$$

where g is proper and lower semicontinuous, f is differentiable with L_f-Lipschitz gradient, and h is convex and differentiable with L_h-Lipschitz gradient. Such a problem is discussed in [26]. If f vanishes, problem (4.65) will reduce to the DC programming [37]

$$\min\{g(x) - h(x)\}. \tag{4.66}$$

A novel DC algorithm is proposed in [2] for (4.65) and the convergence is also proved. The inexact version of this algorithm can be expressed as

$$x^{k+1} \in \mathbf{prox}_{\gamma g}\left(x^k - \gamma(\nabla f(x^k) - \nabla h(x^k)) + e^k\right), \tag{4.67}$$

where γ is the stepsize, and e^k is the noise. The cautious reader may find that iteration (4.67) is actually a special case of (4.5) if regarding $f - h$ as a whole. But with the specific structure, iteration (4.67) enjoys more properties than (4.5), like larger stepsize. It is easy to see that $\nabla(f - h) = \nabla f - \nabla h$ is Lipchitz with $L_f + L_h$. If directly using the convergence results for (4.5) (Theorem 2), the stepsize γ shall satisfy $\lambda < \frac{1}{L_f+L_h}$. However, a larger step can be selected for iteration (4.67); the stepsize can be $\lambda < \frac{2}{L_f}$ (Lemma 16).

Lemma 16. *Let $(x^k)_{k\geq 0}$ is generated by scheme (4.67) and $0<\lambda<\frac{2}{L_f}$. Then, we will have*

$$\Phi(x^k)-\Phi(x^{k+1})\geq (\frac{1}{\lambda}-\frac{L_f}{2})\|x^k-x^{k+1}\|^2-\frac{\|e^k\|^2}{\lambda(2-\lambda L_f)}. \tag{4.68}$$

Direct computations yield

$$\begin{aligned}\Phi(x^k)-\Phi(x^{k+1}) &= f(x^k)-f(x^{k+1})+g(x^k)-g(x^{k+1})+h(x^{k+1})-h(x^k)\\ &\geq \langle\nabla f(x^k),x^k-x^{k+1}\rangle-\frac{L_f}{2}\|x^k-x^{k+1}\|_2^2+g(x^k)-g(x^{k+1})+\langle x^{k+1}-x^k,\nabla h(x^k)\rangle.\end{aligned} \tag{4.69}$$

On the other hand, with Lemma 6, we have

$$\gamma g(x^{k+1})+\frac{\|x^k-\gamma(\nabla f(x^k)-\nabla h(x^k))+e^k-x^{k+1}\|^2}{2}\leq \gamma g(x^k)+\frac{\|-\gamma(\nabla f(x^k)-\nabla h(x^k))+e^k\|^2}{2}. \tag{4.70}$$

Combining (4.69) and (4.70), we derive that

$$\begin{aligned}\Phi(x^k)-\Phi(x^{k+1}) &\geq (\frac{1}{\lambda}-\frac{L_f}{2})\|x^k-x^{k+1}\|^2+\frac{\langle e^k,x^k-x^{k+1}\rangle}{\lambda}\\ &\geq \frac{1}{2}(\frac{1}{\lambda}-\frac{L_f}{2})\|x^k-x^{k+1}\|^2-\frac{\|e^k\|_2^2}{\lambda(2-\lambda L_f)},\end{aligned} \tag{4.71}$$

where we use the inequality $\langle e^k,x^k-x^{k+1}\rangle\geq -\frac{1}{2}(1-\frac{\lambda L_f}{2})\|x^k-x^{k+1}\|^2-\frac{\|e^k\|^2}{2-\lambda L_f}$.

Lemma 17. *Let $(x^k)_{k\geq 0}$ is generated by scheme (4.67). Then, there exist $S,D>0$ such that*

$$\mathrm{dist}(\boldsymbol{0},\partial\Phi(x^{k+1}))\leq S\|x^{k+1}-x^k\|+D\|e^k\|. \tag{4.72}$$

Proof. With scheme of the algorithm,

$$\frac{x^k-x^{k+1}}{\gamma}-\nabla f(x^k)+\frac{e^k}{\gamma}+\nabla h(x^k)\in\partial g(x^{k+1}). \tag{4.73}$$

Thus, we have

$$\frac{x^k-x^{k+1}}{\gamma}+\nabla f(x^{k+1})-\nabla f(x^k)+\frac{e^k}{\gamma}+\nabla h(x^k)-\nabla h(x^{k+1})\in\partial\Phi(x^{k+1}). \tag{4.74}$$

Hence,

$$\begin{aligned}\mathrm{dist}(\mathbf{0},\partial\Phi(x^{k+1})) &\leq \|\frac{x^k-x^{k+1}}{\gamma}+\nabla f(x^{k+1})-\nabla f(x^k)+\frac{e^k}{\gamma}+\nabla h(x^k)-\nabla h(x^{k+1})\|\\ &\leq (\frac{1}{\gamma}+L_f+L_h)\|x^k-x^{k+1}\|+\frac{1}{\gamma}\|e^k\|.\end{aligned} \tag{4.75}$$

□

Lemma 18. *Let $(x^k)_{k\geq 0}$ is generated by scheme (4.67) and $\frac{1}{\lambda}>\frac{L_f}{2}$, and Φ be coercive, and $\sum_k\sqrt{\sum_{l=k}^{+\infty}\|e^l\|^2}<+\infty$. Then, for any $x^*\in\mathrm{crit}(\Phi)$, there exists a subsequence $(x^{k_j})_{j\geq 0}$ converges to x^* satisfying $\Phi(x^{k_j})\to\Phi(x^*)$.*

Proof. Obviously, we have $e^k\to\mathbf{0}$. With Lemma 5, $(x^k)_{k\geq 0}$ is bounded. For any $x^*\in\mathrm{crit}(\Phi)$, there exists a subsequence $(x^{k_j})_{j\geq 0}$ converges to x^*. With Lemmas 5 and 16, we also have

$$x^{k_j-1}\to x^*. \tag{4.76}$$

And in each iteration of updating x^{k_j}, with Lemma 6, we have

$$\gamma g(x^{k_j-1}) + \frac{\|x^k-\gamma(\nabla f(x^{k_j-1})-\nabla h(x^{k_j-1})+e^{k_j-1}-x^{k_j}\|^2}{2} \leq \gamma g(x^{k_j-1})+\frac{\|-\gamma(\nabla f(x^{k_j-1})-\nabla h(x^{k_j-1}))+e^{k_j-1}\|^2}{2}. \tag{4.77}$$

Taking $j\to+\infty$, we have

$$\limsup_{j\to+\infty} \ g(x^{k_j})\leq g(x^*). \tag{4.78}$$

And recalling the lower semi-continuity of g,

$$g(x^*)\leq \liminf_{j\to+\infty} \ g(x^{k_j}). \tag{4.79}$$

That means $\lim_j g(x^{k_j})=g(x^*)$; combining the continuity of f and h, we then prove the result. □

Theorem 5. *Let $(x^k)_{k\geq 0}$ is generated by scheme (4.67). Functions f, g and h are all semi-algebraic, and their domains satisfy* $\mathrm{dom}(\nabla f)=\mathrm{dom}(f)$, $\mathrm{dom}(\partial g)=\mathrm{dom}(g)$, $\mathrm{dom}(\nabla h)=\mathrm{dom}(h)$. *And the stepsize satisfies $\frac{1}{\lambda}>\frac{L_f}{2}$, and Φ is coercive, and $\sum_k\sqrt{\sum_{l=k}^{+\infty}\|e^l\|^2}<+\infty$. Then, the sequence $(x^k)_{k\geq 0}$ has finite length, i.e.*

$$\sum_{k=0}^{+\infty}\|x^{k+1}-x^k\|<+\infty. \tag{4.80}$$

And $\{x^k\}_{k=0,1,2,3,...}$ converges to a critical point x^* of Φ.

4.5 Inexact nonconvex ADMM algorithm

Alternating Direction Method of Multipliers (ADMM) [14, 15] is a powerful tool for the minimization of composite functions with linear constraints. An inexact nonconvex ADMM scheme is considered for the composite optimization

$$\min_{x,y}\{f(x)+g(y), \ \text{s.t.} \ \ x+y=\mathbf{0},\}^1 \tag{4.81}$$

where g is differentiable with L_g-Lipschitz gradient. We consider the following inexact algorithm as

$$\begin{cases} x^{k+1}=\mathbf{prox}_{r_1 f}(x^k-r_1[\gamma^k+\beta(x^k+y^k)]+e_1^k), \\ y^{k+1}=\mathbf{prox}_{r_2 g}(y^k-r_2[\gamma^k+\beta(x^{k+1}+y^k)]+e_2^k), \\ \gamma^{k+1}=\gamma^k+\beta(x^{k+1}+y^{k+1}). \end{cases} \tag{4.82}$$

If $e_1^k\equiv\mathbf{0}$ and $e_2^k\equiv\mathbf{0}$, the scheme is the linearized ADMM [12]. Nonconvex ADMM has been frequently studied in recent years [39, 20, 21, 32, 36, 1, 16]. The augmented Lagrangian function is defined as

$$L_\beta(x,y,\gamma)=f(x)+g(y)+\langle\gamma,x+y\rangle+\frac{\beta}{2}\|x+y\|_2^2, \tag{4.83}$$

where γ is the Lagrangian dual variable. The First, we prove a critical lemma.

Lemma 19. *Let $(x^k,y^k,\gamma^k)_{k\geq 0}$ be generated by (4.82), we then have*

$$\|\gamma^{k+1}-\gamma^k\|^2\leq\rho_1\|y^k-y^{k+1}\|^2+\rho_2\|y^{k-1}-y^k\|^2+\rho_3\|e_2^{k+1}-e_2^k\|^2, \tag{4.84}$$

where $\rho_1=3(\frac{1}{r_2}-\beta+L_g)^2$, $\rho_2=3(\frac{1}{r_2}-\beta)^2$ and $\rho_3=\frac{3}{r_2^2}$.

[1] The reuslt can be easiy extended to a more general constraint $Ax+By=c$. Here, we consider this case just for the simplicity of presentation.

Proof. The second step of each iteration gives

$$y^{k+1} + r_2\nabla g(y^{k+1}) = y^k - r_2[\gamma^k + \beta(x^{k+1} + By^k)] + e_2^k. \tag{4.85}$$

With the fact $\gamma^{k+1} = \gamma^k + \beta(x^{k+1} + y^{k+1})$, we then have

$$\gamma^{k+1} = \frac{y^k - y^{k+1}}{r_2} - \nabla g(y^{k+1}) - \beta(y^k - y^{k+1}) + \frac{e_2^k}{r_2}. \tag{4.86}$$

Substituting $k+1$ with k,

$$\gamma^k = \frac{y^{k-1} - y^k}{r_2} - \nabla g(y^k) - \beta(y^{k-1} - y^k) + \frac{e_2^{k-1}}{r_2}. \tag{4.87}$$

Substraction of the two equalities above yield

$$\begin{aligned} \|\gamma^{k+1} - \gamma^k\|^2 &\leq 3(\frac{1}{r_2} - \beta + L_g)^2\|y^k - y^{k+1}\|^2 + 3(\frac{1}{r_2} - \beta)^2\|y^{k-1} - y^k\|^2 \\ &+ \frac{3}{r_2^2}\|e_2^{k-1} - e_2^k\|^2 \end{aligned} \tag{4.88}$$

□

We define an auxiliary point as

$$d^k := (x^k, y^k, \gamma^k, y^{k-1}), \omega^k := (x^k, y^k), \varepsilon^k = \begin{pmatrix} e_2^{k+1} - e_2^k \\ e_1^k \\ e_2^k \end{pmatrix}$$

and the Lyapunov function as

$$F(d) = F(x, y, \gamma, \bar{y}) := L_\beta(x, y, \gamma) + \frac{\rho_2}{\beta}\|y - \bar{y}\|^2.$$

In the following, we prove the conditions for F.

Lemma 20. *If* $\sum_k \|\omega^k - \omega^{k+1}\| < +\infty$, *we have* $\sum_k \|d^k - d^{k+1}\| < +\infty$.

Proof. Direct basic algebraic computation gives the result. □

Lemma 21. *Let* $(d^k)_{k\geq 0}$ *is generated by scheme (4.82) and*

$$\frac{1}{r_1} - \beta > 0, \frac{1}{r_2} - \beta - \frac{12(\frac{1}{r_2} - \beta + L_g)^2}{\beta} - \frac{12(\frac{1}{r_2} - \beta)^2}{\beta} > 0. \tag{4.89}$$

Then, we will have

$$F(d^k) - F(d^{k+1}) \geq \nu\|z^k - z^{k+1}\|_2^2 - \rho\|\varepsilon^k\|^2 \tag{4.90}$$

for some $\nu, \rho > 0$.

Proof. With Lemma 6,

$$r_1 f(x^{k+1}) + \frac{1}{2}\|x^k - r_1[\gamma^k + \beta(x^k + y^k)] + e_1^k - x^{k+1}\|^2 \leq r_1 f(x^k) + \frac{1}{2}\| - r_1[\gamma^k + \beta(x^k + y^k)] + e_1^k\|^2. \tag{4.91}$$

With direct calculations, (4.91) can be represented as

$$L_\beta(x^{k+1}, y^k, \gamma^k) \leq L_\beta(x^k, y^k, \gamma^k) + \frac{\beta - \frac{1}{r_1}}{2}\|x^{k+1} - x^k\|^2 + \frac{\langle e_1^k, x^k - x^{k+1}\rangle}{r_1}. \tag{4.92}$$

By using the inequality

$$\frac{\langle e_1^k, x^k - x^{k+1}\rangle}{r_1} \le \frac{1}{(1-\beta r_1)r_1}\|e_1^k\|^2 + \frac{\frac{1}{r_1}-\beta}{4}\|x^{k+1}-x^k\|^2, \tag{4.93}$$

in (4.92), we have

$$L_\beta(x^{k+1}, y^k, \gamma^k) \le L_\beta(x^k, y^k, \gamma^k) + \frac{\beta - \frac{1}{r_1}}{4}\|x^{k+1}-x^k\|^2 + \frac{1}{(1-\beta r_1)r_1}\|e_1^k\|^2. \tag{4.94}$$

Similarly, we have

$$L_\beta(x^{k+1}, y^{k+1}, \gamma^k) \quad \le \quad L_\beta(x^{k+1}, y^k, \gamma^k) + \frac{\beta - \frac{1}{r_2}}{4}\|y^{k+1}-y^k\|^2 + \frac{1}{(1-\beta r_2)r_2}\|e_2^k\|^2. \tag{4.95}$$

With Lemma 19,

$$\begin{aligned} L_\beta(x^{k+1}, y^{k+1}, \gamma^{k+1}) \quad &= \quad L_\beta(x^{k+1}, y^{k+1}, \gamma^k) + \frac{\|\gamma^{k+1}-\gamma^k\|^2}{\beta} \\ &\le \quad L_\beta(x^{k+1}, y^{k+1}, \gamma^k) + \frac{\rho_1}{\beta}\|y^k - y^{k+1}\|^2 + \frac{\rho_2}{\beta}\|y^{k-1}-y^k\|^2 \\ &+ \quad \frac{\rho_3}{\beta}\|e_2^{k+1} - e_2^k\|^2. \end{aligned} \tag{4.96}$$

Thus, we have

$$\begin{aligned} F(d^{k+1}) \quad &+ \quad (\frac{1}{4r_1} - \frac{\beta}{4})\|x^{k+1}-x^k\|^2 + [\frac{1}{4r_2} - \frac{\beta}{4} - \frac{\rho_1+\rho_2}{\beta}]\|y^{k+1}-y^k\|^2 \\ &- \quad \max\{\frac{1}{(1-\beta r_1)r_1}, \frac{1}{(1-\beta r_2)r_2}, \frac{\rho_3}{\beta}\}\cdot\|\varepsilon^k\|^2 \le F(d^k). \end{aligned} \tag{4.97}$$

Letting $\nu := \min\{\frac{1}{4r_1} - \frac{\beta}{4}, \frac{1}{4r_2} - \frac{\beta}{4} - \frac{\rho_1+\rho_2}{\beta}\}$ and $\rho := \max\{\frac{1}{(1-\beta r_1)r_1}, \frac{1}{(1-\beta r_2)r_2}, \frac{\rho_3}{\beta}\}$, we then prove the result. □

Remark 3. *We stress that the condition (4.89) always holds. This is because the parameters r_1, r_2 and β are all selected by the user. For example, we choose*

$$r_1 = \frac{1}{\beta+1}, r_2 = \frac{1}{\beta+1}, \beta > 12L_g^2 + 24L_g + 24. \tag{4.98}$$

That means if β is large enough, the condition can be satisfied.

Lemma 22. *Let $(w^k)_{k\ge0}$ is generated by scheme (4.82). Then, there exist $S, D > 0$ such that*

$$\mathrm{dist}(\boldsymbol{0}, \partial F(d^{k+1})) \le S\|z^{k+1}-z^k\| + D\|\varepsilon^k\|. \tag{4.99}$$

Proof. From Lemma 19, we have

$$\|\lambda^{k+1} - \lambda^k\| \le \sqrt{\rho_1}\|y^k - y^{k+1}\| + \sqrt{\rho_2}\|y^{k-1}-y^k\| + \sqrt{\rho_3}\|e_1^{k+1} - e_1^k\|. \tag{4.100}$$

The optimization condition for updating x^{k+1} is

$$\frac{x^k - x^{k+1}}{r_1} - \gamma^{k+1} + \frac{e_1^k}{r_1} \in \partial f(x^{k+1}). \tag{4.101}$$

With direct calculation, we have

$$\partial_x F(d^{k+1}) = \partial f(x^{k+1}) + \gamma^{k+1} + \beta(x^{k+1}+y^{k+1}) \tag{4.102}$$

Thus, we have

$$\begin{aligned}
\text{dist}[\mathbf{0}, \partial_x F(d^{k+1})] &\leq \|\frac{x^k - x^{k+1}}{r_1} + \beta(x^{k+1} + y^{k+1}) + \frac{e_1^k}{r_1}\| \\
&\leq \|\frac{x^k - x^{k+1}}{r_1}\| + \|\beta(x^{k+1} + y^{k+1})\| + \|\frac{e_1^k}{r_1}\| \\
&= \|\frac{x^k - x^{k+1}}{r_1}\| + \|\lambda^{k+1} - \lambda^k\| + \frac{\|e_1^k\|}{r_1} \\
&\leq \frac{1}{r_1}\|x^k - x^{k+1}\| + \sqrt{\rho_1}\|y^k - y^{k+1}\| + \sqrt{\rho_2}\|y^{k-1} - y^k\| + \sqrt{\rho_3}\|e_1^{k+1} - e_1^k\| + \frac{\|e_1^k\|}{r_1} \\
&\leq S_x(\|z^{k+1} - z^k\| + \|z^k - z^{k-1}\|) + D_x(\|e_1^{k+1} - e_1^k\| + \|e_1^k\|), \qquad (4.103)
\end{aligned}$$

where $S_x = \max\{\frac{1}{r_1}, \sqrt{\rho_2}, \sqrt{\rho_3}\}$ and $D_x = \max\{\sqrt{\rho_3}, \frac{1}{r_1}\}$. While in updating y^{k+1}, we have

$$\frac{y^k - y^{k+1}}{r_2} - [\gamma^k + \beta(x^{k+1} + y^k)] + \frac{e_2^k}{r_2} = \nabla g(y^{k+1}). \qquad (4.104)$$

And we have

$$\partial_y F(d^{k+1}) = \nabla g(y^{k+1}) + \gamma^{k+1} + \beta(x^{k+1} + y^{k+1}) + \frac{2\rho_2}{\beta}(y^{k+1} - y^k) \qquad (4.105)$$

Combining (4.104) and (4.105),

$$\begin{aligned}
\text{dist}[\mathbf{0}, \partial_y F(d^{k+1})] &\leq \|\frac{y^k - y^{k+1}}{r_2} + (\gamma^{k+1} - \gamma^k) + \beta(y^{k+1} - y^k) + \frac{2\rho_2}{\beta}(y^{k+1} - y^k) + \frac{e_2^k}{r_2}\| \\
&\leq (\frac{1}{r_2} + \beta + \frac{2\rho_2}{\beta})\|y^{k+1} - y^k\| + \sqrt{\rho_1}\|y^k - y^{k+1}\| + \sqrt{\rho_2}\|y^{k-1} - y^k\| \\
&+ \sqrt{\rho_3}\|e_1^{k+1} - e_1^k\| + \frac{\|e_2^k\|}{r_2} \\
&\leq S_y(\|z^{k+1} - z^k\| + \|z^k - z^{k-1}\|) + D_y(\|e_1^{k+1} - e_1^k\| + \|e_2^k\|), \qquad (4.106)
\end{aligned}$$

where $S_y = \max\{\frac{1}{r_2} + \beta + \frac{2\rho_2}{\beta} + \sqrt{\rho_3}, \sqrt{\rho_2}\}$ and $D_y = \max\{\sqrt{\rho_3}, \frac{1}{r_2}\}$. Noting

$$\partial_\gamma F(d^{k+1}) = x^{k+1} + By^{k+1} = \frac{\gamma^{k+1} - \gamma^k}{\beta}, \qquad (4.107)$$

we have

$$\begin{aligned}
\text{dist}(\mathbf{0}, \partial_\gamma F(d^{k+1})) &\leq \frac{\|\gamma^{k+1} - \gamma^k\|}{\beta} \leq \frac{\sqrt{\rho_1}}{\beta}\|y^k - y^{k+1}\| + \frac{\sqrt{\rho_2}}{\beta}\|y^{k-1} - y^k\| + \frac{\sqrt{\rho_3}}{\beta}\|e_1^{k+1} - e_1^k\| \\
&\leq S_\gamma(\|z^{k+1} - z^k\| + \|z^k - z^{k-1}\|) + D_\gamma\|e_1^{k+1} - e_1^k\|, \qquad (4.108)
\end{aligned}$$

where $D_\gamma = \max\{\frac{\sqrt{\rho_1}}{\beta}, \frac{\sqrt{\rho_2}}{\beta}\}$ and $D_\gamma = \frac{\sqrt{\rho_3}}{\beta}$. The partial differential of F subject to $\bar{y}$ is

$$\partial_{\bar{y}} F(d^{k+1}) = \frac{2\rho_2}{\beta}(y^{k+1} - y^k); \qquad (4.109)$$

thus, we have

$$\text{dist}(\mathbf{0}, \partial_{\bar{y}} F(d^{k+1})) \leq \frac{2\rho_2}{\beta}\|y^{k+1} - y^k\| \leq \frac{2\rho_2}{\beta}(\|z^{k+1} - z^k\| + \|z^k - z^{k-1}\|). \qquad (4.110)$$

Letting $S = S_x + S_y + S_\gamma + \frac{2\rho_2}{\beta}$ and $D = D_x + D_y + D_\gamma$, we then prove the result. □

Then, we prove $\inf F(d^k) > -\infty$. Then, we can obtain the boundedness of the points.

Lemma 23. *If there exists $\sigma_0 > 0$ such that*

$$\inf\{g(y) - \sigma\|\nabla g(y)\|^2\} > -\infty, \tag{4.111}$$

and

$$0 < \sigma \leq \sigma_0, \frac{3}{2\sigma} \leq \beta \leq \frac{3}{\sigma}. \tag{4.112}$$

We also assume that $(e_2^k)_{k\geq 0}$ is bounded and condition (4.89) holds, and $f(x)$ is coercive. Then, the sequence $\{d^k\}_{k=0,1,2,\ldots}$ is bounded.

Proof. From (4.87),

$$\|\gamma^k\|^2 \leq \rho_2\|y^k - y^{k-1}\|^2 + 3\|\nabla g(y^k)\|^2 + 3\frac{\|e_2^{k-1}\|^2}{r_2^2}. \tag{4.113}$$

We have

$$\begin{aligned}
F(d^k) &= f(x^k) + g(y^k) + \langle\gamma^k, x^k + y^k\rangle + \frac{\beta}{2}\|x^k + y^k\|^2 + \frac{\rho_2}{\beta}\|y^k - y^{k-1}\|^2 \\
&= f(x^k) + g(y^k) - \frac{\|\gamma^k\|^2}{2\beta} + \frac{\beta}{2}\|x^k + y^k + \frac{\gamma^k}{\beta}\|^2 + \frac{\rho_2}{\beta}\|y^k - y^{k-1}\|^2 \\
&= f(x^k) + g(y^k) - \frac{\sigma}{3}\|\gamma^k\|^2 + (\frac{\sigma}{3} - \frac{1}{2\beta})\|\gamma^k\|^2 \\
&+ \frac{\beta}{2}\|x^k + y^k + \frac{\gamma^k}{\beta}\|^2 + \frac{\rho_2}{\beta}\|y^k - y^{k-1}\|^2 \\
\overset{(4.113)}{\geq}& f(x^k) + g(y^k) - \sigma_0\|\nabla g(y^k)\|^2 + \rho_2(\frac{1}{\beta} - \frac{\sigma}{3})\|y^k - y^{k-1}\|^2 + (\sigma_0 - \sigma)\|\nabla g(y^k)\|^2 \\
&+ (\frac{\sigma}{3} - \frac{1}{2\beta})\|\gamma^k\|^2 + \frac{\beta}{2}\|x^k + y^k + \frac{\gamma^k}{\beta}\|^2 - \frac{\sigma\|e_2^k\|^2}{r_2^2}.
\end{aligned} \tag{4.114}$$

We then can see $\{f(x^k)\}_{k=0,1,2,\ldots}$, $\{\gamma^k\}_{k=0,1,2,\ldots}$, $\{x^k + y^k + \frac{\gamma^k}{\beta}\}_{k=0,1,2,\ldots}$ are all bounded. It is easy to see that one of the two conditions holds, $\{d^k\}_{k=0,1,2,\ldots}$ will be bounded. □

Remark 4. *The intersection between conditions (4.112) and (4.89) can be always nonempty. This is because we can always choose small σ. For example, we still use the setting (4.98). In this case, we just need to set*

$$\sigma \leq \min\{\frac{1}{12L_f^2 + 24L_f + 24}, \sigma_0\}. \tag{4.115}$$

Remark 5. *The condition (4.111) holds for many quadratical functions [20, 32]. This condition also implies the function g is similar to quadratical function and its property is "good".*

Lemma 24. *Let $(d^k)_{k\geq 0}$ is generated by scheme (4.82) and $\sum_k \sqrt{\sum_{l=k}^{+\infty}\|\varepsilon^l\|^2} < +\infty$. And let conditions of Lemmas 21 and 23 hold. Then, for any stationary point d^* of $(d^k)_{\geq 0}$, there exists a subsequence $(d^{k_j})_{j\geq 0}$ converges to d^* satisfying $F(d^{k_j}) \to F(d^*)$.*

Proof. Obviously, we have $e_2^k \to \mathbf{0}$. With Lemma 23, $(d^k)_{k\geq 0}$ is bounded; so are $(x^k)_{k\geq 0}$ and $(y^k)_{k\geq 0}$. For any stationary point $d^* = (x^*, y^*, \gamma^*, y^*)$, there exists a subsequence $(d^{k_j})_{j\geq 0}$ converges to d^*. With Lemmas 5 and 21, we also have

$$x^{k_j-1} \to x^*, y^{k_j-1} \to y^*. \tag{4.116}$$

Noting $0 \in \partial F(d^*)$, $x^* + y^* = \mathbf{0}$; thus, $\gamma^{k_j-1} \to \gamma^*$. And in each iteration of updating x^{k_j}, with Lemma 6, we have

$$\begin{aligned} r_1 f(x^{k_j}) &+ \frac{\|x^{k_j} - r_1[\gamma^{k_j-1} + \beta(x^{k_j-1} + y^{k_j-1})] + e_1^{k_j-1} - x^{k_j}\|^2}{2} \\ &\leq r_1 f(x^{k_j-1}) + \frac{\| - r_1[\gamma^{k_j-1} + \beta(x^{k_j-1} + y^{k_j-1})] + e^{k_j-1}\|^2}{2}. \end{aligned} \tag{4.117}$$

Taking $j \to +\infty$, we have

$$\limsup_{j \to +\infty} f(x^{k_j}) \leq f(x^*). \tag{4.118}$$

And recalling the lower semi-continuity of f,

$$f(x^*) \leq \liminf_{j \to +\infty} f(x^{k_j}). \tag{4.119}$$

That means $\lim_j f(x^{k_j}) = f(x^*)$; combining the continuity of g, we then prove the result. □

Finally, we present the convergence result for the inexact ADMM (4.82).

Theorem 6. *Let $(x^k)_{k\geq 0}$ is generated by scheme (4.82) and conditions of Lemmas 21 and 23 hold. Assume that f and g are both semi-algebraic, and* $\mathrm{dom}(f) = \mathrm{dom}(\partial f)$, $\mathrm{dom}(g) = \mathrm{dom}(\nabla g)$*. If*

$$\sum_k \sqrt{\sum_{l=k}^{+\infty} (\|e_1^l\|^2 + \|e_2^l\|^2)} < +\infty,$$

then, the sequence $(x^k, y^k)_{k\geq 0}$ has finite length, i.e.

$$\sum_{k=0}^{+\infty} (\|x^{k+1} - x^k\|_2 + \|y^{k+1} - y^k\|_2) < +\infty. \tag{4.120}$$

Proof. Noting

$$\begin{aligned} \sum_k \sqrt{\sum_{l=k}^{+\infty} \|\varepsilon^l\|^2} &\leq \sum_k \sqrt{\sum_{l=k}^{+\infty} (3\|e_1^l\|^2 + 3\|e_2^l\|^2)} \\ &= \sqrt{3} \cdot \sum_k \sqrt{\sum_{l=k}^{+\infty} (\|e_1^l\|^2 + \|e_2^l\|^2)} < +\infty, \end{aligned} \tag{4.121}$$

with the lemmas proved in this part and Theorem 1, we then prove the result. □

5 Conclusion

In this paper, we prove the convergence for a class of inexact nonconvex and nonsmooth algorithms. The sequence generated by the algorithm converges to a critical point of the objective function under finite energy assumption on the noise and the KL property assumption. We apply our theoretical results to many specific algorithms; and obtain the specific convergence results.

Acknowledgments

We are grateful for the support from the National Science Foundation of China (No.61402495) and (No. 61603322).

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] Nguyen Thai An and Nguyen Mau Nam. Convergence analysis of a proximal point algorithm for minimizing differences of functions. *Optimization*, 66(1):129–147, 2017.

[3] Hedy Attouch, Jerome Bolte, Patrick Redont, and Antoine Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

[4] Hedy Attouch, Jerome 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.

[5] Amir Beck. On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes. *SIAM Journal on Optimization*, 25(1):185–209, 2015.

[6] Jerome 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.

[7] Jérôme Bolte, Aris Daniilidis, Adrian Lewis, and Masahiro Shiota. Clarke subgradients of stratifiable functions. *SIAM Journal on Optimization*, 18(2):556–572, 2007.

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

[9] Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14(5):877–905, 2008.

[10] 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.

[11] Soumyadeep Chatterjee, Sheng Chen, and Arindam Banerjee. Generalized dantzig selector: Application to the k-support norm. In *Advances in Neural Information Processing Systems*, pages 1934–1942, 2014.

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

[13] 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.

[14] 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.

[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] 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.

[17] 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.

[18] Ming-Jun Lai, Yangyang Xu, and Wotao Yin. Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization. *SIAM Journal on Numerical Analysis*, 51(2):927–957, 2013.

[19] Chengbo Li, Wotao Yin, Hong Jiang, and Yin Zhang. An efficient augmented lagrangian method with applications to total variation minimization. *Computational Optimization and Applications*, 56(3):507–530, 2013.

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

[21] 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.

[22] Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Convergence rates with inexact non-expansive operators. *Mathematical Programming*, 159(1-2):403–434, 2016.

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

[24] Canyi Lu, Yunchao Wei, Zhouchen Lin, and Shuicheng Yan. Proximal iteratively reweighted algorithm with multiple splitting for nonconvex sparsity optimization. In *AAAI*, pages 1251–1257, 2014.

[25] Zhaosong Lu, Yong Zhang, and Jian Lu. ℓ_p regularized low-rank approximation via iterative reweighted singular value minimization. *Computational Optimization and Applications*, pages 1–24, 2017.

[26] Paul-Emile Maingé and Abdellatif Moudafi. Convergence of new inertial proximal methods for dc programming. *SIAM Journal on Optimization*, 19(1):397–413, 2008.

[27] Jorge Nocedal and Stephen J Wright. *Sequential quadratic programming*. Springer, 2006.

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

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

[30] Mark Schmidt, Nicolas L Roux, and Francis R Bach. Convergence rates of inexact proximal-gradient methods for convex optimization. In *Advances in neural information processing systems*, pages 1458–1466, 2011.

[31] Ron Shefi and Marc Teboulle. On the rate of convergence of the proximal alternating linearized minimization algorithm for convex problems. *EURO Journal on Computational Optimization*, 4(1):27–46, 2016.

[32] Tao Sun, Roberto Barrio, Hao Jiang, and Lizhi Cheng. Precompact convergence of the nonconvex primal-dual hybrid gradient algorithm. *Journal of Computational and Applied Mathematics*, 2018.

[33] Tao Sun and Lizhi Cheng. Little-o convergence rates for several alternating minimization methods. *Communications in Mathematical Sciences*, 15(1):197–211, 2017.

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

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

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

[37] Pham Dinh Tao and Le Thi Hoai An. Convex analysis approach to dc programming: Theory, algorithms and applications. *Acta Mathematica Vietnamica*, 22(1):289–355, 1997.

[38] Silvia Villa, Saverio Salzo, Luca Baldassarre, and Alessandro Verri. Accelerated and inexact forward-backward algorithms. *SIAM Journal on Optimization*, 23(3):1607–1633, 2013.

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

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