SIAM J. OPTIM.
Vol. 35, No. 2, pp. 1089–1109

CONVERGENCE OF ZH-TYPE NONMONOTONE DESCENT METHOD FOR KURDYKA–ŁOJASIEWICZ OPTIMIZATION PROBLEMS*

YITIAN QIAN†, TING TAO‡, SHAOHUA PAN§, AND HOUDUO QI¶

Abstract. We propose a novel iterative framework for minimizing a proper lower semicontinuous Kurdyka–Łojasiewicz (KL) function Φ. It comprises a Zhang–Hager (ZH-type) nonmonotone decrease condition and a relative error condition. Hence, the sequence generated by the ZH-type nonmonotone descent methods will fall within this framework. Any sequence conforming to this framework is proved to converge to a critical point of Φ. If in addition Φ has the KL property of exponent $\theta\in(0,1)$ at the critical point, the convergence has a linear rate for $\theta\in(0,1/2]$ and a sublinear rate of exponent $\frac{1-\theta}{1-2\theta}$ for $\theta\in(1/2,1)$. To the best of our knowledge, this is the first work to establish the full convergence of the iterate sequence generated by a ZH-type nonmonotone descent method for nonconvex and nonsmooth optimization problems. The obtained results are also applied to achieve the full convergence of the iterate sequences produced by the proximal gradient method and Riemannian gradient method with the ZH-type nonmonotone line-search.

Key words. ZH-type nonmonotone descent method, full convergence, KL property, nonconvex and nonsmooth optimization, proximal gradient methods

MSC codes. 90C26, 65K05, 49M27

DOI. 10.1137/24M1669153

1. Introduction.

For a descent method for nonconvex and nonsmooth optimization problems, it has become a common task to establish the full convergence of its iterate sequence under the Kurdyka–Łojasiewicz (KL) property of objective functions. This task has been strongly supported and motivated by the work [1] for gradient-related methods, [2] for proximal point algorithms, [4] for forward-backward algorithms, [3, 5] for proximal alternating minimization algorithms, and [28] for block coordinate descent algorithms, to name just a few of the representative papers in this research direction. The convergence analysis of all these works is based on the (sufficiently) monotone decrease of objective value sequences. Although the (sufficiently) monotone decrease is crucial to achieve the full convergence, it also leads to relatively short step sizes. To overcome this drawback, nonmonotone variants of the sufficient decrease are often considered. In the context of line-search methods, Grippo, Lampariello, and Lucidi [11] first proposed a nonmonotone line-search procedure by monitoring the maximum objective value attained by the latest iterates (referred

*Received by the editors June 20, 2024; accepted for publication (in revised form) February 17, 2025; published electronically May 15, 2025.
https://doi.org/10.1137/24M1669153
Funding: The second author's work was funded by the Natural Science Foundation of Guangdong Province under project 2023A1515111167. The third author's work was funded by the National Natural Science Foundation of China under project 12371299. The first and fourth authors were funded by Hong Kong RGC General Research Fund PolyU/15309223 and PolyU AMA Project 230413007.
†Department of Data Science and Artificial Intelligence, Hong Kong Polytechnic University, Hong Kong (yitian.qian@polyu.edu.hk).
‡School of Mathematics, Foshan University, Foshan, China (taoting@fosu.edu.cn).
§School of Mathematics, South China University of Technology, Guangzhou, China (shhpan@scut.edu.cn).
¶Department of Data Science and Artificial Intelligence, and Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong (houduo.qi@polyu.edu.hk).

to as GLL-type in this paper); Zhang and Hager [29] later proposed a different strategy of nonmonotone line-search (referred to as the ZH-type in this paper) by taking a weighted average of the current objective value and those of the past iterates. These two types of nonmonotone line-search procedures are received as the most popular choices although they can be cast into the general framework proposed in [23]. In the past decade, they have also been aligned with algorithms for nonconvex and nonsmooth optimization problems, such as composite optimization problems of minimizing the sum of a smooth function and a nonsmooth one [13, 17, 14, 25], Riemannian manifold optimization problems [27, 18], and DC programs [8, 16].

On the one hand, nonmonotone line-search procedures bring well-recognized numerical benefits such as increasing the possibility of seeking better critical points, and potentially accelerating the convergence of algorithms [26]. On the other hand, the nonmonotone decrease of objective value sequences brings a great challenge for the full convergence analysis of the iterate sequence of the concerned algorithms. Recently, Qian and Pan [19] investigated the full convergence of the iterate sequence conforming to an iterative framework proposed by the GLL-type decreasing condition for KL optimization problems and achieved its full convergence under a condition that is shown to be sufficient and necessary if objective functions are weakly convex on a neighborhood of the set of critical points. We notice that the ZH-type nonmonotone line-search procedure has been widely applied to nonconvex and nonsmooth optimization problems (see, e.g., [17, 27, 25, 18, 14]), but as far as we know, there is no work to study the full convergence of its iterate sequence even for smooth optimization problems [29, 10, 23]. This paper aims at resolving this open problem via a novel iterative framework.

In the following, we first describe the novel procedure, then explain the main results of the paper. To this end, consider the nonconvex and nonsmooth problem

$$\min_{x\in\mathbb{X}} \Phi(x), \tag{1.1}$$

where $\mathbb{X}$ represents a finite-dimensional real vector space endowed with the inner product $\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\|$, and $\Phi\colon\mathbb{X}\to\overline{\mathbb{R}}:=(-\infty,\infty]$ is a proper lower semicontinuous (lsc) function that is bounded below on its domain $\mathrm{dom}\,\Phi$. We are interested in nonmonotone descent methods for (1.1) to produce an iterate sequence $\{x^k\}_{k\in\mathbb{N}}\subset\mathrm{dom}\,\Phi$ complying with the following conditions:

- H1. For each $k\in\mathbb{N}$, $\Phi(x^k)+a_k\|x^k-x^{k-1}\|^2\le C_{k-1}$, where $a_k\ge \underline{a}$ for some $\underline{a}>0$, and $C_k=(1-\tau_k)C_{k-1}+\tau_k\Phi(x^k)$ with $C_0=\Phi(x^0)$ and $\tau_k\in[\tau,1]$ for a $\tau\in(0,1]$.
- H2. There exists a nonnegative integer k_1 such that for each $k\ge k_1$,

$$\mathrm{dist}(0,\partial\Phi(x^k))\le\frac{1}{b_k}\sum_{i=k-k_1}^{k+k_1}\|x^i-x^{i-1}\|+\varepsilon_k$$

  with $b_k>0$ and $\varepsilon_k\ge 0$, where $\partial\Phi(x^k)$ is the (limiting) subdifferential of Φ at x^k.
- H3. There exists a convergent subsequence $\{x^{k_j}\}_{j\in\mathbb{N}}$ with $\lim_{j\to\infty}x^{k_j}=\overline{x}\in\mathrm{dom}\,\Phi$ such that $\limsup_{j\to\infty}\Phi(x^{k_j})\le\Phi(\overline{x})$.

Unless otherwise stated, the above $\{a_k\}_{k\in\mathbb{N}}, \{b_k\}_{k\in\mathbb{N}}$, and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ are assumed to satisfy

$$\sum_{k=1}^{\infty}b_k=\infty,\ \overline{B}:=\sup_{\mathbb{N}\ni k\ge k_1}\frac{1}{b_k}\sum_{i=k-k_1}^{k+k_1}\frac{1}{\sqrt{a_i}}<\infty,\ \text{and}\ \sum_{k=1}^{\infty}\varepsilon_k<\infty. \tag{1.2}$$

Here, ":=" means "define." The recursion relation of $\{C_k\}_{k\in\mathbb{N}}$ in H1 comes from the ZH-type nonmonotone line-search procedure, and when C_k takes $\Phi(x^k)$ or $\tau_k\equiv 1$, condition H1 degenerates to the sufficiently monotone descent condition. In condition H2, the introduction of a nonnegative integer k_1 aims to bound $\mathrm{dist}(0,\partial\Phi(x^k))$ from above by $2k_1+1$ successive iterate errors. As will be shown in section 4.2, this plays a crucial role in the convergence analysis of nonmonotone Riemannian gradient descent methods. When Φ is a Lipschitz smooth function on $\mathbb{X}=\mathbb{R}^n$, the iterate sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the nonmonotone line-search algorithms in the style of ZH-type [29, 10, 23] under the direction assumption there falls within the above framework with $\varepsilon_k\equiv 0$. We include a proof in the appendix.

The main contribution of this work is to resolve the aforementioned open problem by establishing the full convergence of any sequence $\{x^k\}_{k\in\mathbb{N}}$ conforming to conditions H1–H3 under the KL property of Φ, and its R-linear convergence rate under the KL property of Φ with exponent $1/2$. The difficulty in achieving this goal is how to divide an iterate sequence $\{x^k\}_{k\in\mathbb{N}}$ obeying conditions H1–H3 into appropriate subsequences with the help of the objective values. Different from the iterative framework proposed in [19] for the GLL-type nonmonotone decrease condition, the nonmonotone decrease condition in H1 does not provide any hint on the division. To overcome the difficulty, we make full use of the recursion formula on C_k to skillfully divide $\{x^k\}_{k\in\mathbb{N}}$ into two subsequences, and then establish a recursion relation of the auxiliary sequence $\{\Xi_k\}_{k\in\mathbb{N}}$ in Lemma 3.1. Due to the remarkable difference between these two nonmonotone decrease conditions as showcased in the example [29, (1.3)], the convergence analysis here is also completely different from the one developed in [19]. A technical comparison on the sequence splitting scheme proposed in this paper and that of [19] can be found in Remark 3.2.

As discussed in [3, 12], there are a large number of nonconvex and nonsmooth optimization problems involving the KL functions. Furthermore, by [21, Proposition 1] and [15, Proposition 2(i) and Remark 1(b)], the piecewise linear-quadratic KL functions with the composite structure of [15] necessarily satisfy the KL property of exponent $1/2$. Thus, the obtained convergence results will have a wide range of applications. As a demonstration, section 4 achieves the full convergence of the iterate sequences produced by two existing algorithms: the proximal gradient method of [17] for composite optimization and the Riemannian gradient method of [27, 18] for Riemannian optimization. This result is new and enhances the existing convergence results for the two methods. Section 5 concludes the paper.

2. Notation and preliminary results.

The quantities and notation for describing H1 to H3 will be reserved for use throughout the paper and they are $\{a_k\}$, $\{b_k\}$, $\{C_k\}$, $\{\tau_k\}$, $\{\varepsilon_k\}$, $\overline{B}$, $\underline{a}$, τ, and k_1. We also use other (global) notations. Let m be the smallest positive integer such that $\sqrt{\tau}(m-k_1-1)\ge(1+\sqrt{1-\tau})(2k_1+1)\sqrt{m}$. Obviously, such m exists and $m>k_1+1$. For each $k\in\mathbb{N}$, write $\ell(k):=k+m-1$ and $\Xi_{k-1}:=\sqrt{C_{k-1}-C_k}$ (we will prove in Lemma 2.3(i) that $\{C_k\}$ is a nonincreasing sequence and hence Ξ_k is well defined). For a real number t, the floor operator $\lfloor t\rfloor$ is the largest integer not greater than t and $t_+:=\max\{0,t\}$ (the nonnegative part of t). For a proper $h\colon\mathbb{X}\to\overline{\mathbb{R}}$, denote by $\partial h(\overline{x})$ the (limiting) subdifferential of h at $\overline{x}\in\mathrm{dom}\,h$, and for any $-\infty<\eta_1<\eta_2<\infty$, write $[\eta_1<h<\eta_2]:=\{x\in\mathbb{X}\,|\,\eta_1<h(x)<\eta_2\}$. The point $\overline{x}$ at which $0\in\partial h(\overline{x})$ is called a critical point of h, and the set of all critical points of h is denoted by $\mathrm{crit}\,h$.

Now we introduce the formal definition of the KL property (with exponent $\theta\in[0,1)$).

DEFINITION 2.1 (see [3, Definition 3.1]). *For a given* $\eta\in(0,\infty]$, *denote by* Υ_η *the family of continuous concave* $\varphi\colon[0,\eta)\to\mathbb{R}_+$ *that is continuously differentiable on* $(0,\eta)$ *with* $\varphi'(s)>0$ *for all* $s\in(0,\eta)$ *and* $\varphi(0)=0$. *A proper function* $h\colon\mathbb{X}\to\overline{\mathbb{R}}$ *is said to have the KL property at* $\overline{x}\in\operatorname{dom}\partial h$ *if there exist* $\eta\in(0,\infty]$, *a neighborhood* $\mathcal{U}$ *of* $\overline{x}$, *and a function* $\varphi\in\Upsilon_\eta$ *such that for all* $x\in\mathcal{U}\cap\big[h(\overline{x})<h<h(\overline{x})+\eta\big]$,

$$\varphi'(h(x)-h(\overline{x}))\mathrm{dist}(0,\partial h(x))\geq 1.$$

If φ *can be chosen as* $\varphi(t)=ct^{1-\theta}$ *with* $\theta\in[0,1)$ *for some* $c>0$, *then* h *is said to have the KL property of exponent* θ *at* $\overline{x}$. *If* h *has the KL property (of exponent* θ*) at every point of* $\operatorname{dom}\partial h$, *then it is called a KL function (of exponent* θ*).*

Next we present two technical lemmas used for the subsequent analysis. The first one, proved in [19], says that if a nonnegative and nonincreasing sequence is bounded by a mixture of two special sequences related to the indices of the sequence, then the sequence itself can be bounded by one particular sequence related to its indices.

LEMMA 2.2 (see [19, Lemma 2.7]). *Let* $\{\beta_l\}_{l\in\mathbb{N}}\subset\mathbb{R}_+$ *be a nonincreasing sequence with* $\beta_l\leq\mu_0\max\big\{l^{\frac{1-\varsigma}{1-2\varsigma}},(\beta_{l-m_1}-\beta_l)^{\frac{1-\varsigma}{\varsigma}}\big\}$ *for all* $l\geq\max\{\overline{l},m_1\}$, *where* $\overline{l}$ *and* m_1 *are the nonnegative integers, and* $\mu_0>0$ *and* $\varsigma\in(\frac{1}{2},1)$ *are the constants. Then, there exists* $\widehat{\mu}>0$ *such that for all* $l\geq\overline{l}$, $\beta_l\leq\max\big\{\mu_0,\widehat{\mu}^{\frac{1-\varsigma}{1-2\varsigma}}\big\}\lfloor\frac{l-\overline{l}}{m_1+1}\rfloor^{\frac{1-\varsigma}{1-2\varsigma}}$.

The second one provides the desirable properties of $\{C_k\}_{k\in\mathbb{N}}$ and $\{\Phi(x^k)\}_{k\in\mathbb{N}}$.

LEMMA 2.3. *Let* $\{x^k\}_{k\in\mathbb{N}}$ *be a sequence obeying condition* H1. *Then, the following claims hold.*

(i) *For each* $k\in\mathbb{N}$, $\Phi(x^k)\leq C_k\leq C_{k-1}$, *so the sequence* $\{C_k\}_{k\in\mathbb{N}}$ *is convergent.*
(ii) *The sequence* $\{\Phi(x^k)\}_{k\in\mathbb{N}}$ *is convergent and has the same limit as* $\{C_k\}_{k\in\mathbb{N}}$.
(iii) $\lim_{k\to\infty}(x^{k+1}-x^k)=0$.
(iv) *If in addition* $\{x^k\}_{k\in\mathbb{N}}$ *complies with conditions* H2–H3, *then the sequences* $\{C_k\}_{k\in\mathbb{N}}$ *and* $\{\Phi(x^k)\}_{k\in\mathbb{N}}$ *both converge to* $\Phi(\overline{x})$ *with* $0\in\partial\Phi(\overline{x})$.

Proof. (i)–(iii) For each $k\in\mathbb{N}$, from condition H1, it immediately follows that

$$C_k=(1-\tau_k)C_{k-1}+\tau_k\Phi(x^k)\leq C_{k-1}-a_k\tau_k\|x^k-x^{k-1}\|^2 \tag{2.1}$$

$$\leq C_{k-1}-\underline{a}\tau\|x^k-x^{k-1}\|^2, \tag{2.2}$$

which implies $C_k\leq C_{k-1}$ and $\Phi(x^k)\leq C_{k-1}$. The latter, along with $\Phi(x^k)=(1-\tau_k)(\Phi(x^k)-C_{k-1})+C_k$, leads to $\Phi(x^k)\leq C_k$. Recall that Φ is bounded below on $\operatorname{dom}\Phi$ and $\{x^k\}_{k\in\mathbb{N}}\subset\operatorname{dom}\Phi$. The recursion relation of C_k in condition H1 implies that $\{C_k\}_{k\in\mathbb{N}}$ is lower bounded, so its convergence follows the nonincreasing. Note that $\Phi(x^k)-C_{k-1}=\frac{1}{\tau_k}(C_k-C_{k-1})$ and $\tau_k\in[\tau,1]$ for each $k\in\mathbb{N}$. Then, the sequence $\{\Phi(x^k)\}_{k\in\mathbb{N}}$ is convergent and has the same limit as $\{C_k\}_{k\in\mathbb{N}}$. From inequality (2.2) and the convergence of $\{C_k\}_{k\in\mathbb{N}}$, it is immediate to obtain $\lim_{k\to\infty}(x^{k+1}-x^k)=0$.

(iv) Combining H3 with the lower semicontinuity of Φ yields that $\lim_{j\to\infty}\Phi(x^{k_j})=\Phi(\overline{x})$. Along with part (ii), we have $\lim_{k\to\infty}C_k=\lim_{k\to\infty}\Phi(x^k)=\Phi(\overline{x})$, so that

$$\lim_{k\to\infty}\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}=\lim_{k\to\infty}\sum_{i=k-k_1}^{k+k_1}\sqrt{C_{i-1}-C_i}=0. \tag{2.3}$$

From condition H2, for each $k \geq k_1$, there exists $w^k \in \partial\Phi(x^k)$ satisfying the relations

$$\|w^k\| \leq \frac{1}{b_k} \sum_{i=k-k_1}^{k+k_1} \|x^i - x^{i-1}\| + \varepsilon_k \overset{(2.1)}{\leq} \frac{1}{\sqrt{\tau} b_k} \sum_{i=k-k_1}^{k+k_1} \frac{1}{\sqrt{a_i}} \Xi_{i-1} + \varepsilon_k$$

$$\leq \left(\frac{1}{\sqrt{\tau} b_k} \sum_{i=k-k_1}^{k+k_1} \frac{1}{\sqrt{a_i}} \right) \sum_{i=k-k_1}^{k+k_1} \Xi_{i-1} + \varepsilon_k \overset{(1.2)}{\leq} \frac{\overline{B}}{\sqrt{\tau}} \sum_{i=k-k_1}^{k+k_1} \Xi_{i-1} + \varepsilon_k. \tag{2.4}$$

Passing the limit $k \to \infty$ to the above inequality and using (2.3) lead to $\lim_{k\to\infty} w^k = 0$. Recall that $\partial\Phi$ is outer semicontinuous at $\overline{x}$ with respect to Φ-attentive convergence $x \underset{\Phi}{\to} \overline{x}$ (i.e., $x \to \overline{x}$ and $\Phi(x) \to \Phi(\overline{x})$) by [22, Proposition 8.7]. Together with $\lim_{k\to\infty} \Phi(x^k) = \Phi(\overline{x})$ and $\lim_{j\to\infty} x^{k_j} = \overline{x}$, we obtain $0 \in \partial\Phi(\overline{x})$. ☐

The inequality (2.2) has a useful implication when there are many terms. For any given indices k and k' with $k' \geq k$, it holds that

$$\sum_{i=k}^{k'} \|x^i - x^{i-1}\| \leq \frac{1}{\sqrt{\underline{a}\tau}} \sum_{i=k}^{k'} \sqrt{C_{i-1} - C_i} = \frac{1}{\sqrt{\underline{a}\tau}} \sum_{i=k}^{k'} \Xi_{i-1}. \tag{2.5}$$

This inequality will often be used in the convergence analysis of the next section.

3. Main results.

In this section, we report two results for the iterate sequence satisfying H1–H3. The first one is about its full convergence and the second is about its local convergence rate. We report them in two subsections. For convenience, in the rest of this paper, let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence satisfying conditions H1–H3.

3.1. Full convergence.

In Lemma 2.3(iii), we have proved that the two consecutive terms of $\{x^k\}_{k\in\mathbb{N}}$ get arbitrarily close as k increases. If we further assume that Φ has the KL property, the sequence becomes a Cauchy sequence. We achieve this by establishing $\sum_{i=0}^{\infty} \Xi_i < \infty$. The result $\sum_{i=0}^{\infty} \|x^{i+1} - x^i\| < \infty$ then follows (2.5).

We first establish an iterative bound for the sums of Ξ_i. To this end, we define

$$\mathcal{K}_1 := \left\{ k \in \mathbb{N} \mid \Phi(x^k) \leq C_{k+m} \right\} \text{ and } \mathcal{K}_2 := \left\{ k \in \mathbb{N} \mid \Phi(x^k) > C_{k+m} \right\}.$$

When Φ is assumed to have the KL property at $\overline{x}$, by Definition 2.1, there exist $\delta > 0, \eta > 0$, and $\varphi \in \Upsilon_\eta$ such that for all $x \in \mathbb{B}(\overline{x}, \delta) \cap [\Phi(\overline{x}) < \Phi < \Phi(\overline{x}) + \eta]$,

$$\varphi'(\Phi(x) - \Phi(\overline{x})) \mathrm{dist}(0, \partial\Phi(x)) \geq 1. \tag{3.1}$$

Recall that $C_k \geq \Phi(\overline{x})$ for each $k \in \mathbb{N}$ by Lemma 2.3(i) and (iv). Hence,

$$\Gamma_{k,k+m} := \varphi\big(\Phi(x^k) - \Phi(\overline{x})\big) - \varphi\big(C_{k+m} - \Phi(\overline{x})\big) \text{ for } k \in \mathcal{K}_2$$

is well defined. Now we are ready to state the bound on some partial sums of Ξ_i.

LEMMA 3.1. *Suppose that Φ has the KL property at $\overline{x}$. Let $\delta > 0, \eta > 0$, and $\varphi \in \Upsilon_\eta$ be chosen to satisfy (3.1). Pick any $\rho \in (0, \delta)$. Then, for each $k \in \mathbb{N}$ with $x^k \in \mathbb{B}(\overline{x}, \rho)$ and $\Phi(x^k) < \Phi(\overline{x}) + \eta$, the following inequality holds with $\widehat{c} := \frac{1}{2}\big(\overline{B}/\sqrt{\tau} + 1\big)$:*

$$\frac{\sqrt{\tau}}{\sqrt{m}} \sum_{i=k}^{\ell(k)} \Xi_i \leq \begin{cases} \sqrt{1-\tau}\, \Xi_{k-1} & \text{if } k \in \mathcal{K}_1, \\ \left(\frac{1}{2} + \sqrt{1-\tau}\right) \sum_{i=k-k_1}^{k+k_1} \Xi_{i-1} + \varepsilon_k + \widehat{c}\,\Gamma_{k,k+m} & \text{if } k \in \mathcal{K}_2. \end{cases} \tag{3.2}$$

Proof. We will frequently use the following bound for every $k\in\mathbb{N}$:

$$\frac{1}{m}\sum_{i=k}^{\ell(k)}\Xi_i=\frac{1}{m}\sum_{i=k}^{\ell(k)}\sqrt{C_i-C_{i+1}}\le\sqrt{\sum_{i=k}^{\ell(k)}\frac{1}{m}(C_i-C_{i+1})}=\frac{1}{\sqrt{m}}\sqrt{C_k-C_{k+m}},\tag{3.3}$$

where the inequality is due to the concavity of the function $\mathbb{R}_+\ni t\mapsto\sqrt{t}$. Fix any $k\in\mathbb{N}$ with $x^k\in\mathbb{B}(\overline{x},\rho)$ and $\Phi(x^k)<\Phi(\overline{x})+\eta$. We proceed with the arguments by two cases.

Case 1: $k\in\mathcal{K}_1$. In this case, $\Phi(x^k)\le C_{k+m}$. By the equality in (2.1) and $\tau_k\in[\tau,1]$,

$$\begin{aligned}C_k-C_{k+m}&=(1-\tau_k)C_{k-1}+\tau_k\Phi(x^k)-C_{k+m}\le(1-\tau_k)C_{k-1}+(\tau_k-1)C_{k+m}\\&=(1-\tau_k)(C_{k-1}-C_{k+m})\le(1-\tau)(C_{k-1}-C_k+C_k-C_{k+m})\\&=(1-\tau)(\Xi_{k-1}^2+C_k-C_{k+m}),\end{aligned}$$

which implies that $\sqrt{\tau(C_k-C_{k+m})}\le\sqrt{1-\tau}\,\Xi_{k-1}$. Combining this inequality with the above (3.3), we obtain the inequality (3.2) for $k\in\mathcal{K}_1$.

Case 2: $k\in\mathcal{K}_2$. Now $\Phi(\overline{x})\le C_{k+m}<\Phi(x^k)<\Phi(\overline{x})+\eta$. Using (3.1) with $x=x^k$ yields that $\varphi'(\Phi(x^k)-\Phi(\overline{x}))\mathrm{dist}(0,\partial\Phi(x^k))\ge1$, which by condition H2 implies that

$$\left(\frac{1}{b_k}\sum_{i=k-k_1}^{k+k_1}\|x^i-x^{i-1}\|+\varepsilon_k\right)\varphi'\big(\Phi(x^k)-\Phi(\overline{x})\big)\ge1.\tag{3.4}$$

From the definition of $\Gamma_{k,k+m}$ and the concavity of φ on $[0,\eta)$, we have $\Gamma_{k,k+m}\ge\varphi'\big(\Phi(x^k)-\Phi(\overline{x})\big)\big(\Phi(x^k)-C_{k+m}\big)$. This along with (3.4) leads to the inequalities

$$\Phi(x^k)-C_{k+m}\le\left(\frac{1}{b_k}\sum_{i=k-k_1}^{k+k_1}\|x^i-x^{i-1}\|+\varepsilon_k\right)\Gamma_{k,k+m}\overset{(2.4)}{\le}\left(\frac{\overline{B}}{\sqrt{\tau}}\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}+\varepsilon_k\right)\Gamma_{k,k+m}.$$

In addition, by the equality in (2.1), $\Phi(x^k)-C_{k+m}=C_k-C_{k+m}-\frac{1-\tau_k}{\tau_k}(C_{k-1}-C_k)$. Along with the above inequality, the definition of $\overline{B}$ in (1.2), and $0<\tau_k\le1$, we have

$$\tau_k(C_k-C_{k+m})\le(1-\tau_k)(C_{k-1}-C_k)+\left(\frac{\overline{B}}{\sqrt{\tau}}\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}+\varepsilon_k\right)\Gamma_{k,k+m}.$$

By the definition of $\widehat{c}$, we have $\overline{B}/\sqrt{\tau}=2\widehat{c}-1$. With this in mind, recalling that $\Xi_{k-1}^2=C_{k-1}-C_k$, we continue to bound the term $\sqrt{\tau_k(C_k-C_{k+m})}$:

$$
\begin{aligned}
\sqrt{\tau_k(C_k-C_{k+m})} &\leq \sqrt{(1-\tau_k)\Xi_{k-1}^2+\left((2\widehat{c}-1)\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}+\varepsilon_k\right)\Gamma_{k,k+m}} \\
&\leq \sqrt{1-\tau_k}\Xi_{k-1}+\sqrt{(2\widehat{c}-1)\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}\Gamma_{k,k+m}}+\sqrt{\varepsilon_k\Gamma_{k,k+m}} \\
&= \sqrt{1-\tau_k}\Xi_{k-1}+2\sqrt{\left((\widehat{c}-1/2)\Gamma_{k,k+m}\right)\left(\frac{1}{2}\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}\right)} \\
&\quad +2\sqrt{(\varepsilon_k/2)(\Gamma_{k,k+m}/2)} \\
&\leq \sqrt{1-\tau_k}\Xi_{k-1}+\frac{1}{2}\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}+\frac{1}{2}\varepsilon_k+\widehat{c}\Gamma_{k,k+m} \\
&\leq \left(\frac{1}{2}+\sqrt{1-\tau}\right)\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}+\varepsilon_k+\widehat{c}\Gamma_{k,k+m},
\end{aligned}
$$

where the first inequality is by the definition of $\widehat{c}$, and the second is due to $\sqrt{\alpha+\beta}\leq\sqrt{\alpha}+\sqrt{\beta}$ for $\alpha,\beta\geq 0$. The penultimate inequality is due to twice using the fact $2\sqrt{\alpha\beta}\leq\alpha+\beta$ for $\alpha,\beta\geq 0$. Using (3.3) again yields the inequality (3.2) for $k\in\mathcal{K}_2$. □

Remark 3.2. To fully appreciate the innovative role that the sequence splitting in $\mathcal{K}_1$ and $\mathcal{K}_2$ plays in Lemma 3.1 and also in future results, we like to make a technical note to highlight its difference from the splitting scheme used in [19] for the GLL-type nonmonotone line-search. We recall the search criterion at x^k of GLL-type is

$$
\Phi(x^{k+1})+a\|x^{k+1}-x^k\|^2\leq\max_{j=[k-m_0]_+,\ldots,k}\Phi(x^j),
$$

where $m_0\geq 0$ is a given integer and $a>0$ is a constant. Let $\widetilde{\ell}(k)$ be the maximum index in $\arg\max_{j=[k-m_0]_+,\ldots,k}\Phi(x^j)$. Apparently, we have

$$
\Phi(x^{k+1})\leq\Phi(x^{\widetilde{\ell}(k)})-a\|x^{k+1}-x^k\|^2.
$$

The quality of $\Phi(x^{k+1})$ is further measured against the value of $\Phi(x^{\widetilde{\ell}(k+1)})$. Hence, we define

$$
\widetilde{\mathcal{K}}_1:=\left\{k\in\mathbb{N}\mid\Phi(x^{k+1})\leq\Phi(x^{\widetilde{\ell}(k+1)})-\frac{a}{2}\|x^{k+1}-x^k\|^2\right\}.
$$

This sequence was naturally suggested by the GLL-type search rule and was used in [19] to derive convergence properties on the sequence $\widetilde{\mathcal{K}}_1$. In contrast, the search in H1 uses a weighted average of functional values contained in C_k. There is no obvious way to refer to particular functional values $\{\Phi(x^k)\}$ as they are averaged. Instead, we collect in $\mathcal{K}_1$ all indices k with $\Phi(x^k)\leq C_{k+m}$. Note that $m\geq 2$ in our setting. This means that the value C_{k+m} contains iterate information up to x^{k+m} (the future iterates $x^{k+2},\ldots,x^{k+m}$ were involved). In contrast, $\widetilde{\mathcal{K}}_1$ only used the information up to x^{k+1}, which was already calculated at x^k. The future versus the present information being involved in the sequence splitting requires a completely different set of analysis to derive the respective convergence result. Further difference between our results and those from [19] is discussed in Remark 3.7.

THEOREM 3.3. *If Φ has the KL property at $\overline{x}$, then $\sum_{k=0}^{\infty}\|x^{k+1}-x^k\|<\infty$, and consequently the sequence $\{x^k\}_{k\in\mathbb{N}}$ converges to $\overline{x}$ that is a critical point of Φ.*

Proof. From $\lim_{k\to\infty}\Phi(x^k)=\Phi(\overline{x})$, there exists $\mathbb{N}\ni\widehat{k}>k_1$ such that for all $k\geq\widehat{k}$, $\Phi(x^k)<\Phi(\overline{x})+\eta$. Recall that $\lim_{k\to\infty}C_k=\Phi(\overline{x})$ and $\lim_{k\to\infty}(x^{k+1}-x^k)=0$. Along with condition H3, $\sum_{k=1}^{\infty}\varepsilon_k<\infty$, and the continuity of φ on $[0,\eta)$, if necessary by increasing $\widehat{k}$, we have

$$\|x^{\widehat{k}}-\overline{x}\|+\frac{4(2k_1+1)}{\sqrt{\underline{a}\tau}}\sum_{j=\widehat{k}-k_1}^{\ell(\widehat{k})}\Xi_{j-1}+\frac{2\widehat{c}}{\sqrt{\underline{a}\tau}}\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\varphi\big(C_j-\Phi(\overline{x})\big)+\frac{2}{\sqrt{\underline{a}\tau}}\sum_{j=\widehat{k}}^{\infty}\varepsilon_j<\rho, \tag{3.5}$$

where ρ is the same as in Lemma 3.1. Thus, by Lemma 3.1, inequality (3.2) holds for $k=\widehat{k}$. We claim that the desired conclusion holds if for each $\nu\geq\ell(\widehat{k})$,

$$\begin{cases} x^\nu\in\mathbb{B}(\overline{x},\rho),\\ \sqrt{\tau m}\sum\limits_{j=\ell(\widehat{k})}^{\nu}\Xi_j \leq\Big(\dfrac{1}{2}+\sqrt{1-\tau}\Big)(2k_1+1)\sum\limits_{j=\widehat{k}-k_1}^{\nu+1}\Xi_{j-1}+\sum\limits_{j=\widehat{k}}^{\nu}\varepsilon_j\\ \qquad\qquad+\widehat{c}\sum\limits_{j=\widehat{k}}^{\ell(\widehat{k})}\varphi\big(C_j-\Phi(\overline{x})\big).\end{cases} \tag{3.6}$$

Indeed, by the definition of m, $\sqrt{m\tau}\geq\frac{\sqrt{\tau}(m-k_1-1)}{\sqrt{m}}\geq(1+\sqrt{1-\tau})(2k_1+1)$, so that

$$\widehat{c}_1:=\sqrt{m\tau}-\Big(\frac{1}{2}+\sqrt{1-\tau}\Big)(2k_1+1)\geq\frac{2k_1+1}{2}\geq\frac{1}{2}.$$

According to this inequality, the inequality in (3.6) implies that

$$\widehat{c}_1\sum_{j=\ell(\widehat{k})}^{\nu}\Xi_j\leq\frac{(1+2\sqrt{1-\tau})(2k_1+1)}{2}\sum_{j=\widehat{k}-k_1}^{\ell(\widehat{k})}\Xi_{j-1}+\widehat{c}\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\varphi\big(C_j-\Phi(\overline{x})\big)+\sum_{j=\widehat{k}}^{\nu}\varepsilon_j. \tag{3.7}$$

Passing the limit $\nu\to\infty$ to both sides of (3.7) leads to $\sum_{k=1}^{\infty}\Xi_k<\infty$, which by (2.5) implies that $\sum_{k=0}^{\infty}\|x^{k+1}-x^k\|<\infty$. By condition H3, $\{x^k\}_{k\in\mathbb{N}}$ converges to $\overline{x}$.

Next we prove by induction that the claim (3.6) holds for every $\nu\geq\ell(\widehat{k})$. Indeed, the above (3.5) implies that $\|x^{\widehat{k}}-\overline{x}\|<\rho$. For any $\widehat{k}+1\leq\nu\leq\ell(\widehat{k})$, using $\|x^\nu-\overline{x}\|\leq\|x^{\widehat{k}}-\overline{x}\|+\|x^\nu-x^{\widehat{k}}\|$ and the previous (2.5) and (3.5) leads to

$$\|x^\nu-\overline{x}\|\leq\|x^{\widehat{k}}-\overline{x}\|+\sum_{j=\widehat{k}+1}^{\nu}\|x^j-x^{j-1}\|\overset{(2.5)}{\leq}\|x^{\widehat{k}}-\overline{x}\|+\frac{1}{\sqrt{\underline{a}\tau}}\sum_{j=\widehat{k}+1}^{\nu}\Xi_{j-1}\overset{(3.5)}{<}\rho.$$

Thus, we have $x^\nu\in\mathbb{B}(\overline{x},\rho)$ for $\widehat{k}\leq\nu\leq\ell(\widehat{k})$. Together with $\Phi(x^k)<\Phi(\overline{x})+\eta$ for all $k\geq\widehat{k}$, invoking Lemma 3.1 shows that inequality (3.2) holds for $x^{\widehat{k}},\ldots,x^{\ell(\widehat{k})}$. Summing inequality (3.2) from $\widehat{k}$ to $\ell(\widehat{k})$ and using the nonnegativity of φ leads to

$$\begin{aligned}\frac{\sqrt{\tau}}{\sqrt{m}}\underbrace{\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\sum_{i=j}^{\ell(j)}\Xi_i}_{:=\Delta_1}&\leq\Big(\frac{1}{2}+\sqrt{1-\tau}\Big)\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\sum_{i=j-k_1}^{j+k_1}\Xi_{i-1}+\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\varepsilon_j+\widehat{c}\sum_{\mathcal{K}_2\ni j=\widehat{k}}^{\ell(\widehat{k})}\varphi\big(\Phi(x^j)-\Phi(\overline{x})\big)\\ &\leq\Big(\frac{1}{2}+\sqrt{1-\tau}\Big)\underbrace{\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\sum_{i=j-k_1}^{j+k_1}\Xi_{i-1}}_{:=\Delta_2}+\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\varepsilon_j+\widehat{c}\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\varphi\big(C_j-\Phi(\overline{x})\big),\end{aligned} \tag{3.8}$$

where the second inequality is due to $\Phi(x^j)\le C_j$ by Lemma 2.3(i) and the increasing of φ. Both Δ_1 and Δ_2 contain many terms. We simplify them. Recall that $m>k_1+1$ by the definition of m and $\ell(\widehat{k})=\widehat{k}+m-1$. For Δ_1, there are at least as many as m terms of $\Xi_{\ell(\widehat{k})}$. For each $j\in\{\widehat{k},\widehat{k}+1,\ldots,\ell(\widehat{k})\}$, we consider the terms in the sum $\sum_{i=j}^{\ell(j)}\Xi_i$ that have indices higher than $\ell(\widehat{k})$. Note that $\ell(j)\ge\ell(\widehat{k})+k_1+1$ for $j=\widehat{k}+k_1+1,\ldots,\widehat{k}+m-1$. We have $\sum_{i=j}^{\ell(j)}\Xi_i\ge\sum_{i=\widehat{k}+m}^{\ell(\widehat{k})+k_1+1}\Xi_i$, and consequently,

$$\Delta_1\ge m\Xi_{\ell(\widehat{k})}+(m-k_1-1)\sum_{i=\widehat{k}+m}^{\ell(\widehat{k})+k_1+1}\Xi_i.$$

For Δ_2, we consider the lowest and the highest index for i and they are $i=\widehat{k}-k_1$ and $i=\ell(\widehat{k})+k_1$. Therefore, it holds that

$$\Delta_2\le(2k_1+1)\sum_{i=\widehat{k}-k_1}^{\ell(\widehat{k})+k_1}\Xi_{i-1}.$$

Substitute those bounds back to (3.8) and obtain the following inequalities:

$$\begin{aligned}
&\sqrt{\tau m}\,\Xi_{\ell(\widehat{k})}+\frac{\sqrt{\tau}(m-k_1-1)}{\sqrt{m}}\sum_{j=\widehat{k}+m}^{\ell(\widehat{k})+k_1+1}\Xi_j\\
&\quad\le\left(\frac{1}{2}+\sqrt{1-\tau}\right)(2k_1+1)\sum_{j=\widehat{k}-k_1}^{\ell(\widehat{k})+k_1}\Xi_{j-1}+\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\varepsilon_j+\widehat{c}\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\varphi\big(C_j-\Phi(\overline{x})\big)\\
&\quad\le\left(\frac{1}{2}+\sqrt{1-\tau}\right)(2k_1+1)\left[\sum_{j=\widehat{k}-k_1}^{\ell(\widehat{k})+1}\Xi_{j-1}+\sum_{j=\ell(\widehat{k})+2}^{\ell(\widehat{k})+k_1+2}\Xi_{j-1}\right]\\
&\qquad+\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\varepsilon_j+\widehat{c}\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\varphi\big(C_j-\Phi(\overline{x})\big).
\end{aligned}\tag{3.9}$$

Recall that $\ell(\widehat{k})+1=\widehat{k}+m$ and $\frac{\sqrt{\tau}(m-k_1-1)}{\sqrt{m}}\ge(1+\sqrt{1-\tau})(2k_1+1)$. The above inequality implies that the inequality in (3.6) holds for $\nu=\ell(\widehat{k})$, so the claimed (3.6) holds for $\nu=\ell(\widehat{k})$. Now assume that the claimed (3.6) holds for some $\nu\ge\ell(\widehat{k})$. We argue that it holds for $\nu+1$. From the triangle inequality and (2.5), it holds that

$$\|x^{\nu+1}-\overline{x}\|\le\|x^{\widehat{k}}-\overline{x}\|+\sum_{j=\widehat{k}}^{\nu}\|x^{j+1}-x^j\|\le\|x^{\widehat{k}}-\overline{x}\|+\frac{1}{\sqrt{\underline{a}\tau}}\left[\sum_{j=\widehat{k}}^{\ell(\widehat{k})-1}\Xi_j+\sum_{j=\ell(\widehat{k})}^{\nu}\Xi_j\right].$$

Note that (3.7) holds for this ν. Together with the above inequality, it follows that

$$\begin{aligned}
\|x^{\nu+1}-\overline{x}\|&\le\|x^{\widehat{k}}-\overline{x}\|+\frac{1}{\sqrt{\underline{a}\tau}}\sum_{j=\widehat{k}}^{\ell(\widehat{k})-1}\Xi_j+\frac{(1+2\sqrt{1-\tau})(2k_1+1)}{\sqrt{\underline{a}\tau}}\sum_{j=\widehat{k}-k_1}^{\ell(\widehat{k})}\Xi_{j-1}\\
&\quad+\frac{2}{\sqrt{\underline{a}\tau}}\sum_{j=\widehat{k}}^{\nu}\varepsilon_j+\frac{2\widehat{c}}{\sqrt{\underline{a}\tau}}\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\varphi\big(C_j-\Phi(\overline{x})\big)\quad\text{(also used }\widehat{c}_1\ge 1/2\text{)}.
\end{aligned}$$

Recall that $\tau \in (0,1]$, so $1+2\sqrt{1-\tau} < 3$. Then, we have

$$\|x^{\nu+1} - \overline{x}\| \le \|x^{\widehat{k}} - \overline{x}\| + \frac{4(2k_1+1)}{\sqrt{\underline{a}\tau}} \sum_{j=\widehat{k}-k_1}^{\ell(\widehat{k})} \Xi_{j-1} + \frac{2}{\sqrt{\underline{a}\tau}} \sum_{j=\widehat{k}}^{\nu} \varepsilon_j + \frac{2\widehat{c}}{\sqrt{\underline{a}\tau}} \sum_{j=\widehat{k}}^{\ell(\widehat{k})} \varphi\big(C_j - \Phi(\overline{x})\big) \overset{(3.5)}{<} \rho.$$

This shows $x^{\nu+1} \in \mathbb{B}(\overline{x}, \rho)$, so the rest only proves the inequality in (3.6) for $\nu+1$. Note that Lemma 3.1 holds for $x^{\widehat{k}}$ up to $x^{\nu+1}$. Summing (3.2) from $\widehat{k}$ to $\nu+1$ yields

$$\frac{\sqrt{\tau}}{\sqrt{m}} \sum_{j=\widehat{k}}^{\nu+1} \sum_{i=j}^{\ell(j)} \Xi_i \le \Big(\frac{1}{2} + \sqrt{1-\tau}\Big) \sum_{j=\widehat{k}}^{\nu+1} \sum_{i=j-k_1}^{j+k_1} \Xi_{i-1} + \sum_{j=\widehat{k}}^{\nu+1} \varepsilon_j + \widehat{c} \sum_{\mathcal{K}_2 \ni j=\widehat{k}}^{\nu+1} \big[\varphi\big(\Phi(x^j) - \Phi(\overline{x})\big) - \varphi\big(C_{j+m} - \Phi(\overline{x})\big)\big].$$

As $\Phi(x^j) \le C_j$ for each $j \in \mathbb{N}$ by Lemma 2.3(i), the increasing and nonnegativity of φ implies that

$$\sum_{\mathcal{K}_2 \ni j=\widehat{k}}^{\nu+1} \big[\varphi\big(\Phi(x^j) - \Phi(\overline{x})\big) - \varphi\big(C_{j+m} - \Phi(\overline{x})\big)\big] \le \sum_{j=\widehat{k}}^{\ell(\widehat{k})} \varphi(C_j - \Phi(\overline{x})).$$

From the above two inequalities, it immediately follows that

$$\frac{\sqrt{\tau}}{\sqrt{m}} \underbrace{\sum_{j=\widehat{k}}^{\nu+1} \sum_{i=j}^{\ell(j)} \Xi_i}_{:=\Delta_3} \le \Big(\frac{1}{2} + \sqrt{1-\tau}\Big) \underbrace{\sum_{j=\widehat{k}}^{\nu+1} \sum_{i=j-k_1}^{j+k_1} \Xi_{i-1}}_{:=\Delta_4} + \sum_{j=\widehat{k}}^{\nu+1} \varepsilon_j + \widehat{c} \sum_{j=\widehat{k}}^{\ell(\widehat{k})} \varphi(C_j - \Phi(\overline{x})). \tag{3.10}$$

For Δ_3, there are at least as many as m terms of $\sum_{i=\ell(\widehat{k})}^{\nu+1} \Xi_i$. For each $j \in \{\widehat{k}, \ldots, \nu+1\}$, we consider the terms in the sum $\sum_{i=j}^{\ell(j)} \Xi_i$ that have indices higher than $\nu+1$. Note that $\ell(j) \ge \nu+k_1+2$ for $j = \nu+k_1-m+3, \ldots, \nu+1$. We have $\sum_{i=j}^{\ell(j)} \Xi_i \ge \sum_{i=\nu+2}^{\nu+k_1+2} \Xi_i$, and consequently,

$$\Delta_3 \ge m \sum_{i=\ell(\widehat{k})}^{\nu+1} \Xi_i + (m - k_1 - 1) \sum_{i=\nu+2}^{\nu+k_1+2} \Xi_i.$$

For Δ_4, we consider the lowest and the highest index for i and they are $i = \widehat{k} - k_1$ and $i = \nu + k_1 + 1$. Therefore, it holds that

$$\Delta_4 \le (2k_1 + 1) \sum_{i=\widehat{k}-k_1}^{\nu+k_1+1} \Xi_{i-1}.$$

Substitute those bounds back to (3.10) and get the following:

$$\sqrt{\tau m}\sum_{j=\ell(\widehat{k})}^{\nu+1}\Xi_j+\frac{\sqrt{\tau}(m-k_1-1)}{\sqrt{m}}\sum_{j=\nu+2}^{\nu+k_1+2}\Xi_j$$
$$\leq\Big(\frac{1}{2}+\sqrt{1-\tau}\Big)(2k_1+1)\sum_{j=\widehat{k}-k_1}^{\nu+k_1+1}\Xi_{j-1}+\sum_{j=\widehat{k}}^{\nu+1}\varepsilon_j+\widehat{c}\sum_{j=\widehat{k}}^{\ell(\widehat{k})}\varphi(C_j-\Phi(\overline{x})).$$

This, along with $\frac{\sqrt{\tau}(m-k_1-1)}{\sqrt{m}}\geq(1+\sqrt{1-\tau})(2k_1+1)$, implies that the inequality in (3.6) holds for $\nu+1$. Thus, we complete the induction proof. ∎

Remark 3.4. By the convergence analysis of this section, when H2 is relaxed to H2′. there exist $k_1\in\mathbb{N}$ and $\overline{\delta}>0$ such that for all $k\geq k_1$ and $x^k\in\mathbb{B}(\overline{x},\overline{\delta})$, $\mathrm{dist}(0,\partial\Phi(x^k))\leq\frac{1}{b_k}\sum_{i=k-k_1}^{k+k_1}\|x^i-x^{i-1}\|+\varepsilon_k$ with $b_k>0$ and $\varepsilon_k\geq 0$, any sequence $\{x^k\}_{k\in\mathbb{N}}$ satisfying H1, H2′, and H3 still has the full convergence.

3.2. Convergence rate. Next we establish the convergence rate of $\{x^k\}_{k\in\mathbb{N}}$ under the KL property of Φ with exponent $\theta\in(0,1)$ by Theorem 3.3 and Lemma 2.2. We will require the relative error ε_k in condition H2 to be eventually controlled by a sequence related to the exponent. This requirement is certainly satisfied if we simply set $\varepsilon_k=0$. For this purpose, we first establish the convergence rate of $\{C_k\}_{k\in\mathbb{N}}$.

LEMMA 3.5. *Suppose that Φ has the KL property of exponent $\theta\in(0,1)$ at $\overline{x}$. If there exist $\widetilde{k}_0\in\mathbb{N},\widetilde{\gamma}>0$, and $\widetilde{\varrho}\in(0,1)$ such that for all $k\geq\widetilde{k}_0$,*

$$\varepsilon_k\leq\begin{cases}\widetilde{\gamma}\widetilde{\varrho}^k & \text{if}\,\theta\in(0,1/2],\\ \widetilde{\gamma}k^{\frac{\theta}{1-2\theta}} & \text{if}\,\theta\in(1/2,1),\end{cases}\tag{3.11}$$

then there exist $\gamma>0$ and $\varrho\in(0,1)$ such that for sufficiently large k,

$$\widetilde{\Lambda}_k:=C_k-\Phi(\overline{x})\leq\begin{cases}\gamma\varrho^k & \text{if}\,\theta\in(0,1/2],\\ \gamma k^{\frac{1}{1-2\theta}} & \text{if}\,\theta\in(1/2,1).\end{cases}\tag{3.12}$$

Proof. Since Φ has the KL property of exponent $\theta\in(0,1)$ at $\overline{x}$, for any $x\in\mathbb{B}(\overline{x},\delta)\cap[\Phi(\overline{x})<\Phi<\Phi(\overline{x})+\eta]$, the inequality (3.1) holds with $\varphi(t)=ct^{1-\theta}$ for some $c>0$. Along with $\lim_{k\to\infty}x^k=\overline{x}$ by Theorem 3.3, $\lim_{k\to\infty}\Phi(x^k)=\Phi(\overline{x})$, and condition H2, there is $\widetilde{k}\geq\max\{\ell(\widehat{k}),\widetilde{k}_0\}$ such that for all $k\geq\widetilde{k}$ with $\Phi(x^k)>\Phi(\overline{x})$,

$$(\Phi(x^k)-\Phi(\overline{x}))^{\theta}\leq c(1-\theta)\Bigg(\frac{1}{b_k}\sum_{i=k-k_1}^{k+k_1}\|x^i-x^{i-1}\|+\varepsilon_k\Bigg)\overset{(2.4)}{\leq}c(1-\theta)\Bigg(\frac{\overline{B}}{\sqrt{\tau}}\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}+\varepsilon_k\Bigg)$$
$$\leq c(1-\theta)\max\big\{1,(2\widehat{c}-1)\big\}\Bigg[\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}+\varepsilon_k\Bigg],\tag{3.13}$$

where the third inequality is by the definition of $\widehat{c}$ appearing in Lemma 3.1. By letting $\widehat{c}_2(\theta):=(c(1-\theta)\max\{1,(2\widehat{c}-1)\})^{\frac{1}{\theta}}$, for all $k\geq\widetilde{k}$ with $\Phi(x^k)>\Phi(\overline{x})$, we have

$$\Phi(x^k)-\Phi(\overline{x})\leq\widehat{c}_2(\theta)\Bigg[\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}+\varepsilon_k\Bigg]^{\frac{1}{\theta}},$$

which, together with the equality in (2.1) and $\tau_k \geq \tau$, implies that

$$\begin{aligned}\widetilde{\Lambda}_k = C_k - \Phi(\overline{x}) &= (1-\tau_k)(C_{k-1} - \Phi(\overline{x})) + \tau_k(\Phi(x^k) - \Phi(\overline{x}))\\ &\leq (1-\tau)\widetilde{\Lambda}_{k-1} + \tau\widehat{c}_2(\theta)\left[\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1} + \varepsilon_k\right]^{\frac{1}{\theta}},\end{aligned} \tag{3.14}$$

while for all $k \geq \widetilde{k}$ with $\Phi(x^k) \leq \Phi(\overline{x})$, from the equality in (2.1) and $\tau_k \geq \tau$, we have

$$\widetilde{\Lambda}_k = (1-\tau_k)(C_{k-1} - \Phi(\overline{x})) + \tau_k(\Phi(x^k) - \Phi(\overline{x})) \leq (1-\tau)\widetilde{\Lambda}_{k-1}.$$

Thus, the inequality (3.14) holds for all $k \geq \widetilde{k}$. We proceed with the proof by two cases.

Case 1*:* $\theta \in (0, \frac{1}{2}]$. Note that $\lim_{k\to\infty}\sum_{i=k-k_1}^{k+k_1}\Xi_i = 0$ and $\lim_{k\to\infty}\varepsilon_k = 0$. Hence, $\sum_{i=k-k_1}^{k+k_1}\Xi_i + \varepsilon_k < 0.5$ for all $k \geq \widetilde{k}$ (if necessary by increasing $\widetilde{k}$). For any $k \geq \widetilde{k}$, using the above inequality (3.14), the convexity of $\mathbb{R} \ni t \mapsto t^2$ and $\varepsilon_k \leq \widetilde{\gamma}\widetilde{\varrho}^k$ leads to

$$\begin{aligned}\widetilde{\Lambda}_{k+k_1} \leq \widetilde{\Lambda}_k &\leq (1-\tau)\widetilde{\Lambda}_{k-1} + \tau\widehat{c}_2(\theta)\left[\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1} + \varepsilon_k\right]^2\\ &\leq (1-\tau)\widetilde{\Lambda}_{k-1} + \tau(2k_1+2)\widehat{c}_2(\theta)\left[\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}^2 + \varepsilon_k^2\right]\\ &\leq (1-\tau)\widetilde{\Lambda}_{k-k_1-1} + \tau(2k_1+2)\widehat{c}_2(\theta)\left[\widetilde{\Lambda}_{k-k_1-1} - \widetilde{\Lambda}_{k+k_1} + \widetilde{\gamma}^2\widetilde{\varrho}^{2k}\right],\end{aligned}$$

which, by setting $\varrho_1 := \frac{1-\tau+\tau(2k_1+2)\widehat{c}_2(\theta)}{1+\tau(2k_1+2)\widehat{c}_2(\theta)}$, implies that

$$\widetilde{\Lambda}_{k+k_1} \leq \varrho_1\left(\widetilde{\Lambda}_{k-k_1-1} + \widetilde{\gamma}^2\widetilde{\varrho}^{2k}\right) = \varrho_1\Lambda_{k-k_1-1} + \varrho_1\widetilde{\gamma}^2\widetilde{\varrho}^{2k}. \tag{3.15}$$

Consequently, for any $k \geq \widetilde{k} + k_1$, it holds that

$$\widetilde{\Lambda}_k \leq \varrho_1\widetilde{\Lambda}_{k-2k_1-1} + \varrho_1\widetilde{\gamma}^2\widetilde{\varrho}^{2k-2k_1} \leq \varrho_1\widetilde{\Lambda}_{k-2k_1-1} + \varrho_1\widetilde{\gamma}^2\widetilde{\varrho}^{k-2k_1-1},$$

where the second inequality is due to $\widetilde{\varrho} < 1$. Using the above recursive relation and letting $\overline{k} := \lfloor\frac{k-\widetilde{k}-k_1}{2k_1+1}\rfloor$ and $\beta := \varrho_1\widetilde{\gamma}^2$, we have

$$\widetilde{\Lambda}_k \leq \varrho_1^{\overline{k}}\widetilde{\Lambda}_{k-\overline{k}(2k_1+1)} + \beta\widetilde{\varrho}^{k-2k_1-1}\left[1 + \frac{\varrho_1}{\widetilde{\varrho}^{2k_1+1}} + \cdots + \left(\frac{\varrho_1}{\widetilde{\varrho}^{2k_1+1}}\right)^{\overline{k}-1}\right].$$

After a simple calculation for $\frac{\varrho_1}{\widetilde{\varrho}^{2k_1+1}} > 1$, $\frac{\varrho_1}{\widetilde{\varrho}^{2k_1+1}} = 1$, and $\frac{\varrho_1}{\widetilde{\varrho}^{2k_1+1}} < 1$, respectively, there exist $\gamma > 0$ and $\varrho \in (0,1)$ such that $\widetilde{\Lambda}_k \leq \gamma\varrho^k$ for sufficiently large k.

Case 2*:* $\theta \in (\frac{1}{2}, 1)$. Since $\sum_{i=k-k_1}^{k+k_1}\Xi_i^2 + \varepsilon_k^2 < 0.5$ for all $k \geq \widetilde{k}$ (if necessary by increasing $\widetilde{k}$), invoking inequality (3.14) yields that for any $k \geq \widetilde{k}$,

$$\begin{aligned}\widetilde{\Lambda}_{k+k_1} \leq \widetilde{\Lambda}_k &\leq (1-\tau)\widetilde{\Lambda}_{k-1} + \tau(2k_1+2)^{\frac{1}{2\theta}}\widehat{c}_2(\theta)\left[\sum_{i=k-k_1}^{k+k_1}\Xi_{i-1}^2 + \varepsilon_k^2\right]^{\frac{1}{2\theta}}\\ &\leq (1-\tau)\widetilde{\Lambda}_{k-k_1-1} + \tau(2k_1+2)^{\frac{1}{2\theta}}\widehat{c}_2(\theta)\left[\widetilde{\Lambda}_{k-k_1-1} - \widetilde{\Lambda}_{k+k_1} + \varepsilon_k^2\right]^{\frac{1}{2\theta}}.\end{aligned}$$

By dividing the above inequality by τ, after a suitable rearrangement, for any $k\geq\widetilde{k}$,

$$\begin{aligned}\widetilde{\Lambda}_{k+k_1}&\leq\frac{(1-\tau)}{\tau}[\widetilde{\Lambda}_{k-k_1-1}-\widetilde{\Lambda}_{k+k_1}]+(2k_1+2)^{\frac{1}{2\theta}}\widehat{c}_2(\theta)\Big[\widetilde{\Lambda}_{k-k_1-1}-\widetilde{\Lambda}_{k+k_1}+\varepsilon_k^2\Big]^{\frac{1}{2\theta}}\\&\leq\tau^{-1}[1-\tau+\tau(2k_1+2)^{\frac{1}{2\theta}}\widehat{c}_2(\theta)]\Big((\widetilde{\Lambda}_{k-k_1-1}-\widetilde{\Lambda}_{k+k_1})+\widetilde{\gamma}^2k^{\frac{2\theta}{1-2\theta}}\Big)^{\frac{1}{2\theta}}.\end{aligned}$$

Note that $\frac{k}{k+k_1}\geq 0.5$ for all $k\geq\widetilde{k}$ (if necessary by increasing $\widetilde{k}$). Then, for any $k\geq\widetilde{k}$,

$$\begin{aligned}\widetilde{\Lambda}_{k+k_1}&\leq\tau^{-1}[1-\tau+\tau(2k_1+2)^{\frac{1}{2\theta}}\widehat{c}_2(\theta)]\Big((\widetilde{\Lambda}_{k-k_1-1}-\widetilde{\Lambda}_{k+k_1})+2^{\frac{2\theta}{2\theta-1}}\widetilde{\gamma}^2(k+k_1)^{\frac{2\theta}{1-2\theta}}\Big)^{\frac{1}{2\theta}}\\&\leq M_1\max\big\{(\widetilde{\Lambda}_{k-k_1-1}-\widetilde{\Lambda}_{k+k_1})^{\frac{1}{2\theta}},(k+k_1)^{\frac{1}{1-2\theta}}\big\}\end{aligned}$$

with $M_1:=\tau^{-1}(1+2^{\frac{1}{2\theta-1}}\widetilde{\gamma}^{\frac{1}{\theta}})[1-\tau+\tau(2k_1+2)^{\frac{1}{2\theta}}\widehat{c}_2(\theta)]$. Thus, for all $k\geq\widetilde{k}+k_1$, we have $\widetilde{\Lambda}_k\leq M_1\max\big\{(\widetilde{\Lambda}_{k-2k_1-1}-\widetilde{\Lambda}_k)^{\frac{1}{2\theta}},k^{\frac{1}{1-2\theta}}\big\}$. Invoking Lemma 2.2 leads to the result. ◻

Now we are in a position to achieve the main conclusion of this section.

THEOREM 3.6. *Suppose that Φ has the KL property of exponent $\theta\in(0,1)$ at $\overline{x}$. If there exist $\widetilde{k}_0\in\mathbb{N},\widetilde{\gamma}>0$ and $\widetilde{\varrho}\in(0,1)$ such that inequality (3.11) holds for all $k\geq\widetilde{k}_0$, then there exist $\gamma>0$ and $\varrho\in(0,1)$ such that for sufficiently large k,*

$$\|x^k-\overline{x}\|\leq\Delta_k:=\sum_{j=k}^{\infty}\|x^{j+1}-x^j\|\leq\begin{cases}\gamma\varrho^k & \text{if}\,\theta\in(0,1/2],\\ \gamma k^{\frac{1-\theta}{1-2\theta}} & \text{if}\,\theta\in(1/2,1).\end{cases}\tag{3.16}$$

Proof. By the definition of Δ_k and the triangle inequality, it holds that $\|x^k-\overline{x}\|\leq\Delta_k$. For each $k\in\mathbb{N}$, write $\Lambda_k:=\sum_{j=k}^{\infty}\Xi_j$. From (2.5) and $\ell(k)=k+m-1$, to prove the second inequality in (3.16), it suffices to argue the existence of $\gamma>0$ and $\varrho\in(0,1)$ such that for sufficiently large k,

$$\Lambda_{\ell(k)}\leq\begin{cases}\gamma\varrho^k & \text{if}\,\theta\in(0,1/2],\\ \gamma k^{\frac{1-\theta}{1-2\theta}} & \text{if}\,\theta\in(1/2,1).\end{cases}\tag{3.17}$$

Let $\widetilde{k}$ be the same as before. Invoking the previous (3.7) with $\widehat{k}$ replaced by any $k\geq\widetilde{k}$ and $\varphi(t)=ct^{1-\theta}$ for some $c>0$ and passing the limit $\nu\to\infty$ results in

$$\begin{aligned}\Lambda_{\ell(k)}=\sum_{j=\ell(k)}^{\infty}\Xi_j&\leq\frac{(1+2\sqrt{1-\tau})(2k_1+1)}{2\widehat{c}_1}\sum_{j=k-k_1}^{\ell(k)}\Xi_{j-1}+\frac{c\widehat{c}}{\widehat{c}_1}\sum_{j=k}^{\ell(k)}\big(C_j-\Phi(\overline{x})\big)^{1-\theta}+\frac{1}{\widehat{c}_1}\sum_{j=k}^{\infty}\varepsilon_j\\&\leq\frac{(1+2\sqrt{1-\tau})(2k_1+1)}{2\widehat{c}_1}\sum_{j=k-k_1}^{\ell(k)}\widetilde{\Lambda}_{j-1}^{\frac{1}{2}}+\frac{c\widehat{c}}{\widehat{c}_1}\sum_{j=k}^{\ell(k)}\widetilde{\Lambda}_j^{1-\theta}+\frac{1}{\widehat{c}_1}\sum_{j=k}^{\infty}\varepsilon_j\\&\leq\widetilde{c}_1\widetilde{\Lambda}_{k-k_1-1}^{\frac{1}{2}}+\widetilde{c}_2\widetilde{\Lambda}_k^{1-\theta}+\frac{1}{\widehat{c}_1}\sum_{j=k}^{\infty}\varepsilon_j\end{aligned}\tag{3.18}$$

with $\widetilde{c}_1:=\frac{(1+2\sqrt{1-\tau})(2k_1+1)(m+k_1)}{2\widehat{c}_1}$ and $\widetilde{c}_2:=\frac{c\widehat{c}m}{\widehat{c}_1}$, where the third inequality is due to the nonincreasing of the sequence $\{C_k\}_{k\in\mathbb{N}}$. We proceed with the proof by two cases.

Case 1: $\theta\in(0,\frac{1}{2}]$. Note that $\lim_{k\to\infty}\widetilde{\Lambda}_k=0$, so $\widetilde{\Lambda}_{k-k_1-1}<0.5$ for all $k\geq\widetilde{k}$ (if necessary by increasing $\widetilde{k}$). From (3.11), for any $k\geq\widetilde{k}$, $\sum_{j=k}^{\infty}\varepsilon_j\leq\widetilde{\gamma}\sum_{j=k}^{\infty}\widetilde{\varrho}^k\leq\widetilde{\gamma}\frac{\widetilde{\varrho}^k}{1-\widetilde{\varrho}}$. Thus, together with $1-\theta\geq 1/2$ and inequality (3.18), for any $k\geq\widetilde{k}$, it holds that

$$\Lambda_{\ell(k)}\leq[\widetilde{c}_1+\widetilde{c}_2+\widehat{c}_1^{-1}]\Big[\widetilde{\Lambda}_{k-k_1-1}^{\frac{1}{2}}+\frac{\widetilde{\gamma}\widetilde{\varrho}^k}{1-\widetilde{\varrho}}\Big].$$

By Lemma 3.5, there exist $\widehat{\gamma}>0$ and $\widehat{\varrho}\in(0,1)$ such that $\widetilde{\Lambda}_k\le\widehat{\gamma}\widehat{\varrho}^k$ for sufficiently large k. Then, for all $k\ge\widetilde{k}$ (if necessary by increasing $\widetilde{k}$),

$$\Lambda_{\ell(k)}\le[\widetilde{c}_1+\widetilde{c}_2+\widehat{c}_1^{-1}]\Big(\widehat{\gamma}^{\frac{1}{2}}\widehat{\varrho}^{\frac{k-k_1-1}{2}}+\frac{\widetilde{\gamma}\widetilde{\varrho}^k}{1-\widetilde{\varrho}}\Big).$$

Set $\widetilde{c}_3:=(\widetilde{c}_1+\widetilde{c}_2+\widehat{c}_1^{-1})(\gamma^{\frac{1}{2}}\widehat{\varrho}^{-\frac{k_1+1}{2}}+\frac{\widetilde{\gamma}}{1-\widetilde{\varrho}})$ and $\varrho_1:=\max\{\widehat{\varrho}^{\frac{1}{2}},\widetilde{\varrho}\}$. The above inequality implies $\Lambda_{\ell(k)}\le\widetilde{c}_3\varrho_1^k$ for all $k\ge\widetilde{k}$.

Case 2*:* $\theta\in(\frac{1}{2},1)$. Since $\widetilde{\Lambda}_{k-k_1-1}<0.5$ for all $k\ge\widetilde{k}$ (if necessary by increasing $\widetilde{k}$), from inequality (3.18), it follows that for any $k\ge\widetilde{k}$,

$$\Lambda_{\ell(k)}\le[\widetilde{c}_1+\widetilde{c}_2+\widehat{c}_1^{-1}]\Bigg[\widetilde{\Lambda}_{k-k_1-1}^{1-\theta}+\sum_{j=k}^{\infty}\varepsilon_j\Bigg]. \tag{3.19}$$

By Lemma 3.5, there exists $\widehat{\gamma}>0$ such that $\widetilde{\Lambda}_k\le\widehat{\gamma}k^{\frac{1}{1-2\theta}}$ for sufficiently large k, so $\widetilde{\Lambda}_{k-k_1-1}\le\widehat{\gamma}(k-k_1-1)^{\frac{1}{1-2\theta}}$ for all $k\ge\widetilde{k}$ (if necessary by enlarging $\widetilde{k}$). Note that $\frac{k-k_1-1}{k}>0.5$ for any $k\ge\widetilde{k}$ (if necessary by enlarging $\widetilde{k}$). Then, for any $k\ge\widetilde{k}$, $\widetilde{\Lambda}_{k-k_1-1}\le 2^{\frac{1}{2\theta-1}}\widehat{\gamma}k^{\frac{1}{1-2\theta}}$. In addition, from $\varepsilon_k\le\widetilde{\gamma}k^{\frac{\theta}{1-2\theta}}$, it is not hard to get

$$\sum_{j=k}^{\infty}\varepsilon_j\le\widetilde{\gamma}\sum_{j=k}^{\infty}k^{\frac{\theta}{1-2\theta}}\le\widetilde{\gamma}\int_k^{\infty}t^{\frac{\theta}{1-2\theta}}dt=\frac{\widetilde{\gamma}(2\theta-1)}{1-\theta}k^{\frac{1-\theta}{1-2\theta}}.$$

Together with $\widetilde{\Lambda}_{k-k_1-1}\le 2^{\frac{1}{2\theta-1}}\widehat{\gamma}k^{\frac{1}{1-2\theta}}$ and inequality (3.19), for any $k\ge\widetilde{k}$, it holds that

$$\Lambda_{\ell(k)}\le[\widetilde{c}_1+\widetilde{c}_2+\widehat{c}_1^{-1}]\Big(2^{\frac{1-\theta}{2\theta-1}}\widehat{\gamma}^{1-\theta}k^{\frac{1-\theta}{1-2\theta}}+\frac{\widetilde{\gamma}(2\theta-1)}{1-\theta}k^{\frac{1-\theta}{1-2\theta}}\Big).$$

Set $\widetilde{c}_4:=(\widetilde{c}_1+\widetilde{c}_2+\widehat{c}_1^{-1})(2^{\frac{1-\theta}{2\theta-1}}\widehat{\gamma}^{1-\theta}+\frac{\widetilde{\gamma}(2\theta-1)}{1-\theta})$. We get $\Lambda_{\ell(k)}\le\widetilde{c}_4k^{\frac{1-\theta}{1-2\theta}}$ for all $k\ge\widetilde{k}$. ◻

Remark 3.7. Theorems 3.3 and 3.6 extend the convergence results obtained in [2, 4, 9] for a monotone iterative framework to a nonmonotone case. Compared with the nonmonotone iterative framework proposed in [19] by a GLL-type nonmonotone line-search procedure, the ZH-type iterative framework has better convergence, i.e., the full convergence (local convergence rate) of its iterate sequence does not require additional restriction on Φ except its KL property (KL property of exponent $\theta\in(0,1)$). In contrast, to obtain the global convergence, the paper [19] requires an assumption on the growth of an objective value subsequence, which is shown to hold automatically if Φ is also ρ-weakly convex on a neighborhood of the set of critical points. We also note that the favorable properties of the sequence $\{C^k\}_{k\in\mathbb{N}}$ plays a crucial role in the convergence proof of the iterate sequence $\{x^k\}_{k\in\mathbb{N}}$.

To close this section, we would like to make an important point that the proofs of this section still hold if we modify condition H1. Those variants of H1 were used in some existing research. Therefore, the obtained results may be used to derive new convergence results for some existing algorithms not covered by H1–H3. We briefly discuss two variants of H1 below. The first variant of H1 is the following one:

H1a. For each $k\in\mathbb{N}$, $\Phi(x^k)+a_k\|x^k-x^{k-1}\|^2\le\Phi(x^{k-1})+\nu_{k-1}$, where $0\le\nu_k\le(1-\tau_k)[\Phi(x^{k-1})+\nu_{k-1}-\Phi(x^k)]$ with $\nu_0=0$ and $\tau_k\in[\tau,1]$ for some $\tau\in(0,1]$.

By letting $C_{k-1}:=\Phi(x^{k-1})+\nu_{k-1}$ and using $\nu_k \le (1-\tau_k)[\Phi(x^{k-1})+\nu_{k-1}-\Phi(x^k)]$, it holds that $C_k \le (1-\tau_k)C_{k-1}+\tau_k\Phi(x^k)$. Then, using the same arguments as before we can prove that any sequence satisfying conditions H1a and H2–H3 also has the convergence results of Theorems 3.3 and 3.6. One can check that the iterate sequence of [10, Algorithm 1] satisfies conditions H1a and H2–H3 under Assumptions A1–A3 there.

The second variant of condition H1 is replaced by the following procedure:

H1b. For each $k\in\mathbb{N}$, $\Phi(x^k)+a_k\|x^k-x^{k-1}\|^2\le C_{k-1}+\eta_{k-1}^2$ where $a_k\ge\underline{a}$ for some $\underline{a}>0$ and $C_k=(1-\tau_k)C_{k-1}+\tau_k\Phi(x^k)$ with $C_0=\Phi(x^0)$ and $\tau_k\in[\tau,1]$ for a $\tau\in(0,1]$.

Sun [24] achieved the convergence of the iterate sequence satisfying conditions H1b with $\tau_k\equiv 1$ and H2–H3, by assuming that Φ is a KL function and $\{\eta_k\}_{k\in\mathbb{N}}$ satisfies

$$\{\eta_k\}_{k\in\mathbb{N}}\subset\mathbb{R}_+ \text{ with } \sum_{k=1}^{\infty}\eta_k<\infty \text{ and } \sum_{k=1}^{\infty}\left(\sum_{l=k}^{\infty}\eta_l^2\right)^{\frac{\varsigma-1}{\varsigma}}<\infty \text{ for some } \varsigma>1. \tag{3.20}$$

In fact, for any sequence $\{x^k\}_{k\in\mathbb{N}}$ obeying conditions H1b and H2–H3, by introducing the potential function $\Psi(z):=\Phi(x)+\varsigma^{-1}t^{\varsigma}$ for $z=(x,t)\in\mathbb{X}\times\mathbb{R}_{++}$ and following arguments similar to those for Theorem 3.3, we can obtain the same convergence result under the assumption that Φ is a KL function and $\{\eta_k\}_{k\in\mathbb{N}}$ satisfies (3.20).

4. Applications of the iterative framework.

This section demonstrates that the novel iterative framework encompasses some existing algorithms. We focus on two particular examples: the proximal gradient method (PGM) and the Riemannian gradient method (RGM). We first show that the iterate sequence of a PGM with the ZH-type nonmonotone line-search for nonconvex and nonsmooth composite problems complies with conditions H1–H3. We then apply the conclusions of Theorems 3.3 and 3.6 to provide its full convergence certificate under the KL property of objective functions. To the best of our knowledge, this convergence result is new.

4.1. Convergence of PGM with ZH-type nonmonotone line-search.

Consider the following nonconvex and nonsmooth composite optimization problem:

$$\min_{x\in\mathbb{X}}\Theta(x):=f(x)+g(x), \tag{4.1}$$

where $f:\mathbb{X}\to\overline{\mathbb{R}}$ and $g:\mathbb{X}\to\overline{\mathbb{R}}$ satisfy the following basic assumption.

Assumption 1.

(i) g is a proper, lsc, and prox-bounded function with $\operatorname{dom} g\neq\emptyset$ (for the prox-boundedness, the reader may refer to [22, Definition 1.23]).

(ii) f is differentiable on an open set $\mathcal{O}\supset\operatorname{dom} g$ with ∇f strictly continuous on $\mathcal{O}$.

(iii) the function Θ is lower bounded, i.e., $\inf_{x\in\mathbb{X}}\Theta(x)>-\infty$.

Recently, Marchi [17] proposed a PGM with the ZH-type nonmonotone line-search strategy for solving (4.1), and its iteration steps are described as follows, where

$$\operatorname{Prox}_{\gamma g}(z):=\arg\min_{x\in\mathbb{X}}\left\{\frac{1}{2\gamma}\|x-z\|^2+g(x)\right\}$$

is the proximal mapping of the function g associated with the parameter $\gamma>0$.

From [17, Lemma 4.1], Algorithm 4.1 is well defined, and for its iterate sequence, Marchi [17] only achieved the subsequence convergence. Next we prove that the iterate sequence of Algorithm 4.1 meets conditions H1–H3. Consequently, its full convergence and local convergence rate naturally follows the convergence results of section 3.

Algorithm 4.1 (nonmonotone line-search PGM).

Initialization: Select $0<\gamma_{\min}\le\gamma_{\max}<\infty$, $\alpha,\beta\in(0,1)$, $p_{\min}\in(0,1]$. Choose $x^0\in\mathrm{dom}g$. Set $C_0=\Theta(x^0)$ and $k:=0$.

while the termination condition is not satisfied **do**

1: Choose $\gamma_{k,0}\in[\gamma_{\min},\gamma_{\max}]$.

2: **For** $l=0,1,2,\ldots$ **do**

3: Let $\gamma_k=\gamma_{k,0}\beta^l$ and compute $x^{k+1}\in\mathrm{Prox}_{\gamma_k g}(x^k-\gamma_k\nabla f(x^k))$.

4: If $\Theta(x^{k+1})\le C_k-\frac{\alpha}{2\gamma_k}\|x^{k+1}-x^k\|^2$, then go to step 6.

5: **end for**

6: Choose $p_k\in[p_{\min},1]$, and set $C_{k+1}=(1-p_k)C_k+p_k\Theta(x^{k+1})$.

7: Set $k\leftarrow k+1$ and go to step 1.

end (while)

THEOREM 4.1. *Suppose the sequence $\{x^k\}_{k\in\mathbb{N}}$ of Algorithm 4.1 is bounded. Then,*

(i) *when Θ is a KL function, $\sum_{k=0}^{\infty}\|x^k-x^{k-1}\|<\infty$ and $\{x^k\}_{k\in\mathbb{N}}$ is convergent;*

(ii) *when Θ is a KL function of exponent $\theta\in(0,1)$, $\{x^k\}_{k\in\mathbb{N}}$ converges to a point $\widetilde{x}\in crit\Theta$ and there exist $\overline{k}\in\mathbb{N},\gamma>0$, and $\varrho\in(0,1)$ such that*

$$\|x^k-\widetilde{x}\|\le\sum_{j=k}^{\infty}\|x^{j+1}-x^j\|\le\begin{cases}\gamma\varrho^k & \text{if}\,\theta\in(0,1/2];\\ \gamma k^{\frac{1-\theta}{1-2\theta}} & \text{if}\,\theta\in(1/2,1)\end{cases}\quad \text{for all}\,k\ge\overline{k}.$$

Proof. From the proof of [17, Lemma 4.1], it follows that for every $k\in\mathbb{N}$,

$$\Theta(x^{k+1})\le C_k-\frac{\alpha}{2\gamma_k}\|x^{k+1}-x^k\|^2. \tag{4.2}$$

In addition, from the boundedness of $\{x^k\}_{k\in\mathbb{N}}$ and [17, Corollary 4.5], the argument by contradiction shows that there exists $\underline{\gamma}>0$ such that $\gamma_{\max}\ge\gamma_k\ge\underline{\gamma}$ for all $k\in\mathbb{N}$. Thus, the sequence $\{x^k\}_{k\in\mathbb{N}}$ satisfies condition H1 with $a_k\equiv\frac{\alpha}{2\gamma_{\max}}$. From Lemma 2.3(iii), $\lim_{k\to\infty}(x^k-x^{k-1})=0$. As the sequence $\{x^k\}_{k\in\mathbb{N}}$ is bounded, there exists a convergent subsequence $\{x^{k_j}\}_{j\in\mathbb{N}}$ with the limit, say, $\overline{x}$, i.e., $\lim_{j\to\infty}x^{k_j}=\overline{x}$. Together with $\lim_{k\to\infty}(x^k-x^{k-1})=0$, we obtain $\lim_{k\to\infty}x^{k_j-1}=\overline{x}$. We next argue that $\limsup_{j\to\infty}\Theta(x^{k_j})\le\Theta(\overline{x})$. From the definition of x^{k+1}, for each $j\in\mathbb{N}$,

$$\begin{aligned}&\langle\nabla f(x^{k_j-1}),x^{k_j}-x^{k_j-1}\rangle+\frac{1}{2\gamma_{k_j-1}}\|x^{k_j}-x^{k_j-1}\|^2+\Theta(x^{k_j})\\&\le\langle\nabla f(x^{k_j-1}),\overline{x}-x^{k_j-1}\rangle+\frac{1}{2\gamma_{k_j-1}}\|\overline{x}-x^{k_j-1}\|^2+\Theta(\overline{x})+f(x^{k_j})-f(\overline{x}).\end{aligned}$$

Passing the limit $j\to\infty$ and using $\lim_{j\to\infty}x^{k_j}=\overline{x}=\lim_{j\to\infty}x^{k_j-1}$ and $\gamma_{k_j-1}\ge\underline{\gamma}$, we conclude that $\limsup_{j\to\infty}\Theta(x^{k_j})\le\Theta(\overline{x})$, so condition H3 also holds.

Denote $\omega(x^0)$ by the set of accumulation points of $\{x^k\}_{k\in\mathbb{N}}$. We next prove that condition H2 holds. For each $k\in\mathbb{N}$, from the definition of x^{k+1} in step 3,

$$0\in\nabla f(x^k)+\frac{1}{\gamma_k}(x^{k+1}-x^k)+\partial g(x^{k+1}),$$

which together with the expression of Θ implies that for each $k\in\mathbb{N}$,

$$\partial\Theta(x^{k+1})\ni w^{k+1}:=\nabla f(x^{k+1})-\nabla f(x^k)-\frac{1}{\gamma_k}\big(x^{k+1}-x^k\big). \tag{4.3}$$

We claim that there exists $\overline{L}>0$ such that for all $k\in\mathcal{K}:=\{k\in\mathbb{N}\,|\,x^{k+1}\neq x^k\}$,

$$\limsup_{\mathcal{K}\ni k\to\infty}\frac{\|\nabla f(x^{k+1})-\nabla f(x^k)\|}{\|x^{k+1}-x^k\|}\le\overline{L}. \tag{4.4}$$

If not, there must exist an index set $\widehat{\mathcal{K}}\subset\mathcal{K}$ such that

$$\lim_{\widehat{\mathcal{K}}\ni k\to\infty}\frac{\|\nabla f(x^{k+1})-\nabla f(x^k)\|}{\|x^{k+1}-x^k\|}=\infty. \tag{4.5}$$

We assume that $\lim_{\widehat{\mathcal{K}}\ni k\to\infty}x^k=\widehat{x}^*$ (if necessary by taking a subsequence of $\{x^k\}_{k\in\widehat{\mathcal{K}}}$). Clearly, $\widehat{x}^*\in\omega(x^0)$. By Lemma 2.3(iii), $\lim_{\widehat{\mathcal{K}}\ni k\to\infty}\big[x^k+(x^{k+1}-x^k)\big]=\widehat{x}^*$. Since ∇f is strictly continuous at $\widehat{x}^*$, there exists $\widehat{L}>0$ such that for all $k\in\widehat{\mathcal{K}}$ large enough,

$$\|\nabla f(x^{k+1})-\nabla f(x^k)\|\le\widehat{L}\|x^{k+1}-x^k\|,$$

which is a contradiction to (4.5). Consequently, the claimed inequality (4.4) holds. Thus, for all $k\in\mathcal{K}$, $\|\nabla f(x^{k+1})-\nabla f(x^k)\|\le\overline{L}\|x^{k+1}-x^k\|$. Together with the definition of w^{k+1}, it follows that $\|w^{k+1}\|\le(\frac{1}{\gamma_k}+\overline{L})\|x^{k+1}-x^k\|$ for all $k\in\mathcal{K}$. Recall that for each $k\in\mathbb{N}$, $w^{k+1}\in\partial\Theta(x^{k+1})$ and $\gamma_k\ge\underline{\gamma}$, so $\|w^{k+1}\|\le b\|x^{k+1}-x^k\|$ for each $k\in\mathbb{N}$ with $b=\underline{\gamma}^{-1}+\overline{L}>0$. This shows that condition H2 holds. The desired conclusions follow Theorems 3.3 and 3.6 with $\widetilde{x}=\overline{x}\in(\partial\Phi)^{-1}(0)$. ∎

4.2. Convergence of RGM with ZH-type nonmonotone line-search. Let $\mathcal{M}$ be an embedded submanifold of the finite-dimensional vector space $\mathbb{X}$. Consider

$$\min_{x\in\mathbb{X}}F(x):=f(x)+\delta_{\mathcal{M}}(x), \tag{4.6}$$

where $f:\mathcal{O}\supset\mathcal{M}\to\mathbb{R}$ is an L_f-smooth function and $\mathcal{O}$ is an open set of $\mathbb{X}$, and $\delta_{\mathcal{M}}$ denotes the indicator function of $\mathcal{M}$, i.e., $\delta_{\mathcal{M}}(x)=0$ if $x\in\mathcal{M}$, and ∞ otherwise. Assume that the function F is lower bounded, which automatically holds if $\mathcal{M}$ is compact. For this problem, Wen and Yin [27] and Oviedo [18] proposed the following RGM with the ZH-type nonmonotone line-search strategy, where $\mathrm{grad}f(x^k)$ denotes the Riemannian gradient of f at x^k, $\mathrm{T}_{x^k}\mathcal{M}$ is the tangent space of $\mathcal{M}$ at x^k, and $\mathrm{R}_{x^k}(\cdot)$ denotes the retraction mapping at x^k from the tangent bundle to $\mathcal{M}$ (see [6, Definition 2.1]).

For Algorithm 4.2, as far as we know, only the gradient sequence $\{\mathrm{grad}f(x^k)\}_{k\in\mathbb{N}}$ is proved to converge to zero under the following condition with $c_1>0$ and $c_2>0$:

$$\langle\mathrm{grad}f(x^k),z^k\rangle\le-c_1\|\mathrm{grad}f(x^k)\|^2\ \text{ and }\ \|z^k\|\le c_2\|\mathrm{grad}f(x^k)\|\quad\text{for all } k\in\mathbb{N}. \tag{4.7}$$

Next we establish the full convergence and local convergence rate of the iterate sequence by arguing that it conforms to conditions H1–H3. Consequently, the convergence results in section 3 are applicable to it.

THEOREM 4.2. *Suppose that the submanifold $\mathcal{M}$ is compact and that F is a KL function. Then, for the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by Algorithm 4.2, it holds that $\sum_{k=0}^{\infty}\|x^k-x^{k-1}\|<\infty$. If in addition F is a KL function of exponent $\theta\in(0,1)$, then it converges to some $\widetilde{x}\in\mathrm{crit}F$ and there exist $\overline{k}\in\mathbb{N},\gamma>0$ and $\varrho\in(0,1)$ such that*

$$\|x^k-\widetilde{x}\|\le\sum_{j=k}^{\infty}\|x^{j+1}-x^j\|\le\begin{cases}\gamma\varrho^k & \text{if}\,\theta\in(0,1/2],\\ \gamma k^{\frac{1-\theta}{1-2\theta}} & \text{if}\,\theta\in(1/2,1)\end{cases}\quad\text{for all}\,k\ge\overline{k}.$$

Algorithm 4.2 (nonmonotone line-search RGM).

Initialization: Select $0<\alpha_m\le\alpha_M<\infty$, $\rho_1\in(0,1),\rho_2\in[0,1),\beta\in(0,1)$, and $p_{\min}\in(0,1]$. Choose $x^0\in\mathcal{M}$ and $z^0\in\mathrm{T}_{x^0}\mathcal{M}$ satisfying (4.7). Set $C_0=F(x^0)$ and $k:=0$.

while the termination condition is not satisfied **do**

1: Choose $\alpha_{k,0}\in[\alpha_m,\alpha_M]$.

2: **For** $l=0,1,2,\ldots$ **do**

3: Set $\alpha_k=\beta^l\alpha_{k,0}$.

4: If $f(\mathrm{R}_{x^k}(\alpha_k z^k))\le C_k+\rho_1\alpha_k\langle \mathrm{grad} f(x^k),z^k\rangle-\rho_2\alpha_k^2\|z^k\|^2$, go to step 6.

5: **end for**

6: Let $x^{k+1}=\mathrm{R}_{x^k}(\alpha_k z^k)$ and select a direction $z^{k+1}\in\mathrm{T}_{x^{k+1}}\mathcal{M}$ satisfying (4.7).

7: Choose $p_k\in[p_{\min},1]$, and set $C_{k+1}=(1-p_k)C_k+p_kf(x^{k+1})$.

8: Set $k\leftarrow k+1$ and go to step 1.

end (while)

Proof. Since $\{x^k\}_{k\in\mathbb{N}}\subset\mathcal{M}$, the compactness of $\mathcal{M}$ implies the boundedness of $\{x^k\}_{k\in\mathbb{N}}$. Then, there exists a subsequence $\{x^{k_j}\}_{j\in\mathbb{N}}$ with $\lim_{j\to\infty}x^{k_j}=\overline{x}\in\mathcal{M}$ such that $\lim_{j\to\infty}F(x^{k_j})=F(\overline{x})$, and condition H3 holds. By combining [6, equation (B.3)] with $x^{k+1}=\mathrm{R}_{x^k}(\alpha_k z^k)$, there exists a constant $\widetilde{\alpha}>0$ such that for each $k\in\mathbb{N}$, $\|x^{k+1}-x^k\|^2\le\widetilde{\alpha}^2\alpha_M^2\|z^k\|^2$. From step 4 of Algorithm 4.2 and (4.7), for each $k\in\mathbb{N}$,

$$F(x^{k+1})=f(x^{k+1})\le C_k-\rho_1c_1\alpha_k\|\mathrm{grad} f(x^k)\|^2-\rho_2\alpha_k^2\|z^k\|^2\le C_k-\frac{\rho_1c_1}{c_2}\alpha_k\|z^k\|^2.$$

In addition, from [18, Lemma 2], there exist $k_0\in\mathbb{N}$ and $\underline{\alpha}>0$ such that $\alpha_k\ge\underline{\alpha}$ for all $k\ge k_0$. These two facts show that condition H1 holds. The remaining part is to show that condition H2 holds. By Lemma 2.3(i)–(ii), the sequences $\{C_k\}_{k\in\mathbb{N}}$ and $\{F(x^k)\}_{k\in\mathbb{N}}$ converge to the same limit. Together with step 4 of Algorithm 4.2, we have $\lim_{k\to\infty}\alpha_k\|z^k\|^2=0$, which along with $\alpha_k\ge\underline{\alpha}$ for all $k\in\mathbb{N}$ implies that $\lim_{k\to\infty}\|z^k\|=0$. From $\alpha_k\le\alpha_M$ for all k and [6, equation (B.4)], we have $\|\mathrm{R}_{x^k}(\alpha_kz^k)-(x^k+\alpha_kz^k)\|=o(\|\alpha_kz^k\|)$. If necessary by increasing k_0, for each $k\ge k_0$,

$$\|x^{k+1}-x^k\|\ge\alpha_k\|z^k\|-\|\mathrm{R}_{x^k}(\alpha_kz^k)-(x^k+\alpha_kz^k)\|\ge\frac{1}{2}\underline{\alpha}\|z^k\|\ge\frac{1}{2}\underline{\alpha}c_1\|\mathrm{grad} f(x^k)\|, \tag{4.8}$$

where the third inequality is obtained by using the first inequality in (4.7). Note that $\partial F(x^k)=\nabla f(x^k)+\mathrm{N}_{x^k}\mathcal{M}$ for each $k\in\mathbb{N}$, where $\mathrm{N}_{x^k}\mathcal{M}$ denotes the normal space of $\mathcal{M}$ at x^k. Then, we have for each $k\in\mathbb{N}$,

$$\mathrm{dist}(0,\partial F(x^k))=\|\mathrm{Proj}_{\mathrm{T}_{x^k}\mathcal{M}}(\nabla f(x^k))\|=\|\mathrm{grad}\, f(x^k)\|\le 2/(\underline{\alpha}c_1)\|x^{k+1}-x^k\|,$$

where the inequality is due to (4.8). This shows that condition H2 holds with $k_1=1$ and $\varepsilon_k=0$ for all $k\ge k_0$. Now the desired conclusions follow Theorems 3.3 and 3.6. ◻

From [20, Lemma 2.10], the KL property of F at $x\in\mathcal{M}$ in terms of Definition 2.1 is equivalent to the KL property of $f|_{\mathcal{M}}$, the restriction of f on $\mathcal{M}$, at $x\in\mathcal{M}$ in terms of [7, Definition 3.5].

5. Conclusion. In this paper, we proposed a novel nonmonotone descent iterative framework consisting of the ZH-type nonmonotone decrease condition and a relative error condition. We proved that any iterative sequence complying with this framework enjoys full convergence when Φ is a KL function, and the convergence is linear if Φ is a KL function of exponent $1/2$. This answers the question whether a descent method with ZH-type nonmonotone line-search strategy converges. We also demonstrated the cases of existing algorithms that fall into the proposed framework. As a result, new convergence results are readily available for those algorithms as a consequence of our obtained results. Furthermore, the proofs of the main results may be adapted to modified frameworks that include more existing algorithms as special instances. This shows the potential of the proposed framework to obtain new convergence results for existing algorithms. Compared with the nonmonotone iterative framework [19], the new one possesses the full convergence (local convergence rate) without additional restriction on Φ except its KL property (KL property of exponent $\theta \in (0,1)$). We hope to explore more applications of the proposed framework in our next research project.

Appendix A. ZH-type nonmonotone line search scheme satisfies (H1)–(H3). We first recall the nonmonotone line-search algorithm (NLSA) proposed in [29]. Let $\Phi = f$, which is a continuously differentiable function and is bounded from below on $\mathbb{X} = \mathbb{R}^n$. We follow the notation used in [29]. In particular, NLSA uses x_k with subscript k for its iterates. Furthermore, it uses $\mathbf{g}_k := \nabla f(x_k)$ (gradient of f at x_k) and $\mathbf{d}_k$ for the search direction at x_k. The *direction assumption* used is [29, (2.4), (2.5)]

$$\mathbf{g}_k^\top \mathbf{d}_k \le -c_1 \|\mathbf{g}_k\|^2 \quad \text{and} \quad \|\mathbf{d}_k\| \le c_2 \|\mathbf{g}_k\| \tag{A.1}$$

for some positive constants c_1 and c_2. NLSA uses nonmonotone Wolfe conditions or nonmonotone Armijo conditions to select its steplength α_k. To simplify our validation, we use the latter for the demonstration. The Armijo search inequality on α_k is [29, (1.4)],

$$f(x_k + \alpha_k \mathbf{d}_k) \le C_k + \delta \alpha_k \mathbf{g}_k^\top \mathbf{d}_k \quad \text{and} \quad \alpha_k \le \mu, \tag{A.2}$$

where $0 < \delta < 1$, $\mu > 0$, and C_k is updated as follows [29, (1.6)]:

$$Q_{k+1} = \eta_k Q_k + 1,\ C_{k+1} = (\eta_k Q_k C_k + f(x_{k+1}))/Q_{k+1},\ C_0 = f(x_0), \text{ and } Q_0 = 1. \tag{A.3}$$

Here, $\eta_k \in [\eta_{\min}, \eta_{\max}]$ with $0 \le \eta_{\min} \le \eta_{\max} \le 1$. The update on the iterate is $x_{k+1} = x_k + \alpha_k \mathbf{d}_k$.

With the above setting, [29, Lemma 2.1] shows that there exists $\underline{\alpha} > 0$ such that $\alpha_k \ge \underline{\alpha}$. We now prove all three conditions in (H1)–(H3) are satisfied with the sequence $\{x_k\}$.

On H1. It follows from [29, (2.8)] that for some $\beta > 0$

$$\begin{aligned} f(x_{k+1}) &\le C_k - \beta \|\mathbf{g}_k\|^2 \quad (\text{by [29, (2.8)]}) \\ &\le C_k - \frac{\beta}{c_2^2} \|\mathbf{d}_k\|^2 \quad (\text{by (A.1)}) \\ &= C_k - \frac{\beta}{\alpha_k^2 c_2^2} \|x_{k+1} - x_k\|^2. \end{aligned}$$

Referring to (A.3), H1 holds with the following choices of a_k and τ_k. Let

$$a_k := \beta/(\alpha_{k-1}^2 c_2^2) \overset{(A.2)}{\ge} \beta/(\mu^2 c_2^2) =: \underline{a} > 0$$

and

$$\tau_k := \frac{1}{\eta_{k-1}Q_{k-1}+1} \geq 1-\eta_{\max} := \tau > 0,$$

where the third inequality used the fact $Q_k \leq 1/(1-\eta_{\max})$ established in [29, (2.15)] when $\eta_{\max} < 1$.

On H2. According to the direction assumption (A.1), for sufficiently large k,

$$\|g_k\|^2 \leq -c_1^{-1} g_k^\top d_k \leq c_1^{-1}\|g_k\|\|d_k\| = (\alpha_k c_1)^{-1}\|g_k\|\|x_{k+1}-x_k\|,$$

which implies that condition H2 holds with $k_1 = 1, b_k = c_1\alpha_k$ for sufficiently large k and $\varepsilon_k \equiv 0$. Since $\alpha_k \geq \underline{\alpha}$, we must have $\sum_{k=1}^{\infty} b_k = \infty$.

On H3. Note that $f(x_{k+1}) \leq C_k \leq C_{k-1} \leq \cdots \leq C_0 = f(x_0)$, the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by the NLSA is contained in the level set $\{x \in \mathbb{R}^n \,|\, f(x) \leq f(x_0)\}$. Hence, whenever the level set is bounded, the sequence $\{x_k\}_{k\in\mathbb{N}}$ is bounded and there exists subsequence $\{x_{k_j}\}_{j\in\mathbb{N}}$ with $\lim_{j\to\infty} x_{k_j} = \overline{x}$ such that $\lim_{j\to\infty} f(x_{k_j}) = f(\overline{x})$, so condition H3 holds. Moreover, with the choices a_k and b_k above, we have

$$\overline{B} = \sup_{\mathbb{N}\ni k\geq k_1} \frac{1}{b_k}\sum_{i=k-k_1}^{k+k_1}\frac{1}{\sqrt{a_i}} = \sup_{\mathbb{N}\ni k\geq k_1}\frac{1}{c_1\alpha_k}\sum_{i=k-k_1}^{k+k_1}\frac{c_2\alpha_{i-1}}{\sqrt{\beta}} \leq (2k_1+1)\frac{c_2\mu}{c_1\sqrt{\beta}\underline{\alpha}} < \infty,$$

where the inequality used the fact $\underline{\alpha} \leq \alpha_k \leq \mu$. Therefore, the assumption (1.2) is also satisfied. Thus, the NLSA of [29] with $\eta_{\max} < 1$ falls within our iterative framework for $\varepsilon_k \equiv 0$ whenever the level set $\{x \in \mathbb{R}^n \,|\, f(x) \leq f(x_0)\}$ is bounded.

Similarly, the iterate sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the NLSAs [10, 23] under the direction assumption there also falls within our iterative framework with $\varepsilon_k \equiv 0$.

REFERENCES

[1] P. A. Absil, R. Mahony, and B. Andrews, *Convergence of the iterates of descent methods for analytic cost functions*, SIAM J. Optim., 16 (2005), pp. 531–547, https://doi.org/10.1137/040605266.

[2] H. Attouch and J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Math. Program., 116 (2009), pp. 5–16, https://doi.org/10.1007/s10107-007-0133-5.

[3] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457, https://doi.org/10.1287/moor.1100.0449.

[4] H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods*, Math. Program., 137 (2013), pp. 91–129, https://doi.org/10.1007/s10107-011-0484-9.

[5] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146 (2014), pp. 459–494, https://doi.org/10.1007/s10107-013-0701-9.

[6] N. Boumal, P.-A. Absil, and C. Cartis, *Global rates of convergence for nonconvex optimization on manifolds*, IMA J. Numer. Anal., 39 (2019), pp. 1–33, https://doi.org/10.1093/imanum/drx080.

[7] J. X. Cruz Neto, P. R. Oliveira, P. A. Soares, Jr., and A. Soubeyran, *Learning how to play Nash, potential games and alternating minimization method for structured nonconvex problems on Riemannian manifolds*, J. Convex Anal., 20 (2013), pp. 395–438, https://www.heldermann.de/JCA/JCA20/JCA202/jca20025.htm.

[8] O. P. Ferreira, E. M. Santos, and J. C. O. Souza, *A boosted DC algorithm for non-differentiable DC components with non-monotone line search*, Comput. Optim. Appl., 88 (2024), pp. 783–818, https://doi.org/10.1007/s10589-024-00578-4.

[9] P. Frankel, G. Garrigos, and J. Peypouquet, *Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates*, J. Optim. Theory Appl., 165 (2015), pp. 874–900, https://doi.org/10.1007/s10957-014-0642-3.

[10] G. N. Grapiglia and E. W. Sachs, *On the worst-case evaluation complexity of non-monotone line search algorithms*, Comput. Optim. Appl., 68 (2017), pp. 555–577, https://doi.org/10.1007/s10589-017-9928-3.

[11] L. Grippo, F. Lampariello, and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM J. Numer. Anal., 23 (1986), pp. 707–716, https://doi.org/10.1137/0723046.

[12] A. D. Ioffe, *An invitation to tame optimization*, SIAM J. Optim., 19 (2009), pp. 1894–1917, https://doi.org/10.1137/080722059.

[13] C. Kanzow and P. Mehlitz, *Convergence properties of monotone and nonmonotone proximal gradient methods revisited*, J. Optim. Theory Appl., 195 (2022), pp. 624–646, https://doi.org/10.1007/s10957-022-02101-3.

[14] H. Li and Z. C. Lin, *Accelerated proximal gradient methods for nonconvex programming*, in Proceedings of NeurIPS, 2015, pp. 379–387, https://proceedings.neurips.cc/paper/2015/hash/f7664060cc52bc6f3d620bcedc94a4b6-Abstract.html.

[15] R. Y. Liu, S. H. Pan, Y. Q. Wu, and X. Q. Yang, *An inexact regularized proximal Newton method for nonconvex and nonsmooth optimization*, Comput. Optim. Appl., 88 (2024), pp. 603–641, https://doi.org/10.1007/s10589-024-00560-0.

[16] Z. Lu and Z. Zhou, *Nonmonotone enhanced proximal DC algorithms for a class of structured nonsmooth DC programming*, SIAM J. Optim., 29 (2019), pp. 2725–2752, https://doi.org/10.1137/18M1214342.

[17] A. D. Marchi, *Proximal gradient methods beyond monotony*, J. Nonsmooth Anal. Optim., 4 (2023), pp. 1–18, https://doi.org/10.46298/jnsao-2023-10290.

[18] H. Oviedo, *Global convergence of Riemannian line search methods with a Zhang-Hager-type condition*, Numer. Algorithms, 91 (2022), pp. 1183–1203, https://doi.org/10.1007/s11075-022-01298-8.

[19] Y. Qian and S. Pan, *Convergence of a class of nonmonotone descent methods for Kurdyka-Łojasiewicz optimization problems*, SIAM J. Optim., 33 (2023), pp. 638–651, https://doi.org/10.1137/22M1469663.

[20] Y. T. Qian, S. H. Pan, and L. H. Xiao, *Error bound and exact penalty method for optimization problems with nonnegative orthogonal constraint*, IMA J. Numer. Anal., 44 (2024), pp. 120–156, https://doi.org/10.1093/imanum/drac084.

[21] S. M. Robinson, *Some continuity properties of polyhedral multifunctions*, Math. Program. Stud., 14 (1981), pp. 206–214, https://doi.org/10.1007/BFb0120929.

[22] R. T. Rockafellar and W. R. J.-B., *Variational Analysis*, Springer, New York, 1998.

[23] E. W. Sachs and S. M. Sachs, *Nonmonotone line searches for optimization algorithms*, Control Cybernet., 40 (2011), pp. 1059–1075, http://control.ibspan.waw.pl:3000/contents/show/52?year=2011.

[24] T. Sun, *Sequence convergence of inexact nonconvex and nonsmooth algorithms with more realistic assumptions*, Numer. Funct. Anal. Optim., 42 (2021), pp. 234–250, https://doi.org/10.1080/01630563.2020.1871362.

[25] A. Themelis, L. Stella, and P. Patrinos, *Forward-backward envelope for the sum of two nonconvex functions: Further properties and nonmonotone line search algorithms*, SIAM J. Optim., 28 (2018), pp. 2274–2303, https://doi.org/10.1137/16M1080240.

[26] P. L. Toint, *An assessment of nonmonotone line search techniques for unconstrained optimization*, SIAM J. Sci. Comput., 17 (1996), pp. 725–739, https://doi.org/10.1137/S106482759427021X.

[27] Z. W. Wen and W. T. Yin, *A feasible method for optimization with orthogonality constraints*, Math. Program., 142 (2013), pp. 397–434, https://doi.org/10.1007/s10107-012-0584-1.

[28] Y. Xu and W. Yin, *A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion*, SIAM J. Imaging Sci., 6 (2013), pp. 1758–1789, https://doi.org/10.1137/120887795.

[29] H. Zhang and W. W. Hager, *A nonmonotone line search technique and its application to unconstrained optimization*, SIAM J. Optim., 14 (2004), pp. 1043–1056, https://doi.org/10.1137/S1052623403428208.