

## CONVERGENCE RATE ANALYSIS OF A DYKSTRA-TYPE PROJECTION ALGORITHM\*

XIAOZHOU WANG<sup>†</sup> AND TING KEI PONG<sup>†</sup>

**Abstract.** Given closed convex sets  $C_i$ ,  $i = 1, \dots, \ell$ , and some nonzero linear maps  $A_i$ ,  $i = 1, \dots, \ell$ , of suitable dimensions, the multiset split feasibility problem aims at finding a point in  $\bigcap_{i=1}^{\ell} A_i^{-1}C_i$  based on computing projections onto  $C_i$  and multiplications by  $A_i$  and  $A_i^T$ . In this paper, we consider the associated best approximation problem, i.e., the problem of computing projections onto  $\bigcap_{i=1}^{\ell} A_i^{-1}C_i$ ; we refer to this problem as the best approximation problem in multiset split feasibility settings (BA-MSF). We adapt the Dykstra’s projection algorithm, which is classical for solving the BA-MSF in the special case when all  $A_i = I$ , to solve the general BA-MSF. Our Dykstra-type projection algorithm is derived by applying (proximal) coordinate gradient descent to the Lagrange dual problem, and it only requires computing projections onto  $C_i$  and multiplications by  $A_i$  and  $A_i^T$  in each iteration. Under a standard relative interior condition and a genericity assumption on the point we need to project, we show that the dual objective satisfies the Kurdyka-Lojasiewicz property with an explicitly computable exponent on a neighborhood of the (typically unbounded) dual solution set when each  $C_i$  is  $C^{1,\alpha}$ -cone reducible for some  $\alpha \in (0, 1]$ : this class of sets covers the class of  $C^2$ -cone reducible sets, which include all polyhedrons, second-order cone, and the cone of positive semidefinite matrices as special cases. Using this, explicit convergence rate (linear or sublinear) of the sequence generated by the Dykstra-type projection algorithm is derived. Concrete examples are constructed to illustrate the necessity of some of our assumptions.

**Key words.** Dykstra’s projection algorithm, Kurdyka–Lojasiewicz property, linear convergence,  $C^{1,\alpha}$ -cone reducibility

**MSC codes.** 90C25, 90C30, 90C46, 90C90

**DOI.** 10.1137/23M1545781

**1. Introduction.** The multiset split feasibility (MSF) problems, first introduced by Censor et al. [20], are generalizations of the two-set split feasibility problems [19] and convex feasibility problems [7]. These kinds of problems arise naturally in many contemporary application fields such as image reconstruction; see [20] and references therein. The MSF problem aims at finding a point in the intersection of the linear preimage of a collection of finitely many closed convex sets, i.e.,

$$(1.1) \quad \text{Find } x \in \mathbb{R}^n \text{ such that } A_i x \in C_i \text{ for } i = 1, \dots, \ell,$$

where  $C_i \subseteq \mathbb{R}^{m_i}$ ,  $i = 1, \dots, \ell$ , are closed convex sets and  $A_i \in \mathbb{R}^{m_i \times n}$  for each  $i$ ; moreover, the projections onto  $C_i$  are assumed to be easy to compute, while computing projections onto the sets  $A_i^{-1}C_i$  can be difficult (see [20]). The assumptions concerning projections naturally call for iterative schemes that leverage projections onto  $C_i$  for solving (1.1). One such scheme is the CQ-algorithm proposed in [18] for solving the MSF problem with  $\ell = 2$ , which was later generalized to solve general MSF problems in [20]. Recent works on solution methods for MSF problems can be found in [16, 17, 21, 22, 40]. In this paper, we focus on a natural but relatively less studied variant of the MSF problem. Specifically, given a  $\bar{v} \in \mathbb{R}^n$ , we consider the problem of

\*Received by the editors January 9, 2023; accepted for publication (in revised form) November 15, 2023; published electronically February 6, 2024.

<https://doi.org/10.1137/23M1545781>

**Funding:** The second author is supported partly by the Hong Kong Research Grants Council PolyU153002/21p.

<sup>†</sup>The Hong Kong Polytechnic University, Hong Kong, People’s Republic of China (xzhou.wang@connect.polyu.hk, tk.pong@polyu.edu.hk).

finding the point in  $\bigcap_{i=1}^{\ell} A_i^{-1}C_i$  that is closest to  $\bar{v}$ . In other words, we consider the following best approximation problem in multiset split feasibility settings (BA-MSF):

$$(1.2) \quad \min_{x \in \mathbb{R}^n} f(x) := \frac{1}{2} \|x - \bar{v}\|^2 \quad \text{s.t. } A_i x \in C_i, \quad i = 1, \dots, \ell,$$

where  $\bar{v} \in \mathbb{R}^n$  is a given vector and each  $A_i \in \mathbb{R}^{m_i \times n} \setminus \{0\}$  and each  $C_i$  is a closed convex set; moreover, we assume that

$$\bigcap_{i=1}^{\ell} A_i^{-1}C_i \neq \emptyset.$$

When  $A_i = I$  for all  $i$ , we refer to the corresponding problem (1.2) as the best approximation (BA) problem. In this case, a classical solution method is the Dykstra's projection algorithm proposed in [15, 28]. Each iteration of this algorithm only requires computing projections onto each  $C_i$  instead of  $\bigcap_{i=1}^{\ell} C_i$ , which can be advantageous because the projection onto  $\bigcap_{i=1}^{\ell} C_i$  can be more difficult to compute in general. One remarkable feature of the Dykstra's projection algorithm is that the sequence generated will converge to the unique solution of the BA problem as long as  $\bigcap_{i=1}^{\ell} C_i \neq \emptyset$ ; see [15]. This is in contrast to splitting methods such as Douglas–Rachford splitting, which typically requires additional assumptions to guarantee convergence; see [10, Corollary 28.3]. The Dykstra's projection algorithm and its variants have been studied extensively in recent years concerning its convergence properties and relations to other algorithms; see [6, 9, 11, 24, 27, 29, 33, 34] and references therein. It is now known that the Dykstra's projection algorithm can be derived as a suitable application of the coordinate gradient descent method to the dual problem of (1.2) with  $A_i = I$  for all  $i$ ; see [27, 28]. Moreover, for BA problems with *polyhedral*  $C_i$ , it is known that the sequence generated by the Dykstra's projection algorithm converges linearly [24, 32, 38]. On the other hand, convergence rates in the case of nonpolyhedral  $C_i$  are not very well understood.

In this paper, we will adapt the classical Dykstra's projection algorithm to solve (1.2) and analyze the convergence rate of the resulting algorithm. Following [27, 28], we derive our Dykstra-type projection algorithm by applying (proximal) coordinate gradient descent to a suitable Lagrange dual problem of (1.2), and each iteration of our algorithm only requires projections onto  $C_i$  and multiplications by  $A_i$  and  $A_i^T$ .<sup>1</sup> Then, by imposing the standard relative interior condition  $\bigcap_{i=1}^{\ell} A_i^{-1} \text{ri} C_i \neq \emptyset$  and a genericity assumption on the point  $\bar{v}$ , we show that the objective of the dual problem satisfies Kurdyka–Lojasiewicz (KL) property with exponent  $1/(\alpha + 1)$  for some  $\alpha \in (0, 1]$  on a neighborhood of the (typically unbounded) dual solution set when each  $C_i$  is  $C^{1,\alpha}$ -cone reducible. Based on this, we establish the linear or sublinear convergence of the sequences generated by the Dykstra-type projection algorithm depending on the value of  $\alpha \in (0, 1]$ . The key novelty of our work lies in *both* the convergence rate analysis and the identification of the  $C^{1,\alpha}$ -cone reducibility condition:

- First, our convergence rate results do not follow directly from standard KL-based convergence analysis frameworks such as those in [1, 2, 3]. Indeed, the sequences generated by the Dykstra-type projection algorithm can be unbounded, in general, while the standard convergence rate analysis based on KL property typically requires the boundedness of the sequence generated. Our analysis is made possible thanks to the fact that, under the assumption

<sup>1</sup>Our algorithm reduces to the classical Dykstra's projection algorithm when  $A_i = I$  for all  $i$ .

$\bigcap_{i=1}^{\ell} A_i^{-1} \text{ri} C_i \neq \emptyset$ , the solution set of the dual problem can be written as the sum of a compact convex set and a *subspace*, a key fact which was also established in [4] for the BA problem, i.e., (1.2) with each  $A_i = I$ .

- Second, we would like to point out that  $C^{1,\alpha}$ -cone reducible (with  $\alpha \in (0, 1]$ ) sets are prevalent in applications. Indeed, it covers the notion of  $C^2$ -cone reducible sets introduced in [37],<sup>2</sup> which contains sets such as polyhedrons, second-order cone and positive semidefinite cone. In addition, as we will see later, the  $p$ -norm ball with  $p \in (1, +\infty)$  can be shown to be  $C^{1,\alpha}$ -cone reducible with  $\alpha = \min\{p - 1, 1\}$ ; see Example 4.1 below. Furthermore, when  $p \in (1, 2)$ , the  $p$ -norm ball is  $C^{1,\alpha}$ -cone reducible with  $\alpha = p - 1$  but not  $C^2$ -cone reducible (see Remark 4.8), showing that the class of  $C^{1,\alpha}$ -cone reducible sets is strictly larger than the class of  $C^2$ -cone reducible sets. Our result is thus applicable to analyzing our Dykstra-type projection algorithm for a wide range of sets. Moreover, since our Dykstra-type projection algorithm reduces to the classical Dykstra’s projection algorithm when  $A_i = I$  for all  $i$ , we can deduce convergence rate results for the classical Dykstra’s projection algorithm on a large class of *nonpolyhedral*  $C_i$ , whose convergence rate was previously unknown.

This paper is organized as follows. Section 2 presents some preliminary materials. In section 3, we adapt the classical Dykstra’s projection algorithm to solve the BA-MSF problem (1.2) and present some basic convergence properties. Our key contributions are in sections 4 and 5, where we establish the KL property of the dual objective under suitable assumptions, and use that to study the convergence rate of our Dykstra-type projection algorithm. We also present concrete examples to show the indispensability of some of our assumptions in deriving the error bound and convergence rate results.

**2. Notation and preliminaries.** In this paper, we use  $\mathbb{R}^n$  to denote the  $n$ -dimensional Euclidean space. For an  $x \in \mathbb{R}^n$ , we let  $\|x\|_p = \sqrt[p]{|x_1|^p + \dots + |x_n|^p}$  denote the  $p$ -norm, where  $p \in [1, \infty)$ ; we also use  $\|x\|$  to denote the 2-norm for notational simplicity. For  $x$  and  $y \in \mathbb{R}^n$ , we use  $\langle x, y \rangle$  to denote their inner product. For a matrix  $A \in \mathbb{R}^{m \times n}$ , we use  $\|A\|$  to denote its operator norm. We use  $I$  to denote the identity matrix, whose dimension should be clear from the context. We also use  $B(x, \eta)$  to denote a closed ball centered at  $x$  with radius  $\eta \geq 0$ , i.e.,  $B(x, \eta) = \{u : \|u - x\| \leq \eta\}$ .

An extended-real-valued function  $h : \mathbb{R}^n \rightarrow (-\infty, \infty]$  is said to be proper if  $\text{dom } h := \{x \in \mathbb{R}^n : h(x) < \infty\}$  is nonempty. A proper function is said to be closed if it is lower semicontinuous. For a proper convex function  $h$ , its subdifferential  $\partial h(x)$  at an  $x \in \mathbb{R}^n$  is defined as

$$\partial h(x) = \{\xi \in \mathbb{R}^n : h(y) \geq h(x) + \langle \xi, y - x \rangle \text{ for all } y \in \mathbb{R}^n\}.$$

The domain of  $\partial h$  is defined as  $\text{dom } \partial h := \{x \in \mathbb{R}^n : \partial h(x) \neq \emptyset\}$ , and the convex conjugate of  $h$  is given by

$$h^*(y) = \sup\{\langle x, y \rangle - h(x) : x \in \mathbb{R}^n\}.$$

For a proper closed convex function  $h$ , we have the following equivalences concerning  $\partial h$  and  $h^*$  (see, for example, [36, Proposition 11.3]):

<sup>2</sup>This notion was known as cone reducibility in [37], and as  $C^2$ -cone reducibility in [14, Definition 3.135]. Here, we adopt the latter terminology to highlight the differentiability property.

$$(2.1) \quad h(x) + h^*(y) = \langle x, y \rangle \iff y \in \partial h(x) \iff x \in \partial h^*(y).$$

The proximal operator of a proper closed convex function  $h$  at an  $x \in \mathbb{R}^n$  is defined as

$$\text{prox}_h(x) := \text{Arg min}_{u \in \mathbb{R}^n} \left\{ \frac{1}{2} \|u - x\|^2 + h(u) \right\};$$

recall that the above set of minimizers is a singleton for every  $x \in \mathbb{R}^n$ , and  $\text{prox}_h : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is nonexpansive. Finally, for a proper closed convex function  $h$  with  $\text{Arg min } h \neq \emptyset$ , the following inequality holds (see [36, Proposition 10.59]):

$$(2.2) \quad h(x) - \inf h \leq \text{dist}(0, \partial h(x)) \text{dist}(x, \text{Arg min } h) \quad \text{for all } x \in \mathbb{R}^n,$$

where  $\text{dist}(x, C) := \inf \{ \|x - y\| : y \in C \}$  is the distance of an  $x \in \mathbb{R}^n$  to a set  $C$ .

For a nonempty closed convex set  $C \subseteq \mathbb{R}^n$ , we use  $\text{int } C$  and  $\text{ri } C$  to represent the interior and relative interior of  $C$ , respectively. The indicator function and support function of such a  $C$  are, respectively, defined as

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise,} \end{cases} \quad \text{and} \quad \sigma_C(x) = \sup \{ \langle x, y \rangle : y \in C \}.$$

The normal cone of a nonempty closed convex set  $C$  at  $x \in C$  is defined as

$$\mathcal{N}_C(x) := \partial \delta_C(x) = \{ v \in \mathbb{R}^n : \langle v, u - x \rangle \leq 0 \text{ for all } u \in C \},$$

and the projection of  $x \in \mathbb{R}^n$  onto  $C$  is denoted by  $\text{Proj}_C(x) := \text{prox}_{\delta_C}(x)$ . For a closed convex cone  $K \subseteq \mathbb{R}^n$ , its polar is defined as  $K^\circ = \{ u \in \mathbb{R}^n : \langle u, x \rangle \leq 0 \text{ for all } x \in K \}$ .

Next, we recall the following definition of KL property, which is important for analyzing convergence rate of various first-order methods; see, for example, [1, 2, 3, 13, 30].

**DEFINITION 2.1** (KL property and exponent). *A proper closed convex function  $h : \mathbb{R}^n \rightarrow (-\infty, \infty]$  is said to satisfy the KL property at  $\hat{x} \in \text{dom } \partial h$  if there are  $c \in (0, \infty]$ , a neighborhood  $U$  of  $\hat{x}$  and a continuous concave function  $\varphi : [0, c) \rightarrow [0, \infty)$  with  $\varphi(0) = 0$  such that*

- (i)  $\varphi$  is continuously differentiable on  $(0, c)$  and  $\varphi' > 0$  on  $(0, c)$ ;
- (ii)  $\varphi'(h(x) - h(\hat{x})) \text{dist}(0, \partial h(x)) \geq 1$  whenever  $x \in U \cap \{x : h(\hat{x}) < h(x) < h(\hat{x}) + c\}$ .

If  $h$  has KL property at  $\hat{x}$  with the function  $\varphi(v) = \alpha_0 v^{1-\theta}$  for some  $\alpha_0 > 0$  and  $\theta \in [0, 1)$ , then we say that  $h$  satisfies the KL property at  $\hat{x}$  with exponent  $\theta$ .

A proper closed convex function  $h$  satisfying the KL property with exponent  $\theta \in [0, 1)$  at every point in  $\text{dom } \partial h$  is called a KL function with exponent  $\theta$ .

KL property with exponent  $\theta \in [0, 1)$  is closely related to other notions of error bounds. For example, according to [12, Theorem 5], a proper closed convex function  $h$  satisfies the KL property at an  $\bar{x} \in \text{Arg min } h$  with exponent  $\theta \in [0, 1)$  if and only if there exist positive constants  $c_1$  and  $c_2$  and a neighborhood  $U_{\bar{x}}$  of  $\bar{x}$  such that

$$c_1 \text{dist}(x, \text{Arg min } h) \leq (h(x) - h(\bar{x}))^{1-\theta} \quad \text{for all } x \in U_{\bar{x}} \cap \{x : h(\bar{x}) < h(x) < h(\bar{x}) + c_2\}.$$

Before ending this section, we briefly review the (classical) Dykstra's projection algorithm proposed in [15, 28]. The Dykstra's projection algorithm was developed to solve the following BA problem:

$$(2.3) \quad \min_{x \in \mathbb{R}^n} \frac{1}{2} \|x - \bar{v}\|^2 \quad \text{s.t. } x \in C_i, \quad i = 1, \dots, \ell,$$

---

**Algorithm 2.1.** Dykstra’s projection algorithm for (2.3).

---

Step 1. Choose  $y_1^0 = \dots = y_\ell^0 = 0 \in \mathbb{R}^n$  and  $x^0 = x_\ell^0 = \bar{v} \in \mathbb{R}^n$ . Set  $t = 0$ .

Step 2. Set  $x_0^{t+1} = x_\ell^t$ . For  $i = 1, \dots, \ell$ , compute

$$\begin{aligned} x_i^{t+1} &= \text{Proj}_{C_i}(y_i^t + x_{i-1}^{t+1}), \\ y_i^{t+1} &= y_i^t + x_{i-1}^{t+1} - \text{Proj}_{C_i}(y_i^t + x_{i-1}^{t+1}). \end{aligned}$$

Step 3. Set  $x^{t+1} = x_\ell^{t+1}$ . Update  $t \leftarrow t + 1$  and go to Step 2.

---

where each  $C_i$  is a closed convex set and  $\bigcap_{i=1}^\ell C_i \neq \emptyset$ . Clearly, problem (2.3) is a special instance of (1.2) with  $A_i = I$  for all  $i$ . The Dykstra’s projection algorithm is presented as Algorithm 2.1 below. Note that it makes use of  $\text{Proj}_{C_i}$  in each iteration.

It was shown in [27, 28] that Algorithm 2.1 is equivalent to a (proximal) coordinate gradient descent (CGD) method for solving the (negative of the) dual problem of (2.3), which is given as

$$\min_{y_1, \dots, y_\ell} \frac{1}{2} \left\| \sum_{i=1}^\ell y_i - \bar{v} \right\|^2 - \frac{1}{2} \|\bar{v}\|^2 + \sum_{i=1}^\ell \sigma_{C_i}(y_i).$$

Indeed, with  $y_1^0 = \dots = y_\ell^0 = 0$ , for  $i = 1, \dots, \ell$ , one can show that the  $y_i^{t+1}$  in Algorithm 2.1 can be equivalently obtained as

$$y_i^{t+1} := \text{Arg min}_{y_i \in \mathbb{R}^{m_i}} \left\{ \left\langle \sum_{j=1}^{i-1} y_j^{t+1} + \sum_{j=i}^\ell y_j^t - \bar{v}, y_i - y_i^t \right\rangle + \frac{1}{2} \|y_i - y_i^t\|^2 + \sigma_{C_i}(y_i) \right\}.$$

The following theorem collects some known convergence results of Algorithm 2.1.

**THEOREM 2.2** (convergence properties of Algorithm 2.1). *Consider (2.3). Let  $\{x^t\}$  be generated by Algorithm 2.1. Then the following statements hold.*

- (i) *It holds that  $\lim_{t \rightarrow +\infty} x^t = x^*$ , where  $x^*$  is the unique solution of (2.3).*
- (ii) *If each  $C_i$  is polyhedral, then there exist  $a_1 > 0$ ,  $a_0 \in (0, 1)$ , and a positive integer  $\bar{t}$  such that*

$$\|x^t - x^*\| \leq a_1 a_0^t \quad \forall t \geq \bar{t}.$$

*Proof.* Item (i) was established in [15]. Item (ii) can be deduced from [32, 38]; see also Theorem 5.1 below.  $\square$

**3. A Dykstra-type projection algorithm for BA-MSF problems.** In this section, we describe how Algorithm 2.1 can be adapted to solve the BA-MSF problem (1.2) and discuss some basic convergence properties of the resulting algorithm. We call the resulting algorithm a Dykstra-type projection algorithm. From now on, for notational simplicity, we write

$$(3.1) \quad \mathbf{A} \in \mathbb{R}^{m \times n} \text{ with } \mathbf{A}^T := [A_1^T \quad \dots \quad A_\ell^T] \quad \text{and} \quad D := C_1 \times \dots \times C_\ell \subseteq \mathbb{R}^m,$$

where  $m = \sum_{i=1}^\ell m_i$ ; moreover, we denote the elements in  $\mathbb{R}^m$  in boldface, i.e.,  $\mathbf{y} = (y_1, \dots, y_\ell)$  with  $\mathbf{y} \in \mathbb{R}^m$  and  $y_i \in \mathbb{R}^{m_i}$  for  $i = 1, \dots, \ell$ .

Following the development in [27], by considering a CGD method for solving the dual problem of (1.2), we present a Dykstra-type projection algorithm for solving

the BA-MSF problem (1.2). To this end, we first define  $\psi : \mathbb{R}^{n+m} \rightarrow (-\infty, +\infty]$  as follows:

$$(3.2) \quad \psi(x, \mathbf{y}) := f(x) + \delta_D(\mathbf{A}x + \mathbf{y}).$$

Then it holds that for every  $\bar{x} \in \mathbb{R}^n$  and  $\bar{\mathbf{y}} \in \mathbb{R}^m$  that

$$\begin{aligned} \psi^*(\bar{x}, \bar{\mathbf{y}}) &= \sup_{x, \mathbf{y}} \{ \langle (\bar{x}, \bar{\mathbf{y}}), (x, \mathbf{y}) \rangle - f(x) - \delta_D(\mathbf{A}x + \mathbf{y}) \} \\ &= \sup_{x, \mathbf{y}} \{ \langle -\mathbf{A}^T \bar{\mathbf{y}} + \bar{x}, x \rangle - f(x) + \langle \bar{\mathbf{y}}, \mathbf{A}x + \mathbf{y} \rangle - \delta_D(\mathbf{A}x + \mathbf{y}) \} \\ &= \sup_x \{ \langle -\mathbf{A}^T \bar{\mathbf{y}} + \bar{x}, x \rangle - f(x) \} + \sup_z \{ \langle \bar{\mathbf{y}}, z \rangle - \delta_D(z) \} \\ (3.3) \quad &= f^*(-\mathbf{A}^T \bar{\mathbf{y}} + \bar{x}) + \sigma_D(\bar{\mathbf{y}}). \end{aligned}$$

Therefore, in view of [35, Theorem 31.2] and (3.3), the dual problem of (1.2) is

$$(3.4) \quad \max_{\mathbf{y} \in \mathbb{R}^m} -\psi^*(0, \mathbf{y}).$$

Since  $f^*(x) = \frac{1}{2}\|x + \bar{v}\|^2 - \frac{1}{2}\|\bar{v}\|^2$ , (3.4) can be further written as

$$(3.5) \quad \max_{\mathbf{y} \in \mathbb{R}^m} -\frac{1}{2} \left\| \sum_{i=1}^{\ell} A_i^T y_i - \bar{v} \right\|^2 + \frac{1}{2} \|\bar{v}\|^2 - \sum_{i=1}^{\ell} \sigma_{C_i}(y_i).$$

Note that using the definition of  $\mathbf{A}$  and  $D$  in (3.1), we can rewrite the above dual problem equivalently as

$$(3.6) \quad d^* := \min_{\mathbf{y} \in \mathbb{R}^m} d(\mathbf{y}), \quad \text{where } d(\mathbf{y}) := \underbrace{\frac{1}{2} \left\| \mathbf{A}^T \mathbf{y} - \bar{v} \right\|^2 - \frac{1}{2} \|\bar{v}\|^2}_{g(\mathbf{y})} + \sigma_D(\mathbf{y}).$$

The following proposition states that the duality gap is always zero between (1.2) and its dual problem (3.5). This justifies the rationale of solving (1.2) by considering (3.6).

**PROPOSITION 3.1.** *Consider (1.2). Let the function  $d$  be given in (3.6). Then it holds that  $\inf\{f(x) : x \in \bigcap_{i=1}^{\ell} A_i^{-1}C_i\} = -d^* < \infty$ . Moreover, for any  $\mathbf{y}^* = (y_1^*, \dots, y_{\ell}^*) \in \text{Arg min } d$ , it holds that*

$$x^* = \bar{v} - \sum_{i=1}^{\ell} A_i^T y_i^* = \bar{v} - \mathbf{A}^T \mathbf{y}^*,$$

where  $x^*$  is the unique solution of (1.2).

*Proof.* For the  $\psi$  defined in (3.2), notice that  $\mathbf{y} \mapsto \psi(\cdot, \mathbf{y})$  is a proper, closed convex bifunction.<sup>3</sup> Moreover, the primal problem (i.e.,  $\min_{x \in \mathbb{R}^n} \psi(x, \mathbf{0})$ ) has a unique solution thanks to  $\bigcap_{i=1}^{\ell} A_i^{-1}C_i \neq \emptyset$ . Then we can deduce from [35, Theorem 30.4(i)] that the duality gap is zero, i.e.,

$$\sup_{\mathbf{y}} -\psi^*(0, \mathbf{y}) = \inf_x \psi(x, \mathbf{0}) = \inf \left\{ f(x) : x \in \bigcap_{i=1}^{\ell} A_i^{-1}C_i \right\} = -d^* < \infty.$$

<sup>3</sup>See [35, pages 291–292] for the definition of a bifunction. The bifunction  $\mathbf{y} \mapsto \psi(\cdot, \mathbf{y})$  is said to be proper closed convex if its graph function  $(x, \mathbf{y}) \mapsto \psi(x, \mathbf{y})$  is so; see [35, page 293].

Next, if  $\mathbf{y}^* = (y_1^*, \dots, y_\ell^*) \in \text{Arg min } d$ , then we have from Theorem 10.1 and Exercise 8.8 of [36] that

$$\mathbf{0} \in \partial d(\mathbf{y}^*) = \mathbf{A}(\mathbf{A}^T \mathbf{y}^* - \bar{v}) + \partial \sigma_D(\mathbf{y}^*).$$

The above display together with [36, Example 11.4] implies

$$\mathbf{y}^* \in \mathcal{N}_D(\mathbf{A}(\bar{v} - \mathbf{A}^T \mathbf{y}^*)).$$

Write  $\hat{x} := \bar{v} - \mathbf{A}^T \mathbf{y}^*$  for notational simplicity. Then we have, in view of [36, Proposition 10.5], that  $y_i^* \in \mathcal{N}_{C_i}(A_i \hat{x})$  for all  $i$ . Hence, we have  $A_i^T y_i^* \in A_i^T \mathcal{N}_{C_i}(A_i \hat{x}) \subseteq \mathcal{N}_{A_i^{-1} C_i}(\hat{x})$  for all  $i$ , where the set inclusion follows directly from the definition of normal cone. Consequently, it holds that

$$\bar{v} - \hat{x} = \mathbf{A}^T \mathbf{y}^* = \sum_{i=1}^{\ell} A_i^T y_i^* \in \sum_{i=1}^{\ell} \mathcal{N}_{A_i^{-1} C_i}(\hat{x}) \subseteq \mathcal{N}_{\bigcap_{i=1}^{\ell} A_i^{-1} C_i}(\hat{x}),$$

where the last set inclusion follows directly from definition. The above display shows that  $\hat{x}$  is the projection of  $\bar{v}$  onto  $\bigcap_{i=1}^{\ell} A_i^{-1} C_i$ , and hence we have  $\hat{x} = x^*$  as desired.  $\square$

Now, we derive the Dykstra-type projection algorithm by applying the (proximal) CGD method to solve (3.6). Set the starting point

$$(3.7) \quad \mathbf{y}^0 = (y_1^0, \dots, y_\ell^0) = (0, \dots, 0).$$

For any  $t \geq 0$  and any  $i \in \{1, \dots, \ell\}$ , for notational simplicity, write

$$(3.8) \quad \tilde{\mathbf{y}}_{i-1}^{t+1} := (y_1^{t+1}, \dots, y_{i-1}^{t+1}, y_i^t, \dots, y_\ell^t).$$

By applying the (proximal) CGD in a cyclic order to (3.6) with a proximal term induced by  $\gamma_i I - A_i A_i^T$ , where  $\gamma_i = \lambda_{\max}(A_i A_i^T)$ , we obtain the following update formula for  $y_i$ ,  $i = 1, \dots, \ell$ :

$$(3.9) \quad y_i^{t+1} := \text{Arg min}_{y_i \in \mathbb{R}^{m_i}} \left\{ \langle \nabla_{y_i} g(\tilde{\mathbf{y}}_{i-1}^{t+1}), y_i - y_i^t \rangle + \frac{\gamma_i}{2} \|y_i - y_i^t\|^2 + \sigma_{C_i}(y_i) \right\}.$$

Following a similar argument in [27, pp. 33–34] and using (3.7)–(3.9), one can derive the Dykstra-type projection algorithm for solving (1.2) as we present in Algorithm 3.1. In particular, one can show by induction that

$$(3.10) \quad x^t := x_\ell^t = \bar{v} - \mathbf{A}^T \mathbf{y}^t \quad \text{for all } t \geq 0.$$

The detailed derivation is given in Appendix A.

One can see that Algorithm 3.1 reduces to Algorithm 2.1 if each  $A_i = I$ . As discussed above, Algorithm 3.1 is exactly a CGD method for solving (3.6). The following proposition collects some properties concerning the dual iterates  $\{\mathbf{y}^t\}$  in Algorithm 3.1, which are immediate consequences of known results on CGD; see, for example, [38]. We list them here together with their simple proofs for the convenience of our subsequent convergence rate analysis.

---

**Algorithm 3.1.** Dykstra-type projection algorithm for BA-MSF problem (1.2).

---

Step 1. Choose  $y_i^0 = 0 \in \mathbb{R}^{m_i}$  for  $i = 1, \dots, \ell$ ,  $x^0 = \bar{v} \in \mathbb{R}^n$  and set  $\gamma_i = \lambda_{\max}(A_i^T A_i)$  for  $i = 1, \dots, \ell$ . Set  $\mathbf{y}^0 = (y_1^0, \dots, y_\ell^0)$ ,  $x_\ell^0 = x^0$ , and  $t = 0$ .

Step 2. Set  $x_0^{t+1} = x_\ell^t$ . For  $i = 1, \dots, \ell$ , compute

$$(3.11) \quad \begin{aligned} x_i^{t+1} &= (I - \gamma_i^{-1} A_i^T A_i) x_{i-1}^{t+1} + \gamma_i^{-1} A_i^T \text{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}), \\ y_i^{t+1} &= y_i^t + \gamma_i^{-1} A_i x_{i-1}^{t+1} - \gamma_i^{-1} \text{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}). \end{aligned}$$

Step 3. Set  $x^{t+1} = x_\ell^{t+1}$  and  $\mathbf{y}^{t+1} = (y_1^{t+1}, \dots, y_\ell^{t+1})$ . Update  $t \leftarrow t + 1$  and go to Step 2.

---

**PROPOSITION 3.2.** Consider (1.2). Let the function  $d$  be given in (3.6) and  $\{\mathbf{y}^t\}$  be the sequence generated by Algorithm 3.1. Then the following statements hold:

(i) For each  $i = 1, \dots, \ell$ , it holds that

$$d(\tilde{\mathbf{y}}_i^{t+1}) - d(\tilde{\mathbf{y}}_{i-1}^{t+1}) \leq \frac{1}{2} \Delta_i^t \leq -\frac{\gamma_i}{2} \|y_i^{t+1} - y_i^t\|^2,$$

where  $\Delta_i^t := \langle \nabla_{y_i} g(\tilde{\mathbf{y}}_{i-1}^{t+1}), y_i^{t+1} - y_i^t \rangle + \sigma_D(\tilde{\mathbf{y}}_i^{t+1}) - \sigma_D(\tilde{\mathbf{y}}_{i-1}^{t+1})$ ,  $\tilde{\mathbf{y}}_{i-1}^{t+1}$  is as in (3.8) for  $i = 1, \dots, \ell$  and  $\tilde{\mathbf{y}}_\ell^{t+1} := \mathbf{y}^{t+1}$ .

(ii) There exists  $c > 0$  such that  $d(\mathbf{y}^{t+1}) - d(\mathbf{y}^t) \leq -c \|\mathbf{y}^{t+1} - \mathbf{y}^t\|^2$ .

(iii)  $\lim_{t \rightarrow \infty} \|\mathbf{y}^{t+1} - \mathbf{y}^t\| = 0$ .

(iv)  $\{\mathbf{y}^t\}$  is a stationary sequence, i.e.,  $\lim_{t \rightarrow \infty} \text{dist}(\mathbf{0}, \partial d(\mathbf{y}^t)) = 0$ .

(v) Every accumulation point of  $\{\mathbf{y}^t\}$  is a solution of (3.6).

*Proof.* We first note from the definition of  $\tilde{\mathbf{y}}_{i-1}^{t+1}$  in (3.8), the definition of  $\mathbf{y}_i^{t+1}$  in (3.9), and the strong convexity of the objective in (3.9) that for  $i = 1, \dots, \ell$ ,

$$(3.12) \quad \langle \nabla_{y_i} g(\tilde{\mathbf{y}}_{i-1}^{t+1}), y_i^{t+1} - y_i^t \rangle + \frac{\gamma_i}{2} \|y_i^{t+1} - y_i^t\|^2 + \sigma_D(\tilde{\mathbf{y}}_i^{t+1}) \leq \sigma_D(\tilde{\mathbf{y}}_{i-1}^{t+1}) - \frac{\gamma_i}{2} \|y_i^{t+1} - y_i^t\|^2.$$

In addition, if we let  $z_i^t := \sum_{j=1}^{i-1} A_j^T y_j^{t+1} + \sum_{j=i+1}^{\ell} A_j^T y_j^t - \bar{v}$ , then

$$\begin{aligned} d(\tilde{\mathbf{y}}_i^{t+1}) &= g(\tilde{\mathbf{y}}_i^{t+1}) + \sigma_D(\tilde{\mathbf{y}}_i^{t+1}) = \frac{1}{2} \|A_i^T y_i^{t+1} + z_i^t\|^2 - \frac{1}{2} \|\bar{v}\|^2 + \sigma_D(\tilde{\mathbf{y}}_i^{t+1}) \\ &\stackrel{(a)}{=} \frac{1}{2} \|A_i^T y_i^t + z_i^t\|^2 - \frac{1}{2} \|\bar{v}\|^2 + \langle \nabla_{y_i} g(\tilde{\mathbf{y}}_{i-1}^{t+1}), y_i^{t+1} - y_i^t \rangle + \frac{1}{2} \|A_i^T (y_i^{t+1} - y_i^t)\|^2 + \sigma_D(\tilde{\mathbf{y}}_i^{t+1}) \\ &\leq \frac{1}{2} \|A_i^T y_i^t + z_i^t\|^2 - \frac{1}{2} \|\bar{v}\|^2 + \langle \nabla_{y_i} g(\tilde{\mathbf{y}}_{i-1}^{t+1}), y_i^{t+1} - y_i^t \rangle + \frac{\gamma_i}{2} \|y_i^{t+1} - y_i^t\|^2 + \sigma_D(\tilde{\mathbf{y}}_i^{t+1}) \\ &\stackrel{(b)}{=} d(\tilde{\mathbf{y}}_{i-1}^{t+1}) + \Delta_i^t + \frac{\gamma_i}{2} \|y_i^{t+1} - y_i^t\|^2 \stackrel{(c)}{\leq} d(\tilde{\mathbf{y}}_{i-1}^{t+1}) - \frac{\gamma_i}{2} \|y_i^{t+1} - y_i^t\|^2, \end{aligned}$$

where (a) holds since  $A_i(z_i^t + A_i^T y_i^t) = \nabla_{y_i} g(\tilde{\mathbf{y}}_{i-1}^{t+1})$ , (b) follows from the definition of  $\Delta_i^t$ , and (c) follows from (3.12) and the definition of  $\Delta_i^t$ . This proves item (i). Item (ii) is a direct consequence of (i), and item (iii) follows from item (ii) (since  $\inf d > -\infty$  by Proposition 3.1).

For item (iv), from the first-order optimality condition of (3.9), we see that

$$\begin{aligned} 0 &\in \nabla_{y_i} g(\tilde{\mathbf{y}}_{i-1}^{t+1}) + \gamma_i (y_i^{t+1} - y_i^t) + \partial \sigma_{C_i}(y_i^{t+1}) \\ &= [\gamma_i (y_i^{t+1} - y_i^t) + (\nabla_{y_i} g(\tilde{\mathbf{y}}_{i-1}^{t+1}) - \nabla_{y_i} g(\mathbf{y}^{t+1}))] + \nabla_{y_i} g(\mathbf{y}^{t+1}) + \partial \sigma_{C_i}(y_i^{t+1}). \end{aligned}$$

Using the fact that  $\mathbf{y}^{t+1} - \mathbf{y}^t \rightarrow 0$  from item (iii), the uniform continuity of  $\nabla g$  on  $\mathbb{R}^m$  and the definitions  $\mathbf{y}^{t+1} = (y_1^{t+1}, \dots, y_\ell^{t+1})$  and  $\tilde{\mathbf{y}}_{i-1}^{t+1} = (y_1^{t+1}, \dots, y_{i-1}^{t+1}, y_i^t, \dots, y_\ell^t)$



(see (3.8)), the above display implies that  $\{\mathbf{y}^t\}$  is a stationary sequence. Finally, item (v) follows immediately from item (iv) and the closedness of the subdifferential as a set-valued mapping [35, Theorem 24.4].  $\square$

Proposition 3.2(v) is useful only when accumulation points exist. In the case when each  $A_i = I$  in (1.2), it was shown in [28, Lemma 4.6] that  $d$  is level-bounded if  $\bigcap_{i=1}^\ell \text{int } C_i \neq \emptyset$ , which would further imply the boundedness of  $\{\mathbf{y}^t\}$  and hence the existence of accumulation points. However, in general, it can happen that  $\bigcap_{i=1}^\ell \text{int } C_i = \emptyset$  and the sequence  $\{\mathbf{y}^t\}$  can be unbounded. Fortunately, Proposition 3.2(iv) states that  $\{\mathbf{y}^t\}$  is *always* a stationary sequence. Below, as in [4] which studied the case when  $A_i = I$  for all  $i$ , we will show that every stationary sequence of the function  $d$  in (3.6) is minimizing, under mild conditions on the set of intersection. Recall from [5, Definition 4.2.2] that proper closed convex functions with all stationary sequences being minimizing are said to be asymptotically well-behaved (AWB).

PROPOSITION 3.3 (AWB property of  $d$ ). *Consider (1.2) and assume the condition  $\bigcap_{i=1}^\ell A_i^{-1} \text{ri } C_i \neq \emptyset$ . Let  $d$  be given in (3.6) and  $q(\mathbf{y}) := \inf_x \psi(x, \mathbf{y})$  with  $\psi$  defined in (3.2). Then  $\mathbf{0} \in \text{ri dom } q$ . Moreover, it holds that  $d(\mathbf{y}) = q^*(\mathbf{y})$  for all  $\mathbf{y}$  and*

$$(3.13) \quad \text{Arg min } d = \partial q_E(\mathbf{0}) + E^\perp \neq \emptyset,$$

where  $q_E(\mathbf{y}) := q(\text{Proj}_E(\mathbf{y}))$  and

$$(3.14) \quad E := \text{span}(\text{dom } q) = \text{span}(D - \text{Range}(\mathbf{A}))$$

with  $D$  and  $\mathbf{A}$  given in (3.1). Furthermore, every stationary sequence  $\{\mathbf{z}^t\}$  of  $d$  satisfies  $d(\mathbf{z}^t) \rightarrow d^*$  and  $\text{dist}(\mathbf{z}^t, \text{Arg min } d) \rightarrow 0$ , where  $d^*$  is given in (3.6).

*Proof.* From the definition of  $q$ , we have  $q(\mathbf{y}) < +\infty$  if and only if there exists  $x \in \mathbb{R}^n$  such that  $A_i x + y_i \in C_i$  for all  $i$  (since  $\text{dom } f = \mathbb{R}^n$ ), which implies that  $\text{dom } q = D - \text{Range}(\mathbf{A})$ . According to [36, Exercise 2.45], the above display implies that

$$\text{ri dom } q = \text{ri } D - \text{Range}(\mathbf{A}).$$

Thus, the assumption  $\bigcap_{i=1}^\ell A_i^{-1} \text{ri } C_i \neq \emptyset$  implies  $\mathbf{0} \in \text{ri dom } q$ . It follows from (3.3), (3.6), and the definition of  $q$  that  $d(\mathbf{y}) = \psi^*(0, \mathbf{y}) = q^*(\mathbf{y})$  for all  $\mathbf{y}$ . The desired conclusions now follow from [4, Theorem 3.1].  $\square$

We can now state a global convergence result that is informative even when  $\{\mathbf{y}^t\}$  is possibly unbounded. The proof follows directly from Propositions 3.2(iv) and 3.3.

THEOREM 3.4 (global convergence without boundedness condition). *Consider (1.2) and assume that  $\bigcap_{i=1}^\ell A_i^{-1} \text{ri } C_i \neq \emptyset$ . Let  $\{\mathbf{y}^t\}$  be the sequence generated by Algorithm 3.1. Then it holds that  $d(\mathbf{y}^t) \rightarrow d^*$  and  $\text{dist}(\mathbf{y}^t, \text{Arg min } d) \rightarrow 0$ , where  $d$  and  $d^*$  are defined in (3.6).*

The above theorem states that the sequences  $\{d(\mathbf{y}^t) - d^*\}$  and  $\{\text{dist}(\mathbf{y}^t, \text{Arg min } d)\}$  converge to zero. In this paper, we are interested in their convergence rates. To this end, we recall that the KL property is widely used for analyzing the convergence rate of first-order methods, with the rate usually depending explicitly on the *KL exponent* of a suitable potential function; see, for example, [1, 2, 3, 30, 41]. In the next section, we identify a large class of sets  $C_i$  such that the corresponding function  $d$  can be shown to satisfy the KL property with an *explicitly known exponent*. Moreover, we

show that the KL exponent can be chosen to be *uniform* over a neighborhood of the possibly *unbounded*  $\text{Arg min } d$ . These results will be further developed in section 5 to derive explicit convergence rates of  $\{d(\mathbf{y}^t) - d^*\}$  and  $\{\text{dist}(\mathbf{y}^t, \text{Arg min } d)\}$ .

#### 4. KL property and $C^{1,\alpha}$ -cone reducible sets.

**4.1.  $C^{1,\alpha}$ -cone reducible sets.** In this subsection, we introduce a class of sets for our subsequent convergence rate analysis. Specifically, we define the following notion of  $C^{1,\alpha}$ -cone reducibility for a closed set  $\mathfrak{D} \subseteq \mathbb{X}$ , which can be seen as a generalization of the notion of  $C^2$ -cone reducibility in [37]; here and throughout, we use the typefaces  $\mathbb{X}, \mathbb{Y}, \mathbb{Z}$ , etc., to denote finite-dimensional Hilbert spaces with their associated inner products and norms denoted by  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|$ , respectively, by an abuse of notation. We also let  $\mathcal{T}^*$  denote the adjoint of a linear operator  $\mathcal{T} : \mathbb{X} \rightarrow \mathbb{Z}$ , defined via  $\langle \mathcal{T}x, z \rangle = \langle x, \mathcal{T}^*z \rangle$  for all  $x \in \mathbb{X}$  and  $z \in \mathbb{Z}$ ; and  $\|\mathcal{T}\|$  is the operator norm of  $\mathcal{T}$ .

**DEFINITION 4.1** ( $C^{1,\alpha}$ -cone reducible sets). *Let  $\alpha \in (0, 1]$ . A closed set  $\mathfrak{D} \subseteq \mathbb{X}$  is said to be  $C^{1,\alpha}$ -cone reducible at  $\hat{x} \in \mathfrak{D}$  if there exist positive constant  $\rho$ , a closed convex pointed cone  $K$  in a finite-dimensional Hilbert space  $\mathbb{Y}$ , and a mapping  $\Xi : \mathbb{X} \rightarrow \mathbb{Y}$  that maps  $\hat{x}$  to 0 and is continuously differentiable in  $B(\hat{x}, \rho)$  such that  $D\Xi(\hat{x})$  is surjective,<sup>4</sup> the mapping  $x \mapsto D\Xi(x)$  is  $\alpha$ -Hölder-continuous in  $B(\hat{x}, \rho)$  and*

$$\mathfrak{D} \cap B(\hat{x}, \rho) = \{x : \Xi(x) \in K\} \cap B(\hat{x}, \rho).$$

A closed set  $\mathfrak{D}$  is said to be  $C^{1,\alpha}$ -cone reducible if it is  $C^{1,\alpha}$ -cone reducible at every  $x \in \mathfrak{D}$ .

Clearly, every  $C^2$ -cone reducible set (see [37, Definition 3.1]) is  $C^{1,\alpha}$ -cone reducible for any  $\alpha \in (0, 1]$ . According to [37], all polyhedrons, the second-order cone, and the cone of positive semidefinite matrices are  $C^2$ -cone reducible sets; thus, they are also  $C^{1,\alpha}$ -cone reducible for any  $\alpha \in (0, 1]$ . The next lemma establishes a “calmness-type” property for the normal cone mapping of a  $C^{1,\alpha}$ -cone reducible closed convex set. This can be seen as an extension of [41, Theorem 4.4], which considered  $C^2$ -cone reducible sets. Our proof is also similar to that of [41, Theorem 4.4].

**LEMMA 4.2.** *Let  $\alpha \in (0, 1]$ , and let  $\mathfrak{D} \subseteq \mathbb{X}$  be a closed convex set that is  $C^{1,\alpha}$ -cone reducible at  $\hat{x} \in \mathfrak{D}$ . Let  $\hat{y} \in \mathcal{N}_{\mathfrak{D}}(\hat{x})$ . Then there exist  $\hat{\rho} > 0$ ,  $\hat{\delta} > 0$ , and  $\hat{\kappa} > 0$  such that*

$$(4.1) \quad \mathcal{N}_{\mathfrak{D}}(x) \cap B(\hat{y}, \hat{\delta}) \subseteq \mathcal{N}_{\mathfrak{D}}(\hat{x}) + \hat{\kappa}\|x - \hat{x}\|^\alpha B(0, 1) \quad \text{for all } x \in B(\hat{x}, \hat{\rho}).$$

*Proof.* Since  $\mathfrak{D}$  is  $C^{1,\alpha}$ -cone reducible at  $\hat{x}$ , there exist a positive constant  $\rho$ , closed convex pointed cone  $K \subseteq \mathbb{Y}$ , and continuously differentiable mapping  $\Xi : \mathbb{X} \rightarrow \mathbb{Y}$  satisfying the conditions in Definition 4.1 such that

$$(4.2) \quad \mathfrak{D} \cap B(\hat{x}, \rho) = \{x : \Xi(x) \in K\} \cap B(\hat{x}, \rho).$$

Since  $D\Xi(\hat{x})$  is surjective, one can choose a positive  $\tilde{\rho} \in (0, \rho)$  such that for any  $x \in B(\hat{x}, \tilde{\rho})$ , the linear mapping  $D\Xi(x)$  is surjective (meaning that  $D\Xi(x)D\Xi(x)^*$  is invertible) and that

$$(4.3) \quad \sup_{x \in B(\hat{x}, \tilde{\rho})} \|[D\Xi(x)D\Xi(x)^*]^{-1}\| =: \tau < \infty.$$

<sup>4</sup>Recall that  $D\Xi(x)$  is the derivative mapping of  $\Xi$  at an  $x \in \mathbb{X}$ , which is the linear mapping from  $\mathbb{X}$  to  $\mathbb{Y}$  such that  $[D\Xi(x)](p) = \lim_{t \downarrow 0} \frac{\Xi(x+tp) - \Xi(x)}{t}$  for all  $p \in \mathbb{X}$ .

The relation (4.2) means that

$$\delta_{\mathfrak{D}}(x) = \delta_K(\Xi(x)) \quad \text{for all } x \in B(\hat{x}, \tilde{\rho}).$$

Using the fact that  $D\Xi(x)$  is surjective whenever  $x \in B(\hat{x}, \tilde{\rho})$ , and invoking [36, Exercise 10.7], we deduce from the above display that

$$(4.4) \quad \mathcal{N}_{\mathfrak{D}}(x) = D\Xi(x)^* \mathcal{N}_K(\Xi(x)) \quad \text{for all } x \in \mathfrak{D} \cap B(\hat{x}, \tilde{\rho}).$$

Fix any  $\tilde{\delta} > 0$  and take

$$(4.5) \quad x \in \mathfrak{D} \cap B(\hat{x}, \tilde{\rho}), \quad y \in \mathcal{N}_{\mathfrak{D}}(x) \cap B(\hat{y}, \tilde{\delta}).$$

Relation (4.4) implies that there exists a  $v_x \in \mathcal{N}_K(\Xi(x))$  so that

$$(4.6) \quad y = D\Xi(x)^* v_x.$$

Recall that for any  $z_0 \in K$ , it holds that

$$(4.7) \quad \mathcal{N}_K(z_0) \stackrel{(a)}{=} \{u : \langle u, z - z_0 \rangle \leq 0 \ \forall z \in K\} \stackrel{(b)}{\subseteq} \{u : \langle u, z \rangle \leq 0 \ \forall z \in K\} \stackrel{(c)}{=} K^\circ,$$

where (a) follows from the definition of normal cone, and (b) and (c) hold because  $K$  is a closed convex cone. In particular, equality holds throughout (4.7) if  $z_0 = 0$ .

Now, for the  $v_x$  in (4.6), we have

$$(4.8) \quad D\Xi(\hat{x})^* v_x \stackrel{(a)}{\in} D\Xi(\hat{x})^* \mathcal{N}_K(\Xi(x)) \stackrel{(b)}{\subseteq} D\Xi(\hat{x})^* K^\circ \stackrel{(c)}{=} D\Xi(\hat{x})^* \mathcal{N}_K(\Xi(\hat{x})) \stackrel{(d)}{=} \mathcal{N}_{\mathfrak{D}}(\hat{x}),$$

where (a) holds because  $v_x \in \mathcal{N}_K(\Xi(x))$ , (b) follows from  $\Xi(x) \in K$  and (4.7), (c) holds thanks to  $\Xi(\hat{x}) = 0$ , and (d) follows from (4.4).

Next, for any  $x$  and  $y$  as in (4.5), by (4.6) and the surjectivity of  $D\Xi(x)$ , we have

$$(4.9) \quad v_x = [D\Xi(x)D\Xi(x)^*]^{-1} D\Xi(x)y =: H(x)D\Xi(x)y,$$

where  $H(x) := [D\Xi(x)D\Xi(x)^*]^{-1}$ .

Let  $L$  be the  $\alpha$ -Hölder continuity modulus of  $D\Xi(\cdot)$  over the set  $x \in \mathfrak{D} \cap B(\hat{x}, \tilde{\rho})$ . Then, for any  $x$  and  $y$  chosen as in (4.5), we have

$$\begin{aligned} \|v_x - v_{\hat{x}}\| &= \|[H(x)D\Xi(x)y - H(\hat{x})D\Xi(\hat{x})\hat{y}]\| \\ &\leq \|H(x)D\Xi(x)y - H(\hat{x})D\Xi(\hat{x})y\| + \|H(\hat{x})D\Xi(\hat{x})(y - \hat{y})\| \\ &\leq \|H(x)D\Xi(x) - H(\hat{x})D\Xi(\hat{x}) + H(x)D\Xi(\hat{x}) - H(\hat{x})D\Xi(\hat{x})\| \|y\| \\ &\quad + \|H(\hat{x})D\Xi(\hat{x})(y - \hat{y})\| \\ &\leq (\|H(x)\| \|D\Xi(x) - D\Xi(\hat{x})\| + \|H(x) - H(\hat{x})\| \|D\Xi(\hat{x})\|) \|y - \hat{y}\| + \hat{y} \\ &\quad + \|H(\hat{x})D\Xi(\hat{x})\| \|y - \hat{y}\| \\ &\leq (\tau L \|x - \hat{x}\|^\alpha + 2\tau \|D\Xi(\hat{x})\|) (\|\hat{y}\| + \|y - \hat{y}\|) + \tau \|D\Xi(\hat{x})\| \|y - \hat{y}\| \\ (4.10) \quad &\leq (\tau L \tilde{\rho}^\alpha + 2\tau \|D\Xi(\hat{x})\|) (\|\hat{y}\| + \tilde{\delta}) + \tau \|D\Xi(\hat{x})\| \tilde{\delta}, \end{aligned}$$

where the first equality is due to (4.9), the fourth inequality follows from (4.3) and the  $\alpha$ -Hölder continuity of the mapping  $D\Xi(\cdot)$ , and the last inequality follows from (4.5). Inequality (4.10) means that  $v_x$  is bounded whenever  $x$  and  $y$  are chosen as in (4.5), and we denote  $\tilde{\kappa} := \sup \|v_x\| < +\infty$ .

For any  $x$  and  $y$  chosen according to (4.5), it holds that

$$\text{dist}(y, \mathcal{N}_{\mathfrak{D}}(\hat{x})) \leq \|y - D\Xi(\hat{x})^*v_x\| = \|D\Xi(x)^*v_x - D\Xi(\hat{x})^*v_x\| \leq L\tilde{\kappa}\|x - \hat{x}\|^\alpha,$$

where the first inequality follows from (4.8), the equality follows from (4.6), and the last inequality follows from the  $\alpha$ -Hölder continuity of  $D\Xi$  and the definition of  $\tilde{\kappa}$ . Letting  $\hat{\rho} := \tilde{\rho}$ ,  $\hat{\delta} := \tilde{\delta}$ , and  $\hat{\kappa} := L\tilde{\kappa}$ , the above display implies (4.1).  $\square$

Equipped with Lemma 4.2, we are now ready to prove the main theorem of this subsection, which will be used for deriving KL exponent of the function  $d$  given in (3.6) in the next subsection. It can be seen as an extension of [41, Theorem 4.4], which studied  $C^2$ -cone reducible sets. Before stating the theorem, we first recall that a pair of closed convex sets  $\{\mathfrak{D}_1, \mathfrak{D}_2\}$  is said to be boundedly linearly regular (see [7, Definition 5.6]) at  $\hat{x} \in \mathfrak{D}_1 \cap \mathfrak{D}_2$  if for any bounded neighborhood  $\mathfrak{U}$  of  $\hat{x}$ , there exists  $c > 0$  such that

$$\text{dist}(x, \mathfrak{D}_1 \cap \mathfrak{D}_2) \leq c(\text{dist}(x, \mathfrak{D}_1) + \text{dist}(x, \mathfrak{D}_2)) \quad \text{for all } x \in \mathfrak{U}.$$

It is known that bounded linear regularity holds at any  $\hat{x} \in \mathfrak{D}_1 \cap \mathfrak{D}_2$  when  $\mathfrak{D}_1$  and  $\mathfrak{D}_2$  are polyhedral or when  $\mathfrak{D}_1$  is polyhedral and  $\mathfrak{D}_1 \cap \text{ri} \mathfrak{D}_2 \neq \emptyset$ ; see [8, Corollary 3].

**THEOREM 4.3.** *Let  $\mathfrak{D} \subseteq \mathbb{X}$  be a nonempty  $C^{1,\alpha}$ -cone reducible closed convex set with  $\alpha \in (0, 1]$ ,  $\mathcal{T} : \mathbb{X} \rightarrow \mathbb{Y}$  be a linear mapping, and let  $l : \mathbb{Y} \rightarrow \mathbb{R}$  be strongly convex on any compact convex set with locally Lipschitz gradient, and  $v \in \mathbb{X}$ . Consider*

$$h(x) := l(\mathcal{T}x) + \langle v, x \rangle + \sigma_{\mathfrak{D}}(x).$$

Then, for any  $\bar{x} \in \text{Arg min } h$ , it holds that

$$(4.11) \quad \bar{x} \in \mathcal{N}_{\mathfrak{D}}(\bar{w}) \quad \text{with } \bar{w} := -\mathcal{T}^*\nabla l(\mathcal{T}\bar{x}) - v.$$

Moreover, there exist positive constants  $\rho$ ,  $\delta$ , and  $\kappa$  such that

$$(4.12) \quad \mathcal{N}_{\mathfrak{D}}(w) \cap B(\bar{x}, \delta) \subseteq \mathcal{N}_{\mathfrak{D}}(\bar{w}) + \kappa\|w - \bar{w}\|^\alpha B(0, 1) \quad \text{for all } w \in B(\bar{w}, \rho).$$

Furthermore, if  $\{\mathcal{T}^{-1}\{\mathcal{T}\bar{x}\}, \mathcal{N}_{\mathfrak{D}}(-\mathcal{T}^*\nabla l(\mathcal{T}\bar{x}) - v)\}$  is boundedly linearly regular at  $\bar{x}$ , then  $h$  satisfies the KL property at  $\bar{x}$  with exponent  $\frac{1}{\alpha+1}$ .

*Proof.* The relation (4.11) follows from the same argument as in [41, Theorem 4.4]. The relation (4.12) follows from applying Lemma 4.2 to  $(\bar{x}, \bar{w})$  with  $\bar{x} \in \mathcal{N}_{\mathfrak{D}}(\bar{w})$ .

We now prove the alleged KL property at  $\bar{x}$ . We first show that (4.12) implies

$$(4.13) \quad \text{dist}(x, (\partial\sigma_{\mathfrak{D}})^{-1}(\bar{w})) \leq \kappa \text{dist}(\bar{w}, \partial\sigma_{\mathfrak{D}}(x) \cap B(\bar{w}, \rho))^\alpha \quad \text{for all } x \in B(\bar{x}, \delta).$$

Our argument is similar to the proof of [25, Theorem 3H.3]. Specifically, notice that the above display holds trivially if  $\partial\sigma_{\mathfrak{D}}(x) \cap B(\bar{w}, \rho) = \emptyset$ . Now, consider the case  $\partial\sigma_{\mathfrak{D}}(x) \cap B(\bar{w}, \rho) \neq \emptyset$ . Using  $(\partial\sigma_{\mathfrak{D}})^{-1} = \mathcal{N}_{\mathfrak{D}}$  (see [36, Example 11.4]), we see that

$$w \in \partial\sigma_{\mathfrak{D}}(x) \cap B(\bar{w}, \rho), \quad x \in B(\bar{x}, \delta) \iff w \in B(\bar{w}, \rho), \quad x \in \mathcal{N}_{\mathfrak{D}}(w) \cap B(\bar{x}, \delta).$$

Consequently, for any  $x, w$  satisfying  $w \in \partial\sigma_{\mathfrak{D}}(x) \cap B(\bar{w}, \rho)$  and  $x \in B(\bar{x}, \delta)$ , we have from the above display and (4.12) that

$$\text{dist}(x, (\partial\sigma_{\mathfrak{D}})^{-1}(\bar{w})) \leq \kappa\|w - \bar{w}\|^\alpha.$$

Taking infimum with respect to  $w \in \partial\sigma_{\mathfrak{D}}(x) \cap B(\bar{w}, \rho)$ , we obtain (4.13).

In view of [25, Exercise 3H.4], the inequality (4.13) implies that there exists  $\delta' \in (0, \delta]$  such that

$$(4.14) \quad \text{dist}(x, (\partial\sigma_{\mathfrak{D}})^{-1}(\bar{w})) \leq \kappa \text{dist}(\bar{w}, \partial\sigma_{\mathfrak{D}}(x))^\alpha \quad \text{for all } x \in B(\bar{x}, \delta').$$

Now, define a proper closed convex function

$$F(x) := \sigma_{\mathfrak{D}}(x) - \langle \bar{w}, x \rangle.$$

Using (2.1), the following equivalence holds:

$$(4.15) \quad z \in (\partial\sigma_{\mathfrak{D}})^{-1}(\bar{w}) \Leftrightarrow \bar{w} \in \partial\sigma_{\mathfrak{D}}(z) \Leftrightarrow 0 \in \partial F(z) \Leftrightarrow z \in (\partial F)^{-1}(0).$$

Using (4.15), we can rewrite (4.14) as

$$(4.16) \quad \text{dist}(x, (\partial F)^{-1}(0)) \leq \kappa \text{dist}(0, \partial F(x))^\alpha \quad \text{for all } x \in B(\bar{x}, \delta').$$

Then, we have for any  $x \in B(\bar{x}, \delta')$  that

$$F(x) - F(\bar{x}) \leq \text{dist}(0, \partial F(x)) \text{dist}(x, (\partial F)^{-1}(0)) \leq \kappa \text{dist}(0, \partial F(x))^{1+\alpha},$$

where the first inequality follows from (2.2),<sup>5</sup> the second inequality is due to (4.16). The above display implies that  $F$  satisfies the KL property at  $\bar{x}$  with exponent  $\frac{1}{1+\alpha}$ , which by [12, Theorem 5] also implies the existence of  $\bar{\kappa} > 0$  such that

$$\text{dist}(x, (\partial F)^{-1}(0)) \leq \bar{\kappa} (F(x) - F(\bar{x}))^{1-\frac{1}{1+\alpha}} \quad \text{for all } x \in B(\bar{x}, \delta').$$

Using the definition of  $F$  and (4.15), the above display can be equivalently written as

$$(4.17) \quad \begin{aligned} \text{dist}(x, \mathcal{N}_{\mathfrak{D}}(\bar{w})) &= \text{dist}(x, (\partial\sigma_{\mathfrak{D}})^{-1}(\bar{w})) = \text{dist}(x, (\partial F)^{-1}(0)) \\ &\leq \bar{\kappa} (\sigma_{\mathfrak{D}}(x) - \sigma_{\mathfrak{D}}(\bar{x}) - \langle \bar{w}, x - \bar{x} \rangle)^{1-\frac{1}{1+\alpha}} \quad \text{for all } x \in B(\bar{x}, \delta'), \end{aligned}$$

where the first equality follows from [36, Example 11.4]. Next, notice that we have  $\mathcal{T}x = \mathcal{T}\bar{x}$  for all  $x \in \text{Argmin } h$  thanks to the strict convexity of  $l$ ,<sup>6</sup> and moreover,

$$\begin{aligned} \text{Argmin } h &= \{x : 0 \in \partial h(x)\} = \{x : \mathcal{T}x = \mathcal{T}\bar{x}, x \in \mathcal{N}_{\mathfrak{D}}(-\mathcal{T}^* \nabla l(\mathcal{T}\bar{x}) - v)\} \\ &= \{x : \mathcal{T}x = \mathcal{T}\bar{x}, x \in \mathcal{N}_{\mathfrak{D}}(\bar{w})\}. \end{aligned}$$

Also, for any bounded convex neighborhood  $U \subseteq B(\bar{x}, \delta')$  of  $\bar{x}$ , we have for any  $x \in U$  that

$$(4.18) \quad \begin{aligned} \|\mathcal{T}x - \mathcal{T}\bar{x}\|^\frac{1+\alpha}{\alpha} &= \|\mathcal{T}x - \mathcal{T}\bar{x}\|^\frac{1}{\alpha-1} \|\mathcal{T}x - \mathcal{T}\bar{x}\|^2 \leq \left( \sup_{u \in U} \|\mathcal{T}u - \mathcal{T}\bar{x}\|^\frac{1}{\alpha-1} \right) \|\mathcal{T}x - \mathcal{T}\bar{x}\|^2 \\ &\leq \bar{M} (l(\mathcal{T}x) - l(\mathcal{T}\bar{x}) - \langle \mathcal{T}^* \nabla l(\mathcal{T}\bar{x}), x - \bar{x} \rangle), \end{aligned}$$

<sup>5</sup>Note that  $(\partial F)^{-1}(0) \neq \emptyset$  since it contains  $\bar{x}$  in view of (4.11), (4.15), and the fact that  $(\partial\sigma_{\mathfrak{D}})^{-1} = \mathcal{N}_{\mathfrak{D}}$  (see [36, Example 11.4]).

<sup>6</sup>Indeed, suppose that  $\mathcal{T}x \neq \mathcal{T}\bar{x}$  for some  $x \in \text{Argmin } h$ . Then we have the following contradiction:

$$\inf h = (h(x) + h(\bar{x}))/2 > l((\mathcal{T}x + \mathcal{T}\bar{x})/2) + \langle v, (x + \bar{x})/2 \rangle + \sigma_{\mathfrak{D}}((x + \bar{x})/2) \geq \inf h,$$

where the first inequality follows from the strict convexity of  $l$ , convexity of  $\langle v, \cdot \rangle + \sigma_{\mathfrak{D}}(\cdot)$ , and the assumption that  $\mathcal{T}x \neq \mathcal{T}\bar{x}$ , and the second inequality follows from the fact that  $(x + \bar{x})/2 \in \text{dom } h$  as  $h$  is convex.

where the last inequality holds for some  $\bar{M} > 0$  thanks to the strongly convexity of  $l$  on  $U$  and the fact that  $\sup_{u \in U} \|\mathcal{T}u - \mathcal{T}\bar{x}\|^{\frac{1}{\alpha}-1} < \infty$  (thanks to  $\alpha \in (0, 1]$ ). Now, using the bounded linear regularity condition and the Hoffman's error bound, there exist  $\beta_1 > 0$  and  $\beta_2 > 0$  such that whenever  $x \in U$ , we have

$$(4.19) \quad \text{dist}(x, \mathcal{T}^{-1}\{\mathcal{T}\bar{x}\} \cap \mathcal{N}_{\mathfrak{D}}(\bar{w})) \leq \beta_1 [\text{dist}(x, \mathcal{T}^{-1}\{\mathcal{T}\bar{x}\}) + \text{dist}(x, \mathcal{N}_{\mathfrak{D}}(\bar{w}))],$$

$$(4.20) \quad \text{dist}(x, \mathcal{T}^{-1}\{\mathcal{T}\bar{x}\}) \leq \beta_2 \|\mathcal{T}x - \mathcal{T}\bar{x}\|.$$

Then, it follows that for any  $x \in U$ ,

$$(4.21) \quad \begin{aligned} \text{dist}(x, \text{Arg min } h) &= \text{dist}(x, \mathcal{T}^{-1}\{\mathcal{T}\bar{x}\} \cap \mathcal{N}_{\mathfrak{D}}(\bar{w})) \leq \beta_1 [\beta_2 \|\mathcal{T}x - \mathcal{T}\bar{x}\| + \text{dist}(x, \mathcal{N}_{\mathfrak{D}}(\bar{w}))] \\ &\leq \beta_1 [\beta_2 \bar{M}^{\frac{\alpha}{\alpha+1}} (l(\mathcal{T}x) - l(\mathcal{T}\bar{x}) - \langle \mathcal{T}^* \nabla l(\mathcal{T}\bar{x}), x - \bar{x} \rangle)^{1-\frac{1}{\alpha+1}} \\ &\quad + \bar{\kappa} (\sigma_{\mathfrak{D}}(x) - \sigma_{\mathfrak{D}}(\bar{x}) - \langle \bar{w}, x - \bar{x} \rangle)^{1-\frac{1}{\alpha+1}}], \end{aligned}$$

where the first inequality follows from (4.19)–(4.20), and the last inequality follows from (4.17) and (4.18). Finally, notice that

$$(4.22) \quad \begin{aligned} h(x) - h(\bar{x}) &= l(\mathcal{T}x) - l(\mathcal{T}\bar{x}) - \langle \mathcal{T}^* \nabla l(\mathcal{T}\bar{x}), x - \bar{x} \rangle + \sigma_{\mathfrak{D}}(x) - \sigma_{\mathfrak{D}}(\bar{x}) - \langle \bar{w}, x - \bar{x} \rangle. \end{aligned}$$

Combining (4.21)–(4.22) and the inequality  $a^p + b^p \leq 2^{1-p}(a+b)^p$  for any  $a, b \geq 0$  and  $p \in (0, 1]$ , we deduce that there exists positive constant  $c$  such that

$$\text{dist}(x, \text{Arg min } h) \leq c(h(x) - h(\bar{x}))^{1-\frac{1}{\alpha+1}} \quad \text{for all } x \in U.$$

The desired result now follows upon invoking [12, Theorem 5].  $\square$

Verifying  $C^{1,\alpha}$ -cone reducibility directly from the definition can be nontrivial. The following proposition is handy for checking whether a set is  $C^{1,\alpha}$ -cone reducible. It is an analogue of [37, Proposition 3.2], which studied  $C^2$ -cone reducible sets.

**PROPOSITION 4.4.** *Let  $V := \{x : G(x) \in K\}$ , where  $G : \mathbb{X} \rightarrow \mathbb{Y}$  is continuously differentiable with locally  $\alpha$ -Hölder-continuous derivative for some  $\alpha \in (0, 1]$ , and let  $K \subseteq \mathbb{Y}$  be a closed convex set. If  $K$  is  $C^{1,\alpha}$ -cone reducible at  $y_0 = G(x_0)$  and*

$$DG(x_0)\mathbb{X} + \text{lin } T_K(y_0) = \mathbb{Y},$$

where  $T_K(y_0)$  is the tangent cone of  $K$  at  $y_0$  and  $\text{lin } T_K(y_0) := T_K(y_0) \cap -T_K(y_0)$ , then  $V$  is  $C^{1,\alpha}$ -cone reducible at  $x_0$ .

*Proof.* The proof follows the same argument as in [37, Proposition 3.2].  $\square$

When  $\mathbb{X} = \mathbb{R}^n$ ,  $\mathbb{Y} = \mathbb{R}^m$ ,  $K = -\mathbb{R}_+^m$ , and  $G(x) = (g_1(x), \dots, g_m(x))$  with each  $g_i$  being continuously differentiable with locally  $\alpha$ -Hölder-continuous derivative, since  $K = -\mathbb{R}_+^m$  is  $C^2$ -cone reducible [37], we deduce from Proposition 4.4 that the set

$$V := \{x \in \mathbb{R}^n : g_i(x) \leq 0, i = 1, \dots, m\}$$

is  $C^{1,\alpha}$ -cone reducible at  $x_0$  if  $\{\nabla g_i(x_0) : i \in \mathcal{I}\}$  is linearly independent, where  $\mathcal{I} := \{i : g_i(x_0) = 0\}$ . As further concrete examples utilizing Proposition 4.4, we show below that the  $p$ -cone and the  $p$ -norm ball with  $p \in (1, \infty)$  are  $C^{1,\alpha}$ -cone reducible for some  $\alpha \in (0, 1]$ . Recall that the  $p$ -cone  $\mathcal{K}_p^{n+1} \subseteq \mathbb{R}^{n+1}$  is defined as

$$\mathcal{K}_p^{n+1} := \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : r \geq \|x\|_p\}.$$

*Example 4.1* ( $\mathcal{K}_p^{n+1}$ ,  $p \in (1, \infty)$ , is  $C^{1,\alpha}$ -cone reducible). Notice that  $\mathcal{K}_p^{n+1} = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : G(x, r) \leq 0\}$  with  $G(x, r) := \|x\|_p - r$ . As pointed out in [37], any closed convex cone is  $C^2$ -cone reducible at the origin and its relative interior, and the corresponding  $\Xi$  can be chosen to be linear. So, we need only show that  $\mathcal{K}_p^{n+1}$  is  $C^{1,\alpha}$ -cone reducible at every nonzero boundary point. Note that when  $x \neq 0$ , we have at any  $(x, r)$  satisfying  $r \geq \|x\|_p$  that

$$\nabla G(x, r) = \|x\|_p^{1-p} \cdot [\text{sgn}(x_1)|x_1|^{p-1} \ \cdots \ \text{sgn}(x_n)|x_n|^{p-1} \ -\|x\|_p^{p-1}]^T.$$

Thus, for any  $(x, r) \neq 0$  on the boundary of  $\mathcal{K}_p^{n+1}$ , one has  $r = \|x\|_p$  and  $\nabla G(x, \|x\|_p) \neq 0$ ; hence,  $\{\nabla G(x, \|x\|_p)\}$  is linearly independent.

From the discussion preceding this example, it now remains to show that  $\nabla G$  is locally Hölder continuous at any nonzero boundary point of  $\mathcal{K}_p^{n+1}$ . Using the display above, it is routine to check that  $\nabla G$  is locally  $(p-1)$ -Hölder continuous when  $p \in (1, 2)$  and is locally Lipschitz continuous (and hence 1-Hölder continuous) when  $p \in [2, \infty)$ . Therefore,  $\mathcal{K}_p^{n+1}$  is  $C^{1,\alpha}$ -cone reducible for any  $p \in (1, +\infty)$  with  $\alpha = \min\{1, p - 1\}$ .

*Example 4.2* (The  $p$ -norm ball,  $p \in (1, \infty)$ , is  $C^{1,\alpha}$ -cone reducible). Similar to the discussion in Example 4.1, one can show that the  $p$ -norm ball  $\{x \in \mathbb{R}^n : \|x - x_0\|_p \leq \beta\}$  (center at  $x_0$  with radius  $\beta > 0$ ) is  $C^{1,\alpha}$ -cone reducible with  $\alpha = \min\{1, p - 1\}$ .

**4.2. KL properties of the function  $d$  in (3.6).** In this subsection, we show that the function  $d$  in (3.6) satisfies the KL property with an explicit exponent under suitable assumptions on (1.2). Specifically, we consider the following assumption.

*Assumption 4.1.* Consider (1.2). Suppose the following conditions hold.

- (i) Each  $C_i$  is a  $C^{1,\alpha}$ -cone reducible closed convex set with  $\alpha \in (0, 1]$ ;<sup>7</sup>
- (ii)  $\bigcap_{i=1}^\ell A_i^{-1} \text{ri } C_i \neq \emptyset$ ;
- (iii)  $0 \in x^* - \bar{v} + \text{ri } \partial(\sum_{i=1}^\ell \delta_{A_i^{-1} C_i})(x^*)$ , where  $x^*$  is the unique solution of (1.2).

The following proposition gives the KL exponent of the function  $d$  in (3.6).

**PROPOSITION 4.5** (KL exponent of  $d$ ). *Consider (1.2). Suppose that Assumption 4.1 holds. Then the function  $d$  in (3.6) is a KL function with exponent  $\frac{1}{\alpha+1}$ .*

*Proof.* Fix any  $\mathbf{y}^* = (y_1^*, \dots, y_\ell^*) \in \text{Arg min } d$ . Then Proposition 3.1 implies that

$$(4.23) \quad x^* = \bar{v} - \mathbf{A}^T \mathbf{y}^*,$$

where  $x^*$  is the unique solution of (1.2).

Recall from (1.2) that  $f(x) = \frac{1}{2} \|x - \bar{v}\|^2$ . We first show that

$$(4.24) \quad \{(\mathbf{A}^T)^{-1} \{\mathbf{A}^T \mathbf{y}^*\}, \mathcal{N}_D(-\mathbf{A} \nabla f(\mathbf{A}^T \mathbf{y}^*))\}$$

is boundedly linearly regular at  $\mathbf{y}^*$ . To this end, we first observe that

$$(4.25) \quad (\mathbf{A}^T)^{-1} \{\mathbf{A}^T \mathbf{y}^*\} = (y_1^*, \dots, y_\ell^*) + \ker \mathbf{A}^T.$$

Moreover, using (4.23) and noting  $\nabla f(\mathbf{A}^T \mathbf{y}^*) = \mathbf{A}^T \mathbf{y}^* - \bar{v}$ , we see that

$$(4.26) \quad \mathcal{N}_D(-\mathbf{A} \nabla f(\mathbf{A}^T \mathbf{y}^*)) = \mathcal{N}_D(\mathbf{A} x^*) = \mathcal{N}_{C_1}(A_1 x^*) \times \cdots \times \mathcal{N}_{C_\ell}(A_\ell x^*).$$

<sup>7</sup>Using this, one can verify directly from definition that  $D$  is  $C^{1,\alpha}$ -cone reducible for the same  $\alpha$ .

Furthermore, we deduce using Assumption 4.1(iii) that

$$\bar{v} - x^* \in \text{ri} \partial \left( \sum_{i=1}^{\ell} \delta_{A_i^{-1}C_i} \right) (x^*) = \text{ri} \left( \sum_{i=1}^{\ell} A_i^T \mathcal{N}_{C_i}(A_i x^*) \right) = \sum_{i=1}^{\ell} A_i^T \text{ri} \mathcal{N}_{C_i}(A_i x^*),$$

where the first equality follows from Assumption 4.1(ii) and Theorems 23.8 and 23.9 of [35], and the second equality follows from [35, Theorem 6.6]. Combining the above display with (4.26), we conclude that there exists

$$(4.27) \quad \mathbf{u}^* = (u_1^*, \dots, u_{\ell}^*) \in \text{ri} \mathcal{N}_{C_1}(A_1 x^*) \times \cdots \times \text{ri} \mathcal{N}_{C_{\ell}}(A_{\ell} x^*) = \text{ri} \mathcal{N}_D(-\mathbf{A} \nabla f(\mathbf{A}^T \mathbf{y}^*))$$

such that  $x^* - \bar{v} + \mathbf{A}^T \mathbf{u}^* = 0$ . This last equation together with (4.23) implies that  $\mathbf{u}^* - \mathbf{y}^* \in \ker \mathbf{A}^T$ . Consequently,

$$(4.28) \quad \mathbf{u}^* = \mathbf{y}^* + (\mathbf{u}^* - \mathbf{y}^*) \in \mathbf{y}^* + \ker \mathbf{A}^T.$$

Using (4.25), (4.27), and (4.28), we see that  $\mathbf{u}^* \in (\mathbf{A}^T)^{-1} \{ \mathbf{A}^T \mathbf{y}^* \} \cap \text{ri} \mathcal{N}_D(-\mathbf{A} \nabla f(\mathbf{A}^T \mathbf{y}^*))$ , which implies that

$$(\mathbf{A}^T)^{-1} \{ \mathbf{A}^T \mathbf{y}^* \} \cap \text{ri} \mathcal{N}_D(-\mathbf{A} \nabla f(\mathbf{A}^T \mathbf{y}^*)) \neq \emptyset.$$

This and the polyhedrality of  $(\mathbf{A}^T)^{-1} \{ \mathbf{A}^T \mathbf{y}^* \}$  give the bounded linear regularity for (4.24) at  $\mathbf{y}^*$  (see [8, Corollary 3]). We can then deduce from Assumption 4.1(i) and Theorem 4.3 that  $d$  satisfies the KL property with exponent  $\frac{1}{\alpha+1}$  at  $\mathbf{y}^*$ . The desired conclusion now follows from the arbitrariness of  $\mathbf{y}^* \in \text{Argmin} d$  and [30, Lemma 2.1].  $\square$

The next theorem establishes the uniformized growth condition and KL property for the function  $d$ , which is useful in deriving the Luo–Tseng type error bound in the next section. We remark that existing results on uniformized KL property in [13, Lemma 6] concern the KL property over a *compact* set. Since the set  $\text{Argmin} d$  can be unbounded in general, these existing results cannot be applied directly. Instead, we establish the uniformity result by noting that  $d$  stays constant when moving along  $E^{\perp}$ . This is formally registered in the following auxiliary lemma, which is an immediate consequence of Corollary 2.5.5 and Theorem 2.5.3 of [5] because  $d = q^*$  (see Proposition 3.3).

**LEMMA 4.6.** *Consider (1.2). Let the function  $d$  be given in (3.6), and let  $E$  be given in (3.14). Suppose that Assumption 4.1(ii) holds. Then for any  $\mathbf{y} \in \text{dom} d$ , it holds that  $d(\mathbf{y} + \mathbf{u}) = d(\mathbf{y})$  whenever  $\mathbf{u} \in E^{\perp}$ .*

**THEOREM 4.7** (uniformized KL property and growth condition). *Consider (1.2). Let the function  $d$  be given in (3.6). Suppose that Assumption 4.1 holds. Then there exist positive constants  $\epsilon$  and  $c$  such that*

$$(4.29) \quad \text{dist}(\mathbf{y}, \text{Argmin} d) \leq c(d(\mathbf{y}) - d^*)^{1 - \frac{1}{1+\alpha}}$$

$$(4.30) \quad \text{and} \quad (d(\mathbf{y}) - d^*)^{\frac{1}{1+\alpha}} \leq c \text{dist}(\mathbf{0}, \partial d(\mathbf{y}))$$

whenever  $\mathbf{y} \in Y_{\epsilon} := \{ \mathbf{y} : \text{dist}(\mathbf{y}, \text{Argmin} d) \leq \epsilon, d^* \leq d(\mathbf{y}) \leq d^* + \epsilon \}$ .

*Proof.* Under the assumptions, by Proposition 4.5, we know that  $d$  is a KL function with exponent  $\frac{1}{1+\alpha}$ . In [4, Lemma 2.1(a)], it was shown that

$$\text{dom} q_E = \text{dom} q + E^{\perp} \text{ with } \text{int}(\text{dom} q_E) = \text{ri dom} q + E^{\perp},$$



where  $q_E$  and  $q$  are defined in Proposition 3.3. Now, note that  $\partial q_E(\mathbf{0}) (\subseteq \text{Arg min } d)$  is nonempty and compact, since we have  $\mathbf{0} \in \text{int}(\text{dom } q_E)$  in view of the above display and the fact that  $\mathbf{0} \in \text{ri dom } q$  (see Proposition 3.3). Then, using [13, Lemma 6] and [12, Theorem 5], there exist positive constants  $\epsilon$  and  $c$  so that

$$(4.31) \quad \text{dist}(\mathbf{z}, \text{Arg min } d) \leq c(d(\mathbf{z}) - d^*)^{1 - \frac{1}{1+\alpha}},$$

whenever  $\mathbf{z} \in \bar{Y}_\epsilon := \{\mathbf{z} : \text{dist}(\mathbf{z}, \partial q_E(\mathbf{0})) \leq \epsilon, d^* \leq d(\mathbf{z}) \leq d^* + \epsilon\}$ . Next, we show that

$$(4.32) \quad Y_\epsilon = \bar{Y}_\epsilon + E^\perp.$$

First, recall (3.13). For any  $\mathbf{y} \in Y_\epsilon$ , let  $\text{Proj}_{\text{Arg min } d}(\mathbf{y}) = \hat{\mathbf{z}} + \hat{\mathbf{u}}$  with  $\hat{\mathbf{z}} \in \partial q_E(\mathbf{0})$  and  $\hat{\mathbf{u}} \in E^\perp$ . Then

$$\text{dist}(\mathbf{y} - \hat{\mathbf{u}}, \partial q_E(\mathbf{0})) \leq \|(\mathbf{y} - \hat{\mathbf{u}}) - \hat{\mathbf{z}}\| = \|\mathbf{y} - \text{Proj}_{\text{Arg min } d}(\mathbf{y})\| \leq \epsilon;$$

moreover, we also have  $d(\mathbf{y} - \hat{\mathbf{u}}) = d(\mathbf{y})$  thanks to Lemma 4.6. From these we conclude that  $\mathbf{y} - \hat{\mathbf{u}} \in \bar{Y}_\epsilon$ . Then  $\mathbf{y} = (\mathbf{y} - \hat{\mathbf{u}}) + \hat{\mathbf{u}} \in \bar{Y}_\epsilon + E^\perp$ , which means  $Y_\epsilon \subseteq \bar{Y}_\epsilon + E^\perp$ . To prove the converse inclusion, take any  $\mathbf{z} + \mathbf{u} \in \bar{Y}_\epsilon + E^\perp$  with  $\mathbf{z} \in \bar{Y}_\epsilon$  and  $\mathbf{u} \in E^\perp$ . Then it holds that  $d(\mathbf{z} + \mathbf{u}) = d(\mathbf{z})$  thanks to Lemma 4.6 and

$$\text{dist}(\mathbf{z} + \mathbf{u}, \text{Arg min } d) \leq \text{dist}(\mathbf{z} + \mathbf{u}, \partial q_E(\mathbf{0}) + \mathbf{u}) = \text{dist}(\mathbf{z}, \partial q_E(\mathbf{0})) \leq \epsilon,$$

where the first inequality follows from  $\partial q_E(\mathbf{0}) + \mathbf{u} \subseteq \text{Arg min } d$  (see (3.13)). Then,  $\bar{Y}_\epsilon + E^\perp \subseteq Y_\epsilon$  and relation (4.32) holds.

Now, for any  $\mathbf{y} \in Y_\epsilon$ , according to (4.32), there exist  $\mathbf{z} \in \bar{Y}_\epsilon$  and  $\mathbf{u} \in E^\perp$  such that  $\mathbf{y} = \mathbf{z} + \mathbf{u}$ . It then follows that for any such  $\mathbf{y}$ ,

$$(4.33) \quad \begin{aligned} \text{dist}(\mathbf{y}, \text{Arg min } d) &\stackrel{(a)}{=} \text{dist}(\mathbf{z} + \mathbf{u}, (\text{Arg min } d) + \mathbf{u}) = \text{dist}(\mathbf{z}, \text{Arg min } d) \\ &\stackrel{(b)}{\leq} c(d(\mathbf{z}) - d^*)^{1 - \frac{1}{1+\alpha}} \stackrel{(c)}{\leq} c(d(\mathbf{y}) - d^*)^{1 - \frac{1}{1+\alpha}}, \end{aligned}$$

where (a) follows from  $\text{Arg min } d = (\text{Arg min } d) + \mathbf{u}$  (which holds because  $\text{Arg min } d = \partial q_E(\mathbf{0}) + E^\perp$  and  $\mathbf{u} \in E^\perp$ ), (b) follows from (4.31), and (c) follows from Lemma 4.6. This proves (4.29). Finally, (4.29) together with (2.2) implies (4.30).  $\square$

Before ending this section, we demonstrate that the exponent in (4.29) is the “best possible” under Assumption 4.1 by presenting an instance of BA-MSF (1.2) with a path along which both sides of (4.29) vanish in the same order of magnitude. The following example which argues the tightness of the exponent is in line with the recent research on the study of tight error bounds: see the notions of exact modulus of the generalized concave KL property in [39, Definition 6] and consistent error bound in [31, Definition 3.1].

*Example 4.3* (tightness of the exponent in (4.29)). Consider

$$\min_{x \in \mathbb{R}^2} \frac{1}{2} \|x - \bar{v}\|^2 \quad \text{s.t.} \quad A_1 x \in C_1,$$

where  $\bar{v} = (2, 0)$ ,  $p \in (1, 2]$ ,  $C_1 = \{x \in \mathbb{R}^2 : \|x\|_p \leq 1\}$ , and  $A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ . Since  $A_1^{-1}C_1 = \{x \in \mathbb{R}^2 : A_1 x \in C_1\} = [-1, 1] \times \mathbb{R}$ , it follows that  $x^* = (1, 0)$  is the unique solution. Moreover,  $C_1$  is  $C^{1,\alpha}$ -cone reducible with  $\alpha = p - 1$  in view of Example 4.2, and it holds that  $A_1^{-1} \text{ri } C_1 \neq \emptyset$  as  $(0, 0) \in (\text{ri } C_1) \cap \text{Range}(A_1)$ . Furthermore,

$$\begin{bmatrix} 1 & 0 \end{bmatrix}^T = \bar{v} - x^* \in \left\{ \begin{bmatrix} t & 0 \end{bmatrix}^T : t > 0 \right\} = \text{ri } \mathcal{N}_{[-1,1] \times \mathbb{R}}(x^*) = \text{ri } \mathcal{N}_{A_1^{-1}C_1}(x^*).$$

Thus, Assumption 4.1 is satisfied. Then Theorem 4.7 shows that (4.29) holds with exponent  $1 - \frac{1}{1+\alpha} = 1 - \frac{1}{p} = \frac{p-1}{p}$ , where the  $d$  in (4.29) now takes the form

$$d(y_1) = (1/2)\|A_1 y_1 - \bar{v}\|^2 - (1/2)\|\bar{v}\|^2 + \|y_1\|_{\frac{p}{p-1}}.$$

Next, in view of Proposition 3.1, we have  $[1 \ 0]^T = \bar{v} - x^* = A_1 \hat{y}_1$  whenever  $\hat{y}_1 \in \text{Arg min } d$ . This implies that the first coordinate of  $\hat{y}_1$  is 1. Also, one can see from the definition of  $d$  that the second coordinate of  $\hat{y}_1$  is 0. Thus,  $\text{Arg min } d = \{[1 \ 0]^T\}$ .

Now, let  $y_1^\epsilon = (1, \epsilon)$  for  $\epsilon \downarrow 0$ . By direct computation, we obtain as  $\epsilon \downarrow 0$  that

$$d(y_1^\epsilon) - d(\hat{y}_1) = (1 + \epsilon^{\frac{p}{p-1}})^{\frac{p-1}{p}} - 1 = \Theta(\epsilon^{\frac{p}{p-1}}) \quad \text{and} \quad \text{dist}(y_1^\epsilon, \text{Arg min } d) = \Theta(\epsilon),$$

showing that both sides of (4.29) vanish in the same order of magnitude along  $y_1^\epsilon$ .

*Remark 4.8.* When Assumption 4.1 holds with  $\alpha = 1$  in item (i) (the latter holds when each  $C_i$  is  $C^2$ -cone reducible), Theorem 4.7 asserts that the function  $d$  given in (3.6) satisfies the quadratic growth condition; that is, (4.29) holds with exponent  $1 - \frac{1}{1+\alpha} = \frac{1}{2}$ . However, for the  $p$ -norm ball in Example 4.2 with  $p \in (1, 2)$ , in view of Example 4.3,  $d$  satisfies (4.29) with a tight exponent  $1 - \frac{1}{\alpha+1} = \frac{p-1}{p} < \frac{1}{2}$ . Therefore, the  $p$ -norm ball in Example 4.2 with  $p \in (1, 2)$  is  $C^{1,\alpha}$ -cone reducible with  $\alpha = p-1 \in (0, 1)$  but not  $C^2$ -cone reducible.

**5. Convergence rate analysis.** In this section, we study the convergence rate of Algorithm 3.1. Recall from Theorem 2.2 that the (classical) Dykstra's projection algorithm is known to converge linearly when each  $C_i$  is polyhedral. We will argue that the same conclusion holds for Algorithm 3.1. The proof technique will shed light on how tools from section 4 can be further developed to analyze the convergence rate of Algorithm 3.1 for a more general class of sets.

The analysis below relies on the following mapping  $\mathcal{G} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ :

$$(5.1) \quad \mathcal{G}(\mathbf{y}) := \mathbf{y} - \text{Prox}_{\sigma_D}(\mathbf{y} - \nabla g(\mathbf{y})),$$

where  $g$  and  $D$  are given in (3.6) and (3.1), respectively. Note that the  $\mathcal{G}$  in (5.1) is instrumental in the framework of convergence rate analysis developed in [32, 38] for first-order methods.

**THEOREM 5.1** (linear convergence with polyhedral  $C_i$ ). *Consider (1.2), and let the function  $d$  be given in (3.6). Suppose that each  $C_i$  is polyhedral. Then  $\text{Arg min } d \neq \emptyset$ . Moreover, if  $\{x^t\}$ ,  $\{\mathbf{y}^t\}$  are the sequences generated by Algorithm 3.1, then there exist  $\mathbf{y}^* \in \text{Arg min } d$ ,  $r \in (0, 1)$ ,  $a_0 \in (0, 1)$ ,  $a_1 > 0$ , and positive integer  $\bar{t}$  such that*

$$\|x^t - x^*\| \leq a_1 a_0^t, \quad d(\mathbf{y}^{t+1}) - d^* \leq r(d(\mathbf{y}^t) - d^*), \quad \|\mathbf{y}^t - \mathbf{y}^*\| \leq a_1 a_0^t \quad \forall t \geq \bar{t},$$

where  $x^*$  is the unique solution of (1.2), and  $d^*$  is given in (3.6).

*Proof.* We first note from [36, Proposition 8.29] that  $\sigma_{C_i}$  is a piecewise linear function for each  $i$  and hence the function  $d$  is a piecewise-linear-quadratic function. Since  $d$  is bounded below according to Proposition 3.1, invoking the Frank–Wolfe theorem (see [23, Theorem 2.8.1]), one can see that  $\text{Arg min } d \neq \emptyset$ . Combining this with [38, Theorem 4], we assert that the following first-order error bound condition<sup>8</sup> is satisfied: there exist positive constants  $\epsilon_0$  and  $c_0$  such that

$$(5.2) \quad \text{dist}(\mathbf{y}, \text{Arg min } d) \leq c_0 \|\mathcal{G}(\mathbf{y})\| \quad \text{whenever } \|\mathcal{G}(\mathbf{y})\| \leq \epsilon_0 \text{ and } d(\mathbf{y}) \leq d^* + \epsilon_0,$$

<sup>8</sup>This condition is also known as the Luo–Tseng error bound in the literature; see [32, 38, 30, 42].

where  $\mathcal{G}$  is given in (5.1). Then, the linear convergence results of  $\{d(\mathbf{y}^{t+1})\}$  and  $\{\mathbf{y}^t\}$  follow from [38, Theorem 2(b)].<sup>9</sup> Moreover, from (3.10) and Proposition 3.1, we have

$$(5.3) \quad \|x^t - x^*\| = \left\| \bar{v} - \sum_{j=1}^{\ell} A_j^T y_j^t - \left( \bar{v} - \sum_{j=1}^{\ell} A_j^T y_j^* \right) \right\| \leq \|A^T\| \|\mathbf{y}^t - \mathbf{y}^*\|.$$

The claimed convergence result of  $\{x^t\}$  then follows immediately. □

Notice that in the above proof, the crucial ingredient is the first-order error bound condition (5.2), and it is known in [26, section 3] that this condition is intrinsically related to second-order growth condition. Below, leveraging the study of growth conditions in section 4 for  $C^{1,\alpha}$ -cone reducible sets, we will develop an analogue of (5.2) (with a general exponent on  $\|\mathcal{G}(\mathbf{y})\|$ ) to analyze the convergence rate of Algorithm 3.1 when each  $C_i$  is  $C^{1,\alpha}$ -cone reducible for some  $\alpha \in (0, 1]$ .

**5.1. Convergence rate analysis for  $C^{1,\alpha}$ -cone reducible  $C_i$ .** We first derive a Luo–Tseng type error bound based on the study of growth conditions in section 4.

LEMMA 5.2 (Luo–Tseng type error bound). *Consider (1.2). Let  $d^*$  and the function  $d$  be given in (3.6), and let  $\mathcal{G}$  be defined in (5.1). Suppose that Assumption 4.1 holds. Then there exist  $c > 0$  and  $\epsilon > 0$  such that*

$$\text{dist}(\mathbf{y}, \text{Arg min } d) \leq c \|\mathcal{G}(\mathbf{y})\|^\alpha \quad \text{whenever } \|\mathcal{G}(\mathbf{y})\| \leq \epsilon, \quad d^* \leq d(\mathbf{y}) \leq d^* + \epsilon.$$

*Proof.* By Theorem 4.7, there exist  $\epsilon \in (0, 1/2)$  and  $c > 0$  such that for any  $\mathbf{y} \in Y_\epsilon$

$$\text{dist}(\mathbf{y}, \text{Arg min } d) \leq c(d(\mathbf{y}) - d^*)^{1 - \frac{1}{1+\alpha}} \leq c \text{dist}(\mathbf{0}, \partial d(\mathbf{y}))^{\frac{\alpha}{1+\alpha}} \text{dist}(\mathbf{y}, \text{Arg min } d)^{\frac{\alpha}{1+\alpha}},$$

where the second inequality follows from (2.2). This further implies that

$$(5.4) \quad \text{dist}(\mathbf{y}, \text{Arg min } d) \leq c_1 \text{dist}(\mathbf{0}, \partial d(\mathbf{y}))^\alpha \quad \text{for all } \mathbf{y} \in Y_\epsilon,$$

where  $c_1 := c^{1+\alpha}$ . Next, for any  $\mathbf{y} \in Y_\epsilon$ , we have

$$\left\{ \begin{aligned} d^* &\leq d(\text{prox}_d(\mathbf{y})) \leq d(\text{prox}_d(\mathbf{y})) + \frac{1}{2} \|\mathbf{y} - \text{prox}_d(\mathbf{y})\|^2 \\ &= \inf_{\mathbf{u} \in \mathbb{R}^m} \left\{ d(\mathbf{u}) + \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|^2 \right\} \leq d(\mathbf{y}) \leq d^* + \epsilon, \\ \text{dist}(\text{prox}_d(\mathbf{y}), \text{Arg min } d) &\leq \|\text{prox}_d(\mathbf{y}) - \bar{\mathbf{y}}\| = \|\text{prox}_d(\mathbf{y}) - \text{prox}_d(\bar{\mathbf{y}})\| \\ &\stackrel{(a)}{\leq} \|\mathbf{y} - \bar{\mathbf{y}}\| = \text{dist}(\mathbf{y}, \text{Arg min } d) \leq \epsilon, \end{aligned} \right.$$

where  $\bar{\mathbf{y}} := \text{Proj}_{\text{Arg min } d}(\mathbf{y})$ , (a) holds because the proximal operator is nonexpansive. The above display implies that  $\text{prox}_d(\mathbf{y}) \in Y_\epsilon$ . Similarly, for any  $\mathbf{y} \in Y_\epsilon$ , it holds that

$$(5.5) \quad \|\mathbf{y} - \text{prox}_d(\mathbf{y})\| \leq \|\mathbf{y} - \bar{\mathbf{y}}\| + \|\text{prox}_d(\mathbf{y}) - \bar{\mathbf{y}}\| \leq 2\|\mathbf{y} - \bar{\mathbf{y}}\| \leq 2\epsilon < 1.$$

<sup>9</sup>One has to argue that the proximal CGD (3.7) and (3.9) is a special instance of the algorithm studied in [38] so that [38, Theorem 2(b)] is applicable. In details, we see from Proposition 3.2(i) that if one chooses  $\sigma = 0.5$ ,  $\gamma = 0$ , and  $\alpha_{\text{init}}^k \equiv 1$  in the Armijo rule used in [38] and consider the search direction  $y_i^{t+1} - y_i^t$ , then the Armijo rule will be satisfied with stepsize 1 so that  $y_i^{t+1}$  will appear in the next iterate. Thus, the proximal CGD (3.7) and (3.9) coincides with the algorithm considered in [38] with the aforementioned choices of  $\sigma$ ,  $\gamma$ , and  $\{\alpha_{\text{init}}^k\}$ .

Then we deduce for any  $\mathbf{y} \in Y_\epsilon$  that

$$\begin{aligned} \text{dist}(\mathbf{y}, \text{Arg min } d) &\leq \|\mathbf{y} - \text{prox}_d(\mathbf{y})\| + \text{dist}(\text{prox}_d(\mathbf{y}), \text{Arg min } d) \\ &\stackrel{(a)}{\leq} \|\mathbf{y} - \text{prox}_d(\mathbf{y})\| + c_1 \text{dist}(\mathbf{0}, \partial d(\text{prox}_d(\mathbf{y})))^\alpha \stackrel{(b)}{\leq} \|\mathbf{y} - \text{prox}_d(\mathbf{y})\| + c_1 \|\mathbf{y} - \text{prox}_d(\mathbf{y})\|^\alpha \\ &\stackrel{(c)}{\leq} (1 + c_1) \|\mathbf{y} - \text{prox}_d(\mathbf{y})\|^\alpha \stackrel{(d)}{\leq} (1 + c_1) c_2 \|\mathcal{G}(\mathbf{y})\|^\alpha, \end{aligned}$$

where (a) follows from (5.4) and the fact that  $\text{prox}_d(\mathbf{y}) \in Y_\epsilon$ , (b) follows from the fact that  $\mathbf{y} - \text{prox}_d(\mathbf{y}) \in \partial d(\text{prox}_d(\mathbf{y}))$ , (c) holds as  $\|\mathbf{y} - \text{prox}_d(\mathbf{y})\| < 1$  (thanks to (5.5)) and  $\alpha \in (0, 1]$ , and (d) holds for some  $c_2 > 0$  according to [26, Theorem 3.5].

We can now conclude that there exists  $c_3 > 0$  such that  $\text{dist}(\mathbf{y}, \text{Arg min } d) \leq c_3 \|\mathcal{G}(\mathbf{y})\|^\alpha$  whenever  $\mathbf{y} \in Y_\epsilon$ . To complete the proof, it suffices to show that there exists  $\nu \in (0, \epsilon)$  such that

$$\{\mathbf{y} : \|\mathcal{G}(\mathbf{y})\| \leq \nu\} \subseteq \{\mathbf{y} : \text{dist}(\mathbf{y}, \text{Arg min } d) \leq \epsilon\}.$$

Suppose to the contrary that this is not true. Then there exists  $\{\mathbf{z}^t\}$  such that  $\text{dist}(\mathbf{z}^t, \text{Arg min } d) > \epsilon$  for all  $t$  and  $\mathcal{G}(\mathbf{z}^t) \rightarrow \mathbf{0}$ . In view of the latter limit and [26, Theorem 3.5], we see that  $\|\mathbf{z}^t - \text{prox}_d(\mathbf{z}^t)\| \rightarrow 0$ . Since  $\mathbf{z}^t - \text{prox}_d(\mathbf{z}^t) \in \partial d(\text{prox}_d(\mathbf{z}^t))$ , we then deduce further that  $\text{dist}(\mathbf{0}, \partial d(\text{prox}_d(\mathbf{z}^t))) \rightarrow 0$ , and consequently,  $\text{dist}(\text{prox}_d(\mathbf{z}^t), \text{Arg min } d) \rightarrow 0$  in view of Proposition 3.3. We are now led to the following contradiction:

$$\epsilon < \text{dist}(\mathbf{z}^t, \text{Arg min } d) \leq \|\mathbf{z}^t - \text{prox}_d(\mathbf{z}^t)\| + \text{dist}(\text{prox}_d(\mathbf{z}^t), \text{Arg min } d) \rightarrow 0.$$

This completes the proof.  $\square$

The following theorem establishes the convergence rate of Algorithm 3.1. The steps for deriving upper bounds on  $\|\mathcal{G}(\mathbf{y}^t)\|$  (see (5.12)) and  $d(\mathbf{y}^{t+1}) - d^*$  (see (5.13)) follow a similar argument as in the proof of [38, Theorem 2].

**THEOREM 5.3** (convergence rate with  $C^{1,\alpha}$ -cone reducible  $C_i$ ). *Consider (1.2) and suppose that Assumption 4.1 holds. Let  $d$  be the function given in (3.6), and let  $\{x^t\}$ ,  $\{\mathbf{y}^t\}$  be the sequences generated by Algorithm 3.1. Then it holds that  $x^t \rightarrow x^*$  and  $\text{dist}(\mathbf{y}^t, \text{Arg min } d) \rightarrow 0$ , where  $x^*$  is the unique solution of (1.2). Moreover, the following statements hold:*

- (i) *If each  $C_i$  is  $C^{1,1}$ -cone reducible, then there exist  $\mathbf{y}^* \in \text{Arg min } d$ ,  $a_1 > 0$ ,  $a_0 \in (0, 1)$ , and a positive integer  $\bar{t}$  such that for any  $t \geq \bar{t}$ ,*

$$\|x^t - x^*\| \leq a_1 a_0^t, \quad \|\mathbf{y}^t - \mathbf{y}^*\| \leq a_1 a_0^t.$$

- (ii) *If each  $C_i$  is  $C^{1,\alpha}$ -cone reducible with  $\alpha \in (0, 1)$ , then there exist  $a_1 > 0$  and a positive integer  $\bar{t}$  such that for any  $t \geq \bar{t}$ ,*

$$\begin{aligned} \|x^t - x^*\| &\leq a_1 t^{-\frac{\alpha}{2} \cdot \frac{1-\theta}{2\theta-1}}, \\ d(\mathbf{y}^t) - d^* &\leq a_1 t^{-\frac{1-\theta}{2\theta-1}}, \\ \text{dist}(\mathbf{y}^t, \text{Arg min } d) &\leq a_1 t^{-\frac{\alpha}{2} \cdot \frac{1-\theta}{2\theta-1}}, \end{aligned}$$

where  $d^*$  is given in (3.6) and  $\theta := \frac{1}{1+\alpha} \in (\frac{1}{2}, 1)$  is the KL exponent of  $d$ .

*Proof.* The conclusion that  $\text{dist}(\mathbf{y}^t, \text{Arg min } d) \rightarrow 0$  follows immediately from Theorem 3.4. Recall from (3.1) that  $\mathbf{A}^T = [A_1^T \cdots A_\ell^T]$ . Then the convergence of  $\{x^t\}$  to  $x^*$  can be deduced by noticing

$$(5.6) \quad \|x^t - x^*\| \stackrel{(a)}{=} \left\| \bar{v} - \sum_{j=1}^{\ell} A_j^T y_j^t - \left( \bar{v} - \sum_{j=1}^{\ell} A_j^T \tilde{y}_j^t \right) \right\| \leq \|\mathbf{A}^T\| \text{dist}(\mathbf{y}^t, \text{Arg min } d),$$

where  $\text{Proj}_{\text{Arg min } d}(\mathbf{y}^t) =: (\tilde{y}_1^t, \dots, \tilde{y}_\ell^t)$  and (a) follows from Proposition 3.1 and (3.10).

Recall from (3.1) that  $D = C_1 \times \cdots \times C_\ell$ . Define

$$\mathcal{G}_i^\gamma(\mathbf{y}) := y_i - \text{Prox}_{\gamma^{-1}\sigma_{C_i}}(y_i - \gamma^{-1}\nabla_{y_i}g(\mathbf{y})), \quad i = 1, \dots, \ell,$$

where  $g$  is as in (3.6). Then we can write  $\mathcal{G}(\mathbf{y}) = (\mathcal{G}_1^1(\mathbf{y}), \dots, \mathcal{G}_\ell^1(\mathbf{y}))$  and (3.9) becomes

$$(5.7) \quad y_i^{t+1} = \text{Prox}_{\gamma_i^{-1}\sigma_{C_i}}(y_i^t - \gamma_i^{-1}\nabla_{y_i}g(\tilde{\mathbf{y}}_{i-1}^{t+1})) = y_i^t - \mathcal{G}_i^{\gamma_i}(\tilde{\mathbf{y}}_{i-1}^{t+1}),$$

where  $\tilde{\mathbf{y}}_{i-1}^{t+1}$  is as in (3.8). Let  $\hat{y}_i^{t+1} := y_i^t - \mathcal{G}_i^1(\tilde{\mathbf{y}}_{i-1}^{t+1}) = \text{Prox}_{\sigma_{C_i}}(y_i^t - \nabla_{y_i}g(\tilde{\mathbf{y}}_{i-1}^{t+1}))$ . By the definition of the proximal operator, we see that  $\hat{y}_i^{t+1}$  and  $y_i^{t+1}$  satisfy the respective first-order optimality conditions:

$$\begin{aligned} 0 &\in \nabla_{y_i}g(\tilde{\mathbf{y}}_{i-1}^{t+1}) + (\hat{y}_i^{t+1} - y_i^t) + \partial\sigma_{C_i}(\hat{y}_i^{t+1}), \\ 0 &\in \nabla_{y_i}g(\tilde{\mathbf{y}}_{i-1}^{t+1}) + \gamma_i(y_i^{t+1} - y_i^t) + \partial\sigma_{C_i}(y_i^{t+1}), \end{aligned}$$

which implies that

$$\begin{aligned} \hat{y}_i^{t+1} &\in \text{Arg min}_{z \in \mathbb{R}^{m_i}} \langle \nabla_{y_i}g(\tilde{\mathbf{y}}_{i-1}^{t+1}) + (\hat{y}_i^{t+1} - y_i^t), z \rangle + \sigma_{C_i}(z), \\ y_i^{t+1} &\in \text{Arg min}_{z \in \mathbb{R}^{m_i}} \langle \nabla_{y_i}g(\tilde{\mathbf{y}}_{i-1}^{t+1}) + \gamma_i(y_i^{t+1} - y_i^t), z \rangle + \sigma_{C_i}(z). \end{aligned}$$

Then, it holds that for any  $z \in \mathbb{R}^{m_i}$

$$\begin{aligned} (5.8) \quad &\langle \nabla_{y_i}g(\tilde{\mathbf{y}}_{i-1}^{t+1}) + (\hat{y}_i^{t+1} - y_i^t), \hat{y}_i^{t+1} - z \rangle + \sigma_{C_i}(\hat{y}_i^{t+1}) - \sigma_{C_i}(z) \leq 0, \\ (5.9) \quad &\langle \nabla_{y_i}g(\tilde{\mathbf{y}}_{i-1}^{t+1}) + \gamma_i(y_i^{t+1} - y_i^t), y_i^{t+1} - z \rangle + \sigma_{C_i}(y_i^{t+1}) - \sigma_{C_i}(z) \leq 0. \end{aligned}$$

Substituting  $z = y_i^{t+1}$  and  $z = \hat{y}_i^{t+1}$  into (5.8) and (5.9), respectively, and summing them together, we obtain upon recalling (5.7) and the fact  $\hat{y}_i^{t+1} - y_i^t = -\mathcal{G}_i^1(\tilde{\mathbf{y}}_{i-1}^{t+1})$  that

$$\langle \mathcal{G}_i^1(\tilde{\mathbf{y}}_{i-1}^{t+1}), \mathcal{G}_i^1(\tilde{\mathbf{y}}_{i-1}^{t+1}) - \mathcal{G}_i^{\gamma_i}(\tilde{\mathbf{y}}_{i-1}^{t+1}) \rangle + \gamma_i \langle \mathcal{G}_i^{\gamma_i}(\tilde{\mathbf{y}}_{i-1}^{t+1}), \mathcal{G}_i^{\gamma_i}(\tilde{\mathbf{y}}_{i-1}^{t+1}) - \mathcal{G}_i^1(\tilde{\mathbf{y}}_{i-1}^{t+1}) \rangle \leq 0,$$

which gives  $\|\mathcal{G}_i^1(\tilde{\mathbf{y}}_{i-1}^{t+1})\|^2 + \gamma_i \|\mathcal{G}_i^{\gamma_i}(\tilde{\mathbf{y}}_{i-1}^{t+1})\|^2 \leq (1 + \gamma_i) \langle \mathcal{G}_i^{\gamma_i}(\tilde{\mathbf{y}}_{i-1}^{t+1}), \mathcal{G}_i^1(\tilde{\mathbf{y}}_{i-1}^{t+1}) \rangle$ , and hence

$$(5.10) \quad \|\mathcal{G}_i^1(\tilde{\mathbf{y}}_{i-1}^{t+1})\| \leq (1 + \gamma_i) \|\mathcal{G}_i^{\gamma_i}(\tilde{\mathbf{y}}_{i-1}^{t+1})\| = (1 + \gamma_i) \|y_i^{t+1} - y_i^t\|,$$

where the last equality follows from (5.7). Recall that  $\mathbf{y}^t = (y_1^t, \dots, y_\ell^t)$ . Then

$$\begin{aligned} &\|\mathcal{G}_i^1(\mathbf{y}^t) - \mathcal{G}_i^1(\tilde{\mathbf{y}}_{i-1}^{t+1})\| \\ &= \|y_i^t - \text{prox}_{\sigma_{C_i}}(y_i^t - \nabla_{y_i}g(\mathbf{y}^t)) - (\tilde{\mathbf{y}}_{i-1}^{t+1})_i + \text{prox}_{\sigma_{C_i}}((\tilde{\mathbf{y}}_{i-1}^{t+1})_i - \nabla_{y_i}g(\tilde{\mathbf{y}}_{i-1}^{t+1}))\| \\ (5.11) \quad &\leq L_g \|\mathbf{y}^{t+1} - \mathbf{y}^t\|, \end{aligned}$$

where the inequality follows from the fact that  $(\tilde{\mathbf{y}}_{i-1}^{t+1})_i = \mathbf{y}_i^t$ , the nonexpansiveness of the proximal operator, and the Lipschitz continuity of  $\nabla g$  with Lipschitz modulus  $L_g := \|\mathbf{A}^T \mathbf{A}\|$ . We then deduce further that

$$(5.12) \quad \begin{aligned} \|\mathcal{G}(\mathbf{y}^t)\| &\leq \sum_{i=1}^{\ell} \|\mathcal{G}_i^1(\mathbf{y}^t)\| \leq \sum_{i=1}^{\ell} (\|\mathcal{G}_i^1(\mathbf{y}^t) - \mathcal{G}_i^1(\tilde{\mathbf{y}}_{i-1}^{t+1})\| + \|\mathcal{G}_i^1(\tilde{\mathbf{y}}_{i-1}^{t+1})\|) \\ &\stackrel{(a)}{\leq} \sum_{i=1}^{\ell} (L_g \|\mathbf{y}^{t+1} - \mathbf{y}^t\| + (1 + \gamma_i) \|y_i^{t+1} - y_i^t\|) \leq \hat{M} \|\mathbf{y}^{t+1} - \mathbf{y}^t\|, \end{aligned}$$

where (a) follows from (5.10) and (5.11), and  $\hat{M} := \ell L_g + (\ell + \sum_{i=1}^{\ell} \gamma_i) \sqrt{\ell}$ .

Next, let  $\bar{\mathbf{y}}^t = (\bar{y}_1^t, \dots, \bar{y}_\ell^t) := \text{Proj}_{\text{Arg min } d}(\mathbf{y}^t)$ . Then we see that

$$(5.13) \quad \begin{aligned} d(\mathbf{y}^{t+1}) - d^* &\stackrel{(a)}{=} g(\mathbf{y}^{t+1}) + \sigma_D(\mathbf{y}^{t+1}) - g(\bar{\mathbf{y}}^t) - \sigma_D(\bar{\mathbf{y}}^t) \\ &\stackrel{(b)}{=} \langle \nabla g(\mathbf{u}^t), \mathbf{y}^{t+1} - \bar{\mathbf{y}}^t \rangle + \sum_{i=1}^{\ell} [\sigma_{C_i}(y_i^{t+1}) - \sigma_{C_i}(\bar{y}_i^t)] \\ &= \langle \nabla g(\mathbf{u}^t) - \nabla g(\mathbf{y}^t), \mathbf{y}^{t+1} - \bar{\mathbf{y}}^t \rangle + \sum_{i=1}^{\ell} \langle \nabla_{y_i} g(\mathbf{y}^t) - \nabla_{y_i} g(\tilde{\mathbf{y}}_{i-1}^{t+1}), y_i^{t+1} - \bar{y}_i^t \rangle \\ &\quad + \sum_{i=1}^{\ell} [\langle \nabla_{y_i} g(\tilde{\mathbf{y}}_{i-1}^{t+1}) + \gamma_i (y_i^{t+1} - y_i^t), y_i^{t+1} - \bar{y}_i^t \rangle + \sigma_{C_i}(y_i^{t+1}) - \sigma_{C_i}(\bar{y}_i^t)] \\ &\quad - \sum_{i=1}^{\ell} \gamma_i \langle y_i^{t+1} - y_i^t, y_i^{t+1} - \bar{y}_i^t \rangle \\ &\stackrel{(c)}{\leq} L_g \|\mathbf{u}^t - \mathbf{y}^t\| \|\mathbf{y}^{t+1} - \bar{\mathbf{y}}^t\| + \ell L_g \|\mathbf{y}^{t+1} - \mathbf{y}^t\| \|\mathbf{y}^{t+1} - \bar{\mathbf{y}}^t\| \\ &\quad + \gamma_{\text{sum}} \|\mathbf{y}^{t+1} - \mathbf{y}^t\| \|\mathbf{y}^{t+1} - \bar{\mathbf{y}}^t\| \\ &\leq [(L_g + \ell L_g + \gamma_{\text{sum}}) \|\mathbf{y}^{t+1} - \mathbf{y}^t\| + L_g \|\mathbf{y}^t - \bar{\mathbf{y}}^t\|] (\|\mathbf{y}^{t+1} - \mathbf{y}^t\| + \|\mathbf{y}^t - \bar{\mathbf{y}}^t\|), \end{aligned}$$

where (a) follows from the definition of  $d$  in (3.6), (b) follows from the mean value theorem with  $\mathbf{u}^t = \tau_t \mathbf{y}^{t+1} + (1 - \tau_t) \bar{\mathbf{y}}^t$  for some  $\tau_t \in (0, 1)$ , (c) follows from (5.9) and the Lipschitz continuity of  $\nabla g$  with  $\gamma_{\text{sum}} := \sum_{i=1}^{\ell} \gamma_i$ .

On the other hand, we have  $\|\mathcal{G}(\mathbf{y}^t)\| \rightarrow 0$  (thanks to (5.12) and Proposition 3.2(iii)) and  $d(\mathbf{y}^t) \rightarrow d^*$  (thanks to Theorem 3.4). Thus, there exists sufficiently large positive integer  $\bar{t}$  such that  $\|\mathcal{G}(\mathbf{y}^t)\| \leq \epsilon$  and  $d(\mathbf{y}^t) \leq d^* + \epsilon$  for any  $t \geq \bar{t}$ , where  $\epsilon$  is specified in Lemma 5.2. Then, it holds that for any  $t \geq \bar{t}$ ,

$$(5.14) \quad \|\mathbf{y}^t - \bar{\mathbf{y}}^t\| \stackrel{(a)}{=} \text{dist}(\mathbf{y}^t, \text{Arg min } d) \stackrel{(b)}{\leq} c \|\mathcal{G}(\mathbf{y}^t)\|^\alpha \stackrel{(c)}{\leq} q\tilde{c} \|\mathbf{y}^{t+1} - \mathbf{y}^t\|^\alpha,$$

where (a) holds as  $\bar{\mathbf{y}}^t = \text{Proj}_{\text{Arg min } d}(\mathbf{y}^t)$ , (b) follows from Lemma 5.2, (c) holds for  $\tilde{c} := c\hat{M}^\alpha$  thanks to (5.12). Now, since  $\|\mathbf{y}^{t+1} - \mathbf{y}^t\| \rightarrow 0$  (see Proposition 3.2(iii)) and  $\alpha \in (0, 1]$ , we deduce from (5.13) and (5.14) that there exists  $c_1 > 0$  such that

$$(5.15) \quad d(\mathbf{y}^{t+1}) - d^* \leq c_1 \|\mathbf{y}^{t+1} - \mathbf{y}^t\|^{2\alpha} \quad \text{for all } t \geq \bar{t}.$$

In addition, Proposition 3.2(ii) shows that for all  $t$ ,

$$(5.16) \quad d(\mathbf{y}^{t+1}) - d(\mathbf{y}^t) \leq -c_2 \|\mathbf{y}^{t+1} - \mathbf{y}^t\|^2$$

for some  $c_2 > 0$ . Then, combining the above two displays, we have for all  $t \geq \bar{t}$ ,

$$(d(\mathbf{y}^{t+1}) - d^*) - (d(\mathbf{y}^t) - d^*) \leq -c_2 \|\mathbf{y}^{t+1} - \mathbf{y}^t\|^2 \leq -c_1^{-1/\alpha} c_2 (d(\mathbf{y}^{t+1}) - d^*)^{1/\alpha},$$

which implies that

$$(5.17) \quad (d(\mathbf{y}^{t+1}) - d^*) + c_1^{-1/\alpha} c_2 (d(\mathbf{y}^{t+1}) - d^*)^{1/\alpha} \leq d(\mathbf{y}^t) - d^* \quad \text{for all } t \geq \bar{t}.$$

Next, we consider the cases  $\alpha = 1$  and  $\alpha \in (0, 1)$  separately.

*Case 1.*  $\alpha = 1$ . In this case, from (5.17) we have for  $t \geq \bar{t}$

$$d(\mathbf{y}^{t+1}) - d^* \leq \frac{1}{1 + c_1^{-1} c_2} (d(\mathbf{y}^t) - d^*),$$

which implies that  $\{d(\mathbf{y}^{t+1}) - d^*\}$  is Q-linearly convergent to zero. The last inequality together with (5.16) shows that

$$\|\mathbf{y}^{t+1} - \mathbf{y}^t\|^2 \leq c_2^{-1} [(d(\mathbf{y}^t) - d^*) - (d(\mathbf{y}^{t+1}) - d^*)] \leq c_2^{-1} (d(\mathbf{y}^t) - d^*),$$

which, together with the Q-linearly convergence (to zero) of  $\{d(\mathbf{y}^t) - d^*\}$ , implies that there exists  $c_3 > 0, \alpha_0 \in (0, 1)$  such that  $\|\mathbf{y}^{t+1} - \mathbf{y}^t\| \leq c_3 \alpha_0^t$ . Thus, we have

$$\|\mathbf{y}^{t_1} - \mathbf{y}^{t_2}\| \leq \sum_{j=t_1}^{t_2-1} \|\mathbf{y}^{j+1} - \mathbf{y}^j\| \leq \frac{c_3}{1 - \alpha_0} \alpha_0^{t_1} \quad \text{for all } t_2 > t_1 \geq \bar{t},$$

which means that  $\{\mathbf{y}^t\}$  is convergent. Let  $\mathbf{y}^*$  denote its limit. Then  $\mathbf{y}^* \in \text{Arg min } d$  by Theorem 3.4. Passing to the limit as  $t_2 \rightarrow \infty$  in the above display, we obtain

$$\|\mathbf{y}^{t_1} - \mathbf{y}^*\| \leq (c_3 \alpha_0^{t_1}) / (1 - \alpha_0).$$

Using this, (3.10) and Proposition 3.1, we deduce further that for  $t \geq \bar{t}$

$$(5.18) \quad \|x^t - x^*\| = \left\| \bar{v} - \sum_{j=1}^{\ell} A_j^T y_j^t - \left( \bar{v} - \sum_{j=1}^{\ell} A_j^T y_j^* \right) \right\| \leq \|A^T\| \frac{c_3}{1 - \alpha_0} \alpha_0^t,$$

which implies the linear convergence of  $\{x^t\}$  to  $x^*$ .

*Case 2.*  $\alpha \in (0, 1)$ . We first show that the sequence  $\{d(\mathbf{y}^t) - d^*\}$  is sublinearly convergent to zero. To this end, recalling that  $\theta = \frac{1}{\alpha+1}$ , (5.17) can be written as

$$(5.19) \quad (d(\mathbf{y}^{t+1}) - d^*)^{\frac{\theta}{1-\theta}} \leq c_2^{-1} c_1^{\frac{\theta}{1-\theta}} [(d(\mathbf{y}^t) - d^*) - (d(\mathbf{y}^{t+1}) - d^*)] \quad \text{for all } t \geq \bar{t}.$$

Then, following the same arguments in [1, Theorem 2] starting from [1, equation (13)], we can deduce that there exists  $c_4 > 0$  such that

$$(5.20) \quad d(\mathbf{y}^t) - d^* \leq c_4 t^{-\frac{1-\theta}{2\theta-1}} \quad \text{for all } t \geq \bar{t},$$

where the last inequality corresponds to the first inequality on [1, page 15].

For the sublinear convergence of  $\{\text{dist}(\mathbf{y}^t, \text{Arg min } d)\}$ , we see that for any  $t \geq \bar{t}$

$$(5.21) \quad \begin{aligned} c_2 \tilde{c}^{-\frac{2}{\alpha}} \text{dist}(\mathbf{y}^t, \text{Arg min } d)^{\frac{2}{\alpha}} &\stackrel{(a)}{\leq} c_2 \tilde{c}^{-\frac{2}{\alpha}} (\tilde{c} \|\mathbf{y}^{t+1} - \mathbf{y}^t\|^\alpha)^{2/\alpha} = c_2 \|\mathbf{y}^{t+1} - \mathbf{y}^t\|^2 \\ &\stackrel{(b)}{\leq} d(\mathbf{y}^t) - d(\mathbf{y}^{t+1}) \leq d(\mathbf{y}^t) - d^*, \end{aligned}$$

where (a) follows from (5.14) and (b) follows from (5.16). Thus, the sublinear convergence rate of  $\{\text{dist}(\mathbf{y}^t, \text{Arg min } d)\}$  follows from (5.20). Finally, (5.6) implies that the convergence rate of  $\{x^t\}$  can be obtained from that of  $\{\text{dist}(\mathbf{y}^t, \text{Arg min } d)\}$ .  $\square$

Theorem 5.3(i) shows the linear convergence rate of Algorithm 3.1 under a genericity assumption on  $\bar{v}$ , i.e.,  $0 \in x^* - \bar{v} + \text{ri} \partial(\sum_{i=1}^{\ell} \delta_{A_i^{-1}C_i})(x^*)$ . To see that this condition is indispensable, we give an example that satisfies all assumptions in Theorem 5.3(i) except for Assumption 4.1(iii), and linear convergence fails.

*Example 5.1* (linear convergence fails). Consider the following problem:

$$(5.22) \quad \min_{y_1, y_2 \in \mathbb{R}^3} d(y_1, y_2) := \frac{1}{2} \|y_1 + y_2 - \bar{v}\|^2 + \sigma_{C_1}(y_1) + \sigma_{C_2}(y_2),$$

where  $\bar{v} = (1, -1, 1)$  and

$$C_1 := \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 \leq -\|(x_1, x_2)\|\}, \quad C_2 := \{(0, x_2, x_3) \in \mathbb{R}^3 : x_2, x_3 \in \mathbb{R}\}.$$

Note that (5.22) is the negative of the Lagrange dual (up to an additive constant) of the BA problem  $\min_{x \in C_1 \cap C_2} \frac{1}{2} \|x - \bar{v}\|^2$ . By the definition of  $C_1$  and  $C_2$ , one can see that  $C_1$  and  $C_2$  are  $C^{1,1}$ -cone reducible (see also Example 4.1) and  $\text{ri } C_1 \cap \text{ri } C_2 \neq \emptyset$ .

Notice that  $C_1 = K_1^\circ$  and  $C_2 = K_2^\circ$ , where

$$K_1 := \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 \geq \|(x_1, x_2)\|\}, \quad K_2 := \{(x_1, 0, 0) \in \mathbb{R}^3 : x_1 \in \mathbb{R}\}.$$

Then  $\sigma_{C_1} = \delta_{K_1}$  and  $\sigma_{C_2} = \delta_{K_2}$  according to [36, Example 11.4(b)], and (5.22) can be equivalently written as

$$(5.23) \quad \min_{y_1, y_2 \in \mathbb{R}^3} d(y_1, y_2) = \frac{1}{2} \|y_1 + y_2 - \bar{v}\|^2 + \delta_{K_1}(y_1) + \delta_{K_2}(y_2).$$

One can verify that the optimal value of the above problem is zero, and

$$y_1^* = (0, -1, 1) \in K_1 \quad \text{and} \quad y_2^* = (1, 0, 0) \in K_2$$

is the unique solution of (5.22), i.e.,  $\text{Arg min } d = \{(y_1^*, y_2^*)\}$ , and  $x^* = (0, 0, 0)$  is the unique solution of the primal problem in view of Proposition 3.1. Moreover, we have

$$0 \notin x^* - \bar{v} + \text{ri} \partial(\delta_{C_1} + \delta_{C_2})(x^*),$$

since  $\text{ri} \partial(\delta_{C_1} + \delta_{C_2})(x^*) = \text{ri} \mathcal{N}_{C_1 \cap C_2}(x^*) \subseteq \mathcal{N}_{C_1 \cap C_2}(x^*) = \{(x_1, x_2, x_3) : x_3 \geq |x_2|\}$ .

Recall that the  $y$ -iterates in Algorithm 3.1 can be obtained by applying CGD to (5.23), with the initial points  $y_1^0 = y_2^0 = (0, 0, 0)$ . By induction, the  $y$ -iterates generated by Algorithm 3.1 can be written as<sup>10</sup>

$$(5.24) \quad \begin{cases} y_1^{t+1} = \text{Proj}_{K_1}(\bar{v} - y_2^t) = \left( a_{t+1}, -\frac{1}{2} \left( 1 + \frac{1}{\sqrt{a_t^2 + 1}} \right), \frac{1}{2} \left( 1 + \sqrt{a_t^2 + 1} \right) \right), \\ y_2^{t+1} = \text{Proj}_{K_2}(\bar{v} - y_1^{t+1}) = (1 - a_{t+1}, 0, 0), \\ \text{with } a_{t+1} = \frac{1}{2} \left( 1 + \frac{1}{\sqrt{a_t^2 + 1}} \right) a_t \text{ and } a_0 = 1. \end{cases}$$

<sup>10</sup>The projection of  $(u, t) \in \mathbb{R}^{n+1}$  onto  $K_1$  can be found in [10, Exercise 29.11].



Observing from above that  $\{(y_1^t, y_2^t)\}$  is convergent since  $\{a_t\}$  is bounded and non-increasing, we know that  $(y_1^t, y_2^t) \rightarrow (y_1^*, y_2^*)$  by Proposition 3.2(v). This implies that  $\lim_{t \rightarrow \infty} a_t = 0$ . From the definition of  $a_t$  in (5.24) and noting that  $a_t > 0$ , we have

$$\frac{1}{a_{t+1}^2} - \frac{1}{a_t^2} = \frac{3a_t^2 + 2 - 2\sqrt{a_t^2 + 1}}{a_t^2(1 + \sqrt{a_t^2 + 1})^2} = \frac{3a_t^2 + 2 - 2(1 + 0.5a_t^2 + O(a_t^4))}{a_t^2(1 + \sqrt{a_t^2 + 1})^2} = \frac{2 + O(a_t^2)}{(1 + \sqrt{a_t^2 + 1})^2}.$$

Summing both sides of the above equality from  $t = 0$  to  $t = N$ , we get  $\frac{1}{a_N^2} - \frac{1}{a_0^2} = \Theta(N)$ , which implies that  $a_t = \Theta(1/\sqrt{t})$ . Then, using this relation, (5.24) and the fact  $\sqrt{h^2 + 1} - 1 = \Theta(h^2)$  as  $h \rightarrow 0$ , we deduce that  $\text{dist}((y_1^{t+1}, y_2^{t+1}), \text{Arg min } d)^2 = \|(y_1^{t+1}, y_2^{t+1}) - (y_1^*, y_2^*)\|^2$  is equal to

$$2a_{t+1}^2 + \left(\frac{\sqrt{a_t^2 + 1} - 1}{2\sqrt{a_t^2 + 1}}\right)^2 + \left(\frac{\sqrt{a_t^2 + 1} - 1}{2}\right)^2 = 2a_{t+1}^2 + \Theta(a_t^4) = \Theta\left(\frac{1}{t+1}\right),$$

which means that the convergence rate of  $\{(y_1^t, y_2^t)\}$  to  $(y_1^*, y_2^*)$  is not linear.

**Appendix A. Equivalence between Algorithm 3.1 and a proximal CGD scheme.** We first derive Algorithm 3.1 from (3.7) and (3.9). To this end, note that (3.9) is equivalent to

$$(A.1) \quad y_i^{t+1} = \text{prox}_{\gamma_i^{-1}\sigma_{C_i}}(y_i^t - \gamma_i^{-1}\nabla_{y_i}g(\tilde{y}_{i-1}^{t+1})).$$

Define  $x_\ell^0 := \bar{v}$  and, for  $i = 0, 1, \dots, \ell$ ,

$$(A.2) \quad x_i^{t+1} := \bar{v} - \sum_{j=1}^i A_j^T y_j^{t+1} - \sum_{j=i+1}^{\ell} A_j^T y_j^t.$$

Then according to (A.2) and the definition of  $\tilde{y}_{i-1}^{t+1}$  in (3.8), it holds that for  $i = 1, \dots, \ell$ ,

$$(A.3) \quad A_i x_{i-1}^{t+1} = -\nabla_{y_i}g(\tilde{y}_{i-1}^{t+1}),$$

$$(A.4) \quad x_i^{t+1} = x_{i-1}^{t+1} - A_i^T(y_i^{t+1} - y_i^t).$$

Consequently, relation (A.1) can be further rewritten as

$$(A.5) \quad y_i^{t+1} \stackrel{(a)}{=} y_i^t - \gamma_i^{-1}\nabla_{y_i}g(\tilde{y}_{i-1}^{t+1}) - \gamma_i^{-1}\text{Proj}_{C_i}(\gamma_i y_i^t - \nabla_{y_i}g(\tilde{y}_{i-1}^{t+1}))$$

$$(A.6) \quad \stackrel{(A.3)}{=} y_i^t + \gamma_i^{-1}A_i x_{i-1}^{t+1} - \gamma_i^{-1}\text{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}),$$

where (a) holds since for any closed convex set  $K \subseteq \mathbb{R}^n$ ,  $r > 0$ , and  $u \in \mathbb{R}^n$ , we have  $\text{prox}_{r\sigma_K}(u) = u - \text{prox}_{(\sigma_{rK})^*}(u) = u - \text{Proj}_{rK}(u) = u - r\text{Proj}_K(r^{-1}u)$ . Then, using (A.4) and the expression of  $y_i^{t+1} - y_i^t$  derived from (A.6), we have for  $i = 1, \dots, \ell$ ,

$$(A.7) \quad x_i^{t+1} = (I - \gamma_i^{-1}A_i^T A_i)x_{i-1}^{t+1} + \gamma_i^{-1}A_i^T \text{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}).$$

Thus, we have shown that (3.9) gives (A.6) and (A.7), i.e., (3.11) in Algorithm 3.1. Also, from (3.7), (A.2), and the definition that  $x_\ell^0 := \bar{v}$ , we see that

$$(A.8) \quad x_0^{t+1} = \bar{v} - \sum_{j=1}^{\ell} A_j^T y_j^t = x_\ell^t \quad \text{for all } t \geq 0.$$

Combining (A.6), (A.7), and (A.8), we derive Algorithm 3.1.

Conversely, we show that (3.7) and (3.9) can be deduced from Algorithm 3.1. Thanks to Step 1 of Algorithm 3.1, it suffices to prove (A.1). To this end, let  $\{x^t\}$ ,  $\{x_i^t\}$ , and  $\{y_i^t\}$ ,  $i = 1, \dots, \ell$ , be generated by Algorithm 3.1. Then we obtain (A.4) from (3.11). We claim that (3.10) holds by induction. First, it clearly holds for  $t = 0$  from Step 1 of Algorithm 3.1. Suppose that  $x_\ell^t = \bar{v} - \mathbf{A}^T \mathbf{y}^t$  for some  $t \geq 0$ . Then the beginning of Step 2 of Algorithm 3.1 shows that  $x_0^{t+1} = \bar{v} - \mathbf{A}^T \mathbf{y}^t$ , which, together with (A.4), shows that  $x_\ell^{t+1} = \bar{v} - \mathbf{A}^T \mathbf{y}^{t+1}$ , thus establishes (3.10) by induction.

Now, (3.10), (A.4), and  $x_0^{t+1} = x_\ell^t$  (Step 2 of Algorithm 3.1) give (A.2), and hence (A.3). The  $y$ -update in (3.11) then shows that (A.5) holds, which is just (A.1).

## REFERENCES

- [1] 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.
- [2] H. ATTOUCH, J. BOLTE, P. REDONT, AND A. SOUBEYRAN, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457.
- [3] 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.
- [4] A. AUSLENDER, R. COMINETTI, AND J. CROUZIEUX, *Convex functions with unbounded level sets and applications to duality theory*, SIAM J. Optim., 3 (1993), pp. 669–687, <https://doi.org/10.1137/0803034>.
- [5] A. AUSLENDER AND M. TEBoulLE, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer, New York, 2006.
- [6] H. H. BAUSCHKE AND J. M. BORWEIN, *Dykstra’s alternating projection algorithm for two sets*, J. Approx. Theory, 79 (1994), pp. 418–443.
- [7] H. H. BAUSCHKE AND J. M. BORWEIN, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367–426, <https://doi.org/10.1137/S0036144593251710>.
- [8] H. H. BAUSCHKE, J. M. BORWEIN, AND W. LI, *Strong conical hull intersection property, bounded linear regularity, Jameson’s property (G), and error bounds in convex optimization*, Math. Program., 86 (1999), pp. 135–160.
- [9] H. H. BAUSCHKE, R. S. BURACHIK, D. B. HERMAN, AND C. Y. KAYA, *On Dykstra’s algorithm: Finite convergence, stalling, and the method of alternating projections*, Optim. Lett., 14 (2020), pp. 1975–1987.
- [10] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, Cham, 2017.
- [11] H. H. BAUSCHKE AND A. S. LEWIS, *Dykstra’s algorithm with Bregman projections: A convergence proof*, Optimization, 48 (2000), pp. 409–427.
- [12] J. BOLTE, T. P. NGUYEN, J. PEYPOUQUET, AND B. W. SUTER, *From error bounds to the complexity of first-order descent methods for convex functions*, Math. Program., 165 (2017), pp. 471–507.
- [13] J. BOLTE, S. SABACH, AND M. TEBoulLE, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146 (2014), pp. 459–494.
- [14] J. F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.
- [15] J. P. BOYLE AND R. L. DYKSTRA, *A method for finding projections onto the intersection of convex sets in Hilbert spaces*, in Advances in Order Restricted Statistical Inference, Springer, 1986, pp. 28–47.
- [16] M. BROOKE, Y. CENSOR, AND A. GIBALI, *Dynamic string-averaging CQ-methods for the split feasibility problem with percentage violation constraints arising in radiation therapy treatment planning*, Int. Trans. Oper. Res., 30 (2020), pp. 181–205.
- [17] J. V. BURKE, F. E. CURTIS, H. WANG, AND J. WANG, *Iterative reweighted linear least squares for exact penalty subproblems on product sets*, SIAM J. Optim., 25 (2015), pp. 261–294, <https://doi.org/10.1137/130950239>.
- [18] C. BYRNE, *Iterative oblique projection onto convex sets and the split feasibility problem*, Inverse Problems, 18 (2002), pp. 441–453.

- [19] Y. CENSOR AND T. ELFVING, *A multiprojection algorithm using Bregman projections in a product space*, Numer. Algorithms, 8 (1994), pp. 221–239.
- [20] Y. CENSOR, T. ELFVING, N. KOPF, AND T. BORTFELD, *The multiple-sets split feasibility problem and its applications for inverse problems*, Inverse Problems, 21 (2005), pp. 2071–2084.
- [21] Y. CENSOR, D. REEM, AND M. ZAKNOON, *A generalized block-iterative projection method for the common fixed point problem induced by cutters*, J. Global Optim., 84 (2022), pp. 967–987.
- [22] A. CHAMBOLLE AND T. POCK, *On the ergodic convergence rates of a first-order primal–dual algorithm*, Math. Program., 159 (2016), pp. 253–287.
- [23] R. W. COTTLE, J.-S. PANG, AND R. E. STONE, *The Linear Complementarity Problem*, Academic Press, Boston, 1992.
- [24] F. DEUTSCH AND H. HUNDAL, *The rate of convergence of Dykstra’s cyclic projections algorithm: The polyhedral case*, Numer. Funct. Anal. Optim., 15 (1994), pp. 537–565.
- [25] A. L. DONTCHEV AND R. T. ROCKAFELLAR, *Implicit Functions and Solution Mappings*, Vol. 543, Springer-Verlag New York, 2009.
- [26] D. DRUSVYATSKIY AND A. S. LEWIS, *Error bounds, quadratic growth, and linear convergence of proximal methods*, Math. Oper. Res., 43 (2018), pp. 919–948.
- [27] N. GAFFKE AND R. MATHAR, *A cyclic projection algorithm via duality*, Metrika, 36 (1989), pp. 29–54.
- [28] S.-P. HAN, *A successive projection method*, Math. Program., 40 (1988), pp. 1–14.
- [29] H. HUNDAL AND F. DEUTSCH, *Two generalizations of Dykstra’s cyclic projections algorithm*, Math. Program., 77 (1997), pp. 335–355.
- [30] G. LI AND T. K. PONG, *Calculus of the exponent of Kurdyka–Lojasiewicz inequality and its applications to linear convergence of first-order methods*, Found. Comput. Math., 18 (2018), pp. 1199–1232.
- [31] T. LIU AND B. F. LOURENÇO, *Convergence analysis under consistent error bounds*, Found. Comput. Math., (2022), <https://doi.org/10.1007/s10208-022-09586-4>.
- [32] Z.-Q. LUO AND P. TSENG, *Error bounds and convergence analysis of feasible descent methods: A general approach*, Ann. Oper. Res., 46 (1993), pp. 157–178.
- [33] C. J. PANG, *Set intersection problems: Supporting hyperplanes and quadratic programming*, Math. Program., 149 (2015), pp. 329–359.
- [34] C. J. PANG, *Distributed deterministic asynchronous algorithms in time-varying graphs through Dykstra splitting*, SIAM J. Optim., 29 (2019), pp. 484–510, <https://doi.org/10.1137/18M1168297>.
- [35] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [36] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, Springer, 2009.
- [37] A. SHAPIRO, *Sensitivity analysis of generalized equations*, J. Math. Sci., 115 (2003), pp. 2554–2565.
- [38] P. TSENG AND S. YUN, *A coordinate gradient descent method for nonsmooth separable minimization*, Math. Program., 117 (2009), pp. 387–423.
- [39] X. WANG AND Z. WANG, *The exact modulus of the generalized concave Kurdyka–Lojasiewicz property*, Math. Oper. Res., 47 (2022), pp. 2765–2783.
- [40] Y. YANG, Q. YANG, AND S. ZHANG, *Modified alternating direction methods for the modified multiple-sets split feasibility problems*, J. Optim. Theory Appl., 163 (2014), pp. 130–147.
- [41] P. YU, G. LI, AND T. K. PONG, *Kurdyka–Lojasiewicz exponent via inf-projection*, Found. Comput. Math., 22 (2022), pp. 1171–1217.
- [42] M.-C. YUE, Z. ZHOU, AND A. M.-C. SO, *A family of inexact SQA methods for non-smooth convex minimization with provable convergence guarantees based on the Luo–Tseng error bound property*, Math. Program., 174 (2019), pp. 327–358.