

First Published in SIAM Journal on Optimization in Volume 30, Issue 2, published by the Society for Industrial and Applied Mathematics (SIAM). Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

A SUBGRADIENT-BASED APPROACH FOR FINDING THE MAXIMUM FEASIBLE SUBSYSTEM WITH RESPECT TO A SET

MINGLU YE * AND TING KEI PONG †

Abstract. We propose a subgradient-based method for finding the maximum feasible subsystem in a collection of closed sets with respect to a given closed set C (MFS_C). In this method, we reformulate the MFS_C problem as an ℓ_0 optimization problem and construct a sequence of continuous optimization problems to approximate it. The objective of each approximation problem is the sum of the composition of a nonnegative nondecreasing continuously differentiable concave function with the squared distance function to a closed set. Although this objective function is non-smooth in general, a subgradient can be obtained in terms of the projections onto the closed sets. Based on this observation, we adapt a subgradient projection method to solve these approximation problems. Unlike classical subgradient methods, the convergence (clustering to stationary points) of our subgradient method is guaranteed with a *nondiminishing stepsize* under mild assumptions. This allows us to further study the sequential convergence of the subgradient method under suitable Kurdyka-Lojasiewicz assumptions. Finally, we illustrate our algorithm numerically for solving the MFS_C problems on a collection of halfspaces and a collection of unions of halfspaces, respectively, with respect to the set of s -sparse vectors.

Key words. Maximum feasible subsystem, subgradient methods, Kurdyka-Lojasiewicz property

AMS subject classifications. 90C06, 90C26, 90C30, 90C90

1. Introduction. Let $\{C, D_1, \dots, D_m\}$ be a collection of finitely many nonempty (possibly nonconvex) closed sets in \mathbb{R}^n . We consider the problem of finding the *maximum feasible subsystem with respect to C* (MFS_C):

$$(1) \quad \begin{aligned} \max \quad & \#(I) \\ \text{s.t.} \quad & C \cap \left(\bigcap_{i \in I} D_i\right) \neq \emptyset, \\ & I \subseteq \{1, \dots, m\}, \end{aligned}$$

where $\#(I)$ stands for the cardinality of the index set I . The above problem is a natural generalization of the widely studied problem of finding the *maximum feasible (linear) subsystem* (MFLS), which is just (1) with $C = \mathbb{R}^n$ and D_i being halfspaces and is known to be NP hard; see [11]. The MFLS problem finds applications in many different areas such as image and signal processing [3], operations research [1, 2], machine learning [4] and linear programming [12–14], and various solution methods have been proposed. Many of these methods are based on integer programming techniques and exploit explicitly the fact that each D_i is a halfspace and $C = \mathbb{R}^n$; see, for example, [19–22] and references therein. For instance, the recently proposed branch-and-cut method in [22] builds on the classical branch-and-cut approach for integer programming: it exploits the duality between MFLS and the problem of finding the minimum *irreducible infeasible subsystem* (IIS) cover, and makes use of the structure of a special kind of polytope to identify IIS covers; see Sections 3.1 and 3.2 of [22]. Thus, when it comes to the MFS_C problem (1), it is not clear whether the method in [22] and other existing methods for the MFLS problem can be readily generalized to solve (1) for general sets C and D_i , which are possibly nonconvex.

*Department of Mathematics and Information, China West Normal University, Nanchong, Sichuan, China. yml2002cn@aliyun.com.

†Department of Applied Mathematics, the Hong Kong Polytechnic University, Hong Kong. This author was supported partly by Hong Kong Research Grants Council PolyU153005/17p. tk.pong@polyu.edu.hk.

In this paper, we develop a new approach for approximately solving the MFS_C problem (1). Our method takes advantage of the recent advancement in ℓ_0 minimization such as [26], and is based on the observation that the MFS_C problem (1) is equivalent to the following nonlinear programming problem with an ℓ_0 objective:

$$(2) \quad \min_{x \in C} \Phi_0(x) := \sum_{i=1}^m |d_{D_i}^2(x)|_0$$

where $|\cdot|_0$ is the ℓ_0 norm.¹ In our approach, as in [26], we approximate the ℓ_0 norm in Φ_0 by a sequence of continuous functions. We show that if the sequence of continuous functions $\{\varphi_{\epsilon_k}\}$ is chosen in such a way that it is *both* epi-convergent and pointwise convergent to the ℓ_0 norm, then the sequence of functions

$$(3) \quad \Phi_{\epsilon_k}(x) := \sum_{i=1}^m \varphi_{\epsilon_k}(d_{D_i}^2(x))$$

epi-converges and pointwise converges to Φ_0 . We then explore how to minimize Φ_{ϵ_k} over C , under further differentiability and concavity assumptions on φ_{ϵ_k} (see Section 4 for the assumptions and concrete examples of $\{\varphi_{\epsilon_k}\}$). Notice that the function Φ_{ϵ_k} , though continuous, is still possibly nonsmooth in general: this is because the function $x \mapsto d_{D_i}^2(x)$ is differentiable if and only if D_i is convex. Fortunately, a subgradient of the squared distance function to D_i can be obtained in terms of the projections onto D_i . We thus propose a subgradient projection method for minimizing the Φ_{ϵ_k} in (3) over C . Surprisingly, we are able to show that the projected subgradient direction is indeed a *descent direction*. This enables us to incorporate the standard nonmonotone line-search scheme to empirically accelerate the algorithm. Moreover, under mild assumptions on the collection of closed sets, we show that the stepsizes used are uniformly bounded away from zero and that any accumulation point of the sequence generated by our subgradient projection method is a stationary point of $\Phi_{\epsilon_k} + \delta_C$.² Based on these and some suitable Kurdyka-Lojasiewicz (KL) assumptions, we further show that the whole sequence generated by our method (with monotone linesearch) for minimizing the Φ_{ϵ_k} in (3) over C is convergent to a stationary point of $\Phi_{\epsilon_k} + \delta_C$ when each D_i is convex or $C = \mathbb{R}^n$. We also establish a relationship between the different KL assumptions considered. Finally, we perform numerical experiments on (1) under two different scenarios: we consider C being the set of s -sparse vectors, and $\{D_1, \dots, D_m\}$ being either a collection of halfspaces or a collection of unions of halfspaces. Our experiments on random instances show that our approach is able to identify a reasonably large feasible subsystem with respect to C in a reasonable period of time, even for large-scale problems.

The rest of this paper is organized as follows. We introduce notation and preliminary materials in Section 2. An approximation scheme for solving (1) based on approximately minimizing a bunch of Φ_{ϵ_k} in (3) over C is introduced in Section 3. We then propose and study in Section 4 a subgradient method for minimizing Φ_{ϵ_k} in (3) over C and show that the sequence generated clusters at a stationary point of

¹This equivalence can be easily deduced by noting that if I_* solves (1), then the solution set of (2) is $C \cap \left(\bigcap_{i \in I_*} D_i\right)$, and that if x^* solves (2), then a solution I_* of (1) is given by $I_* = \{i : \text{dist}(x^*, D_i) = 0\}$.

²As a corollary, under mild assumptions on the collection of closed sets, we establish that the sequence generated by the averaged projection method clusters at stationary points of a suitable potential function; see Corollary 4.5.

$\Phi_{\epsilon_k} + \delta_C$ under mild assumptions on the collection of closed sets and some further differentiability and concavity assumptions on φ_{ϵ_k} . Sequential convergence is studied in Section 5 under additional KL assumptions. Finally, numerical experiments are presented in Section 6 and we give some concluding remarks in Section 7.

2. Notation and preliminaries. We let \mathbb{R}^n denote the n -dimensional Euclidean space and let $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ denote the standard inner product and the induced norm, respectively. The nonnegative orthant and positive orthant are denoted by \mathbb{R}_+^n and \mathbb{R}_{++}^n , respectively. For an $x \in \mathbb{R}^n$, we let $\|x\|_0$ denote the ℓ_0 norm of x , and let $\mathbf{B}(x, r)$ denote the closed ball centered at x with radius r . Moreover, for an $s \in \mathbb{R}$, we use $|s|_0$ to denote its ℓ_0 norm and $[s]_+ := \max\{s, 0\}$ to denote its positive part.

Let $C \subseteq \mathbb{R}^n$ be a nonempty closed set. We let $P_C(x)$ denote the set of projections of a vector $x \in \mathbb{R}^n$ onto C : this set is always nonempty, and is a singleton when C is in addition convex. The distance to C from x is denoted by $\text{dist}(x, C)$ or $d_C(x)$, and we use C^∞ to denote the horizon cone of C , which is defined as $C^\infty := \{x : \exists x^t \in C, \lambda_t \downarrow 0 \text{ with } \lambda_t x^t \rightarrow x\}$.³ Finally, we let δ_C denote the indicator function of C , which is defined as

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

For an extended-real-valued function $f : \mathbb{R}^n \rightarrow [-\infty, \infty]$, we let $\text{dom } f := \{x \in \mathbb{R}^n : f(x) < \infty\}$. Such a function is said to be proper if $\text{dom } f \neq \emptyset$ and f is never $-\infty$, and is said to be closed if f is lower semicontinuous. For a proper closed function f , the regular subgradient and (limiting) subgradient [24, Definition 8.3] at a point $\bar{x} \in \text{dom } f$ are defined respectively as

$$\begin{aligned} \widehat{\partial}f(\bar{x}) &:= \left\{ v : \liminf_{x \rightarrow \bar{x}, x \neq \bar{x}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0 \right\}, \\ \partial f(\bar{x}) &:= \left\{ v : \exists x^t \xrightarrow{f} \bar{x} \text{ and } v^t \in \widehat{\partial}f(x^t) \text{ with } v^t \rightarrow v \right\}, \end{aligned}$$

where $x^t \xrightarrow{f} \bar{x}$ means both $x^t \rightarrow \bar{x}$ and $f(x^t) \rightarrow f(\bar{x})$. We define $\partial f(x) = \widehat{\partial}f(x) = \emptyset$ whenever $x \notin \text{dom } f$ by convention, and write $\text{dom } \partial f := \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$. Clearly, we have $\widehat{\partial}f(\bar{x}) \subseteq \partial f(\bar{x})$. It is known that ∂f reduces to the classical subdifferential in convex analysis if f is in addition convex [24, Proposition 8.12], and we have $\partial f(x) = \{\nabla f(x)\}$ if f is continuously differentiable at x . We also define the normal cone of a nonempty closed set C at $x \in C$ as $N_C(x) = \partial \delta_C(x)$. Finally, for a positive number ν , we denote the set of ν -minimizers of f by $\nu\text{-arg min } f$, i.e., $\nu\text{-arg min } f := \{x : f(x) \leq \inf f + \nu\}$. The set of minimizers of f is denoted by $\text{arg min } f$.

We next recall the Kurdyka-Łojasiewicz (KL) property, which is an important property for analyzing convergence of first-order methods; see, for example, [5–7]. For notational simplicity, for any $\nu \in (0, \infty]$, we let Ξ_ν denote the set of continuous concave functions $\phi : [0, \nu] \rightarrow \mathbb{R}_+$ that are continuously differentiable on $(0, \nu)$ with positive derivatives and satisfy $\phi(0) = 0$.

DEFINITION 2.1 (KL property & KL function). *Let f be a proper closed function. We say that f satisfies the Kurdyka-Łojasiewicz (KL) property at $\hat{x} \in$*

³We recall from [24, Theorem 3.5] that $C^\infty = \{0\}$ if and only if C is bounded.

$\text{dom } \partial f$ if there exist a neighborhood \mathcal{N} of \hat{x} , $\nu \in (0, \infty]$ and a $\phi \in \Xi_\nu$ such that

$$(4) \quad \phi'(f(x) - f(\hat{x})) \text{dist}(0, \partial f(x)) \geq 1$$

whenever $x \in \mathcal{N}$ and $f(\hat{x}) < f(x) < f(\hat{x}) + \nu$. If f satisfies the KL property at $\hat{x} \in \text{dom } \partial f$ and the ϕ in (4) can be chosen as $\phi(s) = cs^{1-\theta}$ for some $\theta \in [0, 1)$ and $c > 0$, then f is said to satisfy the KL property at \hat{x} with exponent θ .

A proper closed function f is called a KL function if it satisfies the KL property at every point in $\text{dom } \partial f$, and is called a KL function with exponent $\theta \in [0, 1)$ if it satisfies the KL property with exponent θ at every point in $\text{dom } \partial f$.

It can be shown that the KL property is satisfied by a large class of functions, including all proper closed semialgebraic functions. Indeed, it is known that proper closed semialgebraic functions are KL functions with exponent θ for some $\theta \in [0, 1)$; see, for example, [6, Section 4] and references therein. We next recall the following lemma concerning a *uniformized* KL property, first proved in [10, Lemma 6]. It was used there for establishing convergence of first-order methods for level-bounded objective functions.

LEMMA 2.2 (Uniformized KL property). *Let Ω be a compact set and f be a proper closed function that satisfies the KL property at every point in Ω and is constant on Ω , say, equals l_* . Then there exist $\epsilon > 0$, $\nu \in (0, \infty]$ and a $\phi \in \Xi_\nu$ such that*

$$\phi'(f(x) - l_*) \text{dist}(0, \partial f(x)) \geq 1$$

whenever $d_\Omega(x) < \epsilon$ and $l_* < f(x) < l_* + \nu$.

Finally, we prove that a certain sequence of function is equi-lsc. This will be used in Section 3 to construct an explicit example of sequence $\{\Phi_{\epsilon_t}\}$ (as in (3)) that epi-converges and pointwise converges to Φ_0 in (2). We first recall the following definition; see [24, Page 248].

DEFINITION 2.3. *Let $\{f_t\}$ be a sequence of functions. We say that $\{f_t\}$ is equi-lsc at x_0 if for every $\epsilon > 0$ and $\rho > 0$ there exists $\delta > 0$ such that*

$$f_t(x) \geq \min\{f_t(x_0) - \epsilon, \rho\} \text{ for all } t \text{ and } \|x - x_0\| \leq \delta.$$

We say that $\{f_t\}$ is equi-lsc on \mathbb{R}^n if $\{f_t\}$ is equi-lsc at every point $x_0 \in \mathbb{R}^n$.

LEMMA 2.4. *Let $\{\epsilon_t\}$ be a decreasing positive sequence with $\epsilon_t \downarrow 0$ and define $\varphi_{\epsilon_t}(s) = 1 - \frac{\log(|s| + \epsilon_t)}{\log \epsilon_t}$. Then the sequence of functions $\{\varphi_{\epsilon_t}\}$ is equi-lsc on \mathbb{R} .*

Proof. We prove by contradiction. Let $s_0 \in \mathbb{R}$. Suppose that $\{\varphi_{\epsilon_t}\}$ is not equi-lsc at s_0 . Then we see from Definition 2.3 that there exist $\epsilon_0 > 0$, $\rho_0 > 0$, a sequence $\{t_j\}$, and a sequence $s_j \rightarrow s_0$ such that

$$(5) \quad \varphi_{\epsilon_{t_j}}(s_j) < \min\{\varphi_{\epsilon_{t_j}}(s_0) - \epsilon_0, \rho_0\} \text{ for all } j.$$

If there exists N such that $t_j \leq N$ infinitely often, by passing to a further subsequence if necessary, we may assume that $t_j \equiv N_0$ for some integer N_0 . But this together with (5) contradicts the continuity of $\varphi_{\epsilon_{N_0}}$ at s_0 .

Thus, we must have $t_j \rightarrow \infty$. We then consider the following two cases:

- (a) Suppose that $s_0 \neq 0$. Then clearly $\lim_{j \rightarrow \infty} \varphi_{\epsilon_{t_j}}(s_j) = 1$ and $\lim_{j \rightarrow \infty} \varphi_{\epsilon_{t_j}}(s_0) = 1$. Hence, we have $\lim_{j \rightarrow \infty} \varphi_{\epsilon_{t_j}}(s_j) > \lim_{j \rightarrow \infty} \varphi_{\epsilon_{t_j}}(s_0) - \epsilon_0$. This contradicts (5).

(b) Suppose that $s_0 = 0$. Using the facts that $\epsilon_{t_j} \downarrow 0$ and $s_j \rightarrow s_0 = 0$, we conclude that there exists a positive integer N such that $0 < \epsilon_{t_j} \leq \epsilon_{t_j} + |s_j| < 1$ for all $j \geq N$, which further implies $\log(\epsilon_{t_j}) \leq \log(\epsilon_{t_j} + |s_j|) < 0$ for all $j \geq N$. Thus, we have $1 \geq \frac{\log(\epsilon_{t_j} + |s_j|)}{\log(\epsilon_{t_j})}$ for all $j \geq N$, and hence $\liminf_{j \rightarrow \infty} \varphi_{\epsilon_{t_j}}(s_j) \geq 0$. On the other hand, we have $\varphi_{\epsilon_{t_j}}(s_0) - \epsilon_0 = -\epsilon_0 < 0$ for all j . This contradicts (5). This completes the proof. \square

3. An algorithmic framework for the MFS_C problem. In this section, we introduce an algorithmic framework for solving the MFS_C problem (1). Our approach is to solve the equivalent reformulation (2). We construct a sequence of approximation problems with continuous objectives, and solve those approximate problems successively to approximate the original ℓ_0 optimization problem in (2). A similar approach was previously used in [26] for solving ℓ_0 minimization problems to find sparse solutions of linear systems.

Our algorithm, which is an epigraphical approximation scheme for the MFS_C problem ($\text{EAS}_{\text{MFS}_C}$), is presented below as Algorithm 1.

Algorithm 1 Epigraphical approximation scheme for MFS_C ($\text{EAS}_{\text{MFS}_C}$)

Step 0. Choose a sequence of continuous functions $\{\Phi_{\epsilon_k}\}$ that is epi-convergent and pointwise convergent to Φ_0 . Choose $\tilde{x}^0 \in \mathbb{R}^n$. Let $k = 1$.

Step 1. (a) Find an approximate minimizer \tilde{x}^k of $\Phi_{\epsilon_k} + \delta_C$ by an iterative algorithm initialized at \tilde{x}^{k-1} .

(b) If a termination criterion is not met, set $k \leftarrow k + 1$ and go to Step 1(a).

In $\text{EAS}_{\text{MFS}_C}$, we first construct a sequence of continuous functions $\{\Phi_{\epsilon_k}\}$. However, different from the literature, we require the sequence of functions to be both *epi-convergent* and *pointwise convergent* to Φ_0 in (2); see [24, Chapter 7] for the definition of epi-convergence. Then, in each iteration of our algorithm, we approximately minimize $\Phi_{\epsilon_k} + \delta_C$ and use the approximate minimizer \tilde{x}^k as an initial point for minimizing $\Phi_{\epsilon_{k+1}} + \delta_C$. It can be shown that if $\tilde{x}^k \in \nu_k\text{-arg min}(\Phi_{\epsilon_k} + \delta_C)$ with $\nu_k \downarrow 0$, then any accumulation point of $\{\tilde{x}^k\}$ is a minimizer of (2); see [24, Theorem 7.46(a)] and [24, Theorem 7.31(b)].

In order to make use of $\text{EAS}_{\text{MFS}_C}$, we need to specify how to construct the sequence of continuous functions $\{\Phi_{\epsilon_k}\}$ and how to solve the corresponding subproblem. We postpone the discussion of the subproblem to the next section. In the remainder of this section, we discuss how the sequence of continuous functions $\{\Phi_{\epsilon_k}\}$ can be constructed. We start with the following theorem, which suggests a simple way of constructing such a sequence.

THEOREM 3.1. *Let $\{\varphi_{\epsilon_k}(\cdot)\}$ be a sequence of continuous functions on \mathbb{R} that is both epi-convergent and pointwise convergent to $|\cdot|_0$ on \mathbb{R} . Define $\Phi_{\epsilon_k}(x) := \sum_{i=1}^n \varphi_{\epsilon_k}(d_{D_i}^2(x))$. Then $\{\Phi_{\epsilon_k}\}$ is both epi-convergent and pointwise convergent to Φ_0 in (2).*

Proof. We start by showing that for each i , the sequence of functions $\{\varphi_{\epsilon_k}(d_{D_i}^2(\cdot))\}$ epi-converges to $|d_{D_i}^2(\cdot)|_0$. In view of [24, Proposition 7.2], it suffices to show that, for each i and $x \in \mathbb{R}^n$, it holds that

$$(6) \quad \begin{cases} \liminf_{k \rightarrow \infty} \varphi_{\epsilon_k}(d_{D_i}^2(x^k)) \geq |d_{D_i}^2(x)|_0 & \text{for every sequence } x^k \rightarrow x, \\ \limsup_{k \rightarrow \infty} \varphi_{\epsilon_k}(d_{D_i}^2(x^k)) \leq |d_{D_i}^2(x)|_0 & \text{for some sequence } x^k \rightarrow x. \end{cases}$$

Since $\{\varphi_{\epsilon_k}(\cdot)\}$ converges pointwise to $|\cdot|_0$, the second relation above holds trivially for the constant sequence $x^k \equiv x$. On the other hand, consider any sequence $x^k \rightarrow x$. Then we have $d_{D_i}^2(x^k) \rightarrow d_{D_i}^2(x)$. Using this together with [24, Proposition 7.2] and the fact that $\{\varphi_{\epsilon_k}(\cdot)\}$ epi-converges to $|\cdot|_0$, we conclude that the first relation in (6) also holds. Thus, we have shown that $\{\varphi_{\epsilon_k}(d_{D_i}^2(\cdot))\}$ epi-converges to $|d_{D_i}^2(\cdot)|_0$.

Now, notice that $\{\varphi_{\epsilon_k}(d_{D_i}^2(\cdot))\}$ also pointwise converges to $|d_{D_i}^2(\cdot)|_0$ since $\{\varphi_{\epsilon_k}(\cdot)\}$ pointwise converges to $|\cdot|_0$. The desired conclusion now follows from these and [24, Theorem 7.46]. This completes the proof. \square

Based on the above theorem, in order to construct the desired sequence $\{\Phi_{\epsilon_k}\}$ as required in $\text{EAS}_{\text{MFS}_C}$, it suffices to construct a sequence of continuous functions $\{\varphi_{\epsilon_k}(\cdot)\}$ that is both epi-convergent and pointwise convergent to $|\cdot|_0$ on \mathbb{R} and define Φ_{ϵ_k} accordingly. We now present some concrete examples of such $\{\varphi_{\epsilon_k}\}$.

EXAMPLE 1. Let $\{\epsilon_k\}$ be a decreasing positive sequence with $\epsilon_k \downarrow 0$.

- (a) Consider $\varphi_{\epsilon}(s) = 1 - \frac{\log(|s|+\epsilon)}{\log \epsilon}$, which appeared in [26, Example 2.3(i)]. We claim that the sequence $\{\varphi_{\epsilon_k}(\cdot)\}$ is both epi-convergent and pointwise convergent to $|\cdot|_0$. First, it is routine to show the pointwise convergence. Also, we know from Lemma 2.4 that the sequence $\{\varphi_{\epsilon_k}\}$ is equi-lsc on \mathbb{R} . This together with the pointwise convergence and [24, Theorem 7.10] shows that $\{\varphi_{\epsilon_k}(\cdot)\}$ also epi-converges to $|\cdot|_0$.
- (b) Consider $\varphi_{\epsilon}(s) = \frac{|s|}{|s|+\epsilon} + \epsilon|s|$, which is a modification of the function in [26, Example 2.6]. We claim that the sequence $\{\varphi_{\epsilon_k}(\cdot)\}$ is both epi-convergent and pointwise convergent to $|\cdot|_0$.

Again, it is routine to show the pointwise convergence. Next, define $\varphi_{\epsilon_k,1}(s) := \frac{|s|}{|s|+\epsilon_k}$ and $\varphi_{\epsilon_k,2}(s) := \epsilon_k|s|$. Then the sequence $\{\varphi_{\epsilon_k,1}(\cdot)\}$ is nondecreasing and converges pointwise to $|\cdot|_0$, and the sequence $\{\varphi_{\epsilon_k,2}\}$ is nonincreasing and converges pointwise to 0. Thus, according to Proposition 7.4(c) and (d) of [24], we see that $\{\varphi_{\epsilon_k,1}(\cdot)\}$ epi-converges to $|\cdot|_0$, and $\{\varphi_{\epsilon_k,2}\}$ epi-converges to 0. Since $\varphi_{\epsilon_k} = \varphi_{\epsilon_k,1} + \varphi_{\epsilon_k,2}$, using the above observations and [24, Theorem 7.46], we conclude further that $\{\varphi_{\epsilon}(\cdot)\}$ epi-converges to $|\cdot|_0$.

Suppose a sequence $\{\Phi_{\epsilon_k}\}$ satisfying the requirement of $\text{EAS}_{\text{MFS}_C}$ is constructed as described in Theorem 3.1. Then subproblems in the following form have to be approximately solved to obtain \tilde{x}^k :

$$(7) \quad \min_{x \in C} \sum_{i=1}^m \varphi_{\epsilon_k}(d_{D_i}^2(x)).$$

This optimization problem is hard to solve in general. Indeed, even when φ_{ϵ_k} is chosen to be a smooth function, the objective function in (7) is still nonsmooth and nonconvex in general if some D_i 's are nonconvex. Thus, in the next section, we will restrict our attention to a special class of $\{\varphi_{\epsilon_k}\}$ and discuss how to solve the corresponding problem (7) efficiently.

4. A subgradient method for subproblems in $\text{EAS}_{\text{MFS}_C}$. In this section, we propose an algorithm for solving the subproblem in Step 1(a) of $\text{EAS}_{\text{MFS}_C}$ in the form of (7) for a large class of choices of φ_{ϵ} . Specifically, let Θ denote the set of level-bounded continuous concave functions $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ that satisfy the following properties:

1. ψ is continuously differentiable on \mathbb{R}_{++} with positive derivative and $\psi(0) = 0$;
2. $\lim_{s \downarrow 0} \psi'(s)$ exists and is positive, and ψ'_+ is Lipschitz continuous on \mathbb{R}_+ .

We consider problems of the following form, for a function $\psi \in \Theta$:

$$(8) \quad \min_{x \in C} \Psi(x) := \sum_{i=1}^m \psi(d_{D_i}^2(x)).$$

We would like to point out that the assumption $\psi \in \Theta$ in (8) is general enough to cover the subproblems that arise in Step 1(a) of Algorithm 1 for the two classes of functions studied in Example 1: $\varphi_\epsilon(s) = \frac{|s|}{|s|+\epsilon} + \epsilon|s|$, $\epsilon > 0$, and $\varphi_\epsilon(s) = 1 - \frac{\log(|s|+\epsilon)}{\log \epsilon}$, $\epsilon \in (0, 1)$. Indeed, for $\varphi_\epsilon(s) = \frac{|s|}{|s|+\epsilon} + \epsilon|s|$, $\epsilon > 0$, the corresponding subproblem (7) takes the form of (8) with $\psi(s) = \frac{s}{s+\epsilon} + \epsilon s$; clearly, $\psi \in \Theta$. On the other hand, for $\varphi_\epsilon(s) = 1 - \frac{\log(|s|+\epsilon)}{\log \epsilon}$, $\epsilon \in (0, 1)$, the subproblem (7) becomes

$$\min_{x \in C} \sum_{i=1}^m \left[1 - \frac{\log(d_{D_i}^2(x) + \epsilon)}{\log \epsilon} \right].$$

Since $\epsilon \in (0, 1)$ so that $-\log \epsilon > 0$, the above problem is equivalent to

$$\min_{x \in C} \sum_{i=1}^m [\log(d_{D_i}^2(x) + \epsilon) - \log \epsilon],$$

which takes the form of (8) with $\psi(s) = \log(s + \epsilon) - \log \epsilon$; it is routine to check that $\psi \in \Theta$.

Notice that (8) is a nonconvex nonsmooth problem in general, and it is not obvious at first glance what algorithm should be applied for solving such a problem. However, in the special case when D_i 's are all convex, the functions $x \mapsto d_{D_i}^2(x)$, $i = 1, \dots, m$, are smooth, and (8) can be solved by the classical gradient projection algorithm. This method can be applied efficiently when the projections onto C and D_i 's can be computed efficiently, because the gradient of Ψ can be computed in terms of projections onto D_i 's:

$$(9) \quad \nabla \Psi(x) = 2 \sum_{i=1}^m \psi'_+(d_{D_i}^2(x)) [x - P_{D_i}(x)].$$

In the general case when D_i 's are possibly nonconvex, the function Ψ is not everywhere differentiable in general. Nevertheless, we still have $2\psi'_+(d_{D_i}^2(x))[x - \xi] \in \partial(\psi \circ d_{D_i}^2)(x)$ whenever $\xi \in P_{D_i}(x)$,⁴ and the element $2\psi'_+(d_{D_i}^2(x))[x - \xi]$ can be computed efficiently if a projection onto D_i can be obtained efficiently. Thus, mimicking the framework of gradient projection algorithm, we propose a subgradient projection algorithm with nonmonotone linesearch for solving (8), in which $\nabla \Psi(x)$ is replaced by an element in $\sum_{i=1}^m \partial(\psi \circ d_{D_i}^2)(x)$. Our algorithm, known as subgradient projection algorithm with nonmonotone linesearch (sGP_{ls}), is presented below as Algorithm 2. Even though this is a subgradient type algorithm, surprisingly, we can show that the linesearch can be terminated after finitely many inner iterations (i.e., the linesearch is well defined), and that the stepsize sequence $\{\alpha_t\}$ in the algorithm has a uniform lower bound (under an additional assumption on the collection of closed sets; see Theorem 4.3 below), unlike classical subgradient methods (see, for example, [9]).

We first establish the well-definedness of the linesearch procedure in sGP_{ls}, which is an immediate consequence of the following proposition. For notational convenience,

⁴This can be proved using the definition of subdifferential, [24, Example 8.53] and $\psi \in \Theta$.

Algorithm 2 Subgradient projection algorithm with nonmonotone linesearch (sGP_{ls}) for (8)

Step 0. Choose $\alpha_{\max} > \alpha_{\min} > 0$, $\eta \in (0, 1)$, $\sigma > 0$ and an integer $M \geq 0$. Set $t = 0$ and pick an $x^0 \in C$.

Step 1. (a) Choose $\alpha_t^0 \in [\alpha_{\min}, \alpha_{\max}]$ and set $\alpha = \alpha_t^0$. Pick any $\xi_i^t \in P_{D_i}(x^t)$ for $i = 1, \dots, m$. Set

$$(10) \quad g^t := 2 \sum_{i=1}^m \psi'_+(d_{D_i}^2(x^t))(x^t - \xi_i^t).$$

(b) Choose any

$$(11) \quad \tilde{u} \in \arg \min_{u \in C} \left\{ \langle g^t, u - x^t \rangle + \frac{1}{2\alpha} \|u - x^t\|^2 \right\}.$$

(c) If

$$(12) \quad \Psi(\tilde{u}) - \max_{[t-M]_+ \leq i \leq t} \Psi(x^i) \leq -\frac{\sigma}{2} \|\tilde{u} - x^t\|^2,$$

go to Step 2. Otherwise, update $\alpha \leftarrow \eta\alpha$ and go to Step 1(b).

Step 2. Let $\alpha_t := \alpha$, $x^{t+1} := \tilde{u}$ and go to Step 1.

given $x \in C$ and $\xi_i \in P_{D_i}(x)$ for all i , for each $\alpha > 0$, let $\tilde{u}(\alpha)$ denote a minimizer of the problem

$$(13) \quad \min_{u \in C} \langle g, u - x \rangle + \frac{1}{2\alpha} \|u - x\|^2,$$

where $g := 2 \sum_{i=1}^m \psi'_+(d_{D_i}^2(x))(x - \xi_i)$.

PROPOSITION 4.1 (Sufficient descent). *Let Ψ be defined in (8) with $\psi \in \Theta$, $x \in C$ and $\xi_i \in P_{D_i}(x)$ for all i . Let $\alpha > 0$ and $\tilde{u}(\alpha)$ be defined in (13). Then there exists $\beta > 0$ so that*

$$(14) \quad \Psi(\tilde{u}(\alpha)) - \Psi(x) \leq \left(-\frac{1}{\beta} + \sum_{i=1}^m \psi'_+(d_{D_i}^2(x)) \right) \|\tilde{u}(\alpha) - x\|^2.$$

Indeed, one can take $\beta = \alpha$ when C is in addition convex and set $\beta = 2\alpha$ otherwise.

Proof. Since ψ is concave and ψ'_+ is continuous on \mathbb{R}_+ , we see that

$$(15) \quad \begin{aligned} \Psi(\tilde{u}(\alpha)) &\leq \sum_{i=1}^m \{ \psi(d_{D_i}^2(x)) + \psi'_+(d_{D_i}^2(x)) [d_{D_i}^2(\tilde{u}(\alpha)) - d_{D_i}^2(x)] \} \\ &= \Psi(x) + \sum_{i=1}^m \psi'_+(d_{D_i}^2(x)) [d_{D_i}^2(\tilde{u}(\alpha)) - d_{D_i}^2(x)]. \end{aligned}$$

Next, from the definition of distance function, we see that, for each fixed i ,

$$d_{D_i}^2(x) = \|x\|^2 - 2h_i(x),$$

where $h_i(x) := \sup\{\langle x, y \rangle - \frac{1}{2}\|y\|^2 : y \in D_i\}$. Notice that h_i is finite everywhere and is the pointwise supreme of affine functions. Thus, h_i is a continuous convex function and one can check directly from definition that $P_{D_i}(x) \subseteq \partial h_i(x)$. Then we have

$$\begin{aligned}
& \sum_{i=1}^m \psi'_+(d_{D_i}^2(x)) [d_{D_i}^2(\tilde{u}(\alpha)) - d_{D_i}^2(x)] \\
&= \sum_{i=1}^m \psi'_+(d_{D_i}^2(x)) \{ \|\tilde{u}(\alpha)\|^2 - \|x\|^2 - 2[h_i(\tilde{u}(\alpha)) - h_i(x)] \} \\
(16) \quad &\leq \sum_{i=1}^m \psi'_+(d_{D_i}^2(x)) [\|\tilde{u}(\alpha)\|^2 - \|x\|^2 - 2\langle \xi_i, \tilde{u}(\alpha) - x \rangle] \\
&= \langle g, \tilde{u}(\alpha) - x \rangle + \sum_{i=1}^m \psi'_+(d_{D_i}^2(x)) \|\tilde{u}(\alpha) - x\|^2,
\end{aligned}$$

where the first inequality holds because $\psi'_+ > 0$, h_i is convex and $\xi_i \in P_{D_i}(x) \subseteq \partial h_i(x)$ and the last equality holds because of the relation $\|\tilde{u}(\alpha)\|^2 - \|x\|^2 - 2\langle \xi_i, \tilde{u}(\alpha) - x \rangle = \|\tilde{u}(\alpha) - x\|^2 + 2\langle x - \xi_i, \tilde{u}(\alpha) - x \rangle$ and the definition of g in (13).

Now, using the definition of $\tilde{u}(\alpha)$ as a minimizer of (13) and the fact that $x \in C$, we see that $\langle g, \tilde{u}(\alpha) - x \rangle + \frac{1}{2\alpha} \|\tilde{u}(\alpha) - x\|^2 \leq 0$. Combining this with (15) and (16), we see that (14) holds with $\beta = 2\alpha$.

Finally, suppose C is in addition convex. Then the function $f(u) := \langle g, u - x \rangle + \frac{1}{2\alpha} \|u - x\|^2 + \delta_C(u)$ is strongly convex with modulus $\frac{1}{\alpha}$. Using this and the definition of $\tilde{u}(\alpha)$ as a minimizer of (13), we see that $f(x) - f(\tilde{u}(\alpha)) \geq \frac{1}{2\alpha} \|\tilde{u}(\alpha) - x\|^2$. Rearranging terms, we have

$$\langle g, \tilde{u}(\alpha) - x \rangle \leq -\frac{1}{\alpha} \|\tilde{u}(\alpha) - x\|^2.$$

This together with (15) and (16) implies that (14) holds with $\beta = \alpha$. This completes the proof. \square

Using Proposition 4.1, it is then routine to show the well-definedness of the linesearch procedure in sGP_{ts}.

COROLLARY 4.2 (Well-definedness of linesearch). *Let Ψ be defined in (8) with $\psi \in \Theta$ and suppose that the sGP_{ts} is applied for solving (8). Then, in each iteration, the linesearch criterion in Step 1(c) is satisfied after finitely many inner iterations.*

We next show that the stepsize sequence $\{\alpha_t\}$ generated in sGP_{ts} for solving (8) has a *uniform* lower bound under the additional assumption that $C^\infty \cap (\bigcap_{i=1}^m D_i^\infty) = \{0\}$.⁵ We will also show that the sequence $\{x^t\}$ generated by sGP_{ts} is bounded and any accumulation point is a stationary point of the function $\Psi + \delta_C$ in (8). Here, we say that \bar{x} is a stationary point of $\Psi + \delta_C$ if

$$(17) \quad 0 \in \sum_{i=1}^m [2\psi'_+(d_{D_i}^2(\bar{x}))(\bar{x} - P_{D_i}(\bar{x}))] + N_C(\bar{x}).$$

Note that if \bar{x} is a local minimizer of $\Psi + \delta_C$, then in view of [24, Theorem 10.1], [24, Corollary 10.9], [18, Theorem 1.110(ii)] and [24, Example 8.53], one can show that \bar{x} is a stationary point of $\Psi + \delta_C$.

⁵This assumption is satisfied if C or any D_i is bounded.

THEOREM 4.3. *Suppose that $C^\infty \cap (\bigcap_{i=1}^m D_i^\infty) = \{0\}$ and let Ψ be defined in (8) with $\psi \in \Theta$. Let $\{x^t\}$, $\{\alpha_t\}$ and $\{\xi_i^t\}$, $i = 1, \dots, m$, be the sequences generated by sGP_{ls} . Then the following statements hold.*

- (a) *The sequences $\{x^t\}$ and $\{\xi_i^t\}$, $i = 1, \dots, m$, are all bounded, and $\inf_{t \geq 0} \alpha_t > 0$.*
- (b) *It holds that $\lim_{t \rightarrow \infty} \|x^{t+1} - x^t\| = 0$ and the limit $\lim_{t \rightarrow \infty} \Psi(x^t)$ exists.*
- (c) **(Global subsequential convergence)** *Any cluster point of $\{x^t\}$ is a stationary point of $\Psi + \delta_C$.*

Proof. (a) We first note from (12) and a simple induction argument that $\Psi(x^t) \leq \Psi(x^0)$ for all $t \geq 1$. This together with the nonnegativity and level-boundedness of ψ implies that the sequence $\{d_{D_i}^2(x^t)\}$ is bounded for each $i = 1, \dots, m$. Hence, there exists $M_1 > 0$ such that

$$(18) \quad \|x^t - \xi_i^t\| = d_{D_i}(x^t) \leq M_1 \text{ for all } t \text{ and all } i,$$

where the first equality holds because $\xi_i^t \in P_{D_i}(x^t)$.

Now, we claim that the sequence $\{x^t\}$ is bounded. Suppose to the contrary that $\{x^t\}$ is unbounded. Then there exists a subsequence $\{x^{t_j}\}$ such that $\lim_{j \rightarrow \infty} \|x^{t_j}\| = +\infty$. By passing to a further subsequence if necessary, we may assume without loss of generality that $\|x^{t_j}\| \neq 0$ for all j and $\lim_{j \rightarrow \infty} \frac{x^{t_j}}{\|x^{t_j}\|} = d$ for some d . Then $\|d\| = 1$ and we also have from the definition of horizon cone that $d \in C^\infty$. Next, dividing $\|x^{t_j}\|$ from both sides of (18), we see that

$$0 \leq \left\| \frac{\xi_i^{t_j}}{\|x^{t_j}\|} - \frac{x^{t_j}}{\|x^{t_j}\|} \right\| = \frac{\|\xi_i^{t_j} - x^{t_j}\|}{\|x^{t_j}\|} \leq \frac{M_1}{\|x^{t_j}\|}.$$

Passing to the limit in the above inequality, we see that $\lim_{k \rightarrow \infty} \frac{\xi_i^{t_j}}{\|x^{t_j}\|} = d$ for each i .

Using this together with the definition of horizon cone of D_i^∞ and the fact $\xi_i^{t_j} \in D_i$ for all i and j , we conclude further that $d \in \bigcap_{i=1}^m D_i^\infty$. Since we also have $\|d\| = 1$ and $d \in C^\infty$, this contradicts the assumption that $C^\infty \cap \bigcap_{i=1}^m D_i^\infty = \{0\}$. Thus, the sequence $\{x^t\}$ is bounded. In view of (18), we see that $\{\xi_i^t\}$ is also bounded for all i .

Next, we show that $\inf_{t \geq 0} \alpha_t > 0$. We first note from (18) and the positivity and continuity of ψ'_+ on \mathbb{R}_+ that there exists $M_2 > 0$ so that $0 \leq \psi'_+(d_{D_i}^2(x^t)) \leq M_2$ for all t and i . Using this fact and applying Proposition 4.1 with $x = x^t$ and $\xi_i = \xi_i^t$, we see that

$$\Psi(\tilde{u}(\alpha)) - \Psi(x^t) \leq \left(-\frac{1}{2\alpha} + mM_2 \right) \|\tilde{u}(\alpha) - x^t\|^2,$$

where $\tilde{u}(\alpha)$ is defined as in (13) with $x = x^t$ and $g = g^t$ defined in (10), i.e., $\tilde{u}(\alpha)$ is a minimizer of (11) in the t -th iteration. Thus, the linesearch criterion (12) is satisfied if $\alpha \leq (2mM_2 + \sigma)^{-1}$. Hence, using the definition of α_t , we must have either $\alpha_t = \alpha_t^0 \geq \alpha_{\min}$ (if $\alpha_t^0 \leq (2mM_2 + \sigma)^{-1}$) or $\alpha_t > \eta(2mM_2 + \sigma)^{-1}$. Consequently, it holds that $\inf_{t \geq 0} \alpha_t \geq \min\{\alpha_{\min}, \eta(2mM_2 + \sigma)^{-1}\} > 0$.

(b) This can be proved similarly as in [25, Lemma 4].

(c) Let x^* be a cluster point of the bounded sequence $\{x^t\}$ and let $\{x^{t_j}\}$ be a convergent subsequence with limit x^* . Since $\{\xi_i^t\}$ is also bounded for all i , by passing to a further subsequence if necessary, we may assume without loss of generality that, for each $i = 1, \dots, m$, $\xi_i^{t_j} \rightarrow \xi_i^*$ for some ξ_i^* . Next, using the definition of x^{t_j+1} as a

minimizer of (11) when $\alpha = \alpha_{t_j}$, we see that

$$(19) \quad 0 \in 2 \sum_{i=1}^m \psi'_+(d_{D_i}^2(x^{t_j}))(x^{t_j} - \xi_i^{t_j}) + \frac{1}{\alpha_{t_j}}(x^{t_j+1} - x^{t_j}) + N_C(x^{t_j+1}).$$

Notice that $\psi'_+(d_{D_i}^2(\cdot))$ is continuous, $\lim_{t \rightarrow \infty} \|x^{t+1} - x^t\| = 0$ according to (b) and $\{\frac{1}{\alpha_t}\}$ is bounded according to (a). Then, passing to the limit as $j \rightarrow \infty$ in (19), we have

$$0 \in \sum_{i=1}^m [2\psi'_+(d_{D_i}^2(x^*))(x^* - \xi_i^*)] + N_C(x^*).$$

To complete the proof, it now remains to show that, for each i , $\xi_i^* \in P_{D_i}(x^*)$. To this end, we first note that $\xi_i^* \in D_i$ because D_i is closed and $\{\xi_i^{t_j}\} \subseteq D_i$. In addition, we have $d_{D_i}(x^*) = \|x^* - \xi_i^*\|$ because $d_{D_i}(x^{t_j}) = \|x^{t_j} - \xi_i^{t_j}\|$ (thanks to $\xi_i^{t_j} \in P_{D_i}(x^{t_j})$). Thus, we conclude that $\xi_i^* \in P_{D_i}(x^*)$ and this completes the proof. \square

Before ending this section, as a little digression and an immediate application of Theorem 4.3, we discuss global convergence of the averaged projection algorithm. Averaged projection algorithm is a popular algorithm for finding a point of intersection of several nonempty closed sets D_1, \dots, D_m . In this algorithm, we initialize at an $x^0 \in \mathbb{R}^n$ and update

$$(20) \quad x^{t+1} \in \frac{1}{m} \sum_{i=1}^m P_{D_i}(x^t).$$

When each D_i is convex, the above algorithm is just the proximal gradient algorithm applied to $\frac{1}{2m} \sum_{i=1}^m d_{D_i}^2(x)$ with constant stepsize 1, and its convergence is well known. However, the global convergence of the above algorithm is unknown if D_i 's are non-convex: only local convergence was proved recently in [15] under suitable regularity assumptions.

We next show that the averaged projection algorithm (20) is a special case of sGP_{l_s} when $C = \mathbb{R}^n$ and $\psi(s) = \frac{s}{m}$, which clearly belongs to Θ . Hence, we obtain as an immediate corollary of Theorem 4.3 that the averaged projection algorithm (20) is globally subsequentially convergent when $\bigcap_{i=1}^m D_i^\infty = \{0\}$.

PROPOSITION 4.4. *Suppose that $C = \mathbb{R}^n$, $\psi(s) = \frac{s}{m}$ and let Ψ be defined in (8). Let $\alpha_t^0 = \frac{1}{2}$ for all t and $0 < \sigma \leq 2$ in sGP_{l_s}. Then the sGP_{l_s} reduces to the averaged projection algorithm.*

Proof. It suffices to show that $x^{t+1} = \frac{1}{m} \sum_{i=1}^m \xi_i^t$ in every iteration of sGP_{l_s} under the assumptions. To this end, we first apply Proposition 4.1 with $x = x^t$ and $\xi_i = \xi_i^t$, and invoke $\psi'_+ \equiv \frac{1}{m}$ to obtain

$$\Psi(\tilde{u}(0.5)) - \Psi(x^t) \leq (-2 + 1)\|\tilde{u}(0.5) - x^t\|^2 \leq -\frac{\sigma}{2}\|\tilde{u}(0.5) - x^t\|^2,$$

where $\tilde{u}(\alpha)$ is defined as in (13) with $x = x^t$ and $g = g^t$ defined in (10) (i.e., $\tilde{u}(\alpha)$ is a minimizer of (11) in the t -th iteration.), and the last inequality holds because $0 < \sigma \leq 2$. This implies that $\tilde{u}(0.5)$ satisfies (12) and hence $x^{t+1} = \tilde{u}(0.5)$. Using this, the first-order optimality condition of the subproblem (11) with $\alpha = 0.5$ and the fact that $C = \mathbb{R}^n$, we see further that $0 = g^t + 2(x^{t+1} - x^t)$. Thus

$$x^{t+1} = x^t - \frac{1}{2}g^t = x^t - \left(\frac{1}{m} \sum_{i=1}^m [x^t - \xi_i^t] \right) = \frac{1}{m} \sum_{i=1}^m \xi_i^t,$$

where the second equality follows from (10) and the fact that $\psi'_+ \equiv \frac{1}{m}$ on \mathbb{R}_+ . This completes the proof. \square

COROLLARY 4.5. *Suppose that $\bigcap_{i=1}^m D_i^\infty = \{0\}$ and let $\{x^t\}$ be the sequence generated by the averaged projection algorithm (20). Then the sequence $\{x^t\}$ is bounded, $\lim_{t \rightarrow \infty} \|x^{t+1} - x^t\| = 0$ and any cluster point of $\{x^t\}$ is a stationary point of $\Psi + \delta_C$ in (8) with $\psi(s) = \frac{s}{m}$ and $C = \mathbb{R}^n$.*

5. Global sequential convergence of sGP_{l_s} with $M = 0$. In this section, we study convergence of the whole sequence generated by sGP_{l_s} with $M = 0$ for solving (8). We consider two cases in Sections 5.1 and 5.2, respectively: (1) each D_i is convex; (2) some D_i 's are possibly nonconvex and $C = \mathbb{R}^n$. We establish global convergence of the whole sequence generated by sGP_{l_s} with $M = 0$ in these two cases by assuming KL properties of suitable potential functions. Then, in Section 5.3, we discuss a relationship between the KL properties used in Section 5.1 and Section 5.2.

5.1. Global sequential convergence of sGP_{l_s} with $M = 0$ for convex D_i . In this subsection, we assume that each D_i is convex, but C can still be possibly nonconvex. We show in the next theorem the global convergence of the whole sequence generated by sGP_{l_s} with $M = 0$ under the assumption that $\Psi + \delta_C$ is a KL function.

THEOREM 5.1. *Suppose that each D_i is convex, $C^\infty \cap \bigcap_{i=1}^m D_i^\infty = \{0\}$ and $\Psi + \delta_C$ is a KL function, where Ψ is defined in (8) with $\psi \in \Theta$. Let $\{x^t\}$ be the sequence generated by sGP_{l_s} with $M = 0$. Then $\{x^t\}$ is globally convergent to a stationary point of $\Psi + \delta_C$.*

Proof. In view of Theorem 4.3, it suffices to show that the sequence $\{x^t\}$ is convergent. Since $\text{dom} \partial(\Psi + \delta_C) = C$ thanks to the smoothness of Ψ , and $\Psi + \delta_C$ is a KL function by assumption, according to [7, Theorem 2.9], we only need to check that $\{x^t\}$ satisfies the conditions **H1**, **H2** and **H3** there.

H1: Since $M = 0$ and $\{x^t\} \subseteq C$, we see from (12) that $\Psi(x^{t+1}) + \frac{\sigma}{2} \|x^{t+1} - x^t\|^2 \leq \Psi(x^t)$.

H2: We need to check that for each t there exists $\omega^{t+1} \in \partial(\Psi + \delta_C)(x^{t+1})$ so that $\|\omega^{t+1}\| \leq b \|x^{t+1} - x^t\|$ for some $b > 0$ independent of t .

To this end, we note first that x^{t+1} is a minimizer of (11) when $\alpha = \alpha_t$. Hence, we have $-\nabla \Psi(x^t) - \frac{1}{\alpha_t}(x^{t+1} - x^t) \in N_C(x^{t+1})$. Now, define

$$\omega^{t+1} := \nabla \Psi(x^{t+1}) - \nabla \Psi(x^t) - \frac{1}{\alpha_t}(x^{t+1} - x^t).$$

Then we have $\omega^{t+1} \in \nabla \Psi(x^{t+1}) + N_C(x^{t+1}) = \partial(\Psi + \delta_C)(x^{t+1})$, where the equality follows from [24, Exercise 8.8]. On the other hand, note from (9) that $\nabla \Psi$ is locally Lipschitz because ψ'_+ is Lipschitz. Also, recall from Theorem 4.3(a) that $\{x^t\}$ is bounded. Using these and the definition of w^{t+1} , we conclude that

$$\|w^{t+1}\| = \left\| \nabla \Psi(x^{t+1}) - \nabla \Psi(x^t) - \frac{1}{\alpha_t}(x^{t+1} - x^t) \right\| \leq \left(c + \frac{1}{\alpha_0} \right) \|x^{t+1} - x^t\|,$$

where c is the Lipschitz continuity modulus of $\nabla \Psi$ on a compact set containing $\{x^t\}$, and $\alpha_0 := \inf_{t \geq 0} \alpha_t$, which is positive thanks to Theorem 4.3(a). Thus, **H2** holds.

H3: This follows from the boundedness of $\{x^t\}$ by Theorem 4.3(a), the continuity of Ψ and the closedness of C . \square

Based on the assumptions of Theorem 5.1, it is now routine (see [6, Theorem 3.4] for a similar analysis) to establish the local convergence rate of the sequence generated by sGP_{l_s} with $M = 0$ under the additional assumption that $\Psi + \delta_C$ is a KL

function with exponent $\theta \in [0, 1)$. In particular, an exponent of $\theta = \frac{1}{2}$ implies local linear convergence of the sequence generated. Thus, in the next theorem, we discuss conditions on $\{C, D_1, \dots, D_m\}$ that will guarantee $\Psi + \delta_C$ to be a KL function with exponent $\frac{1}{2}$. Specifically, we assume that $\{C, D_1, \dots, D_m\}$ is a collection of closed convex sets that is boundedly linearly regular. Recall that a collection of closed convex sets $\{C, D_1, \dots, D_m\}$ is boundedly linearly regular if $C \cap (\bigcap_{i=1}^m D_i) \neq \emptyset$ and for every bounded set B , there exists $c > 0$ so that

$$\text{dist} \left(x, C \cap \left(\bigcap_{i=1}^m D_i \right) \right) \leq c \max \{ d_C(x), \max_{1 \leq i \leq m} d_{D_i}(x) \}$$

whenever $x \in B$. It is known that when C and D_i are polyhedra with nonempty intersection, then $\{C, D_1, \dots, D_m\}$ is boundedly linearly regular; see [8, Theorem 5.6.2].

THEOREM 5.2. *Suppose that C and all D_i 's are convex and let Ψ be defined in (8) with $\psi \in \Theta$. If $\{C, D_1, \dots, D_m\}$ is boundedly linearly regular, then $\Psi + \delta_C$ is a KL function with exponent $\frac{1}{2}$.*

Proof. For notational simplicity, we write $D = \bigcap_{i=1}^m D_i$. We first show that the set of stationary points of $\Psi + \delta_C$, denoted by \mathcal{X} , is $C \cap D$. Note that $C \cap D \subseteq \mathcal{X}$ can be shown by a direct verification using the definition of stationary points in (17). Conversely, suppose that $\bar{x} \in \mathcal{X}$, i.e., it satisfies (17). Note that $d_{D_i}^2$ is convex due to the convexity of D_i , C is convex by assumption and $\psi'_+ > 0$ on \mathbb{R}_+ . Hence, the function $x \mapsto \sum_{i=1}^m [\psi'_+(d_{D_i}^2(\bar{x}))d_{D_i}^2(x)] + \delta_C(x)$ is a convex function. Using this, (17) and $\nabla d_{D_i}^2(x) = 2(x - P_{D_i}(x))$, we see further that

$$\bar{x} \in \arg \min_{x \in C} \sum_{i=1}^m \psi'_+(d_{D_i}^2(\bar{x}))d_{D_i}^2(x) = C \cap D,$$

where the equality holds because $\psi'_+ > 0$ on \mathbb{R}_+ and $C \cap D \neq \emptyset$. Thus, we have shown that $C \cap D = \mathcal{X}$.

Since $\mathcal{X} = C \cap D = \arg \min(\Psi + \delta_C)$, in view of [16, Lemma 2.1] and Definition 2.1, we only need to check that, for any fixed $\bar{x} \in C \cap D$, there exist positive numbers c and r so that

$$\text{dist}(0, \partial(\Psi + \delta_C)(x)) \geq c(\Psi(x) - \Psi(\bar{x}))^{\frac{1}{2}} \quad \text{whenever } x \in C \cap \mathbf{B}(\bar{x}, r).$$

To this end, fix any $\bar{x} \in C \cap D$ and any $r > 0$. Since $\psi'_+(d_{D_i}^2(\cdot))$ is continuous on $\mathbf{B}(\bar{x}, r)$ and $\psi'_+ > 0$ on \mathbb{R}_+ , we see that there exist ν_1 and ν_2 such that $0 < \nu_1 \leq \psi'_+(d_{D_i}^2(x)) \leq \nu_2$ for all $x \in \mathbf{B}(\bar{x}, r)$ and all $i = 1, \dots, m$. Now, define $\omega_i(x) := \frac{\psi'_+(d_{D_i}^2(x))}{\sum_{j=1}^m \psi'_+(d_{D_j}^2(x))}$. Then we have for all $x \in \mathbf{B}(\bar{x}, r)$ that

$$(21) \quad \sum_{i=1}^m \omega_i(x) = 1 \quad \text{and} \quad \min_{1 \leq i \leq m} \omega_i(x) \geq \nu := \frac{\nu_1}{m\nu_2} > 0.$$

Let $x \in C \cap \mathbf{B}(\bar{x}, r)$ and $y \in C \cap D$. Since $y \in D_i$ for each i , it follows that

$$(22) \quad \langle P_{D_i}(x) - x, y - x \rangle = \langle P_{D_i}(x) - x, y - P_{D_i}(x) + P_{D_i}(x) - x \rangle \geq d_{D_i}^2(x),$$

where the inequality holds because $\langle P_{D_i}(x) - x, y - P_{D_i}(x) \rangle \geq 0$, which is a consequence of the convexity of D_i and the definition of $P_{D_i}(x)$. Now, for any $\zeta \in N_C(x)$, we have

$$\begin{aligned} \sum_{i=1}^m \omega_i(x) d_{D_i}^2(x) &\leq \left\langle \sum_{i=1}^m \omega_i(x) P_{D_i}(x) - x, y - x \right\rangle \\ &= \left\langle \sum_{i=1}^m \omega_i(x) P_{D_i}(x) - x - \zeta, y - x \right\rangle + \langle \zeta, y - x \rangle \\ &\leq \left\| \sum_{i=1}^m \omega_i(x) P_{D_i}(x) - x - \zeta \right\| \|y - x\|, \end{aligned}$$

where the first inequality is due to (21) and (22), the second inequality holds because $\zeta \in N_C(x)$ and $y \in C$. Taking infimum over $y \in C \cap D$ and $\zeta \in N_C(x)$ in the above inequality and noting that

$$\partial(\Psi + \delta_C)(x) = 2 \left(\sum_{j=1}^m \psi'_+(d_{D_j}^2(x)) \right) \left[x - \sum_{i=1}^m \omega_i(x) P_{D_i}(x) + N_C(x) \right],$$

we obtain further that

$$\sum_{i=1}^m \omega_i(x) d_{D_i}^2(x) \leq c_1 \text{dist}(0, \partial(\Psi + \delta_C)(x)) \cdot d_{C \cap D}(x),$$

where $c_1 = (2m\nu_1)^{-1}$. Using this together with the bounded linear regularity of $\{C, D_1, \dots, D_m\}$, we see further that for any $x \in C \cap \mathbf{B}(\bar{x}, r)$, we have

$$\begin{aligned} \nu \max_{1 \leq j \leq m} \{d_{D_j}^2(x)\} &\leq \sum_{i=1}^m \omega_i(x) d_{D_i}^2(x) \leq c_1 \text{dist}(0, \partial(\Psi + \delta_C)(x)) \cdot d_{C \cap D}(x) \\ &\leq c_2 \text{dist}(0, \partial(\Psi + \delta_C)(x)) \max_{1 \leq j \leq m} \{d_{D_j}(x)\} \end{aligned}$$

for some constant $c_2 > 0$, where the first inequality follows from (21), and the last inequality holds because $\{C, D_1, \dots, D_m\}$ is boundedly linearly regular and $x \in C$. Thus, we have for all $x \in C \cap \mathbf{B}(\bar{x}, r)$ that

$$(23) \quad \max_{1 \leq j \leq m} \{d_{D_j}(x)\} \leq \frac{c_2}{\nu} \text{dist}(0, \partial(\Psi + \delta_C)(x)).$$

On the other hand, note that for any $x \in C \cap \mathbf{B}(\bar{x}, r)$, we have $P_{C \cap D}(x) \in \mathbf{B}(\bar{x}, r)$ since the projection operator is nonexpansive and $\bar{x} \in C \cap D$. Moreover, note from (9) that $\nabla \Psi(x)$ is locally Lipschitz because ψ'_+ is Lipschitz. Thus, we deduce further that for any $x \in C \cap \mathbf{B}(\bar{x}, r)$,

$$\begin{aligned} \Psi(x) - \Psi(\bar{x}) &= \Psi(x) - \Psi(P_{C \cap D}(x)) \\ &\leq \langle \nabla \Psi(P_{C \cap D}(x)), x - P_{C \cap D}(x) \rangle + \frac{L_1}{2} \|x - P_{C \cap D}(x)\|^2 \\ &= \frac{L_1}{2} d_{C \cap D}^2(x) \leq \frac{L_1}{2} c_3 \max_{1 \leq j \leq m} \{d_{D_j}^2(x)\} \leq c_4 \text{dist}^2(0, \partial(\Psi + \delta_C)(x)), \end{aligned}$$

where L_1 is the Lipschitz continuity modulus of $\nabla \Psi$ on $\mathbf{B}(\bar{x}, r)$, the first equality holds because $\Psi(\bar{x}) = \Psi(P_{C \cap D}(x)) = 0$, the second equality holds because $\nabla \Psi(P_{C \cap D}(x)) = 0$ by direct computation, the existence of c_3 follows from the bounded linear regularity of $\{C, D_1, \dots, D_m\}$ and $x \in C$, and the existence of c_4 follows from (23). This completes the proof. \square

5.2. Global sequential convergence of sGP_{l_s} with $M = 0$ for nonconvex D_i . In this subsection, we assume that $C = \mathbb{R}^n$ but allow each D_i to be possibly nonconvex: note that the function $d_{D_i}^2$ is not smooth when D_i is nonconvex. We will study global convergence of the whole sequence generated by sGP_{l_s} with $M = 0$ for solving (8) in this case.

Our analysis below will be based on the following potential function:

$$(24) \quad \mathfrak{M}(x, \boldsymbol{\xi}, u) := \sum_{i=1}^m [\rho^*(u_i) - u_i \|x - \boldsymbol{\xi}_i\|^2 + \delta_{D_i}(\boldsymbol{\xi}_i)],$$

where $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_m) \in \mathbb{R}^{mn}$ with each $\boldsymbol{\xi}_i \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, ρ is the convex function defined by

$$(25) \quad \rho(s) := \begin{cases} -\psi(s) & \text{if } s \geq 0, \\ s^2 - ls & \text{otherwise,} \end{cases}$$

with $\psi \in \Theta$ (see (8)) and $l := \psi'_+(0) > 0$, and ρ^* is the convex conjugate. One can show that

$$(26) \quad \rho'(s) = \begin{cases} -\psi'_+(s) & \text{if } s \geq 0, \\ 2s - l & \text{otherwise.} \end{cases}$$

We then see immediately from (25), (26) and $\psi \in \Theta$ that ρ is a continuously differentiable nonincreasing convex function on \mathbb{R} and ρ' is Lipschitz continuous on \mathbb{R} .

We collect some essential properties of the potential function \mathfrak{M} that will be useful in our subsequent analysis. First, fix any $x \in \mathbb{R}^n$. For each i , pick any $\hat{\boldsymbol{\xi}}_i \in P_{D_i}(x)$ and define $\hat{u}_i := \rho'(d_{D_i}^2(x))$. Then we see from (26) that

$$(27) \quad \hat{u}_i = \rho'(d_{D_i}^2(x)) = -\psi'_+(d_{D_i}^2(x)) < 0.$$

From this, we deduce further that

$$(28) \quad \mathfrak{M}(x, \hat{\boldsymbol{\xi}}, \hat{u}) = \sum_{i=1}^m [\rho^*(\hat{u}_i) - \hat{u}_i d_{D_i}^2(x)] = \sum_{i=1}^m [-\rho(d_{D_i}^2(x))] = \Psi(x) \geq 0,$$

where the first equality holds because $\hat{\boldsymbol{\xi}}_i \in P_{D_i}(x)$, while the second equality follows from (27) and [24, Proposition 11.3]. Finally, using [24, Exercise 8.8] and [24, Proposition 10.5], we have the following formula for the subdifferential of \mathfrak{M} at any $(x, \boldsymbol{\xi}, u) \in \mathbb{R}^n \times \mathbb{R}^{mn} \times \mathbb{R}^m$:

$$(29) \quad \partial \mathfrak{M}(x, \boldsymbol{\xi}, u) = \begin{bmatrix} -2 \sum_{i=1}^m u_i (x - \boldsymbol{\xi}_i) \\ [-2u_i(\boldsymbol{\xi}_i - x) + N_{D_i}(\boldsymbol{\xi}_i)]_{i=1}^m \\ [\partial \rho^*(u_i) - \|x - \boldsymbol{\xi}_i\|^2]_{i=1}^m \end{bmatrix}.$$

We next bound the distance from the origin to $\partial \mathfrak{M}(x, \boldsymbol{\xi}, u)$ along a certain sequence.

LEMMA 5.3. *Suppose that $\bigcap_{i=1}^m D_i^\infty = \{0\}$, $C = \mathbb{R}^n$ and $\psi \in \Theta$. Let \mathfrak{M} be defined in (24) and let $\{x^t\}$ and $\{\boldsymbol{\xi}_i^t\}$, $i = 1, \dots, m$, be the sequences generated by*

sGP_{l_s} . Define, for each t , $u^t := [\rho'(d_{D_i}^2(x^t))]_{i=1}^m$. Then there exists $c > 0$ such that for all $t \geq 0$, we have

$$(30) \quad \text{dist}(0, \partial \mathfrak{M}(x^t, \xi^t, u^t)) \leq c \|x^{t+1} - x^t\|.$$

Proof. Since $C = \mathbb{R}^n$ and x^{t+1} is a minimizer of (11) when $\alpha = \alpha_t$, we have, using the definition of u^t and the expression of ρ' in (26), that

$$0 = g^t + \frac{1}{\alpha_t}(x^{t+1} - x^t) = -2 \sum_{i=1}^m u_i^t(x^t - \xi_i^t) + \frac{1}{\alpha_t}(x^{t+1} - x^t).$$

Combining this with (29), we deduce that

$$\partial_x \mathfrak{M}(x^t, \xi^t, u^t) = -2 \sum_{i=1}^m u_i^t(x^t - \xi_i^t) = -\frac{1}{\alpha_t}(x^{t+1} - x^t).$$

Since $\underline{\alpha} := \inf_{t \geq 0} \alpha_t > 0$ according to Theorem 4.3(a), we obtain further that

$$(31) \quad \text{dist}(0, \partial_x \mathfrak{M}(x^t, \xi^t, u^t)) \leq \underline{\alpha}^{-1} \|x^{t+1} - x^t\|.$$

Next, recall that $\xi_i^t \in P_{D_i}(x^t)$, which implies $x^t - \xi_i^t \in N_{D_i}(\xi_i^t)$, thanks to [24, Example 6.16]. Moreover, using the expression of ρ' in (26) and the definition of u^t together with the assumption that $\psi \in \Theta$, we have $-u_i^t = -\rho'(d_{D_i}^2(x^t)) = \psi'_+(d_{D_i}^2(x^t)) > 0$. Hence, $-2u_i^t(x^t - \xi_i^t) \in N_{D_i}(\xi_i^t)$. This together with (29) gives $0 \in \partial_{\xi} \mathfrak{M}(x^t, \xi^t, u^t)$. Thus,

$$(32) \quad \text{dist}(0, \partial_{\xi} \mathfrak{M}(x^t, \xi^t, u^t)) = 0.$$

Finally, since $u_i^t = \rho'(d_{D_i}^2(x^t))$ implies $d_{D_i}^2(x^t) \in \partial \rho^*(u_i^t)$ according to [24, Proposition 11.3] and note that $\xi_i^t \in P_{D_i}(x^t)$, we see that $\|x^t - \xi_i^t\|^2 \in \partial \rho^*(u_i^t)$. This together with (29) gives $0 \in \partial_u \mathfrak{M}(x^t, \xi^t, u^t)$. The desired bound (30) now follows immediately from this, (31) and (32). \square

We are now ready to present our convergence analysis.

THEOREM 5.4. *Suppose that $\bigcap_{i=1}^m D_i^\infty = \{0\}$, $C = \mathbb{R}^n$, $\psi \in \Theta$ and \mathfrak{M} in (24) is a KL function. Let $\{x^t\}$ be the sequence generated by sGP_{l_s} with $M = 0$. Then $\{x^t\}$ is globally convergent to a stationary point of Ψ .*

Proof. In view of Theorem 4.3, it suffices to show that the sequence $\{x^t\}$ is convergent. We first note that $\max_{[t-M]_+ \leq i \leq t} \Psi(x^i) \equiv \Psi(x^t)$ since $M = 0$. Using this and (12), we have

$$(33) \quad \mathfrak{M}(x^{t+1}, \xi^{t+1}, u^{t+1}) - \mathfrak{M}(x^t, \xi^t, u^t) = \Psi(x^{t+1}) - \Psi(x^t) \leq -\frac{\sigma}{2} \|x^{t+1} - x^t\|^2,$$

where $\{\xi^t\}$ is generated by sGP_{l_s} and $u^t := [\rho'(d_{D_i}^2(x^t))]_{i=1}^m$, and the first equality follows from (28) and the definitions of $\widehat{\xi}$, ξ^t , \widehat{u} and u^t .

Next, we note from Theorem 4.3(a) and the definition of u^t that the sequence $\{(x^t, \xi^t, u^t)\}$ is bounded. Let Ω be the set of cluster points of $\{(x^t, \xi^t, u^t)\}$. Then Ω is nonempty and compact. We now show that \mathfrak{M} is constant on Ω . To this end, we first observe from (28) that $\mathfrak{M}(x^t, \xi^t, u^t) = \Psi(x^t)$ for all t . This together with Theorem 4.3(b) implies that $\lim_{t \rightarrow \infty} \mathfrak{M}(x^t, \xi^t, u^t) = l_*$ for some l_* . Now, take any $(x^*, \xi^*, u^*) \in$

Ω . Then there exists a convergent subsequence $\{(x^{t_j}, \xi^{t_j}, u^{t_j})\}$ converging to it. Using (33), we have

$$\begin{aligned} \mathfrak{M}(x^{t_j}, \xi^{t_j}, u^{t_j}) &\leq \mathfrak{M}(x^{t_j-1}, \xi^{t_j-1}, u^{t_j-1}) - \frac{\sigma}{2} \|x^{t_j} - x^{t_j-1}\|^2 \\ &\leq \mathfrak{M}(x^{t_j-1}, \xi^{t_j-1}, u^*) - \frac{\sigma}{2} \|x^{t_j} - x^{t_j-1}\|^2 \leq \mathfrak{M}(x^{t_j-1}, \xi^*, u^*) - \frac{\sigma}{2} \|x^{t_j} - x^{t_j-1}\|^2, \end{aligned}$$

where the second inequality follows from the definition of $\{u^t\}$ so that u^t is a minimizer of $u \mapsto \mathfrak{M}(x^t, \xi^t, u)$, while the last inequality follows from the facts that $\xi_i^{t_j-1} \in P_{D_i}(x^{t_j-1})$ and that $u_i^t < 0$ (so that $u_i^* \leq 0$) for all i . Passing to the limit in the above inequality and invoking the definition of l_* and Theorem 4.3(b), we deduce that

$$l_* = \lim_{j \rightarrow \infty} \mathfrak{M}(x^{t_j}, \xi^{t_j}, u^{t_j}) \leq \mathfrak{M}(x^*, \xi^*, u^*).$$

Since the converse inequality is an immediate consequence of the lower semicontinuity of \mathfrak{M} , we conclude that $\mathfrak{M} \equiv l_*$ on Ω .

Now, the global convergence of $\{x^t\}$ can be proved based on $\mathfrak{M} \equiv l_*$ on the nonempty compact set $\Omega \subseteq \text{dom} \partial \mathfrak{M}$, (33), (30), Lemma 2.2 and the KL assumption on \mathfrak{M} . The proof is routine (see, for example, [10, Theorem 1]) and we omit the proof for brevity. \square

One can also establish local convergence rate of the sequence generated by sGP_{*l_s*} with $M = 0$ based on the assumptions of Theorem 5.4 and the additional assumption that \mathfrak{M} is a KL function with exponent $\theta \in [0, 1)$; we refer the readers to [6, Theorem 3.4] for a similar analysis.

5.3. Relating the KL exponent of Ψ and \mathfrak{M} when $C = \mathbb{R}^n$. Notice that global convergence of the sequence generated by sGP_{*l_s*} with $M = 0$ was established in Theorems 5.1 and 5.4 under two different KL assumptions. Theorem 5.1 studied the case when each D_i is convex and requires $\Psi + \delta_C$ in (8) to be a KL function, while Theorem 5.4 studied the case when some D_i 's are possibly nonconvex and $C = \mathbb{R}^n$, and requires \mathfrak{M} in (24) to be a KL function. In this section, we study a relationship between these two KL assumptions.

We start with the following lemma, which describes how \mathfrak{M} is related to the stationary points of Ψ in (8) when $C = \mathbb{R}^n$ and each D_i is convex.

LEMMA 5.5. *Suppose that each D_i is convex, $C = \mathbb{R}^n$, Ψ is defined in (8) with $\psi \in \Theta$ and \mathfrak{M} is defined in (24). If $0 \in \partial \mathfrak{M}(\bar{x}, \bar{\xi}, \bar{u})$, then $\mathfrak{M}(\bar{x}, \bar{\xi}, \bar{u}) = \Psi(\bar{x})$, $\nabla \Psi(\bar{x}) = 0$ and*

$$(34) \quad \sum_{i=1}^m \bar{u}_i (\bar{x} - \bar{\xi}_i) = 0, \quad \bar{\xi} = [P_{D_i}(\bar{x})]_{i=1}^m \quad \text{and} \quad \bar{u} = [\rho'(d^2(\bar{x}, D_i))]_{i=1}^m < 0.$$

Proof. We first prove (34). Since $0 \in \partial \mathfrak{M}(\bar{x}, \bar{\xi}, \bar{u})$, we see from (29) that $\bar{\xi}_i \in D_i$, $0 = \sum_{i=1}^m \bar{u}_i (\bar{x} - \bar{\xi}_i)$, $\|\bar{x} - \bar{\xi}_i\|^2 \in \partial \rho^*(\bar{u}_i)$ and $-2\bar{u}_i (\bar{x} - \bar{\xi}_i) \in N_{D_i}(\bar{\xi}_i)$ for each i . Hence, the first relation in (34) holds. Also, combining $\|\bar{x} - \bar{\xi}_i\|^2 \in \partial \rho^*(\bar{u}_i)$ with [24, Proposition 11.3], we obtain

$$(35) \quad \bar{u}_i = \rho'(\|\bar{x} - \bar{\xi}_i\|^2) < 0.$$

Using this and $-2\bar{u}_i (\bar{x} - \bar{\xi}_i) \in N_{D_i}(\bar{\xi}_i)$, we have $\bar{x} - \bar{\xi}_i \in N_{D_i}(\bar{\xi}_i)$. This together with [24, Proposition 6.17] and the convexity of D_i implies that

$$\bar{\xi}_i = (I + N_{D_i})^{-1}(\bar{x}) = P_{D_i}(\bar{x}).$$

In particular, the second relation in (34) holds and $\|\bar{x} - \bar{\xi}_i\|^2 = d^2(\bar{x}, D_i)$. Combining this with (35) gives the third relation in (34). These prove (34).

Next, we deduce from (28), (34) and the definitions of $\hat{\xi}$ and \hat{u} that $\mathfrak{M}(\bar{x}, \bar{\xi}, \bar{u}) = \Psi(\bar{x})$. Finally, from (34) and (26) we see that

$$0 = \sum_{i=1}^m \bar{u}_i(\bar{x} - \bar{\xi}_i) = \sum_{i=1}^m \rho'(d^2(\bar{x}, D_i))(\bar{x} - P_{D_i}(\bar{x})) = \sum_{i=1}^m -\psi'_+(d^2(\bar{x}, D_i))(\bar{x} - P_{D_i}(\bar{x})),$$

i.e., $\nabla \Psi(\bar{x}) = 0$. This completes the proof. \square

We now present our analysis concerning the two different KL assumptions used in Theorems 5.1 and 5.4. In our analysis below, we assume that each D_i is convex, $C = \mathbb{R}^n$ and $\psi \in \Theta$. We also require in addition that ψ is **strict concave** on \mathbb{R}_+ . This latter assumption together with (25) shows that ρ is a **strictly convex** continuously differentiable nonincreasing function on \mathbb{R} . One can check that the functions $s \mapsto \frac{s}{s+\epsilon} + \epsilon s$, $\epsilon > 0$, and $s \mapsto \log(s + \epsilon) - \log(\epsilon)$, $\epsilon \in (0, 1)$, discussed in the beginning of Section 4 are both strictly concave on \mathbb{R}_+ .

Under the additional strict concavity assumption on ψ , we see from [23, Theorem 26.3] that ρ^* is essentially smooth. Thus, $\text{dom } \partial \rho^*$ is open thanks to [23, Theorem 26.1] and ρ^* is indeed continuously differentiable at each $u_i \in \text{dom } \partial \rho^*$ in view of [23, Theorem 25.5]. On the other hand, notice from (29) that $(x, \xi, u) \in \text{dom } \partial \mathfrak{M}$ implies $u_i \in \text{dom } \partial \rho^*$ for each i . Thus, if $(x, \xi, u) \in \text{dom } \partial \mathfrak{M}$, then ρ^* is continuously differentiable at u_i for each i . In the next lemma, we establish some inequalities concerning $(\rho^*)'$ at some special points in $\text{dom } \partial \mathfrak{M}$.

LEMMA 5.6. *Suppose that each D_i is convex, $C = \mathbb{R}^n$, $\psi \in \Theta$ and is strictly concave on \mathbb{R}_+ , Ψ is defined in (8) and \mathfrak{M} is defined in (24). Let $0 \in \partial \mathfrak{M}(\bar{x}, \bar{\xi}, \bar{u})$. Then there exist positive numbers ϵ , L , \bar{c} , c_1 and c_2 so that whenever $(x, \xi, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\xi}, \bar{u}), \epsilon)$, the following inequalities hold:*

$$(36) \quad \mathfrak{M}(x, \xi, u) \leq \Psi(x) + \frac{L}{2} \sum_{i=1}^m [d_{D_i}^2(x) - (\rho^*)'(u_i)]^2 - \sum_{i=1}^m u_i [\|x - \xi_i\|^2 - d_{D_i}^2(x)],$$

$$(37) \quad L|(\rho^*)'(u_i) - d_{D_i}^2(x)| \geq |u_i + \psi'_+(d_{D_i}^2(x))| \text{ for all } i,$$

$$(38) \quad |d_{D_i}^2(x) - (\rho^*)'(u_i)| < m^{-1} \text{ and } 0 \leq \|x - \xi_i\|^2 - d_{D_i}^2(x) < m^{-1} \text{ for all } i,$$

$$(39) \quad -\bar{c} < u_i < 0 \text{ and } u_i^2 \geq c_1 \text{ for all } i,$$

$$(40) \quad \inf_{\mu_i \in N_{D_i}(\xi_i)} \|-2u_i(\xi_i - x) + \mu_i\|^2 \geq c_1 (\|\xi_i - x\|^2 - d_{D_i}^2(x)) \geq 0 \text{ for all } i,$$

$$(41) \quad \inf_{\mu_i \in N_{D_i}(\xi_i)} \|-2u_i(\xi_i - x) + \mu_i\|^2 \geq c_2 \|P_{D_i}(x) - \xi_i\|^2 \text{ for all } i.$$

Proof. Since $(\bar{x}, \bar{\xi}, \bar{u}) \in \text{dom } \partial \mathfrak{M}$, from the discussion preceding this lemma, we see that there exists $\epsilon_0 > 0$ so that $(\rho^*)'$ is continuous at u_i for all i whenever $(x, \xi, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\xi}, \bar{u}), \epsilon_0)$.

Let $(x, \boldsymbol{\xi}, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\boldsymbol{\xi}}, \bar{u}), \epsilon_0)$ and define $\tilde{y}_i = (\rho^*)'(u_i)$ for each i . Then, from [24, Proposition 11.3], we see that

$$(42) \quad u_i = \rho'(\tilde{y}_i) \quad \text{and} \quad \rho^*(u_i) = -\rho(\tilde{y}_i) + \tilde{y}_i u_i \quad \text{for each } i.$$

Moreover, since ρ' is Lipschitz continuous on \mathbb{R} in view of $\psi \in \Theta$ and (26), we have

$$(43) \quad \rho(d_{D_i}^2(x)) \leq \rho(\tilde{y}_i) + u_i(d_{D_i}^2(x) - \tilde{y}_i) + \frac{L}{2}(d_{D_i}^2(x) - \tilde{y}_i)^2,$$

where L is the Lipschitz continuity modulus. Then we have

$$\begin{aligned} \mathfrak{M}(x, \boldsymbol{\xi}, u) &= \sum_{i=1}^m [-u_i \|x - \boldsymbol{\xi}_i\|^2 - \rho(\tilde{y}_i) + \tilde{y}_i u_i] \\ &= \sum_{i=1}^m [-\rho(\tilde{y}_i) - u_i(d_{D_i}^2(x) - \tilde{y}_i) - u_i(\|x - \boldsymbol{\xi}_i\|^2 - d_{D_i}^2(x))] \\ &\leq \sum_{i=1}^m \left[-\rho(d_{D_i}^2(x)) + \frac{L}{2}(d_{D_i}^2(x) - \tilde{y}_i)^2 \right] + \sum_{i=1}^m (-u_i) [\|x - \boldsymbol{\xi}_i\|^2 - d_{D_i}^2(x)] \\ &= \Psi(x) + \frac{L}{2} \sum_{i=1}^m [d_{D_i}^2(x) - (\rho^*)'(u_i)]^2 + \sum_{i=1}^m (-u_i) [\|x - \boldsymbol{\xi}_i\|^2 - d_{D_i}^2(x)], \end{aligned}$$

where the first equality follows from the second relation in (42), the inequality is due to (43) and the last equality follows from the definitions of Ψ and \tilde{y}_i . Furthermore, we deduce that for each i ,

$$|u_i + \psi'_+(d_{D_i}^2(x))| = |\rho'(\tilde{y}_i) - \rho'(d_{D_i}^2(x))| \leq L|\tilde{y}_i - d_{D_i}^2(x)| = L|(\rho^*)'(u_i) - d_{D_i}^2(x)|,$$

where the first equality follows from the first relation in (42) and the expression of ρ' in (26), the inequality follows from the Lipschitz continuity of ρ' and the last equality follows from the definition of \tilde{y}_i . Thus, (36) and (37) hold whenever $(x, \boldsymbol{\xi}, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\boldsymbol{\xi}}, \bar{u}), \epsilon_0)$.

Next, note that (34) holds because $0 \in \partial \mathfrak{M}(\bar{x}, \bar{\boldsymbol{\xi}}, \bar{u})$. Then for each i , it holds that

$$(44) \quad \lim_{u_i \rightarrow \bar{u}_i} (\rho^*)'(u_i) = (\rho^*)'(\bar{u}_i) = (\rho^*)'(\rho'(d_{D_i}^2(\bar{x}))) = d_{D_i}^2(\bar{x}) = \lim_{x \rightarrow \bar{x}} d_{D_i}^2(x),$$

where the second equality follows from the last relation in (34) and the third equality follows from [24, Proposition 11.3]. Moreover, we have for each i ,

$$(45) \quad \lim_{(x, \boldsymbol{\xi}_i) \rightarrow (\bar{x}, \bar{\boldsymbol{\xi}}_i)} \|x - \boldsymbol{\xi}_i\|^2 = \|\bar{x} - \bar{\boldsymbol{\xi}}_i\|^2 = d_{D_i}^2(\bar{x}) = \lim_{x \rightarrow \bar{x}} d_{D_i}^2(x),$$

where the second equality above follows from the second relation in (34). In addition, for any $(x, \boldsymbol{\xi}, u) \in \text{dom } \partial \mathfrak{M}$, we have $\boldsymbol{\xi}_i \in D_i$ for all i and hence $\|x - \boldsymbol{\xi}_i\|^2 \geq d_{D_i}^2(x)$. In view of this, (44) and (45), we conclude that one can further shrink ϵ_0 so that (38) also holds whenever $(x, \boldsymbol{\xi}, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\boldsymbol{\xi}}, \bar{u}), \epsilon_0)$. Now, recall that $\bar{u} < 0$ from the third relation in (34). Thus, one can shrink ϵ_0 further so that (39) holds true for some positive numbers c_1 and \bar{c} whenever $(x, \boldsymbol{\xi}, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\boldsymbol{\xi}}, \bar{u}), \epsilon_0)$.

It now remains to prove (40) and (41). We first prove (40). Let $(x, \boldsymbol{\xi}, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\boldsymbol{\xi}}, \bar{u}), \epsilon_0)$. Fix any i . Then we have $u_i \leq -\sqrt{c_1}$ from (39) and hence

$$(46) \quad \begin{aligned} \inf_{\boldsymbol{\mu}_i \in N_{D_i}(\boldsymbol{\xi}_i)} \left\| -2u_i(\boldsymbol{\xi}_i - x) + \boldsymbol{\mu}_i \right\|^2 &= 4u_i^2 \cdot \inf_{-\frac{1}{2u_i} \boldsymbol{\mu}_i \in N_{D_i}(\boldsymbol{\xi}_i)} \left\| \boldsymbol{\xi}_i - x - \frac{1}{2u_i} \boldsymbol{\mu}_i \right\|^2 \\ &\geq u_i^2 \cdot \inf_{\boldsymbol{\mu}_i \in N_{D_i}(\boldsymbol{\xi}_i)} \left\| \boldsymbol{\xi}_i - x + \boldsymbol{\mu}_i \right\|^2 \geq c_1 \text{dist}^2(0, \partial f(\boldsymbol{\xi}_i)), \end{aligned}$$

where $f(\cdot) := \frac{1}{2}\|\cdot - x\|^2 + \delta_{D_i}(\cdot)$. Since D_i is convex, we see that f is a strongly convex function with modulus 1. This together with $P_{D_i}(x) \in D_i$ and $\xi_i \in D_i$ gives

$$\frac{1}{2}\|P_{D_i}(x) - x\|^2 - \frac{1}{2}\|\xi_i - x\|^2 \geq \langle \eta, P_{D_i}(x) - \xi_i \rangle + \frac{1}{2}\|P_{D_i}(x) - \xi_i\|^2 \quad \text{for all } \eta \in \partial f(\xi_i).$$

Multiplying the above inequality by 2 and rearranging terms, we obtain further that

$$2\langle \eta, \xi_i - P_{D_i}(x) \rangle \geq \|\xi_i - x\|^2 - \|P_{D_i}(x) - x\|^2 + \|P_{D_i}(x) - \xi_i\|^2.$$

This together with the relation $\|\eta\|^2 + \|\xi_i - P_{D_i}(x)\|^2 \geq 2\|\eta\|\|\xi_i - P_{D_i}(x)\|$ gives

$$\|\eta\|^2 \geq \|\xi_i - x\|^2 - \|P_{D_i}(x) - x\|^2 \geq 0 \quad \text{for all } \eta \in \partial f(\xi_i),$$

where the last inequality holds because $\xi_i \in D_i$. This together with (46) shows that (40) holds whenever $(x, \xi, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\xi}, \bar{u}), \epsilon_0)$.

Finally, we show that there exists $\epsilon \in (0, \epsilon_0]$ so that (41) holds whenever $(x, \xi, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\xi}, \bar{u}), \epsilon)$. To this end, we first recall from the second relation in (34) that $\bar{\xi}_i = P_{D_i}(\bar{x})$ for each i . Moreover, recall also from [24, Proposition 6.17] that $(I + N_{D_i})^{-1} = P_{D_i}$, since D_i is convex. Using these together with the nonexpansiveness (and hence, Lipschitz continuity) of P_{D_i} , we see that $(I + N_{D_i})^{-1}$ has the Aubin property at $(\bar{x}, \bar{\xi}_i)$. Hence, $I + N_{D_i}$ is metrically regular at $(\bar{\xi}_i, \bar{x})$ thanks to [24, Theorem 9.43]. Thus, there exist $\epsilon_i > 0$ and $\kappa_i > 0$ such that

$$(47) \quad \kappa_i \|\xi_i - (I + N_{D_i})^{-1}(x)\| \leq \inf_{\omega_i \in (I + N_{D_i})(\xi_i)} \|x - \omega_i\| \quad \text{for all } (x, \xi_i) \in \mathbf{B}((\bar{x}, \bar{\xi}_i), \epsilon_i).$$

On the other hand, let $(x, \xi, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\xi}, \bar{u}), \epsilon_0)$. Then, from (46), we see that $\inf_{\mu_i \in N_{D_i}(\xi_i)} \|-2u_i(\xi_i - x) + \mu_i\|^2$ is bounded below by

$$c_1 \inf_{\mu_i \in N_{D_i}(\xi_i)} \|x - \xi_i - \mu_i\|^2 = c_1 \inf_{\omega_i \in (I + N_{D_i})(\xi_i)} \|x - \omega_i\|^2.$$

Combining this with (47) and recalling that $(I + N_{D_i})^{-1} = P_{D_i}$, we conclude that (41) holds for some $c_2 > 0$ whenever $(x, \xi, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\xi}, \bar{u}), \epsilon)$, where $\epsilon := \min\{\epsilon_i : 0 \leq i \leq m\}$. This completes the proof. \square

We are now ready to prove the main theorem of this section.

THEOREM 5.7. *Suppose that each D_i is convex, $C = \mathbb{R}^n$, $\psi \in \Theta$ and is strictly concave on \mathbb{R}_+ . If Ψ in (8) satisfies the KL property with exponent $\theta \in [\frac{1}{2}, 1)$, then \mathfrak{M} in (24) satisfies the KL property with exponent θ .*

Proof. In view of [16, Lemma 2.1], it suffices to prove that \mathfrak{M} satisfies the KL property with exponent θ at all points $(\bar{x}, \bar{\xi}, \bar{u})$ verifying $0 \in \partial \mathfrak{M}(\bar{x}, \bar{\xi}, \bar{u})$. To this end, fix any $(\bar{x}, \bar{\xi}, \bar{u})$ that satisfies $0 \in \partial \mathfrak{M}(\bar{x}, \bar{\xi}, \bar{u})$. Notice that Ψ is continuously differentiable because each D_i is convex. Since Ψ is also a KL function with exponent θ , we see that there exist positive numbers η and ϵ such that

$$(48) \quad \eta \|\nabla \Psi(x)\|^{\frac{1}{\theta}} \geq \Psi(x) - \Psi(\bar{x}) \quad \text{whenever } \|x - \bar{x}\| \leq \epsilon;$$

here, the condition on the bound on function values is waived by the continuity of Ψ and by choosing a smaller ϵ if necessary. In view of Lemma 5.6, we can shrink this ϵ further so that (36), (37), (38), (39), (40) and (41) hold whenever $(x, \xi, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\xi}, \bar{u}), \epsilon)$.

Now, fix any $(x, \boldsymbol{\xi}, u) \in \text{dom } \partial \mathfrak{M} \cap \mathbf{B}((\bar{x}, \bar{\boldsymbol{\xi}}, \bar{u}), \epsilon)$. By [16, Lemma 2.2] and a suitable scaling, there exists $c_3 > 0$ so that $\text{dist}^{\frac{1}{\theta}}(0, \partial \mathfrak{M}(x, \boldsymbol{\xi}, u))$ is bounded below by

$$(49) \quad c_3 \left[\text{dist}^{\frac{1}{\theta}}(0, \partial_x \mathfrak{M}(x, \boldsymbol{\xi}, u)) + 3 \text{dist}^{\frac{1}{\theta}}(0, \partial_{\boldsymbol{\xi}} \mathfrak{M}(x, \boldsymbol{\xi}, u)) + \text{dist}^{\frac{1}{\theta}}(0, \partial_u \mathfrak{M}(x, \boldsymbol{\xi}, u)) \right].$$

We now derive lower bounds for the three terms on the right hand side of (49). We first derive bounds for the second term. Using [16, Lemma 2.2] and (29), we see that there exists $c_0 > 0$ so that

$$(50) \quad \begin{aligned} 2 \text{dist}^{\frac{1}{\theta}}(0, \partial_{\boldsymbol{\xi}} \mathfrak{M}(x, \boldsymbol{\xi}, u)) &= 2 \left(\sum_{i=1}^m \inf_{\boldsymbol{\mu}_i \in N_{D_i}(\boldsymbol{\xi}_i)} \| -2u_i(\boldsymbol{\xi}_i - x) + \boldsymbol{\mu}_i \|^2 \right)^{\frac{1}{2\theta}} \\ &\geq 2c_0 \sum_{i=1}^m \inf_{\boldsymbol{\mu}_i \in N_{D_i}(\boldsymbol{\xi}_i)} \| -2u_i(\boldsymbol{\xi}_i - x) + \boldsymbol{\mu}_i \|^{\frac{1}{\theta}} \\ &\geq c_0 \sum_{i=1}^m c_1^{\frac{1}{2\theta}} (\|\boldsymbol{\xi}_i - x\|^2 - d_{D_i}^2(x))^{\frac{1}{2\theta}} + c_0 \sum_{i=1}^m c_2^{\frac{1}{2\theta}} \|P_{D_i}(x) - \boldsymbol{\xi}_i\|^{\frac{1}{\theta}} \\ &\geq \bar{\kappa}_0 \left(\sum_{i=1}^m (\|\boldsymbol{\xi}_i - x\|^2 - d_{D_i}^2(x))^{\frac{1}{\theta}} + \sum_{i=1}^m \|P_{D_i}(x) - \boldsymbol{\xi}_i\|^{\frac{1}{\theta}} \right), \end{aligned}$$

where $\bar{\kappa}_0 := \min\{c_0 c_1^{\frac{1}{2\theta}}, c_0 c_2^{\frac{1}{2\theta}}\}$, the second inequality follows from (40) and (41), and the last inequality holds because $0 \leq \|\boldsymbol{\xi}_i - x\|^2 - d_{D_i}^2(x) < 1$ (see (38)). On the other hand, using (40), we also have

$$(51) \quad \begin{aligned} \text{dist}^{\frac{1}{\theta}}(0, \partial_{\boldsymbol{\xi}} \mathfrak{M}(x, \boldsymbol{\xi}, u)) &= \left(\sum_{i=1}^m \inf_{\boldsymbol{\mu}_i \in N_{D_i}(\boldsymbol{\xi}_i)} \| -2u_i(\boldsymbol{\xi}_i - x) + \boldsymbol{\mu}_i \|^2 \right)^{\frac{1}{2\theta}} \\ &\geq \left[\sum_{i=1}^m c_1 (\|\boldsymbol{\xi}_i - x\|^2 - d_{D_i}^2(x)) \right]^{\frac{1}{2\theta}} \geq c_1^{\frac{1}{2\theta}} \sum_{i=1}^m (\|\boldsymbol{\xi}_i - x\|^2 - d_{D_i}^2(x)), \end{aligned}$$

where the last inequality follows from $0 \leq \sum_{i=1}^m (\|\boldsymbol{\xi}_i - x\|^2 - d_{D_i}^2(x)) < 1$ (see (38)) and $\theta \in [\frac{1}{2}, 1)$.

We next derive a lower bound for the third term on the right hand side in (49). In view of [16, Lemma 2.2], there exists $\bar{\kappa}_1 \in (0, \bar{\kappa}_0)$, with $\bar{\kappa}_0$ given in (50), so that

$$(52) \quad \begin{aligned} \text{dist}^{\frac{1}{\theta}}(0, \partial_u \mathfrak{M}(x, \boldsymbol{\xi}, u)) &\geq \bar{\kappa}_1 \sum_{i=1}^m \left| (\rho^*)'(u_i) - \|x - \boldsymbol{\xi}_i\|^2 \right|^{\frac{1}{\theta}} \\ &\geq \bar{\tau}_1 \sum_{i=1}^m \left| (\rho^*)'(u_i) - d_{D_i}^2(x) \right|^{\frac{1}{\theta}} - \bar{\kappa}_0 \eta_1 \sum_{i=1}^m (\|x - \boldsymbol{\xi}_i\|^2 - d_{D_i}^2(x))^{\frac{1}{\theta}}, \end{aligned}$$

where the last inequality follows from [16, Lemma 3.1] for some $\eta_1 \in (0, 1)$ and $\bar{\tau}_1 > 0$.

Finally, we derive a lower bound for the first term on the right hand side in (49). To this end, observe that

$$(53) \quad \hat{c} := \sup \left\{ (|a_1| + \dots + |a_m|)^{\frac{1}{\theta}} : |a_1|^{\frac{1}{\theta}} + \dots + |a_m|^{\frac{1}{\theta}} = 1 \right\} \in (0, \infty).$$

Choose $\bar{\kappa}_2 = \min\{2^{\frac{1}{\theta}}, \bar{c}^{-\frac{1}{\theta}} \widehat{c}^{-1} \bar{\kappa}_0\}$ (with \bar{c} and $\bar{\kappa}_0$ given in (39) and (50), respectively), we then have

$$\begin{aligned}
(54) \quad \text{dist}^{\frac{1}{\theta}}(0, \partial_x \mathfrak{M}(x, \boldsymbol{\xi}, u)) &= \left\| -2 \sum_{i=1}^m u_i(x - \boldsymbol{\xi}_i) \right\|^{\frac{1}{\theta}} \geq \bar{\kappa}_2 \left\| \sum_{i=1}^m u_i(x - \boldsymbol{\xi}_i) \right\|^{\frac{1}{\theta}} \\
&\geq \bar{\tau}_2 \left\| \sum_{i=1}^m u_i(x - P_{D_i}(x)) \right\|^{\frac{1}{\theta}} - \bar{\kappa}_2 \eta_2 \left[\sum_{i=1}^m |u_i| \|P_{D_i}(x) - \boldsymbol{\xi}_i\| \right]^{\frac{1}{\theta}} \\
&\geq \bar{\tau}_2 \left\| \sum_{i=1}^m u_i(x - P_{D_i}(x)) \right\|^{\frac{1}{\theta}} - \bar{\kappa}_0 \eta_2 \sum_{i=1}^m \|P_{D_i}(x) - \boldsymbol{\xi}_i\|^{\frac{1}{\theta}},
\end{aligned}$$

where the second inequality follows from [16, Lemma 3.1] for some $\eta_2 \in (0, 1)$ and $\bar{\tau}_2 > 0$ and the triangle inequality, and the last inequality follows from (39), the definition of \widehat{c} and the fact that $\bar{\kappa}_2 \leq \bar{c}^{-\frac{1}{\theta}} \widehat{c}^{-1} \bar{\kappa}_0$.

Now, we derive a lower bound for the first term on the right hand side of (54).

Let $\widehat{M} := \max_{1 \leq i \leq m} \max_{x \in \mathbf{B}(\bar{x}, \varepsilon)} \|x - P_{D_i}(x)\|$. Choose $\bar{\kappa}_3 = \min\{2^{-\frac{1}{\theta}} \bar{\tau}_2, (2\widehat{M}L)^{-\frac{1}{\theta}} \widehat{c}^{-1} \bar{\tau}_1\}$ (with L , $\bar{\tau}_1$ and \widehat{c} given in (37), (52) and (53), respectively), then we have

$$\begin{aligned}
(55) \quad \bar{\tau}_2 \left\| \sum_{i=1}^m u_i(x - P_{D_i}(x)) \right\|^{\frac{1}{\theta}} &= \frac{\bar{\tau}_2}{2^{\frac{1}{\theta}}} \left\| \nabla \Psi(x) - 2 \sum_{i=1}^m (u_i + \psi'_+(d_{D_i}^2(x)))(x - P_{D_i}(x)) \right\|^{\frac{1}{\theta}} \\
&\geq \bar{\kappa}_3 \left\| \nabla \Psi(x) - 2 \sum_{i=1}^m (u_i + \psi'_+(d_{D_i}^2(x)))(x - P_{D_i}(x)) \right\|^{\frac{1}{\theta}} \\
&\geq \bar{\tau}_3 \|\nabla \Psi(x)\|^{\frac{1}{\theta}} - 2^{\frac{1}{\theta}} \bar{\kappa}_3 \eta_3 \left(\sum_{i=1}^m |u_i + \psi'_+(d_{D_i}^2(x))| \cdot \|x - P_{D_i}(x)\| \right)^{\frac{1}{\theta}} \\
&\geq \bar{\tau}_3 \|\nabla \Psi(x)\|^{\frac{1}{\theta}} - \bar{\kappa}_3 \eta_3 (2\widehat{M})^{\frac{1}{\theta}} \widehat{c} \sum_{i=1}^m |u_i + \psi'_+(d_{D_i}^2(x))|^{\frac{1}{\theta}} \\
&\geq \bar{\tau}_3 \|\nabla \Psi(x)\|^{\frac{1}{\theta}} - \bar{\tau}_1 \eta_3 \sum_{i=1}^m \left| (\rho^*)'(u_i) - d_{D_i}^2(x) \right|^{\frac{1}{\theta}},
\end{aligned}$$

where the second inequality follows from [16, Lemma 3.1] for some $\eta_3 \in (0, 1)$ and $\bar{\tau}_3 > 0$ and the triangle inequality, the third inequality follows from the definitions of \widehat{c} and \widehat{M} , and the last inequality follows from (37) and the fact that $\bar{\kappa}_3 \leq (2\widehat{M}L)^{-\frac{1}{\theta}} \widehat{c}^{-1} \bar{\tau}_1$.

Combining (50), (51), (52), (54) and (55) with the lower bound of the quantity $\text{dist}^{\frac{1}{\theta}}(0, \partial \mathfrak{M}(x, \boldsymbol{\xi}, u))$ in (49), we see that

$$\begin{aligned}
&c_3^{-1} \text{dist}^{\frac{1}{\theta}}(0, \partial \mathfrak{M}(x, \boldsymbol{\xi}, u)) \\
&\geq \bar{\kappa}_0 \left(\sum_{i=1}^m (\|\boldsymbol{\xi}_i - x\|^2 - d_{D_i}^2(x))^{\frac{1}{\theta}} + \sum_{i=1}^m \|P_{D_i}(x) - \boldsymbol{\xi}_i\|^{\frac{1}{\theta}} \right) + c_1^{\frac{1}{2\theta}} \sum_{i=1}^m (\|\boldsymbol{\xi}_i - x\|^2 - d_{D_i}^2(x)) \\
&\quad + \bar{\tau}_1 \sum_{i=1}^m \left| (\rho^*)'(u_i) - d_{D_i}^2(x) \right|^{\frac{1}{\theta}} - \bar{\kappa}_0 \eta_1 \sum_{i=1}^m (\|x - \boldsymbol{\xi}_i\|^2 - d_{D_i}^2(x))^{\frac{1}{\theta}} \\
&\quad + \bar{\tau}_3 \|\nabla \Psi(x)\|^{\frac{1}{\theta}} - \bar{\tau}_1 \eta_3 \sum_{i=1}^m \left| (\rho^*)'(u_i) - d_{D_i}^2(x) \right|^{\frac{1}{\theta}} - \bar{\kappa}_0 \eta_2 \sum_{i=1}^m \|P_{D_i}(x) - \boldsymbol{\xi}_i\|^{\frac{1}{\theta}}
\end{aligned}$$

Grouping like terms and noting that $\eta_1, \eta_2, \eta_3 \in (0, 1)$ and $\|\xi_i - x\|^2 - d_{D_i}^2(x) \geq 0$ (thanks to (40)), we see further that

$$\begin{aligned}
& c_3^{-1} \text{dist}^{\frac{1}{\theta}}(0, \partial \mathfrak{M}(x, \xi, u)) \\
& \geq c_4 \left[\eta \|\nabla \Psi(x)\|^{\frac{1}{\theta}} + \frac{L}{2} \sum_{i=1}^m \left| (\rho^*)'(u_i) - d_{D_i}^2(x) \right|^{\frac{1}{\theta}} + \bar{c} \sum_{i=1}^m (\|\xi_i - x\|^2 - d_{D_i}^2(x)) \right] \\
& \geq c_4 \left[\eta \|\nabla \Psi(x)\|^{\frac{1}{\theta}} + \frac{L}{2} \sum_{i=1}^m \left| (\rho^*)'(u_i) - d_{D_i}^2(x) \right|^{\frac{1}{\theta}} + \sum_{i=1}^m (-u_i) (\|\xi_i - x\|^2 - d_{D_i}^2(x)) \right] \\
& \geq c_4 \left[\Psi(x) - \Psi(\bar{x}) + \frac{L}{2} \sum_{i=1}^m \left| (\rho^*)'(u_i) - d_{D_i}^2(x) \right|^{\frac{1}{\theta}} + \sum_{i=1}^m (-u_i) (\|\xi_i - x\|^2 - d_{D_i}^2(x)) \right] \\
& \geq c_4 (\mathfrak{M}(x, \xi, u) - \mathfrak{M}(\bar{x}, \bar{\xi}, \bar{u})),
\end{aligned}$$

where: the first inequality holds for some $c_4 > 0$ upon a suitable scaling (recall that η is given in (48), \bar{c} is defined in (39) and L is given in (36)); the second inequality follows from (39), the third inequality follows from (48), and the last inequality follows from (36) and $\Psi(\bar{x}) = \mathfrak{M}(\bar{x}, \bar{\xi}, \bar{u})$, thanks to Lemma 5.5 and the assumption that $0 \in \partial \mathfrak{M}(\bar{x}, \bar{\xi}, \bar{u})$. This completes the proof. \square

6. Numerical test. In this section, we perform numerical experiments to study the performance of EAS_{MFS_C} on some large-scale MFS_C problems. All codes are written in Matlab, and the experiments are performed in Matlab 2017b on a 64-bit PC with an Intel(R) Core(TM) i7-6700 CPU (3.40GHz) and 32GB of RAM.

In EAS_{MFS_C}, we take $\varphi_{\epsilon_k}(s) = 1 - \frac{\log(s + \epsilon_k)}{\log \epsilon_k}$ with $\epsilon_k = 0.9(0.1)^{k-1}$, and terminate when $\epsilon_k \leq 10^{-6}$. As for initialization, in our experiments below, we use randomly generated initial points \tilde{x}_0 , which are projections onto C of random vectors with i.i.d. standard Gaussian entries.

As discussed in Section 4, the corresponding subproblems of EAS_{MFS_C} take the form of (8) with $\psi(s) = \psi_{\epsilon_k}(s) := \log(\epsilon_k + s) - \log(\epsilon_k)$, and we use sGP_{ls} for solving them approximately. In the sGP_{ls}, we pick $\alpha_{\min} = 10^{-10}$, $\alpha_{\max} = 10^{10}$, $\eta = 1/2$, $M = 9$ and $\sigma = 10^{-4}$. We initialize the algorithm at \tilde{x}^{k-1} for approximately minimizing $\Phi_{\epsilon_k} + \delta_C$, and terminate the algorithm when $\|x^t - x^{t-1}\| \leq \max\left\{\frac{10^{-5}}{3^{k-1}}, 10^{-7}\right\} \cdot \max\{1, \|x^t\|\}$. As for α_t^0 , we initialize it at $\alpha_1^0 = 1$ and set, for $t \geq 1$,

$$\alpha_t^0 = \begin{cases} P_{[10^{-10}, 10^{10}]} \left(\frac{\|x^t - x^{t-1}\|^2}{\langle x^t - x^{t-1}, g^t - g^{t-1} \rangle} \right) & \text{if } \langle x^t - x^{t-1}, g^t - g^{t-1} \rangle > 10^{-12}, \\ P_{[10^{-10}, 10^{10}]}(2\alpha_{t-1}) & \text{otherwise;} \end{cases}$$

the choice of α_t^0 is motivated by the renowned Barzilai-Borwein stepsize.

We apply EAS_{MFS_C} as described above to two classes of MFS_C problems. In the first class, we set

$$C := \{x \in [-r, r]^n : \|x\|_0 \leq s\}$$

and $D_i = \{x \in \mathbb{R}^n : \langle \mathbf{a}_i, x \rangle \leq b_i\}$, where \mathbf{a}_i is the i th row of an $A \in \mathbb{R}^{m \times n}$. For the second class, we consider the same C as above but we choose $D_i = \{x \in \mathbb{R}^n : \langle \mathbf{a}_i, x \rangle \leq b_i\} \cup \{x \in \mathbb{R}^n : \langle \mathbf{p}_i, x \rangle \leq q_i\}$, where \mathbf{a}_i and \mathbf{p}_i are the i th row of $A \in \mathbb{R}^{m \times n}$ and $P \in \mathbb{R}^{m \times n}$, respectively. The matrices A , P and vectors b and q are randomly generated as follows. We first randomly generate A and P to have i.i.d. standard Gaussian entries. We next generate a $\tilde{w} \in \mathbb{R}^n$ with s i.i.d. standard Gaussian entries

at random positions, and project \tilde{w} onto $[-r, r]^n$ to form w . Set $\tilde{b} = Aw$ and fix a real number $\bar{p} \in [0, 1]$. We then define $b \in \mathbb{R}^n$ by

$$b_i = \begin{cases} \tilde{b}_i + 0.01\varepsilon_i & \text{if } i \leq \lceil \bar{p}m \rceil, \\ \tilde{b}_i - 50\varepsilon_i & \text{otherwise,} \end{cases}$$

where ε_i are chosen uniformly at random from $[0, 1]$. Finally, we set $q = Pw - 50\iota$ for some random vector $\iota \in [0, 1]^m$. By construction, the system $\{C, D_1, \dots, D_{\lceil \bar{p}m \rceil}\}$ is feasible and hence the optimal value of (1) is at least $\lceil \bar{p}m \rceil$ for these problems. Moreover, the vector $w \in C$ is not in $C \cap (\bigcap_{i=1}^m D_i)$. Furthermore, the resulting system $\{C, D_1, \dots, D_m\}$ is conceivably infeasible because of the subtractions of 50ε and 50ι .

In our experiments below, for the two classes of problems, we consider $m = 3000, 5000$, $n = \frac{m}{5}$, $s = \frac{n}{5}$, $\bar{p} = 0.5, 0.6$ and 0.7 , and $r = 10^8$. For each class of problems, for each m and \bar{p} , we randomly generate 10 instances as described above. For each instance, we solve the corresponding MFS_C problem using $\text{EAS}_{\text{MFS}_C}$ from 5 random initial points.⁶ We report the number of iterations (iter) and the CPU time in seconds (CPU) in Table 1, averaged over the 5 random initializations and the 10 instances. We also use the following quantities to evaluate the performance of our algorithm:

- $\text{feas}(x)$: For a given $x \in \mathbb{R}^n$, this corresponds to $\frac{1}{m} \#\{i : x \in D_i\}$.
- $\epsilon\text{-feas}(x)$: For a given $x \in \mathbb{R}^n$, this corresponds to $\frac{1}{m} \#\{i : \langle \mathbf{a}_i, x \rangle < b_i + \frac{10^{-5}m}{4}\}$ when each D_i is a halfspace, and corresponds to $\frac{1}{m} \#\{i : \min\{\langle \mathbf{a}_i, x \rangle - b_i, \langle \mathbf{p}_i, x \rangle - q_i\} < \frac{10^{-5}m}{4}\}$ when each D_i is a union of two halfspaces.

For each of the 10 random instances, we take the maximum of $\text{feas}(x_0)$ and $\epsilon\text{-feas}(x_0)$ and the maximum of $\text{feas}(x_*)$ and $\epsilon\text{-feas}(x_*)$ over 5 random initial points x_0 , where x_* is the approximate solution returned by our algorithm. We report in Table 1 the average of these quantities over the 10 random instances under the columns feas_0 , $\epsilon\text{-feas}_0$, feas_* and $\epsilon\text{-feas}_*$. One can see that our approach is able to identify a reasonably large (approximately) feasible subsystem with respect to C (i.e., $\epsilon\text{-feas}_* \gtrsim \bar{p}$) in a reasonable period of time, even for large-scale problems. Moreover, we always obtain a larger feasible subsystem compared with that identified by the random initial points. Finally, we also observe that our algorithm is faster when each D_i is convex.

7. Concluding remarks. In this paper, we proposed the $\text{EAS}_{\text{MFS}_C}$ for solving (1), which is based on solving a sequence of subproblems of the form (3). We discussed some viable choices of the φ_ϵ in (3), and presented a subgradient projection algorithm with nonmonotone linesearch (sGP_{l_s}) for solving the corresponding subproblems of $\text{EAS}_{\text{MFS}_C}$. Our experiments on large-scale random instances show that our approach is able to identify a reasonably large feasible subsystem with respect to C efficiently.

Note that it has been shown in Theorem 4.3 that any accumulation point of the sequence generated by sGP_{l_s} (with $M \geq 0$) is a stationary point of $\Psi + \delta_C$, where Ψ is defined in (8); moreover, under suitable KL assumptions, convergence of the whole sequence generated by sGP_{l_s} (with $M = 0$) has been established. However, the global convergence property of the whole sequence generated by sGP_{l_s} (with $M > 0$) is still unknown. This is an interesting future research direction. Another interesting research direction is to study when (1) can be provably solved by $\text{EAS}_{\text{MFS}_C}$, based on conditions of the sets $\{C, D_1, \dots, D_m\}$ (for example, when D_i 's are all halfspaces, or when D_i 's are closed balls, etc).

⁶These are projections onto C of random vectors with i.i.d. standard Gaussian entries. A projection onto C can be computed efficiently according to [17, Proposition 3.1].

TABLE 1
Computational results for randomly generated MFS_C problems

	m	\bar{p}	iter	CPU	feas ₀	ϵ -feas ₀	feas _*	ϵ -feas _*
convex D_i	3000	0.5	8034	312.0	0.35	0.35	0.57	0.61
		0.6	2884	119.9	0.38	0.38	0.51	0.59
		0.7	200	6.8	0.41	0.41	0.58	0.70
	5000	0.5	9728	1062.7	0.37	0.37	0.57	0.60
		0.6	2829	329.8	0.40	0.40	0.49	0.60
		0.7	209	17.9	0.42	0.42	0.57	0.70
Nonconvex D_i	3000	0.5	10778	833.6	0.47	0.47	0.85	0.89
		0.6	10753	841.2	0.49	0.49	0.84	0.88
		0.7	9176	713.1	0.52	0.52	0.83	0.87
	5000	0.5	13293	2596.9	0.51	0.51	0.85	0.88
		0.6	13459	2730.9	0.53	0.53	0.84	0.88
		0.7	12919	2732.8	0.55	0.55	0.83	0.87

REFERENCES

- [1] E. Amaldi and V. Kann, The complexity and approximability of finding maximum feasible subsystems of linear relations, *Theor. Comput. Sci.*, 147:181–210, 1995.
- [2] E. Amaldi and V. Kann, On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems, *Theor. Comput. Sci.*, 209:237–260, 1998.
- [3] E. Amaldi, M. E. Pfetsch and L. E. Trotter. Jr, Some structural and algorithmic properties of the maximum feasible subsystem problem, *Integer Programming and Combinatorial Optimization Book Series: Lecture Notes in Computer Science*. 1610:45–59, 1999.
- [4] E. Amaldi, M. E. Pfetsch and L. E. Trotter. Jr, On the maximum feasible subsystem problem, IISs and IIS-hypergraphs, *Math. Program.*, 95:533–554, 2003.
- [5] H. Attouch and J. Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, *Math. Program.*, 116:5–16, 2009.
- [6] 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.*, 35:438–457, 2010.
- [7] 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.*, 137:91–129, 2013.
- [8] H. H. Bauschke, *Projection Algorithms and Monotone Operators*, PhD thesis, Simon Fraser University, 1996.
- [9] D. P. Bertsekas, *Nonlinear Programming*, 2nd edition, Athena Scientific, 1999.
- [10] J. Bolte, S. Sabach and M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program.*, 146:459–494, 2014.
- [11] N. Chakravarti, Some results concerning post-infeasibility analysis, *Eur. J. Oper. Res.*, 73:139–143, 1994.
- [12] J. W. Chinneck, An effective polynomial-time heuristic for the minimum-cardinality IIS setcovering problem, *Ann. Math. Artif. Intell.*, 17:127–144, 1996.
- [13] J. W. Chinneck, Fast heuristics for the maximum feasible subsystem problem, *INFORMS J. Comput.*, 13:210–223, 2001.
- [14] H. J. Greenberg and F. H. Murphy, Approaches to diagnosing infeasible linear programs, *ORSA J. Comput.*, 3:253–261, 1991.
- [15] A. S. Lewis, D. R. Luke and J. Malick, Local linear convergence for alternating and averaged nonconvex projections, *Found. Comput. Math.*, 9:485–513, 2009.
- [16] 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.*, 18:1199–1232, 2018.
- [17] Z. S. Lu and Y. Zhang, Sparse approximation via penalty decomposition methods, *SIAM J. Optim.*, 23:2448–2478, 2012.
- [18] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory*, Springer, 2016.

- [19] M. Parker, *A Set Covering Approach to Infeasibility Analysis of Linear Programming Problems and Related Issues*, Ph.D. thesis, University of Colorado at Denver, 1995.
- [20] M. Parker and J. Ryan, Finding the minimum weight IIS cover of an infeasible system of linear inequalities, *Ann. Math. Artif. Intell.*, 17:107–126, 1996.
- [21] M. E. Pfetsch, *The Maximum Feasible Subsystem Problem and Vertex-Facet Incidence of Polyhedra*, Ph.D. thesis, TU Berlin, Berlin, 2002.
- [22] M. E. Pfetsch, Branch-and-cut for the maximum feasible subsystem problem, *SIAM J. Optim.*, 19:21–38, 2008.
- [23] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
- [24] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, 1998.
- [25] S. J. Wright, R. D. Nowak and M. A. T. Figueiredo, Sparse reconstruction by separable approximation, *IEEE Trans. Signal Process.*, 57:2479–2493, 2009.
- [26] Y. B. Zhao and D. Li, Reweighted ℓ_1 -minimization for sparse solutions to underdetermined linear systems, *SIAM J. Optim.*, 22:1065–1088, 2012.