CONVERGENCE RATE ANALYSIS OF A SEQUENTIAL CONVEX PROGRAMMING METHOD WITH LINE SEARCH FOR A CLASS OF CONSTRAINED DIFFERENCE-OF-CONVEX OPTIMIZATION PROBLEMS [*]

PEIRAN YU[†], TING KEI PONG[‡], AND ZHAOSONG LU[§]

Abstract. In this paper, we study the sequential convex programming method with monotone line search (SCP_{ls}) in [46] for a class of difference-of-convex (DC) optimization problems with multiple smooth inequality constraints. The SCP_{ls} is a representative variant of moving-ball-approximation-type algorithms [6, 10, 13, 54] for constrained optimization problems. We analyze the convergence rate of the sequence generated by SCP_{ls} in both nonconvex and convex settings by imposing suitable Kurdyka-Łojasiewicz (KL) assumptions. Specifically, in the nonconvex settings, we assume that a special potential function related to the objective and the constraints is a KL function, while in the convex settings we impose KL assumptions directly on the extended objective function (i.e., sum of the objective and the indicator function of the constraint set). A relationship between these two different KL assumptions is established in the convex settings under additional differentiability assumptions. We also discuss how to deduce the KL exponent of the extended objective function from its Lagrangian in the convex settings, under additional assumptions on the constraint functions. Thanks to this result, the extended objectives of some constrained optimization models such as minimizing ℓ_1 subject to logistic/Poisson loss are found to be KL functions with exponent $\frac{1}{2}$ under mild assumptions. To illustrate how our results can be applied, we consider SCP_{ls} for minimizing ℓ_{1-2} [60] subject to residual error measured by ℓ_2 norm/Lorentzian norm [21]. We first discuss how the various conditions required in our analysis can be verified, and then perform numerical experiments to illustrate the convergence behaviors of SCP_{ls}.

1. Introduction. Constrained optimization problems naturally arise when one attempts to find a solution that minimizes a certain objective under some restrictions, see [6, 8, 18, 21, 29]. Here, we consider the following specific type of difference-of-convex (DC) constrained optimization problem:

$$(1.1) \qquad \min_{x \in \mathbb{R}^n} \; F(x) := f(x) + P_1(x) - P_2(x) + \delta_{g(\cdot) \leq 0}(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is smooth, $P_1 : \mathbb{R}^n \to \mathbb{R}$ and $P_2 : \mathbb{R}^n \to \mathbb{R}$ are convex continuous (possibly nonsmooth), and $g : \mathbb{R}^n \to \mathbb{R}^m$ is continuous with $\{x : g(x) \leq 0\} \neq \emptyset$. In typical applications, the f in (1.1) arises as measures for data fidelity, g is used for modeling restrictions on the decision variable x, and $P_1 - P_2$ is a regularizer for inducing desirable structures; see [28, Table 1] for examples of such regularizers. In our subsequent algorithmic development for (1.1), we also consider the following additional assumption.

ASSUMPTION 1.1. *Let f, g and F be as in* (1.1).[1]
(i) $f : \mathbb{R}^n \to \mathbb{R}$ *has Lipschitz continuous gradient with Lipschitz modulus L_f.*

[†]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong (peiran.yu@connect.polyu.hk).
[‡]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong (tk.pong@polyu.edu.hk).
[§]Department of Industrial and Systems Engineering, University of Minnesota, USA (zhaosong@umn.edu).

[1]Here and throughout the paper, by referring to (1.1), we mean that the assumptions on f, P_1, P_2 and g stated right after (1.1) are satisfied, i.e., $f : \mathbb{R}^n \to \mathbb{R}$ is smooth, $P_1 : \mathbb{R}^n \to \mathbb{R}$ and $P_2 : \mathbb{R}^n \to \mathbb{R}$ are convex continuous, and $g : \mathbb{R}^n \to \mathbb{R}^m$ is continuous with $\{x : g(x) \leq 0\} \neq \emptyset$.

(ii) *For the mapping $g(x) = (g_1(x), ..., g_m(x))$, each function g_i has Lipschitz continuous gradient with Lipschitz modulus L_{g_i}.*

(iii) *The function F is level-bounded.*

Under Assumption 1.1, the solution set of (1.1) is nonempty and $\inf F > -\infty$.

To design algorithms for solving (1.1) under Assumption 1.1, one common approach is to resort to the majorization-minimization (MM) procedure: in this procedure, one iteratively constructs and minimizes a surrogate function that locally majorizes F; see [13,23–25,39,56] for related models and discussions. For (1.1) under Assumption 1.1, one natural way to construct surrogate function is to make use of the 2nd-order Taylor's expansions of f and g: the resulting algorithms are the moving balls approximation method (MBA) proposed in [6] (for $P_1 = P_2 = 0$) and its variants [10, 13]. In each iteration, these algorithms approximate the constraint $g(x) \leq 0$ in (1.1) by

$$(1.2) \qquad \bar{G}(x, y, w) := \begin{pmatrix} g_1(y) + \langle \nabla g_1(y), x - y \rangle + \frac{w_1}{2}\|x - y\|^2 \\ \vdots \\ g_m(y) + \langle \nabla g_m(y), x - y \rangle + \frac{w_m}{2}\|x - y\|^2 \end{pmatrix} \leq 0$$

for some fixed (y, w): the feasible region of the resulting subproblem is an intersection of m balls. For the sequence generated by MBA, global convergence to a minimizer was established in [6] when $\{f, g_1, \ldots, g_m\}$ are in addition convex and the Slater condition holds. The linear convergence of the sequence generated by MBA was also proved in [6] when f in (1.1) is additionally strongly convex. In [13], when $\{f, g_1, \ldots, g_m\}$ are semi-algebraic and $P_1 = P_2 = 0$ in (1.1), the whole sequence generated by an MBA variant was shown to converge to a critical point and its convergence rate was also established, under the Mangasarian-Fromovitz constraint qualification (MFCQ).

When the DC function $P_1 - P_2$ in (1.1) is nonsmooth (these nonsmooth functions arise naturally as regularizers in applications such as sparse recovery [21, 28, 60]), the MBA method is not directly applicable. Moreover, when P_2 is nonsmooth, the multiprox method in [10] and the majorization-minimization procedure in [13, Section 3] cannot be directly applied to (1.1). Fortunately, under Assumption 1.1, problem (1.1) has DC objective and DC constraints: indeed, one can write f and each g_i in (1.1) as the difference of two convex functions as follows:

$$f(x) = \frac{L_f}{2}\|x\|^2 - \left(\frac{L_f}{2}\|x\|^2 - f(x)\right) \text{ and } g_i(x) = \frac{L_{g_i}}{2}\|x\|^2 - \left(\frac{L_{g_i}}{2}\|x\|^2 - g_i(x)\right).$$

DC algorithms (DCA) (see, for example, [36, 38]) can thus be applied. A variant that specializes in functional constraints is the sequential convex programming (SCP) method proposed in [46] [2]; see also [50, Remark 5]. When applied to (1.1) under Assumption 1.1, this method maintains feasibility at each iteration [3] and each subproblem is constrained over an intersection of balls: thus, this method can also be viewed as a variant of MBA. It was shown that any accumulation point of the sequence generated by SCP is a stationary point under Slater's condition. However, convergence and convergence rate of the whole sequence generated remain unknown. [4]

---

[2] We would like to point out that the methods proposed in [46] (including SCP and its variant) were designed to solve more general models than (1.1). In particular, they can deal with problems with constraints involving nonsmooth functions, and allow for nonmonotone line search.

[3] There are some DCA variants for solving (1.1) under Assumption 1.1 that do not maintain feasibility throughout. We refer the interested readers to [37, 38, 43, 55, 58] for more discussions.

[4] We point out that convergence of the whole sequence and the convergence rate generated by some DCA variants were considered in [5, 36] under suitable Kurdyka-Łojasiewicz (KL) assumptions; however, their problem formulations do not explicitly involve functional constraints as in (1.1).

For empirical acceleration, a variant of MBA that involves a line search scheme was proposed in [10], which is called the Multiproximal method with backtracking step sizes (Multiprox$_{bt}$). When applied to (1.1) under Assumption 1.1, the sequence generated by Multiprox$_{bt}$ converges to a minimizer when $\{f, g_1, \ldots, g_m\}$ are additionally convex, $P_1 = P_2 = 0$ and the Slater condition holds. However, Multiprox$_{bt}$ uses monotone initial step sizes, i.e., $\tilde{\alpha}$ in [10, Eq. (37)] is nondecreasing as the algorithm progresses, which rules out widely used choices such as the truncated Barzilai-Borwein step sizes [7, 9]. On the other hand, the line search variant of SCP proposed in [46] can incorporate flexible line search schemes like the truncated Barzilai-Borwein step size and is general enough to be applied to (1.1) under Assumption 1.1 with possibly nonsmooth $P_1 - P_2$. In [46], the well-definedness of the proposed algorithm was established, and it was also shown that any accumulation point of it is a stationary point under Slater's condition. However, convergence of the whole sequence generated and the corresponding convergence rate is still open.

In this paper, we further study the line search variant of the SCP method proposed in [46] with its line search being monotone, i.e., M in [46, Eq. (22)] being 0. We call this variant SCP$_{ls}$; see Algorithm 2.1 below. We analyze the convergence properties of the sequence generated by SCP$_{ls}$ for solving (1.1) under Assumption 1.1. The main convergence rate analysis of SCP$_{ls}$ is presented in Section 3. We derive global convergence rate of the sequence generated by SCP$_{ls}$ in the following two scenarios:

- F *in* (1.1) *is possibly nonconvex with each g_i being twice continuously differentiable and P_2 being Lipschitz continuously differentiable on an open set Γ that contains the set of stationary points of F.*
  Our analysis is based on the following specially constructed potential function:

  $$(1.3) \qquad \bar{F}(x, y, w) = f(x) + P_1(x) - P_2(x) + \delta_{\bar{G}(\cdot) \leq 0}(x, y, w),$$

  where $\bar{G}$ is defined as in (1.2). Under MFCQ, we characterize the local convergence rate of the sequence generated by SCP$_{ls}$ according to the Kurdyka-Łojasiewicz (KL) exponent of $\bar{F}$. Note the mapping $(x, y) \mapsto \bar{F}(x, y, L)$ with $P_2 = 0$ and L being a constant positive vector (related to the step size) was used previously in [13] for establishing the convergence of an MBA variant when $P_2 = 0$ and $\{f, g_1, \ldots, g_m\}$ in (1.1) are semi-algebraic. This kind of potential functions was called "value function" in [49] and was used there for deducing the global convergence properties of the composite Gauss-Newton method for composite optimization problems. Our potential function $\bar{F}$ allows us to deal with more flexible stepsize rules than those studied in [13, 49].
- $\{f, g_1 \ldots, g_m\}$ *in* (1.1) *are convex and $P_2 = 0$.*
  This same convex setting was considered in [10, Section 3.2.3]. In this setting, we impose KL assumptions directly on F in (1.1) (instead of on $\bar{F}$). In particular, a local *linear* convergence rate is established when F is a KL function with exponent $\frac{1}{2}$, under MFCQ. This is different from many existing analysis (see, for example, [3, 13, 41, 47]), which typically make use of the KL property of a potential function constructed out of F instead of F itself.

In Section 4.1, we study a relationship between the KL property of $\bar{F}$ in (1.3) and that of F in (1.1). Then, we present a "calculus rule" that deduces the KL exponent of F in (1.1) from its Lagrangian in the convex settings, under some mild assumptions. This enables us to deduce that the function F corresponding to minimizing ℓ_1 subject to logistic/Poisson loss is a KL function with exponent $\frac{1}{2}$ under mild conditions.

In Section 5, we discuss some concrete models to which SCP$_{ls}$ can be applied. Specifically, we consider models of the following form:

$$(1.4) \quad \begin{aligned} \min_{x} \quad & \|x\|_1 - \mu\|x\| \\ \text{s.t.} \quad & \ell(Ax - b) \leq \delta, \end{aligned}$$

where $\mu \in [0, 1]$, $A \in \mathbb{R}^{q \times n}$ has *full row rank*, $b \in \mathbb{R}^q$, $\ell : \mathbb{R}^q \to \mathbb{R}_+$ is analytic with Lipschitz continuous gradient and satisfies $\ell(0) = 0$, and $\delta \in (0, \ell(-b))$. This model arises in compressed sensing where the measurements b may be corrupted by different types of noise; see [20]. We focus on two concrete choices of ℓ: the square of norm (for noise following Gaussian distribution) and the Lorentzian norm (for noise following Cauchy distribution). For these two choices, we provide suitable conditions on the problem data so that the assumptions in our convergence results are satisfied. Then we perform numerical tests on solving (1.4) with ℓ being either the square of norm or the Lorentzian norm via two methods: SCP_{ls} and SCP [46]. We observe that SCP_{ls} appears to converge linearly and is much faster.

2. Notation and preliminaries. In this paper, we let $\mathbb{R}$ denote the set of real numbers and $\mathbb{N}_+$ denote the set of positive integers. The n-dimensional Euclidean space is denoted by $\mathbb{R}^n$, and the nonnegative orthant is denoted by $\mathbb{R}^n_+$. For two vectors x and $y \in \mathbb{R}^n$, we write $x \geq y$ if $x_i \geq y_i$ for all i. The Euclidean norm of x is denoted by $\|x\|$, the inner product of x and y is denoted by $\langle x, y \rangle$, and the ℓ_1 norm of x is denoted by $\|x\|_1$. For $x \in \mathbb{R}^n$ and $r \geq 0$, we let $B(x, r)$ denote the closed ball centered at x with radius r, i.e., $B(x, r) = \{y : \|x - y\| \leq r\}$.

We say that an extended-real-valued function $f : \mathbb{R}^n \to (-\infty, \infty]$ is proper if its domain $\text{dom} f := \{x : f(x) < \infty\} \neq \emptyset$. A proper function f is said to be closed if it is lower semicontinuous. For a proper function f, the regular subdifferential of f at $x \in \text{dom} f$ is defined by

$$\hat{\partial} f(x) := \left\{ \zeta : \liminf_{z \to x, z \neq x} \frac{f(z) - f(x) - \langle \zeta, z - x \rangle}{\|z - x\|} \geq 0 \right\}.$$

The (limiting) subdifferential of f at $x \in \text{dom} f$ is defined by

$$\partial f(x) := \left\{ \zeta : \exists x^k \xrightarrow{f} x, \zeta^k \to \zeta \text{ with } \zeta^k \in \hat{\partial} f(x^k) \text{ for each } k \right\},$$

where $x^k \xrightarrow{f} x$ means both $x^k \to x$ and $f(x^k) \to f(x)$. Moreover, we set $\partial f(x) = \hat{\partial} f(x) = \emptyset$ for $x \notin \text{dom} f$ by convention, and we write $\text{dom} \partial f := \{x : \partial f(x) \neq \emptyset\}$. When f is proper convex, thanks to [46, Proposition 8.12], the limiting subdifferential and regular subdifferential of f at an $x \in \text{dom} f$ reduce to the classical subdifferential, which is given by

$$\partial f(x) = \{\zeta : \langle \zeta, y - x \rangle \leq f(y) - f(x) \text{ for all } y\}.$$

For a nonempty set C, the indicator function δ_C is defined as

$$\delta_C(x) := \begin{cases} 0 & x \in C, \\ \infty & x \notin C. \end{cases}$$

The normal cone (resp., regular normal cone) of C at an $x \in C$ is defined as $N_C(x) := \partial \delta_C(x)$ (resp., $\hat{N}_C(x) := \hat{\partial} \delta_C(x)$), and the distance from a point $x \in \mathbb{R}^n$ to C is denoted by $\text{dist}(x, C)$.

We next recall the KL property and the notion of KL exponent; see [2–4, 34, 42, 45]. This property has been used extensively for analyzing convergence properties of first-order methods; see, for example, [2–4, 14, 59].

DEFINITION 2.1 (**Kurdyka-Łojasiewicz property and exponent**). *We say that a proper closed function $h : \mathbb{R}^n \to (-\infty, \infty]$ satisfies the Kurdyka-Łojasiewicz (KL) property at an $\hat{x} \in \text{dom}\partial h$ if there are $a \in (0, \infty]$, a neighborhood V of $\hat{x}$ and a continuous concave function $\varphi : [0, a) \to [0, \infty)$ with $\varphi(0) = 0$ such that*
  (i) *φ is continuously differentiable on $(0, a)$ with $\varphi' > 0$ on $(0, a)$;*
  (ii) *for any $x \in V$ with $h(\hat{x}) < h(x) < h(\hat{x}) + a$, it holds that*

$$(2.1) \qquad\qquad \varphi'(h(x) - h(\hat{x}))\text{dist}(0, \partial h(x)) \geq 1.$$

If h satisfies the KL property at $\hat{x} \in \text{dom}\partial h$ and φ in (2.1) can be chosen as $\varphi(\nu) = a_0 \nu^{1-\alpha}$ for some $a_0 > 0$ and $\alpha \in [0, 1)$, then we say that h satisfies the KL property at $\hat{x}$ with exponent α.

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

There are many examples of KL functions. For instance, proper closed semi-algebraic functions and proper subanalytic functions that have closed domains and are continuous on their domains are KL functions; see [3] and [11, Theorem 3.1], respectively.

Now we recall the definition of stationary points of (1.1) when g_i are smooth.

DEFINITION 2.2 (**Stationary point**). *Consider (1.1) and assume that each g_i is smooth. We say that an $x \in \mathbb{R}^n$ is a stationary point of (1.1) if there exists $\lambda \in \mathbb{R}^m_+$ such that (x, λ) satisfies*

$$g(x) \leq 0, \ \lambda_i g_i(x) = 0 \text{ for all } i, \text{ and } 0 \in \nabla f(x) + \partial P_1(x) - \partial P_2(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x).$$

The following assumption will be used repeatedly throughout this paper.

ASSUMPTION 2.3. *Each g_i in (1.1) is smooth and the Mangasarian-Fromovitz constraint qualification (MFCQ) holds in the whole domain of F in (1.1), i.e., for every x satisfying $g(x) \leq 0$, there exists $d \in \mathbb{R}^n$ such that*

$$\langle \nabla g_i(x), d \rangle < 0 \text{ for each } i \in I(x) := \{j : \ g_j(x) = 0\}.$$

Under Assumptions 1.1 and 2.3, it is routine to show that any local minimizer of (1.1) is a stationary point in the sense of Definition 2.2. In fact, let $\hat{x}$ be a local minimizer of (1.1). Using [53, Theorem 10.1], we have

$$
\begin{aligned}
0 \in \partial F(\hat{x}) &\overset{(a)}{\subseteq} \nabla f(\hat{x}) + \partial P_1(\hat{x}) + \partial(-P_2)(\hat{x}) + \partial \delta_{g(\cdot) \leq 0}(\hat{x}) \\
(2.2) \qquad &\overset{(b)}{\subseteq} \nabla f(\hat{x}) + \partial P_1(\hat{x}) + \bar{\partial}(-P_2)(\hat{x}) + \partial \delta_{g(\cdot) \leq 0}(\hat{x}) \\
&\overset{(c)}{=} \nabla f(\hat{x}) + \partial P_1(\hat{x}) - \bar{\partial} P_2(\hat{x}) + \partial \delta_{g(\cdot) \leq 0}(\hat{x}) \\
&= \nabla f(\hat{x}) + \partial P_1(\hat{x}) - \partial P_2(\hat{x}) + \partial \delta_{g(\cdot) \leq 0}(\hat{x}),
\end{aligned}
$$

where (a) follows from [53, Exercise 10.10], the inclusion (b) uses [17, Theorem 5.2.22], where $\bar{\partial}(-P_2)$ is the Clarke subdifferential of $-P_2$, the equality (c) uses [22, Proposition 2.3.1] and the last equality holds because of the convexity of P_2 and [15, Theorem 6.2.2]. In addition, we can deduce that

$$\partial \delta_{g(\cdot) \leq 0}(\hat{x}) = N_{g(\cdot) \leq 0}(\hat{x}) = \left\{ \sum_{i=1}^{m} \lambda_i \nabla g_i(\hat{x}) : \lambda \in N_{-\mathbb{R}^m_+}(g(\hat{x})) \right\}$$

$$= \left\{ \sum_{i=1}^{m} \lambda_i \nabla g_i(\hat{x}) : \lambda \in \mathbb{R}_+^m, \ \lambda_i g_i(\hat{x}) = 0 \text{ for } i = 1, \ldots, m \right\},$$

where the second equality follows from MFCQ and [53, Theorem 6.14] and the last equality follows from the definition of normal cone. The above display together with (2.2) shows that $\hat{x}$ is a stationary point of (1.1). In passing, we would like to point out that x^* is a stationary point of (1.1) in the sense of Definition 2.2 if and only if there exists ξ^* such that $0 \in \partial \widetilde{F}(x^*, \xi^*)$, where $\widetilde{F}(x, \xi) := f(x) + P_1(x) - \langle \xi, x \rangle + P_2^*(\xi) + \delta_{g(\cdot) \leq 0}(x)$, with $\{P_1, P_2\}$ given in (1.1) and P_2^* being the Fenchel conjugate of P_2. This type of stationary points is widely used in the DC literature; see, for example, [57–59]. Note that there are other concepts of stationarity used in the literature, such as the Clarke stationarity, d-stationarity and B-stationarity; we refer the readers to [1, 32, 48] for more discussions. The notion of stationarity defined in Definition 2.2 is in general weaker than these aforementioned notions.

Before ending this section, we introduce the algorithm we analyze and present some auxiliary results for our subsequent analysis. The algorithm, SCP_{ls} proposed in [46], is presented in Algorithm 2.1, where $\bar{G}$ is defined as in (1.2). Notice that by rearranging terms of the constraint functions of the subproblem (2.4), we can see that the constraint there is equivalent to

$$(2.3) \qquad x \in \bigcap_{i=1}^{m} B\left(\widetilde{s}_i, \sqrt{\widetilde{R}_i} \right),$$

where $\widetilde{s}_i := x^t - \frac{1}{(\widetilde{L}_g)_i} \nabla g_i(x^t)$ and $\widetilde{R}_i := \left\| \frac{\nabla g_i(x^t)}{(\widetilde{L}_g)_i} \right\|^2 - \frac{2}{(\widetilde{L}_g)_i} g_i(x^t)$. Thus, when $m = 1$, the constraint reduces to a *single* ball constraint and a simple root-finding scheme was discussed in [54] for exactly and efficiently solving the subproblem (2.4) with $m = 1$, $P_2 = 0$ and P_1 being the ℓ_1 norm or the nuclear norm, etc. However, solving subproblem (2.4) in general requires an iterative solver; see [6, Section 6] for the case when $P_1 = P_2 = 0$.

In the next lemma, we discuss the well-definedness of SCP_{ls} and also establish some inequalities needed in our analysis below. Note that the well-definedness of SCP_{ls} was already proved in [46, Theorem 3.6] in a more general setting. Here we include its proof for completeness.

LEMMA 2.4. *Consider* (1.1) *and suppose that Assumptions* 1.1 *and* 2.3 *hold. Then the following statements hold:*
  (i) *SCP_{ls} is well defined, i.e., the subproblems* (2.4) *are well defined and there exists a $k_0 \in \mathbb{N}_+$ (independent of t) such that in any iteration $t \geq 0$, the inner loop stops after at most k_0 iterations.*
 (ii) *The sequence $\{(L_f^t, L_g^t)\}$ generated by SCP_{ls} is bounded.*
(iii) *For each $i \in \{1, \ldots, m\}$, each $t \geq 0$ and each $(\widetilde{L}_f, \widetilde{L}_g)$, the $\widetilde{R}_i$ in* (2.3) *is positive.*
(iv) *For each $t \geq 0$ and each $(\widetilde{L}_f, \widetilde{L}_g)$, the problem* (2.4) *has a Lagrange multiplier $\widetilde{\lambda}$. Let $\widetilde{L}_{fg} := \widetilde{L}_f + \langle \widetilde{\lambda}, \widetilde{L}_g \rangle$ and let $\widetilde{x}$ be as in* (2.4). *Then*

$$(2.6) \qquad \widetilde{\lambda}_i \left(g_i(x^t) + \langle \nabla g_i(x^t), \widetilde{x} - x^t \rangle + \frac{(\widetilde{L}_g)_i}{2} \|\widetilde{x} - x^t\|^2 \right) = 0 \text{ for all } i,$$

Algorithm 2.1 Sequential convex programming method with monotone line search (SCP$_{ls}$) for (1.1) under Assumption 1.1

Choose parameters $c > 0$, $0 < \underline{\mathrm{L}} < \bar{\mathrm{L}}$, $\tau > 1$ and an x^0 with $g(x^0) \leq 0$. Set $t = 0$.

Step 1. Pick any $\xi^t \in \partial P_2(x^t)$.

Step 2. Choose $L_f^{t,0} \in [\underline{\mathrm{L}}, \bar{\mathrm{L}}]$ and $L_g^{t,0} \in [\underline{\mathrm{L}}, \bar{\mathrm{L}}]^m$ arbitrarily. Set $\widetilde{L}_f = L_f^{t,0}$ and $\widetilde{L}_g = L_g^{t,0}$.

Step 3. Compute

$$(2.4) \qquad \widetilde{x} = \arg\min_x \left\{ \langle \nabla f(x^t) - \xi^t, x - x^t \rangle + \frac{\widetilde{L}_f}{2} \|x - x^t\|^2 + P_1(x) \right\}$$
$$\text{s.t. } \bar{G}(x, x^t, \widetilde{L}_g) \leq 0.$$

Step 3a) If $g(\widetilde{x}) \leq 0$ and

$$(2.5) \qquad\qquad F(\widetilde{x}) \leq F(x^t) - \frac{c}{2}\|\widetilde{x} - x^t\|^2$$

holds, go to step 4.

Step 3b) If $g(\widetilde{x}) \not\leq 0$, let $\widetilde{L}_g \leftarrow \tau \widetilde{L}_g$ and go to step 3.

Step 3c) If (2.5) does not hold, let $\widetilde{L}_f \leftarrow \tau \widetilde{L}_f$ and go to step 3.

Step 4. If a termination criterion is not met, set $L_g^t = \widetilde{L}_g$, $L_f^t = \widetilde{L}_f$ and $x^{t+1} = \widetilde{x}$. Update $t \leftarrow t + 1$ and go to **Step 1.**

---

and

$$(2.7) \qquad 0 \in \nabla f(x^t) - \xi^t + \widetilde{L}_{fg}(\widetilde{x} - x^t) + \partial P_1(\widetilde{x}) + \sum_{i=1}^m \widetilde{\lambda}_i \nabla g_i(x^t),$$

where $\{x^t\}$ and $\{\xi^t\}$ are generated by SCP$_{ls}$. Moreover, if $g(\widetilde{x}) \leq 0$, then for any $x \in \mathbb{R}^n$ we have

$$(2.8) \quad \begin{aligned} F(\widetilde{x}) &\leq f(x^t) + \langle \nabla f(x^t) - \xi^t, x - x^t \rangle + \frac{\widetilde{L}_{fg}}{2}\|x - x^t\|^2 + P_1(x) - P_2(x^t) \\ &+ \sum_{i=1}^m \widetilde{\lambda}_i \left(g_i(x^t) + \langle \nabla g_i(x^t), x - x^t \rangle \right) - \frac{\widetilde{L}_{fg}}{2}\|x - \widetilde{x}\|^2 - \frac{\widetilde{L}_f - L_f}{2}\|\widetilde{x} - x^t\|^2. \end{aligned}$$

Proof. Let an x^t satisfying $g(x^t) \leq 0$ be given for some $t \geq 0$. We will first show that the corresponding subproblems (2.4) are well defined (for any $(\widetilde{L}_f, \widetilde{L}_g)$) and the conclusions of items (iii) and (iv) hold for this t. Using these, we will then show that there exists k_0 (independent of t) so that the inner loop in Step 3 terminates after k_0 iterations and returns an x^{t+1} that satisfies $g(x^{t+1}) \leq 0$. This together with $g(x^0) \leq 0$ and an induction argument will show that SCP$_{ls}$ is well defined and that items (iii) and (iv) hold for all $t \geq 0$. Finally, we show that $\{(L_f^t, L_g^t)\}$ is bounded.

Suppose that an x^t satisfying $g(x^t) \leq 0$ is given for some $t \geq 0$. Notice that for any $(\widetilde{L}_f, \widetilde{L}_g)$, the feasible region of (2.4) is nonempty (it contains x^t) and the subproblem is to minimize a strongly convex continuous function over a nonempty closed convex set. Thus, $\widetilde{x}$ exists and is unique. Now, fix any $i \in \{1, \ldots, m\}$. Since $g(x^t) \leq 0$ and $(\widetilde{L}_g)_i > 0$, we have $-\frac{2}{(\widetilde{L}_g)_i} g_i(x^t) \geq 0$ and thus $\widetilde{R}_i \geq 0$. Suppose to the contrary that $\widetilde{R}_i = 0$. Then we have $\nabla g_i(x^t) = 0$ and $g_i(x^t) = 0$, contradicting Assumption 2.3. Thus, we must have $\widetilde{R}_i > 0$ at the tth iteration.

Next, using a similar proof of [6, Proposition 2.1(iii)], we deduce using MFCQ that the Slater condition holds for (2.4) for this t. Therefore, using [52, Corollary 28.2.1, Theorem 28.3], for problem (2.4), there exists a Lagrange multiplier $\widetilde{\lambda} \in \mathbb{R}^m_+$ such that (2.6) holds at the tth iteration and $\widetilde{x}$ is a minimizer of the following function:

$$L_t(x, \widetilde{\lambda}) := f(x^t) + \langle \nabla f(x^t), x - x^t \rangle + \frac{\widetilde{L}_f}{2} \|x - x^t\|^2 + P_1(x) - P_2(x^t)$$
$$- \langle \xi^t, x - x^t \rangle + \langle \widetilde{\lambda}, \bar{G}(x, x^t, \widetilde{L}_g) \rangle.$$

This together with [53, Theorem 10.1, Exercise 8.8] shows that (2.7) holds at the tth iteration.

In addition, note that $x \mapsto L_t(x, \widetilde{\lambda})$ is strongly convex with modulus $\widetilde{L}_{fg}$. Then we see that for any $x \in \mathbb{R}^n$,

(2.9)
$$f(x^t) + \langle \nabla f(x^t), \widetilde{x} - x^t \rangle + \frac{\widetilde{L}_f}{2} \|\widetilde{x} - x^t\|^2 + P_1(\widetilde{x}) - P_2(x^t) - \langle \xi^t, \widetilde{x} - x^t \rangle$$

$$= L_t(\widetilde{x}, \widetilde{\lambda}) \leq L_t(x, \widetilde{\lambda}) - \frac{\widetilde{L}_{fg}}{2} \|x - \widetilde{x}\|^2$$

$$= f(x^t) + \langle \nabla f(x^t), x - x^t \rangle + \frac{\widetilde{L}_{fg}}{2} \|x - x^t\|^2 + P_1(x) - P_2(x^t) - \langle \xi^t, x - x^t \rangle$$

$$+ \sum_{i=1}^m \widetilde{\lambda}_i \left(g_i(x^t) + \langle \nabla g_i(x^t), x - x^t \rangle \right) - \frac{\widetilde{L}_{fg}}{2} \|x - \widetilde{x}\|^2,$$

where the first equality makes use of (2.6). On the other hand, since f has Lipschitz continuous gradient (with modulus L_f), if $g(\widetilde{x}) \leq 0$, then we have for any $x \in \mathbb{R}^n$ that

$$F(\widetilde{x}) = f(\widetilde{x}) + P_1(\widetilde{x}) - P_2(\widetilde{x})$$

$$\leq f(x^t) + \langle \nabla f(x^t), \widetilde{x} - x^t \rangle + \frac{L_f}{2} \|\widetilde{x} - x^t\|^2 + P_1(\widetilde{x}) - P_2(\widetilde{x})$$

$$= f(x^t) + \langle \nabla f(x^t), \widetilde{x} - x^t \rangle + \frac{\widetilde{L}_f}{2} \|\widetilde{x} - x^t\|^2 + P_1(\widetilde{x}) - P_2(\widetilde{x}) - \frac{\widetilde{L}_f - L_f}{2} \|\widetilde{x} - x^t\|^2$$

$$\overset{(a)}{\leq} f(x^t) + \langle \nabla f(x^t), \widetilde{x} - x^t \rangle + \frac{\widetilde{L}_f}{2} \|\widetilde{x} - x^t\|^2 + P_1(\widetilde{x})$$

$$- P_2(x^t) - \langle \xi^t, \widetilde{x} - x^t \rangle - \frac{\widetilde{L}_f - L_f}{2} \|\widetilde{x} - x^t\|^2$$

$$\leq f(x^t) + \langle \nabla f(x^t), x - x^t \rangle + \frac{\widetilde{L}_{fg}}{2} \|x - x^t\|^2 + P_1(x) - P_2(x^t) - \langle \xi^t, x - x^t \rangle$$

$$+ \sum_{i=1}^m \widetilde{\lambda}_i \left(g_i(x^t) + \langle \nabla g_i(x^t), x - x^t \rangle \right) - \frac{\widetilde{L}_{fg}}{2} \|x - \widetilde{x}\|^2 - \frac{\widetilde{L}_f - L_f}{2} \|\widetilde{x} - x^t\|^2,$$

where (a) uses the convexity of P_2 and the fact that $\xi^t \in \partial P_2(x^t)$, while the last inequality holds due to (2.9). This shows that (2.8) holds at the tth iteration.

Now we show that there exists k_0 (independent of t) so that the inner loop in Step 3 terminates after finitely many iterations at the tth iteration and returns an x^{t+1} satisfying $g(x^{t+1}) \leq 0$. To this end, let $k_1 \in \mathbb{N}_+$ be such that $\underline{L} \tau^{k_1} > \max\{\frac{1}{2}(c + L_f), L_{g_1}, \ldots, L_{g_m}\}$. Then k_1 does not depend on t and we have

(2.10) $$L_f^{t,0} \tau^{k_1} - \frac{L_f}{2} \geq \underline{L} \tau^{k_1} - \frac{L_f}{2} > \frac{c}{2} \text{ and } (L_g^{t,0})_i \tau^{k_1} \geq \underline{L} \tau^{k_1} \geq L_{g_i} \text{ for } i = 1, \ldots, m.$$

Note that for each i, since g_i has Lipschitz gradient with Lipschitz modulus L_{g_i}, we have for any $(\widetilde{L}_g)_i > 0$ that

$$g_i(\widetilde{x}) \leq g_i(x^t) + \langle \nabla g_i(x^t), \widetilde{x} - x^t \rangle + \frac{L_{g_i}}{2} \|\widetilde{x} - x^t\|^2$$

$$= \bar{G}(\widetilde{x}, x^t, \widetilde{L}_g) + \frac{L_{g_i} - (\widetilde{L}_g)_i}{2} \|\widetilde{x} - x^t\|^2.$$

This together with (2.10) and the update rule of $\widetilde{L}_g$ in Step 3b) shows that after at most k_1 calls of Step 3b), we have $g(\widetilde{x}) \leq 0$. Whenever $\widetilde{x}$ satisfies $g(\widetilde{x}) \leq 0$, we can apply (2.8) with x being x^t to conclude that

$$F(\widetilde{x}) \leq f(x^t) + P_1(x^t) - P_2(x^t) + \langle \widetilde{\lambda}, g(x^t) \rangle - \frac{\widetilde{L}_{fg}}{2} \|x^t - \widetilde{x}\|^2 - \left[\frac{\widetilde{L}_f - L_f}{2} \right] \|x^t - \widetilde{x}\|^2$$

$$\leq f(x^t) + P_1(x^t) - P_2(x^t) - \frac{\langle \widetilde{\lambda}, \widetilde{L}_g \rangle}{2} \|x^t - \widetilde{x}\|^2 - \left[\widetilde{L}_f - \frac{L_f}{2} \right] \|x^t - \widetilde{x}\|^2$$

$$\leq F(x^t) - \left[\widetilde{L}_f - \frac{L_f}{2} \right] \|x^t - \widetilde{x}\|^2,$$

where the second inequality holds because $\widetilde{\lambda} \in \mathbb{R}^n_+$ and $g(x^t) \leq 0$; we also used the fact that $\widetilde{L}_{fg} = \widetilde{L}_f + \langle \widetilde{\lambda}, \widetilde{L}_g \rangle$. Thus, in view of the above two displays, the conditions in Step 3a) must hold when $(\widetilde{L}_g)_i \geq L_{g_i}$ for all i and $\widetilde{L}_f \geq \frac{L_f + c}{2}$; according to the update rules of $\widetilde{L}_f$ and $\widetilde{L}_g$, this happens after at most k_1 calls of Step 3b) and k_1 calls of Step 3c). Thus, at iteration t, the inner loop stops after at most $k_0 := 2k_1$ iterations and outputs an x^{t+1} satisfying $g(x^{t+1}) \leq 0$ and $F(x^{t+1}) \leq F(x^t) - \frac{c}{2} \|x^{t+1} - x^t\|^2$.

Finally, since $g(x^0) \leq 0$ to start with, by induction, we know that for any $t \geq 0$, the inner loop stops after at most k_0 iterations. This together with the fact that $\{(L_f^{t,0}, L_g^{t,0})\} \subseteq [\text{L}, \bar{\text{L}}]^{m+1}$ implies that $\{(L_f^t, L_g^t)\}$ is bounded. Therefore, SCP$_{ls}$ is well defined and items (ii), (iii) and (iv) hold. This completes the proof. □

3. Convergence properties of SCP$_{ls}$.

3.1. Convergence analysis in nonconvex settings. In this section, we analyze SCP$_{ls}$ when F in (1.1) is possibly nonconvex. We first prove some basic properties of the sequence generated by SCP$_{ls}$. Item (iii) in the following theorem was already proved in [46, Theorem 3.7]; we also include its proof here for the ease of the readers.

THEOREM 3.1. *Consider* (1.1) *and suppose that Assumptions 1.1 and 2.3 hold. Let $\{(x^t, L_g^t)\}$ be generated by SCP$_{ls}$. Then the following statements hold:*
  (i) *The sequence $\{x^t\}$ is bounded.*
 (ii) *The sequence $\{\bar{F}(x^{t+1}, x^t, L_g^t)\}$ is nonincreasing and convergent to some real number $\bar{F}^*$, where $\bar{F}$ is defined as in* (1.3). *Moreover, for any $t \geq 1$, we have*

$$(3.1) \qquad \bar{F}(x^{t+1}, x^t, L_g^t) \leq \bar{F}(x^t, x^{t-1}, L_g^{t-1}) - \frac{c}{2} \|x^{t+1} - x^t\|^2.$$

(iii) *It holds that $\lim_{t \to \infty} \|x^{t+1} - x^t\| = 0$.*

Proof. Let F be defined as in (1.1). Then for any $t \geq 0$, we have

$$(3.2) \qquad F(x^{t+1}) - F(x^0) = \sum_{i=0}^{t} [F(x^{i+1}) - F(x^i)] \leq - \sum_{i=0}^{t} \frac{c}{2} \|x^{i+1} - x^i\|^2 \leq 0,$$

where the first inequality follows from (2.5). Since F is level-bounded by Assumption 1.1(iii), we deduce that $\{x^t\}$ is bounded and the conclusion in item (i) holds.

We now prove (ii). Since for any $t \geq 0$, the x^{t+1} belongs to dom F and is feasible for (2.4) with $(\widetilde{L}_f, \widetilde{L}_g) = (L_f^t, L_g^t)$, it holds that

$$(3.3) \qquad \bar{F}(x^{t+1}, x^t, L_g^t) = F(x^{t+1}) \ \text{ for } t \geq 0.$$

This together with (2.5) shows that $\{\bar{F}(x^{t+1}, x^t, L_g^t)\}$ is nonincreasing and (3.1) holds for all $t \geq 1$. Also, thanks to (3.3) and Assumption 1.1, we have

$$\inf_t \bar{F}(x^{t+1}, x^t, L_g^t) = \inf_t F(x^t) \geq \inf F > -\infty,$$

implying that $\{\bar{F}(x^{t+1}, x^t, L_g^t)\}$ is bounded from below. Thus, we conclude that the sequence $\{\bar{F}(x^{t+1}, x^t, L_g^t)\}$ is convergent. We denote this limit by $\bar{F}^*$.

Finally, we prove (iii). Since $\{\bar{F}(x^{t+1}, x^t, L_g^t)\}$ converges to $\bar{F}^*$, passing to the limit as t goes to infinity in (3.2) and invoking (3.3), we have

$$\sum_{i=0}^{\infty} \frac{c}{2}\|x^{i+1} - x^i\|^2 \leq F(x^0) - \lim_{t \to \infty} \bar{F}(x^{t+1}, x^t, L_g^t) = F(x^0) - \bar{F}^* < \infty.$$

Therefore, item (iii) holds. This completes the proof.   □

Next, we show that $\{\lambda^t\}$ with each λ^t being a Lagrange multiplier[5] of (2.4) with $(\widetilde{L}_f, \widetilde{L}_g) = (L_f^t, L_g^t)$ is bounded and any cluster point of the sequence $\{x^t\}$ generated by SCP$_{ls}$ is a stationary point of (1.1) in the sense of Definition 2.2. The latter conclusion was also proved in [46, Theorem 3.7]. We include its proof for completeness.

THEOREM 3.2. *Consider* (1.1) *and suppose that Assumptions 1.1 and 2.3 hold. Let $\{x^t\}$ be the sequence generated by SCP$_{ls}$ and λ^t be a Lagrange multiplier of* (2.4) *with $(\widetilde{L}_f, \widetilde{L}_g) = (L_f^t, L_g^t)$. Then the sequence $\{\lambda^t\}$ is bounded and any accumulation point of $\{x^t\}$ is a stationary point of* (1.1).

Proof. Suppose to the contrary that $\{\lambda^t\}$ is unbounded and let $\{\lambda^{t_j}\}$ be a subsequence of $\{\lambda^t\}$ such that $\|\lambda^{t_j}\| \xrightarrow{j} \infty$. Passing to a further subsequence if necessary, we may assume that there exist $\lambda^* \in \mathbb{R}_+^m$ and x^* such that $\lim_{j \to \infty} \frac{\lambda^{t_j}}{\|\lambda^{t_j}\|} = \lambda^*$ and $\lim_{j \to \infty} x^{t_j} = x^*$, where the existence of x^* is due to Theorem 3.1(i).

Using (2.7), the definition of $\widetilde{L}_{fg}$ there and the fact $(\widetilde{L}_f, \widetilde{L}_g) = (L_f^t, L_g^t)$, we have

$$\eta^t := \sum_{i=1}^{m} \lambda_i^t \left[\nabla g_i(x^t) + (L_g^t)_i(x^{t+1} - x^t)\right] \in -\nabla f(x^t) - L_f^t(x^{t+1} - x^t) - \partial P_1(x^{t+1}) + \xi^t.$$

Since the functions ∇f, P_1 and P_2 are continuous, and $\{(x^t, L_f^t)\}$ is bounded thanks to Theorem 3.1(i) and Lemma 2.4(ii), we deduce from the above display that $\{\eta^t\}$ is bounded. Then, dividing η^{t_j} by $\|\lambda^{t_j}\|$ and letting $j \to \infty$, using the continuity of ∇g and Theorem 3.1(iii) together with Lemma 2.4(ii), we deduce further that

$$(3.4) \qquad \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0.$$

---

[5]The existence of λ^t follows from Lemma 2.4(iv).

On the other hand, using (2.6) with $(\widetilde{x}, \widetilde{\lambda}, \widetilde{L}_g) = (x^{t+1}, \lambda^t, L_g^t)$, the continuity of ∇g_i for each i, Lemma 2.4(ii) and Theorem 3.1(iii), we see that $\lambda_i^* g_i(x^*) = 0$ for all $i = 1, \ldots, m$. This further implies that

$$\lambda_i^* = 0 \text{ for } i \notin I(x^*).$$

The above display and (3.4) imply that

$$\sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*) = 0.$$

Combining this with MFCQ (Assumption 2.3) and recalling that $\lambda^* \in \mathbb{R}_+^m$, we conclude that $\lambda_i^* = 0$ for $i \in I(x^*)$. Therefore, we have $\lambda^* = 0$, contradicting the fact that $\|\lambda^*\| = 1$. Thus, the sequence $\{\lambda^t\}$ is bounded.

For the second conclusion of this theorem, let $\bar{x}$ be an accumulation point of $\{x^t\}$ with $\lim_{k \to \infty} x^{t_k} = \bar{x}$. Since $\{\lambda^t\}$ is bounded, passing to a further subsequence if necessary, we assume without loss of generality that $\lim_{k \to \infty} \lambda^{t_k} = \bar{\lambda}$ for some $\bar{\lambda}$. Since the sequence $\{(L_f^t, L_g^t, \lambda^t)\}$ is bounded thanks to Lemma 2.4(ii) and the boundedness of $\{\lambda^t\}$, using Theorem 3.1(iii), we have that $\lim_{k \to \infty} \left(L_f^{t_k} + \langle \lambda^{t_k}, L_g^{t_k} \rangle \right)(x^{t_k+1} - x^{t_k}) = 0$. Using this fact together with the closedness of ∂P_1 and ∂P_2, the Lipschitz continuity of ∇f and ∇g and Theorem 3.1(iii), we have upon passing to the limit as k goes to infinity in (2.7) with $(\widetilde{x}, \widetilde{\lambda}, \widetilde{L}_f, \widetilde{L}_g) = (x^{t_k+1}, \lambda^{t_k}, L_f^{t_k}, L_g^{t_k})$ and $t = t_k$ that

$$(3.5) \qquad 0 \in \nabla f(\bar{x}) + \partial P_1(\bar{x}) - \partial P_2(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i \nabla g_i(\bar{x}).$$

On the other hand, using (2.6) with $(\widetilde{x}, \widetilde{\lambda}, \widetilde{L}_g) = (x^{t_k+1}, \lambda^{t_k}, L_g^{t_k})$ and $t = t_k$, letting $k \to \infty$, we have upon using the continuity of ∇g, Theorem 3.1(iii) and Lemma 2.4(ii) that

$$(3.6) \qquad \bar{\lambda}_i g_i(\bar{x}) = 0 \text{ for all } i = 1, \ldots, m.$$

Finally, since $\lambda^t \geq 0$ for any $t \geq 0$, we have $\bar{\lambda} \geq 0$. Also, since g_i is continuous for each i and $g(x^t) \leq 0$ thanks to Step 3a) of SCP$_{ls}$, we have $g(\bar{x}) \leq 0$. These together with (3.5) and (3.6) imply that $\bar{x}$ is a stationary point of (1.1). □

LEMMA 3.3. *Consider (1.1) and suppose that Assumptions 1.1 and 2.3 hold. Let $\{(x^t, L_g^t)\}$ be the sequence generated by SCP_{ls} and let Ω be the set of accumulation points of the sequence $\{(x^{t+1}, x^t, L_g^t)\}$. Then $\Omega \neq \emptyset$ and $\bar{F} \equiv \bar{F}^*$ on Ω, where $\bar{F}$ is defined as in (1.3) and $\bar{F}^*$ is given in Theorem 3.1(ii).*

Proof. From Theorem 3.1(i) and Lemma 2.4(ii) we know that $\Omega \neq \emptyset$. Fix any $(x^\Omega, y^\Omega, L^\Omega) \in \Omega$ and let $\{(x^{t_j+1}, x^{t_j}, L_g^{t_j})\}$ be a convergent subsequence with $\lim_{j \to \infty} (x^{t_j+1}, x^{t_j}, L_g^{t_j}) = (x^\Omega, y^\Omega, L^\Omega)$. Since each ∇g_i is continuous and x^{t_j+1} belongs to dom F and is feasible for (2.4) with $t = t_j$ and $(\widetilde{L}_f, \widetilde{L}_g) = (L_f^{t_j}, L_g^{t_j})$, we have

$$(3.7) \qquad g(x^\Omega) = \lim_{j \to \infty} g(x^{t_j+1}) \leq 0, \quad \bar{G}(x^\Omega, y^\Omega, L^\Omega) = \lim_{j \to \infty} \bar{G}(x^{t_j+1}, x^{t_j}, L_g^{t_j}) \leq 0$$

and $F(x^{t_j+1}) = \bar{F}(x^{t_j+1}, x^{t_j}, L_g^{t_j})$ for all j. Then, using the continuity of F on its closed domain, we have

$$F(x^\Omega) = \lim_{j \to \infty} F(x^{t_j+1}) = \lim_{j \to \infty} \bar{F}(x^{t_j+1}, x^{t_j}, L_g^{t_j}) = \bar{F}^*,$$

where the last equality follows from Theorem 3.1(ii). Thus, we deduce that

$$\bar{F}(x^\Omega, y^\Omega, L^\Omega) = F(x^\Omega) = \bar{F}^*,$$

where the first equality follows from (3.7). Since $(x^\Omega, y^\Omega, L^\Omega) \in \Omega$ is arbitrary, we conclude that $\bar{F} \equiv \bar{F}^*$ on Ω. □

To analyze the global convergence properties of SCP_{ls}, we need a bound on the subdifferential of $\bar{F}$ in (1.3). To this end, we consider the following additional differentiability assumption on g_i.

ASSUMPTION 3.4. *Each g_i in (1.1) is twice continuously differentiable.*

LEMMA 3.5. *Consider (1.1) and suppose that Assumption 3.4 holds. Let $(x, y, w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ and assume that P_2 is continuously differentiable around x. Then*

$$(3.8) \quad \partial\bar{F}(x,y,w) \supseteq \begin{pmatrix} \nabla f(x) - \nabla P_2(x) + \partial P_1(x) + \sum_{i=1}^{m} \lambda_i [\nabla g_i(y) + w_i(x-y)] \\ \sum_{i=1}^{m} \lambda_i [\nabla^2 g_i(y)(x-y) - w_i(x-y)] \\ \frac{1}{2}\|x-y\|^2 \lambda \end{pmatrix}$$

whenever $\lambda \in N_{-\mathbb{R}_+^m}(\bar{G}(x,y,w))$, where $\bar{F}$ and $\bar{G}$ are defined as in (1.3).

Proof. We only consider the case where $(x, y, w) \in \mathrm{dom}\bar{F}$, since (3.8) holds trivially otherwise. Using [53, Exercise 8.8, Corollary 10.9, Proposition 10.5], we have

$$\partial\bar{F}(x,y,w) \supseteq \hat{\partial}\bar{F}(x,y,w) \supseteq \begin{pmatrix} \nabla f(x) - \nabla P_2(x) + \hat{\partial}P_1(x) \\ 0 \\ 0 \end{pmatrix} + \hat{\partial}\delta_{\bar{G}(\cdot)\leq 0}(x,y,w)$$

$$\overset{(a)}{=} \begin{pmatrix} \nabla f(x) - \nabla P_2(x) + \partial P_1(x) \\ 0 \\ 0 \end{pmatrix} + \hat{N}_{\bar{G}(\cdot)\leq 0}(x,y,w)$$

$$\overset{(b)}{\supseteq} \begin{pmatrix} \nabla f(x) - \nabla P_2(x) + \partial P_1(x) \\ 0 \\ 0 \end{pmatrix} + \sum_{i=1}^{m} \lambda_i \begin{pmatrix} \nabla g_i(y) + w_i(x-y) \\ \nabla^2 g_i(y)(x-y) - w_i(x-y) \\ \frac{1}{2}\|x-y\|^2 e_i \end{pmatrix},$$

where (a) uses the convexity of P_1 and [53, Proposition 8.12], $e_i \in \mathbb{R}^m$ is the ith standard basis vector and (b) holds for any $\lambda \in \hat{N}_{-\mathbb{R}_+^m}(\bar{G}(x,y,w)) = N_{-\mathbb{R}_+^m}(\bar{G}(x,y,w))$, thanks to [53, Theorem 6.14]. □

We also need the following assumption to derive the desired bound on $\partial\bar{F}$. This assumption was also used in [59] for analyzing the global convergence property of the sequence generated by the proximal DCA with extrapolation (pDCA$_e$).

ASSUMPTION 3.6. *Each g_i in (1.1) is smooth, and the P_2 in (1.1) is continuously differentiable on an open set Γ that contains all stationary points of (1.1). Moreover, the function ∇P_2 is locally Lipschitz continuous on Γ.*

Using this assumption and Lemma 3.5, we can prove the following property of $\partial\bar{F}$.

LEMMA 3.7. *Consider (1.1) and suppose that Assumptions 1.1, 2.3, 3.4 and 3.6 hold. Let $\{(x^t, L_g^t)\}$ be the sequence generated by SCP_{ls} and let $\bar{F}$ be defined as in (1.3). Then there exist $\kappa > 0$ and $\underline{t} \in \mathbb{N}_+$ such that*

$$(3.9) \qquad \mathrm{dist}(0, \partial\bar{F}(x^{t+1}, x^t, L_g^t)) \leq \kappa\|x^{t+1} - x^t\| \quad \text{for all } t > \underline{t}.$$

Proof. From Theorem 3.1(i), we know that $\{x^t\}$ is bounded. Thus, denoting the set of accumulation points of $\{x^t\}$ as Ω_x, we have that Ω_x is compact and $\Omega_x \subseteq \Gamma$ thanks to Theorem 3.2, where Γ is the open set give in Assumption 3.6. Choose an $\epsilon > 0$ so that $\Gamma_\epsilon := \{x : \operatorname{dist}(x, \Omega_x) < \epsilon\} \subseteq \Gamma$ and ∇P_2 is Lipschitz continuous with modulus L_{P_2} on Γ_ϵ, which exists thanks to the compactness of Ω_x and Assumption 3.6. Moreover, since Ω_x is compact, from the definition of cluster points, we see that there exists $t_0 \in \mathbb{N}_+$ such that $\operatorname{dist}(x^t, \Omega_x) < \epsilon$ whenever $t > t_0$. In particular, P_2 is continuously differentiable around each x^t whenever $t > t_0$. In addition, thanks to Theorem 3.1(iii), we can further choose $\underline{t} > t_0 + 1$ such that for $t > \underline{t}$, we have

$$\tag{3.10} \|x^{t+1} - x^t\|^2 \le \|x^{t+1} - x^t\|.$$

Now, let λ^t be a Lagrange multiplier of (2.4) with $(\widetilde{L}_f, \widetilde{L}_g) = (L_f^t, L_g^t)$, which exists thanks to Lemma 2.4(iv). Then it holds that $\lambda^t \in N_{-\mathbb{R}_+^m}(\bar{G}(x^{t+1}, x^t, L_g^t))$. Therefore, using (3.8) with $\lambda = \lambda^t$ for any $t > \underline{t}$, we have that

$$\tag{3.11} \partial \bar{F}(x^{t+1}, x^t, L_g^t) \supseteq \begin{pmatrix} J^t \\ \sum_{i=1}^m \lambda_i^t \left(\nabla^2 g_i(x^t)(x^{t+1} - x^t) - (L_g^t)_i(x^{t+1} - x^t) \right) \\ \frac{1}{2}\|x^{t+1} - x^t\|^2 \lambda^t \end{pmatrix}$$

with $J^t := \nabla f(x^{t+1}) + \partial P_1(x^{t+1}) - \nabla P_2(x^{t+1}) + \sum_{i=1}^m \lambda_i^t \left(\nabla g_i(x^t) + (L_g^t)_i(x^{t+1} - x^t) \right)$. For this J^t, using (2.7) with $\widetilde{x} = x^{t+1}$ and recalling the definition of ξ^t, we have that

$$J^t \ni \nabla f(x^{t+1}) - \nabla P_2(x^{t+1}) + \sum_{i=1}^m \lambda_i^t \left(\nabla g_i(x^t) + (L_g^t)_i(x^{t+1} - x^t) \right)$$

$$+ \left(-\nabla f(x^t) - L_f^t(x^{t+1} - x^t) + \nabla P_2(x^t) - \sum_{i=1}^m \lambda_i^t \left(\nabla g_i(x^t) + (L_g^t)_i(x^{t+1} - x^t) \right) \right)$$

$$= \nabla f(x^{t+1}) - \nabla f(x^t) + \nabla P_2(x^t) - \nabla P_2(x^{t+1}) - L_f^t(x^{t+1} - x^t).$$

Using this together with Cauchy-Schwarz inequality, for $t > \underline{t}$, it holds that

$$\|J^t\|^2 \le 3\left(\|\nabla f(x^{t+1}) - \nabla f(x^t)\|^2 + \|\nabla P_2(x^{t+1}) - \nabla P_2(x^t)\|^2 + \|L_f^t(x^{t+1} - x^t)\|^2 \right)$$

$$\tag{3.12} \overset{(a)}{\le} 3L_f^2\|x^{t+1} - x^t\|^2 + 3L_{P_2}^2\|x^{t+1} - x^t\|^2 + 3(L_f^t)^2\|x^{t+1} - x^t\|^2$$

$$= \left(3L_f^2 + 3(L_f^t)^2 + 3L_{P_2}^2 \right)\|x^{t+1} - x^t\|^2,$$

where (a) makes use of the fact that $t > \underline{t}$ (so that $x^t \in \Gamma_\epsilon$) and the Lipschitz continuity of ∇f and ∇P_2.

On the other hand, since $\{(x^t, L_g^t, \lambda^t)\}$ is bounded thanks to Theorem 3.1(i), Lemma 2.4(ii) and Theorem 3.2, using the continuity of $\nabla^2 g_i$ for each i, there exists $D_1 > 0$ such that

$$\tag{3.13} \left\| \sum_{i=1}^m \lambda_i^t \left(\nabla^2 g_i(x^t)(x^{t+1} - x^t) - (L_g^t)_i(x^{t+1} - x^t) \right) \right\|^2$$

$$\le m \sum_{i=1}^m (\lambda_i^t)^2 \|\nabla^2 g_i(x^t)(x^{t+1} - x^t) - (L_g^t)_i(x^{t+1} - x^t)\|^2 \le D_1 \|x^{t+1} - x^t\|^2,$$

where the first inequality uses the Cauchy-Schwarz inequality.

Therefore, since $\{(L_f^t, \lambda^t)\}$ is bounded thanks to Lemma 2.4(ii) and Theorem 3.2, combining (3.10), (3.11), (3.12) and (3.13), we conclude that there exists $\kappa > 0$ such that (3.9) holds. This completes the proof. $\square$

Now, if we suppose in addition that $\bar{F}$ is a KL function with exponent $\alpha \in [0, 1)$, then using the results above and following the analysis in [2–4, 14, 44, 59], we can deduce the convergence of the sequence $\{x^t\}$ generated by SCP_{ls} to a stationary point of (1.1) and estimate its local convergence rate. Specifically, using similar proofs as in [44, 59], we have the following results. The lines of arguments are standard and we omit its proof for brevity.

THEOREM 3.8 (**Convergence rate of SCP_{ls} in nonconvex settings**). *Consider* (1.1). *Suppose that Assumptions 1.1, 2.3, 3.4 and 3.6 hold, and $\bar{F}$ in (1.3) is a KL function. Let $\{(x^t, L_g^t)\}$ be the sequence generated by SCP_{ls} and let Ω be the set of accumulation points of the sequence $\{(x^{t+1}, x^t, L_g^t)\}$. Then $\{x^t\}$ converges to a stationary point x^* of* (1.1). *Moreover, if $\bar{F}$ satisfies the KL property with exponent $\alpha \in [0, 1)$ at every point in Ω, then there exists $\underline{t} \in \mathbb{N}_+$ such that the following statements hold:*

(i) *If $\alpha = 0$, then $\{x^t\}$ converges finitely, i.e., $x^t \equiv x^*$ for $t > \underline{t}$.*
(ii) *If $\alpha \in (0, \frac{1}{2}]$, then there exist $a_0 \in (0, 1)$ and $a_1 > 0$ such that*

$$\|x^t - x^*\| \le a_1 a_0^t \quad \text{for } t > \underline{t}.$$

(iii) *If $\alpha \in (\frac{1}{2}, 1)$, then there exists $a_2 > 0$ such that*

$$\|x^t - x^*\| \le a_2 t^{-\frac{1-\alpha}{2\alpha - 1}} \quad \text{for } t > \underline{t}.$$

3.2. Convergence analysis in convex settings. In this section, we study the convergence properties of SCP_{ls} under the following convex settings:

ASSUMPTION 3.9. *Suppose that in* (1.1), *$P_2 = 0$ and $\{f, g_1, \ldots, g_m\}$ are convex.*

Assumption 3.9 was also considered in [10, Section 3.2.3] for analyzing MBA, and in [10, Section 4] for its line search variant Multiprox$_{bt}$ [10, Eq. (37)]. Here, we would like to point out that the line search criterion in Multiprox$_{bt}$ [10, Eq. (37)] is different from the criterion (2.5) used in SCP_{ls}. The criterion in Multiprox$_{bt}$ relies on a local majorant of the objective function, while (2.5) uses the objective function directly, and is originated from SpaRSA; see [51, Eq. (22)]. We will establish global convergence of the whole sequence generated by SCP_{ls} in the above convex settings, under suitable assumptions. Unlike the analysis in the previous subsection, our analysis here is based on KL property of F in (1.1) instead of that of $\bar{F}$, and we will *not* assume g to be twice continuously differentiable (i.e., we do not require Assumption 3.4). We start with two auxiliary lemmas. The first lemma is an analogue of [14, Lemma 6] and follows immediately from an application of [12, Theorem 5] and standard compactness argument. We omit the proof for brevity.

LEMMA 3.10. *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a level-bounded proper closed convex function with $\Lambda := \text{Arg min} f \ne \emptyset$. Let $\underline{f} := \inf f$. Suppose that f satisfies the KL property at each point in Λ with exponent $\alpha \in [0, 1)$. Then there exist $\epsilon > 0$, $r_0 > 0$ and $c_0 > 0$ such that*

$$\text{dist}(x, \Lambda) \le c_0 (f(x) - \underline{f})^{1-\alpha}$$

for any $x \in \text{dom}\partial f$ satisfying $\text{dist}(x, \Lambda) \le \epsilon$ and $\underline{f} \le f(x) < \underline{f} + r_0$.

The next lemma is an analogue of Lemma 3.3 for F in (1.1).

LEMMA 3.11. *Consider* (1.1) *and suppose that Assumptions 1.1 and 2.3 hold. Let* $\{x^t\}$ *be the sequence generated by SCP$_{ls}$ for* (1.1) *and let* Ω_x *be the set of accumulation point of* $\{x^t\}$. *Then the following statements hold:*
  (i) *It holds that* $\Omega_x \neq \emptyset$ *and* $F \equiv \bar{F}^*$ *on* Ω_x, *where* F *is defined as in* (1.1) *and* $\bar{F}^*$ *is given in Theorem 3.1(ii).*
 (ii) *The sequence* $\{F(x^t)\}$ *is nonincreasing and convergent to* $\bar{F}^*$.

Proof. We note first from Theorem 3.1(i) that $\Omega_x \neq \emptyset$. In addition, since $x^t \in \operatorname{dom} F$ and is feasible for (2.4) (with $(t-1, L_f^{t-1}, L_g^{t-1})$ in place of $(t, \widetilde{L}_f, \widetilde{L}_g)$), we have

$$(3.14) \qquad F(x^t) = f(x^t) + P_1(x^t) - P_2(x^t) = \bar{F}(x^t, x^{t-1}, L_g^{t-1}), \text{ for all } t \geq 1.$$

Fix any $x^* \in \Omega_x$ and let $\lim_{j\to\infty} x^{t_j} = x^*$. Using the continuity of F on its closed domain and (3.14), we see that

$$F(x^*) = \lim_{j\to\infty} f(x^{t_j}) + P_1(x^{t_j}) - P_2(x^{t_j}) = \lim_{j\to\infty} \bar{F}(x^{t_j}, x^{t_j-1}, L_g^{t_j-1}) = \bar{F}^*,$$

where the last equality makes use of Theorem 3.1(ii). This proves (i). The conclusion in (ii) now follows immediately upon combining the above display and (3.14) with Theorem 3.1(ii). This completes the proof. □

Now we present our main result in this subsection.

THEOREM 3.12 (**Convergence rate of SCP$_{ls}$ in convex settings**). *Consider* (1.1) *and suppose that Assumptions 1.1, 2.3 and 3.9 hold. Let* $\{x^t\}$ *be the sequence generated by SCP$_{ls}$. Then* $\{x^t\}$ *converges to a minimizer* x^* *of* (1.1). *If in addition* F *in* (1.1) *is a KL function with exponent* $\alpha \in [0, 1)$, *then the following statements hold:*
  (i) *If* $\alpha \in [0, \frac{1}{2}]$, *then there exist* $c_0 > 0$, $Q_1 \in (0, 1)$ *and* $\underline{t} \in \mathbb{N}_+$, *such that*

$$\|x^t - x^*\| \leq c_0 Q_1^t \quad \text{for } t > \underline{t}.$$

 (ii) *If* $\alpha \in (\frac{1}{2}, 1)$, *then there exist* $c_0 > 0$ *and* $\underline{t} \in \mathbb{N}_+$ *such that*

$$\|x^t - x^*\| \leq c_0\, t^{-\frac{1-\alpha}{2\alpha-1}} \text{ for } t > \underline{t}.$$

Proof. Let $S := \operatorname{Arg\,min} F$ for notational simplicity. Note that $S \neq \emptyset$ thanks to Assumption 1.1. Since $P_2 = 0$ and $\{f, g_1, \ldots, g_m\}$ are convex by Assumption 3.9, using Theorem 3.2 and [52, Theorem 28.3], we see that

$$(3.15) \qquad\qquad \emptyset \neq \Omega_x \subseteq S,$$

where Ω_x is as in Lemma 3.11. This together with Lemma 3.11 implies that $\bar{F}^* = \inf F$.

Next, let λ^t be a Lagrange multiplier of (2.4) with $(\widetilde{L}_f, \widetilde{L}_g) = (L_f^t, L_g^t)$, which exists thanks to Lemma 2.4(iv). Since $P_2 = 0$ and $g(x^t) \leq 0$ for all t, for any $\bar{x} \in S$, using (2.8) with $x = \bar{x}$, $\widetilde{x} = x^{t+1}$, $\widetilde{\lambda} = \lambda^t$, $\widetilde{L}_f = L_f^t$ and $\widetilde{L}_{fg} = L_{fg}^t := L_f^t + \langle \lambda^t, L_g^t \rangle$, we deduce that

$$F(x^{t+1}) \leq f(x^t) + \left\langle \nabla f(x^t), \bar{x} - x^t \right\rangle + P_1(\bar{x}) + \frac{L_{fg}^t}{2}\|\bar{x} - x^t\|^2 - \frac{L_{fg}^t}{2}\|\bar{x} - x^{t+1}\|^2$$

$$+ \sum_{i=1}^m \lambda_i^t \left(g_i(x^t) + \langle \nabla g_i(x^t), \bar{x} - x^t \rangle \right) - \frac{L_f^t - L_f}{2}\|x^{t+1} - x^t\|^2$$

$$\overset{(a)}{\le} f(\bar{x}) + P_1(\bar{x}) + \frac{L_{fg}^t}{2}\|\bar{x} - x^t\|^2 - \frac{L_{fg}^t}{2}\|\bar{x} - x^{t+1}\|^2 - \frac{L_f^t - L_f}{2}\|x^{t+1} - x^t\|^2$$

$$\overset{(b)}{\le} f(\bar{x}) + P_1(\bar{x}) + \frac{L_{fg}^t}{2}\|\bar{x} - x^t\|^2 - \frac{L_{fg}^t}{2}\|x^{t+1} - \bar{x}\|^2 + \frac{(L_f - L_f^t)_+}{c}(F(x^t) - F(x^{t+1}))$$

$$\le f(\bar{x}) + P_1(\bar{x}) + \frac{L_{fg}^t}{2}\|\bar{x} - x^t\|^2 - \frac{L_{fg}^t}{2}\|x^{t+1} - \bar{x}\|^2 + \frac{M_0}{c}(F(x^t) - F(x^{t+1})),$$

where (a) holds because $\{f, g_1, \ldots, g_m\}$ are convex, and $\lambda_i^t \ge 0$ and $g_i(\bar{x}) \le 0$ for all i, (b) follows from (2.5), and the M_0 in the last inequality is an upper bound of $\{(L_f - L_f^t)_+\}$, which exists thanks to Lemma 2.4(ii). Rearranging terms in the above inequality and noting $\bar{F}^* = \inf F = f(\bar{x}) + P_1(\bar{x})$ whenever $\bar{x} \in S$, we have for any $\bar{x} \in S$ that

$$\frac{F(x^{t+1}) - \bar{F}^*}{L_{fg}^t} \le \frac{1}{2}\|\bar{x} - x^t\|^2 - \frac{1}{2}\|x^{t+1} - \bar{x}\|^2 + \frac{M_0}{cL_{fg}^t}\left(F(x^t) - F(x^{t+1})\right).$$

Let $L_{\max}$ be the upper bound of $\{L_{fg}^t\}$ (which exists according to Lemma 2.4(ii) and Theorem 3.2) and recall that $L_{fg}^t \ge L_f^t \ge \underline{L} > 0$ for all t, where $\underline{L}$ is the one used in Step 2 of SCP$_{ls}$. Then we have from the above display that for any $\bar{x} \in S$,

$$\gamma\left(F(x^{t+1}) - \bar{F}^*\right) \le \frac{1}{2}\|\bar{x} - x^t\|^2 - \frac{1}{2}\|x^{t+1} - \bar{x}\|^2 + \theta\left(F(x^t) - F(x^{t+1})\right),$$

where $\gamma := \frac{1}{L_{\max}}$ and $\theta := \frac{M_0}{c\underline{L}}$. Rearranging terms in the above inequality, we have

$$(3.16) \quad (\gamma + \theta)\left(F(x^{t+1}) - \bar{F}^*\right) \le \frac{1}{2}\|\bar{x} - x^t\|^2 - \frac{1}{2}\|x^{t+1} - \bar{x}\|^2 + \theta\left(F(x^t) - \bar{F}^*\right).$$

The inequality above in particular implies that for any $\bar{x} \in S$,

$$(3.17) \quad \begin{aligned} \frac{1}{2}\|x^{t+1} - \bar{x}\|^2 &\le \frac{1}{2}\|\bar{x} - x^t\|^2 + \theta\left(F(x^t) - \bar{F}^*\right) - (\gamma + \theta)\left(F(x^{t+1}) - \bar{F}^*\right) \\ &\le \frac{1}{2}\|\bar{x} - x^t\|^2 + (\gamma + \theta)\left(F(x^t) - F(x^{t+1})\right), \end{aligned}$$

where the last inequality holds because $\bar{F}^* = \inf F \le F(x^t)$. Since $\{F(x^t) - F(x^{t+1})\}$ is nonnegative and summable thanks to Lemma 3.11(ii), using (3.15), (3.17) and [31, Proposition 1], we conclude that $\{x^t\}$ converges to a minimizer x^* of (1.1).

Now, we suppose in addition that F is a KL function with exponent $\alpha \in [0, 1)$. Let $\bar{x}^t \in S$ satisfy $\|x^t - \bar{x}^t\| = \mathrm{dist}(x^t, S)$. Since $\bar{x}^t \in S$, it holds that $-\|x^{t+1} - \bar{x}^t\|^2 \le -\mathrm{dist}^2(x^{t+1}, S)$. Using this and applying (3.16) with $\bar{x}^t$ in place of $\bar{x}$ gives

$$(3.18) \quad (\gamma + \theta)\left(F(x^{t+1}) - \bar{F}^*\right) \le \frac{1}{2}\mathrm{dist}^2(x^t, S) - \frac{1}{2}\mathrm{dist}^2(x^{t+1}, S) + \theta\left(F(x^t) - \bar{F}^*\right).$$

For notational simplicity, let

$$(3.19) \quad \beta_t := F(x^t) - \bar{F}^* + \frac{1}{2(\gamma + \theta)}\mathrm{dist}^2(x^t, S).$$

Using this, rearranging terms and dividing $\gamma + \theta$ from both sides of (3.18), we have

$$(3.20) \quad \beta_{t+1} \le \frac{\theta}{\gamma + \theta}\left(F(x^t) - \bar{F}^*\right) + \frac{1}{2(\gamma + \theta)}\mathrm{dist}^2(x^t, S).$$

Since F is a proper closed convex level-bounded KL function with exponent $\alpha \in [0, 1)$, using Lemma 3.10, there exist $0 < \bar{a} < 1$, $\bar{c} > 0$ and $0 < \epsilon < 1$ such that

$$(3.21) \qquad \mathrm{dist}(x, S)^{\frac{1}{1-\alpha}} \le \bar{c}\left(F(x) - \bar{F}^*\right)$$

for any $x \in \mathrm{dom}\partial F$ satisfying $\mathrm{dist}(x, S) \le \epsilon$ and $\bar{F}^* \le F(x) < \bar{F}^* + \bar{a}$.

Clearly, $\{x^t\} \subset \mathrm{dom}\partial F = \{x : \ g(x) \le 0\}$. Next, since $\{x^t\}$ is bounded thanks to Theorem 3.1(i), using (3.15), there exists t_1 such that

$$(3.22) \qquad \mathrm{dist}(x^t, S) \le \mathrm{dist}(x^t, \Omega_x) < \epsilon, \ \text{for } t > t_1.$$

On the other hand, using Lemma 3.11(ii), we see that there exists t_2 such that

$$(3.23) \qquad \bar{F}^* \le F(x^t) < \bar{F}^* + \bar{a}, \ \text{for } t > t_2.$$

We now consider the cases when $\alpha \in [0, \frac{1}{2}]$ and $\alpha \in (\frac{1}{2}, 1)$ separately.

Case (i). $\alpha \in [0, \frac{1}{2}]$. Combining (3.21), (3.22) and (3.23), we conclude that for any $t > t_3 := \max\{t_1, t_2\}$,

$$(3.24) \qquad \mathrm{dist}^2(x^t, S) \le \mathrm{dist}^{\frac{1}{1-\alpha}}(x^t, S) \le \bar{c}\left(F(x^t) - \bar{F}^*\right),$$

where the first inequality holds because $\frac{1}{1-\alpha} \le 2$ and $\mathrm{dist}(x^t, S) < \epsilon < 1$. Next, let $\zeta := \frac{2\theta + \bar{c}}{2(\gamma + \theta) + \bar{c}} \in (0, 1)$. Then one can show that

$$(3.25) \qquad \frac{\theta}{\gamma + \theta} + \frac{(1-\zeta)\bar{c}}{2(\gamma + \theta)} = \zeta.$$

Using this and (3.20), we have for all $t > t_3$ that

$$
\begin{aligned}
\beta_{t+1} &\le \frac{\theta}{\gamma + \theta}(F(x^t) - \bar{F}^*) + \frac{1-\zeta}{2(\gamma + \theta)}\mathrm{dist}^2(x^t, S) + \frac{\zeta}{2(\gamma + \theta)}\mathrm{dist}^2(x^t, S) \\
&\overset{(a)}{\le} \left(\frac{\theta}{\gamma + \theta} + \frac{(1-\zeta)\bar{c}}{2(\gamma + \theta)}\right)(F(x^t) - \bar{F}^*) + \frac{\zeta}{2(\gamma + \theta)}\mathrm{dist}^2(x^t, S) \\
&\overset{(b)}{=} \zeta\left(F(x^t) - \bar{F}^* + \frac{1}{2(\gamma + \theta)}\mathrm{dist}^2(x^t, S)\right) = \zeta\beta_t,
\end{aligned}
$$

where (a) follows from (3.24) and (b) follows from (3.25). Combining the above inequality with the definition of β_t in (3.19) gives

$$(3.26) \qquad F(x^t) - \bar{F}^* \le \beta_t \le \zeta^{t-t_3-1}\beta_{t_3+1} \ \text{ for } t > t_3.$$

Then, for $t > t_3$, we have

$$
\begin{aligned}
\|x^* - x^t\| &\le \sum_{j=t+1}^{\infty} \|x^j - x^{j-1}\| \le \sum_{j=t+1}^{\infty} \sqrt{\frac{2}{c}}\sqrt{F(x^{j-1}) - F(x^j)} \\
&\le \sum_{j=t+1}^{\infty} \sqrt{\frac{2}{c}}\sqrt{F(x^{j-1}) - \bar{F}^*} \le \sum_{j=t+1}^{\infty} \sqrt{\frac{2}{c}}\sqrt{\zeta^{j-t_3-2}\beta_{t_3+1}} = \sqrt{\frac{2\beta_{t_3+1}}{c\zeta^{t_3+1}}}\frac{(\sqrt{\zeta})^t}{1 - \sqrt{\zeta}},
\end{aligned}
$$

where the second inequality follows from (2.5), the third inequality follows from Lemma 3.11(ii) and the last inequality follows from (3.26). This proves (i).

Case (ii). $\alpha \in (\frac{1}{2}, 1)$. Using (3.20) and the definition of β_t in (3.19), for any $t > t_3 = \max\{t_1, t_2\}$, we have

$$\beta_{t+1} \leq \beta_t - \frac{\gamma}{\gamma + \theta} \left(F(x^t) - \bar{F}^* \right)$$

$$= \beta_t - \frac{1}{2} c_3 \left[F(x^t) - \bar{F}^* + \bar{c} \left(\frac{1}{2(\gamma + \theta)} \right)^{\frac{1}{2(1-\alpha)}} \left(F(x^t) - \bar{F}^* \right) \right]$$

$$\overset{(a)}{\leq} \beta_t - \frac{1}{2} c_3 \left[F(x^t) - \bar{F}^* + \left(\frac{1}{2(\gamma + \theta)} \right)^{\frac{1}{2(1-\alpha)}} \mathrm{dist}(x^t, S)^{\frac{1}{1-\alpha}} \right]$$

$$\overset{(b)}{\leq} \beta_t - \frac{1}{2} c_3 \left[\left(F(x^t) - \bar{F}^* \right)^{\frac{1}{2(1-\alpha)}} + \left(\frac{1}{2(\gamma + \theta)} \mathrm{dist}^2(x^t, S) \right)^{\frac{1}{2(1-\alpha)}} \right],$$

where $c_3 = 2 \frac{\frac{\gamma}{\gamma + \theta}}{1 + \bar{c} \left(\frac{1}{2(\gamma + \theta)} \right)^{\frac{1}{2(1-\alpha)}}}$, (a) follows from (3.21), (3.22), (3.23) and the fact that $\{x^t\} \subset \mathrm{dom}\partial F = \{x : g(x) \leq 0\}$, and (b) holds because $0 \leq F(x^t) - \bar{F}^* < \bar{a} < 1$ (thanks to (3.23)) and $\frac{1}{2(1-\alpha)} > 1$. Since the mapping $w \mapsto w^{\frac{1}{2(1-\alpha)}}$ is convex, for $t > t_3$, we obtain further that

$$\beta_{t+1} \leq \beta_t - c_3 c_4 \left(F(x^t) - \bar{F}^* + \frac{1}{2(\gamma + \theta)} \mathrm{dist}^2(x^t, S) \right)^{\frac{1}{2(1-\alpha)}}$$

$$= \beta_t - c_3 c_4 \beta_t^{\frac{1}{2(1-\alpha)}} = \beta_t \left(1 - c_3 c_4 \beta_t^{\frac{1}{2(1-\alpha)} - 1} \right),$$

where $c_4 := 2^{-\frac{1}{2(1-\alpha)}}$. Since $\frac{1}{2(1-\alpha)} - 1 = \frac{2\alpha - 1}{2(1-\alpha)} > 0$, using the above inequality and [16, Lemma 4.1], we have

$$(3.27) \qquad \beta_t \leq \left(\beta_{t_3+1}^{-\frac{2\alpha-1}{2(1-\alpha)}} + \frac{2\alpha - 1}{2(1-\alpha)} c_3 c_4 (t - t_3 - 1) \right)^{-\frac{2(1-\alpha)}{2\alpha-1}} \quad \text{for } t > t_3.$$

Then, for any $t > t_3$ and $t' \geq 0$, we have

$$\|x^t - x^{t+t'}\|^2 \leq 2 \left(\|x^t - \bar{x}^t\|^2 + \|\bar{x}^t - x^{t+t'}\|^2 \right)$$

$$\overset{(a)}{\leq} 2 \left(\|x^t - \bar{x}^t\|^2 + \|\bar{x}^t - x^t\|^2 + 2(\gamma + \theta) \left(F(x^t) - F(x^{t+t'}) \right) \right)$$

$$= 2 \left(2\mathrm{dist}^2(x^t, S) + 2(\gamma + \theta) \left(F(x^t) - F(x^{t+t'}) \right) \right)$$

$$\overset{(b)}{\leq} 2 \left(2\mathrm{dist}^2(x^t, S) + 4(\gamma + \theta) \left(F(x^t) - \bar{F}^* \right) \right)$$

$$\overset{(c)}{=} 8(\gamma + \theta) \beta_t \leq 8(\gamma + \theta) \left(\beta_{t_3+1}^{-\frac{2\alpha-1}{2(1-\alpha)}} + \frac{2\alpha - 1}{2(1-\alpha)} c_3 c_4 (t - t_3 - 1) \right)^{-\frac{2(1-\alpha)}{2\alpha-1}},$$

where (a) follows from (3.17) and the first equality uses the definition of $\bar{x}^t$ (i.e., the projection of x^t onto S), (b) follows from Lemma 3.11(ii), (c) uses the definition of β_t and the last inequality follows from (3.27). Letting $t' \to \infty$ and recalling that $x^t \to x^*$, we see that the conclusion in (ii) holds. This completes the proof. □

Remark 3.13. From the proof of the above theorem, we can actually deduce that the sequence $\left\{ F(x^t) - \bar{F}^* + c_0 \mathrm{dist}^2(x^t, S) \right\}$ (with some suitable $c_0 > 0$) is Q-linearly convergent when F is a KL function with exponent $\alpha \in [0, \frac{1}{2}]$, and is sublinearly convergent when F is a KL function with exponent $\alpha \in (\frac{1}{2}, 1)$; see (3.26) and (3.27).

4. KL properties of $\bar{F}$ and F. In Section 3, we deduced the rate of convergence of the sequence $\{x^t\}$ generated by SCP$_{ls}$ under nonconvex and convex settings by imposing KL assumptions on $\bar{F}$ in (1.3) and F in (1.1), respectively; see Theorem 3.8 and Theorem 3.12. Note that the assumptions in Theorem 3.8 and Theorem 3.12 for (1.1) are different as follows:

- Assumptions 1.1, 2.3, 3.4 and 3.6 are used in Theorem 3.8.
- Assumptions 1.1, 2.3, 3.9 are used in Theorem 3.12.

Thus, it is interesting to find a relationship between KL exponent of $\bar{F}$ and that of F when all the above assumptions hold. In this regard, we have the following theorem.

THEOREM 4.1 (**Relation between the KL exponents of $\bar{F}$ and F**). *Let F be defined as in* (1.1) *and suppose that Assumptions 1.1, 2.3, 3.4 and 3.9 hold. If $\bar{F}$ defined in* (1.3) *is a KL function with exponent $\alpha \in [0, 1)$, then F is also a KL function with exponent α.*

Proof. Fix any $x_0 \in \mathrm{dom}\partial F$ and $w_0 \in \mathbb{R}$. Using (3.8) and noting that $P_2 = 0$ (Assumption 3.9), we have for any $x \in \mathrm{dom}\partial F$ that

$$
\begin{aligned}
&\partial \bar{F}(x, x, w_0) \\
&\supseteq \left\{ \begin{pmatrix} \nabla f(x) + \partial P_1(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) \\ 0 \\ 0 \end{pmatrix} : \ \lambda \in N_{-\mathbb{R}_+^m}(\bar{G}(x, x, w_0)) \right\} \\
(4.1) \quad &\overset{(a)}{=} \left\{ \begin{pmatrix} \nabla f(x) + \partial P_1(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) \\ 0 \\ 0 \end{pmatrix} : \ \lambda \in N_{-\mathbb{R}_+^m}(g(x)) \right\} \\
&\overset{(b)}{=} \begin{pmatrix} \nabla f(x) + \partial P_1(x) + N_{g(\cdot) \le 0}(x) \\ 0 \\ 0 \end{pmatrix} \overset{(c)}{=} \begin{pmatrix} \partial F(x) \\ 0 \\ 0 \end{pmatrix},
\end{aligned}
$$

where (a) follows from the fact that $g(x) = \bar{G}(x, x, w_0)$, (b) follows from Assumption 2.3 and [53, Theorem 6.14], and (c) holds due to [53, Exercise 8.8] and [52, Theorem 23.8] together with the convexity of P_1 and g and the continuity of P_1. Using this together with the assumption that $x_0 \in \mathrm{dom}\partial F$, we have $(x_0, x_0, w_0) \in \mathrm{dom}\partial \bar{F}$. Then, from the KL assumption on $\bar{F}$, we see that there exist $a > 0$, $\epsilon > 0$ and $c_0 > 0$ such that

$$
(4.2) \qquad \mathrm{dist}(0, \partial \bar{F}(x, y, w)) \ge a(\bar{F}(x, y, w) - \bar{F}(x_0, x_0, w_0))^\alpha
$$

whenever $0 < \bar{F}(x, y, w) - \bar{F}(x_0, x_0, w_0) < c_0$ and $\|(x, y, w) - (x_0, x_0, w_0)\| \le \epsilon$.

In addition, thanks to the fact that $g(x) = \bar{G}(x, x, w_0)$, for any $x \in \mathrm{dom}\partial F$ satisfying $F(x_0) < F(x) < F(x_0) + c_0$, we have

$$
(4.3) \qquad \bar{F}(x_0, x_0, w_0) < \bar{F}(x, x, w_0) < \bar{F}(x_0, x_0, w_0) + c_0.
$$

On the other hand, for x such that $\|x - x_0\| \le \frac{1}{2}\epsilon$, we have $\|(x, x, w_0) - (x_0, x_0, w_0)\| \le \epsilon$. Using this and (4.3), for $x \in \mathrm{dom}\partial F$ satisfying $\|x - x_0\| \le \frac{1}{2}\epsilon$ and $F(x_0) < F(x) < F(x_0) + c_0$, we have

$$
\mathrm{dist}(0, \partial F(x)) \overset{(a)}{\ge} \mathrm{dist}(0, \partial \bar{F}(x, x, w_0)) \overset{(b)}{\ge} a(\bar{F}(x, x, w_0) - \bar{F}(x_0, x_0, w_0))^\alpha
$$

$$\overset{(c)}{=} a(F(x) - F(x_0))^\alpha,$$

where (a) follows from (4.1), (b) uses (4.2) and (c) holds thanks to $g(x) = \bar{G}(x, x, w_0)$. This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

4.1. KL exponent for a concrete model. In this subsection, we study the KL exponent of F in (1.1) with additional assumptions on the functions involved. Specifically, we consider the following multiply constrained optimization problem:

$$(4.4) \qquad\qquad \min_{x \in \mathbb{R}^n} P_1(x) + \delta_{g(\cdot) \leq 0}(x),$$

where P_1 is convex continuous, the function $g(x) = (l_1(A_1 x), \ldots, l_m(A_m x))$ with each $A_i \in \mathbb{R}^{q_i \times n}$ and $l_i : \mathbb{R}^{q_i} \to \mathbb{R}$ being *strictly* convex, and $\{x : g(x) \leq 0\} \neq \emptyset$. Clearly, (4.4) is a special case of (1.1) with $f = P_2 = 0$, $g_i(x) = l_i(A_i x)$, for $i = 1, \ldots, m$ and

$$(4.5) \qquad\qquad F(x) = P_1(x) + \delta_{g(\cdot) \leq 0}(x) = P_1(x) + \sum_{i=1}^m \delta_{l_i(\cdot) \leq 0}(A_i x).$$

We will derive rules to deduce the KL exponent of F in (4.5) from its Lagrangian. Similar rules were introduced in [42] and [61], which studied the KL exponent of F in (4.5) respectively when $m = 1$ and when the constraint set is defined by equality constraints, under suitable assumptions. Here, we look at (4.5) that involves multiple inequality constraints.

THEOREM 4.2 (**KL exponent of** (4.5) **from its Lagrangian**). *Let F be as in* (4.5) *and $\bar{x} \in \operatorname{Arg\,min} F$. Suppose the following conditions hold:*
  (i) *There exists a Lagrange multiplier $\bar{\lambda} \in \mathbb{R}_+^m$ for* (4.4) *and $x \mapsto P_1(x) + \langle \bar{\lambda}, g(x) \rangle$ is a KL function with exponent $\alpha \in (0, 1)$.*
  (ii) *The strict complementarity condition holds at $(\bar{x}, \bar{\lambda})$, i.e., for every i satisfying $\bar{\lambda}_i = 0$, it holds that $l_i(A_i \bar{x}) < 0$.*
Then F satisfies the KL property with exponent α at $\bar{x}$.

Proof. Let $F_{\bar{\lambda}}(x) := P_1(x) + \langle \bar{\lambda}, g(x) \rangle$. By the definition of Lagrange multiplier, we have

$$(4.6) \qquad\qquad F(\bar{x}) = \inf F = P_1(\bar{x}) = \inf F_{\bar{\lambda}} \leq F_{\bar{\lambda}}(\bar{x}) \leq F(\bar{x}),$$

where the second inequality holds because $g(\bar{x}) \leq 0$ and $\bar{\lambda} \in \mathbb{R}_+^m$. On the other hand, thanks to (ii), it holds that $\{i : \bar{\lambda}_i > 0\} = I(\bar{x})$. This together with [52, Theorem 28.1] gives

$$(4.7) \quad \bar{x} \in \operatorname{Arg\,min} F = \bigcap_{i \in I(\bar{x})} \{x : l_i(A_i x) = 0\} \cap \bigcap_{i \notin I(\bar{x})} \{x : l_i(A_i x) \leq 0\} \cap \operatorname{Arg\,min} F_{\bar{\lambda}}.$$

Since l_i is strictly convex and $\bar{\lambda}_i > 0$ for $i \in I(\bar{x})$, we see that $A_i x$ is constant over $\operatorname{Arg\,min} F_{\bar{\lambda}}$ for each $i \in I(\bar{x})$. This together with the fact that $l_i(A_i \bar{x}) = 0$ for $i \in I(\bar{x})$ and (4.7) implies that

$$(4.8) \qquad\qquad \bar{x} \in \operatorname{Arg\,min} F = \bigcap_{i \notin I(\bar{x})} \{x : l_i(A_i x) \leq 0\} \cap \operatorname{Arg\,min} F_{\bar{\lambda}}.$$

Next, since $l_i(A_i \bar{x}) < 0$ for each $i \notin I(\bar{x})$, there exists $\epsilon_0 > 0$ such that

$$l_i(A_i x) < 0, \ \forall x \in B(\bar{x}, \epsilon_0), \ \forall i \notin I(\bar{x}).$$

This together with (4.8) implies that

$$\text{(4.9)} \qquad \bar{x} \in \operatorname{Arg\,min} F \cap B(\bar{x}, \epsilon_0) = \operatorname{Arg\,min} F_{\bar{\lambda}} \cap B(\bar{x}, \epsilon_0).$$

Now, using (i) and [12, Theorem 5(i)] together with the fact that $\bar{x} \in \operatorname{Arg\,min} F_{\bar{\lambda}}$, we see that there exist $\bar{a} > 0$, $\bar{c} > 0$ and $0 < \epsilon < \epsilon_0$ such that

$$\text{(4.10)} \qquad \operatorname{dist}(x, \operatorname{Arg\,min} F_{\bar{\lambda}}) \leq \bar{c}(F_{\bar{\lambda}}(x) - F_{\bar{\lambda}}(\bar{x}))^{1-\alpha}$$

whenever $\|x - \bar{x}\| \leq \epsilon$ and $F_{\bar{\lambda}}(\bar{x}) \leq F_{\bar{\lambda}}(x) < F_{\bar{\lambda}}(\bar{x}) + \bar{a}$. Note that for any x satisfying $F(\bar{x}) < F(x) < F(\bar{x}) + \bar{a}$, we have $l_i(A_i x) \leq 0$ for each i and

$$\text{(4.11)} \qquad F(\bar{x}) = F_{\bar{\lambda}}(\bar{x}) \leq F_{\bar{\lambda}}(x) \leq F(x) < F(\bar{x}) + \bar{a} = F_{\bar{\lambda}}(\bar{x}) + \bar{a},$$

where the first and the last equalities follow from (4.6) and the second inequality holds because $\bar{\lambda}_i \geq 0$ and $l_i(A_i x) \leq 0$ for each $i = 1, \ldots, m$. Therefore, for any x satisfying $F(\bar{x}) < F(x) < F(\bar{x}) + \bar{a}$ and $\|x - \bar{x}\| \leq \epsilon$, we have

$$\operatorname{dist}(x, \operatorname{Arg\,min} F) \leq \operatorname{dist}(x, \operatorname{Arg\,min} F \cap B(\bar{x}, \epsilon_0)) \overset{\text{(a)}}{=} \operatorname{dist}(x, \operatorname{Arg\,min} F_{\bar{\lambda}} \cap B(\bar{x}, \epsilon_0))$$

$$\overset{\text{(b)}}{\leq} 4 \max \left\{ \operatorname{dist}(x, \operatorname{Arg\,min} F_{\bar{\lambda}}), \operatorname{dist}(x, B(\bar{x}, \epsilon_0)) \right\} \overset{\text{(c)}}{=} 4\operatorname{dist}(x, \operatorname{Arg\,min} F_{\bar{\lambda}})$$

$$\overset{\text{(d)}}{\leq} 4\bar{c}(F_{\bar{\lambda}}(x) - F_{\bar{\lambda}}(\bar{x}))^{1-\alpha} \leq 4\bar{c}(F(x) - F(\bar{x}))^{1-\alpha},$$

where (a) follows from (4.9), (b) follows from [40, Lemma 4.10], (c) holds because $\epsilon < \epsilon_0$, (d) follows from (4.10) and (4.11) and the last inequality holds because of (4.6) (so that $F_{\bar{\lambda}}(\bar{x}) = F(\bar{x})$), $l_i(A_i x) \leq 0$ for each i and $\bar{\lambda} \in \mathbb{R}_+^m$. The desired conclusion now follows immediately from this and [12, Theorem 5(ii)]. □

Now, we give a corollary that deals with (4.4) with $m = 1$. This result is different from [42, Theorem 3.5] because, here, it is the constraint function that is a composition of strictly convex function and a linear map, but not the objective function.

COROLLARY 4.3. *Let F be defined as in* (4.5) *with $m = 1$. Suppose the following conditions hold:*
  (i) *It holds that* $\inf P_1 < \inf F$.
  (ii) *There exists a Lagrange multiplier*[6] *$\bar{\lambda} \geq 0$ for* (4.4) *and $x \mapsto P_1(x) + \bar{\lambda} l_1(A_1 x)$ is a KL function with exponent $\alpha \in (0, 1)$.*
Then F is KL function with exponent α.

Proof. Let $F_{\bar{\lambda}}(x) := P_1(x) + \bar{\lambda} l_1(A_1 x)$. In view of [42, Lemma 2.1] and the convexity of F, it suffices to show that F has KL property at every point in $\{x : 0 \in \partial F(x)\} = \operatorname{Arg\,min} F$ with exponent α. Fix any $\bar{x}$ with $0 \in \partial F(\bar{x})$. Then one can see from condition (i) and the definition of Lagrange multiplier that $\bar{\lambda} > 0$ and thus $l_1(A_1 \bar{x}) = 0$. Therefore, Assumption (ii) of Theorem 4.2 is satisfied. This together with (ii) and Theorem 4.2 shows that F satisfies the KL property at $\bar{x}$ with exponent α. □

Remark 4.4. When $P_1(\cdot) = \| \cdot \|_1$ in (4.4), we deduce from [42, Corollary 5.1] and Corollary 4.3 that the KL exponent of F in (4.5) is $\frac{1}{2}$ if $m = 1$ and l_1 takes one of the following forms with $b \in \mathbb{R}^q$ and $\delta > 0$ chosen so that the Slater condition holds and the origin is not feasible:

---

[6]Following [52, Page 274], we say that $\bar{\lambda}$ is a Lagrange multiplier for (4.4) with $m = 1$ if $\bar{\lambda} \geq 0$ and $\inf\limits_{x \in \mathbb{R}^n} \{P_1(x) + \bar{\lambda} g_1(x)\} = \inf\limits_{x \in \mathbb{R}^n} \{P_1(x) + \delta_{g_1(\cdot) \leq 0}(x)\} > -\infty$.

(i) (Basis pursuit denoising [19]) $l_1(z) = \frac{1}{2}\|z - b\|^2 - \delta$.

(ii) (Logistic loss [30,33]) $l_1(z) = \sum_{i=1}^q \log(1 + \exp(b_i z_i)) - \delta$ for some $b \in \{-1, 1\}^q$.

(iii) (Poisson loss [27,35,62]) $l_1(z) = \sum_{i=1}^q (-b_i z_i + \exp(z_i)) - \delta$ for some $b \in \mathbb{R}^q$.

5. Applications in compressed sensing. In this section, we consider applications of (1.1) and discuss how the various assumptions required in our analysis of SCP_{ls} can be verified. We focus on the problem of compressed sensing, which attempts to reconstruct sparse signals from possibly noisy low-dimensional measurements; see [21] for a recent review. We specifically look at the following model:

$$(5.1) \qquad \begin{array}{ll} \min_x & \|x\|_1 - \mu\|x\| \\ \text{s.t.} & \ell(Ax - b) \leq \delta, \end{array}$$

where $\mu \in [0, 1]$, $A \in \mathbb{R}^{q \times n}$ has *full row rank*, $b \in \mathbb{R}^q$, $\ell : \mathbb{R}^q \to \mathbb{R}_+$ is an analytic function whose gradient is Lipschitz continuous with modulus L_ℓ and satisfies $\ell(0) = 0$, and $\delta \in (0, \ell(-b))$. The ℓ in (5.1) is typically chosen according to different types of noise. We will look at two specific choices in Section 5.1 and Section 5.2, respectively.

Problem (5.1) is a special case of (1.1) with $f = 0$, $P_1(x) = \|x\|_1$, $P_2(x) = \mu\|x\|$ and $g(x) = \ell(Ax - b) - \delta$.[7] Then the F from (1.1) corresponding to (5.1) is

$$(5.2) \qquad F(x) = \|x\|_1 - \mu\|x\| + \delta_{\ell(A\cdot - b) \leq \delta}(x),$$

and the $\bar{F}$ from (1.3) corresponding to (5.1) is

$$(5.3) \qquad \bar{F}(x, y, w) = \|x\|_1 - \mu\|x\| + \delta_{\bar{G}(\cdot) \leq 0}(x, y, w)$$

with

$$(5.4) \qquad \bar{G}(x, y, w) = \ell(Ay - b) + \langle A^T \nabla \ell(Ay - b), x - y \rangle + \frac{w}{2}\|x - y\|^2 - \delta.$$

Our next theorem concerns the KL conditions needed in Theorems 3.8 and 3.12.

THEOREM 5.1. *Let F and $\bar{F}$ be defined as in (5.2) and (5.3), respectively, and let $\Xi \subseteq \mathrm{dom}\,\partial F$ and $\Upsilon \subseteq \mathrm{dom}\,\partial\bar{F}$ be compact sets. Then there exists $\alpha \in [0, 1)$ so that F (resp., $\bar{F}$) satisfies the KL property with exponent α at every point in Ξ (resp., in Υ).*

Proof. Let $\mathfrak{D}_0 := \{x : \ell(Ax - b) \leq \delta\}$ and $\mathfrak{D}_1 = \{(x, y, w) : \bar{G}(x, y, w) \leq 0\}$, where $\bar{G}$ is as in (5.4). Since ℓ and $\bar{G}$ are analytic, we have that $\mathfrak{D}_0$ and $\mathfrak{D}_1$ are semianalytic; see [26, Page 596] for the definition.

On the other hand, since $x \mapsto \|x\|_1 - \mu\|x\|$ is semialgebraic, it holds that $\mathfrak{F}_0 := \{(x, z) : z = \|x\|_1 - \mu\|x\|\}$ and $\mathfrak{F}_1 := \{(x, y, w, z) : z = \|x\|_1 - \mu\|x\|\}$ are subanalytic (see [26, Page 597(p2)] for the subanalyticity of $\mathfrak{F}_1$). Therefore,

$$\mathrm{gph}(F) = \mathfrak{F}_0 \cap (\mathfrak{D}_0 \times \mathbb{R}) \quad \text{and} \quad \mathrm{gph}(\bar{F}) = \mathfrak{F}_1 \cap (\mathfrak{D}_1 \times \mathbb{R})$$

are subanalytic, thanks to [26, Page 597(p1)&(p2)]. Also, the functions F and $\bar{F}$ have closed domains and are continuous on their respective domains. Thus, the desired conclusion follows from [11, Theorem 3.1] and a standard compactness argument as in the proof of [2, Lemma 1]. □

We next focus on two common choices of ℓ in (5.1): $\ell(\cdot) = \frac{1}{2}\|\cdot\|^2$ (for Gaussian noise [8]) and $\ell(\cdot) = \|\cdot\|_{LL_2,\gamma}$ being the Lorentzian norm (for Cauchy noise [20]) for some $\gamma > 0$. We will discuss how to verify the other assumptions necessary for the applications of Theorem 3.8 or Theorem 3.12 to (5.1) with these two choices of ℓ.

---

[7] Note that $\{x : g(x) \leq 0\} \neq \emptyset$ because A has full row rank and $\ell(0) = 0 < \delta$.

5.1. When $\ell(\cdot) = \frac{1}{2}\|\cdot\|^2$. In this case, the model (5.1) becomes

$$(5.5) \qquad \begin{aligned} \min_{x} & \quad \|x\|_1 - \mu\|x\| \\ \mathrm{s.t.} & \quad \tfrac{1}{2}\|Ax - b\|^2 \leq \delta, \end{aligned}$$

and the corresponding F in (1.1) becomes:

$$(5.6) \qquad F(x) = \|x\|_1 - \mu\|x\| + \delta_{g(\cdot) \leq 0}(x),$$

with $f = 0$, $P_1(x) = \|x\|_1$, $P_2(x) = \mu\|x\|$ and $g(x) = \frac{1}{2}\|Ax - b\|^2 - \delta$ for A, b, δ and μ as in (5.1). Then, for (5.5), P_1 and P_2 are convex continuous, and Assumption 1.1(i) and (ii) and Assumption 3.4 are satisfied. Moreover, A having full row rank and $\delta \in (0, \frac{1}{2}\|b\|^2)$ imply that Slater condition holds for (5.5). Hence, it follows that $\{x : g(x) \leq 0\} \neq \emptyset$ and Assumption 2.3 holds. In addition, this P_2 satisfies Assumption 3.6 since its only possible point of nondifferentiability (the origin) is not feasible thanks to the fact that $\delta < \frac{1}{2}\|b\|^2$. Furthermore, the required KL conditions follow from Theorem 5.1.[8] In order to apply Theorem 3.8 (or Theorem 3.12), we now demonstrate how conditions can be imposed so that Assumption 1.1(iii) (level-boundedness) is satisfied.

PROPOSITION 5.2. *Let F be defined as in (5.6). The following statements hold:*
(i) *If $\mu \in [0, 1)$, then F is level-bounded.*
(ii) *If $\mu = 1$ and A does not have zero columns, then F is level-bounded.*

Proof. Note first that if $0 \leq \mu < 1$, then $x \mapsto \|x\|_1 - \mu\|x\|$ is level-bounded and hence (i) holds trivially. We next focus on the case where $\mu = 1$.

Suppose to the contrary that there exists σ and $\{x^t\} \subseteq \{x : F(x) \leq \sigma\}$ such that $\|x^t\| \to \infty$. By passing to a further subsequence if necessary, we may assume that there exists d with $\|d\| = 1$ and $\lim_{t \to \infty} \frac{x^t}{\|x^t\|} = d$. Since $\frac{1}{2}\|Ax^t - b\|^2 \leq \delta$ thanks to $F(x^t) \leq \sigma$ for each t, we have

$$(5.7) \qquad \frac{1}{2}\|Ad\|^2 = \lim_{t \to \infty} \frac{1}{2} \frac{\|Ax^t - b\|^2}{\|x^t\|^2} \leq \lim_{t \to \infty} \frac{\delta}{\|x^t\|^2} = 0.$$

On the other hand, since $F(x^t) \leq \sigma$, it holds that

$$0 \leq \|x^t\|_1 - \|x^t\| \leq \sigma \implies 0 \leq \lim_{t \to \infty} \frac{\|x^t\|_1 - \|x^t\|}{\|x^t\|} = \|d\|_1 - 1 \leq 0.$$

This together with $\|d\| = 1$ implies that exactly one coordinate of d is nonzero. Since A does not have zero columns, we obtain that $\|Ad\| \neq 0$, which contradicts (5.7). Thus, the statement in (ii) holds. □

Therefore, if the assumptions in the above proposition hold, one can apply Theorem 3.8 or Theorem 3.12 to deducing the convergence rate of the sequence generated by SCP$_{ls}$ when applied to solving (5.5). When $\mu = 0$ in (5.5), since we assumed $\delta \in (0, \frac{1}{2}\|b\|^2)$ and A has full row rank, we know from Remark 4.4 that $x \mapsto \|x\|_1 + \delta_{\frac{1}{2}\|A(\cdot)-b\|^2 \leq \delta}(x)$ is a KL function with exponent $\frac{1}{2}$. Consequently, the

---

[8]Specifically, if $\mu = 0$, then F is convex and level-bounded, and the set of stationary points (minimizers) is compact. We can then deduce from Theorem 5.1 that F is a KL function with some exponent $\alpha \in [0, 1)$. On the other hand, the KL property required in the nonconvex case (see Theorem 3.8) follows directly from Theorem 5.1.

sequence $\{x^t\}$ generated by SCP_{ls} for (5.5) converges locally linearly. When $\mu \in (0, 1]$, although no explicit KL exponent is known for the corresponding $\bar{F}$, we still observe in our numerical experiments below that the sequence $\{x^t\}$ generated by SCP_{ls} for (5.5) appears to converge linearly.

5.2. When ℓ is the Lorentzian norm. Recall that, given $\gamma > 0$, the Lorentzian norm of a vector $y \in \mathrm{I\!R}^q$ is defined as

$$\|y\|_{LL_2,\gamma} := \sum_{i=1}^q \log\left(1 + \frac{y_i^2}{\gamma^2}\right).$$

In this case, the model (5.1) becomes

$$(5.8) \qquad \begin{aligned} &\min_x \|x\|_1 - \mu\|x\| \\ &\text{s.t. } \|Ax - b\|_{LL_2,\gamma} \le \delta, \end{aligned}$$

and the corresponding F in (1.1) now takes the following form:

$$(5.9) \qquad F(x) = \|x\|_1 - \mu\|x\| + \delta_{g(\cdot)\le 0}(x),$$

with $f = 0$, $P_1(x) = \|x\|_1$, $P_2(x) = \mu\|x\|$ and $g(x) = \|Ax - b\|_{LL_2,\gamma} - \delta$ for A, b, δ and μ defined as in (5.1). One can show that the mapping $z \mapsto \|z\|_{LL_2,\gamma} - \delta$ has Lipschitz gradient with modulus $L_\ell = \frac{2}{\gamma^2}$ and is twice continuously differentiable. From these one can readily see that P_1 and P_2 are convex continuous, and Assumption 1.1(i) and (ii) and Assumption 3.4 are satisfied. Also, since A has full row rank and $\delta \in (0, \|b\|_{LL_2,\gamma})$, we see that $\{x : g(x) \le 0\} \ne \emptyset$. In addition, this P_2 satisfies Assumption 3.6 since its only possible point of nondifferentiability is not feasible, thanks to $\delta \in (0, \|b\|_{LL_2,\gamma})$. Furthermore, the required KL conditions follow from Theorem 5.1. In order to apply Theorem 3.8, we show below that Assumption 2.3 holds and impose conditions so that Assumption 1.1(iii) is satisfied.

PROPOSITION 5.3. *Let F be defined as in* (5.9). *The following statements hold:*
(i) *The MFCQ holds in the whole feasible set of* (5.8).
(ii) *If $\mu \in [0, 1)$, then F is level-bounded.*
(iii) *If $\mu = 1$ and A does not have zero columns, then F is level-bounded.*

Proof. For (i), using the definition of MFCQ, it suffices to show that for every feasible point x with $g(x) = 0$, it holds that $\nabla g(x) \ne 0$. Suppose to the contrary that there exists $\hat{x}$ such that $g(\hat{x}) = 0$ and

$$\nabla g(\hat{x}) = A^T \left(\frac{2(A\hat{x} - b)_1}{\gamma^2 + (A\hat{x} - b)_1^2}, \ldots, \frac{2(A\hat{x} - b)_q}{\gamma^2 + (A\hat{x} - b)_q^2}\right)^T = 0.$$

Since A is surjective, we deduce that $\left(\frac{2(A\hat{x}-b)_1}{\gamma^2+(A\hat{x}-b)_1^2}, \ldots, \frac{2(A\hat{x}-b)_q}{\gamma^2+(A\hat{x}-b)_q^2}\right) = 0$. This shows that $A\hat{x} - b = 0$ and thus $g(\hat{x}) = \|A\hat{x} - b\|_{LL_2,\gamma} - \delta = -\delta \ne 0$, a contradiction. Therefore, the MFCQ holds in the whole feasible set of (5.8).

The assertion in (ii) holds trivially. We now prove (iii). Suppose to the contrary that there exist σ and $\{x^t\} \subseteq \{x : F(x) \le \sigma\}$ such that $\|x^t\| \to \infty$. By passing to a further subsequence if necessary, we may assume that there exists d with $\|d\| = 1$ and $d = \lim_{t\to\infty} \frac{x^t}{\|x^t\|}$. Since $\ell(Ax^t - b) \le \delta$ thanks to $F(x^t) \le \sigma$ for each t, and the Lorentzian norm is level-bounded, we see that there exists ξ such that $\|Ax^t - b\| \le \xi$ for all t. The rest of the proof is then the same as that of Proposition 5.2(ii). $\square$

Therefore, if the assumptions in the above proposition hold, one can apply Theorem 3.8 to deducing the convergence rate of the sequence $\{x^t\}$ generated by SCP$_{ls}$ when applied to solving (5.8). Although no explicit KL exponent is known for the corresponding $\bar{F}$, in our numerical experiments below, we observe empirically that the sequence $\{x^t\}$ generated by SCP$_{ls}$ for (5.8) appears to converge linearly.

5.3. Numerical experiments. In this subsection, we perform numerical experiments to illustrate the convergence results of SCP$_{ls}$ established in Section 3. We apply SCP$_{ls}$ to (5.1) with ℓ being either $\frac{1}{2}\|\cdot\|^2$ (as in (5.5)) or the Lorentzian norm (as in (5.8)). We also consider the SCP in [46] in our experiments below.

Algorithms and their parameters. We consider the following algorithms:
 (i) **SCP$_{ls}$**: We solve the corresponding subproblem (2.4) through a root-finding scheme outlined in Appendix A. Moreover, we let $\tau = 2$, $c = 10^{-4}$, $\underline{\mathrm{L}} = 10^{-8}$, $\bar{\mathrm{L}} = 10^8$. For $t = 0$, we choose $L_f^{t,0} = 1$ and $L_g^{t,0} = 1$. For $t \geq 1$, we choose:

$$L_f^{t,0} = 1, \; L_g^{t,0} = \begin{cases} \max\left\{10^{-8}, \min\left\{\frac{\langle \Delta x, \Delta g \rangle}{\|\Delta x\|^2}, 10^8\right\}\right\} & \text{if } \langle \Delta x, \Delta g \rangle \geq 10^{-12}, \\ \max\left\{10^{-8}, \min\left\{L_g^{t-1}/\tau, 10^8\right\}\right\} & \text{else,} \end{cases}$$

where $\Delta x = x^t - x^{t-1}$ and $\Delta g = \nabla g(x^t) - \nabla g(x^{t-1})$. We initialize SCP$_{ls}$ at $A^\dagger b$ and terminate it when $\|x^{t+1} - x^t\| < 10^{-8} \max\{1, \|x^{t+1}\|\}$.
 (ii) **SCP**: This was proposed in [46]. The subproblem of SCP is solved using a root-finding scheme outlined in Appendix A. We initialize SCP at $A^\dagger b$ and terminate it when $\|x^{t+1} - x^t\| < 10^{-8} \max\{1, \|x^{t+1}\|\}$.

Numerical results. All codes are written in Matlab, and the experiments are performed in Matlab 2019b on a 64-bit PC with an Intel(R) Core(TM) i7-4790 CPU (3.60GHz) and 32GB of RAM.

For both models (5.5) and (5.8), we consider either $\mu = 0$ or 1. In our tests, we let $q = 720i$ and $n = 2560i$ with $i = 5$. We generate an $A \in \mathbb{R}^{q \times n}$ with i.i.d standard Gaussian entries, and then normalize this matrix so that each column of A has unit norm. Then we choose a subset T of size $s_0 = \lceil \frac{q}{9} \rceil$ uniformly at random from $\{1, 2, \ldots, n\}$ and an s_0-sparse vector x_{orig} having i.i.d. standard Gaussian entries on T is generated.

For (5.5), we let $b = Ax_{\mathrm{orig}} + 0.01 \cdot \hat{n}$ with $\hat{n} \in \mathbb{R}^q$ being a random vector with i.i.d. standard Gaussian entries. We then set the δ in (5.5) to be $\frac{1}{2}\sigma^2$ with $\sigma = 1.1\|0.01 \cdot \hat{n}\|$.

For (5.8), we let $b = Ax_{\mathrm{orig}} + 0.01 \cdot \bar{n}$ with $\bar{n}_i \sim \mathrm{Cauchy}(0,1)$, i.e., $\bar{n}_i := \tan(\pi(\widetilde{n}_i - 1/2))$ with $\widetilde{n} \in \mathbb{R}^m$ being a random vector with i.i.d. entries uniformly chosen in $[0,1]$. We set the δ in (5.8) to be $1.1\|0.01\bar{n}\|_{LL_2,\gamma}$ with $\gamma = 0.02$.

We compare the approximate solution obtained by SCP$_{ls}$ and the original sparse solution in Figures 1 and 2 to illustrate the recovery ability of SCP$_{ls}$. In Figures 3 and 4, we plot $\|x^t - x^{out}\|$ (in logarithmic scale) against the number of iterations, where x^t and x^{out} are respectively the tth iterate and the approximate solution obtained by the algorithm under study. As we can see, SCP$_{ls}$ always appears to converge linearly and is also faster than SCP.

Appendix A. Solving the subproblem of SCP$_{ls}$ with P_1 being the ℓ_1 norm, $P_2 = 0$ and $m = 1$. We discuss how the subproblem (2.4) that arises in our numerical tests when SCP$_{ls}$ is applied to (5.1) can be solved efficiently. Our approach is based on a root-finding strategy for solving the dual, which was also adopted in [54] for solving the subproblem that arises in the MBA variant there. Comparing with the subproblem considered in [54], our subproblem has an additional quadratic term, which slightly complicates the derivation and implementation.
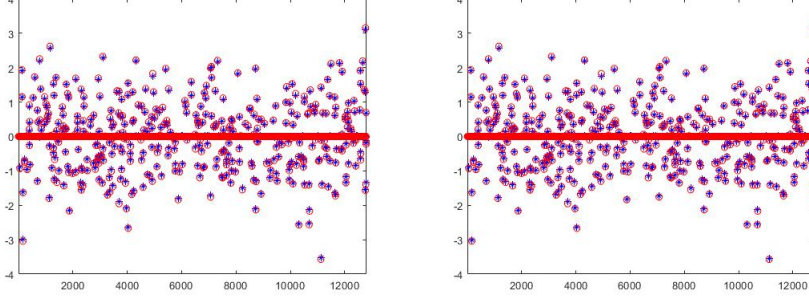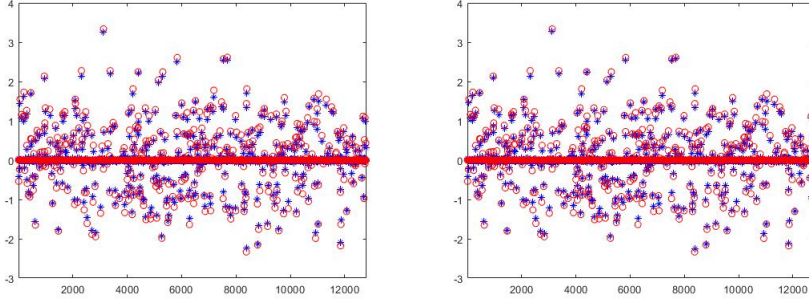
FIG. 1. *Recovery results by solving model (5.5) with $\mu = 0$ (left) and $\mu = 1$ (right) via SCP_{ls}. The approximate solution obtained by SCP_{ls} is marked by asterisk, and x_{orig} is marked by circle.*

FIG. 2. *Recovery results by solving model (5.8) with $\mu = 0$ (left) and $\mu = 1$ (right) via SCP_{ls}. The approximate solution obtained by SCP_{ls} is marked by asterisk, and x_{orig} is marked by circle.*

At the tth iteration, the corresponding subproblem (2.4) that arises when SCP_{ls} is applied to (5.1) takes the following form:

$$(A.1) \qquad \begin{aligned} \min_{x} \quad & \|x\|_1 + \frac{\alpha}{2}\|x - y\|^2 \\ \mathrm{s.t.} \quad & \|x - s\|^2 \le r, \end{aligned}$$

where $y, s \in \mathbb{R}^n$, $\alpha > 0$ and $r > 0$.[9]

Recall that the Lagrangian function for (A.1) is given by

$$\widetilde{L}(x, \lambda) = \|x\|_1 + \frac{\alpha}{2}\|x - y\|^2 + \lambda(\|x - s\|^2 - r).$$

Using [52, Corollary 28.2.1, Theorem 28.3], we know that there exists (x^*, λ^*) with $\lambda^* \ge 0$ such that x^* is optimal for (A.1) and

$$\min_{x \in \mathbb{R}^n} \widetilde{L}(x, \lambda^*) = \min_{x \in \mathbb{R}^n} \|x\|_1 + \frac{\alpha}{2}\|x - y\|^2 + \delta_{\|(\cdot) - s\|^2 \le r}(x).$$

If $\lambda^* = 0$, then the solution $\check{x}$ of $\min_{x \in \mathbb{R}^n} \|x\|_1 + \frac{\alpha}{2}\|x - y\|^2$ lies in $\{x : \|x - s\|^2 \le r\}$ and $\check{x}$ solves (A.1). Moreover, $\check{x}$ is given explicitly as $\mathrm{sign}(y) \circ \max\{|y| - \frac{1}{\alpha}, 0\}$, where $\circ$

---

[9]The fact that $r > 0$ follows from Lemma 2.4(iii).

FIG. 3. *Plot of* $\|x^t - x^{\mathrm{out}}\|$ *(in log scale) for model* (5.5) *with* $\mu = 0$ *(left) and* $\mu = 1$ *(right).* *The number in the parenthesis is the CPU time taken.*
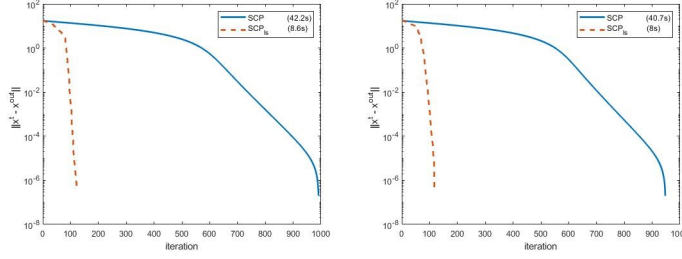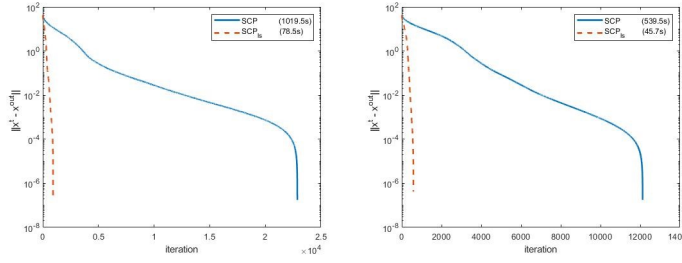

FIG. 4. *Plot of* $\|x^t - x^{\mathrm{out}}\|$ *(in log scale) for model* (5.8) *with* $\mu = 0$ *(left) and* $\mu = 1$ *(right).* *The number in the parenthesis is the CPU time taken.*

denotes the entrywise product, and the sign function, absolute value and maximum are taken componentwise.

If $\lambda^* > 0$, using [52, Theorem 28.3], we obtain that

$$(A.2) \qquad 0 \in \partial \|x^*\|_1 + \alpha(x^* - y) + 2\lambda^*(x^* - s) \text{ and } \|x^* - s\|^2 = r.$$

Using the first relation in (A.2), we have

$$(A.3) \qquad x^* = \mathrm{Prox}_{\frac{1}{\alpha + 2\lambda^*} \|\cdot\|_1} \left(\frac{\alpha}{\alpha + 2\lambda^*} y + \frac{2\lambda^*}{\alpha + 2\lambda^*} s \right),$$

where $\mathrm{Prox}_h(u) := \arg\min_{v \in \mathbb{R}^n} \left\{ h(v) + \frac{1}{2}\|u - v\|^2 \right\}$ for a proper closed convex function h.

Plugging this into the second relation in (A.2), we see that λ^* can be obtained by solving the following one-dimensional nonsmooth equation and the solution x^* can then be recovered via (A.3):

$$\left\| \mathrm{Prox}_{\frac{1}{\alpha + 2\lambda^*} \|\cdot\|_1} \left(\frac{\alpha}{\alpha + 2\lambda^*} y + \frac{2\lambda^*}{\alpha + 2\lambda^*} s \right) - s \right\|^2 = r.$$

Upon the transformation $t^* = \frac{\alpha}{\alpha + 2\lambda^*}$, the above equation becomes piecewise linear quadratic and can be solved efficiently by a standard root-finding procedure.

In passing, we note that a solution procedure for the subproblem that arises when SCP is applied to (5.1) can be derived similarly, where the subproblem takes the form

$$\min_x \quad \|x\|_1 - \langle \xi, x \rangle$$
$$\text{s.t.} \quad \|x - s\|^2 \le r,$$

for some ξ, $s \in \mathbb{R}^n$ and $r > 0$. We omit the details for brevity.

REFERENCES

[1] W. van Ackooij and W. de Oliveira. Non-smooth DC-constrained optimization: constraint qualification and minimizing methodologies. *Optim. Method. Softw.* 34:890–920, 2019.
[2] H. Attouch and J. Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math. Program.* 116:5–16, 2009.
[3] H. Attouch, J. Bolte, P. Redont and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.* 35:438–457, 2010.
[4] H. Attouch, J. Bolte and B. F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. *Math. Program.* 137:91–129, 2013.
[5] F. J. Aragón Artacho and P. T. Vuong. The boosted difference of convex functions algorithm for nonsmooth functions. *SIAM J. Optim.* 30:980–1006, 2020.
[6] A. Auslender, R. Shefi and M. Teboulle. A moving balls approximation method for a class of smooth constrained minimization problems. *SIAM J. Optim.* 20:3232–3259, 2010.
[7] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Ana.* 8:141–148, 1988.
[8] E. van den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.* 31:890–912, 2009.
[9] E. G. Birgin, J. M. Martínez and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM. J. Optim.* 10:1196–1211, 2000.
[10] J. Bolte, Z. Chen and E. Pauwels. The multiproximal linearization method for convex composite problems. *Math. Program.* https://doi.org/10.1007/s10107-019-01382-3.
[11] J. Bolte, A. Daniilidis and A. Lewis. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.* 17:1205–1223, 2007.
[12] J. Bolte, T. P. Nguyen, J. Peypouquet and B. W. Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Math. Program.* 165:471–507, 2017.
[13] J. Bolte and E. Pauwels. Majorization-minimization procedures and convergence of SQP methods for semi-algebraic and tame programs. *Math. Oper. Res.* 41:442–465, 2016.
[14] J. Bolte, S. Sabach and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* 146:459–494, 2014.
[15] J. Borwein and A. Lewis. *Convex Analysis and Nonlinear Optimization.* 2nd edition, Springer, 2006.
[16] J. Borwein, G. Li and L. Yao. Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets. *SIAM J. Optim.* 24:498–527, 2014.
[17] J. M. Borwein and Q. J. Zhu. *Techniques of Variational Analysis.* Springer, 2005.
[18] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, Cambridge, UK, 2004.
[19] E. J. Candés. The restricted isometry property and its implications for compressed sensing. *C. R. Math.* 346:589–592, 2008.
[20] R. E. Carrillo, K. E. Barner and T. C. Aysal. Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise. *IEEE J. Sel. Topic Signal Process.* 4:392–408, 2010.
[21] R. E. Carrillo, A. B. Ramirez, G. R. Arce, K. E. Barner and B. M. Sadler. Robust compressive sensing of sparse signals: a review. *EURASIP J. Adv. Signal Process.* 108, 2016.
[22] F. H. Clarke. *Optimization and Nonsmooth Analysis.* SIAM, 1983.
[23] D. Drusvyatskiy, A. D. Ioffe and A. S. Lewis. Nonsmooth optimization using Taylor-like models: error bounds, convergence, and termination criteria. *Math. Program.* https://doi.org/10.1007/s10107-019-01432-w.
[24] D. Drusvyatskiy and C. Paquette. Efficiency of minimizing compositions of convex functions and smooth maps. *Math. Program.* 178:503–558, 2019.
[25] D. Drusvyatskiy and A. S. Lewis. Error bounds, quadratic growth, and linear convergence of proximal methods. *Math. Oper. Res.* 43:919–948, 2018.
[26] F. Facchinei and J.-S. Pang. *Finite-dimensional Variational Inequalities and Complementarity Problems* Part I. Springer, New York, 2003.
[27] E. L. Frome. The analysis of rates using Poisson regression models. *Biometrics.* 39:665–674, 1983.

[28] P. Gong, C. Zhang, Z. Lu, J. Huang and J. Ye. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In: ICML, 2013.

[29] L. J. Hong, Y. Yang and L. Zhang. Sequential convex approximations to joint chance constrained programs: a Monte Carlo approach. *Oper. Res-ger.* 59:617–630, 2011.

[30] Jr. D. W. Hosmer, S. Lemeshow and R. X. Sturdivant. *Applied Logistic Regression.* John Wiley & Sons, 3rd edition, 2013.

[31] A. N. Iuesm. On the convergence properties of the projected gradient method for convex optimization. *Comput. Appl. Math.* 22:37–52, 2003.

[32] K. Joki, A. M. Bagirov, N. Karmitsa, M. M. Mäkelä and S. Taheri. Double bundle method for finding Clarke stationary points in nonsmooth DC programming. *SIAM J. Optim.* 28:1892–1919, 2018.

[33] D. G. Kleinbaum and M. Klein. *Logistic Regression: a Self-Learning Text*, 2nd edition, Springer-Verlag, New York, 2002.

[34] K. Kurdyka. On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier* 48:769–783, 1998.

[35] D. Z. Lambert. Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics.* 34:1–14, 1992.

[36] H. A. Le Thi and P. D. Tao. DC programming and DCA: thirty years of developments. *Math. Program.* 169:5–68, 2018.

[37] H. A. Le Thi, P. D. Tao and H. V. Ngai. Exact penalty and error bounds in DC programming. *J. Glob. Optim.* 52:509–535, 2012.

[38] H. A. Le Thi, H. V. Ngai, and P. D. Tao, DC programming and DCA for general DC programs. in *Advanced Computational Methods for Knowledge Engineering.* Proceedings of the 2nd Inter- national Conference on Computer Science, Applied Mathematics and Applications (ICCSAMA 2014), T. Van Do, H.A. Le Thi, and N.T. Nguyen, eds., Springer International Publishing, Warsaw, Poland, 15–35, 2004.

[39] A. S. Lewis and S. J. Wright. A proximal method for composite minimization. *Math. Program.* 158:501–546, 2016.

[40] C. Li, K. F. Ng and T. K. Pong. The SECQ, linear regularity, and the strong CHIP for an infinite system of closed convex sets in normed linear spaces. *SIAM J. Optim.* 18:643–665, 2007.

[41] G. Li and T. K. Pong. Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems. *Math. Program.* 159:371–401, 2016.

[42] G. Li and T. K. Pong. Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods. *Found. Comput. Math.* 18:1199–1232, 2018.

[43] T. Lipp and S. Boyd. Variations and extension of the convex-concave procedure. *Optim. Eng.* 17:263–287, 2016.

[44] T. Liu, T. K. Pong and A. Takeda. A refined convergence analysis of pDCA$_e$ with applications to simultaneous sparse recovery and outlier detection. *Comput. Optim. and Appl.* 73:69–100, 2019.

[45] S. Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. In *Les Équations aux Dérivées Partielles*, Éditions du Centre National de la Recherche Scientifique, Paris, 87–89, 1963.

[46] Z. Lu. Sequential convex programming methods for a class of structured nonlinear programming. Submitted on 10 Oct 2012. Available at https://arxiv.org/abs/1210.3039.

[47] P. Ochs. Unifying abstract inexact convergence theorems and block coordinate variable metric iPiano. *SIAM J. Optim.* 29:541–570, 2019.

[48] J.-S. Pang, M. Razaviyayn and A. Alvarado. Computing B-stationary points of nonsmooth DC programs. *Math. Oper. Res.* 42:95–118, 2017.

[49] E. Pauwels. The value function approach to convergence analysis in composite optimization. *Oper. Res. Lett.* 44:790-795, 2016.

[50] T. D. Quoc and M. Diehl. Sequential convex programming methods for solving nonlinear optimization problems with DC constraints. Available at https://arxiv.org/abs/1107.5841.

[51] S. J. Wright, R. D. Nowak and M. A. T. Figueiredo. Sparse reconstruction by separable approximation. *IEEE T. Signal Proces.* 57:2479–2493, 2009.

[52] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, 1970.

[53] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Springer, Berlin, 1997.

[54] R. Shefi and M. Teboulle. A dual method for minimizing a nonsmooth objective over one smooth inequality constraint. *Math. Program.* 159:137–164, 2016.

[55] A. S. Strekalovsky and I. M. Minarchenko. A local search method for optimization problem with d.c. inequality constraints. *Appl. Math. Model.* 58:229–244, 2018.

[56] Y. Sun, P. Babu and D. P. Palomar. Majorization-minimization algorithms in signal, communications, and machine learning processing. *IEEE T. Signal Process.* 65:794–816, 2017.

[57] P. D. Tao and L. T. H. An. Convex analysis approach to DC programming: theory, algorithms and applications. *Acta Math. Vietnamica.* 22:289–355, 1997.

[58] P. D. Tao and L. T. H. An. Recent advances in DC programming and DCA. In: Nguyen NT., Le-Thi H.A. (eds) *Transactions on Computational Intelligence XIII. Lecture Notes in Computer Science*, vol 8342, 1–37. Springer, Berlin, Heidelberg.

[59] B. Wen, X. Chen and T. K. Pong. A proximal difference-of-convex algorithm with extrapolation. *Comput. Optim. Appl.* 69:297–324, 2018.

[60] P. Yin, Y. Lou, Q. He and J. Xin. Minimization of ℓ_{1-2} for compressed sensing. *SIAM J. Sci. Comput.* 37:A536–A563, 2015.

[61] P. Yu, G. Li and T. K. Pong. Deducing Kurdyka-Łojasiewicz exponent via inf-projection. Submitted. Available at https://arxiv.org/abs/1902.03635.

[62] G. Zou. A modified Poisson regression approach to prospective studies with binary data. *Am. J. Epidemiol.* 159:702–706, 2004.