Sequential Convex Programming Methods for A Class of Structured Nonlinear Programming

Zhaosong Lu*

October 10, 2012 (Revised: June 17, 2022)

Abstract

In this paper we study a broad class of structured nonlinear programming (SNLP) problems. In particular, we first establish the first-order optimality conditions for them. Then we propose sequential convex programming (SCP) methods for solving them in which each iteration is obtained by solving a convex programming problem. Under some suitable assumptions, we establish that any accumulation point of the sequence generated by the methods is a KKT point of the SNLP problems. In addition, we propose a variant of the SCP method for SNLP in which nonmonotone scheme and "local" Lipschitz constants of the associated functions are used. A similar convergence result as mentioned above is established.

Key words: Sequential convex programming, structured nonlinear programming, first-order methods

1 Introduction

In this paper we consider a class of structured nonlinear programming problems in the form of

$$\begin{array}{ll} \min & f(x) + p(x) - u(x) \\ \text{s.t.} & g_i(x) + q_i(x) - v_i(x) \le 0, \quad i = 1, \ldots, m, \\ & x \in \mathcal{X}, \end{array} \tag{1}$$

where $\mathcal{X} \subseteq \Re^n$ is a nonempty closed convex set, f, g_i's are differentiable in $\mathcal{X}$, and p, u, q_i's, v_i's are convex (but not necessarily smooth) in $\mathcal{X}$.

Throughout this paper we make the following assumption.

*Department of Industrial and Systems Engineering, University of Minnesota, USA (email: zhaosong@umn.edu).

Assumption 1 *The gradients of f and g_i's are Lipschitz continuous in $\mathcal{X}$ with constants $L_f \geq 0$ and $L_{g_i} \geq 0$ for $i = 1, \ldots, m$, that is,*

$$\begin{aligned} \|\nabla f(x) - \nabla f(y)\| &\leq L_f \|x - y\|, \ \forall x, y \in \mathcal{X}, \\ \|\nabla g_i(x) - \nabla g_i(y)\| &\leq L_{g_i} \|x - y\|, \ \forall x, y \in \mathcal{X}, \ i = 1, \ldots, m. \end{aligned}$$

Some special cases of problem (1) have received considerable attention in the literature (see, for example, [18, 3, 15, 20, 22, 1, 12, 14]). In particular, Nesterov [15] and Beck and Teboulle [3] considered a special case of (1) with $m = 0$, $u \equiv 0$ and f being smooth convex with Lipschitz continuous gradient, and they proposed accelerated gradient methods for solving it. Tseng and Yun [20], Wright et al. [22], and Lu and Zhang [14] proposed efficient first-order methods for the similar problems as studied in [3, 15] with f being smooth but not necessarily convex. Recently, Auslender et al. [1] studied another special case of (1), where $\mathcal{X} = \Re^n$, $p \equiv 0$, $u \equiv 0$, $q_i \equiv 0$, $v_i \equiv 0$ for all i, and f and g_i's are smooth with Lipschitz continuous gradient. They proposed a gradient-based method so called the moving balls approximation (MBA) method for solving the problem. Very recently, Hong et al. [12] studied a sequential convex programming (SCP) approach for solving a special case of (1) with $m = 1$, $f \equiv 0$, $g_1 \equiv 0$, and p, u, q_1, u_1 being smooth convex functions in $\mathcal{X}$. In addition, a broad subclass of (1) with $m = 0$, $f \equiv 0$, known as DC (difference of convex functions) programming, was extensively studied and efficient first-order method was proposed for it (see, for example, [18, 13]).

Recently, a class of nonlinear programming models were widely used for finding a sparse approximate solution to a system or a function. They can also be viewed as special cases of (1). In particular, they are in the form of

$$\min_{x \in \Omega} l(x) + \sum_{i=1}^{n} h(|x_i|), \tag{2}$$

where l is a loss function, $\Omega \subseteq \Re^n$ is a nonempty closed convex set, and $h : \Re_+ \to \Re_+$ is a sparsity-induced penalty function. Some popular h's used in the literature are listed as follows:

(i) (l_1 penalty [19, 6, 5]): $h(t) = \lambda t \ \ \forall t \geq 0$;

(ii) (SCAD penalty [7]): $h(t) = \begin{cases} \lambda t & \text{if } 0 \leq t \leq \lambda, \\ \frac{-t^2 + 2a\lambda t - \lambda^2}{2(a-1)} & \text{if } \lambda < t \leq a\lambda, \\ \frac{(a+1)\lambda^2}{2} & \text{if } \ t > a\lambda; \end{cases}$

(iii) (l_q penalty [8, 11]): $h(t) = \lambda(t + \epsilon)^q \ \ \forall t \geq 0$;

(iv) (Log penalty [21]): $h(t) = \lambda \log(t + \epsilon) - \lambda \log(\epsilon) \ \ \forall t \geq 0$;

(v) (Capped-l_1 penalty [23]): $h(t) = \begin{cases} \lambda t & \text{if } 0 \leq t < \eta, \\ \lambda \eta & \text{if } t \geq \eta, \end{cases}$

where $\lambda > 0$, $0 < q < 1$, $a > 1$, $\eta > 0$ and $\epsilon > 0$ are parameters. One can observe that the above h's are monotonically increasing functions in $[0, \infty)$. Moreover, $\lambda t - h(t)$ is convex in $[0, \infty)$ (see [9]). It implies that $u(y) = \sum_{i=1}^n (\lambda y_i - h(y_i))$ is convex in $\Re^n_+$. Using the monotonicity of h, we can see that (2) can be equivalently reformulated as

$$\min\{l(x) + \sum_{i=1}^n h(y_i) : y \geq |x|, x \in \Omega\}.$$

Further, by using the definition of u, we observe that (2) is equivalent to

$$\min\{l(x) + \lambda\|y\|_1 - u(y) : y \geq |x|, x \in \Omega\},$$

which clearly is a special case of (1) with $\mathcal{X} = \{(x, y) : y \geq |x|, x \in \Omega\}$.

In this paper we provide a comprehensive study on problem (1). In particular, we first establish the first-order optimality conditions for (1). Then we propose SCP methods for solving (1) in which each iteration is obtained by solving a convex programming problem. Under some suitable assumptions, we establish that any accumulation point of the sequence generated by the methods is a KKT point of (1). In addition, we propose a variant of the SCP method for (1) in which nonmonotone scheme and "local" Lipschitz constants of the associated functions are used. A similar convergence result as mentioned above is established.

The outline of this paper is as follows. In Subsection 1.1 we introduce some notations that are used in the paper. In Section 2 we establish the first-order optimality conditions for problem (1). In Section 3 we propose an SCP method and its variant for solving (1) and establish their convergence.

1.1 Notation

Given a nonempty closed convex $\Omega \subseteq \Re^n$, cone(Ω) denotes the cone generated by Ω. Given an arbitrary point $x \in \Omega$, $\mathcal{N}_\Omega(x)$ and $\mathcal{T}_\Omega(x)$ denote the normal and tangent cones of Ω at x, respectively. In addition, dist(y, Ω) denotes the distance between $y \in \Re^n$ and Ω. For a function $h : \Omega \to \Re$, $d \in \Re^n$ and $x \in \Omega$, $h'(x; d)$ is the directional derivative of h at x along d. For a convex function h, $\partial h(x)$ denotes the subdifferential of h at x. Finally, given any $t \in \Re$, we denote its nonnegative part by t^+, that is, $t^+ = \max(t, 0)$.

2 First-order optimality conditions

In this section we establish the first-order optimality conditions for problem (1). Given any $x \in \mathcal{X}$, the set of indices corresponding to the active constraints of (1) at x is denoted by $\mathcal{A}(x)$, that is,

$$\mathcal{A}(x) = \{1 \leq i \leq m : \; g_i(x) + q_i(x) - v_i(x) = 0\}.$$

Theorem 2.1 *Suppose that x^* is a local minimizer of problem* (1)*. Assume that the cone*

$$\sum_{i \in \mathcal{A}(x^*)} \text{cone}(\nabla g_i(x^*) + \partial q_i(x^*) - \partial v_i(x^*)) + \mathcal{N}_{\mathcal{X}}(x^*) \tag{3}$$

is closed, and moreover, there exists $\bar{d} \in \mathcal{T}_{\mathcal{X}}(x^*)$ *such that*

$$g_i'(x^*;\bar{d}) + q_i'(x^*;\bar{d}) - \inf_{s\in\partial v_i(x^*)} s^T\bar{d} \ < \ 0, \quad \forall i \in \mathcal{A}_{\mathcal{T}}(x^*), \tag{4}$$

where

$$\mathcal{A}_{\mathcal{T}}(x^*) = \{i \in \mathcal{A}(x^*) : g_i'(x^*;d) + q_i'(x^*;d) - v_i'(x^*;d) = 0 \textit{ for some } 0 \neq d \in \mathcal{T}_{\mathcal{X}}(x^*)\}. \tag{5}$$

Then, there exists $\lambda^* \in \Re^m$ *together with* x^* *satisfying the KKT conditions*

$$\begin{array}{c} 0 \in \nabla f(x^*) + \partial p(x^*) - \partial u(x^*) + \sum\limits_{i=1}^m \lambda_i^*[\nabla g_i(x^*) + \partial q_i(x^*) - \partial v_i(x^*)] + \mathcal{N}_{\mathcal{X}}(x^*), \\ \lambda_i^* \geq 0, \quad \lambda_i^*[g_i(x^*) + q_i(x^*) - v_i(x^*)] = 0, \quad i = 1,\ldots,m. \end{array}$$

Proof. For convenience, let

$$\begin{array}{rcl} A & = & -\nabla f(x^*) - \partial p(x^*) + \partial u(x^*), \\ B & = & \sum\limits_{i\in\mathcal{A}(x^*)} \mathrm{cone}(\nabla g_i(x^*) + \partial q_i(x^*) - \partial v_i(x^*)) + \mathcal{N}_{\mathcal{X}}(x^*). \end{array}$$

In view of the assumption, one can observe that A and B are closed convex sets. We first show that $A \cap B \neq \emptyset$. Suppose for contradiction that $A \cap B = \emptyset$. It then follows from the well-known separation theorem that there exists $0 \neq d \in \Re^n$ such that

$$\inf_{s\in A} d^T s \geq 1, \qquad \sup_{s\in B} d^T s \leq 0. \tag{6}$$

By the definition of A and the first inequality of (6), one has

$$\begin{array}{rcl} f'(x^*;d) + p'(x^*;d) - u'(x^*;d) & = & d^T\nabla f(x^*) + \sup\limits_{s\in\partial p(x^*)} d^T s - \sup\limits_{s\in\partial u(x^*)} d^T s \\ & \leq & \sup\limits_{s\in A}(-d^T s) \ \leq \ -1 \ < \ 0. \end{array} \tag{7}$$

In addition, it follows from the definition of B and the second inequality of (6) that $d \in (\mathcal{N}_{\mathcal{X}}(x^*))^\circ = \mathcal{T}_{\mathcal{X}}(x^*)$ and

$$\sup\{d^T s : s \in \nabla g_i(x^*) + \partial q_i(x^*) - \partial v_i(x^*)\} \ \leq \ 0, \quad \forall i \in \mathcal{A}(x^*),$$

which implies that

$$g_i'(x^*;d) + q_i'(x^*;d) - v_i'(x^*;d) \ \leq \ 0, \ \forall i \in \mathcal{A}(x^*).$$

Since $d \in \mathcal{T}_{\mathcal{X}}(x^*)$, there exist a positive sequence $\{t_k\} \downarrow 0$ and a sequence $\{x^k\} \subseteq \mathcal{X}$ such that $x^k = x^* + t_k d + o(t_k)$. We next consider two cases to derive a contradiction.

Case 1): Suppose that $g_i'(x^*;d) + q_i'(x^*;d) - v_i'(x^*;d) < 0$ for all $i \in \mathcal{A}(x^*)$. It then follows that for every $i \in \mathcal{A}(x^*)$,

$$\begin{array}{rcl} g_i(x^k) + q_i(x^k) - v_i(x^k) & = & g_i(x^k) - g_i(x^*) + q_i(x^k) - q_i(x^*) - [v_i(x^k) - v_i(x^*)], \\ & = & t_k[g_i'(x^*;d) + q_i'(x^*;d) - v_i'(x^*;d)] + o(t_k) \ < \ 0 \end{array}$$

when $k \gg 1$. Hence, x^k is a feasible point when k is sufficiently large. Using (7) and a similar argument as above, we have

$$f(x^k) + p(x^k) - u(x^k) < f(x^*) + p(x^*) - u(x^*)$$

for all sufficiently large k. In addition, notice that $x^k \to x^*$ as $k \to \infty$. These results imply that x^* is not a local minimizer, which is a contradiction to the assumption.

Case 2): Suppose that there exists some $i_0 \in \mathcal{A}(x^*)$ such that

$$g'_{i_0}(x^*; d) + q'_{i_0}(x^*; d) - v'_{i_0}(x^*; d) = 0.$$

It then together with (5) implies that $i_0 \in \mathcal{A}_\mathcal{T}(x^*)$. By the assumption, there exists $0 \neq \bar{d} \in \mathcal{T}_\mathcal{X}(x^*)$ such that (4) holds. Since $\bar{d} \in \mathcal{T}_\mathcal{X}(x^*)$, there exist a positive sequence $\{\eta_l\} \downarrow 0$ and a sequence $\{y^l\} \subseteq \mathcal{X}$ such that $y^l = x^* + \eta_l \bar{d} + o(\eta_l)$. Let $\bar{d}^l = (y^l - x^*)/\eta_l$. Clearly, $\|\bar{d}^l - \bar{d}\| = o(1)$. It follows that for all i,

$$g'_i(x^*; \bar{d}^l) - g'_i(x^*; \bar{d}) + q'_i(x^*; \bar{d}^l) - q'_i(x^*; \bar{d}) - [\inf_{s \in \partial v_i(x^*)} s^T \bar{d}^l - \inf_{s \in \partial v_i(x^*)} s^T \bar{d}] = O(\|\bar{d}^l - \bar{d}\|) = o(1),$$

which together with (4) implies that for sufficiently large l,

$$g'_i(x^*; \bar{d}^l) + q'_i(x^*; \bar{d}^l) - \inf_{s \in \partial v_i(x^*)} s^T \bar{d}^l \ < \ 0, \quad \forall i \in \mathcal{A}_\mathcal{T}(x^*). \tag{8}$$

Let $\{\alpha_l\} \subset (0, 1]$ be a sequence such that $\alpha_l \downarrow 0$, and let

$$d^l = (1 - \alpha_l) d + \alpha_l \bar{d}^l.$$

Claim that for sufficiently large l,

$$g'_i(x^*; d^l) + q'_i(x^*; d^l) - v'_i(x^*; d^l) \ < \ 0, \ \forall i \in \mathcal{A}(x^*). \tag{9}$$

Indeed, we arbitrarily choose $i \in \mathcal{A}(x^*)$. If $g'_i(x^*; d) + q'_i(x^*; d) - v'_i(x^*; d) < 0$, we then have

$$\lim_{l \to \infty} g'_i(x^*; d^l) + q'_i(x^*; d^l) - v'_i(x^*; d^l) = g'_i(x^*; d) + q'_i(x^*; d) - v'_i(x^*; d) < 0,$$

which immediately implies that (9) holds for sufficiently large l. We now suppose that

$$g'_i(x^*; d) + q'_i(x^*; d) - v'_i(x^*; d) = 0. \tag{10}$$

Hence, $i \in \mathcal{A}_\mathcal{T}(x^*)$. Let $s^* \in \operatorname{Arg}\max_s \{s^T d : s \in \partial v_i(x^*)\}$. Using (8), (10), convexity, and the definition of $\{d^l\}$, we have

$$\begin{aligned}
& g'_i(x^*; d^l) + q'_i(x^*; d^l) - v'_i(x^*; d^l) \\
& \leq (1 - \alpha_l) g'_i(x^*; d) + \alpha_l g'_i(x^*; \bar{d}^l) + (1 - \alpha_l) q'_i(x^*; d) + \alpha_l q'_i(x^*; \bar{d}^l) - (s^*)^T [(1 - \alpha_l) d + \alpha_l \bar{d}^l] \\
& = (1 - \alpha_l)[g'_i(x^*; d) + q'_i(x^*; d) - v'_i(x^*; d)] + \alpha_l [g'_i(x^*; \bar{d}^l) + q'_i(x^*; \bar{d}^l) - (s^*)^T \bar{d}^l] \\
& = \alpha_l [g'_i(x^*; \bar{d}^l) + q'_i(x^*; \bar{d}^l) - (s^*)^T \bar{d}^l] \ \leq \ \alpha_l [g'_i(x^*; \bar{d}^l) + q'_i(x^*; \bar{d}^l) - \inf_{s \in \partial v_i(x^*)} s^T \bar{d}^l] \ < \ 0,
\end{aligned}$$

and hence (9) again holds for sufficiently large l. Now let the sequence $\{x^{k,l}\}$ be defined as

$$x^{k,l} = (1-\alpha_l)x^k + \alpha_l(x^* + t_k\bar{d}^l), \ \forall k, l \geq 1. \tag{11}$$

By the definition of $\bar{d}^l$, one can observe that $x^* + t_k\bar{d}^l \in \mathcal{X}$ for sufficiently large k. It then follows that for each l, $\tilde{x}^{k,l} \in \mathcal{X}$ when $k \gg 1$ due to $x^k \in \mathcal{X}$ and convexity of $\mathcal{X}$. Recall that $x^k = x^* + t_k d + o(t_k)$, which together with (11) yields

$$x^{k,l} = x^* + t_k d^l + o(t_k).$$

Using this relation and (9), one can obtain that, for any $i \in \mathcal{A}(x^*)$ and sufficiently large l,

$$\begin{aligned} g_i(x^{k,l}) + q_i(x^{k,l}) - v_i(x^{k,l}) &= g_i(x^{k,l}) - g_i(x^*) + q_i(x^{k,l}) - q_i(x^*) - [v_i(x^{k,l}) - v_i(x^*)], \\ &= t_k[g_i'(x^*; d^l) + q_i'(x^*; d^l) - v_i'(x^*; d^l)] + o(t_k) \ < 0 \end{aligned}$$

whenever $k \geq n_l$ for some sequence $\{n_l\}$. Hence, $x^{k,l}$ is a feasible point for $k \geq n_l$ and sufficiently large l. Using (7) and the fact $d^l \to d$ as $l \to \infty$, we know that

$$f'(x^*; d^l) + p'(x^*; d^l) - u'(x^*; d^l) \ < \ 0.$$

Using this relation and a similar argument as above, we obtain that for sufficiently large l,

$$f(x^{k,l}) + p(x^{k,l}) - u(x^{k,l}) < f(x^*) + p(x^*) - u(x^*)$$

whenever $k \geq \bar{n}_l$ for some sequence $\{\bar{n}_l\}$. Notice that $x^{k,l} \to x^*$ as $k, l \to \infty$. The above results again contradicts with the assumption that x^* is a local minimizer. Therefore, $A \cap B \neq \emptyset$. The conclusion of this theorem then immediately follows from this relation and the definitions of A and B. ■

Remark.

(a) Condition (3) is satisfied if $\mathcal{X}$ is a polyhedron and $\sum_{i \in \mathcal{A}(x^*)} \text{cone}(\nabla g_i(x^*) + \partial q_i(x^*) - \partial v_i(x^*))$ is a finitely generated cone or if

$$\left(- \sum_{i \in \mathcal{A}(x^*)} \text{cone}(\nabla g_i(x^*) + \partial q_i(x^*) - \partial v_i(x^*)) \right) \cap \mathcal{N}_{\mathcal{X}}(x^*) = \{0\}.$$

It thus follows that, if $\mathcal{X}$ is a polyhedron and q_i and v_i are differentiable or piecewise convex functions (e.g., $\|x\|_1$) for each $i \in \mathcal{A}(x^*)$, condition (3) holds.

(b) When f and g_i's are convex, condition (4) holds if there exists a generalized Slater point $\bar{x} \in \mathcal{X}$, that is, $\bar{x}$ satisfies

$$g_i(\bar{x}) + q_i(\bar{x}) - v_i(x^*) - \inf_{s \in \partial v_i(x^*)} s^T(\bar{x} - x^*) \ < \ 0, \quad \forall i \in \mathcal{A}(x^*).$$

Indeed, let $\bar{d} = \bar{x} - x^*$. Clearly, $\bar{d} \in \mathcal{T}_{\mathcal{X}}(x^*)$. Moreover, for each $i \in \mathcal{A}(x^*)$,

$$\begin{aligned} g_i'(x^*;\bar{d}) + q_i'(x^*;\bar{d}) - \inf_{s\in\partial v_i(x^*)} s^T\bar{d} \;&\le\; g_i(x^*+\bar{d}) - g_i(x^*) + q_i(x^*+\bar{d}) - q_i(x^*) + v_i(x^*) \\ &\quad - v_i(x^*) - \inf_{s\in\partial v_i(x^*)} s^T(\bar{x}-x^*), \\ &= g_i(\bar{x}) + q_i(\bar{x}) - v_i(x^*) - \inf_{s\in\partial v_i(x^*)} s^T(\bar{x}-x^*) \;<\; 0, \end{aligned}$$

and hence condition (4) holds.

■

3 A sequential convex programming method

In this section we propose a sequential convex programming (SCP) method for solving problem (1) in which each iteration is obtained by solving a convex programming problem. We also propose a variant of it for solving (1). Before proceeding, we introduce some notations that will be used subsequently.

For each $x \in \mathcal{X}$, s_f, s_u, s_{g_i}, $s_{v_i} \in \Re^n$ for $i = 1,\ldots,m$, we define

$$\mathcal{C}(x,\{s_{g_i}\}_{i=1}^m,\{s_{v_i}\}_{i=1}^m) = \left\{ y \in \mathcal{X} : \begin{array}{l} g_i(x) + s_{g_i}^T(y-x) + \frac{L_{g_i}}{2}\|y-x\|^2 + q_i(y) \\ -[v_i(x) + s_{v_i}^T(y-x)] \;\le\; 0 \end{array} \right\}, \qquad (12)$$

$$h(y;x,s_f,s_u) = f(x) + s_f^T(y-x) + \frac{L_f}{2}\|y-x\|^2 + p(y) - [u(x) + s_u^T(y-x)].$$

In addition, we denote by $\mathcal{F}$ the feasible region of problem (1).

We are now ready to present an SCP method for solving problem (1).

Exact sequential convex programming method for (1)**:**

Let $x^0 \in \mathcal{F}$ be arbitrarily chosen. Set $k = 0$.

1) Compute $s_f^k = \nabla f(x^k)$, $s_u^k \in \partial u(x^k)$, $s_{g_i}^k = \nabla g_i(x^k)$, $s_{v_i}^k \in \partial v_i(x^k)$ for all i.

2) Solve
$$x^{k+1} \in \operatorname{Arg}\min_y \{h(y;x^k,s_f^k,s_u^k) : y \in \mathcal{C}(x^k,\{s_{g_i}^k\}_{i=1}^m,\{s_{v_i}^k\}_{i=1}^m)\}. \qquad (13)$$

3) Set $k \leftarrow k+1$ and go to step 1).

end

Remark.

(a) When $\mathcal{X} = \Re^n$, $p \equiv 0$, $u \equiv 0$, $L_f > 0$, $q_i \equiv 0$, $v_i \equiv 0$ and $L_{g_i} > 0$ for all i, the above method becomes the MBA method proposed in [1].

(b) When $m = 1$, $f \equiv 0$, $g_1 \equiv 0$, and p, u, q_1, u_1 are smooth convex functions in $\mathcal{X}$, the above method becomes the method studied in [12].

(c) When $m = 0$ and $f \equiv 0$, the above method becomes the well-known method [18, 13] for DC programming.

■

In what follows, we will establish that under some assumptions, any accumulation point of the sequence $\{x^k\}$ generated above is a KKT point of problem (1). Before proceeding, we state several lemmas that will be used subsequently.

The following lemma is well known (see, for example, [16]), which provides an upper bound for a smooth function with Lipschitz continuous gradient.

Lemma 3.1 *Let $\Omega \subseteq \Re^n$ be a closed convex set, and h a differentiable function in Ω. Suppose that there exists some constant $L_h \geq 0$ such that*

$$\|\nabla h(x) - \nabla h(y)\| \leq L_h \|x - y\|, \quad \forall x, y \in \Omega.$$

Then, for any $L \geq L_h$,

$$h(y) \;\leq\; h(x) + \nabla h(x)^T (y - x) + \frac{L}{2}\|y - x\|^2, \quad \forall x, y \in \Omega.$$

The following lemma is due to Robinson [17], which provides an error bound for a class of convex inequalities.

Lemma 3.2 *Let X be a closed convex set in $\Re^n$, and $\mathcal{K}$ a nonempty closed convex cone in $\Re^m$. Suppose that $g : X \to \Re^m$ is a $\mathcal{K}$-convex function, that is,*

$$\lambda g(x^1) + (1 - \lambda) g(x^2) \in g(\lambda x^1 + (1 - \lambda) x^2) + \mathcal{K}.$$

Assume that $x^s \in X$ is a generalized Slater point for the set $\Omega := \{x \in X : 0 \in g(x) + \mathcal{K}\}$, that is, there exists $\delta > 0$ such that $\mathcal{B}(0;\delta) \subseteq g(x^s) + \mathcal{K}$, where $\mathcal{B}(0;\delta)$ is the closed ball centered at 0 with radius δ. Then,

$$\mathrm{dist}(x, \Omega) \;\leq\; \delta^{-1} \|x - x^s\| \mathrm{dist}(0, g(x) + \mathcal{K}), \quad \forall x \in X.$$

The following lemma states a simple property of the set $\mathcal{C}$ that is defined in (12).

Lemma 3.3 *For each $x \in \mathcal{F}$, let $s_{g_i} = \nabla g_i(x)$ and $s_{v_i} \in \partial v_i(x)$. Then, $\mathcal{C}(x, \{s_{g_i}\}_{i=1}^m, \{s_{v_i}\}_{i=1}^m)$ is a nonempty closed convex set in $\mathcal{F}$.*

Proof. Since $x \in \mathcal{F}$, one can clearly see that $x \in \mathcal{C}(x, \{s_{g_i}\}_{i=1}^m, \{s_{v_i}\}_{i=1}^m)$. Hence, $\mathcal{C}(x, \{s_{g_i}\}_{i=1}^m, \{s_{v_i}\}_{i=1}^m) \neq \emptyset$. Due to $s_{v_i} \in \partial v_i(x)$, we know that $v_i(y) \geq v_i(x) + s_{v_i}^T(y - x), \forall y \in \Re^n$. Using this relation and Lemma 3.1, one can see that for any $y \in \mathcal{C}(x, \{s_{g_i}\}_{i=1}^m, \{s_{v_i}\}_{i=1}^m)$, y is in $\mathcal{X}$ and $g_i(y) + q_i(y) - v_i(y) \leq 0$ for $i = 1, \ldots, m$. Hence, $y \in \mathcal{F}$. It implies that $\mathcal{C}(x, \{s_{g_i}\}_{i=1}^m, \{s_{v_i}\}_{i=1}^m) \subseteq \mathcal{F}$. Finally, it is easy to see that $\mathcal{C}(x, \{s_{g_i}\}_{i=1}^m, \{s_{v_i}\}_{i=1}^m)$ is a closed convex set. ∎

We are now ready to establish that under some assumptions, any accumulation point of the sequence $\{x^k\}$ generated by the above SCP method is a KKT point of problem (1).

Theorem 3.4 *Let $\{(x^k, s_f^k, s_u^k, \{s_{g_i}^k\}_{i=1}^m, \{s_{v_i}^k\}_{i=1}^m)\}$ be the sequence generated by the above SCP method. The following statements hold:*

(i) *$\{x^k\} \subset \mathcal{F}$ and $\{f(x^k) + p(x^k) - u(x^k)\}$ is monotonically nonincreasing.*

(ii) *Suppose further that $(x^*, s_f^*, s_u^*, \{s_{g_i}^*\}_{i=1}^m, \{s_{v_i}^*\}_{i=1}^m)$ is an accumulation point of $\{(x^k, s_f^k, s_u^k, \{s_{g_i}^k\}_{i=1}^m, \{s_{v_i}^k\}_{i=1}^m)\}$. Assume that Slater's condition holds for the set $\mathcal{C}(x^*, \{s_{g_i}^*\}_{i=1}^m, \{s_{v_i}^*\}_{i=1}^m)$, that is, there exists $\bar{y} \in \mathcal{X}$ such that*

$$g_i(x^*) + (s_{g_i}^*)^T(\bar{y} - x^*) + \frac{L_{g_i}}{2}\|\bar{y} - x^*\|^2 + q_i(\bar{y}) - [v_i(x^*) + (s_{v_i}^*)^T(\bar{y} - x^*)] \;<\; 0, \; i = 1, \ldots, m. \tag{14}$$

Then, x^* is a KKT point of problem (1).

Proof. (i) We know that $x^0 \in \mathcal{F}$. Since $x^1 \in \mathcal{C}(x^0, \{s_{g_i}^0\}_{i=1}^m, \{s_{v_i}^0\}_{i=1}^m)$, it follows from Lemma 3.3 that $x^1 \in \mathcal{F}$. By repeating this argument, we can conclude that $\{x^k\} \subset \mathcal{F}$. In addition, notice that $x^k \in \mathcal{C}(x^k, \{s_{g_i}^k\}_{i=1}^m, \{s_{v_i}^k\}_{i=1}^m)$. Hence, we have

$$h(x^{k+1}; x^k, s_f^k, s_u^k) \;\leq\; h(x^k; x^k, s_f^k, s_u^k) \;=\; f(x^k) + p(x^k) - u(x^k).$$

Since $s_u^k \in \partial u(x^k)$, we know that $u(x^{k+1}) \geq u(x^k) + (s_u^k)^T(x^{k+1} - x^k)$. Using this relation and Lemma 3.1, one can see that

$$f(x^{k+1}) + p(x^{k+1}) - u(x^{k+1}) \;\leq\; h(x^{k+1}; x^k, s_f^k, s_u^k).$$

It then follows that

$$f(x^{k+1}) + p(x^{k+1}) - u(x^{k+1}) \;\leq\; h(x^{k+1}; x^k, s_f^k, s_u^k) \;\leq\; f(x^k) + p(x^k) - u(x^k). \tag{15}$$

Thus, $\{f(x^k) + p(x^k) - u(x^k)\}$ is monotonically nonincreasing.

(ii) Let $w := (x, \{s_{g_i}\}_{i=1}^m, \{s_{v_i}\}_{i=1}^m)$, $w^k := (x^k, \{s_{g_i}^k\}_{i=1}^m, \{s_{v_i}^k\}_{i=1}^m)$, $w^* := (x^*, \{s_{g_i}^*\}_{i=1}^m, \{s_{v_i}^*\}_{i=1}^m)$. By the assumption, there exists a subsequence K such that $\{(s_f^k, s_u^k, w^k)\}_K \to (s_f^*, s_u^*, w^*)$. We first show that for any $z \in \mathcal{C}(w^*)$, there exists $z^k \in \mathcal{C}(w^k)$ such that $\{z^k\}_K \to z$, where $\mathcal{C}$ is defined in (12). Indeed, let

$$\mathcal{G}_i(y, w) := g_i(x) + s_{g_i}^T(y - x) + \frac{L_{g_i}}{2}\|y - x\|^2 + q_i(y) - [v_i(x) + s_{v_i}^T(y - x)] \quad \forall i,$$

and $\mathcal{G}(y,w) := (\mathcal{G}_1(y,w), \ldots, \mathcal{G}_m(y,w))$. It follows from (14) that $\mathcal{G}(\bar{y}, w^*) < 0$. Hence, there exists $\delta > 0$ such that

$$\mathcal{B}(0;\delta) \subseteq \mathcal{G}(\bar{y}, w^*) + \Re^m_+. \tag{16}$$

Notice that $\mathcal{G}(\bar{y}, w)$ is continuous in w and $\{w^k\}_K \to w^*$. Hence, when $k \in K$ is sufficiently large, $\|\mathcal{G}(\bar{y}, w^k) - \mathcal{G}(\bar{y}, w^*)\| \leq \delta/2$ holds. It immediately implies that, for sufficiently large $k \in K$,

$$\mathcal{G}(\bar{y}, w^*) - \mathcal{G}(\bar{y}, w^k) + \mathcal{B}(0;\delta/2) \subseteq \mathcal{B}(0;\delta).$$

This relation together with (16) yields that, for sufficiently large $k \in K$,

$$\begin{array}{rcl} \mathcal{G}(\bar{y}, w^k) + \Re^m_+ & = & \mathcal{G}(\bar{y}, w^k) - \mathcal{G}(\bar{y}, w^*) + \mathcal{G}(\bar{y}, w^*) + \Re^m_+ \supseteq \mathcal{G}(\bar{y}, w^k) - \mathcal{G}(\bar{y}, w^*) + \mathcal{B}(0;\delta) \\ & \supseteq & \mathcal{G}(\bar{y}, w^k) - \mathcal{G}(\bar{y}, w^*) + [\mathcal{G}(\bar{y}, w^*) - \mathcal{G}(\bar{y}, w^k) + \mathcal{B}(0;\delta/2)] \;=\; \mathcal{B}(0;\delta/2). \end{array}$$

Hence, $\bar{y}$ is also a generalized Slater point for the set $\mathcal{C}(w^k)$ when $k \in K$ is sufficiently large. In addition, it is not hard to verify that $\mathcal{G}(y, w^k)$ is $\Re^m_+$-convex. Letting $g(\cdot) = \mathcal{G}(\cdot, w^k)$, $\mathcal{K} = \Re^m_+$, $\Omega = \mathcal{C}(w^k)$, $X = \mathcal{X}$, and using Lemma 3.2, we obtain that, for sufficiently large $k \in K$,

$$\operatorname{dist}(y, \mathcal{C}(w^k)) \;\leq\; 2\delta^{-1}\|y - \bar{y}\| \operatorname{dist}(0, \mathcal{G}(y, w^k) + \Re^m_+), \quad \forall y \in \mathcal{X}. \tag{17}$$

Let $z \in \mathcal{C}(w^*)$ be arbitrarily given, and let $z^k = \arg\min_y\{\|z - y\| : y \in \mathcal{C}(w^k)\}$. Notice that $z \in \mathcal{X}$. It then follows from (17) with $y = z$ that, when $k \in K$ is sufficiently large,

$$\|z^k - z\| = \operatorname{dist}(z, \mathcal{C}(w^k)) \;\leq\; 2\delta^{-1}\|z - \bar{y}\| \operatorname{dist}(\mathcal{G}(z, w^k), -\Re^m_+).$$

Since $z \in \mathcal{C}(w^*)$, we can observe that $\{\operatorname{dist}(\mathcal{G}(z, w^k), -\Re^m_+)\}_K \to \operatorname{dist}(\mathcal{G}(z, w^*), -\Re^m_+) = 0$. Using this relation and the above inequality, we obtain that $\{z^k\}_K \to z$ and $z^k \in \mathcal{C}(w^k)$.

Since $\{x^k\}_K \to x^*$, by continuity we have $\{f(x^k) + p(x^k) - u(x^k)\}_K \to f(x^*) + p(x^*) - u(x^*)$. Notice that $\{f(x^k) + p(x^k) - u(x^k)\}$ is monotonically nonincreasing. Hence, we have $f(x^k) + p(x^k) - u(x^k) \to f(x^*) + p(x^*) - u(x^*)$, which together with (15) implies that $h(x^{k+1}; x^k, s^k_f, s^k_u) \to f(x^*) + p(x^*) - u(x^*)$. Recall that $x^{k+1} \in \operatorname{Arg\,min}\{h(y; x^k, s^k_f, s^k_u) : y \in \mathcal{C}(w^k)\}$. Since $z^k \in \mathcal{C}(w^k)$, we obtain that $h(x^{k+1}; x^k, s^k_f, s^k_u) \leq h(z^k; x^k, s^k_f, s^k_u)$. Upon taking limits on both sides of this inequality as $k \in K \to \infty$, we have

$$f(x^*) + p(x^*) - u(x^*) \leq h(z; x^*, s^*_f, s^*_u), \quad \forall z \in \mathcal{C}(w^*).$$

In addition, since $\{x^k\} \subset \mathcal{F}$ and $\{x^k\}_K \to x^*$, we know that $x^* \in \mathcal{F}$, which yields $x^* \in \mathcal{C}(w^*)$. Also, $f(x^*) + p(x^*) - u(x^*) = h(x^*; x^*, s^*_f, s^*_u)$. Therefore,

$$x^* \in \operatorname{Arg\,min}\{h(z; x^*, s^*_f, s^*_u) : z \in \mathcal{C}(w^*)\}. \tag{18}$$

Since Slater's condition holds for $\mathcal{C}(w^*)$, the first-order optimality condition of (18) immediately implies that x^* is a KKT point of (1). ∎

Remark. Since $s^k_f = \nabla f(x^k)$, $s^k_u \in \partial u(x^k)$, $s^k_{g_i} = \nabla g_i(x^k)$, and $s^k_{v_i} \in \partial v_i(x^k)$ for all i, we observe that if $\{x^k\}$ has an accumulation point, so is $\{s^k_f, s^k_u, \{s^k_{g_i}\}_{i=1}^m, \{s^k_{v_i}\}_{i=1}^m)\}$. Therefore, the first assumption in statement (ii) is mild. We next provide a sufficient condition for the second assumption to hold. In particular, we show that the assumption (14) holds if the following generalized Mangasarian-Fromovitz constraint qualification (MFCQ) holds at x^*. ■

Proposition 3.5 *Let x^* be a point in $\mathcal{F}$. If the generalized MFCQ holds at x^*, that is, $\exists d \in \mathcal{T}_{\mathcal{X}}(x^*)$ such that*

$$g_i'(x^*;d) + q_i'(x^*;d) - \inf_{s\in\partial v_i(x^*)} s^T d \ < \ 0, \quad \forall i \in \mathcal{A}(x^*). \tag{19}$$

Then, (14) holds at x^* for $s^*_{g_i} = \nabla g_i(x^*)$ and every $s^*_{v_i} \in \partial v_i(x^*)$.

Proof. Let d be given above, $s^*_{g_i} = \nabla g_i(x^*)$ and $s^*_{v_i} \in \partial v_i(x^*)$. Then, there exist a positive sequence $\{t_k\} \downarrow 0$ and a sequence $\{x^k\} \subseteq \mathcal{X}$ such that $x^k = x^* + t_k d + o(t_k)$. For each $i \in \mathcal{A}(x^*)$, we have that, for sufficiently large k,

$$\begin{aligned}
&g_i(x^*) + (s^*_{g_i})^T(x^k - x^*) + \tfrac{L_{g_i}}{2}\|x^k - x^*\|^2 + q_i(x^k) - [v_i(x^*) + (s^*_{v_i})^T(x^k - x^*)] \\
&= (s^*_{g_i})^T(x^k - x^*) + \tfrac{L_{g_i}}{2}\|x^k - x^*\|^2 + q_i(x^k) - q_i(x^*) - (s^*_{v_i})^T(x^k - x^*) \\
&= t_k(s^*_{g_i})^T d + q_i(x^* + t_k d) - q_i(x^*) - t_k(s^*_{v_i})^T d + o(t_k) \\
&= t_k[(s^*_{g_i})^T d + q_i'(x^*;d) - (s^*_{v_i})^T d + o(1)] \ \le \ t_k[g_i'(x^*;d) + q_i'(x^*;d) - \inf_{s\in\partial v_i(x^*)} s^T d + o(1)] \ < \ 0,
\end{aligned}$$

where the last inequality follows from (19). In addition, for each $i \notin \mathcal{A}(x^*)$, we know that $g_i(x^*) + q_i(x^*) - v_i(x^*) < 0$. Notice that $x^k \to x^*$ as $k \to \infty$. Hence, for sufficiently large k, we have

$$g_i(x^*) + (s^*_{g_i})^T(x^k - x^*) + \frac{L_{g_i}}{2}\|x^k - x^*\|^2 + q_i(x^k) - [v_i(x) + (s^*_{v_i})^T(x^k - x^*)] < 0, \quad i = 1, \ldots, m.$$

■

The above SCP method uses the global Lipschitz constants of ∇f and ∇g_i's, which may be too conservative. To improve its practical performance, we can use "local" Lipschitz constants that are updated dynamically. In addition, the above method is a monotone method since $\{f(x^k)+p(x^k)-u(x^k)\}$ is nonincreasing. As mentioned in [10, 4, 22, 14], nonmonotone methods generally outperform monotone counterparts for many nonlinear programming problems. We next propose a variant of the SCP in which "local" Lipschitz constants and nonmonotone scheme are used. Before proceeding, we introduce some notations as follows.

For each $x \in \mathcal{F}$, l_f, $l_{g_i} \in \Re$, s_f, s_u, s_{g_i}, $s_{v_i} \in \Re^n$ for $i = 1, \ldots, m$, we define

$$\bar{\mathcal{C}}(x, \{l_{g_i}\}_{i=1}^m, \{s_{g_i}\}_{i=1}^m, \{s_{v_i}\}_{i=1}^m) \ = \ \left\{ y \in \mathcal{X} : \begin{array}{l} g_i(x) + s_{g_i}^T(y-x) + \frac{l_{g_i}}{2}\|y-x\|^2 + q_i(y) \\ -[v_i(x) + s_{v_i}^T(y-x)] \ \le \ 0 \end{array} \right\}, \tag{20}$$

$$\bar{h}(y; x, l_f, s_f, s_u) \ = \ f(x) + s_f^T(y-x) + \frac{l_f}{2}\|y-x\|^2 + p(y) - [u(x) + s_u^T(y-x)],$$

$$F(x) \ := \ f(x) + p(x) - u(x).$$

We are now ready to present a variant of the above SCP method.

A variant of SCP method for (1)**:**

Choose parameters $c > 0$, $0 < L_{\min} < L_{\max}$, $\tau > 1$, and integer $M \geq 0$. Set $k = 0$ and choose an arbitrary $x^0 \in \mathcal{F}$.

1) Compute $s_f^k = \nabla f(x^k)$, $s_u^k \in \partial u(x^k)$, $s_{g_i}^k = \nabla g_i(x^k)$, $s_{v_i}^k \in \partial v_i(x^k)$ for all i.

2) Choose $l_f^{k,0}$, $l_{g_i}^{k,0} \in [L_{\min}, L_{\max}]$ arbitrarily, and set $l_f^k = l_f^{k,0}$ and $l_{g_i}^k = l_{g_i}^{k,0}$ for all i.

3) Find

$$x^{k+1} = \arg\min_y \{\bar{h}(y; x^k, l_f^k, s_f^k, s_u^k) : y \in \bar{\mathcal{C}}(x^k, \{l_{g_i}^k\}_{i=1}^m, \{s_{g_i}^k\}_{i=1}^m, \{s_{v_i}^k\}_{i=1}^m)\}. \tag{21}$$

   3a) If $x^{k+1} \in \mathcal{F}$ and

$$F(x^{k+1}) \leq \max_{[k-M]^+ \leq i \leq k} F(x^i) - \frac{c}{2}\|x^{k+1} - x^k\|^2 \tag{22}$$

   holds, go to step 4).

   3b) If $x^{k+1} \notin \mathcal{F}$, set $l_{g_i}^k \leftarrow \tau l_{g_i}^k$ for all i and go to step 3).

   3c) If (22) does not hold, set $l_f^k \leftarrow \tau l_f^k$ and go to step 3).

4) Set $k \leftarrow k + 1$ and go to step 1).

end

Remark.

(i) When $M = 0$, the above method becomes a monotone method.

(ii) In practical computation, $l_f^{k,0}$, $l_{g_i}^{k,0}$ can be updated by the similar strategy as used in [2, 4], that is,

$$\begin{aligned} l_f^{k,0} &= \max\left\{L_{\min}, \min\left\{L_{\max}, \frac{\Delta x^T \Delta f}{\|\Delta x\|^2}\right\}\right\}, \\ l_{g_i}^{k,0} &= \max\left\{L_{\min}, \min\left\{L_{\max}, \frac{\Delta x^T \Delta g_i}{\|\Delta x\|^2}\right\}\right\}, \quad \forall i, \end{aligned}$$

where $\Delta x = x^k - x^{k-1}$, $\Delta f = \nabla f(x^k) - \nabla f(x^{k-1})$, and $\Delta g_i = \nabla g_i(x^k) - \nabla g_i(x^{k-1})$ for all i.

(iii) l_f^k and $\{l_{g_i}^k\}_{i=1}^m$ can be updated by some other strategies. For example,

   1) we may update l_f^k and $\{l_{g_i}^k\}_{i=1}^m$ simultaneously, that is, steps 3b) and 3c) can be replaced by:

if $x^{k+1} \notin \mathcal{F}$ or (22) does not hold, set $l_f^k \leftarrow \tau l_f^k$ and $l_{g_i}^k \leftarrow \tau l_{g_i}^k$ for all i;

2) in step 3b), each $l_{g_i}^k$ can be updated individually. In particular, for each i, we can update $l_{g_i}^k$ only if the ith constraint of (1) is violated at x^{k+1}, that is, $g_i(x^{k+1}) + q_i(x^{k+1}) - v_i(x^{k+1}) > 0$.

■

We first show that for each outer iteration, its number of inner iterations is finite.

Theorem 3.6 *At each kth outer iteration, its associated inner iterations terminate after at most*

$$\left\lfloor \frac{\log(L_f + c) + \log(\max_i L_{g_i}) - 2\log(2L_{\min})}{\log \tau} + 4 \right\rfloor \tag{23}$$

loops.

Proof. Let $\bar{l}_f^k$ and $\bar{l}_{g_i}^k$ denote the final value of l_f^k and $l_{g_i}^k$ at the kth outer iteration, respectively. Note that $\bar{h}(\cdot; x^k, l_f^k, s_f^k, s_u^k)$ is a strongly convex function with modulus $l_f^k > 0$. It then follows from (21) that

$$F(x^k) = f(x^k) + p(x^k) - u(x^k) = \bar{h}(x^k; x^k, l_f^k, s_f^k, s_u^k) \;\geq\; \bar{h}(x^{k+1}; x^k, l_f^k, s_f^k, s_u^k) + \frac{l_f^k}{2}\|x^{k+1} - x^k\|^2.$$

Since $s_u^k \in \partial u(x^k)$, we know that $u(x^{k+1}) \geq u(x^k) + (s_u^k)^T(x^{k+1} - x^k)$. Using this relation and Lemma 3.1, one can see that

$$F(x^{k+1}) \;=\; f(x^{k+1}) + p(x^{k+1}) - u(x^{k+1}) \;\leq\; \bar{h}(x^{k+1}; x^k, l_f^k, s_f^k, s_u^k) + \frac{L_f - l_f^k}{2}\|x^{k+1} - x^k\|^2.$$

The above two inequalities yield

$$F(x^{k+1}) \;\leq\; F(x^k) - (l_f^k - \frac{L_f}{2})\|x^{k+1} - x^k\|_2^2 \;\leq\; \max_{[k-M]^+ \leq i \leq k} F(x^i) - (l_f^k - \frac{L_f}{2})\|x^{k+1} - x^k\|_2^2.$$

Similarly, one can show that

$$g_i(x^{k+1}) + q_i(x^{k+1}) - v_i(x^{k+1}) \;\leq\; g_i(x^k) + q_i(x^k) - v_i(x^k) - (l_{g_i}^k - \frac{L_{g_i}}{2})\|x^{k+1} - x^k\|_2^2, \;\forall i,$$

which together with $x^k \in \mathcal{F}$ implies that

$$g_i(x^{k+1}) + q_i(x^{k+1}) - v_i(x^{k+1}) \;\leq -(l_{g_i}^k - \frac{L_{g_i}}{2})\|x^{k+1} - x^k\|_2^2, \;\forall i.$$

Hence, $x^{k+1} \in \mathcal{F}$ and (22) holds whenever $l_f^k \geq (L_f + c)/2$ and $\min_i l_{g_i}^k \geq (\max_i L_{g_i})/2$, which, together with the definitions of $\bar{l}_k$ and $\bar{l}_{g_i}^k$, implies that $\bar{l}_k/\tau < (L_f + c)/2$ and $\min_i \bar{l}_{g_i}^k/\tau <$

$(\max_i L_{g_i})/2$, that is, $\bar{l}_k < \tau(L_f + c)/2$ and $\min_i \bar{l}^k_{g_i} < \tau(\max_i L_{g_i})/2$. Let n^k_f and n^k_g denote the number of inner iterations for updating l^k_f and $l^k_{g_i}$ at the kth outer iteration. Then, we have

$$\begin{array}{rcccl} L_{\min}\tau^{n^k_f-1} & \leq & L^{k,0}_f\tau^{n^k_f-1} & = & \bar{l}^k_f \; < \; \tau(L_f+c)/2, \\ L_{\min}\tau^{n^k_g-1} & \leq & (\min_i L^{k,0}_{g_i})\tau^{n^k_g-1} & = & \min_i \bar{l}^k_{g_i} \; < \; \tau(\max_i L_{g_i})/2. \end{array}$$

Hence, the total number of inner iterations, $n^k_f + n^k_g$, is bounded above by the quantity given in (23) and the conclusion holds. ∎

We next establish that under some assumptions, any accumulation point of the sequence $\{x^k\}$ generated by the above variant of the SCP method is a KKT point of problem (1).

Theorem 3.7 *Let $\{(x^k, s^k_f, s^k_u, \{s^k_{g_i}\}^m_{i=1}, \{s^k_{v_i}\}^m_{i=1})\}$ be the sequence generated by the above variant of the SCP method. Assume that $F(x) := f(x)+p(x)-u(x)$ is uniformly continuous in the level set $\mathcal{L} = \{x \in \mathcal{F} : F(x) \leq F(x^0)\}$. Suppose that $(x^*, l^*_f, \{l^*_{g_i}\}^m_{i=1}, s^*_f, s^*_u, \{s^*_{g_i}\}^m_{i=1}, \{s^*_{v_i}\}^m_{i=1})$ is an accumulation point of $\{(x^k, l^k_f, \{l^k_{g_i}\}^m_{i=1}, s^k_f, s^k_u, \{s^k_{g_i}\}^m_{i=1}, \{s^k_{v_i}\}^m_{i=1})\}$. Then the following statements hold:*

(i) $\|x^{k+1} - x^k\| \to 0$ *and* $f(x^k) + p(x^k) - u(x^k) \to f(x^*) + p(x^*) - u(x^*)$.

(ii) *Suppose further that Slater's condition holds for the constraint set $\mathcal{C}(x^*, \{l^*_{g_i}\}^m_{i=1}, \{s^*_{g_i}\}^m_{i=1}, \{s^*_{v_i}\}^m_{i=1})$, that is, there exists $\bar{y} \in \mathcal{X}$ such that*

$$g_i(x^*)+(s^*_{g_i})^T(\bar{y}-x^*)+\frac{l^*_{g_i}}{2}\|\bar{y}-x^*\|^2+q_i(\bar{y})-[v_i(x^*)+(s^*_{v_i})^T(\bar{y}-x^*)] \; < \; 0, \; i=1,\ldots,m. \quad (24)$$

Then, x^* is a KKT point of problem (1).

Proof. (i) By the definition of x^k, we observe that $\{x^k\} \subseteq \mathcal{L}$. Let $d^k := x^{k+1} - x^k$, and $l(k)$ an integer between $[k-M]^+$ and k such that

$$F(x^{l(k)}) = \max\{F(x^i) : [k-M]^+ \leq i \leq k\}, \quad \forall k \geq 0.$$

It follows from (22) that $F(x^{k+1}) \leq F(x^{l(k)})$ for all $k \geq 0$, which together with the definition of $l(k)$ implies that $\{F(x^{l(k)})\}$ is monotonically nonincreasing. Further, by continuity of F and $\{x^k\}_K \to x^*$, we know that $\{F(x^k)\}_K \to F(x^*)$. This together with the fact $F(x^{l(k)}) \geq F(x^k)$ implies that $\{F(x^{l(k)})\}_K$ is bounded below. Using this result and the monotonicity of $\{F(x^{l(k)})\}$, we see that $\{F(x^{l(k)})\}$ is bounded below. Hence, there exists some $F^* \in \Re$ such that

$$\lim_{k\to\infty} F(x^{l(k)}) = F^*. \quad (25)$$

We can prove by induction that the following limits hold for all $j \geq 1$:

$$\lim_{k\to\infty} d^{l(k)-j} = 0, \qquad \lim_{k\to\infty} F(x^{l(k)-j}) = F^*. \quad (26)$$

Indeed, replacing k by $l(k)-1$ in (22) and using the definition of $l(k)$, we obtain that

$$F(x^{l(k)}) \leq F(x^{l(l(k)-1)}) - \frac{c}{2}\|d^{l(k)-1}\|^2,$$

which together with (25) implies that $\lim_{k\to\infty} d^{l(k)-1} = 0$. Using this relation, (25) and uniform continuity of F in $\mathcal{L}$, we have

$$\lim_{k\to\infty} F(x^{l(k)-1}) = \lim_{k\to\infty} F(x^{l(k)} - d^{l(k)-1}) = \lim_{k\to\infty} F(x^{l(k)}) = F^*.$$

Therefore, (26) holds for $j = 1$. Now, we assume that (26) holds for j. We need to show that it also holds for $j+1$. Replacing k by $l(k)-j-1$ in (22) and using the definition of $l(k)$, we have

$$F(x^{l(k)-j}) \ \leq \ F(x^{l(l(k)-j-1)}) - \frac{c}{2}\|d^{l(k)-j-1}\|^2,$$

which, together with (25) and the induction assumption $\lim_{k\to\infty} F(x^{l(k)-j}) = F^*$, implies that $\lim_{k\to\infty} d^{l(k)-j-1} = 0$. Using this result, $\lim_{k\to\infty} F(x^{l(k)-j}) = F^*$ and uniform continuity of F in $\mathcal{L}$, we see that $\lim_{k\to\infty} F(x^{l(k)-j-1}) = F^*$. Hence, (26) holds for $j+1$. It then follows from the induction that (26) holds for all $j \geq 1$. Further, by the definition of $l(k)$, we see that for $k \geq M+1$, $k-M-1 = l(k)-j$ for some $1 \leq j \leq M+1$, which together with the first limit in (26), implies that $\lim_{k\to\infty} d^k = \lim_{k\to\infty} d^{k-M-1} = 0$. Additionally, we observe that

$$x^{l(k)} = x^{k-M-1} + \sum_{j=1}^{\bar{l}_k} d^{l(k)-j} \quad \forall k \geq M+1,$$

where $\bar{l}_k = l(k) - (k-M-1) \leq M+1$. Using the above identity, (26), and uniform continuity of F in $\mathcal{L}$, we see that $\lim_{k\to\infty} F(x^k) = \lim_{k\to\infty} F(x^{k-M-1}) = F^*$, which, together with $\{F(x^k)\}_K \to F(x^*)$, implies that $F(x^k) \to F(x^*)$. Hence, the statement (i) holds.

(ii) Let $w := (x, \{l_{g_i}\}_{i=1}^m, \{s_{g_i}\}_{i=1}^m, \{s_{v_i}\}_{i=1}^m)$, $w^k := (x^k, \{l_{g_i}^k\}_{i=1}^m, \{s_{g_i}^k\}_{i=1}^m, \{s_{v_i}^k\}_{i=1}^m)$, $w^* := (x^*, \{l_{g_i}^*\}_{i=1}^m, \{s_{g_i}^*\}_{i=1}^m, \{s_{v_i}^*\}_{i=1}^m)$. By the assumption, there exists a subsequence K such that $\{(l_f^k, s_f^k, s_u^k, w^k)\}_K \to (l_f^*, s_f^*, s_u^*, w^*)$. We first show that for any $z \in \bar{\mathcal{C}}(w^*)$, there exists $z^k \in \bar{\mathcal{C}}(w^k)$ such that $\{z^k\}_K \to z$, where $\bar{\mathcal{C}}$ is defined in (20). Indeed, let

$$\bar{\mathcal{G}}_i(y,w) := g_i(x) + s_{g_i}^T(y-x) + \frac{l_{g_i}}{2}\|y-x\|^2 + q_i(y) - [v_i(x) + s_{v_i}^T(y-x)] \quad \forall i,$$

and $\bar{\mathcal{G}}(y,w) := (\bar{\mathcal{G}}_1(y,w), \ldots, \bar{\mathcal{G}}_m(y,w))$. Notice that $\bar{\mathcal{G}}(\bar{y},w)$ is continuous in w. Using this fact, (24), Lemma 3.2, and the similar arguments as in the proof of Theorem 3.4 (ii), one can show that there exists some $\delta > 0$ such that for sufficiently large $k \in K$,

$$\mathrm{dist}(y, \bar{\mathcal{C}}(w^k)) \ \leq \ 2\delta^{-1}\|y-\bar{y}\|\mathrm{dist}(0, \bar{\mathcal{G}}(y,w^k) + \Re_+^m), \quad \forall y \in \mathcal{X}. \tag{27}$$

Let $z \in \bar{\mathcal{C}}(w^*)$ be arbitrarily given, and let $z^k = \arg\min_y\{\|z-y\| : y \in \bar{\mathcal{C}}(w^k)\}$. Clearly, $z \in \mathcal{X}$ and $\mathrm{dist}(\bar{\mathcal{G}}(z,w^*), -\Re_+^m) = 0$. Using these facts and letting $y = z$ in (27), one can obtain that $\{z^k\}_K \to z$ and $z^k \in \bar{\mathcal{C}}(w^k)$.

Recall from statement (i) that $\|x^{k+1} - x^k\| \to 0$. Since $\{x^k\}_K \to x^*$, it then follows that $\{x^{k+1}\}_K \to x^*$. Let $\bar{l}_f^k$ denote the final value of l_f^k at the kth outer iteration. From the proof of Theorem 3.6, we know that $\bar{l}^k \in [L_{\min}, \tau(L_f + c)/2]$. Using these facts and $\{F(x^k)\} \to F(x^*)$, we observe that

$$\{\bar{h}(x^{k+1}; x^k, \bar{l}_f^k, s_f^k, s_u^k)\}_K \to F(x^*).$$

Recall that $x^{k+1} = \arg\min\{\bar{h}(y; x^k, \bar{l}_f^k, s_f^k, s_u^k) : y \in \bar{\mathcal{C}}(w^k)\}$. Since $z^k \in \bar{\mathcal{C}}(w^k)$, we have $\bar{h}(x^{k+1}; x^k, \bar{l}^k, s_f^k, s_u^k) \leq \bar{h}(z^k; x^k, \bar{l}^k, s_f^k, s_u^k)$. Upon taking limits on both sides of this inequality as $k \in K \to \infty$, we obtain that

$$F(x^*) \leq \bar{h}(z; x^*, l_f^*, s_f^*, s_u^*), \quad \forall z \in \bar{\mathcal{C}}(w^*).$$

In addition, we know that $x^* \in \mathcal{F}$, which implies that $x^* \in \bar{\mathcal{C}}(w^*)$. Also, $F(x^*) = \bar{h}(x^*; x^*, l_f^*, s_f^*, s_u^*)$. Hence, we have

$$x^* \in \operatorname{Arg\,min}\{\bar{h}(z; x^*, l_f^*, s_f^*, s_u^*) : z \in \bar{\mathcal{C}}(w^*)\}. \tag{28}$$

Since Slater's condition holds for $\bar{\mathcal{C}}(w^*)$, the first-order optimality condition of (28) immediately implies that x^* is a KKT point of (1). ■

Remark. For $M = 0$, Theorem 3.7 still holds without the uniform continuity of $F(x)$ in the level set $\mathcal{L} = \{x \in \mathcal{F} : F(x) \leq F(x^0)\}$. ■

Acknowledgment

The author would like to thank Ting Kei Pong for bringing his attention to the reference [17].

References

[1] A. Auslender, R. Shefi, and M. Teboulle. A Moving balls approximation method for a class of smooth constrained minimization problems. *SIAM J. Optim.*, 20(6): 3232–3259, 2010.

[2] J. Barzilai and J.M. Borwein. Two point step size gradient methods. *IMA J. Numer. Anal.*, 8:141–148, 1988.

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2:183–202, 2009.

[4] E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optimiz*, 4:1196–1211, 2000.

[5] E. Candès and T. Tao. Decoding by linear programming. *IEEE T. Inform. Theory*, 51(12):4203–4215, 2005.

[6] S. Chen, D. Donoho and M. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20:33-61, 1998.

[7] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.*, 96:1348–1360, 2001.

[8] I. Frank and J. Friedman. A statistical view of some chemometrics regression tools (with discussion). *Technometrics*, 35:109–148, 1993.

[9] G. Gasso, A. Rakotomamonjy, and S. Canu. Recovering sparse signals with a certain family of non-convex penalties and DC programming. *IEEE T. Image Process.*, 57:4686–4698, 2009.

[10] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.*, 23:707–716, 1986.

[11] W. J. Fu. Penalized regression: the bridge versus the lasso. *J. Comput. Graph. Stat.*, 7:397–416, 1998.

[12] L. Hong, Y. Yang, and L. Zhang. Sequential convex approximations to joint chance constrained programs: a Monte Carlo approach. *Oper. Res.*, 59(3): 617–630, 2011.

[13] R. Horst and N. Y. Thoai. DC programming: overview. *J. Optim. Theory App*, 103: 1–41, 1999.

[14] Z. Lu and Y. Zhang. An augmented Lagrangian approach for sparse principal component analysis. *Math. Program.*, 135: 145–193, 2012.

[15] Y. E. Nesterov. Gradient methods for minimizing composite objective function. CORE report, 2007; available at http://www.ecore.be/DPs/dp 1191313936.pdf.

[16] J. M. Ortega and W. C. Rheinboldt. Iterative Solution of Nonlinear Equations in Several Variables. *Classics Appl. Math.* 30, SIAM, Philadelphia, PA, 2000.

[17] S. Robinson. An application of error bounds for convex programming in a linear space. *SIAM J. Optim.*, 13(2): 271–273, 1975.

[18] P. D. Tao and L. T. H. An. DC optimization algorithms for solving the trust region subproblem. *SIAM J. Optim.*, 8(2): 476–505, 1998.

[19] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. B*, 58(1):267-288, 1996.

[20] P. Tseng and S. Yun. A coordinate gradient descent method for nonsmooth separable minimization. *Math. Prog.*, 117:387–423, 2009.

[21] J. Weston, A. Elisseeff, B. Scholkopf, and M. Tipping. The use of zero-norm with linear models and kernel methods. *J. Mach. Learn. Res.*, 3:1439–1461, 2003.

[22] S. J. Wright, R. Nowak, and M. Figueiredo. Sparse reconstruction by separable approximation. *IEEE T. Image Process.*, 57:2479–2493, 2009.

[23] T. Zhang. Analysis of multi-stage convex relaxation for sparse regularization. *J. Mach. Learn. Res.*, 11:1081–1107, 2010.