November 12, 2001 Technical Report 2001-011

A Sequential Quadratically Constrained Quadratic Programming Method for Differentiable Convex Minimization*

Masao Fukushima[1], Zhi-Quan Luo[2] and Paul Tseng[3]

ABSTRACT

This paper presents a Sequential Quadratically Constrained Quadratic Programming (SQCQP) method for solving smooth convex programs. The SQCQP method solves at each iteration a subproblem that involves convex quadratic inequality constraints as well as a convex quadratic objective function. Such a quadratically constrained quadratic programming problem can be formulated as a second-order cone program, which can be solved efficiently by using interior point methods. We consider the following three fundamental issues on the SQCQP method: the feasibility of subproblems, the global convergence, and the quadratic rate of convergence. In particular, we show that the Maratos effect is avoided without any modification to the search direction, even though we use an ordinary ℓ_1 exact penalty function as the line search merit function.

---

* This research is supported by a Grant-in-Aid for Scientific Research (B) from the Ministry of Education, Science, Sports and Culture of Japan. In addition, the second author is supported by a grant from NSERC and by the Canada Research Chair Program, and the third author is also supported by the National Science Foundation Grant CCR-9731273.

[1] Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan. Email: fuku@i.kyoto-u.ac.jp

[2] Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, L8S 4L7, Canada. Email: luozq@mcmaster.ca

[3] Department of Mathematics, University of Washington, Seattle, WA 98195, U.S.A. Email: tseng@math.washington.edu

1 Introduction

This paper presents a *Sequential Quadratically Constrained Quadratic Programming* (SQCQP) method for solving smooth convex programs. Unlike Sequential Quadratic Programming (SQP) methods [3, 4, 5, 10], the SQCQP method solves at each iteration a subproblem that involves convex quadratic inequality constraints and a convex quadratic objective function. Such a quadratically constrained quadratic programming problem can be formulated as a second-order cone program [8, 12], and be solved efficiently by using interior point methods [1, 8, 11, 12, 19].

In this paper, we consider the following three fundamental issues concerning the SQCQP method: the feasibility of subproblems, the global convergence, and the quadratic rate of convergence. First, we note that a straightforward quadratic approximation of the constraints may yield an infeasible subproblem even though we are dealing with a convex program. This is in contrast with SQP methods, in which linearized constraints in a subproblem always admit a feasible solution as long as the original convex program is feasible. To maintain feasibility of the subproblem, we propose a strategy of switching between the linear and the quadratic approximation for each constraint, taking into account the constraint violation and curvature at the current iterate. It is shown that the proposed strategy always yields a feasible subproblem and that the quadratic approximation will eventually be used for every constraint. To ensure global convergence, we adopt a standard line search technique using the ℓ_1 penalty function. A major difference between the SQCQP method and the classical SQP method lies in that the former uses information up to second-order for both the constraints and the objective function in generating a search direction. As a result, the SQCQP method not only enjoys local quadratic convergence, but also is free of the usual Maratos effect known to cause difficulty in an SQP method employing a nondifferentiable exact penalty function as the line search merit function [17]. More importantly, the elimination of the Maratos effect is achieved without performing any additional correction step in the SQCQP method as required in the SQP-type methods [5, 10].

Related SQCQP-type methods have been considered by several authors. Panin [14] studied an SQCQP method and proved its global and superlinear convergence under the somewhat restrictive assumptions that the objective and constraint functions are strongly convex functions with uniformly bounded Hessian and that the sequence of Lagrange multiplier estimates generated by the method is uniformly bounded (see also [15] for a related work). Polak, Mayne and Higgins [16]

proposed, for semi-infinite minimax problems, an iterative method that solves at each iteration a minimax subproblem constructed from quadratic approximations of component functions. Global convergence and local superlinear convergence of order 1.5 for this method were shown under the assumption that the component functions are strongly convex with uniformly bounded Hessian. Wiest and Polak [20] proposed a phase I-phase II SQCQP method in which the quadratic objective and constraint functions involved in the subproblem have the identical Hessian which is a positive multiple of the identity matrix. Global convergence and local linear convergence for this method were shown under the Mangasarian-Fromovitz constraint qualification (MFCQ) plus the assumptions of second-order sufficiency and strict complementarity. More recently, Kruk and Wolkowicz [7] discuss, for convex and nonconvex problems, (approximate) solution of the trust region subproblem by semidefinite relaxation. As this paper was being written, we learned of a recent work of Anitescu [2] proposing a trust-region SQCQP method for solving (nonconvex) nonlinear minimization problems. Local superlinear convergence of order 1.5 for this method is shown under the MFCQ plus a mild quadratic growth condition.

Throughout this paper, $\Re^n$ denotes the set of n-dimensional column vectors, $\mathcal{S}^n$ denotes the set of $n \times n$ symmetric real matrices, and ^{T} denotes transpose. For $x \in \Re^n$, $\|x\| = \sqrt{x^T x}$. For A and B in $\mathcal{S}^n$, $A \succeq B$ if and only if $A - B \succeq O$ (the zero matrix), i.e., $A - B$ is positive semidefinite. Also, ":=" means "define".

2 Algorithm Description

We are interested in solving the convex minimization problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c_i(x) \leq 0, \quad i = 1, \ldots, m, \end{array} \tag{2.1}$$

where $f : \Re^n \to \Re$ and $c_i : \Re^n \to \Re$, $i = 1, \ldots, m$, are twice continuously differentiable convex functions. For simplicity, we will confine ourselves to the inequality constrained problem, although the subsequent argument can be extended naturally to problems involving additional linear equality constraints. Throughout we make the following assumptions on problem (2.1):

Assumption 1

(a) *Problem* (2.1) *has a nonempty optimal solution set.*

(b) *There exists a* $\bar{x} \in \Re^n$ *satisfying the Slater condition* $c_i(\bar{x}) < 0$, $i = 1, \ldots, m$.

Notice that, for the convex problem (2.1), the Slater condition is equivalent to the MFCQ. Define the ℓ_1 exact penalty function $F_r : \Re^n \to \Re$ for problem (2.1) as

$$F_r(x) := f(x) + r \sum_{i=1}^{m} [c_i(x)]_+,$$

where $r > 0$ is a penalty parameter and $[\cdot]_+$ denotes the projection onto the set of nonnegative real numbers, i.e., $[\cdot]_+ = \max\{0, \cdot\}$. Under Assumption 1, for any $r > 0$ sufficiently large, the set of unconstrained minimizers of F_r coincides with the solution set of problem (2.1) [3, §5.5].

At each iteration of the SQCQP method, given a current iterate $x \in \Re^n$, we solve the following subproblem in d:

$$\begin{array}{ll} \text{minimize} & g(x)^T d + \dfrac{1}{2} d^T B d \\ \text{subject to} & c_i(x) + g_i(x)^T d + \dfrac{\alpha_i}{2} d^T G_i(x) d \leq 0, \quad i = 1, \ldots, m, \end{array} \tag{2.2}$$

where $\alpha_i \in [0, 1]$, $g(x) := \nabla f(x) \in \Re^n$, $g_i(x) := \nabla c_i(x) \in \Re^n$, $G_i(x) := \nabla^2 c_i(x) \in \mathcal{S}^n$, $i = 1, \ldots, m$, and $B \in \mathcal{S}^n$ is positive semidefinite. Subproblem (2.2) is a convex quadratically constrained quadratic program and, as such, can be formulated as a second-order cone program [8, 12],[1] and be solved efficiently by using interior point methods [1, 8, 11, 12, 19].

The nonnegative parameters $(\alpha_i)_{i=1}^m$ are introduced to ensure the feasibility of the subproblem (2.2). Intuitively, if $\alpha_i = 1$, then the local quadratic approximation $c_i(x) + g_i(x)^T d + \frac{1}{2} d^T G_i(x) d$ of $c_i(x + d)$ may be too "aggressive" to admit a feasible solution for (2.2). As an example, for $m = n = 1$ and $c_1(x) = x^4 - \frac{1}{4}$, it is readily checked that (2.2) with $x = 1$ and $\alpha_1 = 1$ has no feasible solution. The less-than-1 weighting factor α_i relaxes the local quadratic cut in order to preserve feasibility. The lemma below quantifies those α_i that ensure the feasibility of (2.2). In particular, at every x, the linearized constraints admit a solution; and if x is nearly feasible or the constraint functions have small curvature at x, then quadratic constraints can be used. Here, each weight α_i is set to be either 0 or 1. We can alternatively set α_i to be the largest value in $[0, 1]$ for which (2.7) below holds (with t chosen as in (a) or (b) or (c) in the lemma). This makes the subproblems harder to solve (more quadratic constraints), but the number of subproblems solved may be less.

[1] In particular, as is noted in [12, p. 221], the quadratic inequality $\|y\|^2 \leq z$ can be rewritten as $\left(\|2y\|^2 + (1 - z)^2\right)^{1/2} \leq 1 + z$ or, equivalently, $(1 + z, 1 - z, 2y)$ belongs to the second-order cone $\{(t, w) \mid \|w\| \leq t\}$.

Lemma 2.1 *Let $\bar{x}$ be the Slater point of Assumption* 1(b). *Fix any $\theta \in [0,1)$. For each $x \in \Re^n$, define $\kappa_i := (\bar{x}-x)^T G_i(x)(\bar{x}-x)$ for all i and*

$$\begin{aligned}
\mathcal{J} &:= \{i \,|\, \theta c_i(\bar{x}) \le c_i(x)\}, && (2.3)\\
s_1 &:= \max_{i\in\mathcal{J}} \frac{c_i(x)-\theta c_i(\bar{x})}{c_i(x)-c_i(\bar{x})}, && (2.4)\\
s_2 &:= \min\left\{\min_{i\in\mathcal{J}} \frac{c_i(x)-c_i(\bar{x})}{\kappa_i}, 1\right\}, && (2.5)\\
s_3 &:= \min\left\{s_2, \min_{i\notin\mathcal{J}} \frac{-2(1-\theta)c_i(\bar{x})}{\kappa_i}\right\}, && (2.6)
\end{aligned}$$

where s_1 is understood to be $-\infty$ if there is no i such that $c_i(x) > 0$, i.e., if x is a feasible solution of problem (2.1), *and the fractions in* (2.5) *and* (2.6) *are understood to be $+\infty$ if $\kappa_i = 0$. Then, the vector $d := t(\bar{x}-x)$ satisfies*

$$c_i(x) + g_i(x)^T d + \frac{\alpha_i}{2} d^T G_i(x) d \;\le\; \theta c_i(\bar{x}), \quad i = 1,\ldots,m, \tag{2.7}$$

under any of the following choices of t and $\alpha_1,\ldots,\alpha_m$:

(a) *Set $t = 1$ and $\alpha_i = 0$ for all $i = 1,\ldots,m$.*

(b) *If $2s_1 < s_2$, set $t = s_2$ and $\alpha_i = 1$ for $i \in \mathcal{J}$ and $\alpha_i = 0$ for $i \notin \mathcal{J}$.*

(c) *If $2s_1 < s_3$, set $t = s_3$ and $\alpha_i = 1$ for all $i = 1,\ldots,m$.*

Proof. From (2.5) and (2.6), we see that $0 < s_3 \le s_2 \le 1$, so that $0 < t \le 1$ for any choice of t in (a) or (b) or (c). The convexity of functions c_i, $i = 1,\ldots,m$, then yields

$$\begin{aligned}
c_i(x) + g_i(x)^T d + \frac{\alpha_i}{2} d^T G_i(x) d &= (1-t)c_i(x) + t(c_i(x) + g_i(x)^T(\bar{x}-x)) + t^2\frac{\alpha_i\kappa_i}{2}\\
&\le (1-t)\eta_i + t\bar{\eta}_i + t^2\frac{\alpha_i\kappa_i}{2},
\end{aligned}$$

where for simplicity we denote $\eta_i := c_i(x)$ and $\bar{\eta}_i := c_i(\bar{x})$. Let $R_i(t,\alpha_i)$ denote the right-hand side quadratic polynomial in the above inequality. We will show that $R_i(t,\alpha_i) \le \theta\bar{\eta}_i$ for any of the choice of t and α_i's in (a)–(c). Clearly, this, together with the above inequality, immediately proves the lemma.

We need to first develop two inequalities. Fix some $i \in \mathcal{J}$. It follows from (2.3) that $\eta_i - \bar{\eta}_i > 0$. Moreover, we have

$$R_i(t,\alpha_i) - \theta\bar{\eta}_i = (1-t)\eta_i + t\bar{\eta}_i + t^2\frac{\alpha_i\kappa_i}{2} - \theta\bar{\eta}_i$$

$$
\begin{aligned}
&= \eta_i - \theta\bar{\eta}_i + t(\bar{\eta}_i - \eta_i) + t^2\frac{\alpha_i\kappa_i}{2} \\
&= (\eta_i - \bar{\eta}_i)\left[\frac{\eta_i - \theta\bar{\eta}_i}{\eta_i - \bar{\eta}_i} - t + t^2\frac{\alpha_i\kappa_i}{2(\eta_i - \bar{\eta}_i)}\right] \\
&\leq (\eta_i - \bar{\eta}_i)\left[s_1 - t + t^2\frac{\alpha_i}{2s_2}\right], \quad \forall i \in \mathcal{J}, \qquad (2.8)
\end{aligned}
$$

where the last step follows from (2.4) and (2.5). Next we consider any $i \notin \mathcal{J}$. By (2.3) we have $\eta_i < \theta\bar{\eta}_i$. Then, for $t \in (0, 1]$, we have

$$
\begin{aligned}
R_i(t, \alpha_i) - \theta\bar{\eta}_i &= (1-t)\eta_i + t\bar{\eta}_i + t^2\frac{\alpha_i\kappa_i}{2} - \theta\bar{\eta}_i \\
&\leq (1-t)\theta\bar{\eta}_i + t\bar{\eta}_i + t^2\frac{\alpha_i\kappa_i}{2} - \theta\bar{\eta}_i \\
&= t(1-\theta)\bar{\eta}_i + t^2\frac{\alpha_i\kappa_i}{2} \\
&\leq -(1-\theta)\bar{\eta}_i\left[-t + t^2\frac{\alpha_i}{s_3}\right], \quad \forall i \notin \mathcal{J}, \qquad (2.9)
\end{aligned}
$$

where the last step is due to (2.6). We now consider the three choices of t and α_i's in (a)–(c).

(a) By (2.3) and (2.4), we have $s_1 < 1$. Setting $t = 1$ and $\alpha_i = 0$ in both (2.8) and (2.9), we see that $R_i(1, 0) - \theta\bar{\eta}_i < 0$ for all i.

(b) Suppose $2s_1 < s_2$, and let $t = s_2 \in (0, 1]$ and $\alpha_i = 1$ for $i \in \mathcal{J}$ and $\alpha_i = 0$ for $i \notin \mathcal{J}$. Then, we substitute these choices of t and α_i's into the inequality (2.8) to obtain

$$
R_i(s_2, 1) - \theta\bar{\eta}_i \leq (\eta_i - \bar{\eta}_i)\left[s_1 - s_2 + s_2^2\frac{1}{2s_2}\right] = (\eta_i - \bar{\eta}_i)\left[s_1 - \frac{s_2}{2}\right] < 0, \quad \forall i \in \mathcal{J},
$$

where we used $2s_1 < s_2$, and into the inequality (2.9) to obtain

$$
R_i(s_2, 0) - \theta\bar{\eta}_i \leq -(1-\theta)\bar{\eta}_i\,[-s_2] < 0, \quad \forall i \notin \mathcal{J}.
$$

(c) Suppose $2s_1 < s_3$, and let $t = s_3 \in (0, 1]$ and $\alpha_i = 1$ for all i. Then, we substitute these choices of t and α_i's into the inequality (2.8) to obtain

$$
R_i(s_3, 1) - \theta\bar{\eta}_i \leq (\eta_i - \bar{\eta}_i)\left[s_1 - s_3 + s_3^2\frac{1}{2s_2}\right] \leq (\eta_i - \bar{\eta}_i)\left[s_1 - \frac{s_3}{2}\right] < 0, \quad \forall i \in \mathcal{J},
$$

where we used $s_3 \leq s_2$, $2s_1 < s_3$, and into the inequality (2.9) to obtain

$$
R_i(s_3, 1) - \theta\bar{\eta}_i \leq -(1-\theta)\bar{\eta}_i\,[-s_3 + s_3] = 0, \quad \forall i \notin \mathcal{J}.
$$

This completes the proof of the lemma. □

We remark that Lemma 2.1 offers a specific way of picking the parameters α_i to ensure the feasibility of subproblem (2.2). In particular, given an iterate x, we compute the values of s_1, s_2 and s_3 by (2.4)–(2.6), and determine α_i as follows:

$$\begin{array}{lcl} s_2 \le 2s_1 & \Longrightarrow & \alpha_i = 0,\ i = 1, \ldots, m \\ s_3 \le 2s_1 < s_2 & \Longrightarrow & \alpha_i = \left\{ \begin{array}{ll} 1, & i \in \mathcal{J} \\ 0, & i \notin \mathcal{J} \end{array} \right. \\ s_3 > 2s_1 & \Longrightarrow & \alpha_i = 1,\ i = 1, \ldots, m \end{array} \tag{2.10}$$

Then, by Lemma 2.1, subproblem (2.2) is feasible at every iteration with α_i chosen according to the rule (2.10). In general, linear constraints with $\alpha_i = 0$ may have to be used (case (a)) in the early stage of the SQCQP method in order to maintain subproblem feasibility. However, as we will argue later (see Lemma 4.1), quadratic constraints with $\alpha_i = 1$ (case (c)) can eventually be adopted when the iterates get close to the optimal solution. The intermediate case of mixed linear and quadratic constraints (case (b)) provides a mechanism for quadratic constraints to be gracefully phased into the SQCQP method, thus potentially improving the convergence of this method in practice. However, we can alternatively set $\alpha_i = 0$ for *all* i in this intermediate case without affecting the subsequent theoretical convergence analysis of the SQCQP method.

The choice (2.4)–(2.6) of s_1, s_2, s_3 in (2.10) can be replaced by the alternative choice

$$\begin{aligned} s_1 &:= \max_{i: c_i(x) > 0} \frac{c_i(x)}{c_i(x) - \vartheta c_i(\bar{x})}, \\ s_2 &:= \min \left\{ \min_{i \in \mathcal{J}} \frac{c_i(x) - \vartheta c_i(\bar{x})}{\kappa_i}, 1 \right\}, \\ s_3 &:= \min \left\{ s_2, \min_{i \notin \mathcal{J}} \frac{-2(\vartheta - \theta) c_i(\bar{x})}{\kappa_i} \right\}, \end{aligned}$$

where ϑ is any constant in $(\theta, 1)$. It can be shown that Lemma 2.1 still holds, provided "θ" in (2.7) is replaced by "$t(1-\vartheta)$". This does not affect the subsequent convergence analysis. By choosing ϑ near 1, this alternative choice gives an s_1 that is smaller than (2.4) and s_2, s_3 that are near (2.5), (2.6), so that (2.10) will set more α_i to 1, i.e., more quadratic constraints will be used.

Given its feasibility, subproblem (2.2) has an optimal solution if the objective function is bounded from below over the feasible region; see [9]. This condition is obviously satisfied if either B or at least one of $G_i(x)$, $i = 1, \ldots, m$, is positive definite. In our global convergence analysis (Theorem 4.1), we will essentially assume that B is positive definite.

Let d be an optimal solution of subproblem (2.2). Then the next iterate x^{new} is generated by

$$x^{new} := x + \beta d,$$

where $\beta \in \Re$ is the step size. To ensure global convergence, we choose $\beta > 0$ small enough to satisfy

$$F_r(x + \beta d) - F_r(x) \leq \sigma\beta \left(\bar{F}_r(x, d, \alpha) - F_r(x)\right), \tag{2.11}$$

where $\sigma \in (0, 1)$ is a user chosen constant, $\alpha = (\alpha_i)_{i=1}^m$, and

$$\bar{F}_r(x, d, \alpha) := f(x) + g(x)^T d + \frac{1}{2} d^T G(x) d + r \sum_{i=1}^m \left[c_i(x) + g_i(x)^T d + \frac{\alpha_i}{2} d^T G_i(x) d \right]_+, \tag{2.12}$$

with $G(x) := \nabla^2 f(x)$. Notice that the weights $\alpha_1, \ldots, \alpha_m$ given by (2.10) depend on the current iterate x as well as the Slater point $\bar{x}$. Since $c_i(x) + g_i(x)^T d + \frac{1}{2}\alpha_i d^T G_i(x) d \leq 0$, we have

$$\bar{F}_r(x, d, \alpha) - F_r(x) = g(x)^T d + \frac{1}{2} d^T G(x) d - r \sum_{i=1}^m [c_i(x)]_+. \tag{2.13}$$

By (2.7), subproblem (2.2) has a feasible solution satisfying the Slater condition. Then it is known [18, Theorem 28.2] that there exists a vector of Lagrange multipliers $v = (v_1, \ldots, v_m)^T$ satisfying the KKT conditions for subproblem (2.2), namely,

$$\begin{aligned} & Bd + g(x) + \sum_{i=1}^m v_i(\alpha_i G_i(x) d + g_i(x)) = 0, \\ & c_i(x) + g_i(x)^T d + \frac{\alpha_i}{2} d^T G_i(x) d \leq 0, \quad v_i \geq 0, \\ & v_i \Big(c_i(x) + g_i(x)^T d + \frac{\alpha_i}{2} d^T G_i(x) d \Big) = 0, \quad i = 1, \ldots, m. \end{aligned} \tag{2.14}$$

We call (d, v) a *KKT pair* of subproblem (2.2).

We state our method formally below:

SQCQP Method

Step 0: Choose an initial point $x^1 \in \Re^n$, an initial penalty parameter $r^0 \in (0, +\infty)$, constants $\gamma \in (0, 1)$, $\delta \in (0, +\infty)$, $\theta \in [0, 1)$ and $\sigma \in (0, 1)$, and a point $\bar{x} \in \Re^n$ such that $c_i(\bar{x}) < 0$, $i = 1, \ldots, m$. Initialize $k = 1$.

Step 1: Choose a positive semidefinite matrix $B^k \in \mathcal{S}^n$. Determine $\alpha^k = (\alpha_i^k)_{i=1}^m$ according to the rule (2.10) with $x = x^k$, and solve subproblem (2.2) associated with x^k, B^k, α^k to obtain a KKT pair $(d^k, v^k) \in \Re^n \times \Re^m$. If $d^k = 0$, terminate.

Step 2: Update the penalty parameter as

$$r^k := \begin{cases} r^{k-1} & \text{if } r^{k-1} \geq \max_i v_i^k + \delta; \\ \max_i v_i^k + 2\delta & \text{otherwise.} \end{cases} \tag{2.15}$$

Let β^k be the largest $\beta \in \{1, \gamma, (\gamma)^2, (\gamma)^3, \ldots\}$ satisfying

$$F_{r^k}(x^k + \beta d^k) - F_{r^k}(x^k) \leq \sigma\beta \left(\bar{F}_{r^k}(x^k, d^k, \alpha^k) - F_{r^k}(x^k) \right). \tag{2.16}$$

Update $x^{k+1} := x^k + \beta^k d^k$, increment k by 1, and go to Step 1.

The updating rule (2.15) maintains the penalty parameter r^k large enough so to guarantee d^k to be a descent direction for the function F_{r^k}, while allowing for r^k to eventually stay constant as the iterates x^k converge to a solution; see Lemma 3.1 below.

The termination criterion in Step 1 is justified by the following lemma.

Lemma 2.2 *For any $x \in \Re^n$ and $\alpha_i \geq 0$, $i = 1, \ldots, m$, suppose that subproblem (2.2) has a KKT pair (d, v). If $d = 0$, then (x, v) is a KKT pair for problem (2.1), and hence x is an optimal solution of problem (2.1).*

Proof. This can easily be verified by substituting $d = 0$ in (2.14). □

3 Global Convergence

In this section, we analyze the global convergence of the SQCQP method. We begin with the following lemma giving conditions for an optimal solution d of (2.2) to be a descent direction for the penalty function F_r at x.

Lemma 3.1 *For any $x \in \Re^n$ and $\alpha = (\alpha_i)_{i=1}^m \geq 0$, suppose that subproblem (2.2) has a KKT pair (d, v). Also, suppose there exists a $\mu > 0$ such that*

$$2B - G(x) + \sum_{i=1}^{m} \alpha_i v_i G_i(x) \succeq \mu I \tag{3.1}$$

and $r \geq \max_i v_i$. Then

$$\bar{F}_r(x, d, \alpha) - F_r(x) \leq -\frac{1}{2}\mu \|d\|^2. \tag{3.2}$$

Proof. We have from (2.14) that

$$\begin{aligned} g(x)^T d + \frac{1}{2} d^T G(x) d &= -\Big(Bd + \sum_{i=1}^m v_i(\alpha_i G_i(x)d + g_i(x))\Big)^T d + \frac{1}{2} d^T G(x) d \\ &= -d^T \Big(B + \sum_{i=1}^m \alpha_i v_i G_i(x) - \frac{1}{2} G(x)\Big) d - \sum_{i=1}^m v_i g_i(x)^T d \\ &= -\frac{1}{2} d^T \Big(2B + \sum_{i=1}^m \alpha_i v_i G_i(x) - G(x)\Big) d + v^T c(x), \end{aligned} \tag{3.3}$$

where $c(x) = (c_1(x), \ldots, c_m(x))^T$. Then, from (2.13) we obtain

$$\begin{aligned} \bar{F}_r(x, d, \alpha) - F_r(x) &= g(x)^T d + \frac{1}{2} d^T G(x) d - r \sum_{i=1}^m [c_i(x)]_+ \\ &= -\frac{1}{2} d^T \Big(2B - G(x) + \sum_{i=1}^m \alpha_i v_i G_i(x)\Big) d + v^T c(x) - r \sum_{i=1}^m [c_i(x)]_+ \\ &\le -\frac{1}{2} \mu \|d\|^2 - \sum_{i=1}^m (r - v_i)[c_i(x)]_+ \\ &\le -\frac{1}{2} \mu \|d\|^2, \end{aligned}$$

where the second equality follows from (3.3), the first inequality is due to (3.1) and $v^T c(x) \le \sum_i v_i [c_i(x)]_+$, and the last inequality is due to $r \ge \max_i v_i$. □

Since each $G_i(x)$ is positive semidefinite and v_i and α_i are nonnegative, condition (3.1) is satisfied if B is chosen to satisfy

$$2B \succeq G(x) + \mu I$$

for some constant $\mu > 0$.

The next lemma shows that the line search for the penalty function F_r at x along the direction d is well defined.

Lemma 3.2 *Suppose that the conditions in Lemma* 3.1 *are satisfied. If* $d \neq 0$, *then the descent condition* (2.11) *holds for all* $\beta > 0$ *small enough.*

Proof. Since each $G_i(x)$ is positive semidefinite, then

$$c_i(x) + g_i(x)^T d \;\le\; c_i(x) + g_i(x)^T d + \frac{\alpha_i}{2} d^T G_i(x) d \;\le\; 0, \qquad i = 1, \ldots, m.$$

Thus, for any $\beta \in [0, 1]$, we have

$$c_i(x) + \beta g_i(x)^T d \le (1 - \beta) c_i(x),$$

implying

$$[c_i(x) + \beta g_i(x)^T d]_+ \leq (1-\beta)[c_i(x)]_+, \qquad i = 1, \ldots, m.$$

As a result, we obtain

$$\sum_{i=1}^{m} \Big([c_i(x) + \beta g_i(x)^T d]_+ - [c_i(x)]_+ \Big) \leq -\beta \sum_{i=1}^{m} [c_i(x)]_+. \tag{3.4}$$

Then, using the Taylor series approximation of $f(x+\beta d)$ and $c_i(x+\beta d)$ at x, we have

$$\begin{aligned}
& F_r(x+\beta d) - F_r(x) \\
= \;& \beta g(x)^T d + \frac{\beta^2}{2} d^T G(x) d + o(\beta^2 \|d\|^2) \\
& + r \sum_{i=1}^{m} \Big[c_i(x) + \beta g_i(x)^T d + O(\beta^2 \|d\|^2) \Big]_+ - r \sum_{i=1}^{m} [c_i(x)]_+ \\
= \;& \beta g(x)^T d + \frac{\beta^2}{2} d^T G(x) d + r \sum_{i=1}^{m} \Big([c_i(x) + \beta g_i(x)^T d]_+ - [c_i(x)]_+ \Big) + O(\beta^2 \|d\|^2) \\
\leq \;& \beta g(x)^T d + \frac{\beta^2}{2} d^T G(x) d - \beta r \sum_{i=1}^{m} [c_i(x)]_+ + O(\beta^2 \|d\|^2),
\end{aligned}$$

where the second equality uses the Lipschitzian property of $[\cdot]_+$ and the inequality uses (3.4). Since $G(x)$ is positive semidefinite and $0 < \beta \leq 1$, the above relation implies

$$\begin{aligned}
F_r(x+\beta d) - F_r(x) &\leq \beta \Big(g(x)^T d + \frac{1}{2} d^T G(x) d - r \sum_{i=1}^{m} [c_i(x)]_+ \Big) + O(\beta^2 \|d\|^2) \\
&= \beta \Big(\bar{F}_r(x, d, \alpha) - F_r(x) \Big) + O(\beta^2 \|d\|^2),
\end{aligned}$$

where the equality follows from (2.13). Then by (3.2) and $\sigma \in (0,1)$, we have

$$F_r(x+\beta d) - F_r(x) \leq \sigma \beta \left(\bar{F}_r(x, d, \alpha) - F_r(x) \right)$$

for all sufficiently small $\beta > 0$. □

The next lemma shows that a KKT pair (d, v) obtained by solving subproblem (2.2) is bounded when (x, B) lies in a bounded set.

Lemma 3.3 *Let X be any nonempty bounded subset of $\Re^n$. Let $\Omega := \{B \in \mathcal{S}^n \,|\, \mu_1 I \succeq B \succeq \mu_2 I\}$ for some constants $\mu_1 \geq \mu_2 > 0$. Then there exists a bounded subset $D \times V$ of $\Re^n \times \Re^m$ such that, for any $(x, B) \in X \times \Omega$ and any KKT pair (d, v) of subproblem (2.2) with $(\alpha_i)_{i=1}^m$ given by (2.10), we have $(d, v) \in D \times V$.*

Proof. First we show the boundedness of d. Fix any $x \in X$ and $B \in \Omega$. Since $(\alpha_i)_{i=1}^m$ is given by (2.10), we see from Lemma 2.1 that there exists a $t \in (0,1]$ such that $t(\bar{x} - x)$ is a feasible solution of subproblem (2.2). So d satisfies

$$\begin{aligned} -\|g(x)\|\|d\| + \frac{\mu_2}{2}\|d\|^2 &\leq g(x)^T d + \frac{1}{2} d^T B d \\ &\leq t g(x)^T (\bar{x} - x) + \frac{t^2}{2} (\bar{x} - x)^T B (\bar{x} - x) \\ &\leq \|g(x)\| \, \|\bar{x} - x\| + \frac{\mu_1}{2} \|\bar{x} - x\|^2, \end{aligned}$$

where the first and last inequalities follow from $B \in \Omega$ and $t \in (0,1]$. Notice that the right-hand side of this inequality, as well as $\|g(x)\|$ on the left-hand side, are bounded for $x \in X$. This implies that d is bounded for $x \in X$. Also, $B \in \Omega$ is bounded since B is symmetric and its eigenvalues are bounded.

Next we show the boundedness of v. Suppose to the contrary that there exists a sequence of KKT pair $\{(d^k, v^k)\}$ of subproblem (2.2) associated with some $(x^k, B^k) \in X \times \Omega$ and $(\alpha_i^k)_{i=1}^m$ given by (2.10) for x^k, $k = 1, 2, \ldots$, such that $\|v^k\| \to +\infty$. By passing to a subsequence if necessary, we may assume that (x^k, B^k, d^k) converges to some $(x^\infty, B^\infty, d^\infty)$, and $v^k / \|v^k\| \to w$ with $\|w\| = 1$ and $w \geq 0$. By the KKT conditions (2.14), we have for each k that

$$\begin{aligned} & B^k d^k + g(x^k) + \sum_{i=1}^m v_i^k \nabla q_i^k(d^k) = 0, \\ & q_i^k(d^k) \leq 0, \quad v_i^k \geq 0, \quad v_i^k q_i^k(d^k) = 0, \qquad i = 1, \ldots, m, \end{aligned} \tag{3.5}$$

where $q_i^k : \Re^n \to \Re$ are quadratic functions defined by

$$q_i^k(d) := c_i(x^k) + g_i(x^k)^T d + \frac{\alpha_i^k}{2} d^T G_i(x^k) d.$$

By further passing to a subsequence if necessary, we may assume that α_i^k converges to some $\alpha_i^\infty \in \{0, 1\}$ for each i. Define the quadratic functions $q_i^\infty : \Re^n \to \Re$, $i = 1, \ldots, m$, by

$$q_i^\infty(d) = c_i(x^\infty) + g_i(x^\infty)^T d + \frac{\alpha_i^\infty}{2} d^T G_i(x^\infty) d.$$

Then, upon dividing both sides of the first equality in (3.5) by $\|v^k\|$ and taking limit, we obtain

$$\sum_{i=1}^m w_i \nabla q_i^\infty(d^\infty) = 0. \tag{3.6}$$

Moreover, the rest of (3.5) implies $q_i^\infty(d^\infty) \leq 0$, $i = 1, \ldots, m$, and

$$w_i > 0 \quad \Longrightarrow \quad q_i^\infty(d^\infty) = 0. \tag{3.7}$$

Note that there is at least one index i such that $w_i > 0$.

By Lemma 2.1, for each k, there exists $t^k \in (0,1]$ such that $\hat{d}^k := t^k(\bar{x} - x^k)$ satisfies

$$q_i^k(\hat{d}^k) \leq \theta c_i(\bar{x}), \quad i = 1, \ldots, m.$$

Then, any accumulation point $\hat{d}$ of $\{\hat{d}^k\}$ satisfies

$$q_i^\infty(\hat{d}) \leq \theta c_i(\bar{x}) < 0, \quad i = 1, \ldots, m.$$

Then, for every i such that $w_i > 0$, (3.7) implies $q_i^\infty(d^\infty) = 0$ so that the convexity of q_i^∞ yields

$$\nabla q_i^\infty(d^\infty)^T(\hat{d} - d^\infty) \leq q_i^\infty(\hat{d}) - q_i^\infty(d^\infty) < 0.$$

Since $w \geq 0$ and there is at least one index i such that $w_i > 0$, this implies

$$\sum_{i=1}^m w_i \nabla q_i^\infty(d^\infty)^T(\hat{d} - d^\infty) < 0.$$

This contradicts (3.6), thus completing the proof. □

It is well known that, for a convex program, the Slater condition implies the boundedness of the Lagrange multipliers. Thus, Lemma 3.3 can be viewed as a generalization of this result to the uniform boundedness of the multipliers for all the subproblems which are convex and have nonempty interior. Now we are ready to establish a global convergence result for the SQCQP method.

Theorem 3.1 *Let $\{x^k\}, \{B^k\}, \{\alpha^k\}, \{v^k\}$ be generated by the SQCQP method. Suppose that $\{x^k\}$ is bounded. Also, suppose that*

$$2B^k - G(x^k) + \sum_{i=1}^m \alpha_i^k v_i^k G_i(x^k) \succeq \mu I \quad \textit{and} \quad \mu_1 I \succeq B^k \succeq \mu_2 I$$

for all k, where $\mu > 0$ and $\mu_1 \geq \mu_2 > 0$ are some constants. Then the SQCQP method either terminates at an optimal solution of problem (2.1) or generates an infinite sequence $\{x^k\}$ of which every accumulation point is an optimal solution of problem (2.1).

Proof. For each k, by Lemma 2.1 and the choice of α^k, subproblem (2.1) associated with x^k, B^k, α^k is feasible and, by B^k being positive definite, has a unique optimal solution d^k. Since $r^k \geq \max_i v_i^k$ by (2.15), Lemma 3.2 implies β^k in the line search is well defined. If the method terminates at

some iteration k, then Lemma 2.2 implies that the iterate x^k is an optimal solution of problem (2.1). Suppose that the method generates an infinite sequence $\{x^k\}$. Since $\{x^k\}$ is assumed to be bounded and $\mu_1 I \succeq B^k \succeq \mu_2 I$ for all k, Lemma 3.3 implies that $\{(d^k, v^k)\}$ is also bounded, which in particular implies that the penalty parameter r^k is constant for all k sufficiently large. Moreover, we have from Lemma 3.1 that

$$\bar{F}_{r^k}(x^k, d^k, \alpha^k) - F_{r^k}(x^k) \leq -\frac{1}{2}\mu\|d^k\|^2 \tag{3.8}$$

for all k.

We consider two cases: (i) The sequence of step sizes $\{\beta^k\}$ is bounded away from zero, say by $\hat{\beta} > 0$; and (ii) $\{\beta^k\}$ contains a subsequence converging to zero.

In case (i), since (2.16) holds with $\beta = \beta^k$, (3.8) yields

$$F_{r^k}(x^k + \beta^k d^k) - F_{r^k}(x^k) \leq -\frac{1}{2}\sigma\hat{\beta}\mu\|d^k\|^2$$

for all k. Since $\{x^k\}$ is bounded so that $\{F_{r^k}(x^k)\}$ is bounded from below, this inequality implies that $\{d^k\}$ tends to 0. Since (d^k, v^k) satisfies the KKT conditions (2.14) for subproblem (2.2) associated with x^k, B^k, α^k, by taking the limit as $k \to \infty$ and using the boundedness of $\{x^k\}$, $\{B^k\}$, $\{v^k\}$ and $\{\alpha^k\}$, we see that any accumulation point (x^*, v^*) of $\{(x^k, v^k)\}$ satisfies the KKT conditions for problem (2.1). Since (2.1) is a convex program, x^* is an optimal solution of (2.1).

In case (ii), there exists a subsequence $\{\beta^k\}_{k\in K}$ converging to 0, where $K \subset \{0, 1, \ldots\}$. Since $\{(x^k, d^k, \alpha^k)\}$ is bounded and $\{r^k\}$ has a constant tail, by further passing to a subsequence if necessary, we can assume that $\{(x^k, d^k, \alpha^k)\}_{k\in K}$ converges to some limit (x^*, d^*, α^*) and $r^k = r$, $\beta^k < 1$ for all $k \in K$, where $r > 0$ is a constant. For each $k \in K$, since $\beta^k < 1$ and $r^k = r$, the Armijo-type line search rule in Step 2 implies

$$F_r(x^k + (\beta^k/\gamma)d^k) - F_r(x^k) > \sigma(\beta^k/\gamma)\left(\bar{F}_r(x^k, d^k, \alpha^k) - F_r(x^k)\right). \tag{3.9}$$

Using the simple property $[u_1 + u_2]_+ - [u_1]_+ \leq [u_2]_+$ for any real numbers u_1, u_2, we deduce

$$\begin{aligned}[c_i(x + \beta d)]_+ - [c_i(x)]_+ &= [c_i(x) + \beta\nabla c_i(x)^T d + o_i(x; \beta d)]_+ - [c_i(x)]_+ \\ &\leq \beta[\nabla c_i(x)^T d]_+ + [o_i(x; \beta d)]_+,\end{aligned}$$

for any i, x, d and $\beta > 0$, with $\lim_{(x,d,\beta)\to(x^*,d^*,0^+)} o_i(x; \beta d)/\beta = 0$. Dividing both sides of the above inequality by β and taking limit give

$$[\nabla c_i(x^*)^T d^*]_+ \geq \limsup_{(x,d,\beta)\to(x^*,d^*,0^+)} ([c_i(x + \beta d)]_+ - [c_i(x)]_+)/\beta, \quad \text{for all } i.$$

Notice that the directional derivative of $[c_i(x)]_+$ at x^* in the direction of d^* is $[\nabla c_i(x^*)^T d^*]_+$ if $c_i(x^*) = 0$, and is zero if $c_i(x^*) < 0$. Combining this with the above relation, we obtain

$$F_r'(x^*; d^*) = \nabla f(x^*)^T d^* + \sum_{\{i: c_i(x^*)=0\}} [\nabla c_i(x^*)^T d^*]_+ \geq \limsup_{(x,d,\beta)\to(x^*,d^*,0^+)} (F_r(x+\beta d) - F_r(x))/\beta,$$

where $F_r'(x^*; d^*)$ denotes the directional derivative of F_r at x^* in the direction d^*. Thus, dividing both sides of (3.9) by β^k/γ and taking limit as $k \in K$, $k \to \infty$, yield

$$F_r'(x^*; d^*) \geq \sigma \left(\bar{F}_r(x^*, d^*, \alpha^*) - F_r(x^*)\right). \tag{3.10}$$

On the other hand, we note from (2.12) that $\bar{F}_r(x^*, 0, \alpha^*) = F_r(x^*)$ and $\bar{F}_r(x^*, d, \alpha^*)$ is convex in d and its directional derivative at $d = 0$ in the direction d^* coincides with $F_r'(x^*; d^*)$. Therefore

$$F_r'(x^*; d^*) \leq \bar{F}_r(x^*, d^*, \alpha^*) - F_r(x^*). \tag{3.11}$$

Combining (3.10) and (3.11), and using $\sigma \in (0, 1)$, we obtain

$$\bar{F}_r(x^*, d^*, \alpha^*) - F_r(x^*) \geq 0.$$

Also, using $r^k = r$ and passing to the limit in (3.8) as $k \in K$, $k \to \infty$, yield

$$\bar{F}_r(x^*, d^*, \alpha^*) - F_r(x^*) \leq -\frac{1}{2}\mu \|d^*\|^2.$$

The last two inequalities imply $d^* = 0$. Hence, by the same reasoning as in case (i), we obtain that x^* is an optimal solution of problem (2.1). □

4 Local Quadratic Rate of Convergence

In this section, we analyze the local quadratic convergence of the SQCQP method under the following further assumptions:

Assumption 2

(a) *Problem* (2.1) *has an optimal solution* x^* *satisfying*

$$H(x^*, v^*) \succeq \mu^* I, \quad \forall v^* \in V^*$$

for some constant $\mu^* > 0$, *where* $H(x, v) := G(x) + \sum_{i=1}^m v_i G_i(x)$ *and* $V^* := \{v^* \in \Re^m \mid (x^*, v^*)$ *is a KKT pair of problem* (2.1)$\}$.[2]

[2] V^* is nonempty and bounded by Assumption 1 and [18, Theorem 28.2].

(b) *G and $G_1, \ldots, G_m$ are Lipschitz continuous in a neighborhood of x^*.*

In addition, we will choose $B = G(x)$ in the SQCQP method, so subproblem (2.2) becomes

$$\begin{aligned} &\text{minimize} \quad g(x)^T d + \frac{1}{2} d^T G(x) d \\ &\text{subject to} \quad c_i(x) + g_i(x)^T d + \frac{\alpha_i}{2} d^T G_i(x) d \le 0, \quad i = 1, \ldots, m. \end{aligned} \tag{4.1}$$

Assumption 2 (a) implies that x^* is the unique solution of the convex program (2.1).

The next lemma shows that, for x near x^*, all constraints in subproblem (4.1) use the second-order information.

Lemma 4.1 *If x is sufficiently close to x^*, then $(\alpha_i)_{i=1}^m$ given by (2.10) satisfies $\alpha_i = 1$ for all i and subproblem (4.1) has at least one KKT pair.*

Proof. Since $\alpha_1, \ldots, \alpha_m$ are chosen by (2.10), Lemma 2.1 implies (4.1) is feasible for any x. Suppose x is close to the optimal solution x^* (thus nearly feasible). Let s_1, s_3 be given by (2.4)–(2.6). Then, s_1 is close to 0 if x is infeasible, and $s_1 = -\infty$ if x is feasible. Moreover, in a neighborhood of x^*, s_3 is bounded from below by a positive constant. Therefore there exists a neighborhood X of x^* such that, for $x \in X$, we have $2s_1 \le s_3$ and the rule (2.10) yields $\alpha_i = 1$ for all i.

Fix any $v^* \in V^* \neq \emptyset$. Assumption 2 (a) and the continuity property of $G, G_1, \ldots, G_m$ and their eigenvalues imply $H(x, v^*)$ is positive definite for all $x \in X$ sufficiently close to x^*. Then $d = 0$ is the only solution of $G(x)d = 0$, $G_i(x)d = 0$, $i = 1, \ldots, m$, from which we can deduce that the objective function of (4.1) is bounded from below on the feasible region. This implies (4.1) has an optimal solution [9]. In addition, Lemma 2.1 ensures that the Slater condition (2.7) holds with $d = s_3(\bar{x} - x)$, so (4.1) has at least one Lagrange multiplier vector by [18, Theorem 28.2]. □

By Lemma 4.1, for all x sufficiently close to x^*, subproblem (4.1) reduces to

$$\begin{aligned} &\text{minimize} \quad g(x)^T d + \frac{1}{2} d^T G(x) d \\ &\text{subject to} \quad c_i(x) + g_i(x)^T d + \frac{1}{2} d^T G_i(x) d \le 0, \quad i = 1, \ldots, m. \end{aligned} \tag{4.2}$$

Then, (d, v) forms a KKT pair for (4.2) if and only if

$$g(x) + G(x)d + \sum_{i \in \mathcal{I}(x;d)} v_i (g_i(x) + G_i(x)d) = 0, \tag{4.3}$$

$$c_i(x) + g_i(x)^T d + \frac{1}{2} d^T G_i(x) d = 0, \quad v_i \geq 0, \quad i \in \mathcal{I}(x;d), \tag{4.4}$$

$$c_i(x) + g_i(x)^T d + \frac{1}{2} d^T G_i(x) d < 0, \quad v_i = 0, \quad i \notin \mathcal{I}(x;d), \tag{4.5}$$

for some $\mathcal{I}(x;d) \subseteq \{1, \ldots, m\}$.

Lemma 4.2 *Problem* (4.2) *with* $x = x^*$ *has the unique optimal solution* $d = 0$.

Proof. Fix any $v^* \in V^* \neq \emptyset$. Then, it is readily verified using (4.3)–(4.5) that $d = 0$ together with v^* satisfies the KKT conditions for subproblem (4.2) with $x = x^*$. Since (4.2) is a convex program, this shows $d = 0$ is an optimal solution. To show uniqueness, consider any optimal solution d^* of (4.2) with $x = x^*$. Then d^* is a feasible solution with objective function value of zero, so that

$$g(x^*)^T d^* + \frac{1}{2}(d^*)^T G(x^*) d^* = 0, \tag{4.6}$$

$$g_i(x^*)^T d^* + \frac{1}{2}(d^*)^T G_i(x^*) d^* \leq 0, \quad i \in \mathcal{I}^*, \tag{4.7}$$

where $\mathcal{I}^* := \{i \,|\, c_i(x^*) = 0\}$. Multiplying (4.7) by v_i^* (which is nonnegative) and adding to (4.6) yield

$$\begin{aligned} 0 &\geq g(x^*)^T d^* + \frac{1}{2}(d^*)^T G(x^*) d^* + \sum_{i \in \mathcal{I}^*} v_i^* \left(g_i(x^*)^T d^* + \frac{1}{2}(d^*)^T G_i(x^*) d^* \right) \\ &= \frac{1}{2}(d^*)^T H(x^*, v^*) d^*, \end{aligned}$$

where the equality uses $g(x^*) + \sum_{i=1}^m v_i^* g_i(x^*) = 0$ and $v_i^* = 0$ for $i \notin \mathcal{I}^*$. Since $H(x^*, v^*) \succ 0$ by Assumption 2 (a), this implies $d^* = 0$. □

Lemma 4.3 *There exist a neighborhood* X *of* x^* *and a constant* $\rho_1 > 0$ *such that, for each* $x \in X$, *problem* (4.2) *has a KKT pair and each such KKT pair* (d, v) *satisfies* $\|d\| \leq \rho_1$ *and* $\|v\| \leq \rho_1$.

Proof. By Lemma 4.1, there exists a neighborhood X_0 of x^* such that, for each $x \in X_0$, (4.2) has a KKT pair. Moreover, the proof of Lemma 4.1 shows that the vector $s_3(\bar{x} - x)$ is a feasible solution of (4.2), where s_3 is given by (2.5)–(2.6). For each KKT pair (d, v) of (4.2), since d is an optimal solution of (4.2), we then have

$$\begin{aligned} g(x)^T d + \frac{1}{2} d^T G(x) d &\leq s_3 g(x)^T (\bar{x} - x) + \frac{(s_3)^2}{2} (\bar{x} - x)^T G(x) (\bar{x} - x) \\ &\leq \|g(x)\| \|\bar{x} - x\| + \frac{1}{2} (\bar{x} - x)^T G(x) (\bar{x} - x), \end{aligned}$$

where the last inequality follows from $0 < s_3 \leq 1$. Moreover, the feasibility of d for (4.2) implies

$$c_i(x) + g_i(x)^T d + \frac{1}{2} d^T G_i(x) d \leq 0, \quad i = 1, \ldots, m.$$

Fix any $v^* \in V^* \neq \emptyset$. Then the above inequalities and $v^* \geq 0$ yield

$$\begin{aligned} \sum_{i=1}^m v_i^* c_i(x) + \left(g(x) + \sum_{i=1}^m v_i^* g_i(x) \right)^T d + \frac{1}{2} d^T H(x, v^*) d \\ \leq \|g(x)\| \|\bar{x} - x\| + \frac{1}{2} (\bar{x} - x)^T G(x) (\bar{x} - x). \end{aligned} \tag{4.8}$$

By Assumption 2 (a) and the continuity property of $G, G_1, \ldots, G_m$ and their eigenvalues, there exists a neighborhood $X \subseteq X_0$ of x^* such that

$$H(x, v^*) \succeq \frac{\mu^*}{2} I$$

for every $x \in X$. Thus, the left-hand side of (4.8) is a convex quadratic function of d with bounded coefficients and uniformly positive definite Hessian. Since the right-hand side of (4.8) is clearly bounded for $x \in X$, this implies the boundedness of d.

By using Lemma 4.1, the boundedness of v can be proved in the same manner as in Lemma 3.3 and thus is omitted. □

In the next lemma, we show that the solution d of subproblem (4.2) tends to zero as the iterate x approaches x^*. In what follows, for any nonempty closed set $V \subseteq \Re^m$, we denote $\mathrm{dist}(v, V) := \min_{v' \in V} \|v - v'\|$.

Lemma 4.4 *There exist a neighborhood X of x^* and a constant $\rho_2 > 0$ such that for each $x \in X$ and each KKT pair (d, v) of (4.2), conditions (4.3)–(4.5) hold for some $\mathcal{I}(x; d) \subseteq \{1, \ldots, m\}$ and*

$$\max\{\|d\|, \mathrm{dist}(v, V^*)\} \leq \rho_2 \|x - x^*\|. \tag{4.9}$$

Proof. Let X be the neighborhood from Lemma 4.3. Consider any sequence $\{x^k\} \subset X$ converging to x^*. Let (d^k, v^k) be any KKT pair of (4.2) with $x = x^k$. By Lemma 4.3, $\{(d^k, v^k)\}$ is bounded. Then it follows from (4.3)–(4.5) that every accumulation point of $\{(d^k, v^k)\}$ is a KKT pair for (4.2) with $x = x^*$. By Lemma 4.2, $d = 0$ is the unique optimal solution of (4.2) with $x = x^*$. Hence it follows that $d^k \to 0$. For any subset $\mathcal{I}$ of $\{1, \ldots, m\}$, let $V_{\mathcal{I}}$ denote the set of vectors v satisfying

$$g(x^*) + \sum_{i \in \mathcal{I}} v_i g_i(x^*) = 0, \quad v_i \geq 0, \quad i \in \mathcal{I}, \quad v_i = 0, \quad i \notin \mathcal{I}. \tag{4.10}$$

Then we can deduce that, if $\mathcal{I}(x^k; d^k) = \mathcal{I}$ holds for an infinite number of k's, then $\mathcal{I} \subseteq \mathcal{I}^* := \{i \,|\, c_i(x^*) = 0\}$ and any accumulation point v of $\{v^k\}$ along this subsequence belongs to $V_{\mathcal{I}}$. Consequently, by taking X small enough, we can assume that

$$\mathcal{I}(x; d) \subseteq \mathcal{I}^* \quad \text{and} \quad V_{\mathcal{I}(x;d)} \neq \emptyset \tag{4.11}$$

for every $x \in X$ and every optimal solution d of subproblem (4.2).

We claim that

$$\mathrm{dist}(v, V^*) = O(\|d\| + \|x - x^*\|) \tag{4.12}$$

for all $x \in X$ and all (d, v) satisfying the KKT conditions (4.3)–(4.5) for problem (4.2). To see this, note that $V^* \neq \emptyset$ coincides with the solution set of the following linear system in v:

$$g(x^*) + \sum_{i \in \mathcal{I}^*} v_i g_i(x^*) = 0, \quad v_i \geq 0, \quad i \in \mathcal{I}^*, \quad v_i = 0, \quad i \notin \mathcal{I}^*. \tag{4.13}$$

Since (d, v) satisfies (4.3)–(4.5) and $\mathcal{I}(x; d) \subseteq \mathcal{I}^*$ by (4.11), v satisfies the last two conditions of (4.13). Moreover, v satisfies the first equation of (4.13) approximately in the sense that

$$\begin{aligned} g(x^*) + \sum_{i \in \mathcal{I}^*} v_i g_i(x^*) &= g(x^*) - g(x) - G(x)d + \sum_{i \in \mathcal{I}^*} v_i (g_i(x^*) - g_i(x)) - \sum_{i \in \mathcal{I}^*} v_i G_i(x) d \\ &= O(\|d\| + \|x - x^*\|), \end{aligned}$$

where the first equality follows from (4.3) and the second equality follows from the local Lipschitz continuity of g and g_i, $i = 1, \ldots, m$ (cf. Assumption 2 (c)) and the uniform boundedness of (x, v) for $x \in X$ (cf. Lemma 4.3). Thus, by applying a well-known Hoffman's error bound [6] to the linear system (4.13) in v, we obtain (4.12), where the constant in the big "O" notation is independent of x and (d, v).

We now show that, by taking a smaller neighborhood X if necessary, we further have that

$$\|d\| = O(\|x - x^*\|),$$

which together with (4.12) would complete the proof. To show this, suppose to the contrary that there exist a sequence $\{x^k\} \subset X$ converging to x^* and a sequence $\{(d^k, v^k)\}$ satisfying (4.3)–(4.5) with $x = x^k$ such that $\|d^k\|/\|x^k - x^*\| \to \infty$. By passing to a subsequence if necessary, we can assume that $\mathcal{I}(x^k; d^k)$ in (4.3)–(4.5) is fixed at some $\mathcal{I}$ for all k. Since $x^k \to x^*$, we have $d^k \to 0$ as was shown earlier. Moreover, by (4.11), we have $\mathcal{I} \subseteq \mathcal{I}^*$ and $V_{\mathcal{I}} \neq \emptyset$. By replacing $\mathcal{I}^*$ by $\mathcal{I}$ in the proof of (4.12), we have that

$$\mathrm{dist}(v, V_{\mathcal{I}}) = O(\|d\| + \|x - x^*\|)$$

for all $x \in X$ and all (d, v) satisfying (4.3)–(4.5) with $\mathcal{I}(x; d) = \mathcal{I}$. Since $\mathcal{I}(x^k; d^k) = \mathcal{I}$, this together with the assumption $\|d^k\|/\|x^k - x^*\| \to \infty$ implies $\|v^k - \hat{v}^k\| = O(\|d^k\|)$, where $\hat{v}^k$ is an element of $V_{\mathcal{I}}$ nearest to v^k. Since $\hat{v}^k \in V_{\mathcal{I}}$, there holds

$$g(x^*) + \sum_{i \in \mathcal{I}} \hat{v}_i^k g_i(x^*) = 0, \quad \hat{v}_i^k \geq 0, \quad i \in \mathcal{I}, \quad \hat{v}_i^k = 0, \quad i \notin \mathcal{I}.$$

Then it follows from (4.3) that

$$g(x^k) - g(x^*) + G(x^k)d^k + \sum_{i \in \mathcal{I}} v_i^k (g_i(x^k) - g_i(x^*) + G_i(x^k)d^k) = -\sum_{i \in \mathcal{I}} (v_i^k - \hat{v}_i^k) g_i(x^*). \tag{4.14}$$

Also, $\mathcal{I}(x^k; d^k) = \mathcal{I}$ together with (4.4) implies

$$c_i(x^k) - c_i(x^*) + g_i(x^k)^T d^k + \frac{1}{2}(d^k)^T G_i(x^k) d^k = 0, \quad i \in \mathcal{I}.$$

By Lemma 4.3, $\{v^k\}$ is bounded. Then, by further passing to a subsequence if necessary, we can assume that $\{(d^k/\|d^k\|, (v^k - \hat{v}^k)/\|d^k\|, v^k)\}$ converges to some (u, z, w) with $\|u\| = 1$. Dividing (4.14) and the above equality by $\|d^k\|$, and using $(x^k, d^k) \to (x^*, 0)$, $\|x^k - x^*\|/\|d^k\| \to 0$ and local Lipschitz continuity of g and c_i, g_i, $i \in \mathcal{I}$, yield in the limit that

$$\begin{aligned} G(x^*)u + \sum_{i \in \mathcal{I}} w_i G_i(x^*)u &= -\sum_{i \in \mathcal{I}} z_i g_i(x^*), \\ g_i(x^*)^T u &= 0, \quad i \in \mathcal{I}. \end{aligned}$$

The above two equalities imply $u^T(G(x^*) + \sum_{i \in \mathcal{I}} w_i G_i(x^*))u = 0$. Since it is readily seen from $\mathcal{I} \subseteq \mathcal{I}^*$ that $w \in V_{\mathcal{I}} \subseteq V^*$, Assumption 2 (a) implies $G(x^*) + \sum_{i \in \mathcal{I}} w_i G_i(x^*) \succ 0$ and hence $u = 0$, contradicting $\|u\| = 1$. □

Like the ordinary SQP method, the SQCQP method is locally convergent at a quadratic rate if the unit step size is used.

Lemma 4.5 *For any x sufficiently close to x^* and any KKT pair (d, v) of (4.2), we have $H(x, v) \succeq \frac{1}{2}\mu^* I$ and*

$$\|x + d - x^*\| = O(\|x - x^*\|^2), \quad \operatorname{dist}(v, V^*) = O(\|x - x^*\|^2). \tag{4.15}$$

Proof. For each x and each KKT pair (d, v) of (4.2), we have that (d, v) satisfies (4.3) and

$$c_i(x) + g_i(x)^T d + \frac{1}{2} d^T G_i(x) d = 0, \quad i \in \mathcal{I}(x; d), \tag{4.16}$$

and $v_i = 0$ for $i \notin \mathcal{I}(x;d)$, for some $\mathcal{I}(x;d) \subseteq \{1,\ldots,m\}$.

Let X be a sufficiently small neighborhood of x^* so that Lemmas 4.3 and 4.4 apply. As was shown in the proof of Lemma 4.4, by taking X smaller if necessary, we can further assume that (4.11) holds for every $x \in X$ and every KKT pair (d,v) of (4.2), where $V_{\mathcal{I}(x;d)}$ denotes the set of vectors v satisfying (4.10) with $\mathcal{I}$ being replaced by $\mathcal{I}(x;d)$. Since $\mathcal{I}(x;d)$ takes on only a finite number of distinct values, it suffices to prove the lemma for those x and d such that $\mathcal{I}(x;d) = \mathcal{I}$ for some fixed index set $\mathcal{I} \subseteq \{1,\ldots,m\}$. Then, we have from (4.11) that $\mathcal{I} \subseteq \mathcal{I}^*$ and $V_{\mathcal{I}} \neq \emptyset$. Let v^* be an arbitrary element of $V_{\mathcal{I}}$. Then we have

$$g(x^*) + \sum_{i\in\mathcal{I}} v_i^* g_i(x^*) = 0, \tag{4.17}$$

$$c_i(x^*) = 0, \quad i \in \mathcal{I}. \tag{4.18}$$

It follows from (4.3) and (4.17) that

$$\begin{aligned}
0 &= g(x) - g(x^*) + G(x)d + \sum_{i\in\mathcal{I}} \Big(v_i g_i(x) - v_i^* g_i(x^*) + v_i G_i(x)d\Big) \\
&= \Big(G(x) + \sum_{i\in\mathcal{I}} v_i G_i(x)\Big)(x + d - x^*) + \sum_{i\in\mathcal{I}} (v_i - v_i^*) g_i(x^*) \\
&\quad + g(x) - g(x^*) - G(x)(x - x^*) + \sum_{i\in\mathcal{I}} v_i \Big(g_i(x) - g_i(x^*) - G_i(x)(x - x^*)\Big) \\
&= \Big(G(x) + \sum_{i\in\mathcal{I}} v_i G_i(x)\Big)(x + d - x^*) + \sum_{i\in\mathcal{I}} (v_i - v_i^*) g_i(x^*) + O(\|x - x^*\|^2),
\end{aligned} \tag{4.19}$$

where the last equality follows from Assumption 2 (b) and the uniform boundedness of v shown by Lemma 4.3. On the other hand, by (4.16) and (4.18), we have for each $i \in \mathcal{I}$ that

$$\begin{aligned}
0 &= c_i(x) + g_i(x)^T d + \frac{1}{2} d^T G_i(x) d - c_i(x^*) \\
&= g_i(x^*)^T (x + d - x^*) + c_i(x) - c_i(x^*) - g_i(x^*)^T (x - x^*) \\
&\quad + (g_i(x) - g_i(x^*))^T d + \frac{1}{2} d^T G_i(x) d \\
&= g_i(x^*)^T (x + d - x^*) + O(\|x - x^*\|^2) + O(\|x - x^*\| \cdot \|d\|) + O(\|d\|^2) \\
&= g_i(x^*)^T (x + d - x^*) + O(\|x - x^*\|^2),
\end{aligned} \tag{4.20}$$

where the third equality follows from Assumption 2 (b) and the last equality follows from (4.9). Since $v_i = 0$ for all $i \notin \mathcal{I}$, then

$$H(x,v) \;=\; G(x) + \sum_{i\in\mathcal{I}} v_i G_i(x)$$

Let $A_{\mathcal{I}}$ be the $n \times |\mathcal{I}|$ matrix whose columns are $g_i(x^*)$, $i \in \mathcal{I}$. Then (4.19) and (4.20) yield

$$\begin{pmatrix} H(x,v) & A_{\mathcal{I}} \\ A_{\mathcal{I}}^T & 0 \end{pmatrix} \begin{pmatrix} x+d-x^* \\ v_{\mathcal{I}} - v_{\mathcal{I}}^* \end{pmatrix} = O(\|x-x^*\|^2), \tag{4.21}$$

where $v_{\mathcal{I}}^* := (v_i^*)_{i\in\mathcal{I}}$ and $v_{\mathcal{I}} := (v_i)_{i\in\mathcal{I}}$. Let $A_{\mathcal{I}_1} \in \Re^{n\times|\mathcal{I}_1|}$, with $\mathcal{I}_1 \subseteq \mathcal{I}$, be any submatrix of $A_{\mathcal{I}}$ whose columns are linearly independent and span the column space of $A_{\mathcal{I}}$. Then $A_{\mathcal{I}}(v_{\mathcal{I}} - v_{\mathcal{I}}^*) = A_{\mathcal{I}_1} u$ for some $u \in \Re^{|\mathcal{I}_1|}$, so that (4.21) implies

$$\begin{pmatrix} H(x,v) & A_{\mathcal{I}_1} \\ A_{\mathcal{I}_1}^T & 0 \end{pmatrix} \begin{pmatrix} x+d-x^* \\ u \end{pmatrix} = O(\|x-x^*\|^2). \tag{4.22}$$

Also, Assumption 2 (a) together with Lemma 4.4 implies $H(x,v) \succeq \frac{1}{2}\mu^* I$ for all x sufficiently close to x^*. Thus, by taking X sufficiently small, we can assume that $H(x,v) \succeq \frac{1}{2}\mu^* I$ for all $x \in X$ and all KKT pairs (d,v) of (4.2). Since the matrix $A_{\mathcal{I}_1}^T$ is constant with full column rank, this implies that the coefficient matrix in (4.22) is nonsingular and its inverse is uniformly bounded in operator norm for $x \in X$. Then it follows from (4.22) that

$$\|x+d-x^*\| = O(\|x-x^*\|^2), \qquad \|u\| = O(\|x-x^*\|^2).$$

Consequently, we have

$$\sum_{i\in\mathcal{I}} (v_i - v_i^*) g_i(x^*) = A_{\mathcal{I}}(v_{\mathcal{I}} - v_{\mathcal{I}}^*) = A_{\mathcal{I}_1} u = O(\|x-x^*\|^2). \tag{4.23}$$

Recall that v^* is an arbitrary element of the solution set $V_{\mathcal{I}}$ of the linear system (4.10) in v. Then (4.23) shows that v is an approximate solution of the first equation in (4.10) with residual being $O(\|x-x^*\|^2)$. Since v satisfies the remaining equations and inequalities in (4.10), by invoking Hoffman's error bound [6] and using $V_{\mathcal{I}} \subseteq V^*$, we obtain

$$\operatorname{dist}(v, V^*) \le \operatorname{dist}(v, V_{\mathcal{I}}) = O(\|x-x^*\|^2),$$

as desired. □

The next lemma shows that the unit step size is accepted whenever an iterate x is sufficiently close to the optimal solution x^*. In other words, the Maratos effect will not occur.

Lemma 4.6 *Let X and ρ_1 be given by Lemma 4.3. Fix $r^0 > 0$ and $\delta > 0$. Let $\rho_3 := \max\{r^0, \rho_1 + 2\delta\}$. If $x \in X$ is sufficiently close to x^* and (d,v) is any KKT pair of (4.2) satisfying $\max_i v_i \le r \le \rho_3$, then the step size $\beta = 1$ is accepted by the line search criterion (2.11) with $\alpha_i = 1$ for all i.*

Proof. Since $x \in X$, Lemma 4.3 implies (4.2) has a KKT pair. Any KKT pair (d, v) of (4.2) satisfies $c_i(x) + g_i(x)^T d + \frac{1}{2} d^T G_i(x) d \leq 0$ for all i. Thus

$$\begin{aligned} |[c_i(x+d)]_+| &= \left| [c_i(x+d)]_+ - [c_i(x) + g_i(x)^T d + \frac{1}{2} d^T G_i(x) d]_+ \right| \\ &\leq \left| c_i(x+d) - \left(c_i(x) + g_i(x)^T d + \frac{1}{2} d^T G_i(x) d \right) \right| \\ &= o(\|d\|^2) \end{aligned}$$

for all i, so that

$$\begin{aligned} F_r(x+d) - F_r(x) &= g(x)^T d + \frac{1}{2} d^T G(x) d - r \sum_{i=1}^{m} [c_i(x)]_+ + r \cdot o(\|d\|^2) \\ &= \left(\bar{F}_r(x, d, \alpha) - F_r(x) \right) + r \cdot o(\|d\|^2), \end{aligned} \tag{4.24}$$

where $\alpha_i = 1$ for all i and the second equality follows from (2.13). Moreover, Lemma 4.5 implies that, for x sufficiently close to x^*, condition (3.1) with $B = G(x)$ is satisfied with $\mu = \frac{1}{2}\mu^*$. Since $r \geq \max_i v_i$, Lemma 3.1 yields that (3.2) holds. Let $\sigma \in (0, 1)$ be the constant in the line search criterion (2.11). Then, combining (3.2) with (4.24) yields

$$F_r(x+d) - F_r(x) \leq \sigma \left(\bar{F}_r(x, d, \alpha) - F_r(x) \right),$$

provided $r \cdot o(\|d\|^2) \leq (1-\sigma)\mu \|d\|^2/2$. Since $r \leq \rho_3$, the latter holds whenever $\|d\|$ is sufficiently small. Notice that $\max_i v_i \leq \rho_1 < \rho_3$, so that r exists. Since Lemma 4.4 shows that $\|d\|$ becomes arbitrarily small as x approaches x^*, this implies that (2.11) is satisfied by $\beta = 1$ whenever $x \in X$ is sufficiently close to x^*. □

We can now establish the local quadratic convergence of the SQCQP method.

Theorem 4.1 *There exists a neighborhood X of x^* (depending on r^0 and δ) such that the sequence $\{x^k\}$ generated by the SQCQP method, with $x^1 \in X$ and $B^k := G(x^k)$ for all k, converges to x^* at a Q-quadratic rate. In this case, we also have* $\mathrm{dist}(v^k, V^*) \to 0$ *at an R-quadratic rate.*[3]

Proof. By Lemma 4.1, there exists a neighborhood X of x^* such that, whenever $x^k \in X$, we have $\alpha_i^k = 1$ for all i and (d^k, v^k) is a well defined KKT pair of (4.2) with $x = x^k$. By Lemmas 4.3 and 4.6, we can assume by taking X smaller if necessary that there exists a $\rho_1 > 0$ such that,

[3]For the definition of Q-quadratic and R-quadratic rate of convergence, see [13].

whenever $x^k \in X$ and $\max_i v_i^k \le r^k \le \rho_3 := \max\{r^0, \rho_1 + 2\delta\}$, we have $\|v^k\| \le \rho_1$ and $\beta^k = 1$. By Lemma 4.5, we can further assume by taking X smaller if necessary that, whenever $x^k \in X$, we have $x^k + d^k \in X$ and $\|x^k + d^k - x^*\| \le \frac{1}{2}\|x^k - x^*\|$.

Then, whenever

$$x^k \in X \quad \text{and} \quad r^{k-1} \le \rho_3, \tag{4.25}$$

we have that $\alpha_i^k = 1$ for all i and the KKT pair (d^k, v^k) is well defined. Moreover, $\|v^k\| \le \rho_1$ so the updating rule (2.15) and the definition of ρ_3 imply $r^k \le \rho_3$ as well as $r^k \ge \max_i v_i^k$. This then implies $\beta^k = 1$ and hence $x^{k+1} \in X$ and $\|x^{k+1} - x^*\| \le \frac{1}{2}\|x^k - x^*\|$. Since $x^1 \in X$ and $r^0 \le \rho_3$, it follows by induction on k that (4.25) holds for all k and that $x^k \to x^*$. Moreover, Lemma 4.5 implies $\|x^{k+1} - x^*\| = O(\|x^k - x^*\|^2)$ and $\mathrm{dist}(v^k, V^*) = O(\|x^k - x^*\|^2)$, so that $x^k \to x^*$ at a Q-quadratic rate and $\mathrm{dist}(v^k, V^*) \to 0$ at an R-quadratic rate. □

If $G(x) \succeq \mu I$ for all x, where $\mu > 0$ is some constant, then Assumption 2 (a) holds automatically and $B^k = G(x^k)$ satisfies the assumptions of Theorem 3.1. Thus, in this case, we conclude from Theorems 3.1 and 4.1 that if $\{x^k\}$ generated by the SQCQP method is bounded (in addition to Assumptions 1 and 2 (b)), then $\{x^k\}$ either terminates finitely at the unique optimal solution x^* or converges to x^* at a quadratic rate locally.

We remark that Lemmas 4.1–4.6 and Theorem 4.1 still hold if Assumption 2 (a) is replaced by the weaker assumption that, for each x sufficiently close to x^* and each $u \in \Re^n$, there exists a Lagrange multiplier vector v of (4.2) satisfying

$$u^T H(x, v) u \quad \ge \quad \mu^* \|u\|^2$$

for some constant $\mu^* > 0$. This assumption is more restrictive than the quadratic growth assumption made by Anitescu [2, Equations (1.1), (1.12)], although our conclusion of quadratic rate of convergence is stronger than the conclusion of superlinear rate of convergence obtained in [2, Theorem 3.6].

5 Conclusion

In this paper we have presented a Sequential Quadratically Constrained Quadratic Programming (SQCQP) method for solving smooth convex programs. At each iteration, this method solves a single convex quadratically constrained quadratic minimization subproblem. The latter can be

formulated as a second-order cone program and solved efficiently. A key advantage of the SQCQP method as compared to the classical SQP methods is that the Maratos effect can be avoided when using an ordinary ℓ_1 exact penalty function for the line search.

As future work, numerical experimentation with the SQCQP method is needed in order to assess the computational saving (if any) that can be achieved over classical SQP methods. Also, it would be interesting to see if some of the assumptions (such as the Slater condition and Assumption 2) used in the convergence analysis can be relaxed.

Acknowledgment. The authors thank Jim Burke and Steve Wright for bringing to their attention the references [16] and [2], respectively.

References

[1] F. Alizadeh and S. Schmieta, *Symmetric cones, potential reduction methods*, in Handbook of Semidefinite Programming, H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., Kluwer Academic Publishers, Boston, MA, 2000, pp. 195–233.

[2] M. Anitescu, *A superlinearly convergent sequential quadratically constrained programming algorithm for degenerate nonlinear programming*, Department of Mathematics, University of Pittsburgh, submitted.

[3] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, NY, 1982.

[4] B.T. Boggs and J.W. Tolle, *Sequential quadratic programming*, Acta Numerica, 4 (1996), pp. 1–51.

[5] M. Fukushima, *A successive quadratic programming algorithm with global and superlinear convergence properties*, Math. Programming, 35 (1986), pp. 253–264.

[6] A.J. Hoffman, *On approximate solutions of systems of linear inequalities*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 263–265.

[7] S. Kruk and H. Wolkowicz, *Sequential, quadratic constrained, quadratic programming for general nonlinear programming*, in Handbook of Semidefinite Programming, H. Wolkowicz,

R. Saigal and L. Vandenberghe, eds., Kluwer Academic Publishers, Boston, MA, 2000, pp. 563–575.

[8] M.S. Lobo, L. Vandenberghe, S. Boyd and H. Lebret, *Applications of second-order cone programming*, Linear Algebra Appl., 284 (1998), pp. 193–228.

[9] Z.-Q. Luo and S. Zhang, *On extensions of Frank-Wolfe Theorems*, Comput. Optim. Appl., 13 (1999), pp. 87–110.

[10] D.Q. Mayne and E. Polak, *A superlinearly convergent algorithm for constrained optimization problems*, Math. Programming Stud., 16 (1982), pp. 45–61.

[11] R.D.C. Monteiro and T. Tsuchiya, *Polynomial convergence of primal-dual algorithms for the second-order cone programs based on the MZ-family of directions*, Math. Programming, 88 (2000), pp. 61–83.

[12] Y.E. Nesterov and A.S. Nemirovskii, *Interior Point Polynomial Methods in Convex Programming: Theory and Applications*, SIAM, Philadelphia, PA, 1993.

[13] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY, 1970.

[14] V.M. Panin, *A second-order method for the discrete min-max problem*, USSR Comput. Math. Math. Phys., 19 (1) (1979), pp. 90–100.

[15] V.M. Panin, *Some methods of solving convex programming problems*, USSR Comput. Math. Math. Phys., 21 (2) (1981), pp. 57–72.

[16] E. Polak, D.Q. Mayne and J.E. Higgins, *On the extension of Newton's method to semi-infinite minimax problems*, SIAM J. Control Optim., 30 (1992), pp. 367–389.

[17] M.J.D. Powell, *Variable metric methods for constrained optimization*, in Mathematical Programming: The State of Art, A. Bachem, M. Grötschel, and B. Korte, eds., Springer-Verlag, Berlin, 1983, pp. 288–311.

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

[19] T. Tsuchiya, *A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming*, Optim. Methods Softw., 11/12 (1999), pp. 141–182.

[20] E.J. Wiest and E. Polak, *A generalized quadratic programming-based phase*-I–*phase*-II *method for inequality-constrained optimization*, Appl. Math. Optim., 26 (1992), pp. 223–252.