Mathematics of Operations Research
(accepted for publication)

ON THE SEQUENTIAL QUADRATICALLY CONSTRAINED QUADRATIC PROGRAMMING METHODS*

M. V. Solodov†

May 14, 2002 (Revised March 17, 2003)

ABSTRACT

An iteration of the sequential quadratically constrained quadratic programming method (SQCQP) consists of minimizing a quadratic approximation of the objective function subject to quadratic approximation of the constraints, followed by a linesearch in the obtained direction. Methods of this class are receiving attention due to the development of efficient interior point techniques for solving subproblems with this structure, via formulating them as second-order cone programs. Recently, Fukushima, Luo and Tseng [8] proposed a SQCQP method for convex minimization with twice continuously differentiable data. Their method possesses global and locally quadratic convergence, and it is free of the Maratos effect. The feasibility of subproblems in their method is enforced by switching between the linear and quadratic approximations of the constraints. This strategy requires computing a strictly feasible point, as well as choosing some further parameters. We propose a SQCQP method where feasibility of subproblems is ensured by introducing a slack variable and, hence, is automatic. In addition, we do not assume convexity of the objective function or twice differentiability of the problem data. While our method has all the desirable convergence properties, it is easier to implement. Among other things, it does not require computing a strictly feasible point, which is a nontrivial task. In addition, its global convergence requires weaker assumptions.

Key words. Quadratically constrained quadratic programming, nonsmooth penalty function, Maratos effect.
AMS subject classifications. 90C30, 90C33.

---

* Research of the author is partially supported by CNPq Grant 300734/95-6, by PRONEX–Optimization, and by FAPERJ.
† Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email : `solodov@impa.br` .

1 Introduction

We consider the problem

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

where $f : \Re^n \to \Re$ and $g : \Re^n \to \Re^m$ are differentiable functions with Lipschitz-continuous derivatives. We further assume that g is convex, and that the Slater constraint qualification [11] is satisfied:

$$\exists\, \hat{x} \in \Re^n \text{ such that } g_i(\hat{x}) < 0,\ i = 1, \ldots, m. \tag{2}$$

For local rate of convergence results, f, g will further be assumed to be twice continuously differentiable, and f will also be assumed to be convex. We emphasize that those additional assumptions will not be needed for global convergence of our algorithm.

For the case where both f and g are convex twice continuously differentiable functions, Fukushima, Luo and Tseng [8] proposed an SQCQP method based on solving the subproblems of the following structure:

$$\begin{array}{ll} \text{minimize}_d & \langle f'(x^k), d\rangle + \frac{1}{2}\langle H^k d, d\rangle \\ \text{subject to} & g_i(x^k) + \langle g_i'(x^k), d\rangle + \frac{\alpha_i^k}{2}\langle g_i''(x^k) d, d\rangle \leq 0, \quad i = 1, \ldots, m, \end{array} \tag{3}$$

where $x^k \in \Re^n$ is the current iterate, H^k is an $n \times n$ symmetric positive definite matrix, and $\alpha_i^k \in [0, 1]$ are parameters which control the feasibility of (3) according to the rules described below. Subproblem (3) is a convex quadratically constrained quadratic program, which can be cast as a second-order cone program [10, 15]. The latter can be efficiently solved by interior point algorithms (such as [14, 19]), which justifies the interest in methods based on second-order approximation of the constraints. In fact, in [2] nonconvex QCQPs were also handled quite efficiently by using other nonlinear programming techniques.

Let $d^k \in \Re^n$ be the solution of (3), and let $\mu^k \in \Re^m_+$ be some associated Lagrange multiplier. The next iterate in the method of [8] is given by $x^{k+1} = x^k + \gamma_k d^k$, where the stepsize $\gamma_k > 0$ is computed using an Armijo-type linesearch procedure for the standard l_1-penalty function $f(x) + \beta_k \sum_{i=1}^m [g_i(x)]_+$, with $\beta_k > \|\mu^k\|_\infty$ being the penalty parameter and $[\cdot]_+ = \max\{0, \cdot\}$. In [8], the method outlined above is shown to be globally convergent to a solution of (1) under certain natural assumptions. Local quadratic rate of convergence is established under the assumption of positive definiteness of the Hessian of the Lagrangian of (1) at the KKT points. As discussed in [8], one attractive feature of this SQCQP method is that it is free from the Maratos effect. The Maratos effect [12, 18] is a notorious difficulty for sequential quadratic programming methods (SQP) [3, 4, 6], which can be thought of as a special case of SQCQP with $\alpha_i^k = 0$ for all i and k. It refers to the situation where, even though conditions for local superlinear convergence of the pure SQP iterations are satisfied, the linesearch procedure based on a nonsmooth penalty function (such as l_1) does not accept the unit stepsize. As a result, superlinear convergence is not achieved. Known remedies to avoid the Maratos effect in SQP methods are quite complex and require careful implementation [13, 7, 5, 6]. By contrast, no special care is needed in the SQCQP method of [8]. In particular, the unit stepsize is always accepted under the assumptions required for the local superlinear rate of convergence.

One important detail which may not be so easy in the implementation of the method of [8] concerns the control of parameters α_i^k in (3). Note that quadratic approximation of constraints can be "too optimistic", in the sense that the feasible set in (3) can be empty if $\alpha_i^k = 1$ for all i. By contrast, the linear approximation corresponding to $\alpha_i^k = 0$ for all i, is guaranteed to yield feasible subproblems (in the case of convex g). The following rule is proposed in [8] to take care of the feasibility problem:

> Compute $\hat{x} \in \Re^n$ satsifying the Slater condition (2). Choose $\theta \in [0,1)$ and $\nu \in (\theta, 1)$.
> (This is done once, to initialize the algorithm.)
>
> Compute $\eta_i^k := \langle g_i''(x^k)(x^k - \hat{x}), x^k - \hat{x}\rangle$, $i = 1, \ldots, m$, and
>
> $$K_1(x^k) := \{i \mid g_i(x^k) > 0\}, \quad K_2(x^k) := \{i \mid \theta g_i(\hat{x}) \le g_i(x^k)\},$$
>
> $$s_1(x^k) := \max_{i \in K_1(x^k)} \frac{g_i(x^k)}{g_i(x^k) - \nu g_i(\hat{x})} \quad (s_1(x^k) := -\infty \text{ if } K_1(x^k) = \emptyset),$$
>
> $$s_2(x^k) := \min\left\{1, \min_{i \in K_2(x^k)} \frac{g_i(x^k) - \nu g_i(\hat{x})}{\eta_i^k}\right\},$$
>
> $$s_3(x^k) := \min\left\{s_2(x^k), \min_{i \notin K_2(x^k)} \frac{-2(\nu - \theta) g_i(\hat{x})}{\eta_i^k}\right\}.$$
>
> If $s_2(x^k) \le 2s_1(x^k)$ then set $\alpha_i^k = 0$ for all i.
> If $s_3(x^k) > 2s_1(x^k)$ then set $\alpha_i^k = 1$ for all i.
> If $s_3(x^k) \le 2s_1(x^k) < s_2(x^k)$, then $\alpha_i^k = 1$ for $i \in K_2(x^k)$ and $\alpha_i^k = 0$ for $i \notin K_2(x^k)$.

The above rule for choosing α_i^k is based on testing constraints violation and curvature at the current iterate, and it guarantees that the subproblem (3) is feasible. Another potentially useful feature of this rule is that it allows to balance the numbers of quadratic and linear constraints in the subproblem at the early stages of the algorithm (subproblems with fewer quadratic constraints are in general easier to solve). But it is guaranteed that eventually, $\alpha_i^k = 1$ for all i, so that all the second-order information is being used. As discussed in [8], many other rules are possible, including the continuous choice in the interval $[0,1]$. Furthermore, there are some choices to be made even for the given rule, namely the parameters $\theta\,, \nu$ and, especially, $\hat{x}$. The implementation is therefore not straightforward, and may require a fair bit of fine-tuning. Most importantly, computing $\hat{x}$ is a general convex feasibility problem, which is a computationally nontrivial task. Essentially, the method of [8] is a "Phase I - Phase II" method. Phase I corresponds to computing a Slater point, while Phase II is the SQCQP method for solving problem (1). Computing a Slater point can be done in a finite number of iterations by various means, see [8]. Nevertheless, this Phase I problem has complexity comparable to the original problem itself, and so it increases the computational burden considerably. As an alternative, [8] proposes to use in the rule for choosing α_i^k some estimates c_1 and c_2 such that $0 > c_1 \geq \max_i\, g_i(\hat{x}),\ c_2 \geq \max_i\, \langle g_i''(x^k)(x^k - \hat{x}), x^k - \hat{x}\rangle$ for some (unknown) Slater point $\hat{x}$. However, this requires either essentially heuristic considerations or computing the estimates iteratively (if c_1 and c_2 are not adequate and the subproblems are still infeasible, one has to increase c_1 and c_2 and try again).

In this paper, we propose to deal with the feasibility of subproblems by introducing a slack variable, which simplifies the implementation considerably. Among other things, computing a strictly feasible point is no longer necessary. In addition, for the global convergence result we do not assume the convexity of f or twice differentiability of f and g. In fact, our global convergence result appears to be stronger than that in [8] also in a number of other ways (see the comments preceding Theorem 6).

Our proposal is to consider, instead of (3), the subproblem

$$\begin{array}{ll} \text{minimize}_{(d,t)} & \langle f'(x^k), d\rangle + \frac{1}{2}\langle H^k d, d\rangle + \beta_k t \\ \text{subject to} & g_i(x^k) + \langle g_i'(x^k), d\rangle + \frac{1}{2}\langle G_i^k d, d\rangle \leq t,\ i \in I_k, \quad t \geq 0, \end{array} \tag{4}$$

where $\beta_k > 0$ is a penalty parameter, G_i^k, $i \in I_k$, are $n \times n$ symmetric positive semidefinite matrices (possibly different from $g_i''(x^k)$ even when the latter exists), and I_k is any index set which satisfies

$$\begin{array}{c} I(x^k) \subset I_k \subset \{1, \ldots, m\}, \\ I(x^k) = \{i \mid g_i(x^k) = p(x^k)\}, \quad p(x^k) = \max_{i=1,\ldots,m}\, [g_i(x^k)]_+. \end{array} \tag{5}$$

Note that the possibility to choose the index set I_k smaller than $\{1, \ldots, m\}$ is very useful (especially at the early stages of the algorithm), because it reduces the number of constraints thus leading to simpler subproblems. This effect can be considered similar to setting $\alpha_i^k = 0$ for some i in (3), but can have a further advantage of removing altogether the constraints which are locally irrelevant.

As suggested by a referee, instead of (4) we could also consider subproblems with a different slack and a different penalty parameter for each of the constraints:

$$\begin{array}{ll} \text{minimize}_{(d,t)} & \langle f'(x^k), d\rangle + \frac{1}{2}\langle H^k d, d\rangle + \sum_{i \in I_k} \beta_k^i t_i \\ \text{subject to} & g_i(x^k) + \langle g_i'(x^k), d\rangle + \frac{1}{2}\langle G_i^k d, d\rangle \leq t_i\,,\ t_i \geq 0,\ i \in I_k. \end{array}$$

Numerically, this can be advantageous in some situations. For the subsequent analysis we prefer the simpler form of (4). It can be seen that the analysis extends to the case of different slacks and penalty parameters after simple modifications.

Let $(d^k, t^k) \in \Re^n \times \Re$ be the solution of (4), which is easily seen to be unique (Section 2). Then the next iterate of our method is given by $x^{k+1} = x^k + \gamma_k d^k$, where the stepsize $\gamma_k > 0$ is computed using an Armijo-type linesearch procedure (see Algorithm 1) for the l_∞-penalty function

$$\psi_{\beta_k}(x) = f(x) + \beta_k p(x).$$

The idea of using a slack variable to ensure feasibility comes from a similar strategy for SQP [3]. In general SQP methods, removing the feasibility concerns using this strategy comes at the expense of the difficulty in choosing the penalty parameter β_k (Essentially, one needs to ensure that eventually $\beta_k > \|\mu^k\|_1$, where μ^k is a Lagrange multiplier associated with d^k. But β_k has to be chosen before the new multiplier μ^k is computed, which leads to certain difficulties. See, for example, the the last comment in Section 3 of [1]). For our algorithm, we are able to give a novel rule for updating the penalty parameter β_k (see Algorithm 1), which guarantees convergence.

To conclude this section, we cite some previous work on SQCQP and related methods. For the first time methods of this class were probably considered in [16, 17], where strong convexity of f and g is assumed. In [2], local superlinear rate of convergence of a trust-region SQCQP method is obtained without the convexity assumptions, under the Mangasarian-Fromovitz constraint qualification and a certain quadratic growth condition. This reference also provides some numerical testing, which highlights the following fact. Although SQCQP subproblems are more computationally difficult than subproblems in SQP, they can be solved reasonably efficiently even in the nonconvex case (in the nonconvex case, subproblems have to be solved by approaches other than the interior point methods). Moreover, since SQCQP subproblems are a better approximation of the original problem, fewer subproblems have to be solved. Thus this extra effort (compared to SQP) may be well worth it. However, global convergence is not discussed in [2]. As for some other SQCQP-related work, we mention [20, 9].

Some final words about our notation. For a directionally differentiable function $\psi : \Re^n \to \Re$, by $\psi'(x; d)$ we denote the usual directional derivative of ψ at the point $x \in \Re^n$ in the direction $d \in \Re^n$. By $\|\cdot\|$ we denote the 2-norm, by $\|\cdot\|_1$ we denote the l_1-norm, and by $\|\cdot\|_\infty$ the l_∞-norm. The space is always clear from the context, and will not be specified. For two symmetric positive semidefinite matrices A and B, we write $A \succeq B$ if the matrix $A - B$ is positive semidefinite. By E we denote the identity matrix and by O the zero matrix. Finally, the operation $t/0$ where $t > 0$ is considered well-defined, with the result being $+\infty$.

2 SQCQP Method and Its Global Convergence

First, note that (4) is always feasible and has unique solution. Indeed, as is easy to see, for each $d \in \Re^n$ fixed, the minimum with respect to t in (4) is attained at

$$t^k(d) = \max_{i \in I_k} \left[g_i(x^k) + \langle g_i'(x^k), d\rangle + \frac{1}{2}\langle G_i^k d, d\rangle\right]_+. \tag{6}$$

Hence, (4) is equivalent to

$$\text{minimize}_d \ \langle f'(x^k), d\rangle + \frac{1}{2}\langle H^k d, d\rangle + \beta_k t^k(d).$$

Since H^k is positive definite, the objective function of this program is strongly convex and has the unique minimizer d^k. It follows that (d^k, t^k) with $t^k = t^k(d^k)$ is the unique solution of (4). Furthermore, since the constraints in (4) obviously satisfy the Slater constraint qualification, the Karush-Kuhn-Tucker (KKT) optimality conditions [11] hold: there exist some $\mu_i^k \in \Re$, $i \in I_k$, and $\nu^k \in \Re$ such that

$$f'(x^k) + H^k d^k + \sum_{i \in I_k} \mu_i^k (g_i'(x^k) + G_i^k d^k) = 0, \tag{7}$$

$$\beta_k - \sum_{i \in I_k} \mu_i^k - \nu^k = 0, \tag{8}$$

$$g_i(x^k) + \langle g_i'(x^k), d^k\rangle + \frac{1}{2}\langle G_i^k d^k, d^k\rangle \le t^k, \quad \mu_i^k \ge 0, \quad i \in I_k, \tag{9}$$

$$\mu_i^k\left(g_i(x^k) + \langle g_i'(x^k), d^k\rangle + \frac{1}{2}\langle G_i^k d^k, d^k\rangle - t^k\right) = 0, \quad i \in I_k, \tag{10}$$

$$t^k \ge 0, \quad \nu^k \ge 0, \quad t^k \nu^k = 0. \tag{11}$$

We proceed to formally state our algorithm.

Algorithm 1 (SQCQP)
Choose some $x^0 \in \Re^n$, β_0, δ_1, $\delta_2 \in (0, +\infty)$ and σ, $\theta \in (0, 1)$. Set $k := 0$.

1. `QCQP subproblem`.
   Choose an index set I_k according to (5). Choose $n \times n$ symmetric matrices H^k (positive definite) and G_i^k, $i \in I_k$ (positive semidefinite). Compute (d^k, t^k) as the solution of (4), and the associated Lagrange multipliers (μ^k, ν^k).

2. `Stopping test`.
   Stop if $d^k = 0$ and $t^k = 0$.

3. `Linesearch`*.*
If $d^k = 0$ (but $t_k > 0$), set $j_k := 0$ and go to Step 4.

Otherwise, find j_k, the smallest nonnegative integer j, such that

$$\psi_{\beta_k}(x^k + \theta^j d^k) \leq \psi_{\beta_k}(x^k) + \sigma\theta^j \Delta_k, \tag{12}$$

where

$$\Delta_k := \langle f'(x^k), d^k\rangle + \frac{1}{2}\langle H^k d^k, d^k\rangle + \beta_k(t^k - p(x^k)). \tag{13}$$

4. `Variable update`*.*
Set $\gamma_k := \theta^{j_k}$ and $x^{k+1} := x^k + \gamma_k d^k$.

5. `Penalty parameter update`*.*
Compute $r_k := \min\{\|d^k\|^{-1}, \|\mu^k\|_1 + \delta_1\}$. Set

$$\beta_{k+1} := \begin{cases} \beta_k, & \text{if } \beta_k \geq r_k \\ \beta_k + \delta_2, & \text{if } \beta_k < r_k \end{cases} . \tag{14}$$

6. Set $k := k + 1$ and go to Step 1.

Note that our penalty parameter rule is somewhat different from the usual strategies in SQP (where typically $r_k = \|\mu^k\|_1 + \delta_1$ is used). As already mentioned above, choosing this parameter in the framework where feasibility is controlled using slacks is problematic [3] (see also the last comment in Section 3 of [1]). In the specific setting of this paper, we are able to give the novel rule that guarantees convergence. The idea is to ensure that if $\{\beta^k\}$ were to be unbounded, then necessarily two things happen: $\|\mu^k\| \to +\infty$ *and* $d^k \to 0$. As it will be established (see Proposition 5), this situation cannot occur, thus leading to boundedness of $\{\beta_k\}$.

The following simple lemma will be used several times in the sequel.

Lemma 1 *Let $\{x \mid g_i(x) \leq 0,\ i = 1, \ldots, m\} \neq \emptyset$, where $g : \Re^n \to \Re^m$ is convex and continuously differentiable. Then for any $x \in \Re^n$ such that*

$$J(x) := \{i \mid g_i(x) > 0\} \neq \emptyset,$$

it holds that

$$\sum_{i \in J(x)} \mu_i g_i'(x) = 0, \quad \mu_i \geq 0,\ i \in J(x) \quad \Leftrightarrow \quad \mu_i = 0,\ i \in J(x).$$

Proof. By Gordan Theorem of the Alternatives [11], the assertion follows if we establish that

$$\exists\, d \in \Re^n \text{ such that } \langle g_i'(x), d\rangle < 0, \quad i \in J(x).$$

We proceed to exhibit this d. Take any y such that $g_i(y) \leq 0$, $i = 1, \ldots, m$. By the convexity of g, we have that

$$0 > g_i(y) - g_i(x) \geq \langle g_i'(x), y - x \rangle, \quad i \in J(x),$$

thus establishing the claim. ∎

We start with showing that the method is well-defined. This is done in two steps: for the case where $d^k = 0$ (but $t_k > 0$), and when $d^k \neq 0$.

Proposition 2 *Let $\{x \mid g_i(x) \leq 0,\ i = 1, \ldots, m\} \neq \emptyset$, where $g : \Re^n \to \Re^m$ is convex and continuously differentiable.*

If in Algorithm 1 we have $d^k = 0$ and $t^k = 0$, then (x^k, μ^k) is a KKT point of (1).

If $d^k = 0$ and $t^k > 0$, then there exists a finite integer q such that Algorithm 1 generates $d^{k+q} \neq 0$.

Proof. Suppose that $d^k = 0$ and $t^k = 0$. Then (9) and (5) imply that $g_i(x^k) \leq 0$ for $i \in I(x^k) \subset I_k$. Hence, $p(x^k) = 0$. Using again (5), it is easy to see that $g_i(x^k) < 0$ for all $i \notin I_k$. With this observation, setting $d^k = 0$ and $t^k = 0$ in (7)-(11), and further $\mu_i^k = 0$ for $i \notin I_k$, we obtain KKT conditions for (1).

Suppose now that $d^k = 0$, $t^k > 0$, and $d^{k+q} = 0$, $t^{k+q} > 0$ for all integers q. Obviously, we then have $x^{k+q} = x^k$, by Step 4 of Algorithm 1. By (6), $p(x^{k+q}) = t^{k+q}(0) = t^{k+q} = p(x^k) = t^k > 0$. Then (11) implies that $\nu^{k+q} = 0$. Using further (8), we have that

$$\beta_{k+q} = \sum_{i \in I_{k+q}} \mu_i^{k+q}. \tag{15}$$

In Step 5 of Algorithm 1, since $\|d^{k+q}\|^{-1} = +\infty$, (15) implies that $\beta_{k+q} < r_{k+q}$. Hence, by (14),

$$\beta_{k+q+1} = \beta_{k+q} + \delta_2.$$

This shows that $\beta_{k+q} \to +\infty$ as $q \to \infty$, and by (15),

$$+\infty = \lim_{q \to \infty} \|\mu^{k+q}\|.$$

Observe that since $t^{k+q} = p(x^{k+q})$, (9) and (10) imply that $\mu_i^{k+q} = 0$ for $i \in I_{k+q} \setminus I(x^{k+q})$. Since $I(x^k) = I(x^{k+q}) \subset I_{k+q}$, (7) reduces to

$$f'(x^k) + \sum_{i \in I(x^k)} \mu_i^{k+q} g_i'(x^k) = 0.$$

Dividing both sides of the above equality by $\|\mu^{k+q}\|$ and passing onto the limit as $q \to \infty$, we conclude that

$$\sum_{i \in I(x^k)} \tilde{\mu}_i g_i'(x^k) = 0, \quad \tilde{\mu}_i \geq 0,\ i \in I(x^k), \quad \|\tilde{\mu}\| = 1.$$

Since $p(x^k) > 0$ implies that $\emptyset \neq I(x^k) \subset J(x^k)$, the latter contradicts Lemma 1, completing the proof. ∎

The following result shows that whenever $d^k \neq 0$, it is a descent direction for ψ_{β_k} at x^k. This, in turn, implies that the linesearch step is well-defined. Combining this fact with Proposition 2, it follows that the whole Algorithm 1 is well-defined.

Proposition 3 *Let f, $g_i : \Re^n \to \Re$, $i = 1, \ldots, m$, be continuously differentiable. Then in Algorithm 1 it holds that*

$$\begin{aligned} \psi'_{\beta_k}(x^k; d^k) &\leq \Delta_k - \frac{1}{2}\langle H^k d^k, d^k\rangle \\ &\leq -\langle H^k d^k, d^k\rangle - \frac{1}{2}\sum_{i\in I_k} \mu_i^k \langle G_i^k d^k, d^k\rangle - \nu^k p(x^k). \end{aligned} \tag{16}$$

As a consequence, Step 3 of Algorithm 1 is well-defined and terminates with some finite integer j_k.

Proof. As is well known and easy to see,

$$\psi'_{\beta_k}(x^k; d^k) = \langle f'(x^k), d^k\rangle + \beta_k \begin{cases} 0, & \text{if } I(x^k) = \emptyset \\ \max_{i\in I(x^k)}\langle g_i'(x^k), d^k\rangle, & \text{if } p(x^k) > 0 \\ \max_{i\in I(x^k)}[\langle g_i'(x^k), d^k\rangle]_+, & \text{if } p(x^k) = 0,\ I(x^k) \neq \emptyset \end{cases} . \tag{17}$$

Consider the three possible cases in (17).

If $I(x^k) = \emptyset$, then it must hold that $p(x^k) = 0$ (and $g_i(x^k) < 0$ for all i). Thus $0 \leq t^k = t^k - p(x^k)$.

If $I(x^k) \neq \emptyset$, by (9), for all $i \in I(x^k) \subset I_k$ it holds that

$$\begin{aligned} \langle g_i'(x^k), d^k\rangle &\leq t^k - g_i(x^k) - \frac{1}{2}\langle G_i^k d^k, d^k\rangle \\ &\leq t^k - p(x^k), \end{aligned}$$

where the second inequality follows from $g_i(x^k) = p(x^k)$ for $i \in I(x^k)$, and $G_i^k \succeq O$.

If further $p(x^k) = 0$, then $\langle g_i'(x^k), d^k\rangle \leq t^k$. Then by the monotonicity of $[\cdot]_+$,

$$[\langle g_i'(x^k), d^k\rangle]_+ \leq [t^k]_+ = t^k - p(x^k).$$

Combining the three cases, (17) gives that

$$\psi'_{\beta_k}(x^k; d^k) \leq \langle f'(x^k), d^k\rangle + \beta_k(t^k - p(x^k)), \tag{18}$$

which is the first inequality in the assertion (recalling the definition of Δ_k in (13)).

Multiplying both sides of (7) by d^k, we have that

$$\langle f'(x^k), d^k\rangle = -\langle H^k d^k, d^k\rangle - \sum_{i\in I_k} \mu_i^k(\langle g_i'(x^k), d^k\rangle + \langle G_i^k d^k, d^k\rangle). \tag{19}$$

We further obtain

$$\begin{aligned}
-\sum_{i\in I_k} \mu_i^k \langle g_i'(x^k), d^k\rangle &= \sum_{i\in I_k} \mu_i^k (g_i(x^k) + \frac{1}{2}\langle G_i^k d^k, d^k\rangle - t^k) \\
&\leq (p(x^k) - t^k)\sum_{i\in I_k} \mu_i^k + \frac{1}{2}\sum_{i\in I_k} \mu_i^k \langle G_i^k d^k, d^k\rangle \\
&= (\beta_k - \nu^k)(p(x^k) - t^k) + \frac{1}{2}\sum_{i\in I_k} \mu_i^k \langle G_i^k d^k, d^k\rangle \\
&= \beta_k(p(x^k) - t^k) - \nu^k p(x^k) + \frac{1}{2}\sum_{i\in I_k} \mu_i^k \langle G_i^k d^k, d^k\rangle,
\end{aligned}$$

where (10) was used in the first equality, (8) was used in the second equality, and (11) in the last. Combining the latter relation with (19) and (18), we obtain (16).

If $d^k = 0$, Step 3 of Algorithm 1 returns $j_k = 0$. If $d^k \neq 0$, for $\gamma \in [0,1]$ we have that

$$\begin{aligned}
\psi_{\beta_k}(x^k + \gamma d^k) &= \psi_{\beta_k}(x^k) + \gamma\psi'_{\beta_k}(x^k; d^k) + o(\gamma) \\
&\leq \psi_{\beta_k}(x^k) + \gamma\Delta_k + o(\gamma),
\end{aligned}$$

where the inequality is by (16). It follows that (12) is guaranteed to hold whenever $\gamma = \theta^j > 0$ satisfies

$$(1-\sigma)\gamma\Delta_k \leq o(\gamma).$$

Since $\Delta_k < 0$ when $d^k \neq 0$, the inequality above clearly holds for all $\gamma = \theta^j$ sufficiently small (i.e., all j sufficiently large). Hence, Step 3 of Algorithm 1 terminates with some finite integer j_k. ∎

We next establish that when close to the feasible region of (1), the solution of subproblem (4) is given by the solution of the subproblem without the slack variable:

$$\begin{array}{ll}
\text{minimize}_d & \langle f'(x^k), d\rangle + \frac{1}{2}\langle H^k d, d\rangle \\
\text{subject to} & g_i(x^k) + \langle g_i'(x^k), d\rangle + \frac{1}{2}\langle G_i^k d, d\rangle \leq 0, \quad i \in A_k,
\end{array} \tag{20}$$

where $A_k \subset \{1, \ldots, m\}$. This fact will be used later to establish that the penalty parameters $\{\beta_k\}$ stay fixed from some point on, as well as for the local rate of convergence analysis. The part of Proposition 4 concerning boundedness of solutions and Lagrange multipliers of (20) is closely related to [8, Lemma 3.3]. However, we consider the case where possibly $G_i^k \neq g_i''(x^k)$ and $A_k \neq \{1, \ldots, m\}$.

Proposition 4 *Let f, $g_i : \Re^n \to \Re$, $i = 1, \ldots, m$, be continuously differentiable. Suppose that g is convex and the Slater condition (2) is satisfied.*

Let $\{x^k\}$ be any sequence converging to some $\bar{x} \in \Re^n$ such that $p(\bar{x}) = 0$, and let the matrices in (20) satisfy $\rho_2 E \succeq H^k \succeq \rho_1 E$ and $\rho_2 E \succeq G_i^k \succeq O, i \in \{1, \ldots, m\}$, for all k, where $\rho_2 \geq \rho_1 > 0$.

Then problem (20) is feasible (hence, solvable) for all k sufficiently large and any $A_k \subset \{1, \ldots, m\}$. Moreover, if (s^k, λ^k) is a KKT point of (20), then for any $\beta_k \geq \|\lambda^k\|_1$, $(s^k, 0)$ is the unique solution of (4) with $I_k = A_k$. Conversely, if $(d^k, 0, \mu^k, \nu^k)$ is a KKT point of (4), then (d^k, μ^k) is a KKT point of (20).

Furthermore, the sequences $\{s^k\}$ and $\{\lambda^k\}$ are bounded.

Proof. Since $p(\bar{x}) = 0$, we have that $g_i(\bar{x}) = 0$ for $i \in I(\bar{x})$. By the convexity of g and (2), it then holds that

$$0 > g_i(\hat{x}) - g_i(\bar{x}) \geq \langle g_i'(\bar{x}), \hat{x} - \bar{x} \rangle, \quad i \in I(\bar{x}). \tag{21}$$

Taking $d(\eta) = \eta(\hat{x} - \bar{x})$, $\eta \in (0, 1]$, and using $\rho_2 E \succeq G_i^k$, we have that

$$\begin{aligned} & g_i(x^k) + \eta \langle g_i'(x^k), \hat{x} - \bar{x} \rangle + \frac{\eta^2}{2} \langle G_i^k (\hat{x} - \bar{x}), \hat{x} - \bar{x} \rangle \\ \leq \;& \eta(\langle g_i'(\bar{x}), \hat{x} - \bar{x} \rangle + \eta \rho_2 \|\hat{x} - \bar{x}\|^2/2) + \epsilon_i^k, \end{aligned} \tag{22}$$

where, as $k \to \infty$,

$$\epsilon_i^k = g_i(x^k) + \eta \langle g_i'(x^k) - g_i'(\bar{x}), \hat{x} - \bar{x} \rangle \to \left\{ \begin{array}{ll} g_i(\bar{x}) = 0, & i \in I(\bar{x}) \\ g_i(\bar{x}) < 0, & i \notin I(\bar{x}) \end{array} \right. . \tag{23}$$

For each $i \in \{1, \ldots, m\}$ there exists $\eta_i > 0$, and for $i \in I(\bar{x})$ also some $c_i > 0$, such that

$$\eta_i(\langle g_i'(\bar{x}), \hat{x} - \bar{x} \rangle + \eta_i \rho_2 \|\hat{x} - \bar{x}\|^2/2) \leq \left\{ \begin{array}{ll} -2c_i, & i \in I(\bar{x}) \\ -g_i(\bar{x})/4, & i \notin I(\bar{x}) \end{array} \right. ,$$

where (21) was used for $i \in I(\bar{x})$. Given those choices, for each $i \in \{1, \ldots, m\}$ there further exists an index k_i such that

$$\epsilon_i^k \leq \left\{ \begin{array}{ll} c_i, & i \in I(\bar{x}) \\ g_i(\bar{x})/2, & i \notin I(\bar{x}) \end{array} \right. , \quad \forall\, k \geq k_i.$$

Denoting

$$\bar{k} = \max_{i=1,\ldots,m} k_i, \quad \eta = \min_{i=1,\ldots,m} \eta_i, \quad c = \min\{\min_{i \in I(\bar{x})} c_i, -\max_{i \notin I(\bar{x})} g_i(\bar{x})/4\},$$

and using the two relations above and (22), we have that $d = \eta(\hat{x} - \bar{x})$ satisfies

$$g_i(x^k) + \langle g_i'(x^k), d\rangle + \frac{1}{2}\langle G_i^k d, d\rangle \leq -c < 0 \quad \forall\, k \geq \bar{k},\ i = 1, \ldots, m. \tag{24}$$

In particular, this shows that this d is (strictly) feasible in (20), for any choice of $A_k \subset \{1, \ldots, m\}$ and all $k \geq \bar{k}$.

Therefore, (20) is a feasible strongly convex program. Hence, it is (uniquely) solvable. Furthermore, by (24), it satisfies the Slater constraint qualification. Hence, there exist some $\lambda_i^k \in \Re$, $i \in A_k$, such that

$$f'(x^k) + H^k s^k + \sum_{i \in A_k} \lambda_i^k (g_i'(x^k) + G_i^k s^k) = 0, \tag{25}$$

$$g_i(x^k) + \langle g_i'(x^k), s^k\rangle + \frac{1}{2}\langle G_i^k s^k, s^k\rangle \leq 0, \quad \lambda_i^k \geq 0, \quad i \in A_k, \tag{26}$$

$$\lambda_i^k (g_i(x^k) + \langle g_i'(x^k), s^k\rangle + \frac{1}{2}\langle G_i^k s^k, s^k\rangle) = 0, \quad i \in A_k. \tag{27}$$

If $\beta_k \geq \|\lambda^k\|_1$, then conditions (25)-(27) imply that $d^k = s^k$, $t^k = 0$, $\nu^k = \beta_k - \|\lambda^k\|_1$ and $\mu_i^k = \lambda_i^k$, $i \in A_k = I_k$, satisfy KKT conditions (7)-(11). Hence, $(s^k, 0)$ is the solution of (4). Conversely, if $(d^k, 0, \mu^k, \nu^k)$ is a KKT point of (4), then (d^k, μ^k) is a KKT point of (20) for $A_k = I_k$, by direct inspection.

Next, note that since $d = \eta(\hat{x} - \bar{x})$ is feasible in (20) for all $k \geq \bar{k}$, we have that

$$\begin{aligned} \langle f'(x^k), d\rangle + \frac{1}{2}\langle H^k d, d\rangle &\geq \langle f'(x^k), s^k\rangle + \frac{1}{2}\langle H^k s^k, s^k\rangle \\ &\geq \|s^k\|(\rho_1\|s^k\|/2 - \|f'(x^k)\|), \end{aligned}$$

where the second inequality is by $H^k \succeq \rho_1 E$. Since $\{f'(x^k)\}$ is bounded and $\rho_2 E \succeq H^k$, the above relation implies that $\{s^k\}$ must be bounded.

Suppose now that $\{\lambda^k\}$ is unbounded. We augment (25) by formally setting $\lambda_i^k = 0$ for $i \in \{1, \ldots, m\} \setminus A_k$. Passing onto a subsequence, if necessary, we can assume that $\|\lambda^k\| \to \infty$ and $\{s^k\} \to \bar{s}$, $\{H^k\} \to H$, $\{G_i^k\} \to G_i$. Dividing both sides of (25) by $\|\lambda^k\|$ and passing onto the limit as $k \to \infty$, we obtain that

$$0 = \sum_{i=1}^m \bar{\lambda}_i (g_i'(\bar{x}) + G_i \bar{s}), \quad \bar{\lambda} \geq 0, \quad \|\bar{\lambda}\| = 1.$$

By Gordan Theorem of the Alternatives [11], the latter is equivalent to

$$\nexists\, s \in \Re^n \text{ such that } \langle g_i'(\bar{x}) + G_i \bar{s}, s\rangle < 0 \text{ for } i \in \{i \mid \bar{\lambda}_i > 0\}. \tag{28}$$

Clearly, $\bar{\lambda}_i > 0$ means that there exists an infinite subsequence of indices $\{k_j\}$ such that $i \in A_{k_j}$ and $\lambda_i^{k_j} > 0$. For such i, (27) implies that

$$g_i(x^{k_j}) + \langle g_i'(x^{k_j}), s^{k_j}\rangle + \frac{1}{2}\langle G_i^{k_j} s^{k_j}, s^{k_j}\rangle = 0 \quad \forall\, j.$$

Passing onto the limit as $j \to \infty$, we obtain that

$$g_i(\bar{x}) + \langle g_i'(\bar{x}), \bar{s}\rangle + \frac{1}{2}\langle G_i \bar{s}, \bar{s}\rangle = 0, \quad i \in \{i \mid \bar{\lambda}_i > 0\}. \tag{29}$$

Passing onto the limit as $k \to \infty$ in (24), we also have that

$$g_i(\bar{x}) + \langle g_i'(\bar{x}), d\rangle + \frac{1}{2}\langle G_i d, d\rangle \le -c < 0.$$

Substracting (29) from the latter inequality, we have that

$$\begin{aligned} 0 &> \langle g_i'(\bar{x}), d - \bar{s}\rangle + \frac{1}{2}\langle G_i d, d\rangle - \frac{1}{2}\langle G_i \bar{s}, \bar{s}\rangle \\ &= \langle g_i'(\bar{x}) + G_i \bar{s}, d - \bar{s}\rangle, \quad i \in \{i \mid \bar{\lambda}_i > 0\}, \end{aligned}$$

which contradicts (28). We conclude that $\{\lambda^k\}$ is bounded.

■

By (14), either β_k is constant starting from some iteration index k_0, or it diverges to $+\infty$. We next show that the latter case cannot occur.

Proposition 5 *Let f, $g_i : \Re^n \to \Re$, $i = 1, \ldots, m$, be continuously differentiable. Suppose that g is convex and that the Slater condition* (2) *is satisfied.*

Let the sequence $\{x^k\}$ generated by Algorithm 1 be bounded, and $\rho_2 E \succeq H^k \succeq \rho_1 E$, $\rho_2 E \succeq G_i^k \succeq O, i \in I_k$, for all k, where $\rho_2 \ge \rho_1 > 0$.

Then there exists some iteration index k_0 such that

$$\beta_k = \beta_{k_0} \quad \forall\, k \ge k_0.$$

Proof. Suppose the opposite, i.e., $\beta_k \to +\infty$ as $k \to \infty$. Then (14) implies that $\beta_k < r_k = \min\{\|d^k\|^{-1}, \|\mu^k\|_1 + \delta_1\}$ happens an infinite number of times. It then further follows that there exists a subsequence of iteration indices $\{k_j\}$ such that

$$0 = \lim_{j\to\infty} d^{k_j} \quad \text{and} \quad +\infty = \lim_{j\to\infty} \|\mu^{k_j}\|. \tag{30}$$

Taking a further subsequence, if necessary, we can assume that $\{x^{k_j}\} \to \tilde{x}$ as $j \to \infty$. We next consider the two possible cases: $p(\tilde{x}) > 0$ and $p(\tilde{x}) = 0$.

Let $p(\tilde{x}) > 0$, and denote

$$J(\tilde{x}) := \{i \mid g_i(\tilde{x}) > 0\} \ne \emptyset.$$

Note that, by the continuity of g, $J(\tilde{x}) \cap I_{k_j} \ne \emptyset$ for all j large enough. By (9),

$$g_i(x^{k_j}) + \langle g_i'(x^{k_j}), d^{k_j}\rangle + \frac{1}{2}\langle G_i^{k_j} d^{k_j}, d^{k_j}\rangle \le t^{k_j},\ i \in I_{k_j}.$$

For $i \in I_{k_j} \setminus J(\tilde{x})$, as $j \to \infty$ the left-hand side of the inequality above tends to $g_i(\tilde{x}) \le 0$ (taking into account (30) and $\rho_2 E \succeq G_i^k \succeq O$), while the right-hand side tends to $p(\tilde{x}) > 0$ (recall (6)). Hence, such constraints are inactive for all j large enough and, by (10),

$$\mu_i^{k_j} = 0, \quad i \in I_{k_j} \setminus J(\tilde{x}).$$

Formally setting $\mu_i^{k_j} = 0$ and $G_i^{k_j} = 0$ for $i \in J(\tilde{x}) \setminus I_{k_j}$, we can write (7) as

$$f'(x^{k_j}) + H^{k_j} d^{k_j} + \sum_{i \in J(\tilde{x})} \mu_i^{k_j} (g_i'(x^{k_j}) + G_i^{k_j} d^{k_j}) = 0.$$

Dividing both sides of this equality by $\|\mu^{k_j}\|$ and passing onto the limit as $j \to \infty$, by (30) and $\rho_2 E \succeq H^k, G_i^k \succeq O$, we obtain that

$$0 = \sum_{i \in J(\tilde{x})} \tilde{\mu} g_i'(\tilde{x}), \quad \tilde{\mu}_i \ge 0,\, i \in J(\tilde{x}), \quad \|\tilde{\mu}\| = 1,$$

which contradicts Lemma 1.

Suppose now that $p(\tilde{x}) = 0$. By Proposition 4, subproblems (20) with $A_{k_j} = I_{k_j}$ are solvable for all indices j large enough. Let (s^{k_j}, λ^{k_j}) be a KKT point of (20). In particular, by Proposition 4, $\{\lambda^{k_j}\}$ is bounded. Since $\beta_k \to +\infty$, we have that $\beta_{k_j} > \|\lambda^{k_j}\|_1$ and, again by Proposition 4, the unique solution of (4) is $d^{k_j} = s^{k_j}$, $t^{k_j} = 0$, with multipliers (μ^{k_j}, ν^{k_j}) generated by Algorithm 1. But then (s^{k_j}, μ^{k_j}) is a KKT point of (20), again by Proposition 4. However, unboundedness of $\{\mu^{k_j}\}$ (see (30)) contradicts Proposition 4.

The proof is complete. ∎

We are now ready to establish global convergence of Algorithm 1. Our results are stronger than [8, Theorem 3.4] in some respects. First, we do not assume the convexity of f or twice differentiability of f and g. Second, we prove that the stepsizes stay bounded away from zero, which is a desirable property not established in [8]. Additionally, even if the data is twice differentiable, we do not assume that $G_i^k = g_i''(x^k)$ and/or that the matrices $2H^k - f''(x^k) + \sum_{i=1}^m \mu_i^k g_i''(x^k)$ are uniformly positive definite, as required in [8, Theorem 3.4]. Both Theorem 6 below and [8, Theorem 3.4], assume the uniform boundedness and positive definiteness of H^k, and boundedness of the sequence $\{x^k\}$.

Theorem 6 *Let f, g_i, $i = 1, \ldots, m$, be differentiable functions with Lipschitz-continuous derivatives. Let g be convex, and the Slater condition (2) be satisfied.*

Suppose further that in Algorithm 1 we have $\rho_2 E \succeq H^k \succeq \rho_1 E$ and $\rho_2 E \succeq G_i^k \succeq O, i \in I_k$, for all k, where $\rho_2 \ge \rho_1 > 0$, and that $I_k = \{1, \ldots, m\}$ for $k \ge k_1$.

If the sequence $\{x^k\}$ generated by Algorithm 1 is bounded, then the sequence $\{\mu^k\}$ is also bounded and every accumulation point of $\{(x^k, \mu^k)\}$ is a KKT point for problem

(1). *In particular, if f is convex then every accumulation point of $\{x^k\}$ is a solution of* (1).

Proof. We first show that the sequence of stepsizes $\{\gamma_k\}$ is bounded away from zero. Let $L > 0$ be the Lipschitz constant of f', g_i', $i = 1, \dots, m$. For all $\gamma \in [0, 1]$ and any i, we have that

$$\begin{aligned} & [g_i(x^k + \gamma d^k)]_+ - [g_i(x^k) + \gamma\langle g_i'(x^k), d^k\rangle]_+ \\ \le\ & [g_i(x^k + \gamma d^k) - g_i(x^k) - \gamma\langle g_i'(x^k), d^k\rangle]_+ \\ \le\ & |g_i(x^k + \gamma d^k) - g_i(x^k) - \gamma\langle g_i'(x^k), d^k\rangle| \\ \le\ & \frac{L}{2}\gamma^2\|d^k\|^2, \end{aligned} \tag{31}$$

where the first inequality follows from the fact that $[a - b]_+ \ge [a]_+ - [b]_+\ \forall\, a, b \in \Re$; and the last inequality is by the Lipschitz-continuity of g_i' (e.g., [3, Proposition A.24]). For $i \in I_k$, we further obtain

$$\begin{aligned} & [g_i(x^k) + \gamma\langle g_i'(x^k), d^k\rangle]_+ \\ =\ & [\gamma(g_i(x^k) + \langle g_i'(x^k), d^k\rangle) + (1-\gamma)g_i(x^k)]_+ \\ \le\ & \gamma[g_i(x^k) + \langle g_i'(x^k), d^k\rangle]_+ + (1-\gamma)[g_i(x^k)]_+ \\ \le\ & \gamma[g_i(x^k) + \langle g_i'(x^k), d^k\rangle + \frac{1}{2}\langle G_i^k d^k, d^k\rangle]_+ + (1-\gamma)[g_i(x^k)]_+ \\ \le\ & \gamma t^k + (1-\gamma)[g_i(x^k)]_+, \end{aligned} \tag{32}$$

where the second inequality is by the convexity of $[\cdot]_+$; and the last two are by the monotonicity of $[\cdot]_+$ (taking also into account $G_i^k \succeq O$ and (9), (11)). Combining (31) and (32), we conclude that

$$[g_i(x^k + \gamma d^k)]_+ \le \gamma t^k + (1-\gamma)[g_i(x^k)]_+ + \frac{L}{2}\gamma^2\|d^k\|^2, \quad i \in I_k.$$

By the Lipschitz-continuity of f', it also holds that

$$f(x^k + \gamma d^k) \le f(x^k) + \gamma\langle f'(x^k), d^k\rangle + \frac{L}{2}\gamma^2\|d^k\|^2.$$

For $k \ge k_1$, we then have that

$$\begin{aligned} \psi_{\beta_k}(x^k + \gamma d^k) &= f(x^k + \gamma d^k) + \beta_k p(x^k + \gamma d^k) \\ &\le f(x^k) + \gamma\langle f'(x^k), d^k\rangle + \beta_k(1-\gamma)p(x^k) + \gamma\beta_k t^k + \frac{L(1+\beta_k)}{2}\gamma^2\|d^k\|^2 \\ &= \psi_{\beta_k}(x^k) + \gamma(\Delta_k - \frac{1}{2}\langle H^k d^k, d^k\rangle) + \frac{L(1+\beta_k)}{2}\gamma^2\|d^k\|^2 \\ &\le \psi_{\beta_k}(x^k) + \gamma\Delta_k + \frac{L(1+\beta_k)}{2}\gamma^2\|d^k\|^2, \end{aligned}$$

where $I_k = \{1, \ldots, m\}$ was used in the first inequality, the second equality is by (13), and the last inequality is by $H^k \succeq O$.

By a direct comparison of the latter relation with (12), we conclude that (12) is guaranteed to be satisfied once j is large enough, so that $\gamma = \theta^j$ falls within the set of γ satisfying

$$\frac{L(1+\beta_k)}{2}\gamma\|d^k\|^2 \leq -(1-\sigma)\Delta_k.$$

In particular, since Step 3 of Algorithm 1 did not accept the stepsize value $\theta^{k_j-1} > \theta^{k_j}$, it must be the case that

$$k_j = 0 \quad \text{or} \quad \gamma_k/\theta = \theta^{k_j-1} > \frac{-2(1-\sigma)\Delta_k}{L(1+\beta_k)\|d^k\|^2}.$$

Hence,

$$\gamma_k \geq \min\left\{1,\ \frac{-2\theta(1-\sigma)\Delta_k}{L(1+\beta_k)\|d^k\|^2}\right\}, \quad k \geq k_1.$$

By Proposition 5, $\beta_k = \beta_{k_0}$ for all $k \geq k_0$ and some iteration index k_0. Furthermore, by (16) and $H^k \succeq \rho_1 E$,

$$-\Delta_k \geq \frac{1}{2}\langle H^k d^k, d^k\rangle \geq \rho_1\|d^k\|^2/2. \tag{33}$$

We conclude that

$$\gamma_k \geq \min\left\{1,\ \frac{(1-\sigma)\theta\rho_1}{L(1+\beta_{k_0})}\right\} = \bar{\gamma} > 0 \quad \forall k \geq k_2 := \max\{k_0, k_1\}. \tag{34}$$

Now taking into account (12), (33) and (34), for $k \geq k_2$ we obtain that

$$\begin{array}{rcl} \psi_{\beta_{k_0}}(x^k) - \psi_{\beta_{k_0}}(x^{k+1}) & \geq & -\sigma\gamma_k\Delta_k \\ & \geq & \sigma\bar{\gamma}\rho_1\|d^k\|^2/2. \end{array} \tag{35}$$

Since $\{x^k\}$ is bounded, $\{\psi_{\beta_{k_0}}(x^k)\}$ is also bounded, by continuity. By (35), it is also nonincreasing. Hence, it converges. Then (35) further implies that

$$0 = \lim_{k\to\infty} d^k. \tag{36}$$

Furthermore, by Step 5 of Algorithm 1, the fact that $\beta_{k+1} = \beta_k = \beta_{k_0}$ implies that

$$\beta_{k_0} \geq r_k = \min\{\|d^k\|^{-1},\ \|\mu^k\|_1 + \delta_1\}.$$

Taking into account (36), we conclude that there exists some $k_3 \geq k_0$ such that

$$\beta_{k_0} > \sum_{i\in I_k} \mu^k \quad \forall\, k \geq k_3.$$

In particular, this means that $\{\mu^k\}$ is bounded and, using (8), that

$$\nu^k > 0 \quad \forall\, k \geq k_3.$$

By (11), we then also have that

$$t^k = 0 \quad \forall\, k \geq k_3. \tag{37}$$

Now passing onto the limit as $k \to \infty$ in (7)-(11) (where $I_k = \{1, \ldots, m\}$), we obtain that every accumulation point of $\{(x^k, \mu^k)\}$ satisfies the KKT conditions for problem (1). If f is further convex, then every accumulation point of $\{x^k\}$ is a solution of (1). ■

As an immediate consequence of Theorem 6, we have the following fact.

Corollary 7 *Suppose that the assumptions of Theorem 6 are satisfied (except for the boundedness of $\{x^k\}$). If problem (1) has no KKT points, then the sequence $\{x^k\}$ generated by Algorithm 1 must be unbounded.*

3 Local Convergence

We next show that if the matrices H^k and G_i^k are (in a certain sense) asymptotically good approximations of $f''(x^k)$ and $g_i''(x^k)$, respectively, then in Algorithm 1 we have $\gamma_k = 1$ for all k sufficiently large. This implies that the Maratos effect does not occur. Then, under certain further assumptions, the quadratic rate of convergence is achieved.

Note that given the assumptions on H^k needed for global convergence, the first condition in (38) implicitly subsumes that $f''(x^k)$ is positive definite for k sufficiently large (unless some further analysis involving d^k is performed). The latter is consistent with the assumptions of Theorem 6 below on the quadratic rate of convergence. In Algorithm 1, we can take the positive definite regularization $H^k = f''(x^k) + \rho_k E$, decreasing $\rho_k \geq 0$ appropriately as x^k enters the region where $f''(x)$ is positive definite. This strategy is consistent with both the global and local convergence theory.

Approximation conditions (38) also resemble the quasi-Newton conditions of Dennis-Moré, but specific update rules are beyond the scope of this paper. Here, we can mention the situation where computing the second derivatives involves some numerical work (e.g., the approximation is done by finite differences). In that case, (38) imposes the natural requirement on the accuracy of such approximation.

Proposition 8 *Let the assumptions of Theorem 6 be satisfied. Suppose, in addition, that f and g are twice continuously differentiable, and*

$$\langle (f''(x^k) - H^k)d^k, d^k\rangle = o(\|d^k\|^2), \quad \langle (g_i''(x^k) - G_i^k)d^k, d^k\rangle = o(\|d^k\|^2),\ i \in I_k. \tag{38}$$

Then there exists an iteration index k_4 such that

$$\gamma_k = 1 \quad \forall\, k \geq k_4.$$

Proof. By Theorem 6, $\beta_k = \beta_{k_0}$, $t^k = 0$ for $k \geq k_3$ (where k_3 is defined above), and $d^k \to 0$.

For $i \in I_k$ and $k \geq k_3$, we have that

$$\begin{aligned}
[g_i(x^k + d^k)]_+ &= [g_i(x^k + d^k)]_+ - [g_i(x^k) + \langle g_i'(x^k), d^k\rangle + \frac{1}{2}\langle G_i^k d^k, d^k\rangle]_+ \\
&\leq |g_i(x^k + d^k) - g_i(x^k) - \langle g_i'(x^k), d^k\rangle - \frac{1}{2}\langle g_i''(x^k) d^k, d^k\rangle| \\
&\quad + \frac{1}{2}|\langle (g_i''(x^k) - G_i^k) d^k, d^k\rangle| \\
&= o(\|d^k\|^2),
\end{aligned} \tag{39}$$

where the first equality follows from (9) where $t^k = 0$, and the inequality follows from the fact that $[a]_+ - [b]_+ \leq |a - b|$ for any $a, b \in \Re$. Since $I_k = \{1, \ldots, m\}$ for $k \geq k_3$, we further obtain that

$$\begin{aligned}
\psi_{\beta_{k_0}}(x^k + d^k) &= f(x^k + d^k) + \beta_{k_0} p(x^k + d^k) \\
&= f(x^k + d^k) + o(\|d^k\|^2) \\
&= f(x^k) + \langle f'(x^k), d^k\rangle + \frac{1}{2}\langle f''(x^k) d^k, d^k\rangle + o(\|d^k\|^2) \\
&= \psi_{\beta_{k_0}}(x^k) - \beta_{k_0} p(x^k) + \langle f'(x^k), d^k\rangle + \frac{1}{2}\langle H^k d^k, d^k\rangle + o(\|d^k\|^2) \\
&= \psi_{\beta_{k_0}}(x^k) + \Delta_k + o(\|d^k\|^2),
\end{aligned}$$

where the second equality is by (39), the fourth equality is by (38), and the last is by the definition of Δ_k given in (13) and $t^k = 0$.

In view of the last relation and (12), to establish that $\gamma_k = 1$ in Algorithm 1, it is enough to show that

$$(1 - \sigma)\Delta_k + o(\|d^k\|^2) \leq 0 \tag{40}$$

for all k sufficiently large. By (16) and $H^k \succeq \rho_1 E$, we have that

$$\Delta_k \leq -\frac{1}{2}\langle H^k d^k, d^k\rangle \leq -\rho_1 \|d^k\|^2/2,$$

which verifies (40). ∎

Remark 9 *If we* assume *that the method converges, then some of the assumptions in Proposition 8 can be relaxed. First, one no longer has to ask that* $I_k = \{1, \ldots, m\}$ *for* k *large enough. This is easy to see, because* $d^k \to 0$*, and for* $i \notin I(x^k)$ *we then have* $g_i(x^k) + \langle g_i'(x^k), d^k \rangle + \frac{1}{2}\langle G_i^k d^k, d^k \rangle < 0$ *for* k *large enough. Hence, the first equality leading to* (39) *is then valid also for* $i \in \{1, \ldots, m\} \setminus I_k$*. Then the rest of the proof applies.*

Additionally, instead of $H^k \succeq \rho_1 E$*, we could assume that* $H^k + \sum_{i \in I_k} \mu_i^k G_i^k \succeq \rho_1 E$*. Indeed,* (16) *implies*

$$\Delta_k \leq -\frac{1}{2}\langle (H^k + \sum_{i \in I_k} \mu_i^k G_i^k) d^k, d^k \rangle,$$

from which then (40) *also follows.*

Theorem 6 (recall in particular (37)) and Proposition 4 imply that Algorithm 1 eventually reduces to SQCQP method based on subproblems (20) without the slack variable. Furthermore, by Proposition 8, the unit stepsize is always accepted from some point on. The local quadratic rate of convergence now follows essentially from [8, Theorem 4.7].

Theorem 10 *Let* f, g_i, $i = 1, \ldots, m$, *be convex twice continuously differentiable functions with Lipschitz-continuous second derivatives, and suppose that the Slater condition* (2) *holds. Let* $\bar{x} \in \Re^n$ *be the (unique) solution of* (1) *such that* $f''(\bar{x}) + \sum_{i=1}^m \bar{\mu} g_i''(\bar{x})$ *is positive definite for any* $\bar{\mu} \in M$*, where* $M \subset \Re^m$ *is the set of Lagrange multipliers assiciated with* $\bar{x}$*.*

Then there exists a neighbourhood X *of* $\bar{x}$ *such that the sequence* $\{x^k\}$ *generated by Algorithm 1 with* $x^0 \in X$ *and* $H^k = f''(x^k)$, $G_i^k = g_i''(x^k)$ *for all* k*, converges to* $\bar{x}$ *at a quadratic rate. Furthermore,* $\mathrm{dist}(\mu^k, M) \to 0$ *at a quadratic rate.*

4 Concluding Remarks

We presented a sequential quadratically constrained quadratic programming algorithm with a number of desirable convergence and implementation features. Specifically,

- The feasibility of subproblems is automatic.
- Convexity of the objective function is not needed for global convergence.
- Twice differentiability of the problem data is not needed for global convergence.
- The stepsize in the linesearch procedure is always bounded away from zero.
- If the problem data is twice continuously differentiable and the method uses asymptotically good approximations of second-order derivatives, then the unit stepsize is eventually always accepted.

- Thus the Maratos effect does not occur, and quadratic convergence is achieved under certain assumptions.

Acknowledgement. The author thanks the two anonymous referees for their comments which led to an improvement of the original version of this paper.

References

[1] M. Anitescu. On the rate of convergence of sequential quadratic programming with nondifferentiable exact penalty function in the presence of constraint degeneracy. *Mathematical Programming*, 92:359–386, 2002.

[2] M. Anitescu. A superlinearly convergent sequential quadratically constrained quadratic programming algorithm for degenerate nonlinear programming. *SIAM Journal on Optimization*, 12:949–978, 2002.

[3] D.P. Bertsekas. *Nonlinear programming.* Athena Scientific, Belmont, Massachusetts, 1995.

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

[5] J. F. Bonnans. Asymptotic admissibility of the unit stepsize in exact penalty methods. *SIAM Journal on Control and Optimization*, 27:631–641, 1989.

[6] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Numerical Optimization: Theoretical and Practical aspects.* Springer–Verlag, Berlin, Germany, 2003.

[7] M. Fukushima. A successive quadratic programming algorithm with global and superlinear convergence properties. *Mathematical Programming*, 35:253–264, 1986.

[8] M. Fukushima, Z.-Q. Luo, and P. Tseng. A sequential quadratically constrained quadratic programming method for differentiable convex minimization. *SIAM Journal on Optimization*, 13:1098–1119, 2003.

[9] S. Kruk and H. Wolkowicz. Sequential, quadratically constrained, quadratic programming for general nonlinear programming. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming*, pages 563–575. Kluwer Academic Publishers, 2000.

[10] M.S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284:193–228, 1998.

[11] O.L. Mangasarian. *Nonlinear Programming.* McGraw–Hill, New York, 1969.

[12] N. Maratos. *Exact Penalty Function Algorithms for Finite Dimensional and Control Optimization Problems.* PhD thesis, Imperial College, University of London, 1978.

[13] D.Q. Mayne and E. Polak. A superlinearly convergent algorithm for constrained optimization problems. *Mathematical Programming Study*, 16:45–61, 1982.

[14] 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. *Mathematical Programming*, 88:61–83, 2000.

[15] Y.E. Nesterov and A.S. Nemirovskii. *Interior point polynomial methods in convex programming: theory and applications.* SIAM Publications, Philadelphia, PA, 1993.

[16] V.M. Panin. A second-order method for discrete min-max problem. *USSR Computational Mathematics and Mathematical Physics*, 19:90–100, 1979.

[17] V.M. Panin. Some methods of solving convex programming problems. *USSR Computational Mathematics and Mathematical Physics*, 21:57–72, 1981.

[18] M.J.D. Powell. Variable metric methods for constrained optimization. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical Programming: The State of Art*, pages 288–311. Springer–Verlag, Berlin, 1983.

[19] T. Tsuchiya. A convergence analysis of the scale-invariant primal-dual path-following algorithms for second-order cone programming. *Optimization Methods and Software*, 11:141–182, 1999.

[20] E.J. Wiest and E. Polak. A generalized quadratic programming-based phase-I–phase-II method for inequality-constrained optimization. *Applied Mathematics and Optimization*, 26:223–252, 1992.