

Global convergence of an SQP method without boundedness assumptions on any of the iterative sequences

Mikhail V. Solodov

Received: 24 April 2006 / Accepted: 14 June 2007 / Published online: 27 July 2007
© Springer-Verlag 2007

Abstract Usual global convergence results for sequential quadratic programming (SQP) algorithms with linesearch rely on some a priori assumptions about the generated sequences, such as boundedness of the primal sequence and/or of the dual sequence and/or of the sequence of values of a penalty function used in the linesearch procedure. Different convergence statements use different combinations of assumptions, but they all assume boundedness of at least one of the sequences mentioned above. In the given context boundedness assumptions are particularly undesirable, because even for non-pathological and well-behaved problems the associated penalty functions (whose descent is used to produce primal iterates) may not be bounded below for any value of the penalty parameter. Consequently, boundedness assumptions on the iterates are not easily justifiable. By introducing a very simple and computationally cheap safeguard in the linesearch procedure, we prove boundedness of the primal sequence in the case when the feasible set is nonempty, convex, and bounded. If, in addition, the Slater condition holds, we obtain a complete global convergence result without any a priori assumptions on the iterative sequences. The safeguard consists of not accepting a further increase of constraints violation at iterates which are infeasible beyond a chosen threshold, which can always be ensured by the proposed modified SQP linesearch criterion.

Keywords Sequential quadratic programming · Global convergence · Nonsmooth penalty function · Linesearch · Slater condition

The author is supported in part by CNPq Grants 301508/2005-4, 490200/2005-2, 550317/2005-8, by PRONEX–Optimization, and by FAPERJ Grant E-26/151.942/2004.

M. V. Solodov (✉)
IMPA, Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110,
Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil
e-mail: solodov@impa.br

Mathematics Subject Classification (2000) 90C30 · 65K05

1 Introduction

Consider the nonlinear programming problem

$$\begin{aligned} & \text{minimize } f(x) \\ & \text{subject to } x \in D = \{x \in \mathfrak{R}^n \mid g_i(x) \leq 0, i = 1, \dots, m\}, \end{aligned} \quad (1)$$

where $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ and $g : \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ are differentiable. One of the classical, yet practically efficient, approaches to solving (1) is the sequential quadratic programming (SQP) method (see, e.g., [1, Chap. 4] and [4, Part III]). The development and analysis of the method go back to [12] and to [5, 6, 8–11], as well as numerous subsequent references.

Given some current primal iterate $x^k \in \mathfrak{R}^n$, SQP computes the primal direction of change $d^k \in \mathfrak{R}^n$ as the solution of

$$\begin{aligned} & \text{minimize } \langle f(x^k), d \rangle + \frac{1}{2} \langle H_k d, d \rangle \\ & \text{subject to } d \in D_k = \{d \in \mathfrak{R}^n \mid g_i(x^k) + \langle g'_i(x^k), d \rangle \leq 0, i = 1, \dots, m\}, \end{aligned} \quad (2)$$

where H_k is an $n \times n$ symmetric positive definite matrix (For the purposes of this discussion, we assume that $D_k \neq \emptyset$. We note that this is always the case when g is convex and $D \neq \emptyset$. Also, having in mind global convergence properties, we shall not discuss any specifics concerning the choice of H_k . Those specifics are much more relevant for local convergence rates). In one of the most common classes of SQP methods, the next iterate x^{k+1} is computed by a linesearch procedure in the obtained direction d^k , in order to decrease the value (with respect to x^k) of some nonsmooth penalty function

$$\psi_{\beta_k} : \mathfrak{R}^n \rightarrow \mathfrak{R}, \quad \psi_{\beta_k}(x) = f(x) + \beta_k p(x), \quad \beta_k > 0, \quad (3)$$

where β_k is the current penalty parameter and the measure p of constraints violation can be given, for example, by

$$p : \mathfrak{R}^n \rightarrow \mathfrak{R}_+, \quad p(x) = \sum_{i=1}^m \max\{0, g_i(x)\}. \quad (4)$$

To ensure that d^k is a direction of descent for ψ_{β_k} at x^k , the standard way is to choose

$$\beta_k \geq \|\mu^k\|_\infty + \delta, \quad \delta > 0, \quad (5)$$

where $\mu^k \in \mathfrak{R}_+^m$ is a Lagrange multiplier associated to the solution d^k of (2).

Ideally (and to be really satisfactory from the mathematical standpoint), a global convergence result should show that the iterates approach stationary points of the problem under reasonable assumptions on the problem data (perhaps boundedness

of some level sets or of the feasible set, perhaps some convexity, etc.). Despite rich literature and widely acknowledged practical usefulness, this is not the case for SQP. To prove global convergence of the method outlined above, some of the following a priori assumptions on the iterative sequences, in this or that combination, are made in the literature:

- (A1) $\beta_k = \beta > 0$ for all $k \geq k_0$ (or equivalently, $\{\mu^k\}$ is bounded);
- (A2) $\lim_{k \rightarrow \infty} \psi_\beta(x^k) > -\infty$;
- (A3) $\{x^k\}$ is bounded.

Assuming (A3), the assumption (A2) becomes automatic, of course (for β fixed). Furthermore, (A3) goes a long way to help justify (A1) in some situations, as will be discussed below. However, assuming (A2) or (A3) cannot be regarded really satisfactory in the context of SQP for the following reason. It is perfectly possible that for non-pathological and well-behaved problems,

$$\inf_{x \in \mathbb{R}^n} \psi_\beta(x) = -\infty,$$

no matter what $\beta > 0$ is. Therefore, as $\{x^k\}$ is constructed by descent steps for ψ_{β_k} , assuming (A2) or (A3) cannot be considered justifiable in this context. Consider, for example, the problem

$$\text{minimize } x^3 \text{ subject to } x^2 \leq 1. \tag{6}$$

Evidently, $\psi_\beta(x) = x^3 + \beta \max\{0, x^2 - 1\}$ is unbounded below for any $\beta > 0$. Hence, given that each SQP iteration is a descent step for ψ_{β_k} , one cannot take the possibility of $\lim_{k \rightarrow \infty} \psi_{\beta_k}(x^k) = -\infty$ lightly. This is quite disturbing, since the problem above satisfies just about any reasonable assumption one may want: we would like to compute a stationary point of a polynomial function on a feasible set which is compact, convex and satisfies the Slater condition. Hardly anything more could be asked. However, to the best of our knowledge, there is no result in the literature which could claim convergence of SQP for the presented example *by looking at the problem data only*, at least when it comes to the basic SQP scheme outlined above. We shall give a theoretical justification of why SQP works for problems like (6).

Before proceeding, we survey some of the typical global convergence statements for basic (i.e., without complex modifications) SQP methods.

In [4, Theorem 15.2], the following situations are listed as possible for global behaviour of SQP (we simplify the statement by assuming that the problem data is defined on the whole space and that (1) has stationary points):

- (i) $\{\beta_k\}$ is unbounded, in which case so is $\{\mu^k\}$;
- (ii) $\beta_k = \beta > 0$ for all $k \geq k_0$, and one of the following holds
 - (a) $\lim_{k \rightarrow \infty} \psi_\beta(x^k) = -\infty$,
 - (b) $\lim_{k \rightarrow \infty} L'_x(x^k, \mu^k) = 0$ and $\lim_{k \rightarrow \infty} \max\{\mu_i^k, -g_i(x^k)\} = 0$ for all $i = 1, \dots, m$.

Clearly, out of all the possibilities, only the last one is satisfactory: it means that the residual of the Karush-Kuhn-Tucker (KKT) optimality conditions for problem (1) goes to zero, i.e., the iterates approach the set of stationary points of (1). Therefore,

to claim convergence, (A1) and (A2) need to be assumed or established, in order to eliminate the situations (i) and (ii a).

There are many other SQP global convergence statements in the literature, but it seems that all of them (again, at least for simple schemes) use some of (A1)–(A3), perhaps proving the rest, in order to eliminate the unsatisfactory outcomes (i) and (ii a). We shall cite only a few more. In [1, Proposition 4.14(a)], (A1) is assumed and the statement deals with accumulation points of $\{x^k\}$ without proving their existence (therefore, (A3) is needed for the result not to become vacuous). As already mentioned, (A1) is actually not difficult to guarantee under reasonable assumptions, *if (A3) is assumed*. An example of this is [1, Proposition 4.14(b)], where g is convex and satisfies the Slater condition

$$\exists \hat{x} \in \mathfrak{N}^n \text{ such that } g_i(\hat{x}) < 0, \quad i = 1, \dots, m. \quad (7)$$

Under convexity of g , (7) and (A3), convergence follows. When, in addition to convexity of g and the Slater condition (7), the feasible set D is compact and f is bounded below on \mathfrak{N}^n , (A3) had been established in [6, Theorem 3.3] (assuming also (A1) and using approximate minimization of ψ_β in the direction d^k instead of an implementable linesearch). Of course, boundedness below of f on \mathfrak{N}^n is just another way of assuming (A2). It should be noted here that while boundedness below of f on the feasible set D is a natural assumption necessary for existence of solutions of the problem, boundedness of f on \mathfrak{N}^n is a completely artificial requirement for a constrained problem. Summarizing, (A2) is still needed as an assumption even in the presence of (A1).

Let us go back to example (6). The possible trouble, of course, could come from points $x^k < -1$. Observe, however, that if $x^k < -1$ then $d^k \geq (1 - (x^k)^2)/(2x^k) > 0$. In particular, from the point $x^k < -1$ the direction d^k points to the feasible set D of (1). It can be further seen that for $x^{k+1} = x^k + \alpha_k d^k$ with *any* $\alpha_k \in (0, 1]$, it holds that $p(x^{k+1}) < p(x^k)$. We note that the latter property is independent of the value of the stepsize α_k which produced x^{k+1} , and in particular, independent of the value of β_k chosen for the penalty function ψ_{β_k} . The above observation indicates that the primal iterates are getting closer to the bounded feasible set D , which makes the iterates bounded. And this is just due to the structure of SQP constraints and regardless of any other details of the method. From this, convergence of SQP when applied to the example (6) comes as no surprise.

What has been observed for the example (6) is not accidental, of course. As is easily seen, the directional derivative of p in the SQP direction $d^k \in D_k$ is strictly negative at any infeasible point $x^k \notin D$. Indeed, for arbitrary $x \in \mathfrak{N}^n$ and $d \in \mathfrak{N}^n$, we have that

$$p'(x; d) = \sum_{i \in I_+(x)} \langle g'_i(x), d \rangle + \sum_{i \in I_0(x)} \max\{0, \langle g'_i(x), d \rangle\},$$

where $I_+(x) = \{i = 1, \dots, m \mid g_i(x) > 0\}$, $I_0(x) = \{i = 1, \dots, m \mid g_i(x) = 0\}$. Since for any $x^k \in \mathfrak{N}^n$ and $d^k \in D_k$ it holds that $\langle g'_i(x^k), d^k \rangle \leq -g_i(x^k) = 0$,

$\forall i \in I_0(x^k)$, we have that

$$p'(x^k; d^k) = \begin{cases} \sum_{i \in I_+(x^k)} \langle g'_i(x^k), d^k \rangle, & \text{if } x^k \notin D, \\ 0, & \text{if } x^k \in D \ (I_+(x^k) = \emptyset). \end{cases} \tag{8}$$

Furthermore, for $x^k \notin D$ and $d^k \in D_k$, we obtain that $\langle g'_i(x^k), d^k \rangle \leq -g_i(x^k) < 0$, $\forall i \in I_+(x^k)$, so that

$$p'(x^k; d^k) \leq - \sum_{i \in I_+(x^k)} g_i(x^k) = -p(x^k) < 0. \tag{9}$$

This indicates that when we make a reasonably small step in the direction d^k from an infeasible point, then the measure of infeasibility p is “likely” to decrease [because of (9)]. Of course, the latter is not quite automatic if we ignore in the course of linesearch the values of p and look at ψ_{β_k} only. But we can guarantee the needed property by combining the linesearch for ψ_{β_k} with the linesearch for p . Actually, the only essential modification we shall make to the standard SQP method consists in ensuring that

$$p(x^{k+1}) \leq p(x^k) \quad \text{when } p(x^k) > \bar{p},$$

where the infeasibility threshold $\bar{p} \geq 0$ is a user-chosen parameter. Descent is not even required. The idea is that this safeguard of not accepting an increase of constraints violation at points which are already infeasible beyond a chosen (could be large) threshold, should be enough to make $\{x^k\}$ bounded under reasonable assumptions about the problem.

We should mention that the idea of limiting infeasibility of SQP iterates is certainly not new. Related conditions with similar goals can be found, e.g., in [7], and possibly other publications as well, but in the context of significantly modified SQP schemes. Our goal here is to produce a complete global convergence result for the classical SQP method, perhaps very slightly modified.

Recall that assuming regularity of constraints (e.g., the Slater condition (7) in the case when g is convex), stationary points of (1) are characterized by the KKT conditions in variables $(x, \mu) \in \mathfrak{R}^n \times \mathfrak{R}^m$, which are

$$\left. \begin{aligned} f'(x) + \sum_{i=1}^m \mu_i g'_i(x) &= 0, \\ \mu_i \geq 0, \ g_i(x) \leq 0, \ \mu_i g_i(x) &= 0, \ i = 1, \dots, m. \end{aligned} \right\} \tag{10}$$

2 SQP algorithm and its global convergence

We shall now proceed to formally describe the algorithm details. The unique solution d^k of (2) and an associated Lagrange multiplier μ^k are characterized by the KKT

optimality conditions

$$f'(x^k) + H_k d^k + \sum_{i=1}^m \mu_i^k g'_i(x^k) = 0, \quad (11)$$

$$\mu_i^k \geq 0, \quad g_i(x^k) + \langle g'_i(x^k), d^k \rangle \leq 0, \quad (12)$$

$$\mu_i^k (g_i(x^k) + \langle g'_i(x^k), d^k \rangle) = 0, \quad i = 1, \dots, m. \quad (13)$$

Define

$$\Delta_k = \langle f'(x^k), d^k \rangle - \beta_k p(x^k) + \frac{1}{2} \langle H_k d^k, d^k \rangle, \quad (14)$$

which is an upper bound for the derivative of ψ_{β_k} at x^k in the direction d^k ; specifically:

$$\begin{aligned} \psi'_{\beta_k}(x^k; d^k) &= \langle f'(x^k), d^k \rangle + \beta_k \sum_{i \in I_+(x^k)} \langle g'_i(x^k), d^k \rangle \\ &\leq \Delta_k - \frac{1}{2} \langle H_k d^k, d^k \rangle, \end{aligned} \quad (15)$$

where the first equality follows from (8) and the inequality follows from (12) and (14).

Algorithm 1 below is the same as a standard SQP method, except for the update rule of the penalty parameter and the linesearch procedure. We shall discuss those differences after stating the method.

Algorithm 1 (SQP with an infeasibility safeguard)

Choose some $x^0 \in \mathfrak{N}^n$, $\bar{p} \geq 0$, $\delta > 0$, $\beta_{-1} = 0$, $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Set $k := 0$.

1. Choose an $n \times n$ positive definite matrix H_k and compute $(d^k, \mu^k) \in \mathfrak{N}^n \times \mathfrak{N}^m$ as the solution of (2) and an associated Lagrange multiplier.
Stop if $d^k = 0$. Otherwise,
2. Set

$$\gamma_k = \begin{cases} \max_{i \in I_+(x^k)} \mu_i^k, & \text{if } x^k \notin D, \\ 0, & \text{if } x^k \in D \quad (I_+(x^k) = \emptyset). \end{cases} \quad (16)$$

If $\beta_{k-1} \leq \gamma_k$ then set $\beta_k := \gamma_k + \delta$; otherwise set $\beta_k := \beta_{k-1}$.

3. Compute Δ_k by (14). Find j_k , the smallest nonnegative integer j , such that

$$\psi_{\beta_k}(x^k + \theta^j d^k) \leq \psi_{\beta_k}(x^k) + \sigma \theta^j \Delta_k, \quad \text{if } p(x^k) \leq \bar{p}, \quad (17)$$

$$\left. \begin{aligned} \psi_{\beta_k}(x^k + \theta^j d^k) &\leq \psi_{\beta_k}(x^k) + \sigma \theta^j \Delta_k \\ p(x^k + \theta^j d^k) &\leq p(x^k) \end{aligned} \right\} \quad \text{if } p(x^k) > \bar{p}. \quad (18)$$

4. Set $x^{k+1} = x^k + \alpha_k d^k$, where $\alpha_k = \theta^{j_k}$. Set $k := k + 1$ and got to Step 1.

We note that the rule presented in Step 2 of Algorithm 1 for the choice of the penalty parameter β_k is somewhat different from the common requirement based on (5) and $\|\mu^k\|_\infty$. We cannot be completely sure whether our rule is new, but we were not able to find it in the literature. The advantage of the rule based on ensuring that $\beta_k > \gamma_k$,

instead of $\beta_k > \|\mu^k\|_\infty$, is that it may increase β_{k-1} by a smaller value, when it increases it at all. Since $\gamma_k \leq \|\mu^k\|_\infty$, quite possibly strictly, if the increase is needed it may be smaller. For example, at feasible points $\gamma_k = 0$ and β_{k-1} is not increased (except on the first iteration, if it is feasible), while the standard rule may produce an increase. Since keeping the penalty parameter as small as possible is a well-known requirement for good numerical performance of SQP, the rule used in our algorithm appears to be preferable to the standard one.

The main difference that shall allow us to prove stronger-than-usual convergence results is the modification (18) of the standard linesearch procedure. Including the second inequality in (18) is a necessary safeguard for our analysis. But as a practical matter, choosing $\bar{p} > 0$ very large one may expect/hope that $p(x^k) > \bar{p}$ will not occur and the method would therefore work just as the usual SQP. Even if $p(x^k) > \bar{p}$ does occur, one can still expect that when the first inequality in (18) would be satisfied, very likely so would be the second. This is because d^k is a direction of “strong” descent for p at $x^k \notin D$ (the directional derivative is smaller than the value $-p(x^k)$, see (9), which is in turn smaller than $-\bar{p}$, which can be chosen as a very negative number), while (18) does not even require strict descent for p . Thus the interval of acceptable stepsizes for the second inequality in (18) should typically be quite large. One can therefore expect that (18) would often result in the same stepsize value as standard SQP. This is why, even though our convergence results are established for the (slightly) modified SQP method, we believe they are relevant for the usual SQP as well. For example, our results certainly explain convergence of SQP for problems like (6), where the second inequality in (18) always holds automatically, as discussed above.

We next show that Algorithm 1 is well-defined.

Proposition 1 *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be continuously differentiable in a neighbourhood of $x^k \in \mathbb{R}^n$. Assume that the subproblem (2) is feasible (In particular, this holds if g is convex and $D \neq \emptyset$).*

If $d^k = 0$ then Algorithm 1 terminates at a KKT point of (1).

If $d^k \neq 0$ then the linesearch procedure in Step 3 of Algorithm 1 terminates finitely with some integer j_k . In particular, the k -th iteration of Algorithm 1 is well defined.

Proof Since H_k is positive definite, whenever the subproblem (2) is feasible it has the unique solution d^k . If $d^k = 0$ then (11)–(13) show that the point (x^k, μ^k) satisfies the KKT conditions (10) for problem (1).

Suppose that $d^k \neq 0$. Using (14), we have that

$$\begin{aligned} \Delta_k &= -\langle H_k d^k, d^k \rangle - \sum_{i=1}^m \mu_i^k \langle g_i'(x^k), d^k \rangle - \beta_k p(x^k) + \frac{1}{2} \langle H_k d^k, d^k \rangle \\ &\leq -\frac{1}{2} \langle H_k d^k, d^k \rangle - \sum_{i \in I_+(x^k)} (\beta_k - \mu_i^k) g_i(x^k), \end{aligned}$$

where the first equality follows from (11), and the inequality follows from (13) and the fact that $\mu_i^k g_i(x^k) \leq 0$ for all $i \notin I_+(x^k)$.

Furthermore, by Step 2 of Algorithm 1, either $I_+(x^k) = \emptyset$ or $\beta_k \geq \mu_i^k$ for all $i \in I_+(x^k)$. Hence,

$$\Delta_k \leq -\frac{1}{2} \langle H_k d^k, d^k \rangle. \quad (19)$$

By (15), we further have that

$$\psi'_{\beta_k}(x^k; d^k) \leq \Delta_k - \frac{1}{2} \langle H_k d^k, d^k \rangle \leq \Delta_k < 0, \quad (20)$$

where the last inequality holds whenever $d^k \neq 0$, due to (19).

By standard argument (20) implies that there exists $a_k > 0$ such that (17) would hold once j is such that $\theta^j \leq a_k$. If $x^k \notin D$, by (9) there exists $b_k > 0$ such that the second inequality in (18) would hold once j is large enough to satisfy $\theta^j \leq b_k$. Consequently, in this case (18) holds for $\theta^j \leq \min\{a_k, b_k\}$. \square

We assume from now on that $d^k \neq 0$ for all k , so that an infinite sequence of iterates is generated. We next prove that the primal sequence is bounded.

Proposition 2 *Let $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ and $g : \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ be continuously differentiable on \mathfrak{R}^n .*

In the case of $\bar{p} = 0$, suppose that the subproblems (2) are feasible for all k and that the level sets of the function p are bounded (In particular, those assumptions hold when g is convex and D is nonempty and bounded).

In the case of $\bar{p} > 0$, suppose that g is convex and D is nonempty and bounded.

Suppose that in Algorithm 1 the matrices H_k are chosen uniformly bounded and uniformly positive definite, i.e.,

$$c_1 \|d\|^2 \leq \langle H_k d, d \rangle \leq c_2 \|d\|^2 \quad \forall d \in \mathfrak{R}^n \quad (21)$$

for all k , where $c_2 \geq c_1 > 0$.

Then any sequence $\{x^k\}$ generated by Algorithm 1 is bounded.

Proof Denote the level sets of p by $L(c) = \{x \in \mathfrak{R}^n \mid p(x) \leq c\}$, $c \in \mathfrak{R}$. Obviously, $D = L(0)$. Note that when g is convex, the subproblems (2) are feasible automatically. Also, in this case p is convex. If we assume that the level set $D = L(0)$ of p is nonempty and bounded, it follows that all the level sets of p are bounded (e.g., [3, Proposition 2.3.1]).

If there exists some iteration index k_1 such that $p(x^k) \leq \bar{p}$ for all $k \geq k_1$, then $\{x^k\}$ is bounded by the boundedness of $L(\bar{p})$.

If there exists some iteration index k_1 such that $p(x^k) > \bar{p}$ for all $k \geq k_1$, then (18) implies that $p(x^k) \leq p(x^{k_1})$ for all $k > k_1$. Hence, it holds that $x^k \in L(p(x^{k_1}))$ for all $k \geq k_1$. Since the level sets of p are bounded, we conclude that $\{x^k\}$ is bounded.

Suppose now that there is an infinite number of iterates such that $p(x^k) \leq \bar{p}$ and such that $p(x^k) > \bar{p}$. Then we can define an infinite subsequence $\{k_j\}$, consisting of all iteration indices such that

$$p(x^{k_j}) \leq \bar{p}, \quad p(x^{k_j+1}) > \bar{p}. \quad (22)$$

Since the subsequence $\{x^{k_j}\} \subset L(\bar{p})$ is bounded, so is $\{f'(x^{k_j})\}$. Take $M > 0$ such that $\|f'(x^{k_j})\| \leq M$ for all j . We next show that $\{d^{k_j}\}$ is bounded, separately for the cases of $\bar{p} = 0$ and $\bar{p} > 0$.

Consider first the case of $\bar{p} = 0$. The fact that $x^{k_j} \in L(0) = D$ means that $g(x^{k_j}) \leq 0$. Hence, $d = 0$ is feasible in subproblem (2). Using the optimality of d^{k_j} in this subproblem, we obtain that $\langle f'(x^{k_j}), d^{k_j} \rangle + \frac{1}{2} \langle H_{k_j} d^{k_j}, d^{k_j} \rangle < 0$. Hence, $\frac{c_1}{2} \|d^{k_j}\| \leq \|f'(x^{k_j})\| \leq M$, where the Cauchy-Schwarz inequality and (21) have been used. We conclude that $\{d^{k_j}\}$ is bounded.

Consider now $\bar{p} > 0$, in which case g is assumed to be convex. Let $\tilde{x} \in D$ be any fixed feasible point. By the convexity of g_i , for all $i = 1, \dots, m$, we have that $0 \geq g_i(\tilde{x}) \geq g_i(x^{k_j}) + \langle g'_i(x^{k_j}), \tilde{x} - x^{k_j} \rangle$, so that $(\tilde{x} - x^{k_j}) \in D_{k_j}$. Hence, by the optimality of d^{k_j} in subproblem (2), we obtain that

$$\begin{aligned} M \|\tilde{x} - x^{k_j}\| + \frac{1}{2} c_2 \|\tilde{x} - x^{k_j}\|^2 &\geq \langle f'(x^{k_j}), \tilde{x} - x^{k_j} \rangle + \frac{1}{2} \langle H_{k_j}(\tilde{x} - x^{k_j}), \tilde{x} - x^{k_j} \rangle \\ &\geq \langle f'(x^{k_j}), d^{k_j} \rangle + \frac{1}{2} \langle H_{k_j} d^{k_j}, d^{k_j} \rangle \\ &\geq \|d^{k_j}\| (c_1 \|d^{k_j}\| / 2 - M), \end{aligned} \tag{23}$$

where the Cauchy-Schwarz inequality has been used twice, in the first and last inequalities, together with (21). Taking into account the boundedness of $\{x^{k_j}\}$, this shows that $\{d^{k_j}\}$ must be bounded.

We next consider the cases of $\bar{p} = 0$ and $\bar{p} > 0$ again together. Since $x^{k_{j+1}} = x^{k_j} + \alpha_{k_j} d^{k_j}$, where $\alpha_{k_j} \in (0, 1]$ and $\{x^{k_j}\} \subset L(\bar{p})$ and $\{d^{k_j}\}$ are bounded, the subsequence $\{x^{k_{j+1}}\}$ is also bounded. Then by the continuity of p , $\{p(x^{k_{j+1}})\}$ is bounded. Take $C > 0$ such that $p(x^{k_{j+1}}) \leq C$ for all j .

In addition to $\{k_j\}$, define also an infinite subsequence $\{k_l\}$, consisting of all iteration indices such that

$$p(x^{k_l}) > \bar{p}, \quad p(x^{k_l+1}) \leq \bar{p}. \tag{24}$$

By the definitions of $\{k_j\}$ and of $\{k_l\}$, it holds that

$$\begin{aligned} C \geq p(x^{k_{j+1}}) &\geq p(x^{k_j+2}) \geq \dots \geq p(x^k) > \bar{p}, \quad k_j + 1 \leq k \leq k_l, \\ \bar{p} \geq p(x^k), \quad &k_l + 1 \leq k \leq k_{j+1}, \end{aligned} \tag{25}$$

where we have used the facts that p is nonincreasing on iterations with $x^k \notin L(\bar{p})$ [recall (18)] until the iterates enter $L(\bar{p})$, after which they stay in $L(\bar{p})$ until the k_{j+1} -st iterate leaves this set ($x^{k_{j+1}} \in L(\bar{p})$ and $x^{k_{j+1}+1} \notin L(\bar{p})$). This iterate $x^{k_{j+1}+1}$ again satisfies $p(x^{k_{j+1}+1}) \leq C$ and, by induction, the chain of inequalities (25) holds for any j . This means that $\{x^k\} \subset L(C)$, so that $\{x^k\}$ is bounded. \square

We next show that the penalty parameter is fixed from some iteration on. The key to the proof is boundedness of the primal sequence established in Proposition 2.

Proposition 3 *Let $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ and $g : \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ be continuously differentiable on \mathfrak{R}^n . Let g be convex, D bounded and the Slater condition (7) be satisfied.*

Then any sequence $\{\mu^k\}$ generated by Algorithm 1 is bounded, and it holds that $\beta_k = \beta_{k_0} = \beta$ for all $k \geq k_0$.

Proof By Proposition 2, under the stated assumptions $\{x^k\}$ is bounded. Let $M > 0$ be such that $\|f'(x^k)\| \leq M$ for all k . Similarly to the proof of Proposition 2, we can now deduce that the whole sequence $\{d^k\}$ is bounded, by writing relations analogous to (23), for $(\hat{x} - x^k) \in D_k$ and d^k .

We next show that the sequence $\{\mu^k\}$ is bounded. Suppose the opposite, i.e., that there exists $\{k_j\}$ such that $\|\mu^{k_j}\| \rightarrow \infty$ as $j \rightarrow \infty$. Taking into account the boundedness of $\{x^k\}$ and of $\{\|\mu^{k_j}\|^{-1}\mu^{k_j}\}$, and taking a further subsequence, if necessary, we can assume that $\{x^{k_j}\} \rightarrow \tilde{x}$ and $\{\|\mu^{k_j}\|^{-1}\mu^{k_j}\} \rightarrow \tilde{\mu} \neq 0$. Dividing (11) by $\|\mu^{k_j}\| > 0$, we have that

$$\|\mu^{k_j}\|^{-1}(f'(x^{k_j}) + H_{k_j}d^{k_j}) + \sum_{i=1}^m \|\mu^{k_j}\|^{-1}\mu^{k_j}g'_i(x^{k_j}) = 0.$$

Now taking into account the boundedness of $\{H_k\}$ and of $\{d^k\}$ when passing onto the limit as $j \rightarrow \infty$, we obtain that

$$0 = \sum_{i=1}^m \tilde{\mu}_i g'_i(\tilde{x}), \quad \tilde{\mu} \in \mathfrak{R}_+^m \setminus \{0\}. \tag{26}$$

Define $J = \{i = 1, \dots, m \mid \tilde{\mu}_i > 0\}$. Evidently, it must hold that $\mu_i^{k_j} > 0$ for all $i \in J$ and all j sufficiently large. By (13) we then have that

$$g_i(x^{k_j}) + \langle g'_i(x^{k_j}), d^{k_j} \rangle = 0 \quad \forall i \in J. \tag{27}$$

By the Slater condition (7), we have that $0 > g_i(\hat{x}) \geq g_i(x^{k_j}) + \langle g'_i(x^{k_j}), \hat{x} - x^{k_j} \rangle \forall i \in J$. Subtracting (27) from the latter relation, we conclude that $0 > g_i(\hat{x}) \geq \langle g'_i(x^{k_j}), \hat{x} - x^{k_j} - d^{k_j} \rangle \forall i \in J$. Taking possibly a further subsequence of $\{k_j\}$ so that $\{d^{k_j}\} \rightarrow \tilde{d}$, and passing onto the limit as $j \rightarrow \infty$, we obtain $0 > g_i(\hat{x}) \geq \langle g'_i(\tilde{x}), \hat{x} - \tilde{x} - \tilde{d} \rangle \forall i \in J$. But then, using that $\tilde{\mu}_i > 0$ for all $i \in J$, that $\tilde{\mu}_i = 0$ for all $i \in \{1, \dots, m\} \setminus J$, and (26), we obtain

$$0 > \sum_{i \in J} \tilde{\mu}_i g_i(\hat{x}) \geq \sum_{i \in J} \tilde{\mu}_i \langle g'_i(\tilde{x}), \hat{x} - \tilde{x} - \tilde{d} \rangle = \sum_{i=1}^m \langle \tilde{\mu}_i g'_i(\tilde{x}), \hat{x} - \tilde{x} - \tilde{d} \rangle = 0,$$

which is a contradiction.

We have therefore established that $\{\mu^k\}$ is bounded. This obviously implies that $\{\gamma_k\}$ given by (16) is also bounded. Consequently, $\{\beta_k\}$ is bounded. It then easily follows from the construction of Step 2 of Algorithm 1 that β_k is modified only a finite number of times, i.e., $\beta_k = \beta_{k_0} = \beta$ for all $k \geq k_0$. □

We next give the final convergence result, not assuming boundedness of any algorithmic sequences. Here, we have to choose \bar{p} strictly positive in Algorithm 1, which was not important in previous results.

Theorem 1 *Let $f : \mathfrak{N}^n \rightarrow \mathfrak{R}$ and $g : \mathfrak{N}^n \rightarrow \mathfrak{R}^m$ be continuously differentiable on \mathfrak{N}^n , with derivatives which are Lipschitz-continuous on bounded sets. Let g be convex, D bounded and the Slater condition (7) be satisfied.*

Then any sequence $\{x^k\}$ generated by Algorithm 1, where $\bar{p} > 0$, is bounded and each of its accumulation points is a KKT point of problem (1).

Proof Having established the preceding results, the remaining argument is close to standard, but we need to take into account the modified stepsize rule.

We first show that the sequence of stepsizes $\{\alpha_k\}$ is bounded away from zero. From now on, let $k \geq k_0$ (so that $\beta_k = \beta$, by Proposition 3). Since the sequences $\{x^k\}$ and $\{d^k\}$ are bounded (by Proposition 2 and by the proof of Proposition 3, respectively), all the functions involved are Lipschitz-continuous (with modulus $L > 0$) on the relevant bounded set. Let also $M > 0$ be such that $\|d^k\| \leq M$ for all k .

For all $\alpha \in [0, 1]$ and $i = 1, \dots, m$, we have that

$$\begin{aligned} & \max \{0, g_i(x^k + \alpha d^k)\} - \max \{0, g_i(x^k) + \alpha \langle g'_i(x^k), d^k \rangle\} \\ & \leq \max \{0, g_i(x^k + \alpha d^k) - g_i(x^k) - \alpha \langle g'_i(x^k), d^k \rangle\} \\ & \leq |g_i(x^k + \alpha d^k) - g_i(x^k) - \alpha \langle g'_i(x^k), d^k \rangle| \\ & \leq \frac{L}{2} \alpha^2 \|d^k\|^2, \end{aligned} \tag{28}$$

where the first inequality follows from the fact that $\max \{0, a - b\} \geq \max \{0, a\} - \max \{0, b\} \forall a, b \in \mathfrak{R}$; and the last inequality is by the Lipschitz-continuity of g'_i (e.g., [2, Proposition A.24]). We further have that

$$\begin{aligned} & \max \{0, g_i(x^k) + \alpha \langle g'_i(x^k), d^k \rangle\} \\ & = \max \{0, \alpha(g_i(x^k) + \langle g'_i(x^k), d^k \rangle) + (1 - \alpha)g_i(x^k)\} \\ & \leq \alpha \max \{0, g_i(x^k) + \langle g'_i(x^k), d^k \rangle\} + (1 - \alpha) \max \{0, g_i(x^k)\} \\ & = (1 - \alpha) \max \{0, g_i(x^k)\}, \end{aligned} \tag{29}$$

where the inequality is by the convexity of $\max\{0, \cdot\}$, and the last equality is by (12). Combining (28) and (29), and summing them up for $i = 1, \dots, m$, we conclude that

$$p(x^k + \alpha d^k) \leq (1 - \alpha)p(x^k) + \frac{Lm}{2} \alpha^2 \|d^k\|^2.$$

It follows that for iterations k such that $p(x^k) > \bar{p}$, the second relation in (18) is guaranteed to be satisfied for $\alpha = \theta^j$ such that $Lm\alpha\|d^k\|^2/2 \leq p(x^k)$, and in particular, for α such that $\alpha \leq 2\bar{p}/(LM^2m)$ (note that $\bar{p} > 0$ is important here). It is also well known that there exists $\hat{\alpha} > 0$ such that the inequality (17) is satisfied for $\alpha = \theta^j$ such

that $\alpha \leq \hat{\alpha}$. From here, by standard argument, we obtain that Step 3 of Algorithm 1 generates $\alpha_k \geq \bar{\alpha} > 0$ for all k .

Therefore, using also (19) and (21), for $k \geq k_0$, we obtain that

$$\psi_\beta(x^k) - \psi_\beta(x^{k+1}) \geq -\sigma \alpha_k \Delta_k \geq \sigma \bar{\alpha} c_1 \|d^k\|^2/2.$$

This implies that $0 = \lim_{k \rightarrow \infty} d^k$, and the assertion follows by passing onto the limit in (11)–(13) along appropriately chosen convergent subsequences. \square

References

1. Bertsekas, D.: *Constrained Optimization and Lagrange Multiplier Methods*. Academic, New York (1982)
2. Bertsekas, D.: *Nonlinear Programming*. Athena Scientific, Belmont (1995)
3. Bertsekas, D.: *Convex Analysis and Optimization*. Athena Scientific, Belmont (2003)
4. Bonnans, J., Gilbert, J., Lemaréchal, C., Sagastizábal, C.: *Numerical Optimization: Theoretical and Practical Aspects*. Springer, Berlin (2003)
5. Han, S.P.: Superlinearly convergent variable metric algorithms for general nonlinear programming problems. *Math. Program.* **11**, 263–282 (1976)
6. Han, S.P.: A globally convergent method for nonlinear programming. *J. Optim. Theory Appl.* **22**, 297–309 (1977)
7. Murray, W., Prieto, F.J.: A sequential quadratic programming algorithm using an incomplete solution of the subproblem. *SIAM J. Optim.* **5**, 590–640 (1995)
8. Palomares, U.G., Mangasarian, O.: Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems. *Math. Program.* **11**, 1–13 (1976)
9. Powell, M.: The convergence of variable metric methods for nonlinearly constrained optimization calculations. In: Mangasarian, O., Meyer, R., Robinson, S. (eds.) *Nonlinear Programming*, pp. 27–63. Academic, London (1978)
10. Robinson, S.: A quadratically convergent algorithm for general nonlinear programming problems. *Math. Program.* **3**, 145–156 (1972)
11. Robinson, S.: Perturbed Kuhn–Tucker points and rates of convergence for a class of nonlinear programming algorithms. *Math. Program.* **7**, 1–16 (1974)
12. Wilson, R.B.: A simplicial method for concave programming. Ph.D. thesis, Graduate School of Business Administration, Harvard University, Cambridge (1963)