Journal of Global Optimization **21:** 157–184, 2001.

A Globally and Superlinearly Convergent SQP Algorithm for Nonlinear Constrained Optimization ⋆

LIQUN QI[1] and YU-FEI YANG[2,†]
[1]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (e-mail: maqilq@inet.polyu.edu.hk);* [2]*College of Mathematics and Econometrics, Hunan University, Changsha 410082, China.*

(Received 25 April 2000; accepted in revised form 8 May 2001)

Abstract. Based on a continuously differentiable exact penalty function and a regularization technique for dealing with the inconsistency of subproblems in the SQP method, we present a new SQP algorithm for nonlinear constrained optimization problems. The proposed algorithm incorporates automatic adjustment rules for the choice of the parameters and makes use of an approximate directional derivative of the merit function to avoid the need to evaluate second order derivatives of the problem functions. Under mild assumptions the algorithm is proved to be globally convergent, and in particular the superlinear convergence rate is established without assuming that the strict complementarity condition at the solution holds. Numerical results reported show that the proposed algorithm is promising.

Key words: SQP method, constrained optimization, exact penalty function, global convergence, superlinear convergence

1. Introduction

We consider the following nonlinear inequality constrained optimization

$$(P) \qquad \begin{array}{ll} \min & f(x), \\ s.t. & g(x) \leqslant 0, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are assumed to be twice continuously differentiable. Such problems arise in a variety of applications in science, engineering and management.

Since the late 1970s, the sequential quadratic programming (SQP) method is considered among the most effective methods for solving nonlinear programming problems. Many techniques for solving problem (P) have been proposed, e.g., see

⋆ This work is supported by the Research Grant Council of Hong Kong and the Australian Research Council.

† This work was done while he was visiting School of Mathematics, the University of New South Wales, Sydney, Australia.

Boggs and Tolle [1] for a review of these techniques. The basic idea of a typical SQP method is as follows: Given an approximate solution, say x, problem (P) is modeled by the quadratic programming (QP) subproblem

$$QP(x; H) \qquad \begin{array}{ll} \min & \frac{1}{2}d^T Hd + \nabla f(x)^T d, \\ \text{s.t.} & g(x) + \nabla g(x)^T d \leqslant 0, \end{array} \tag{1}$$

where $H \in \mathbb{R}^{n\times n}$ is symmetric positive definite; then the solution d (referred to as the search direction) to subproblem (1) is used to generate a better approximation $\tilde{x}$:

$$\tilde{x} := x + \alpha d.$$

A suitable merit function, which usually depends on some positive parameter, is chosen to measure progress towards a solution to problem (P) and the steplength α is determined to yield a sufficient decrease in the merit function.

A serious limitation in the application of the SQP method to problem (P) is the possible inconsistency of the constraints in (1), i.e., the feasible set of subproblem (1) may be empty. To address this difficulty, various avenues have been proposed, e.g., see [2, 3, 12, 13, 18].

Recently, Spellucci [17] presented a new regularization technique for dealing with inconsistent QP subproblems in the SQP method. The modified QP subproblem in [17] is as follows:

$$\begin{array}{ll} \min & \frac{1}{2}d^T Hd + \nabla f(x)^T d + \tau e_A^T u_A + \frac{1}{2}\beta \|u_A\|^2, \\ \text{s.t.} & g_A(x) + \nabla g_A(x)^T d - u_A \leqslant 0, u_A \geqslant 0, \end{array} \tag{2}$$

where $A := A(x, \delta)$ is an index set to be introduced later and e is a vector whose components are all one. The method has the following merits:

- Subproblem (2) is always feasible with $u_i = \max\{g_i(x), 0\},\ i \in A$ and $d = 0$.
- Only constraints from A are considered.
- A simple automatic adjustment rule for the parameter τ is used.

By using l_1-exact penalty function as the merit function, the method was shown to be globally convergent. However, rate of convergence results can be obtained under stronger regularity conditions only. In particular, the strict complementarity condition for problem (P) at the solution has to be assumed so that a wide variety of practical applications may not be treated in this framework.

By using the differentiable exact penalty function developed by Lucidi [10] as the merit function, Facchinei [6] proposed a hybrid method for solving problem (P). Its basic idea is: if subproblem (1) is consistent and its solution is acceptable, then the solution is used as the search direction; otherwise, a first order direction, which is an approximation of the gradient of the exact penalty function introduced in [10], is used. The method was globally and superlinearly convergent. In particular, the superlinear convergence rate was obtained without requiring that the strict complementarity condition of problem (P) holds at the solution.

In this paper, based on the continuously differentiable exact merit function proposed by Lucidi [10], we present a new SQP algorithm for problem (P). The subproblem of the proposed algorithm is a slight modification of problem (2) and hence the new algorithm can be regarded as an application of Spellucci's approach to the merit function proposed by Lucidi [10]. Under mild assumptions we show that the algorithm is globally convergent. Without requiring the strict complementarity condition we prove that the convergence rate is superlinear. At the same time our approach preserves all the merits of Spellucci's method in [17]. Here we summarize some relevant features of the new algorithm as follows:

- The subproblem of the new algorithm is always feasible.
- Only constraints from a subset of $I := \{1, \dots, m\}$ are considered.
- Simple automatic adjustment rules for the parameters are used.
- Only first order derivatives of the problem functions are needed.
- Global and superlinear convergence are obtained.
- Strict complementarity condition of problem (P) at the solution is not assumed.

The paper is organized as follows. In the next section we discuss some basic assumptions and recall some important results on the merit function proposed in [10]. In Section 3 we give a detailed description of the subproblem and of the new algorithm, and prove that the proposed algorithm is well defined. Section 4 devotes to prove global convergence of the algorithm towards KKT points of problem (P). In Section 5 we deal with superlinear convergence of the algorithm. Some numerical results are reported in Section 6 and some conclusive remarks are given in the last section.

We give a list of notation employed. Throughout the paper, the symbol $\|\cdot\|$ will refer to the Euclidean vector norm or its associated matrix norm. For all $x \in \mathbb{R}^n$, we define the following index sets:

$$\begin{aligned} I_0(x) &:= \{i \in I \,:\, g_i(x) = 0\}, \\ P(x) &:= \{i \in I \,:\, g_i(x) > 0\}, \\ P_0(x) &:= P(x) \cup I_0(x), \\ A(x, \delta) &:= \{i \in I \,:\, g_i(x) \geqslant -\delta\}, \end{aligned} \tag{3}$$

where $\delta > 0$ is a parameter. We denote by $g^+(x)$ the vector with components $g_i^+(x) := \max[0, g_i(x)]$, $i \in I$.

Given $h : \mathbb{R}^n \to \mathbb{R}^m$ and a subset B of I, we denote by $h_B(x)$ the subvector of $h(x)$ with components $h_i(x)$, $i \in B$ and by $\nabla h_B(x)$ the transposed Jacobian of h_B at x. Specially, if the index set depends on parameters, we do not reflect this dependence because the given parameters will always be clear from the context. For example, we denote by $h_A(x)$ the subvector of $h(x)$ with components $h_i(x)$, $i \in A(x, \delta)$.

2. Preliminaries

In this section we discuss some basic assumptions and review some properties of the continuously differentiable exact penalty function developed in [10].

We denote by

$$\mathcal{F} := \{x \in \mathbb{R}^n \ : \ g(x) \leqslant 0\}$$

the feasible set of problem (P). The Lagrangian function associated with problem (P) is the function

$$L(x, \lambda) := f(x) + \lambda^T g(x).$$

A KKT pair for problem (P) is a pair $(\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^{n \times m}$ such that KKT conditions

$$\nabla L(\tilde{x}, \tilde{\lambda}) = 0, \quad g(\tilde{x}) \leqslant 0, \quad \tilde{\lambda} \geqslant 0, \quad G(\tilde{x})\tilde{\lambda} = 0 \tag{4}$$

hold, where

$$G(x) := \text{diag}(g_i(x))$$

and

$$\nabla L(x, \lambda) := \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x). \tag{5}$$

The point $\tilde{x}$ is said to be a KKT point of problem (P). If all conditions in (4) except for $\tilde{\lambda} \geqslant 0$ are satisfied, the point $\tilde{x}$ is said to be a stationary point of problem (P).

Letting $\alpha > 0$ be a given scalar, we consider an open perturbation of the feasible set $\mathcal{F}$ defined by

$$\mathcal{A} := \left\{ x \in \mathbb{R}^n \ : \ \sum_{i=1}^{m} g_i^+(x)^3 < \alpha \right\},$$

and we denote by $\bar{\mathcal{A}}$ the closure of $\mathcal{A}$ and by $\partial\mathcal{A}$ its boundary. Moreover, we introduce the function

$$a(x) := \alpha - \sum_{i=1}^{m} g_i^+(x)^3,$$

which takes positive values on $\mathcal{A}$ and is zero on $\partial\mathcal{A}$. Set $\psi(x) := \sum_{i=1}^{m} g_i^+(x)$.

In the sequel we make the following hypotheses:

ASSUMPTION A1. The set $\mathcal{A}$ is bounded.

ASSUMPTION A2. At every $x \in \mathcal{F}$, the vectors $\nabla g_i(x), \ i \in I_0(x)$ are linearly independent.

ASSUMPTION A3. At every $x \in \bar{\mathcal{A}} \backslash \mathcal{F}$, the Mangasarian-Fromowitz condition holds; i.e., there exists some $z \in \mathbb{R}^n$ such that $\nabla g_{P_0}(x)^T z < 0$.

Assumptions A1 and A2 are the same as those of [6]. Assumption A1 is often substituted by the assumption: the level sets of the objective function in unconstrained optimization are compact or the sequence of points produced by the algorithm is bounded; Assumption A2 is a common assumption in all algorithms dealing with global convergence of methods for solving problem (P). Assumption A3 involves the behaviour of the constraint functions outside the feasible set. Here Assumption A3 is slightly stronger than the related assumption in [6], this is due to the requirement of our subproblems. Moreover, under Assumptions A1 and A3, $\mathcal{F}$ is nonempty.

We now recall the class of multiplier functions proposed in [10]; this class is a generalization of the multiplier function proposed by Glad and Polak [7].

PROPOSITION 2.1. *The following statements hold.*

(i) *For every $x \in \mathcal{A}$, there exists a unique minimizer $\lambda(x)$ of the quadratic function in λ*

$$\|\nabla_x L(x, \lambda)\|^2 + \|G(x)\lambda\|^2 + \sum_{i=1}^{m} g_i^+(x)^3 \|\lambda\|^2$$

over $\mathbb{R}^m$, given by

$$\lambda(x) = -M^{-1}(x)\nabla g(x)^T \nabla f(x), \tag{6}$$

where $M(x)$ is the $m \times m$ matrix defined by

$$M(x) = \nabla g(x)^T \nabla g(x) + G^2(x) + \sum_{i=1}^{m} g_i^+(x)^3 I_m,$$

and I_m is the $m \times m$ identity matrix.

(ii) *$\lambda(x)$ is continuously differentiable in $\mathcal{A}$.*

(iii) *If $(\tilde{x}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m$ is a KKT pair for problem (P), we have $\lambda(\tilde{x}) = \tilde{\lambda}$.*

We then can define the following exact penalty function for problem (P):

$$Z(x; \epsilon) := f(x) + \lambda(x)^T c(x; \epsilon) + \frac{1}{\epsilon a(x)} \|c(x; \epsilon)\|^2, \tag{7}$$

where

$$\begin{aligned}
&c(x; \epsilon) := g(x) + Y(x; \epsilon) y(x; \epsilon),\\
&y_i(x; \epsilon) := \left\{ -\min\left[0, g_i(x) + \frac{\epsilon a(x)}{2}\lambda_i(x)\right]\right\}^{1/2}, \qquad i \in I,\\
&Y(x; \epsilon) := \mathrm{diag}(y_i(x; \epsilon)).
\end{aligned} \tag{8}$$

The penalty function $Z(x;\epsilon)$ was introduced by Lucidi [10] and the introduction of such a peculiar function was intended to weaken the conditions which ensures the equivalence between the solutions of problem (P) and the solutions of the unconstrained minimization problem of $Z(x;\epsilon)$ on $\mathcal{A}$. Later, by using the penalty function $Z(x;\epsilon)$ as the merit function, we will remove the strict complementarity condition in the analysis of the superlinear convergence of the proposed algorithm.

Note that the particular form of the barrier term $\frac{1}{\epsilon a(x)}$ causes the penalty function Z to go to infinity on $\partial\mathcal{A}$, so that its level sets (contained in $\mathcal{A}$) are compact. We now collect some properties of Z in the next proposition, more details can be found in [10].

PROPOSITION 2.2. *The following statements hold.*

(i) *For any $\epsilon > 0$, $Z(x;\epsilon)$ is continuously differentiable at all $x \in \mathcal{A}$, with gradient*

$$\nabla Z(x;\epsilon) = \nabla f(x) + \nabla g(x)\lambda(x) + \nabla\lambda(x)c(x;\epsilon) + \frac{1}{\epsilon a(x)}\sum_{i=1}^{m} r_i(x;\epsilon)\nabla g_i(x), \tag{9}$$

where for $i \in I$,

$$r_i(x;\epsilon) := 2c_i(x;\epsilon) + 3[\|c(x;\epsilon)\|^2/a(x)]g_i^+(x)^2. \tag{10}$$

(ii) *For any $\epsilon > 0$, $Z(x;\epsilon) \leqslant f(x)$ for all $x \in \mathcal{F}$.*

(iii) *For every $\epsilon > 0$, let x be such that $c(x;\epsilon) \leqslant 0$. Then $x \in \mathcal{F}$.*

(iv) *If f and g are three times continuously differentiable, then for any $\epsilon > 0$, $Z(x;\epsilon)$ is twice continuously differentiable for all $x \in \mathcal{A}$ except at the points where $g_i(x) + \epsilon a(x)\lambda_i(x)/2 = 0$ for some i.*

(v) *Let $(\tilde{x}, \tilde{\lambda})$ be a KKT pair for problem (P). Then for every $\epsilon > 0$, we have $\nabla Z(\tilde{x};\epsilon) = 0$, $c(\tilde{x};\epsilon) = 0$ and $Z(\tilde{x};\epsilon) = f(\tilde{x})$.*

(vi) *Let $\tilde{x} \in \mathcal{A}$ be a stationary point of $Z(x;\epsilon)$ and assume that $c(\tilde{x};\epsilon) = 0$. Then $(\tilde{x}, \lambda(\tilde{x}))$ is a KKT pair for problem (P).*

3. Algorithm

In this section we propose a new SQP algorithm for solving problem (P) and prove that the proposed algorithm is well defined. At each iteration, we first solve a subproblem of the following form

$$QP_1(x, H; A, \tau, \beta) \qquad \begin{array}{ll} \min & \frac{1}{2}d^T Hd + \nabla f(x)^T d + \tau e_P^T u_P + \frac{1}{2}\beta\|u_P\|^2, \\ \text{s.t.} & g_P(x) + \nabla g_P(x)^T d - u_P \leqslant 0, u_P \geqslant 0, \\ & g_{A\setminus P}(x) + \nabla g_{A\setminus P}(x)^T d \leqslant 0, \end{array} \tag{11}$$

where the sets $A := A(x, \delta)$ and $P := P(x)$ are defined in Section 1 and $H \in \mathbb{R}^{n\times n}$ is symmetric positive definite. The KKT conditions of (11) are as follows:

$$\begin{aligned}
&Hd + \nabla f(x) + \nabla g_{\mathcal{A}}(x)\rho_{\mathcal{A}} = 0,\\
&\beta u_i + \tau - \rho_i - \nu_i = 0, \quad \forall i \in P,\\
&\nu_P \geqslant 0, \quad u_P \geqslant 0, \quad \nu_P^T u_P = 0,\\
&\rho_P \geqslant 0, \quad g_P(x) + \nabla g_P(x)^T d - u_P \leqslant 0,\\
&\rho_P^T[g_P(x) + \nabla g_P(x)^T d - u_P] = 0,\\
&\rho_{A\backslash P} \geqslant 0, \quad g_{A\backslash P} + \nabla g_{A\backslash P}(x)^T d \leqslant 0,\\
&\rho_{A\backslash P}^T[g_{A\backslash P}(x) + \nabla g_{A\backslash P}(x)^T d] = 0.
\end{aligned} \tag{12}$$

Subproblem (11) is different from (2) in that: in (11) a slack variable u_i is used only for every constraint satisfying $g_i(x) > 0$ so that slack variables will disappear when the iterate is close to the feasible region. Since we cannot guarantee that a solution d to (11) is a descent direction of the merit function Z, we have to solve another subproblem of the form

$$\begin{aligned}
QP_2(x, H; A, \tau, \beta) \qquad &\min\ \frac{1}{2}d^T Hd + \nabla f(x)^T d + \sum_{i\in P} \tau v_i(x) u_i + \tfrac{1}{2}\beta\|u_P\|^2,\\
&\text{s.t.}\ \ g_P(x) + \nabla g_P(x)^T d - u_P \leqslant 0, u_P \geqslant 0,\\
&\qquad\ g_{A\backslash P}(x) + \nabla g_{A\backslash P}(x)^T d \leqslant 0,
\end{aligned} \tag{13}$$

where $v_i(x) := g_i(x) + \frac{3}{2}\|g^+(x)\|^2 g_i(x)^2/a(x),\ i \in P(x)$. It is obvious that (11) is a special case of (13) with $v_P(x) = e_P$.

Similar to the analysis performed in [17], we can deduce the following properties of problem $QP_1(x, H; A, \tau, \beta)$, whose proofs are given in Appendix A.

PROPOSITION 3.1. *If* (1) *is consistent and has a solution d with $\|d\|$ small enough, then $(d, u_P = 0)$ solves* (11) *for any $\beta > 0$, provided τ large enough. Conversely, if $(d, 0)$ is a solution of* (11) *and $\|d\|$ is small enough, then d is a solution of* (1).

PROPOSITION 3.2. *Under Assumptions A1-A3, there exists some $\bar\eta > 0$ such that for any $x \in \Omega(\bar\eta) := \{x \in \mathbb{R}^n : \psi(x) \leqslant \bar\eta\}$,* (1) *is consistent. Furthermore, if $\|H\|$ and $\|H^{-1}\|$ are bounded from above, then there exists some $\bar\tau > 0$ such that for any $x \in \Omega(\bar\eta)$, $\beta > 0$ and $\tau \geqslant \bar\tau$, $(d, u_P = 0)$ solves* (11)*; i.e.,* (11) *is equivalent to the following problem:*

$$\begin{aligned}
QP_0(x, H; A) \qquad &\min\ \frac{1}{2}d^T Hd + \nabla f(x)^T d,\\
&s.t.\ \ g_A(x) + \nabla g_A(x)^T d \leqslant 0.
\end{aligned} \tag{14}$$

PROPOSITION 3.3. *Under Assumptions A1-A3, there exists some $\tau^* = \tau^*(\|H\|, \|H^{-1}\|, \beta) > 0$, such that for any $x \in \mathcal{A}$, $\delta > 0$ and $\tau \geqslant \tau^*$ the solution (d, u_P)*

of problem $QP_1(x, H;\ A(x,\delta),\ \tau, \beta)$ *satisfies*

$$\sum_{i \in P(x)} u_i \leqslant L\psi(x)$$

with some $0 < L < 1$ *independent of* x *and* δ. τ^* *is bounded from above if* $\|H\|, \|H^{-1}\|, \beta$ *are.*

COROLLARY 3.1. *Under Assumptions A1–A3, for every* $\eta > 0$*, there exists some* $\tau^* = \tau^*(\|H\|, \|H^{-1}\|, \beta, \eta) > 0$ *such that for any* $x \in \{x \in \mathcal{A} : \psi(x) \geqslant \eta\}$, $\delta > 0$ *and* $\tau \geqslant \tau^*$ *the solution* (d, u_P) *of problem* $QP_2(x, H;\ A(x,\delta), \tau, \beta)$ *satisfies*

$$\sum_{i \in P(x)} v_i(x) u_i \leqslant L \sum_{i \in P(x)} v_i(x) g_i(x)$$

with some $0 < L < 1$ *independent of* x *and* δ*. Moreover, if* $\|H\|, \|H^{-1}\|, \beta$ *are bounded from above and* η *is bounded away from zero, then* τ^* *is bounded from above.*

The next proposition establishes the connection between the solutions of the original problem (P) and of the subproblem $QP_1(x, H;\ A, \tau, \beta)$.

PROPOSITION 3.4. *Assume that* $\tilde{A} \supseteq P_0(\tilde{x})$, $\tilde{\tau} > 0$ *and* $\tilde{\beta} > 0$ *are given. Then*

(i) If $(\tilde{x}, \tilde{\lambda})$ *is a KKT pair for problem* (P)*, then* $P(\tilde{x}) = \emptyset$ *and for any positive definite matrix* $H \in \mathbb{R}^{n\times n}$, $(0, \tilde{\lambda}_{\tilde{A}})$ *is a KKT pair for problem* $QP_1(\tilde{x}, H;\ \tilde{A}, \tilde{\tau}, \tilde{\beta})$.

(ii) If $(d, u_P) = (0, 0)$ *is a KKT point of problem* $QP_1(\tilde{x}, H;\ \tilde{A}, \tilde{\tau}, \tilde{\beta})$*, where* $H \in \mathbb{R}^{n\times n}$ *is positive definite, then* $\tilde{x}$ *is a KKT point of problem* (P).

Proof. (i) The assertion follows directly from the KKT conditions of problem (P) and of problem $QP_1(\tilde{x}, H;\ \tilde{A}, \tilde{\tau}, \tilde{\beta})$.

(ii) Since $(d, u_P) = (0, 0)$ is a KKT point of problem $QP_1(\tilde{x}, H;\ \tilde{A}, \tilde{\tau}, \tilde{\beta})$, it follows from (12) that

$$\nabla f(\tilde{x}) + \nabla g_{\tilde{A}}(\tilde{x})\rho_{\tilde{A}} = 0, \quad g_{\tilde{A}}(\tilde{x}) \leqslant 0, \quad \rho_{\tilde{A}}(\tilde{x}) \geqslant 0, \quad \text{and} \quad g_{\tilde{A}}(\tilde{x})^T \rho_{\tilde{A}}(\tilde{x}) = 0,$$

which implies that $(\tilde{x}, \rho)$ with $\rho_{I\setminus\tilde{A}} := 0$ satisfies KKT conditions (4). Hence $\tilde{x}$ is a KKT point of problem (P). □

By Proposition 3.4 (ii), we deduce that: if $\tilde{x} \in \mathcal{F}$ and $d = 0$ is a KKT point of problem $QP_0(\tilde{x}, H;\ \tilde{A})$, then $\tilde{x}$ is a KKT point of problem (P). Moreover, Proposition 3.4 is still valid if $QP_1(\tilde{x}, H;\ \tilde{A}, \tilde{\tau}, \tilde{\beta})$ is replaced by $QP_2(\tilde{x}, H;\ \tilde{A}, \tilde{\tau}, \tilde{\beta})$.

In this paper we make use of an approximate directional derivative which requires only the evaluation of first order derivatives of the problem functions, as was already considered in [6]. By approximating the term $w^T\nabla\lambda(x)$ (which contains the second-order derivatives of the problem functions) by finite differences, we approximate the directional derivative of $\nabla Z(x;\epsilon)$ along the direction w by

$$DZ(x,w;\epsilon,t) := \nabla L(x,\lambda(x))^T w + \frac{1}{\epsilon a(x)}\sum_{i=1}^{m} r_i(x;\epsilon)\nabla g_i(x)^T w + \frac{1}{t}[\lambda(x+tw)-\lambda(x)]^T c(x;\epsilon). \tag{15}$$

It is obvious that we have

$$DZ(x,w;\epsilon,t) = \nabla Z(x;\epsilon)^T w + w^T(\nabla\lambda(x+\xi w)-\nabla\lambda(x))c(x;\epsilon)$$

for some $\xi \in (0,t)$, and

$$\lim_{\substack{y\to x\\ v\to w\\ t\downarrow 0}} DZ(y,v;\epsilon,t) = \nabla Z(x;\epsilon)^T w. \tag{16}$$

We now state an algorithm for solving problem (P) and then give some further comments.

ALGORITHM 3.1.

Data: $x^1 \in \mathcal{A}$, $H^1 \in \mathbb{R}^{n\times n}$, $\theta \in (0,1)$, $\mu \in (0,\frac{1}{2})$, $\sigma \in (0,1)$, $\sigma_1 > 1$, $\omega \in (0,1)$, $\epsilon^1 > 0$, $\tau^1 > 0$, $\gamma^1 > 0$ *and* $\beta^1 > 0$.

Set $k := 1$.

Step 0. *Set* $\delta^k := \frac{\epsilon^k a(x^k)}{2}\max\left\{|\lambda_i(x^k)| \ : \ g_i(x^k)+\frac{\epsilon^k a(x^k)}{2}\lambda_i(x^k) \geqslant 0, i \in I\right\}$. *Denote*

$$P^k := P(x^k) \quad \textit{and} \quad A^k := A(x^k,\delta^k).$$

Step 1. *Compute* $(d^k, u^k_{P^k})$, *the solution of problem* $QP_1(x^k,H^k;A^k,\tau^k,\beta^k)$. *If* $(d^k,u^k_{P^k}) = (0,0)$, *stop.*

Step 2. *If* $\sum_{i\in P^k} u_i^k \leqslant (1-\omega/\tau^k)\psi(x^k)$, *go to Step 4.*

Step 3. *Set* $\tau^k := \sigma_1\tau^k$, *go to Step 1.*

Step 4. *If* $\psi(x^k) \leqslant \gamma^k$ *and* $\sum_{i\in P^k} u_i^k \neq 0$, *then set* $\gamma^k := \gamma^k/\sigma_1$ *and* $\tau^k := \sigma_1\tau^k$.

Step 5. *If*

$$\sum_{i\in P^k} v_i(x^k)u_i^k \leqslant (1-\omega/\tau^k)\sum_{i\in P^k} v_i(x^k)g_i(x^k), \tag{17}$$

go to Step 8.

Step 6. *Compute* $(d^k, u^k_{P^k})$, *the solution of problem* $QP_2(x^k,H^k;A^k,\tau^k,\beta^k)$. *If* (17) *is satisfied, go to Step 8.*

Step 7. *Set* $\tau^k := \sigma_1\tau^k$, *go to Step 6.*

Step 8. *Set* $\alpha^k := 1$.

Step 9. *If*

$$DZ(x^k, d^k; \epsilon^k, \alpha^k) \leqslant -\frac{1}{2}\max[(d^k)^T H^k d^k, \|c(x^k; \epsilon^k)\|^2], \tag{18}$$

go to Step 11.

Step 10. *Set* $\epsilon^k := \sigma\epsilon^k$. *If* $Z(x^k; \epsilon^k) > Z(x^1; \epsilon^k)$, *go to Step 13; otherwise, go to Step 9.*

Step 11. *If* $x^k + \alpha^k d^k \in \mathcal{A}$ *and*

$$Z(x^k + \alpha^k d^k; \epsilon^k) \leqslant Z(x^k; \epsilon^k) + \mu\alpha^k DZ(x^k, d^k; \epsilon^k, \alpha^k), \tag{19}$$

go to Step 12; otherwise, set $\alpha^k := \theta\alpha^k$, *and go to Step 9.*

Step 12. *Set* $\tau^{k+1} := \tau^k$, $\gamma^{k+1} := \gamma^k$, $\epsilon^{k+1} := \epsilon^k$, $x^{k+1} := x^k + \alpha^k d^k$; *generate* H^{k+1} *and* β^{k+1}. *Set* $k := k+1$, *and go to Step 0.*

Step 13. *Set* $x^k := x^1$; *generate new* H^k *and* β^k, *go to Step 0.*

In the above algorithm we use automatic adjustment rules for parameters τ, γ and ϵ. At early steps, these parameters possibly change many times at one iteration, but we do not reflect the change in our algorithm. We will prove that after a finite number of steps these parameters will stay fixed.

In Step 0 to Step 7, we compute the search direction d^k. To this end, we first solve subproblem (11). If (17) is not satisfied, we then solve another subproblem (13). We will prove later that d^k, whether from (11) or (13), satisfies the descent condition (18) after at most finite times of the reduction of ϵ as long as (17) holds. We will also prove that eventually we only require solving subproblem (14) at each iteration, so that (11) or (13) is solved only in early stages of the algorithm, when the iterate is far from the feasible region.

Step 9 to Step 11 are similar to the associated steps of the algorithm in [6]. At Step 9, we decide whether ϵ has to be reduced. If (18) is not satisfied, we reduce ϵ (Step 10); otherwise, we perform an Armijo-like test (Step 11). By using the Armijo-like test, we can force the approximation directional derivative DZ to 0, so that (18) will ensure that d^k and $c(x^k; \epsilon^k)$ tend to 0. By Proposition 2.2 (iii), $c(x^k; \epsilon) \to 0$ for some fixed $\epsilon > 0$ shows that we are converging to a feasible point and hence, by Proposition 3.4 (ii), $d^k \to 0$ implies that such point is a KKT point of problem (P).

To prove the well-definedness and convergence of the above algorithm, we assume that the next hypothesis holds.

ASSUMPTION A4. $H^k \in \mathbb{R}^{n\times n}$ is symmetric positive definite and there exists some constant $C > 0$ such that for all k,

$$\|H^k\| \leqslant C, \quad \|(H^k)^{-1}\| \leqslant C \quad \text{and} \quad \frac{1}{C} \leqslant \beta^k \leqslant C.$$

By quasi-Newton updates of an augmented Lagrangian and $\beta^k := trace(H^k)$, the above assumption holds, e.g., see [4] for details. From now on, we assume that Assumptions A1–A4 hold.

PROPOSITION 3.5. *Algorithm 3.1 cannot cycle infinitely between Step 1 and Step 3, Step 6 and Step 7, respectively.*

Proof. The assertion follows directly from Proposition 3.3 and Corollary 3.1. □

LEMMA 3.1. $d^k \neq 0$ *whenever we come to Step 9.*

Proof. We only need to consider the case $\psi(x^k) > 0$. Suppose by contradiction that $d^k = 0$. Then $u^k_{P^k} \geqslant g_{P^k}(x^k)$ and hence it follows from (17) that

$$(1 - \omega/\tau^k) \sum_{i \in P^k} v_i(x^k) g_i(x^k) \geqslant \sum_{i \in P^k} v_i(x^k) u_i^k \geqslant \sum_{i \in P^k} v_i(x^k) g_i(x^k) > 0,$$

which contradicts $1 - \omega/\tau^k < 1$. □

PROPOSITION 3.6. *Algorithm 3.1 cannot cycle infinitely between Step 9 and Step 10.*

The proof of the above proposition is given in Appendix A. The proof of the next proposition can be omitted, since it is similar to that of Proposition 3.5 in [6].

PROPOSITION 3.7. *Algorithm 3.1 cannot cycle infinitely between Step 9 and Step 11.*

The above analysis shows that Algorithm 3.1 is well-defined.

4. Global Convergence

This section devotes to prove global convergence of Algorithm 3.1. First, we show that the parameters τ and ϵ are updated only a finite number of times and that eventually the search direction d^k is obtained by solving subproblem $QP_0(x^k, H^k; A^k)$. Then we prove $d^k \to 0$ and $c(x^k, \epsilon^k) \to 0$ so that every limit point of the sequence $\{x^k\}$ generated by Algorithm 3.1 is a KKT point of problem (P). From Proposition 3.4 (ii), we deduce immediately the following proposition.

PROPOSITION 4.1. *If Algorithm 3.1 terminates at Step 1, i.e.,* $(d^k, u^k_{P^k}) = (0, 0)$, *then* x^k *is a KKT point of problem* (P).

Without loss of generality, we assume that Algorithm 3.1 generates an infinite iterative sequence $\{x^k\}$. The next lemma was proved in [6].

LEMMA 4.1. *Let* $\{x^k\}$ *and* $\{\epsilon^k\}$ *be two sequences such that* $x^k \in \mathcal{A}$, $Z(x^k; \epsilon^k) \leqslant Z(x^1; \epsilon^k)$ *for any k and* $\{\epsilon^k\} \downarrow 0$. *Then, every limit point of the sequence* $\{x^k\}$ *belongs to* $\mathcal{A}$.

PROPOSITION 4.2. *The parameters τ and γ can be changed for only a finite number of times; i.e., there exists an iteration index $\tilde{k}$ such that*

$$\tau^k = \tilde{\tau} \quad \text{and} \quad \gamma^k = \tilde{\gamma}, \quad \text{for all} \quad k \geqslant \tilde{k}.$$

Proof. By the construction of Algorithm 3.1, the parameter τ can be increased only at Step 3, Step 4 and Step 7, while the parameter γ can be reduced only at Step 4. The assertion then follows from Proposition 3.2, Proposition 3.3 and Corollary 3.1. □

The above proposition, altogether with Proposition 3.2 and Step 4 of the algorithm, shows that: if $\psi(x^k) \leqslant \tilde{\gamma}$, then $\sum_{i\in P^k} u_i^k = 0$. Hence the search direction d^k is actually a KKT point of problem $QP_0(x^k, H^k; A^k)$.

Similar to the proof of Theorem 5.3 in [17], we can deduce the following results.

PROPOSITION 4.3. *The solution $(d^k, u^k_{P^k})$ and the related multiplier ρ^k of problem $QP_1(x^k, H^k; A^k, \tau^k, \beta^k)$ are uniformly bounded with respect to $(\nabla f(x^k), \tau^k)$. Moreover, for all $x^k \in \{x \in \mathcal{A} : \psi(x) \geqslant \tilde{\gamma}\}$, the solution $(d^k, u^k_{P^k})$ and the related multiplier ρ^k of problem $QP_2(x^k, H^k; A^k, \tau^k, \beta^k)$ are also uniformly bounded with respect to $(\nabla f(x^k), \tau^k)$.*

The proof of the next proposition is put in Appendix A.

PROPOSITION 4.4. *The penalty parameter ϵ can be reduced only a finite number of times at Step 10.*

Without loss of generality, we assume $\epsilon^k = \tilde{\epsilon}$ for all k by the above proposition.

PROPOSITION 4.5. *After a finite number of steps we will always obtain the direction d^k from Step 1, i.e., Step 6 and Step 7 are never used.*

Proof. The proof is by contradiction. If the proposition does not hold, there exists an infinite subsequence $\{x^r\}$ of $\{x^k\}$ such that, at x^r, we get d^r from Step 6 and $x^r \to \tilde{x} \in \bar{\mathcal{A}}$. Noting that $\{Z(x^k; \tilde{\epsilon})\}$ is monotonically decreasing and that the level sets of $Z(x; \tilde{\epsilon})$ are compact, we have $\tilde{x} \notin \partial\mathcal{A}$ and $\lim_{k\to\infty} Z(x^k; \tilde{\epsilon}) = Z(\tilde{x}; \tilde{\epsilon})$. By (18) and (19), we have

$$\alpha^r DZ(x^r, d^r; \tilde{\epsilon}, \alpha^r) \to 0.$$

We consider the two possible cases below:

Case 1. $DZ(x^r, d^r; \tilde{\epsilon}, \alpha^r) \to 0$.

By (18), we have

$$c(x^r; \tilde{\epsilon}) \to 0. \tag{20}$$

This implies $\tilde{x} \in \mathcal{F}$ and hence for r sufficiently large, $\sum_{i\in P^r} u_i^r = 0$, i.e., d^r is obtained from Step 1. So, we obtain a contradiction.

Case 2. $\alpha^r \to 0$.

Without loss of generality, we assume that $\{d^r\}$ converges to $\tilde{d}$. From the construction of Algorithm 3.1, we deduce that for r sufficiently large, $\tilde{\alpha}^r := \alpha^r/\theta$ does not satisfy (19), i.e.,

$$[Z(x^r + \tilde{\alpha}^r d^r; \tilde{\epsilon}) - Z(x^r; \tilde{\epsilon})]/\tilde{\alpha}^r > \mu DZ(x^r, d^r; \tilde{\epsilon}, \tilde{\alpha}^r). \tag{21}$$

Hence, it follows from the continuity assumptions and (16) that

$$(1-\mu)\nabla Z(\tilde{x}; \tilde{\epsilon})^T \tilde{d} \geqslant 0. \tag{22}$$

Moreover, by (16) and (18), we also have

$$\nabla Z(\tilde{x}; \tilde{\epsilon})^T \tilde{d} \leqslant 0. \tag{23}$$

So, we obtain

$$\nabla Z(\tilde{x}; \tilde{\epsilon})^T \tilde{d} = 0,$$

which shows $DZ(x^r, d^r; \tilde{\epsilon}, \alpha^r) \to 0$. But then we can argue as in Case 1, and hence complete the proof. □

We have the following convergence results.

THEOREM 4.1. *(i)*

$$\lim_{k\to\infty} \|d^k\| = 0 \quad \textit{and} \quad \lim_{k\to\infty} \|c(x^k; \tilde{\epsilon})\| = 0. \tag{24}$$

(ii) Each accumulation point of the sequence $\{x^k\}$ generated by Algorithm 3.1 is a KKT point of problem (P).

Proof. (i) From the fact that $\{Z(x^k; \tilde{\epsilon})\}$ is monotonically decreasing and bounded from below, it follows that $\{Z(x^k; \tilde{\epsilon})\}$ is convergent and hence, by (18) and (19), we have

$$\lim_{k\to\infty} \alpha^k DZ(x^k, d^k; \tilde{\epsilon}, \alpha^k) = 0. \tag{25}$$

Suppose that (24) is false. Then there exist a constant $C_0 > 0$ and a subsequence $\{x^r\}$ of $\{x^k\}$, such that for all r,

$$\|d^r\| > C_0 \tag{26}$$

or

$$\|c(x^r; \tilde{\epsilon})\| > C_0 \quad \text{and} \quad \lim_{k\to\infty} \|d^k\| = 0. \tag{27}$$

By (18), (25)–(27) and Assumption A4, we deduce

$$\lim_{r\to\infty} \alpha^r = 0. \tag{28}$$

Moreover, without loss of generality, we can assume by the compactness of the level sets of $Z(x;\tilde{\epsilon})$ and Proposition 4.3 that $\lim_{r\to\infty} x^r = \tilde{x} \in \mathcal{A}$ and $\lim_{r\to\infty} d^r = \tilde{d}$.

Consider first the case $\|d^r\| > C_0$ for all r. In this case, $\tilde{d} \neq 0$. It follows from (28), (16) and (18) that

$$\lim_{r\to\infty} DZ(x^r, d^r; \tilde{\epsilon}, \alpha^r) = \nabla Z(\tilde{x}; \tilde{\epsilon})^T \tilde{d} < 0. \tag{29}$$

Let $\tilde{\alpha}^r := \alpha^r/\theta$. By the construction of Algorithm 3.1, we have, for r sufficiently large,

$$[Z(x^r + \tilde{\alpha}^r d^r; \tilde{\epsilon}) - Z(x^r; \tilde{\epsilon})]/\tilde{\alpha}^r > \mu DZ(x^r, d^r; \tilde{\epsilon}, \tilde{\alpha}^r). \tag{30}$$

Passing to the limit in (30) and taking into account (16) and $\lim_{r\to\infty} \tilde{\alpha}^r = 0$, we have

$$(1-\mu)\nabla Z(\tilde{x}; \tilde{\epsilon})^T \tilde{d} \geqslant 0.$$

This contradicts (29) and $0 < \mu < 1$.

Then consider the case $\|c(x^r; \tilde{\epsilon})\| > C_0$ for all r and $\lim_{k\to\infty} \|d^k\| = 0$. In this case, $\tilde{d} = 0$. It follows from (29) that

$$\lim_{r\to\infty} DZ(x^r, d^r; \tilde{\epsilon}, \alpha^r) = 0.$$

This contradicts to (18) and $\|c(x^r; \tilde{\epsilon})\| > C_0$ for all r. Therefore, (24) is proved.

(ii) Without loss of generality, assume that $\lim_{k\to\infty} x^k = \tilde{x} \in \mathcal{A}$ and $\lim_{k\to\infty} H^k = \tilde{H}$. By continuity and the definition of $c(x;\epsilon)$, it follows from (24) that

$$c(\tilde{x}; \tilde{\epsilon}) = 0.$$

This implies $\tilde{x} \in \mathcal{F}$ and hence for k sufficiently large, d^k is a KKT point of problem $QP_0(x^k, H^k; A^k)$.

Moreover, there exists an infinite subsequence $\{A^r\}$ of $\{A^k\}$, such that for all r,

$$A^r = \hat{A}$$

where $\hat{A}$ is some subset of I. Therefore, $\lim_{k\to\infty} \|d^k\| = 0$ shows that $d = 0$ is a KKT point of problem $QP_0(\tilde{x}, \tilde{H}; \hat{A})$. The assertion then follows from Proposition 3.4 (ii). □

5. Superlinear Convergence

In this section we study the rate of convergence result of Algorithm 3.1. We will deduce that under mild assumptions the whole sequence $\{x^k\}$ generated by the algorithm converges and that if an unit steplength ensures superlinear convergence, then eventually the unit steplength is obtained so that the Maratos effect does not occur. To this end, we need to strengthen the regularity assumptions on the functions involved. Let x^* be an accumulation point of the sequence $\{x^k\}$ generated by the algorithm and λ^* be the related multiplier. Then it follows from Theorem 4.1 (ii) that (x^*, λ^*) is a KKT pair for problem (P). Moreover, a function $h : \mathbb{R}^n \to \mathbb{R}$ is said to be an SC^1-function on an open set $\mathcal{A}$ if h is continuously differentiable on $\mathcal{A}$ and ∇h is semismooth (see [15]) on $\mathcal{A}$.

ASSUMPTION A5. *Every component of the gradient of $f(x)$ and of the Jacobian of $g(x)$ is an SC^1-function on $\mathcal{A}$.*

ASSUMPTION A6. *Strong second-order sufficient condition holds at (x^*, λ^*), i.e., the Hessian $\nabla^2 L(x^*, \lambda^*)$ is positive definite on the space $\{u : \nabla g_i(x^*)^T u = 0,\ \forall i \in I_+(x^*)\}$, where $I_+(x^*) := \{i \in I_0(x^*) : \lambda_i^* > 0\}$.*

PROPOSITION 5.1. *Under the stated assumptions, the whole sequence $\{x^k\}$ converges to x^*.*

Proof. Assumption A2 and A6 mean that x^* is an isolated accumulation point of $\{x^k\}$, see [16]. The assertion then follows from [11] and Theorem 4.1 (i). □

It follows from Lemma 5.1 in [6] that for every positive ϵ, $Z(x; \epsilon)$ is an SC^1-function on $\mathcal{A}$. Then applying Theorem 3.2 in [5] and similar to the proof of Theorem 5.2 in [6], we deduce the next theorem.

THEOREM 5.1. *Suppose that the stated assumptions hold. If*

$$\lim_{k\to\infty} \frac{\|x^k + d^k - x^*\|}{\|x^k - x^*\|} = 0,$$

then for all sufficiently large k, the unit steplength is accepted by the algorithm. That is,

$$x^{k+1} = x^k + d^k.$$

6. Numerical Results

In this section we report some numerical results to show viability of Algorithm 3.1, using a subset of test problems from Hock and Schittkowski [8]. The algorithm was coded in Matlab and run on a DEC George Server 8200. A brief sketch of our implementation is given as follows:

The method of Gill and Murray [9] was used to solve QP subproblems (11) and (13). The initial guess for the Lagrangian Hessian was $H^1 = I_n$. Updating of the Hessian approximation H^k was done using the damped BFGS formula described in Powell [14] with $\lambda(x)$ as the KKT multiplier estimation at x. We took $\beta^k = \text{trace}(H^k)$.

Table 1. Starting points for test problems

Problem	n	m	Starting point (a)	Starting point (b)
HS34	3	8	(0,1.05,2.9)	(0,0,0)
HS35	3	4	(0.5,0.5,0.5)	(1,1,1)
HS36	3	7	(10,10,10)	(10,11.5,10)
HS37	3	8	(10,10,10)	(20,20,20)
HS43	4	3	(0,0,0,0)	(2,2,2,2)
HS44	4	10	(0,0,0,0)	(2,2,2,2)
HS65	3	7	(0,0,0)	(-5,5,0)
HS66	3	8	(0,1.05,2.9)	(1,1,1)
HS93	6	8	(5.54,4.4,12.02,11.82,0.702,0.852)	(3,3,3,3,3,3)
HS100	7	4	(1,2,0,4,0,1,1)	(3,2,1,4,1,1,2)

The parameters used in this implementation were set as

$$\beta^1 = 1, \quad \mu = 0.1, \quad \omega = 0.5, \quad \tau^1 = 1, \quad \theta = 0.5, \quad \gamma^1 = 10000,$$
$$\sigma = 0.1, \quad \sigma_1 = 10 \quad \text{and} \quad \epsilon^1 = 3.0.$$

All the test problems in this implementation have inequality constraints only. We use two initial values for the primal variable: points (a) and (b), where (a) is feasible and (b) is infeasible, see Table 1; and one of them was taken from [8]. In this table the first column gives the problem number from [8], n and m are the numbers of the variables and of the constraints respectively.

The termination criterion for the algorithm was

$$\|(d^k, u^k_{p^k})\| \leqslant 10^{-10}.$$

The precision for the solution of QP subproblems was also 10^{-10}. We summarize the numerical results on 10 test problems in Table 2. In this table $Iter$ is the number of iterations, Nf and Ng denote the numbers of evaluations of f and g respectively. Fv is the final value of the function f and $Prec$ denotes the precision of the termination criterion in Algorithm 3.1. $Point$ denotes starting point (a) or (b). Moreover, the numbers of evaluations of ∇f and ∇g are the same as that of g.

All test problems with starting points chosen in Table 1 were solved successfully. The results of the numerical experiments show that the proposed algorithm works well and is quite promising.

7. Conclusion

By applying Spellucci's technique for dealing with inconsistent problems in the SQP method into the merit function developed by Lucidi, we present a new SQP

Table 2. Numerical results on test problems

Problem	Iter	Nf	Ng	Fv	Prec	Point
HS34	34	161	162	−0.83403244524796	0	(a)
	13	32	33	−0.83403244526530	0	(b)
HS35	10	46	46	0.11111111111111	0	(a)
	9	11	12	0.11111111111111	0	(b)
HS36	3	7	7	−3300	0	(a)
	5	10	10	−3300	0	(b)
HS37	12	34	34	−3456.000000000001	0	(a)
	29	113	119	−3456.000000000001	0	(b)
HS43	14	29	29	−44.00000000000001	0	(a)
	19	60	65	−44.00000000000011	0	(b)
HS44	7	13	13	−15	0	(a)
	8	18	21	−15	0	(b)
HS65	13	38	39	0.95352885680478	0	(a)
	15	41	44	0.95352885680188	0	(b)
HS66	8	8	8	0.51816327418154	0	(a)
	14	26	28	0.51816327418153	0	(b)
HS93	21	62	66	135.0759628317741	0	(a)
	34	79	84	135.0759628315913	0	(b)
HS100	28	96	96	680.6300573744018	0	(a)
	26	90	92	680.6300573743961	0	(b)

algorithm for solving nonlinear constrained optimization problems. The proposed algorithm possesses both the merits of the method in [6] and the merits of the method in [17]. In the algorithm, the subproblems are always consistent and do not necessarily contain all constraints. Under suitable assumptions we established global convergence. Superlinear convergence result was also obtained without assuming the strict complementarity. Moreover, we use an approximated direction derivative of the merit function so that only first order derivatives of the problem functions are required to evaluate.

We observe that Assumption A3 is a stronger condition. It might be an interesting topic for future research how to weaken Assumption A3 to the weak Mangasarian–Fromowitz condition as in [10] within the framework of our algorithm.

Appendix A

Proof of Proposition 3.1. Assume that (1) is consistent. The KKT conditions of (1) are as follows:

$$\begin{aligned} Hd + \nabla f(x) + \nabla g(x)\lambda &= 0, \\ \lambda \geqslant 0, \quad g(x) + \nabla g(x)^T d &\leqslant 0, \\ \lambda^T[g(x) + \nabla g(x)^T d] &= 0. \end{aligned} \tag{31}$$

Since (11) has always a unique optimal solution, the KKT conditions (12) are necessary and sufficient for optimality. Due to $g_i(x) < 0$ for $i \notin A$, the term $\|d\|$ small enough means

$$g_i(x) + \nabla g_i(x)^T d < 0 \quad \text{for} \quad i \notin A. \tag{32}$$

Hence it follows from (31) that $(d, u_P = 0)$ solves (12) provided that

$$\tau \geqslant \max\{\lambda_i : i \in P\},$$

with $\rho_A := \lambda_A$ and $\nu_P := \tau e_P - \lambda_P \geqslant 0$. Conversely, if $(d, 0)$ is a solution of (11) and $\|d\|$ is small enough such that (32) is satisfied, then it follows from (12) that d, $\lambda_A := \rho_A$, $\lambda_i := 0$ for $i \notin A$ satisfy (31). The proof is complete. □

The following two lemmas are due to Theorems 3.4 and 3.5 of Spellucci [17] and play a crucial role in the subsequent proof. It is not difficult to deduce that there exists some $\psi_0 > 0$ such that $\Omega(\psi_0) \subseteq \mathcal{A}$.

LEMMA 1. *Under Assumptions A1–A3, there exists some pair* $\Delta > 0$ *and* $\bar{\delta} > 0$ *independent of* x, *such that for any* $x \in \mathcal{A}$ *and for* $0 \leqslant \delta \leqslant \bar{\delta}$, *there exists some* $d \neq 0$ *satisfying*

$$\|d\| \leqslant \Delta \quad \textit{and} \quad \nabla g_{A(x,\delta)}(x)^T d \leqslant -e_{A(x,\delta)}.$$

LEMMA 2. *Under Assumptions A1-A3, there exists some triple* $\xi^* \in [0, 1]$, $\nu^* > 0$, $\Delta > 0$, *such that for any* $x \in \mathcal{A}$ *and for* $0 < \xi \leqslant \xi^*$ *there exists some* $d \neq 0$ *satisfying*

$$\begin{aligned} \|d\| &\leqslant \Delta, \\ \xi g_{P_0}(x) + \nabla g_{P_0}(x)^T d &\leqslant -\nu^* e_{P_0}, \\ g_{\bar{P}_0}(x) + \nabla g_{\bar{P}_0}(x)^T d &\leqslant -\nu^* e_{\bar{P}_0}, \end{aligned}$$

where $\bar{P}_0 := I \backslash P_0$.

Proof of Proposition 3.2. It follows from Lemma 1 that there exists some pair $\Delta > 0$ and $\bar{\delta} > 0$ independent of x, such that

$$\nabla g_{A(x,\bar{\delta})}(x)^T d(x) \leqslant -e_{A(x,\bar{\delta})}$$

for some $d(x)$ with $\|d(x)\| \leqslant \Delta$. For $i \in A(x, \bar{\delta})$, we have

$$\begin{aligned} &g_i(x) + \nabla g_i(x)^T d(x)(\psi(x) + \bar{\delta}/(4M_1\Delta)) \\ &\quad \leqslant \psi(x) - \psi(x) - \bar{\delta}/(4M_1\Delta) \\ &\quad = -\bar{\delta}/(4M_1\Delta). \end{aligned}$$

For $i \notin A(x, \bar{\delta})$, we get

$$\begin{aligned} &g_i(x) + \nabla g_i(x)^T d(x)(\psi(x) + \bar{\delta}/(4M_1\Delta)) \\ &\quad \leqslant -\bar{\delta} + M_1\Delta\psi(x) + M_1\Delta\bar{\delta}/(4M_1\Delta) \\ &\quad \leqslant -\bar{\delta}/4, \end{aligned}$$

if $\psi(x) \leqslant \bar{\delta}/(2M_1\Delta)$. Let $\bar{\eta} := \min\{\psi_0, \bar{\delta}/(2M_1\Delta)\}$. Then for any $x \in \Omega(\bar{\eta})$, $d := d(x)(\psi(x) + \bar{\delta}/(4M_1\Delta))$ satisfies

$$g(x) + \nabla g(x)^T d \leqslant -\nu e$$

with $\nu := \min\{\bar{\delta}/4, \bar{\delta}/(4M_1\Delta)\}$. This means that the problem (1) is consistent and satisfies a Slater-condition uniformly in x. Since its Hessian and their inverse are uniformly bounded if $\|H\|$ and $\|H^{-1}\|$ are, its multiplier λ is also uniformly bounded. It follows from the proof of Proposition 3.1 that for any $\beta > 0$ and $\tau > \max\{\lambda_i : i \in P\}$, $(d, u_P = 0)$ solves (11). The proof is complete. □

Proof of Proposition 3.3. In the following proof we use the constants Δ, ξ^*, $\bar{\delta}$ and ν^* given in Lemma 1 and Lemma 2. We proved in Proposition 3.2 the following assertion: If $\psi(x) \leqslant \bar{\eta}$ with $\bar{\eta} := \min\{\psi_0, \bar{\delta}/(2M_1\Delta)\}$, then there exists some τ_1^* such that for $\tau \geqslant \tau_1^*$, $u_P = 0$.

We now consider another case: $\psi(x) > \bar{\eta}$. It follows from Lemma 2 that for every $x \in \mathcal{A}$ there exists some $\tilde{d}$ such that

$$\begin{aligned} \|\tilde{d}\| &\leqslant \Delta, \\ \xi^* g_{P_0}(x) + \nabla g_{P_0}(x)^T \tilde{d} &\leqslant -\nu^* e_{P_0}, \\ g_{\bar{P}_0}(x) + \nabla g_{\bar{P}_0}(x)^T \tilde{d} &\leqslant -\nu^* e_{\bar{P}_0}. \end{aligned}$$

Let

$$\tilde{u}_P := (1 - \xi^*) g_P(x).$$

Then $(\tilde{d}, \tilde{u}_P)$ is feasible for (11) and we have

$$\begin{aligned} &\frac{1}{2}\tilde{d}^T H \tilde{d} + \nabla f(x)^T \tilde{d} + \tau e_P^T \tilde{u}_P + \frac{1}{2}\beta\|\tilde{u}_P\|^2 \\ &\quad \leqslant C + (1 - \xi^*)\tau\psi(x) + \frac{1}{2}\beta\psi(x)^2, \end{aligned} \tag{33}$$

where

$$C := M_1\Delta + \frac{1}{2}\Delta^2\|H\|$$

is uniformly bounded if $\|H\|$ is.

For every $d \in \mathbb{R}^n$, we get

$$\begin{aligned} \frac{1}{2}d^T H d + \nabla f(x)^T d &= \frac{1}{2}(H^{1/2}d)^T H^{1/2} d + \nabla f(x)^T d \\ &\geqslant -\frac{1}{2}(H^{-1/2}\nabla f(x))^T (H^{-1/2}\nabla f(x)) \\ &= -\frac{1}{2}\nabla f(x)^T H^{-1} \nabla f(x) \\ &\geqslant -\frac{1}{2}M_1^2\|H^{-1}\|. \end{aligned} \tag{34}$$

Moreover, the solution (d, u_P) of (11) satisfies

$$\frac{1}{2}d^T Hd + \nabla f(x)^T d + \tau e_P^T u_P + \frac{1}{2}\beta\|u_P\|^2 \leqslant \frac{1}{2}\tilde{d}^T H\tilde{d} + \nabla f(x)^T \tilde{d} + \tau e_P^T \tilde{u}_P + \frac{1}{2}\beta\|\tilde{u}_P\|^2,$$

which together with (33) and (34), implies

$$\tau \sum_{i\in P(x)} u_i \leqslant (1-\xi^*)\tau\psi(x) + C_1, \tag{35}$$

where C_1 satisfies

$$C_1 \geqslant C + \frac{1}{2}M_1^2\|H^{-1}\| + \frac{1}{2}\beta\psi(x)^2$$

for $x \in \mathcal{A}$. Let

$$\tau_2^* := \frac{2C_1}{\xi^*\bar{\eta}}.$$

Then for $\tau \geqslant \tau_2^*$, (35) implies

$$\sum_{i\in P(x)} u_i \leqslant (1-\xi^*/2)\psi(x).$$

The assertion is satisfied with $\tau^* := \max\{\tau_1^*, \tau_2^*\}$ and $L := 1 - \xi^*/2$. □

Proof of Corollary 3.1. Repeating with minor modifications the proof of Proposition 3.3, we can deduce

$$\tau \sum_{i\in P(x)} v_i(x)u_i \leqslant (1-\xi^*)\tau \sum_{i\in P(x)} v_i(x)g_i(x) + C_1 \tag{36}$$

for $x \in \mathcal{A}$. If $\psi(x) \geqslant \eta$, then

$$\sum_{i\in P(x)} v_i(x)g_i(x) \geqslant \sum_{i\in P(x)} g_i(x)^2 \geqslant \psi(x)^2/m \geqslant \eta^2/m.$$

It follows from (36) that the first assertion is satisfied with $L := 1 - \xi^*/2$ and

$$\tau^* := \frac{2mC_1}{\xi^*\eta^2},$$

which implies the second assertion. □

Proof of Proposition 3.6. We only need to show that: if $\tilde{H} \in \mathbb{R}^{n\times n}$ is positive definite, $\tilde{\delta} > 0$, $\tilde{\tau} > 0$, $\tilde{\beta} > 0$ and $\tilde{x} \in \mathcal{A}$, then for every $\eta \in [0, 1)$ and $\tilde{t} > 0$, there exists an $\tilde{\epsilon} > 0$ such that for all $\epsilon \in (0, \tilde{\epsilon}]$ and all $t \in (0, \tilde{t}]$,

$$DZ(\tilde{x}, \tilde{d}; \epsilon, t) \leqslant -\eta \max[\tilde{d}^T \tilde{H}\tilde{d}, \|c(\tilde{x}; \epsilon)\|^2],$$

where $(\tilde{d}, \tilde{u}_{\tilde{P}})$ is the solution of problem $QP_1(\tilde{x}, \tilde{H}; \tilde{A}, \tilde{\tau}, \tilde{\beta})$ or $QP_2(\tilde{x}, \tilde{H}; \tilde{A}, \tilde{\tau}, \tilde{\beta})$ with $\tilde{A} := A(\tilde{x}, \tilde{\delta})$ and $\tilde{P} := P(\tilde{x})$.

If we come to Step 9, it follows from Lemma 3.1 that $\tilde{d} \neq 0$. We can write

$$\begin{aligned} DZ(\tilde{x}, \tilde{d}; \epsilon, t) = & \nabla f(\tilde{x})^T \tilde{d} + \sum_{i=1}^{m} \left[\lambda_i(\tilde{x}) + \frac{1}{\epsilon a(\tilde{x})} r_i(\tilde{x}; \epsilon) \right] \nabla g_i(\tilde{x})^T \tilde{d} \\ & + \sum_{i=1}^{m} \frac{1}{t} [\lambda_i(\tilde{x} + t\tilde{d}) - \lambda_i(\tilde{x})] c_i(\tilde{x}; \epsilon). \end{aligned} \tag{37}$$

Moreover, we deduce from the continuous differentiability of $\lambda(x)$ that there exists a constant $\Lambda > 0$ such that

$$\frac{1}{t}[\lambda_i(\tilde{x} + t\tilde{d}) - \lambda_i(\tilde{x})] \leqslant \Lambda, \qquad \forall i \in I \text{ and } t \in (0, \tilde{t}].$$

If $g_i(\tilde{x}) > 0$, then $\nabla g_i(\tilde{x})^T \tilde{d} \leqslant \tilde{u}_i - g_i(\tilde{x})$, $c_i(\tilde{x}; 0) = g_i(\tilde{x})$, and $r_i(\tilde{x}; 0) = 2v_i(\tilde{x}) > 0$. Hence by (17), we have

$$\begin{aligned} & \lim_{\epsilon \to 0} \sum_{i \in \tilde{P}} \left\{ \left[\lambda_i(\tilde{x}) + \frac{1}{\epsilon a(\tilde{x})} r_i(\tilde{x}; \epsilon) \right] \nabla g_i(\tilde{x})^T \tilde{d} + \frac{1}{t}[\lambda_i(\tilde{x} + t\tilde{d}) - \lambda_i(\tilde{x})] c_i(\tilde{x}; \epsilon) \right\} \\ & \leqslant \lim_{\epsilon \to 0} \sum_{i \in \tilde{P}} \left\{ \lambda_i(\tilde{x}) \nabla g_i(\tilde{x})^T \tilde{d} + \frac{1}{\epsilon a(\tilde{x})} r_i(\tilde{x}; \epsilon)[\tilde{u}_i - g_i(\tilde{x})] + c_i(\tilde{x}; \epsilon)\Lambda \right\} \\ & = -\infty. \end{aligned} \tag{38}$$

If $g_i(\tilde{x}) = 0$, then $\nabla g_i(\tilde{x})^T \tilde{d} \leqslant 0$ and

$$c_i(\tilde{x}; \epsilon) = y_i^2(\tilde{x}; \epsilon) = -\min\left[0, \frac{\epsilon a(\tilde{x})}{2} \lambda_i(\tilde{x})\right].$$

Hence, $c_i(\tilde{x}, 0) = 0$ and

$$\begin{aligned} & \lim_{\epsilon \to 0} \left\{ \left[\lambda_i(\tilde{x}) + \frac{1}{\epsilon a(\tilde{x})} r_i(\tilde{x}; \epsilon) \right] \nabla g_i(\tilde{x})^T \tilde{d} + \frac{1}{t}[\lambda_i(\tilde{x} + t\tilde{d}) - \lambda_i(\tilde{x})] c_i(\tilde{x}; \epsilon) \right\} \\ & \quad = \lim_{\epsilon \to 0} \{\lambda_i(\tilde{x}) - \min[0, \lambda_i(\tilde{x})]\} \nabla g_i(\tilde{x})^T \tilde{d} \leqslant 0. \end{aligned} \tag{39}$$

If $g_i(\tilde{x}) < 0$, then $g_i^+(\tilde{x}) = 0$ and $g_i(\tilde{x}) + \frac{\epsilon a(\tilde{x})}{2} \lambda_i(\tilde{x}) < 0$ for all sufficiently small $\epsilon > 0$. Therefore,

$$c_i(\tilde{x}; \epsilon) = -\frac{\epsilon a(\tilde{x})}{2} \lambda_i(\tilde{x}) \quad \text{and} \quad r_i(\tilde{x}; \epsilon) = -\epsilon a(\tilde{x}) \lambda_i(\tilde{x}).$$

So, we have

$$\lim_{\epsilon\to 0}\left\{\left[\lambda_i(\tilde{x})+\frac{1}{\epsilon a(\tilde{x})}r_i(\tilde{x};\epsilon)\right]\nabla g_i(\tilde{x})^T\tilde{d}+\frac{1}{t}[\lambda_i(\tilde{x}+t\tilde{d})-\lambda_i(\tilde{x})]c_i(\tilde{x};\epsilon)\right\}=0. \tag{40}$$

We shall show the assertion by distinguishing the following two possible cases:

Case 1. $\tilde{x}\in\mathcal{F}$.

Note that $g(\tilde{x})\leqslant 0$. By (37), (39), (40) and (12), we get

$$\begin{aligned}\lim_{\epsilon\to 0} DZ(\tilde{x},\tilde{d};\epsilon,t) &\leqslant \nabla f(\tilde{x})^T\tilde{d}\\ &= -\tilde{d}^T\tilde{H}\tilde{d}-\tilde{d}^T\nabla g_{\tilde{A}}(\tilde{x})\rho_{\tilde{A}}\\ &= -\tilde{d}^T\tilde{H}\tilde{d}+g_{\tilde{A}}(\tilde{x})^T\rho_{\tilde{A}}\\ &\leqslant -\tilde{d}^T\tilde{H}\tilde{d}.\end{aligned}$$

The assertion then follows from the fact that $0\leqslant\eta<1$ and $c(\tilde{x};0)=0$.

Case 2. $\tilde{x}\in\mathcal{A}\backslash\mathcal{F}$.

In this case, by (37)–(40), we have, for all $t\in(0,\tilde{t}]$,

$$\lim_{\epsilon\to 0} DZ(\tilde{x},\tilde{d};\epsilon,t)=-\infty,$$

and hence the assertion is obvious. □

Proof of Proposition 4.4. Suppose that the assertion does not hold. Then, there exist sequences $\{x^r\}\subseteq\mathcal{A}$, $\{H^r\}\subseteq\mathbb{R}^{n\times n}$, $\{\epsilon^r\}$, $\{\alpha^r\}$ generated by Algorithm 3.1, such that

$$x^r\to\tilde{x},\quad H^r\to\tilde{H},\quad \epsilon^r\to 0,\quad \alpha^r\in(0,1],$$

and

$$DZ(x^r,d^r;\epsilon^r,\alpha^r)>-\frac{1}{2}\max[(d^r)^T H^r d^r,\|c(x^r;\epsilon^r)\|^2]. \tag{41}$$

By Lemma 4.1, we only need to consider the following two possible cases. We will obtain a contradiction in each case.

Case 1. $\tilde{x}\in\mathcal{A}\backslash\mathcal{F}$.

Using (15), we can write

$$\begin{aligned}DZ(x^r,d^r;\epsilon^r,\alpha^r) =&\nabla f(x^r)^T d^r+\sum_{i=1}^m\left[\lambda_i(x^r)+\frac{1}{\epsilon^r a(x^r)}r_i(x^r;\epsilon^r)\right]\nabla g_i(x^r)^T d^r\\ &+\sum_{i=1}^m\frac{1}{\alpha^r}[\lambda_i(x^r+\alpha^r d^r)-\lambda_i(x^r)]c_i(x^r;\epsilon^r).\end{aligned} \tag{42}$$

Since $f(x)$ and $\lambda(x)$ are continuously differentiable, it follows from Proposition 4.3 that the first and third terms of the expression (42) are bounded.

We now consider the second term of (42). Noting that the number of subsets in I is finite, without loss of generality, we can assume $A^r = \hat{A}$ and $P^r = \hat{P}$ for all r, where $\hat{A}$ and $\hat{P}$ are subsets of I. Moreover, we assume that $\tau^r = \tilde{\tau}$ for all r, $d^r \to \tilde{d}$ and $\beta^r \to \tilde{\beta}$. From (12) and (17), we deduce

$$\begin{aligned}\sum_{i\in\hat{P}} v_i(x^r)\nabla g_i(x^r)^T d^r &\leqslant \sum_{i\in\hat{P}} v_i(x^r)[u_i^r - g_i(x^r)]\\ &\leqslant -\frac{\omega}{\tilde{\tau}}\sum_{i\in\hat{P}} v_i(x^r)g_i(x^r).\end{aligned}$$

Therefore, we have

$$\begin{aligned}\lim_{r\to\infty}\sum_{i\in\hat{P}} r_i(x^r;\epsilon^r)\nabla g_i(x^r)^T d^r &= 2\sum_{i\in\hat{P}} v_i(\tilde{x})\nabla g_i(\tilde{x})^T \tilde{d}\\ &\leqslant -\frac{2\omega}{\tilde{\tau}}\sum_{i\in\hat{P}} v_i(\tilde{x})g_i(\tilde{x}) < 0.\end{aligned}$$

This shows

$$\lim_{r\to\infty}\sum_{i\in\hat{P}}\left[\lambda_i(x^r) + \frac{1}{\epsilon^r a(x^r)} r_i(x^r;\epsilon^r)\right]\nabla g_i(x^r)^T d^r = -\infty.$$

On the other hand, for $i \in I\backslash\hat{P}$, $g_i^+(x^r) = 0$ and

$$\begin{aligned}&\left[\lambda_i(x^r) + \frac{1}{\epsilon^r a(x^r)} r_i(x^r;\epsilon^r)\right]\nabla g_i(x^r)^T d^r\\ &\quad = \left[\lambda_i(x^r) + \frac{1}{\epsilon^r a(x^r)} 2c_i(x^r;\epsilon^r)\right]\nabla g_i(x^r)^T d^r \qquad (43)\\ &\quad = \frac{2}{\epsilon^r a(x^r)}\max\left\{g_i(x^r) + \frac{\epsilon^r a(x^r)}{2}\lambda_i(x^r), 0\right\}\nabla g_i(x^r)^T d^r.\end{aligned}$$

If $g_i(x^r) + \frac{\epsilon^r a(x^r)}{2}\lambda_i(x^r) \geqslant 0$, then $g_i(x^r) \geqslant -\frac{\epsilon^r a(x^r)}{2}\lambda_i(x^r) \geqslant -\delta^r$; i.e., $i \in \hat{A}\backslash\hat{P}$ and

$$\nabla g_i(x^r)^T d^r \leqslant -g_i(x^r) \leqslant \frac{\epsilon^r a(x^r)}{2}\lambda_i(x^r).$$

Hence, by (43), we have

$$\begin{aligned}&\left[\lambda_i(x^r) + \frac{1}{\epsilon^r a(x^r)} r_i(x^r;\epsilon^r)\right]\nabla g_i(x^r)^T d^r\\ &\quad \leqslant \lambda_i(x^r)\max\left\{g_i(x^r) + \frac{\epsilon^r a(x^r)}{2}\lambda_i(x^r), 0\right\}.\end{aligned}$$

While if $g_i(x^r) + \frac{\epsilon^r a(x^r)}{2}\lambda_i(x^r) < 0$, then by (43), we get

$$\left[\lambda_i(x^r) + \frac{1}{\epsilon^r a(x^r)} r_i(x^r;\epsilon^r)\right] \nabla g_i(x^r)^T d^r = 0.$$

So, for $i \in I\backslash\hat{P}$, $[\lambda_i(x^r) + \frac{1}{\epsilon^r a(x^r)} r_i(x^r;\epsilon^r)]\nabla g_i(x^r)^T d^r$ is bounded from above.

The above analysis shows that

$$DZ(x^r, d^r; \epsilon^r, \alpha^r) \to -\infty,$$

which contradicts (41).

Case 2. $\tilde{x} \in \mathcal{F}$.

In this case, for r sufficiently large, d^r is the solution of problem $QP_0(x^r, H^r; A^r)$ by Proposition 3.2. Moreover, passing to a subsequence if necessary, without loss of generality, we can assume that for all r,

$$A^r = \hat{A} \tag{44}$$

where $\hat{A}$ is some subset of I.

Let $(\tilde{d}, \tilde{\rho}_{\hat{A}})$ be a KKT pair for problem $QP_0(\tilde{x}, \tilde{H}; \hat{A})$. By Proposition 3.2 in [6], $d(x, H)$ is continuous in a neighborhood of $(\tilde{x}, \tilde{H})$. Hence, $d^r \to \tilde{d}$.

Consider first the case $\tilde{d} \neq 0$. By the proof of Proposition 3.6, we deduce that

$$\lim_{\epsilon\to 0} DZ(\tilde{x}, \tilde{d}; \epsilon, t) \leqslant -\tilde{d}^T \tilde{H}\tilde{d}$$

for all $t \in (0, 1]$. From the fact that $DZ(x, d; \epsilon, t)$ is continuous and $\lim_{r\to\infty} c(x^r;\epsilon^r) = 0$, we have, for r sufficiently large,

$$DZ(x^r, d^r; \epsilon^r, \alpha^r) \leqslant -\frac{1}{2}\max[(d^r)^T H^r d^r, \|c(x^r;\epsilon^r)\|^2],$$

which contradicts (41).

Then consider the case $\tilde{d} = 0$. Assumption A4 implies that for all r,

$$(d^r)^T H^r d^r \geqslant \frac{1}{C}\|d^r\|^2. \tag{45}$$

Let $\rho^r_{I\backslash\hat{A}} := 0$ for all r. Similar to the proof of Lemma 4.3 in [6], we can deduce that there exists a constant $\gamma_2 > 0$ such that

$$\|\lambda(x^r) - \rho^r\| \leqslant \gamma_2\|d^r\|. \tag{46}$$

Moreover, the differentiability of $\lambda(x)$ implies that there exists a constant $\gamma_3 > 0$ such that

$$\frac{1}{\alpha^r}\|\lambda(x^r + \alpha^r d^r) - \lambda(x^r)\| \leqslant \gamma_3\|d^r\|. \tag{47}$$

For r sufficiently large, if $i \in P^r$, then $g_i^+(x^r) = g_i(x^r) > 0$ and $\nabla g_i(x^r)^T d^r \leqslant -g_i(x^r) < 0$; while if $i \in I \backslash P^r$, then $g_i^+(x^r) = 0$. And renumbering if necessary, we can suppose without loss of generality that we can partition $\hat{A}$ into two subsets B and $\hat{A} \backslash B$ such that

$$\rho_B^r \geqslant 0, \quad g_B(x^r) + \nabla g_B(x^r)^T d^r = 0, \tag{48}$$

and

$$\rho_{\hat{A}\backslash B}^r = 0, \quad g_{\hat{A}\backslash B}(x^r) + \nabla g_{\hat{A}\backslash B}(x^r)^T d^r < 0. \tag{49}$$

Let $\bar{B} := I \backslash B$. Then $\rho_{\bar{B}}^r = 0$. From (15), (41), (48) and the KKT system of $QP_0(x^r, H^r; \hat{A})$, we obtain

$$\begin{aligned}
0 &< DZ(x^r, d^r; \epsilon^r, \alpha^r) + \frac{1}{2}\max[(d^r)^T H^r d^r, \|c(x^r; \epsilon^r)\|^2] \\
&\leqslant -\frac{1}{2}(d^r)^T H^r d^r + \frac{1}{2}\|c(x^r; \epsilon^r)\|^2 \\
&\quad + (d^r)^T \nabla g(x^r)(\lambda(x^r) - \rho^r) + \frac{2}{\epsilon^r a(x^r)}(d^r)^T \nabla g(x^r) c(x^r; \epsilon^r) \\
&\quad + \frac{1}{\alpha^r}[\lambda(x^r + \alpha^r d^r) - \lambda(x^r)]^T c(x^r; \epsilon^r) \\
&= -\frac{1}{2}(d^r)^T H^r d^r + \frac{1}{2}\|c(x^r; \epsilon^r)\|^2 \\
&\quad - g_B(x^r)^T(\lambda_B(x^r) - \rho_B^r) + (d^r)^T \nabla g_{\bar{B}}(x^r)\lambda_{\bar{B}}(x^r) \\
&\quad - \frac{2}{\epsilon^r a(x^r)} g_B(x^r)^T c_B(x^r; \epsilon^r) + \frac{2}{\epsilon^r a(x^r)}(d^r)^T \nabla g_{\bar{B}}(x^r) c_{\bar{B}}(x^r; \epsilon^r) \\
&\quad + \frac{1}{\alpha^r}[\lambda(x^r + \alpha^r d^r) - \lambda(x^r)]^T c(x^r; \epsilon^r).
\end{aligned} \tag{50}$$

Denote

$$\begin{aligned}
B_+(x^r) &:= \{i \in B \mid 2g_i(x^r) + \epsilon^r a(x^r)\lambda_i(x^r) \geqslant 0\}, \\
B_-(x^r) &:= \{i \in B \mid 2g_i(x^r) + \epsilon^r a(x^r)\lambda_i(x^r) < 0\}, \\
\bar{B}_+(x^r) &:= \{i \in \bar{B} \mid 2g_i(x^r) + \epsilon^r a(x^r)\lambda_i(x^r) \geqslant 0\}, \\
\bar{B}_-(x^r) &:= \{i \in \bar{B} \mid 2g_i(x^r) + \epsilon^r a(x^r)\lambda_i(x^r) < 0\}.
\end{aligned}$$

If $i \in B_+(x^r) \cup \bar{B}_+(x^r)$, then $c_i(x^r;\epsilon^r) = g_i(x^r)$; while if $i \in B_-(x^r) \cup \bar{B}_-(x^r)$, then $c_i(x^r;\epsilon^r) = -\frac{\epsilon^r a(x^r)}{2}\lambda_i(x^r)$. Therefore,

$$\begin{aligned} & -g_B(x^r)^T(\lambda_B(x^r)-\rho_B^r) - \frac{2}{\epsilon^r a(x^r)} g_B(x^r)^T c_B(x^r;\epsilon^r) \\ &= -c_B(x^r;\epsilon^r)^T(\lambda_B(x^r)-\rho_B^r) - \frac{2}{\epsilon^r a(x^r)}\|c_B(x^r;\epsilon^r)\|^2 \\ &\quad - (c_{B_-}(x^r;\epsilon^r) - g_{B_-}(x^r))^T \rho_{B_-}^r \\ &\leqslant \gamma_2 \|c_B(x^r;\epsilon^r)\|\|d^r\| - \frac{2}{\epsilon^r a(x^r)}\|c_B(x^r;\epsilon^r)\|^2. \end{aligned} \tag{51}$$

If $i \in \bar{B}_+(x^r)$, then $g_i(x^r) \geqslant -\frac{\epsilon^r a(x^r)}{2}\lambda_i(x^r) \geqslant -\delta^r$. This shows $i \in \hat{A}\backslash B$. Hence, (49) holds and

$$\begin{aligned} & (d^r)^T \nabla g_{\bar{B}}(x^r)\lambda_{\bar{B}}(x^r) + \frac{2}{\epsilon^r a(x^r)}(d^r)^T \nabla g_{\bar{B}}(x^r) c_{\bar{B}}(x^r;\epsilon^r) \\ &\quad = (d^r)^T \nabla g_{\bar{B}_+}(x^r)\left[\lambda_{\bar{B}_+}(x^r) + \frac{2}{\epsilon^r a(x^r)} g_{\bar{B}_+}(x^r)\right] \\ &\quad \leqslant -g_{\bar{B}_+}(x^r)^T\left[\lambda_{\bar{B}_+}(x^r) + \frac{2}{\epsilon^r a(x^r)} g_{\bar{B}_+}(x^r)\right] \\ &\quad \leqslant -\frac{2}{\epsilon^r a(x^r)}\|g_{\bar{B}_+}(x^r)\|^2 + \gamma_2\|g_{\bar{B}_+}(x^r)\|\|d^r\|. \end{aligned} \tag{52}$$

Moreover, it is not difficult to deduce that

$$\begin{aligned} & \frac{1}{\alpha^r}[\lambda(x^r+\alpha^r d^r) - \lambda(x^r)]^T c(x^r;\epsilon^r) \\ &\quad \leqslant \frac{1}{\alpha^r}\|\lambda_B(x^r+\alpha^r d^r) - \lambda_B(x^r)\|\|c_B(x^r;\epsilon^r)\| \\ &\qquad + \frac{1}{\alpha^r}\|\lambda_{\bar{B}}(x^r+\alpha^r d^r) - \lambda_{\bar{B}}(x^r)\|\|c_{\bar{B}}(x^r;\epsilon^r)\| \\ &\quad \leqslant \gamma_3\|d^r\|[\|c_B(x^r;\epsilon^r)\| + \|g_{\bar{B}_+}(x^r)\| + \|c_{\bar{B}_-}(x^r;\epsilon^r)\|] \\ &\quad \leqslant \gamma_3\|c_B(x^r;\epsilon^r)\|\|d^r\| + \gamma_3\|g_{\bar{B}_+}(x^r)\|\|d^r\| + \frac{\epsilon^r a(x^r)\gamma_2\gamma_3}{2}\|d^r\|^2 \end{aligned} \tag{53}$$

and

$$\begin{aligned} \frac{1}{2}\|c(x^r;\epsilon^r)\|^2 &= \frac{1}{2}\|c_B(x^r;\epsilon^r)\|^2 + \frac{1}{2}\|g_{\bar{B}_+}(x^r)\|^2 + \frac{1}{2}\|c_{\bar{B}_-}(x^r;\epsilon^r)\|^2 \\ &\leqslant \frac{1}{2}\|c_B(x^r;\epsilon^r)\|^2 + \frac{1}{2}\|g_{\bar{B}_+}(x^r)\|^2 + \frac{(\epsilon^r a(x^r)\gamma_2)^2}{8}\|d^r\|^2. \end{aligned} \tag{54}$$

Combining (50) and (51)–(54), we obtain

$$-\left(\frac{1}{2C}-\frac{\epsilon^r a(x^r)\gamma_2\gamma_3}{2}-\frac{(\epsilon^r a(x^r)\gamma_2)^2}{8}\right)\|d^r\|^2$$
$$+(\gamma_2+\gamma_3)\|c_B(x^r;\epsilon^r)\|\|d^r\|-\left(\frac{2}{\epsilon^r a(x^r)}-\frac{1}{2}\right)\|c_B(x^r;\epsilon^r)\|^2 \tag{55}$$
$$+(\gamma_2+\gamma_3)\|g_{\bar{B}_+}(x^r)\|\|d^r\|-\left(\frac{2}{\epsilon^r a(x^r)}-\frac{1}{2}\right)\|g_{\bar{B}_+}(x^r)\|^2>0.$$

Since $\lim_{r\to\infty}\epsilon^r=0$, for ϵ^r sufficiently small, the left-hand side of (55) is a negative-definite quadratic form in $\|d^r\|$, $\|c_B(x^r;\epsilon^r)\|$, $\|g_{\bar{B}_+}(x^r)\|$. So, we obtain a contradiction, completing the proof. □

Acknowledgements

The authors would like to thank the referees and the associate editor for their helpful comments and suggestions.

References

1. Boggs, P.T. and Tolle, J.W. (1995), Sequential quadratic programming, *Acta Numerica*, Cambridge University Press, 1995, pp. 1–51.
2. Burke, J.V. (1989), A sequential quadratic programming algorithm for potentially infeasible mathematical programs, *Journal of Mathematical Analysis and Applications* 139: 319–351.
3. Burke, J.V. and Han, S.P. (1989), A robust sequential quadratic programming method, *Mathematical Programming* 43: 277–303.
4. Byrd, R.H., Tapia, R.A. and Zhang, Y. (1992), An SQP augmented Lagrangian BFGS algorithm for constrained optimization, *SIAM Journal on Optimization* 2: 210–241.
5. Facchinei, F. (1995), Minimization of SC^1-functions and the Maratos effect, *Operations Research Letters* 17: 131–137.
6. Facchinei, F. (1997), Robust recursive quadratic programming algorithm model with global and superlinear convergence properties, *Journal of Optimization Theory and Applications* 92: 543–579.
7. Glad, T. and Polak, E. (1979), A multiplier method with automatic limitation of penalty growth, *Mathematical Programming* 17: 140–155.
8. Hock, W. and Schittkowski, K. (1981), *Test examples for nonlinear programming codes*, Lecture Notes in Economics and Mathematical Systems 187, Springer-Verlag, Berlin, New York.
9. Gill, P.E. and Murray, W. (1978), Numerically stable methods for quadratic programming, *Mathematical Programming* 14: 349–372.
10. Lucidi, S. (1992), New results on a continuously differentiable penalty function, *SIAM Journal on Optimization* 2: 558–574.
11. Moré, J.J. and Sorensen, D.C. (1983), Computing a trust region step, *SIAM Journal on Scientific and Statistical Computing* 4: 553–572.

12. Panier, E.R. and Tits, A.L. (1987), A superlinearly convergent feasible method for the solution of inequality constrained optimization problems, *SIAM Journal on Control and Optimization* 25: 934–950.
13. Pantoja, J.F.A. and Mayne, D.Q. (1991), Exact penalty function algorithm with simple updating of the penalty parameter, *Journal of Optimization Theory and Applications* 69: 441–467.
14. Powell, M.J.D. (1978), A fast algorithm for nonlinearly constrained optimization calculations, in *Numerical analysis*, Proceedings, Biennial Conference, Dundee 1977, Lecture Notes in Math. 630, Springer-Verlag, Berlin, New York, pp. 144–157.
15. Qi, L. and Sun, J. (1993), A nonsmooth version of Newton's method, *Mathematical Programming* 58: 353–367.
16. Robinson, S.M. (1980), Strongly regular generalized equations, *Mathematics of Operations Research* 5: 43–62.
17. Spellucci, P. (1998), A new technique for inconsistent QP problems in the SQP methods, *Mathematical Methods of Operations Research* 47: 355–400.
18. Tone, K. (1983), Revision of constraint approximations in the successive QP-method for nonlinear programming problems, *Mathematical Programming* 26: 144–152.