A Sequential Quadratic Programming Method for Potentially Infeasible Mathematical Programs*

J. V. BURKE

*Department of Mathematics, GN-50, University of Washington,
Seattle, Washington 98195*

The paper describes a modification to the Wilson–Han–Powell sequential quadratic programming technique that is applicable to potentially infeasible nonlinear programs. The method is designed to locate stationary points of a nonlinear program in a generalized sense (Definition 2.8). These stationary points include the familiar Kuhn–Tucker and Fritz John points, but also include what we call *external and strong-external* stationary points. The external and strong external stationary points are not feasible. Thus the problem may be infeasible and yet have stationary points. Loosely speaking, an external stationary point is one that satisfies a certain first-order necessary optimality condition for being as *close* to feasibility as possible. The proposed technique is similar to those described by J. V. Burke and S-P. Han ("A Modified Sequential Quadratic Programming Method," manuscript) and M. Sahba (*J. Optim. Theory Appl.* **52** (1987), 291–309); however, it is shown to possess substantially enhanced stability properties. The global convergence properties of the method are described along with some continuity results for the quadratic programming subproblems. © 1989 Academic Press, Inc.

1. INTRODUCTION

We consider the constrained optimization problem

NLP: minimize $f(x)$

subject to $g(x) \in C$ and $x \in X$,

where it is assumed that the functions $f: \mathbb{R}^n \to \mathbb{R}$ and $g: \mathbb{R}^n \to \mathbb{R}^m$ are C^1 and the sets $C \subset \mathbb{R}^m$ and $X \subset \mathbb{R}^n$ are nonempty closed and convex. Throughout the remainder of the paper it is assumed that the model NLP satisfies the conditions posited in the above statement. However, it is not assumed that NLP has a solution or is even feasible. The reason for not imposing these restrictions is that it is often the case, when one attempts to model a large

319

nonlinear system via a nonlinear program, that the consistency of the non-linear program is not known beforehand. In this situation, one would greatly appreciate some information as to why the system is inconsistent or, subject to infeasibility, what is the best one can do in the sense of (1.1). Consequently, our goal is to present an iterative method for locating "stationary" points of NLP in a generalized sense. Loosely speaking, we say that a point is a stationary point for NLP if it is a stationary point for the problem

$$\text{minimize } f(x) \tag{1.1}$$

$$\text{subject to } x \in \operatorname{argmin}\{\operatorname{dist}(g(x)\,|\,C)\,|\,x \in X\},$$

where $\operatorname{dist}(y\,|\,C) := \inf\{|y-z| : z \in C\}$ and $|\cdot|$ is a given norm on $\mathbb{R}^m$. A precise meaning is given to this statement in Section 2.

The algorithm that we propose is quite similar to the one presented in [3], which is turn is based upon the work in [2]. Recall that our motivation in [3] was to develop a robust extension to the well-known sequential quadratic programming (SQP) method of Wilson [22], Han [9], and Powell [12]. By robust we mean that the method is designed to overcome the difficulties associated with infeasible quadratic programming sub-problems and the possible occurrence of a divergent sequence of search directions. Furthermore, we were interested in developing a modification to the standard SQP technique that was amenable to a trust region-like methodology. The objective of this paper is similar. However, we now consider the more general problem NLP. The approach that we take is quite different and allows a more complete analysis of the stability and continuity properties of the method.

A straightforward application of the SQP technique to the problem NLP requires that one iteratively solve a direction-finding subproblem of the form

$$\text{minimize } \nabla f(x)^T d + \tfrac{1}{2} d^T H d$$
$$\text{subject to } g(x) + g'(x)\,d \in C \text{ and } x + d \in X. \tag{1.2}$$

The primary difficulty with these subproblems is their potential infeasibility. Moreover, this difficulty is exacerbated if one wishes to imple-ment a trust region-like variation of the procedure. One way to alleviate this problem and yet maintain the flavor of the SQP approach would be to modify the constraints appearing in (1.2). Two such modifications are

$$g(x) + g'(x)\,d \in C + s,$$
$$x + d \in X, \tag{1.3}$$
$$d \in \beta \mathbb{B}^n,$$

and

$$g(x) + g'(x)\, d \in C + \kappa \mathbb{B}^m,$$

$$x + d \in X, \tag{1.4}$$

$$d \in \beta \mathbb{B}^n,$$

where $\mathbb{B}^n$ and $\mathbb{B}^m$ are the closed unit balls in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, associated with given norms. Given $x \in \mathbb{R}^n$ and $\beta > 0$, the shift $s \in \mathbb{R}^m$ and the collar radius $\kappa \geqslant 0$ are chosen so as to make the modified constraint regions described by (1.3) and (1.4) nonempty. Clearly the choice of s and κ determines the usefulness of these modifications. The modification (1.3) was carefully studied in [3], where the sets C and X were taken to be the sets $\mathbb{R}^{m_1}_- \times \{0\}_{\mathbb{R}^{m_2}}$ and $\mathbb{R}^n$, respectively. Due to the choice of the shift s in [3] we call (1.3) the residual method. In this paper we study the modification (1.4) which we call the collaring method. Our procedure for determining the collar radius κ will ensure that any solution $d \in \mathbb{R}^n$ to the system (1.4) is a descent direction for the distance function

$$\varphi(x) := \operatorname{dist}(g(x) \,|\, C) + \psi(x \,|\, X), \tag{1.5}$$

where $\psi(\,\cdot\,|\,X)$ is the convex indicator of X and is given by

$$\psi(x \,|\, X) := \begin{cases} +\infty & \text{if} \quad x \notin X \\ 0 & \text{if} \quad x \in X. \end{cases}$$

A precise description of how κ is chosen is given in Section 3.

As in [3] the algorithm of this paper generates iterates of the form

$$x_{i+1} := x_i + \tau_i d_i, \tag{1.6}$$

where d_i is the search direction computed as the solution of a convex programming subproblem whose constraint region takes the form (1.4). The subproblem may be solved for various choices of the parameter β in order to determine an appropriate trust region radius. Then, if required, a stepsize is chosen in order to guarantee a sufficient reduction in the value of the exact penalty function

$$P_\alpha(x) := f(x) + \alpha \varphi(x). \tag{1.7}$$

As is usual, a great deal of effort is devoted to an appropriate selection of the penalty parameter α. Our choice of α guarantees that the direction d_i is a descent direction for P_α. The details of this iteration scheme are carefully developed in Sections 3, 4, 5, and 6.

Some of the key ideas for this method can be found in a recent article by Sahba [19]. Sahba considers the case in which $C = \mathbb{R}^{m_1}_- \times \{0\}_{\mathbb{R}^{m_2}}$, $X = \mathbb{R}^n$, both $\mathbb{R}^n$ and $\mathbb{R}^m$ are endowed with the l_∞-norm, and the mapping g satisfies the Mangasarian–Fromowitz constraint qualification at every x in $\mathbb{R}^n$. In this case Sahba's algorithm [19, Algorithm 5.1] differs very little from ours. The differences are, however, essential. In particular, they allow us to derive a valid convergence theory. Although Sahba's article is important for its numerical experiments, the application of second-order correction techniques, and the introduction of a procedure for reducing the value of the penalty parameter, the theoretical claims stated in Section 6 of [19] concerning the convergence properties of his algorithm remain unsubstantiated. The primary causes of the difficulties in [19] are the invalid use of Han [9, Theorem 3] in the proof of [19, Proposition 6.2, part (iii)] and of Robinson [14, Theorem 2.1] in the proof of [19, Proposition 6.4]. The results of Han and Robinson do not apply to the situation that Sahba is considering. The issues that Sahba is able to overlook in this way consume the majority of the effort in this paper.

The paper proceeds as follows. In Section 2 the stationarity conditions that we employ for NLP are precisely described. In Sections 3, 4, and 5 the modified constraint region, the modified subproblem, and the rule for updating the penalty parameter, respectively, are discussed. In Section 6, the local boundedness of the multiplier set associated with our subproblems is addressed. The purpose of this section is to identify the conditions under which the updating rule for the penalty parameter can produce an unbounded sequence of penalty parameters. In Section 7, the algorithm is described, and, in Section 8, its convergence properties are described. We conclude the paper in Section 9 with a discussion of some continuity properties associated with our approach.

The notation that we employ is for the most part standard; however, a partial list is provided for the reader's convenience.

- For $C \subset \mathbb{R}^m$, $\overline{\text{co}}\, C$ is the closed convex hull of C, cl C is the closure of C, int C is the interior of C, rec C is the recession cone of C, ri C is the relative interior of C, bdry C is the boundary of C, $C^0 := \{x^*: \langle x^*, x \rangle \leqslant 1\ \forall x \in C\}$ is the polar of C, $\psi(\cdot\,|\,C)$ is the convex indicator function for C, and $\psi^*(\cdot\,|\,C)$ is the support function for C. If C is moreover convex, then $N(x\,|\,C)$ and $T(x\,|\,C)$ are respectively the normal and tangent cones to C at $x \in C$.

- For $f: \mathbb{R}^n \to \mathbb{R}$, $\partial f(x)$ is the Clarke subdifferential [7] for f at x,

$$\text{argmin}\{f(x): x \in C\} = \{\bar{x} \in C: f(\bar{x}) = \min\{f(x): x \in C\}\},$$

and

$$f'(x; d) := \lim_{\lambda \downarrow 0} \frac{f(x + \lambda d) - f(x)}{\lambda}$$

when this limit exists.

- For $g: \mathbb{R}^n \to \mathbb{R}^m$, $\text{Ran}(g)$ and $\text{Nul}(g)$ are respectively the range and nullity of g, and $g'(x)$ is the Fréchet derivative of g at x.

- The symbol $| \cdot |$ denotes a given norm on $\mathbb{R}^\kappa$ and $\mathbb{B}$ denotes the associated closed unit ball. The symbol $| \cdot |_0$ denotes the associated dual norm, i.e., $|x|_0 := \psi^*(x \,|\, \mathbb{B})$, whose closed unit ball is $\mathbb{B}^0$. For $C \subset \mathbb{R}^n$, define

$$\text{dist}(x \,|\, C) := \inf\{|x - y| : y \in C\}$$

and

$$\text{dist}_0(x \,|\, C) := \inf\{|x - y|_0 : y \in C\}.$$

2. Stationarity Conditions for NLP

As stated in the Introduction, we say that a point $\bar{x} \in \mathbb{R}^n$ is a stationary point for NLP if it is a stationary point for the problem (1.1). By this we mean that $\bar{x}$ satisfies first-order necessary conditions for optimality for both of the problems

$$\text{minimize } \varphi(x), \; x \in \mathbb{R}^n \tag{2.1}$$

and

$$\text{minimize } f(x)$$
$$\text{subject to } g(x) \in C + \varphi(\bar{x}) \, \mathbb{B} \text{ and } x \in X. \tag{2.2}$$

Regarding (2.2), we employ the first-order conditions derived in [4].

PROPOSITION 2.3. [4, Theorem 3.3]. *Let $\bar{x}$ be a local solution to NLP. Then there exist $\lambda^* \geq 0$, $y^* \in N(g(x) \,|\, C)$, and $z^* \in N(x \,|\, X)$ with*

$$\lambda^* + |y^*|_0 + |z^*|_0 = 1$$

such that

$$0 = \lambda^* \nabla f(x^*) + g'(x^*)^T y^* + z^*.$$

Observe that one can easily derive stationary conditions for (2.2) from Proposition 2.3 by simply replacing the set C with the set $C + \varphi(\bar{x})\,\mathbb{B}$. For this reason, we define the sets

$$M_1(x) := \left\{ \begin{pmatrix} y^* \\ z^* \end{pmatrix} \middle| \begin{array}{c} y^* \in N(g(x)|\,C + \varphi(x)\,\mathbb{B}),\, z^* \in N(x\,|\,X) \\ 0 + \nabla f(x) + g'(x)^T y^* + z^* \end{array} \right\}$$

and

$$M_0(x) := \left\{ \begin{pmatrix} y^* \\ z^* \end{pmatrix} \middle| \begin{array}{c} y^* \in N(g(x)|\,C + \varphi(x)\,\mathbb{B}),\, z^* \in N(x\,|\,X) \\ 0 = g'(x)^T y^* + z^* \end{array} \right\}$$

for each point $x \in X$. Then, by Proposition 2.3, if the point $\bar{x}$ is a local solution to (2.2), then either $M_1(\bar{x}) \neq \varnothing$, or $M_0(\bar{x}) \neq \{0\}$, or both. Following Clarke [7, Chap. 6], the sets $M_1(\bar{x})$ and $M_0(\bar{x})$ are called the set of normal multipliers and the set of abnormal multipliers for (2.2) at $\bar{x}$, respectively. These sets are well-defined for each $x \in X$. Moreover, they are related by the equation

$$M_0(x) = \text{rec}(M_1(x)) \tag{2.4}$$

whenever $x \in X$ and $M_1(x) \neq \varnothing$ (see [4, Proposition 3.7]).

We will say that the point $\bar{x}$ satisfies the stationary conditions for (2.1) if $0 \in \partial \varphi(\bar{x})$, where $\partial \varphi(x)$ is the Clarke subdifferential [7] of φ at x.

PROPOSITION 2.5. *For each $x \in X$, the Clarke subdifferential of φ at x is given by*

$$\partial \varphi(x) := g'(x)^T [\partial \, \text{dist}(\cdot|C)(g(x))] + N(x\,|\,X), \tag{2.6}$$

where $\partial \, \text{dist}(\cdot|C)$ is the usual subdifferential of the convex function $\text{dist}(\cdot|C)$ and is given by the expression

$$\partial \, \text{dist}(y\,|\,C) := \begin{cases} \mathbb{B}^0 \cap N(y|C), & \text{if } y \in C \\ (\text{bdry } \mathbb{B}^0) \cap N(y\,|\,C + \text{dist}(y|C)\,\mathbb{B}), & \text{if } y \notin C. \end{cases} \tag{2.7}$$

Proof. Formula (2.6) follows from [5, Sect. 6; 7, Theorem 2.9.8], while formula (2.7) was established in [4, Sect. 2]. ∎

From representations (2.6) and (2.7), observe that when $x \in X$ is such that $g(x) \notin C$ the condition $0 \in \partial \varphi(x)$ is equivalent to the statement $M_0(x) \neq \{0\}$. Thus our stationarity criteria for NLP can be entirely specified by the sets $M_1(x)$ and $M_0(x)$. In order to fix these concepts and to be able to identify the various types of stationary points, we provide the following definition.

DEFINITION 2.8. We say that the point $x \in X$ is

(i) a Kuhn–Tucker point for NLP if $g(x) \in C$ and $M_1(x) \neq \varnothing$,

(ii) a Fritz John point for NLP if $g(x) \in C$ and $M_0(x) \neq \{0\}$,

(iii) an external stationary (ES) point for NLP if $g(x) \notin C$ and $M_0(x) \neq \{0\}$, and

(iv) a strong external stationary (s-ES) point for NLP if x is an ES point and $M_1(x) \neq \varnothing$.

If $x \in X$ is any one of the above types of points, we say that x is a stationary point for NLP.

Remark. Due to the structure of the sets $M_0(x)$ and $M_1(x)$ and the natural relationships between the various types of stationary points, one is tempted to call ES points external Fritz John points, and s–ES points external Kuhn–Tucker points. Nonetheless, we do not use this terminology, since it may lead to unnecessary confusion.

Because of the way we approach the study of NLP, certain elementary facts concerning the distance function $\operatorname{dist}(y|C)$, the support function $\psi^*(y^*|C)$, and normal cones are of fundamental importance. One of these facts has already appeared as formula (2.7). We state these results in the following lemma; for their proof, we refer the reader to [4, 7, 16].

LEMMA 2.9, *Let $K \subset \mathbb{R}^n$ be a nonempty closed convex set.*

(1) *If $x \in K$, then $x^* \in N(x|K)$ if and only if*

$$\langle x^*, x \rangle = \psi^*(x^* | K).$$

(2) *For $x \in \mathbb{R}^n$ we have*

$$\operatorname{dist}(x|K) = \sup\{ \langle x^*, x \rangle - \psi^*(x^*|K) \,|\, |x^*|_0 \leqslant 1 \}$$

and

$$\partial \operatorname{dist}(x|K) = \begin{cases} \mathbb{B}^0 \cap N(x|K), & \text{if } x \in K \\ (\text{bdry } \mathbb{B}^0) \cap N(x|K + \operatorname{dist}(x|K)), & \text{if } x \in K. \end{cases}$$

(3) *For any $x \in \mathbb{R}^n$ and $x^* \in \mathbb{R}^n$, it is always the case that*

$$\langle x^*, x \rangle - \psi^*(x^*|K) \leqslant |x^*|_0 \operatorname{dist}(x|K).$$

3. THE MODIFIED CONSTRAINT REGION

As stated in the Introduction, we consider iterates of the form (1.6) where the search direction d_i is the solution to a certain convex programming subproblem whose constraint region is of the form (1.4). The parameter κ in (1.4) is identified by the mapping

$$\kappa: X \times \mathbb{R}_+ \times [0, 1] \to \mathbb{R}$$

defined by the relation

$$\kappa(x, \rho, \lambda) := (1 - \lambda)\, \varphi(x) + \lambda \min \{ \text{dist}(g(x) + g'(x)\, d | C): d \in [X - x] \cap \rho \mathbb{B} \} \tag{3.1}$$

for all $(x, \rho, \lambda) \in X \times \mathbb{R}_+ \times [0, 1]$. Observe that $\kappa(x, 0, \lambda) = \varphi(x)$ for all $(x, \lambda) \in X \times [0, 1]$. Also, if we were to set $\kappa := \kappa(x, \rho, 1)$ for some $(x, \rho) \in X \times \mathbb{R}_+$, then the system (1.4) would be consistent for all choices of $\beta \geqslant \rho$. However, as we will see, this choice of κ is unsatisfactory. In particular, it leads to problems concerning the existence of Kuhn–Tucker multipliers for the convex programming subproblems that we employ (see Theorem 4.4). It is for this reason that we require the parameter λ in the definition of the mapping κ.

An essential feature of the mapping κ is its Lipschitz continuity.

PROPOSITION 3.2.   *The mapping κ is locally Lipschitz on $X \times \mathbb{R}_+ \times [0, 1]$.*

Proof. It is sufficient to show that the mapping $\kappa(\cdot, \cdot, 1): X \times \mathbb{R}_+ \to \mathbb{R}_+$ is locally Lipschitz on $X \times \mathbb{R}_+$ since $\varphi: X \to \mathbb{R}_+$ is locally Lipschitzian on X. First we show that $\kappa(\cdot, \cdot, 1)$ is locally Lipschitz on $X \times (0, \infty)$. By Rockafellar [17, Theorem 3.1], $\kappa(\cdot, \cdot, 1)$ is locally Lipschitz on $X \times (0, \infty)$ if

$$N(x + d | X) \cap (-N(d | \rho \mathbb{B})) = \{0\}$$

for every $\rho > 0$, $x \in X$, and

$$d \in \text{argmin} \{ \text{dist}(g(x) + g'(x)\, d | C): d \in \rho \} =: S(x, \rho).$$

Let $x \in X$, $\beta > 0$, $d \in S(x, \rho)$, and

$$z^* \in N(x + d | X) \cap (-N(d | \rho \mathbb{B})).$$

Then, by Lemma 2.9,

$$\langle z^*, x + d \rangle = \psi^*(z^* | X)$$

and

$$- \langle z^*, d \rangle = \psi^*(z^* \,|\, \rho \mathbb{B}) = \rho |z^*|_0.$$

Therefore,

$$\rho |z^*|_0 = \langle z^*, x \rangle - \psi^*(z^* \,|\, X)$$
$$\leqslant |z^*|_0 \operatorname{dist}(x \,|\, X)$$
$$= 0.$$

Hence $z^* = 0$ since $\rho > 0$.

Finally, it is clear that $\kappa(\cdot, \cdot, 1)$ is locally Lipschitz at points of the form $(x, 0)$ where $x \in X$, since, for $(\bar{x}, \rho) \in X \times \mathbb{R}_+$, we have

$$|\kappa(x, 0, 1) - \kappa(\bar{x}, \rho, 1)| \leqslant |g(x) - g(\bar{x})| + |g'(\bar{x})| \, \beta. \quad \blacksquare$$

Remark. Examples illustrating how the mapping κ can be evaluated in various situations are provided in [3].

The constraint regions that will be used in our modified subproblems are described by the multifunction $D: X \times \operatorname{cl} T \times [0, 1] \Rightarrow \mathbb{R}^n$ defined as

$$D(x, r, \lambda) := \left\{ d \in \mathbb{R}^n \,\middle|\, \begin{matrix} g(x) + g'(x) \, d \in C + \kappa(x, \rho, \lambda) \, \mathbb{R} \\ d \in [X - x] \cap \beta \mathbb{B} \end{matrix} \right\}$$
$$= [g'(x)]^{-1} [C + \kappa(x, \rho, \lambda) \, \mathbb{B} - g(x)] \cap [X - x] \cap \beta \mathbb{B},$$

where $r =: (\rho, \beta)$ and T is the parameter set

$$T := \{ r := (\rho, \beta) \,|\, 0 < \rho \leqslant \beta \}.$$

Here the multifunction $[g'(x)]^{-1}: \mathbb{R}^m \Rightarrow \mathbb{R}^n$ is the multivalued inverse of $g'(x)$, i.e.,

$$[g'(x)]^{-1} y := \{ z \in \mathbb{R}^n : g'(x) \, z = y \}.$$

Since the mapping κ is locally Lipschitz on $X \times \mathbb{R}_+ \times [0, 1]$ it is trivial to verify that the multifunction D is upper semi-continuous on $X \times \operatorname{cl} T \times [0, 1]$, that is, the set

$$\operatorname{graph} D := \{ (x, r, \lambda, d) : (x, r, \lambda) \in X \times \operatorname{cl} T \times [0, 1], d \in D(x, r, \lambda) \}$$

is closed. Since this fact is used in later sections, we record it now in the following proposition.

PROPOSITION 3.3. *The multifunction D is upper semi-continuous on $X \times \operatorname{cl} T \times [0, 1]$.*

We close this section by showing that for each $(x, r, \lambda) \in T \times \operatorname{Cl} T \times$ $[0, 1]$ the set $D(x, r, \lambda)$ is a subset of the set of descent directions for φ at x.

PROPOSITION 3.4. *If* $d \in D(x, r, \lambda)$ *for some* $(x, r, \lambda) \in X \times \operatorname{cl} T \times [0, 1]$, *then*

$$\varphi'(x; d) \leqslant \operatorname{dist}(g(x) + g'(x) \, d | C) - \varphi(x)$$

$$\leqslant \kappa(x, \rho, \lambda) - \varphi(x). \tag{3.5}$$

Moreover, if λ *and* ρ *are nonzero, then* $\kappa(x, \rho, \lambda) = \varphi(x)$ *if and only if* $0 \in \partial \varphi(x)$.

Proof. Since $x \in X$ and $x + d \in X$, we have $x + \tau d \in X$ for $\tau \in [0, 1]$. Consequently, $\varphi(x + \tau d) = \operatorname{dist}(g(x + \tau d) | C)$ for $\tau \in [0, 1]$, and so

$$\varphi'(x; d) = (\operatorname{dist}(g(\cdot) | C))'(x; d).$$

Hence, by [5, Sect. 2],

$$\varphi'(x; d) \leqslant \operatorname{dist}(g(x) + g'(x) \, d | C) - \operatorname{dist}(g(x) | C).$$

Now since $\operatorname{dist}(z | C) \leqslant \kappa$ if and only if $z \in C + \kappa \mathbb{B}$, inequality (3.5) follows from the preceding inequality and the definition of $D(x, r, \lambda)$.

Next suppose that λ and ρ are nonzero so that $0 < \rho \leqslant \beta$ and $0 < \lambda \leqslant 1$. Then $\kappa(x, \rho, \lambda) = \varphi(x)$ if and only if $d = 0$ solves the convex program $\min\{\operatorname{dist}(g(x) + g'(x) \, d | C) : d \in [X - x]\}$. The proof is concluded by observing that the Kuhn–Tucker conditions for this convex program, when the solution is known to be $d = 0$, are equivalent to the statement $0 \in \partial \varphi(x)$. ∎

4. THE MODIFIED SUBPROBLEMS

The modification to the standard Wilson–Han–Powell convex programming subproblem (1.2) that we study is given by

$$Q(x, r, \lambda, H): \min \nabla f(x)^T d + \tfrac{1}{2} d^T H d$$

$$\text{subject to } d \in [X - x] \cap \beta \mathbb{B}$$

$$g(x) + g'(x) \, d \in C + \kappa(x, \rho, \lambda) \, \mathbb{B},$$

where $(x, r, \lambda, H) \in \operatorname{cl} \Omega$ with

$$\Omega := X \times T \times (0, 1] \times \Gamma.$$

and

$$\Gamma := \{H \in \mathbb{R}^{n \times n}: H \text{ is symmetric and positive definite}\}.$$

For each $\omega = (x, r, \lambda, H) \in \text{cl } \Omega$ the feasible region for $Q(\omega)$, namely $D(x, r, \lambda)$, is a non-empty compact convex subset of the set of directions of nonascent for φ at x. Hence the convex program $Q(\omega)$ is guaranteed to have a unique solution whenever $\omega \in \Omega$. Henceforth we denote this solution by $d(\omega)$ so that d is a mapping from Ω to $\mathbb{R}^n$.

In the analysis that follows, an understanding of the continuity and stability properties of the convex programs $Q(\omega)$ is essential. In particular, the multiplier set associated with the solution to $Q(\omega)$ plays a key role. Thus we must be able to characterize this multiplier set for various choices of $\omega \in \text{cl } \Omega$. In preparation for this study we have the following lemmas.

LEMMA 4.1. *Let S be a subspace of $\mathbb{R}^m$ and let $z \in U$, where $U \subset \mathbb{R}^m$ is a nonempty closed convex set. If $\text{ri}(U) \cap S = \varnothing$ while $U \cap S \neq \varnothing$, then for each $z \in U \cap S$*

$$N(z|U) \cap S^{\perp} \neq \{0\}.$$

Proof. Let $z \in S \cap U$. If $N(z|U) \cap S^{\perp} = \{0\}$, *then, by* [16, Corollary 16.4.2], we have $S + T(z|U) = \mathbb{R}^m$. It is straightforward to show that $S + T(z|U) = R^m$ if and only if

$$S + \bigcup_{\lambda > 0} \lambda^{-1}(\text{ri } U - z) = \mathbb{R}^m.$$

Hence there is a $\lambda > 0$, $y_1 \in S$, and $y_2 \in \text{ri } U$ such that $y_2 + \lambda^{-1}(y_2 - z) = 0$. But then $y_2 = z - \lambda y_1 \in S$, contradicting the hypothesis that $\text{ri}(U) \cap S = \varnothing$. ∎

LEMMA 4.2. *Let $x \in X$, $0 < \rho$, $\lambda \in (0, 1)$, and let $\kappa: \mathbb{R}^n \times \mathbb{R}_+ \times [0, 1] \to \mathbb{R}_+$ be as defined in (3.1).*

(1) *If $\varphi(x) > 0$ and $\kappa(x, \rho, \lambda) < \varphi(x)$, then for all $\varepsilon > 0$ there is a $d \in (\rho + \varepsilon)\mathbb{B}$ such that*

$$g(x) + g'(x) d \in \text{int}[C + \kappa(x, \rho, \lambda) \mathbb{B}]$$

and

$$x + d \in \text{ri}(X).$$

(2) *If $\varphi(x) > 0$ and $\kappa(x, \rho, \lambda) = \varphi(x)$, then x is an ES point for NLP.*

(3)   *If $\varphi(x) = 0$, then either x is a Fritz John point for NLP or for all $\varepsilon > 0$ there is a $d \in \varepsilon \mathbb{B}$ such that*

$$g(x) + g'(x)\,d \in \text{ri } C$$

and

$$x + d \in \text{ri } X.$$

Proof.   By definition there is always a $d \in \rho \mathbb{B}$ such that $x + d \in X$ and $g(x) + g'(x)\,d \in C + \kappa(x, \rho, \lambda)\,\mathbb{B}$.

(1)   Since $\kappa(x, \rho, \lambda) \neq \varphi(x)$ and $0 < \lambda < 1$, there is a $d \in \rho \mathbb{B}$ such that

$$g(x) + g'(x)\,d \in \text{int}[C + \kappa(x, \rho, \lambda)\,\mathbb{B}]$$

and

$$x + d \in X.$$

The result now follows by continuity.

(2)   By Proposition 3.4, $0 \in \partial \varphi(x)$. Also, by hypothesis, $\varphi(x) \neq 0$. Thus x is an ES point for NLP by Definition 2.8.

(3)   Since $\kappa(x, \rho, \lambda) = 0$, we have $g(x) \in C$ and $x \in X$. Hence, if there is a vector $d \in \mathbb{R}^n$ such that $g(x) + g'(x)\,d \in \text{ri } C$ and $x + d \in \text{ri } X$, then $g(x) + g'(x)(\tau d) \in \text{ri } C$ and $x + \tau d \in \text{ri } X$ for all $\tau \in (0, 1]$, thereby establishing the result. Thus we suppose that no such vector d exists and show that x is a Fritz John point for NLP. If no such d exists, then

$$\text{Ran}\begin{bmatrix} g'(x) \\ I \end{bmatrix} \cap \text{ri}[(C - g(x)) \times (X - x)] = \varnothing$$

while

$$\text{Ran}\begin{bmatrix} g'(x) \\ I \end{bmatrix} \cap [(C - g(x)) \times (X - x)] \neq \varnothing.$$

Therefore, by Lemma 4.1,

$$N\left(\begin{pmatrix} g(x) \\ x \end{pmatrix} \middle| C \times X\right) \cap \text{Nul}[g'(x)^T, I] \neq \{0\}.$$

But this set is exactly $M_0(x)$. Hence the result is established.   ∎

In the next two results, we describe the types of multiplier sets one can obtain for the convex programs $Q(\omega)$.

THEOREM 4.3. *Let $\omega = (x, r, \lambda, H) \in \Omega$ be such that $\varphi(x) = 0$. If $d \in \mathbb{R}^n$ is a Fritz John point for $Q(\omega)$, then x is a Fritz John point for NLP.*

Proof. Since d is a Fritz John point for $Q(\omega)$, there are vectors $y^* \in N(g(x) + g'(x) d \mid C)$, $z_1^* \in N(d \mid X - x)$, and $x_2^* \in N(d \mid \beta \mathbb{B})$ such that

$$0 = g'(x)^T y^* + z_1^* + z_2^*,$$

and

$$|y^*|_0 + |z_1^*|_0 + |z_2^*|_0 = 1.$$

Hence, by Lemma 2.9,

$$\langle y^*, g(x) + g'(x) d \rangle = \psi^*(y^* \mid C)$$

and

$$-\langle y^*, g'(x) d \rangle = \langle z_1^* + z_2^*, d \rangle.$$

Therefore, again by Lemma 2.9,

$$\begin{aligned}
\langle z_1^* + z_2^*, d \rangle &= \langle y^*, g(x) \rangle - \psi^*(y^* \mid C) \\
&\leqslant |y^*|_0 \operatorname{dist}(g(x) \mid C) \\
&= 0.
\end{aligned}$$

Moreover, by Lemma 2.9,

$$\begin{aligned}
\langle z_1^* + z_2^*, d \rangle &= \langle z_1^*, d \rangle + \langle z_2^*, d \rangle \\
&= \psi^*(z_1^* \mid X - x) + \psi^*(z_2^* \mid \beta \mathbb{B}) \\
&= \psi^*(z_1^* \mid X) - \langle z_1^*, x \rangle + \beta |z_2^*|_0 \\
&\geqslant -|z_2^*|_0 \operatorname{dist}(x \mid X) + \beta |z_2^*|_0 \\
&= \beta |z_2^*|_0.
\end{aligned}$$

Hence,

$$\begin{aligned}
0 \geqslant \langle y^*, g(x) \rangle - \psi^*(y^* \mid C) \\
= \langle z_1^* + z_2^*, d \rangle \\
\geqslant \psi^*(z_1^*, \mid X) - \langle z_1^*, x \rangle + \beta |z_2^*|_0 \\
\geqslant \beta |z_2^*|_0.
\end{aligned}$$

Consequently, $z_2^* = 0$, and so

$$\langle z_1^*, x \rangle = \psi^*(z^* \mid X)$$

and

$$\langle y^*, g(x) \rangle = \psi^*(y^*|C).$$

Therefore, by Lemma 2.9, $z_1^* \in N(x|X)$, $y^* \in N(g(x)|C)$, and $0 = '(x)^T y^* + z_1^*$.

Hence $(y^*, z_1^*) \in M_0(x)$ with $|y^*|_0 + |z_1^*|_0 = 1$, so that x is a Fritz John point for NLP. ∎

Remarks. A result similar to Lemma 4.3 for the case $\varphi(x) > 0$ is, in general, not available. The absence of such a result will be the cause of some technical difficulty in later sections. Specifically, it forces us to onsider the two cases $\varphi(x) = 0$ and $\varphi(x) > 0$ separately.

In the next theorem, one should pay particular attention to the essential role played by the parameters (ρ, β, λ) in establishing the existence of a Kuhn–Tucker solution to $Q(x, r, \lambda, H)$.

THEOREM 4.4. *Let* $\omega = (x, r, \lambda, H) \in \Omega$.

(1) *If the sets and X are polyhedral, and the norms chosen for* $\mathbb{R}^n$ *and* $\mathbb{R}^m$ *are also polyhedral, then* $\bar{d}(\omega)$ *is a Kuhn–Tucker point for* $Q(\omega)$.

(2) *If* $(r, \lambda) \in \text{int}(T \times [0, 1])$, *then either*

(a) *x is an ES point for NLP, or*

(b) *x is a Fritz John point for NLP, or*

(c) $\bar{d}(\omega)$ *is a Kuhn–Tucker point for* $Q(\omega)$, *or*

(d) *some combination of* (a), (b), *and* (c) *holds.*

Proof. Since $\bar{d}(\omega)$ solves $Q(\omega)$ we have

$$-(\nabla f(x) + H\bar{d}(\omega)) \in N(\bar{d}(\omega)|D(x, r, \lambda)).$$

Thus, if we can show that

$$N(\bar{d}(\omega)|D(x, r, \lambda)) = g'(x)^T N(g(x) + g'(x)\bar{d}(\omega)|C + \kappa(x, \rho, \lambda)\mathbb{B})$$
$$+ N(x + \bar{d}(\omega)|X) + N(\bar{d}(\omega)|\beta\mathbb{B}), \qquad (4.5)$$

then $\bar{d}(\omega)$ is a Kuhn–Tucker point for $Q(\omega)$. Recall that

$$D(x, r, \lambda) := [g'(x)]^{-1}[C + \kappa(x, \rho, \lambda)\mathbb{B} - g(x)] \cap [X - x] \cap [\beta\mathbb{B}].$$

Hence, by [16, Corollary 23.8.1],

$$N(\bar{d}(\omega)|D(x, r, \lambda)) = N(\bar{d}(\omega)|[g'(x)]^{-1}[C + \kappa(x, \rho, \lambda)\mathbb{B} - g(x)])$$
$$+ N(x + \bar{d}(\omega)|X) + N(\bar{d}(\omega)|\beta\mathbb{B}) \qquad (4.6)$$

if either the intersection

$$ri([g'(x)]^{-1}[C + \kappa(x, \rho, \lambda)\,\mathbb{B} - g(x)]) \cap ri[X - x] \cap ri[\beta\mathbb{B}] \quad (4.7)$$

is nonempty, or the sets C and X are polyhedral, the norms on both $\mathbb{R}^m$ and $\mathbb{R}^n$ are polyhedral, and the intersection

$$[g'(x)]^{-1}(C + \kappa(x, \rho, \lambda)\,\mathbb{B}) \cap [X - x] \cap [\beta\mathbb{B}]$$

is nonempty. Hence (4.6) holds under the hypotheses of part (1). Also, since in part (2) $(r, \lambda) \in \text{int}(T \times (0, 1])$, Lemma 4.2 implies that either (4.7) holds, or one of (a) and (b) is true.

It remains to show that

$$N(\bar{d}(\omega) | [g'(x)]^{-1}[C + \kappa(x, \rho, \lambda)\,\mathbb{B} - g(x)])$$
$$= g'(x)^T N(g'(x)\,\bar{d}(\omega) | C + \kappa(x, \rho, \lambda)\,\mathbb{B} - g(x)).$$

Observe that for all $d \in \mathbb{R}^n$,

$$\psi^*(d | [g'(x)]^{-1}[C + \kappa(x, \rho, \lambda)\,\mathbb{B} - g(x)])$$
$$= \psi^*(g'(x)\,d | C + \kappa(x, \rho, \lambda)\,\mathbb{B} - g(x)).$$

Consequently, for $d \in D(x, r, \lambda)$,

$$N(d | [g'(x)]^{-1}[C + \kappa(x, \rho, \lambda)\,\mathbb{B} - g(x)])$$
$$= \partial\psi(d | [g'(x)]^{-1}[C + \kappa(x, \rho, \lambda)\,\mathbb{B} - g(x)])$$
$$= \partial\psi(g'(x)(\cdot) | C + \kappa(x, \rho, \lambda)\,\mathbb{B} - g(x))(d).$$

By [16, Theorem 23.9], the last term in this expression equals

$$g'(x)^T N(g'(x)\,d | C + \kappa(x, \rho, \lambda)\,\mathbb{B} - g(x))$$

if either the intersection

$$\text{Ran}[g'(x)] \cap ri[C + \kappa(x, \rho, \lambda)\,\mathbb{B} - g(x)] \quad (4.8)$$

is nonempty, or the set C is polyhedral, the norm on $\mathbb{R}^m$ is polyhedral, and the intersection

$$\text{Ran}[g'(x)] \cap [C + \kappa(x, \rho, \lambda)\,\mathbb{B} - g(x)]$$

is nonempty. These conditions are satisfied bpart (1) of the theorem, and are also satisfied by part (2) as long as $\kappa(x, \rho, \lambda) \neq \varphi(x)$. If $\kappa(x, \rho, \lambda) = \varphi(x)$ and $\varphi(x) > 0$, then x is an ES point for NLP, as was noted in part (2) of Lemma 4.2. On the other hand, if $\varphi(x) = 0$, then, by part (3) of Lemma 4.2,

either x is a Fritz John point for NLP or the intersection (4.8) is non-empty, whereby the result is established. ∎

COROLLARY 4.9. *Let* $\omega = (x, r, \lambda, H) \in \Omega$ *be such that* $\lambda \neq 1$.

   (1) *If the hypotheses of Theorem 4.4, part (1), are satisfied and* $d(\omega) = 0$, *then x is either a Kuhn-Tucker point for NLP or an s–ES point for NLP.*

   (2) *If the hypotheses of Theorem 4.4, part (2), are satisfied and* $d(\omega) = 0$, *then x is a stationary point for NLP.*

Proof. If $d(\omega) = 0$, then $g(x) \in C + \kappa(x, \rho, \lambda)\, \mathbb{B}$ and $x \in X$. Since $\lambda \in (0, 1)$, this implies that $\kappa(x, \rho, \lambda) = \varphi(x)$. Hence $0 \in \partial\varphi(x)$.

   (1) By part (1) of Theorem 4.4, $d = 0$ is a Kuhn–Tucker point for $Q(\omega)$, that is,

$$-\nabla f(x) \in g'(x)^T N(g(x) | C + \varphi(x)\, \mathbb{B}) + N(x | X),$$

since $0 \in \operatorname{int} \beta\mathbb{B}$ and $\varphi(x) = \kappa(x, \rho, \lambda)$. Hence x is a Kuhn–Tucker point for NLP if $\varphi(x) = 0$, and x is an s–ES point for NLP if $\varphi(x) > 0$.

   (2) From the hypotheses, we know that some combination of (a), (b), and (c) in part (2) of Theorem 4.4 must hold. If $d = 0$ is a Kuhn–Tucker point for $Q(\omega)$, then, as in part (1), x is either a Kuhn–Tucker point or an s–ES point for NLP. ∎

5. ADJUSTING THE PENALTY PARAMETER

   The convergence theory that we develop is similar to that given in [2, 3]. At each iteration one determines the parameters $r_i \in T$, $\lambda_i \in (0, 1]$, $h_i \in \Gamma$, and $\tau_i > 0$ so that the update,

$$x_{i+1} := x_i + \tau_i d_i, \tag{5.1}$$

where $d_i = d(\omega_i)$ is the solution to $Q(\omega_i)$, induces a sufficient decrease in the penalty function

$$P_{\alpha_{i+1}}(x) := f(x) + \alpha_{i+1} \varphi(x)$$

for an "appropriate" choice of the penalty parameter α_{i+1}. In this context, the key to the analysis is the rule for adjusting the penalty parameter α. The rule that we consider is based upon the insight obtained in the following lemma.

LEMMA 5.2. *Let* $\omega = (x, r, \lambda, H) \in \Omega$ *and* $d \in D(x, r, \lambda)$. *Then*

$$P'_\alpha(x; d) \leqslant \nabla f(x)^T d + \alpha[\operatorname{dist}(g(x) + g'(x) d | C) - \operatorname{dist}(g(x) | C)]$$
$$\leqslant \nabla f(x)^T d + \alpha[\kappa(x, \rho, \lambda) - \varphi(x)]. \tag{5.3}$$

Moreover, if $\bar{d}(\omega)$ *is a Kuhn–Tucker point for* $Q(\omega)$, *then*

$$\nabla f(x)^T \bar{d}(\omega) \leqslant -\bar{d}(\omega)^T H \bar{d}(\omega) - |y^*|_0 [\kappa(x, \rho, \lambda) - \varphi(x)], \tag{5.4}$$

where y^* *is any element from the set of multipliers*

$$M_C(\omega) := \left\{ y^* \left| \begin{array}{c} y^* \in N(g(x) + g'(x) d | C + \kappa(x, \rho, \lambda) \mathbb{B}), \\ 0 \in \nabla f(x) + Hd + g'(x)^T y^* + N(d | [X - x] \cap \beta\mathbb{B}) \\ \text{for some } d \in D(x, r, \lambda) \end{array} \right. \right\}.$$

Remark. The multifunction $M_C(\omega)$ defined above is well-defined on all of cl Ω even though it may happen that $M_C(\omega)$ is empty for a particular choice of $\omega = (x, r, \lambda, H) \in$ cl Ω. Also, observe that if $H \in \Gamma$, then the vector $d \in D(x, r, \lambda)$ in the definition of $M_C(\omega)$ is the same for each element of $M_C(\omega)$, by [16, Corollary 28.1.1]. However, if $H \in$ (cl Γ)\\$\Gamma$, this may not be the case.

Proof. The inequalities (5.3) follow directly from Proposition 3.4. For (5.4), note that $M_C(\omega)$ is nonempty since $\bar{d}(\omega)$ is assumed to be a Kuhn–Tucker point for $Q(\omega)$. Since

$$N(\bar{d}(\omega) | [X - x] \cap \beta\mathbb{B}) = N(x + \bar{d}(\omega) | X) + N(\bar{d}(\omega) | \beta\mathbb{B})$$

[16, Corollary 23.8.1], for any $y^* \in M_C(\omega)$ there exist $x^* \in N(x + \bar{d}(\omega) | X)$ and $z^* \in N(\bar{d}(\omega) | \beta\mathbb{B})$ such that

$$\nabla f(x) = -H \bar{d}(\omega) - g'(x)^T y^* - x^* - z^*.$$

Inequality (5.3) will be established once the inequalities
  (a) $- \langle y^*, g'(x) \bar{d}(\omega) \rangle \leqslant -|y^*|_0 [\kappa(x, \rho, \lambda) - \varphi(x)]$,
  (b) $- \langle x^*, \bar{d}(\omega) \rangle \leqslant 0$, and
  (c) $\langle z^*, \bar{d}(\omega) \rangle = \beta |z^*|_0$,

have been verified. To see (a), observe that

$$- \langle y^*, g'(x) \bar{d}(\omega) \rangle = \langle y^*, g(x) \rangle - \langle y^*, g(x) + g'(x) \bar{d}(\omega) \rangle$$
$$= \langle y^*, g(x) \rangle - \psi^*(y^* | C + \kappa(x, \rho, \lambda) \mathbb{B})$$
$$\leqslant |y^*|_0 \operatorname{dist}(g(x) | C - \kappa(x, \rho, \lambda) \mathbb{B})$$
$$\leqslant |y^*|_0 [\operatorname{dist}(g(x) | C) - \kappa(x, \rho, \lambda)]$$
$$= -|y^*|_0 [\kappa(x, \rho, \lambda) - \varphi(x)]$$

by Lemma 2.9. For (b), note that

$$
\begin{aligned}
-\langle x^*, \tilde{d}(\omega)\rangle &= \langle x^*, x\rangle - \langle x^*, x + \tilde{d}(\omega)\rangle \\
&= \langle x^*, x\rangle - \psi^*(x^*|X) \\
&\leqslant |x^*|_0 \,\operatorname{dist}(x|X) \\
&= 0
\end{aligned}
$$

again by Lemma 2.9. Finally, for (c), observe that

$$
\langle z^*, \tilde{d}\rangle = \psi^*(z^*|\beta\mathbb{B}) = \beta|z^*|_0. \quad \blacksquare
$$

At each of the iterations described by (5.1), we wish to adjust the parameter α_i to α_{i+1} so as to guarantee the validity of the inequality

$$
P'_{\alpha_{i+1}}(x_i; d_i) \leqslant -d_i^T H_i d_i. \tag{5.5}
$$

According to Lemma 5.2, inequality (5.5) is satisfied whenever

$$
\nabla f(x_i)^T d_i + \alpha_{i+1}[\kappa(x_i, \rho_i, \lambda_i) - \varphi(x_i)] \leqslant -d_i^T H_i d_i.
$$

The adjustment rule that we employ is based upon this last inequality and is defined as a mapping $\bar{\alpha}: \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$.

5.6. THE PARAMETER ADJUSTMENT RULE. Given $\omega = (x, r, \lambda, H) \in \Omega$ with $r =: (\rho, \beta)$ and $\alpha > 0$, define $\bar{\alpha}(\omega, \alpha)$ as follows:

    (i)   if

$$
\kappa(x, \rho, \lambda) = \varphi(x) \text{ and } \nabla f(x)^T \tilde{d}(\omega) > -\tilde{d}(\omega)^T H \tilde{d}(\omega), \tag{5.7}
$$

set $\bar{\alpha}(\omega, \alpha) = 0$;

    (ii) if

$$
\nabla f(x)^T \tilde{d}(\omega) + \alpha[\kappa(x, \rho, \lambda) - \varphi(x)] \leqslant -\tilde{d}(\omega)^T H \tilde{d}(\omega), \tag{5.8}
$$

set $\bar{\alpha}(\omega, \alpha) = \alpha$;

    (iii) if neither (5.7) nor (5.8) holds, set

$$
\bar{\alpha}(\omega, \alpha) := \max\left\{\frac{\nabla f(x)^T \tilde{d}(\omega) + \tilde{d}(\omega)^T H \tilde{d}(\omega)}{\varphi(x) - \kappa(x, \rho, \lambda)}, 2\alpha\right\}.
$$

Observe that, as long as (5.7) does not occur, the adjustment rule described in (5.6) guarantees the inequalities (5.8) and $\alpha \leqslant \bar{\alpha}(\omega, \alpha)$. The exceptional case, when (5.7) occurs, provides a stopping rule for our algorithm. Note that if (5.7) occurs, then, by Lemma 5.2, $\tilde{d}(\omega)$ cannot be

a Kuhn–Tucker point for $Q(\omega)$. Hence, if $(r, \lambda) \in \operatorname{int}(T \times [0, 1])$, then, by Theorem 4.4, the point x is necessarily either a Fritz John point or an ES point for NLP. Since these are stationary points for NLP, it is appropriate that the algorithm should terminate.

In the convergence analysis of Section 8, the local behavior of the mapping $\bar{\alpha} \colon \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$ plays a key role. Specifically, we are interested in characterizing those points $\omega \in \Omega$ at which the mapping $\bar{\alpha}$ is locally unbounded. In this regard, it is clear that we need only analyze the quotient

$$\frac{\nabla f(x)^T \bar{d}(\omega) + \bar{d}(\omega)^T H \bar{d}(\omega)}{\varphi(x) - \kappa(x, \rho, \lambda)}. \tag{5.9}$$

From inequality (5.4) in Lemma 5.2, we obtain the inequality

$$\frac{\nabla f(x)^T \bar{d}(\omega) + \bar{d}(\omega)^T H \bar{d}(\omega)}{\varphi(x) - \kappa(x, \rho, \lambda)} \leqslant \operatorname{dist}_0(0 \,|\, M_C(\omega)). \tag{5.10}$$

Thus we can study the local boundedness of (5.9) by studying the local boundedness of $\operatorname{dist}_0(0 \,|\, M_C(\omega))$.

6. The Local Boundedness of $\operatorname{dist}_0(0 \,|\, M_C(\omega))$

In order to understand the distance function $\operatorname{dist}_0(0 \,|\, M_C(\omega))$ one must first carefully study the multifunction M_C. For this reason we introduce the multifunctions

$$QM_1(x, r, \lambda, H) := \left\{ (y^*, z^*) \left| \begin{array}{c} y^* \in N(g(x) + g'(x)\, d \,|\, C + \kappa(x, \rho, \lambda)\, \mathbb{B}) \\ z^* \in N(d \,|\, (X - x) \cap \beta \mathbb{B}) \\ 0 = \nabla f(x) + Hd + g'(x)^T y^* + z^* \\ \text{for some } d \in D(x, r, \lambda) \end{array} \right. \right\}$$

and

$$QM_0(x, r, \lambda, H) := \left\{ (y^*, z^*) \left| \begin{array}{c} y^* \in N(g(x) + g'(x)\, d \,|\, C + \kappa(x, \rho, \lambda)\, \mathbb{B}) \\ z^* \in N(d \,|\, (X - x) \cap \beta \mathbb{B}) \\ 0 = g'(x)^T y^* + z^* \\ \text{for some } d \in D(x, r, \lambda) \end{array} \right. \right\}$$

for each $(x, r, \lambda, H) \in \Omega$. These multifunctions describe the normal and abnormal multipliers for $Q(\omega)$, respectively. Also,

$$M_C(\omega) = \{ y^* \colon (y^*, z^*) \in QM_1(\omega) \text{ for some } z^* \}.$$

As with the multifunction M_C, one should note that $QM_1(\omega)$ may be empty for some values of $\omega \in \text{cl}\, \Omega$ (see Theorem 4.4). Moreover, if $H \in \Gamma$, then the vector $d \in D(x, r, \lambda)$ in the definition of $QM_1(x, r, \lambda, H)$ and $QM_0(x, r, \lambda, H)$ is just $\bar{d}(\omega)$, where $\omega := (x, r, \lambda, H)$, by [16, Corollary 28.1.1]. However, if $H \in (\text{cl}\, \Gamma) \backslash \Gamma$, there may be different vectors $d \in D(x, r, \lambda)$ for different multipliers (y^*, z^*) in either $QM_1(\omega)$ or $QM_0(\omega)$. Also note that

$$\text{rec}(QM_1(\omega)) = QM_0(\omega)$$

by [4, Proposition 3.7], whenever $QM_1(\omega) \neq \varnothing$.

The most fundamental property of these multifunctions is their upper semi-continuity on $\text{cl}\, \Omega$.

PROPOSITION 6.1. *The multifunctions QM_1, QM_0, and M_C are all upper semi-continuous on $\text{cl}\, \Omega$, i.e., they all have closed graphs over $\text{cl}\, \Omega$.*

Proof. Since the upper semi-continuity of M_C follows from the upper semi-continuity of QM_1, and the proof of the upper semi-continuity of QM_0 is almost identical to that for QM_1, we only establish the upper semi-continuity of QM_1 on $\text{cl}\, \Omega$.

Let $\{(\omega_i, y_i^*, z_i^*)\} \subset \text{cl}\, \Omega \times \mathbb{R}^m \times \mathbb{R}^n$ be a sequence such that $(\omega_i, y_i^*, z_i^*) \to (\omega, y^*, z^*)$ and $(y_i^*, z_i^*) \in QM_1(\omega_i)$ for each $i = 1, 2, \dots$ We need to show that $(y^*, z^*) \in QM_1(\omega)$. First note that, since the sequence $\{\bar{d}(\omega_i)\}$ is bounded, we may assume, with no loss of generality, that there is a $d \in \mathbb{R}^n$ such that $\bar{d}(\omega_i) \to d$ with $d \in D(x, r, \lambda)$, by Proposition 3.3, where $\omega_i =: (x_i, r_i, \lambda_i, H_i)$ and $\omega =: (x, r, \lambda, H)$. Now, let $y \in C + \kappa(x, \rho, \lambda)\, \mathbb{B}$, and $z \in \beta \mathbb{B}$, where $r =: (\rho, \beta)$. Then, since φ is continuous on X and, by Proposition 3.2, κ is continuous on $X \times \mathbb{R}_+ \times [0, 1]$, there are sequences $\{y_i\} \subset \mathbb{R}^m$ and $\{z_i\} \subset \mathbb{R}^n$ such that $y_i \in C + \kappa(x_i, \rho_i, \lambda_i)\, \mathbb{B}$ and $z_i \in \beta_i \mathbb{B}$, for all $i = 1, 2, \dots$, and $y_i \to y$ and $z_i \to z$, where $r_i =: (\rho_i, \beta_i)$. Hence

$$\langle y_i^*, y_i - (g(x_i) + g'(x_i)\, \bar{d}(\omega_i)) \rangle \leqslant 0,$$

$$\langle z_i^*, z_i - \bar{d}(\omega_i) \rangle \leqslant 0,$$

and

$$0 = \nabla f(x_i) + H_i \bar{d}(\omega_i) + g'(x_i)^T y_i^* + z_i^*$$

for each $i = 1, 2, \dots$. Taking limits, we find that

$$\langle y^*, y - (g(x) + g'(x)\, d) \rangle \leqslant 0,$$

$$\langle z^*, z - d \rangle \leqslant 0,$$

and

$$0 = \nabla f(x) + Hd + g'(x)^T y^* + z^*.$$

Now, since y and z were chosen arbitrarily from $C + \kappa(x, \rho, \lambda) \mathbb{B}$ and $\beta \mathbb{B}$, respectively, we have that $y^* \in N(g(x) + g'(x) d \mid C + \kappa(x, \rho, \lambda) \mathbb{B})$, $z^* \in N(d \mid \beta \mathbb{B})$, and $0 = \nabla f(x) + Hd + g'(x)^T y^* + z^*$. Therefore $(y^*, z^*) \in QM_1(\omega)$. ∎

We now address the local boundedness of the multifunction QM_1. We say that the multifunction QM_1 is *locally bounded* at a point $\omega \in \Omega$ if there is a neighborhood U of ω and a constant $\eta > 0$ such that $QM_1(U) \subset \eta \mathbb{B}$, where

$$QM_1(U) := \bigcup_{\omega \in U \cap \Omega} QM_1(\omega).$$

PROPOSITION 6.2. *Let* $\omega = (x, r, \lambda, H) \in \mathrm{cl}\, \Omega$. *If the multifunction* $QM_1 : \mathrm{cl}\, \Omega \Rightarrow \mathbb{R}^m \times \mathbb{R}^n$ *is not locally bounded at* ω, *then there exists* $(y^*, z^*) \in QM_0(\omega)$ *such that* $y^* \neq 0$.

Proof. If QM_1 is not locally bounded at ω, then there exist sequences $\{\omega_i\} \subset \Omega$ and $\{(y_i^*, z_i^*)\} \subset \mathbb{R}^m \times \mathbb{R}^n$ such that $\omega_i \to \omega$, $\max\{|y_i^*|_0, |z_i^*|_0\} \uparrow \infty$, and $(y_i^*, z_i^*) \in QM_1(\omega_i)$ for all $i = 1, 2, \ldots$. If $\{y_i^*\}$ is bounded, then $|z_i^*|_0 \uparrow \infty$ and we can assume that $z_i^* |z_i^*|_0^{-1} \to z^*$. But, since

$$0 = \nabla f(x_i) + H_i \bar{d}(\omega_i) + g'(x_i)^T y_i^* + z_i^*,$$

where $\omega_i = (x_i, r_i, H_i)$ for all $i = 1, 2, \ldots$, we have that $z^* = 0$, due to the boundedness of $\{\omega_i, \bar{d}(\omega_i)\}$. This contradicts the fact that $|z^*|_0 = 1$. Hence the sequence $\{y_i^*\}$ is unbounded and we can assume that $y_i^* |y_i^*|_0^{-1} \to y^*$ for some $y^* \in \mathbb{R}^m \times \{0\}$. Dividing the relation

$$0 = \nabla f(x_i) + H_i \bar{d}(\omega_i) + g'(x_i)^T y_i^* + z_i^*$$

through by $|y_i^*|_0$ and taking the limit as $i \to \infty$ yields the existence of a $z^* \in \mathbb{R}^n$ such that $z_i^* |y_i^*|_0^{-1} \to z^*$. Now, due to the continuity of g, g', and κ, and the upper semi-continuity of the multifunction D, it is straightforward to show that $(y^*, z^*) \in QM_0(\omega)$. ∎

THEOREM 6.3. *Let* $w = (x, r, \lambda, H) \in \Omega$ *be such that* $(r, \lambda) \in \mathrm{int}(T \times [0, 1])$. *Then either* $M_0(x) \neq \{0\}$ *or the multifunction* QM_1 *is locally bounded at* ω.

Proof. First suppose that $\varphi(x) = 0$. If QM_1 is not locally bounded at ω, then, by Proposition 6.2, $QM_0(\omega) \neq \{0\}$. Hence, by Theorem 4.3, $M_0(x) \neq \{0\}$. Next, if $\varphi(x) > 0$ and $\kappa(x, \rho, \lambda) = \varphi(x)$, then x is an ES point for NLP and so $M_0(x) \neq \{0\}$. Thus, for the remainder of the proof, we will assume that $\varphi(x) > 0$ and $\kappa(x, r, \lambda) \neq \varphi(x)$. The result will then follow from Proposition 6.2 once we have shown that $QM_0(\omega) = \{0\}$.

By part (1) of Lemma 4.2, there is a vector $d_1 \in \mathbb{R}^n$ such that

$$
\begin{gathered}
g'(x) d_1 \in \mathrm{int}[C + \kappa(x, \rho, \lambda)\, \mathbb{B} - g(x)], \\
d_1 \in (\mathrm{ri}[X - x]) \cap \mathrm{int}(\beta\mathbb{B}).
\end{gathered}
\tag{6.4}
$$

Setting $d_2 := \frac{1}{2}(d_1 - \bar{d}(\omega))$, we have that $\bar{d}(\omega) + d_2$ also satisfies (6.4). Let $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ be given. Then there is a $\bar{\mu} > 0$ such that

$$
\mu y + g'(x)(\bar{d}(\omega) + d_2 - \mu z) \in \mathrm{int}[C + \kappa(x, \rho, \lambda)\, \mathbb{B} - g(x)]
$$

and

$$
\bar{d}(\omega) + d_2 - \mu z \in \mathrm{int}(\beta\mathbb{B})
$$

for all $\mu \in [0, \bar{\mu}]$. Set $d_3 := \mu^{-1} d_2 - z$ for some $\mu \in [0, \bar{\mu}]$. Then

$$
\mu y + g'(x)(\bar{d}(\omega) + \mu d_3) \in \mathrm{int}[C + \kappa(x, \rho, \lambda)\, \mathbb{B} - g(x)]
$$

and

$$
\mu z + \bar{d}(\omega) + \mu d_3 \in (\mathrm{ri}[X - x]) \cap \mathrm{int}(\beta\mathbb{B}).
$$

Therefore,

$$
\begin{aligned}
\binom{y}{z} &+ \begin{bmatrix} g'(x) \\ I \end{bmatrix} d_3 \\
&\in T\left(\left. \begin{matrix} g(x) + g'(x)\,\bar{d}(\omega) \\ \bar{d}(\omega) \end{matrix} \right| (C + \kappa(x, \rho, \lambda)\, \mathbb{B}) \times ((X - x) \cap \beta\mathbb{B}) \right).
\end{aligned}
$$

Now since $\binom{y}{z} \in \mathbb{R}^m \times \mathbb{R}^n$ was arbitrary, we have

$$
\begin{aligned}
\mathbb{R}^m \times \mathbb{R}^n = \,&\mathrm{Ran} \begin{bmatrix} g'(x) \\ I \end{bmatrix} \\
&+ T\left(\left. \begin{matrix} g(x) + g'(x)\,\bar{d}(\omega) \\ \bar{d}(\omega) \end{matrix} \right| (C + \kappa(x, \rho, \lambda)\, \mathbb{B}) \times ((X - x) \cap \beta\mathbb{B}) \right).
\end{aligned}
$$

Thus, by duality, $QM_0(\omega) = \{0\}$. ∎

We now conclude this section with our main result concerning the local boundedness of the distance function $\text{dist}_0(0 | M_C(\omega))$.

THEOREM 6.5. *Let* $U_2 \times \Gamma'$ *be a compact subset of* $\text{int}(T \times [0, 1]) \times \Gamma$ *and let* $\bar{x} \in X$. *Then either* $M_0(\bar{x}) \neq \{0\}$ *or there is a neighborhood* U_1 *of* $\bar{x}$ *such that*

$$\sup\{\max\{|y^*|_0, |z^*|_0\} \mid (y^*, z^*) \in QM_1(\omega), \omega \in \Omega'\} < \infty \qquad (6.6)$$

and

$$\sup\{\text{dist}_0(0 | M_C(\omega)): \omega \in \Omega'\} < \infty, \qquad (6.7)$$

where $\Omega' := (U_1 \cap X) \times U_2 \times \Gamma'$.

Proof. Clearly (6.7) follows from (6.6). Thus, we only establish (6.6). If (6.6) does not hold, then there is a sequence $\{\omega_i\} \subset \Omega'$, with $\omega_i =: (x_i, r_i, H_i)$ for all $i = 1, 2, \ldots$, such that $\omega_i \to \bar{\omega} = (\bar{x}, \bar{r}, \bar{\lambda}, \bar{H})$ with $(\bar{r}, \bar{\lambda}, \bar{H}) \in \text{int}[T \times [0, 1]] \times \Gamma$ and $\text{dist}_0(0 | QM_1(\omega_i)) \to \infty$. But then QM_1 is not locally bounded at $\bar{\omega}$. Hence, by Theorem 6.3, we know that $M_0(\bar{x}) \neq \{0\}$. ∎

From inequality (5.10) we have the following corollary.

COROLLARY 6.8. *Let* $U_2 \times \Gamma'$ *be a compact subset of* $\text{int}(\Gamma \times [0, 1]) \times \Gamma$ *and let* $\bar{x} \in X$. *Then either* $M_0(\bar{x}) \neq \{0\}$ *or there is a neighborhood* U_1 *of* $\bar{x}$ *such that*

$$\sup\left\{ \left| \frac{\nabla f(x)^T \bar{d}(\omega) + \bar{d}(\omega)^T H \bar{d}(\omega)}{\varphi(x) - \kappa(x, \rho, \lambda)} \right| \mid \omega \in \Omega' \right\} < \infty,$$

where $\Omega' := (U_1 \cap X) \times U_2 \times \Gamma'$ *and* $\omega =: (x, r, \lambda, H)$ *and* $r =: (\rho, \beta)$ *for each* $\omega \in \Omega$.

7. THE ALGORITHM

Initialization: Choose $x_0 \in X$, $\alpha_0 > 0$, and a compact subset, Π, of $\text{int}(T \times [0, 1]) \times \Gamma$. Select step length parameters $\gamma_1 > 0$, $1 > \gamma_2 > 0$, and $0 < \mu_1 \leqslant \mu_2 < 1$.

The Iteration: Let $x_i \in X$ and $\alpha_i > 0$ be given. If x_i is a stationary point of NLP stop; otherwise choose parameters $(r_i, \lambda_i, H_i) \in \Pi$ and the step length τ_i so that the update

$$x_{i+1} := x_i + \tau_i \bar{d}(\omega_i)$$

satisfies the inequality

$$P_{\alpha_{i+1}}(x_{i+1}) \leqslant P_{\alpha_{i+1}}(x_i)$$
$$+ \mu_1 \tau_i [\nabla f(x_i)^T \bar{d}(\omega_i) + \alpha_{i+1}(\text{dist}(g(x_i) + g'(x_i) \bar{d}(\omega_i)|C) - \varphi(x_i))]$$

and either $\tau_i \leqslant \gamma_1$ there is a $\bar{\tau}_i > 0$ such that $\tau_i \geqslant \gamma_2 \bar{\tau}_i > 0$, $x_i + \bar{\tau}_i \bar{d}(\omega_i) \in X$, and $\bar{x}_{i+1} := x_i + \bar{\tau}_i \bar{d}(\omega_i)$ satisfies the inequality

$$P_{\alpha_{i+1}}(\bar{x}_{i+1}) > P_{\alpha_{i+1}}(x_i)$$
$$+ \mu_1 \bar{\tau}_i [\nabla f(x_i)^T \bar{d}(\omega_i) + \alpha_{i+1}(\text{dist}(g(x_i) + g'(x_i) \bar{d}(\omega_i)|C) - \varphi(x_i))],$$

where $\omega_i := (x_i, r_i, \lambda_i, H_i)$, $r_i =: (\rho_i, \beta_i)$, and $\alpha_{i+1} := \tilde{\alpha}(\omega_i, \alpha_i)$ is defined in (5.6).

Remarks. (1) The step length criteria employed by the algorithm was introduced in Calamai and Moré [6]. It follows from the results of Section 5 that for each $(r_i, \lambda_i, H_i) \in \Pi$ there is a $\tau_i > 0$ for which the update x_{i+1} satisfies the requirements of the step length choice as long as x_i is not a stationary point for NLP. In this regard, it is important to note that if $\bar{d}(\omega_i)$ is not a Kuhn–Tucker point of $Q(\omega_i)$ for some choice of $(r_i, \lambda_i, H_i) \in \Pi$, then it is not a Kuhn–Tucker point for any choice of $(r_i, \lambda_i, H_i) \in \Pi$ and consequently is a stationary point for NLP. Thus the stopping criteria can be checked in the process of implementing the update. In particular, the procedure must stop if at any point one finds that $\tilde{\alpha}(\omega_i, \alpha_i) = 0$.

(2) As noted in the previous remark, the stopping criteria for the algorithm can be checked in the process of implementing the update. In this regard, it can happen that a nontrivial update is computed even though x_i is an ES point for NLP. This is not a drawback, since in this case it may be possible to continue progress toward either a feasible stationary point for NLP or an s–ES point for NLP. But, as noted at the end of Section 5, if one proceeds in this way, then one must terminate the algorithm if it is ever the case that $\tilde{\alpha}(\omega_i, \alpha_i) = 0$.

(3) Although in this paper we are only concerned with the global convergence properties of the alorithm, for reasons associated with its local convergence properties it is desirable to allow the parameter λ_i to converge to 1 as $[\kappa(x_i, \rho_i, 1) - \varphi(x)] \to 0$. A straightforward modification to the analysis that we provide allows one to incorporate such a modification. Specifically, this can be done by replacing λ with a locally Lipschitz function $\bar{\lambda}: \mathbb{R}^n \times \mathbb{R}_1 \to [\lambda_0, 1]$ for some $\lambda_0 \in (0, 1)$, where the mapping $\bar{\lambda}$ satisfies the property that $\bar{\lambda}(x, \rho) = 1$ only if $\kappa(x, \rho, 1) = \varphi(x)$. Then at each iteration one sets $\lambda_i = \bar{\lambda}(x_i, \rho_i)$. An example of such a function $\bar{\lambda}$ is given by

$$\bar{\lambda}(x, \rho) := \max\{\lambda_0, 1 + (\kappa(x, \rho, 1) - \varphi(x))\}.$$

(4) The algorithm is general enough to allow at least a partial implementation of a trust region-like strategy with second-order corrections, as is done in Sahba [19]. The limitations are that the matrices H_i remain uniformly symmetric positive definite and bounded, and that the trust region radii neither converge to zero nor become unbounded.

(5) In the implementation of the algorithm described above, the sequence of penalty parameters $\{\alpha_i\}$ is necessarily nondecreasing. However, one may employ a clever device proposed by Sahba [19] for reducing the penalty parameter on certain iterations. Specifically, at the end of the kth iteration one evaluates

$$\tilde{\varphi}_k := \min\{\varphi(x_k), \tilde{\varphi}_{k-1}\},$$

thus keeping track of the minimum value of the $\varphi(x_i)$'s. If $\tilde{\varphi}_k \leqslant \tilde{\varphi}_{k-1} - \varepsilon$ for some prespecified $\varepsilon > 0$, then one resets α_{i+1} to α_0. Clearly, this re-initialization of α can only occur a finite number of times. Hence the convergence analysis remains unaltered.

(6) When implementing the above procedure, it appears as though one must accurately compute the value $\kappa(x, \rho, 1)$. This can be a drawback to the method, particularly if one intends to implement the procedure using a trust region approach. Fortunately, it is possible to avoid much of this work. The key observation is that it is sufficient to compute a direction $d \in [X - x] \cap \rho \mathbb{B}$ such that

$$\text{dist}(g(x) + g'(x) d | C) - \varphi(x) \leqslant \lambda[\kappa(x, \rho, 1) - \varphi(x)] \qquad (7.1)$$

and then replace the term $\lambda[\kappa(x, \rho, 1) - \varphi(x)]$ in the definition of $\kappa(x, \rho, \lambda)$ with the term $[\text{dist}(g(x) + g'(x) d | C) - \varphi(x)]$. It is straightforward, although tedious, to show that this modification to the algorithm does not alter its convergence characteristics. We now briefly describe one way to take advantage of this variation. Suppose one employs a method for evaluating $\kappa(x, \rho, 1)$ that produces a sequence $\{(l_i, d_i)\} \subset \mathbb{R} \times ([X - x] \cap \rho \mathbb{B})$ such that

$$\text{dist}(g(x) + g'(x) d_i | C) \downarrow \kappa(x, \rho, 1),$$
$$l_i \uparrow \kappa(x, \rho, 1). \qquad (7.2)$$

If one terminates this procedure when

$$\text{dist}(g(x) + g'(x) d_i | C) - \varphi(x) \leqslant \lambda[l_i - \varphi(x)]$$

(which must occur after a finite number of iterations), then inequality (7.1) is satisfied with $d_i \in [X - x] \cap \rho \mathbb{B}$, and the value $\kappa(x, \rho, 1)$ was never computed. Observe that if the sets C and X are polyhedral, and the norms on

$\mathbb{R}^n$ and $\mathbb{R}^m$ are polyhedral, then the computation of $\kappa(x, \rho, 1)$ is a linear program. One can then use either the Anstreicher [1], the Gay [8], or the Todd–Burrell [21] variation of Karmarkar's algorithm [10] to solve this linear program, thereby producing vectors $d_i \in [X - x] \cap \rho \mathbb{B}$ and scalars l_i satisfying (7.2). Since Karmarkar's algorithm computes good approximate solutions to linear programs very rapidly, this modification to the algorithm is quite attractive.

8. CONVERGENCE

THEOREM 8.1. *Let* $\{x_i\}$ *be a sequence generated by the algorithm of Section 7 and suppose that the mappings* ∇f *and* g' *are bounded on* $\{x_i\}$ *and uniformly continuous on* $\overline{\mathrm{co}}\{x_i\}$.

(1)   *If* $\alpha_i \uparrow \infty$, *then*

$$\frac{\nabla f(x_i)^T d(\omega_i) + d(\omega_i)^T H_i d(\omega_i)}{\varphi(x_i) - \kappa(x_i, p_i, \lambda_i)} \uparrow_I \infty$$

and

$$[\varphi(x_i) - \kappa(x_i, \rho_i, \lambda_i)] \xrightarrow{I} 0,$$

where $I := \{i : \alpha_i < \alpha_{i+1}\}$.

(2)   *If* $\sup_i \alpha_i =: \alpha$, *then either*

(a) $\inf_i P_{\bar{\alpha}}(x_i) = -\infty$ *for all* $\bar{\alpha} \leqslant \alpha$, *or*

(b) $\{x_i\}$ *is finitely terminating at a stationary point NLP, or*

(c) $|d(\omega_i)| \to 0$, $[\varphi(x_i) - \kappa(x_i, p_i, \lambda_i)] \to 0$, *and* $\nabla_x L(x_i, y_i^*, z_i^*)$ $\to 0$, *where* $L(x, y^*, z^*) = f(x) + g(x)^T y^* + z^*$ *and* (y_i^*, z_i^*) *is any element of* $QM_i(\omega_i)$ *for each* $i = 1, 2, \dots$.

Proof.   For the remainder of the proof we denote $d(\omega_i)$ by d_i.

(1)   This follows immediately from the statement of the algorithm, rule (5.6), and the hypotheses.

(2)   Statement (b) is obvious from the statement of the algorithm. Thus we assume that neither (a) nor (b) occurs, and show that (c) must occur. Clearly there is no loss of generality by assuming that $\alpha_i = \alpha$ for all $i < 1, 2, \dots$. Now, since $P_\alpha(x_i)$ is bounded below, we have that $[P_\alpha(x_{i+1}) - P_\alpha(x_i)] \to 0$. Therefore $\tau_i d_i^T H_i d_i \to 0$. If $d_i^T H_i d_i \to 0$, then $|d_i| \to 0$. Also, $[\kappa(x_i, \rho_i, \lambda_i) - \varphi(x_i)] \to 0$ since

$$0 \leqslant [\varphi(x_i) - \kappa(x_i, \rho_i, \lambda_i)]$$

$$\leqslant \varphi(x_i) - \text{dist}(g(x_i) + g'(x_i) \, d_i | C)$$

$$\leqslant |g'(x_i) \, d_i|,$$

where $|g'(x_i) \, d_i| \to 0$ due to the boundedness of $\{g'(x_i)\}$. Moreover, for each $(y_i^*, z_i^*) \in QM_1(\omega_i)$,

$$\nabla_x L(x_i, y_i^*, z_i^*) = -H_i d_i$$

so that $\nabla_x L(x_i, y_i^*, z_i^*) \to 0$. Hence the result is established if $d_i^T H_i d_i \to 0$. Thus let us suppose that $d_i^T H_i d_i \nrightarrow 0$. Then there is a subsequence $I \subset \mathbb{N}$ such that $\tau_i \to^I 0$ with $\tau_i < \gamma_i$ for all $i \in I$, and

$$\inf_I \{d_i^T H_i d_i\} =: \theta > 0.$$

Consequently, the associated subsequence $\{\bar{\tau}_i\}_I$ satisfies $\tau_i \geqslant \gamma_2 \bar{\tau}_i > 0$ and

$$P_\alpha(\bar{x}_{i+1}) - P_\alpha(x_i) > \mu_2 \bar{\tau}_i [\nabla f(x_i)^T d_i + \alpha(\text{dist}(g(x_i) + g'(x_i) \, d_i | C) - \varphi(x_i))] \tag{8.2}$$

for all $i \in I$. We assume, with no loss of generality, that $\bar{\tau}_i < 1$ for all $i \in I$. Now observe that

$$\varphi(\bar{x}_{i+1}) - \varphi(x_i) \leqslant |g(\bar{x}_{i+1}) - (g(x_i) + \bar{\tau}_i g'(x_i) \, d_i)|$$

$$+ [\text{dist}(g(x_i) + \bar{\tau}_i g'(x_i) \, d_i | C) - \varphi(x_i)]$$

$$\leqslant \bar{\tau}_i |d_i| \, w_2(\bar{\tau}_i d_i) + \bar{\tau}_i [\text{dist}(g(x_i) + \bar{\tau}_i g'(x_i) \, d_i | C) - \phi(x_i)],$$

where w_2 is the modulus of continuity for g'. Hence, by expanding on (8.2), we obtain

$$\mu_2 \bar{\tau}_i [\nabla f(x_i)^T d_i + \alpha(\text{dist}(g(x_i) + g'(x_i) \, d_i | C) - \varphi(x_i))]$$

$$\leqslant \bar{\tau}_i [\nabla f(x_i)^T d_i + \alpha(\text{dist}(g(x_i) + g'(x_i) \, d_i | C) - \varphi(x_i))]$$

$$+ |f(\bar{x}_{i+1}) - (f(x_i) + \bar{\tau}_i \nabla f(x_i)^T d_i)| + \bar{\tau}_i |d_i| \, w_2(\bar{\tau}_i d_i)$$

$$\leqslant \bar{\tau}_i [\nabla f(x_i)^T d_i + \alpha(\text{dist}(g(x_i) + g'(x_i) \, d_i | C) - \varphi(x_i))]$$

$$+ \bar{\tau}_i |d_i| [w_1(\bar{\tau}_i d_i) + w_2(\bar{\tau}_i d_i)],$$

where w_1 is the modulus of continuity for ∇f. Therefore,

$$0 \leqslant (1 - \mu_2)[\nabla f(x_i)^T d_i + \alpha(\text{dist}(g(x_i) + g'(x_i) \, d_i | C) - \varphi(x_i))]$$

$$+ |d_i| \, w(\bar{\tau}_i d_i)$$

$$\leqslant (\mu_2 - 1) \theta + |d_i| \, w(\bar{\tau}_i d_i)$$

for all $i \in I$, where $w := w_1 + w_2$. Taking the limit as $i \to^I \infty$ yields the contradiction $0 \leqslant (\mu_2 - 1) \, \theta$. Hence $d_i^T H_i d_i \to 0$. ∎

COROLLARY 8.3. *Let* $\{x_n\}, f,$ *and* g *satisfy the hypotheses of Theorem* 8.1.

(1) *If* $\alpha_i \uparrow \infty$, *then every cluster point of the subsequence* $\{x_i : \alpha_i < \alpha_{i+1}\}$ *is either an ES point or a Fritz John point for NLP.*

(2) *If* $\sup_i \alpha_i < \infty$, *then every cluster point of* $\{x_i\}$ *is a stationary point for NLP.*

Proof. (1) By part (1) of Theorem 8.1 and Corollary 6.6, we know that $M_0(\bar{x}) \neq 0$ for any cluster point $\bar{x}$ of $\{x_i : \alpha_i < \alpha_{i+1}\}$. Hence $\bar{x}$ is either an ES point or a Fritz John point for NLP.

(2) By part (2) of Theorem 8.1, any cluster point $\bar{x}$ of the sequence $\{\alpha_i\}$ must satisfy $\kappa(\bar{x}, \bar{\rho}, \bar{\lambda}) = \varphi(\bar{x})$ for some $\bar{\rho} > 0$ and $\bar{\lambda} \in (0, 1)$, due to the compactness of Π and the continuity of κ and φ.

If $\varphi(\bar{x}) > 0$, then $\bar{x}$ is an ES point for NLP.

If $\varphi(\bar{x}) = 0$ and $\bar{x}$ is not a Fritz John pont for NLP, then, by Theorem 6.3 and the compactness of Π, there is a subsequence $\{x_i : i \in I\}$ of $\{x_i\}$ for which $x_i \to^I \bar{x}$ and there exist $(y_i^*, z_i^*) \in QM_1(\omega_i)$ for each $i \in I$ such that $(y_i^*, z_i^*) \to^I (y^*, z^*)$ for some (y^*, z^*) in $\mathbb{R}^m \times \mathbb{R}^n$. By Proposition 6.1, $(y^*, z^*) \in QM_1(\omega)$.

Moreover, by part (2) of Theorem 8.1,

$$0 = \nabla_x L(\bar{x}, y^*, z^*).$$

Hence $\bar{x}$ is a Kuhn–Tucker point for NLP. ∎

The next corollary follows immediately from Corollary 8.3.

COROLLARY 8.4. *Let* $\{x_n\}$, f, *and* g *satisfy the hypotheses of Theorem* 8.1. *If the sequence* $\{x_n\}$ *is bounded, then the sequence* $\{x_n\}$ *has a cluster point that is a stationary point for NLP. If, moreover, the sequence* $\{\alpha_i\}$ *is also bounded, then every cluster point of the sequence* $\{x_i\}$ *is a stationary point for NLP.*

The convergence results that are stated above are almost identical to those obtained for the algorithm presented in [3]. Thus it is reasonable to ask which of these algorithms is better. Ultimately, the answer to this question can only be given after extensive numerical testing. Nonetheless, there is theoretical evidence indicating that the procedure proposed in this paper is superior. Specifically, one can establish certain continuity results for the programs $Q(\omega)$ that are not available for the modified subproblems employed in [3].

9. SOME CONTINUITY PROPERTIES OF $Q(\omega)$

In this section we provide some continuity results for the convex programs $Q(\omega)$. These results indicate that the algorithm of Section 7 which is based upon the programs $Q(\omega)$ possesses better stability and continuity properties than the procedure proposed in [3].

In order to more fully appreciate these results in the context of the existing literature we introduce the following constraint qualification.

DEFINITION 9.1. Let $g: \mathbb{R}^n \to \mathbb{R}^m$, $C \subset \mathbb{R}^m$, and $X \subset \mathbb{R}^n$ be as posited in the statement of NLP. Then g is said to satisfy the CQ condition at a point $x \in X$ if and only if $M_0(x) = \{0\}$.

Remarks. (1) It is important to note that, since

$$M_0(x) = \text{Nul}[g'(x)^T, I] \cap N\left(\begin{pmatrix} g(x) \\ x \end{pmatrix}\middle| [C + \varphi(x)\,\mathbb{B}] \times X\right),$$

the statement $M_0(x) = \{0\}$ is equivalent to the statement

$$\mathbb{R}^m \times \mathbb{R}^n = (M_0(x))^0$$

$$= \text{Ran}\begin{bmatrix} g'(x) \\ I \end{bmatrix} + T\left(\begin{pmatrix} g(x) \\ x \end{pmatrix}\middle| [C + \varphi(x)\,\mathbb{B}] \times X\right).$$

This latter condition was used in the definition of the CQ condition given in [4, Definition 6.1]. For a definition that generalizes Definition 9.1 to more general situations see Rockafellar [18].

(2) If $C = \mathbb{R}^s_+ \times \{0\}_{\mathbb{R}^l}$ and $X = \mathbb{R}^n$, then the CQ condition is equivalent to the Mangasarian–Fromowitz constraint qualification.

Observe, from Definition 2.8, that a point $x \in \mathbb{R}^n$ is either an ES point or a Fritz John point for NLP if and only if g does not satisfy the CQ condition at x. In [4], it is shown that the CQ condition is equivalent to the regularity condition of Maguregui [11]. Consequently, the CQ condition is a powerful tool for the local analysis of the constrained system $g(x) \in C$ and $x \in X$. In this regard, we need the following result which is essentially due to Maguregui [11].

PROPOSITION 9.2. *If g satisfies the CQ condition at a point $x \in X$, then there is a neighborhood U of x such that g satisfies the CQ condition at every point of $U \cap X$.*

Proof. By [4, Proposition 6.5], g satisfies the CQ condition at $x \in X$ if

and only if x is a *regular* point (in the sense of Maguregui [11, Chap. 2])
of the system

$$g(x) \in C + \text{dist}(g(x) \mid C) \, \mathbb{B},$$

$$x \in X.$$

Therefore the result follows from Maguregui [11, Chap. 2, Theorem 2]. ∎

We now employ the CQ condition in the description of certain con-
tinuity properties of the convex programs $Q(\omega)$. The key result in this con-
text concerns the continuity of the multifunction $D: X \times \text{cl } T \times [0, 1] \Rightarrow \mathbb{R}^n$
defined in Section 3. Recall that in Proposition 3.3 it was shown that D is
upper semi-continuous on $X \times \text{cl } T \times [0, 1]$. We now show that D is also
lower semi-continuous at a point (x, r, λ) if g satisfies the CQ condition at
x. Hence at such points D is continuous. This result extends some of the
work in Robinson and Meyer [13], Robinson [15], and Rockafellar [17].

THEOREM 9.3.   Let $(\bar{x}, \bar{r}, \bar{\lambda}) \in X \times T \times [0, 1]$ be such that $(\bar{r}, \bar{\lambda}) \in \text{int}(T \times [0, 1])$ and g satisfies the CQ condition at $\bar{x}$. Then the multifunction D is
continuous at $(\bar{x}, \bar{r}, \bar{\lambda})$ with respect to $X \times T \times [0, 1]$. More specifically, we
have the following:

(1)   If $\varphi(\bar{x}) = 0$, then the multifunction D is sub-Lipschitzian at $(\bar{x}, \bar{r}, \bar{\lambda})$
with respect to $X \times T \times [0, 1]$, that is, for every compact set $K \subset \mathbb{R}^{n+3}$ there
is a neighborhood U of $(\bar{x}, \bar{r}, \bar{\lambda})$ and a constant $\theta \geqslant 0$ such that

$$D(x, r, \lambda) \cap K \subset D(x_2, r_2, \lambda_2) + \theta |(x_1, r_1, \lambda_1) - (x_2, r_2, \lambda_2)| \mathbb{B}$$

for all $(x_i, r_i, \lambda_i) \in U \cap (X \times T \times [0, 1])$, $i = 1, 2$.

(2)   If $\varphi(x) > 0$, then the multifunction D is both upper and lower
semi-continuous at $(\bar{x}, \bar{r}, \bar{\lambda})$ with respect to $X \times T \times [0, 1]$.

Remarks. (1) Recall that D is said to be lower semi-continuous at $(\bar{x}, \bar{r}, \bar{\lambda})$
if for every $d \in D(\bar{x}, \bar{r}, \bar{\lambda})$ and $\varepsilon > 0$ there is a $\delta > 0$ such that $D(x, r, \lambda) \cap
(d + \varepsilon \mathbb{B}) \neq 0$ whenever $(x, r, \lambda) \in ((\bar{x}, \bar{r}, \bar{\lambda}) + \delta \mathbb{B}) \cap (X \times T \times [0, 1])$.

(2)   Lipschitzian properties of multifunctions were extensively studied
in Rockafellar [17], wherein the term "sup-Lipschitzian" first appears. We
make strong use of these results in part (1) of Theorem 9.3. In some sense,
part (1) of the theorem is the classical result in this area; it represents a
modest generalization of Robinson and Meyer [13, Theorem 3] and
Robinson [15, Theorem 2] and, due to Theorem 4.3, is a direct conse-
quence of Rockafellar [17, Corollary 3.3]. On the other hand, part (2) of
the theorem is of a somewhat different nature and so requires a different
approach.

Proof. Clearly, the result will be established once (1) and (2) have been verified.

(1) Since g satisfies the CQ condition at $\bar{x}$, we know that $\bar{x}$ is not a Fritz John point for NLP. Hence, by Theorem 4.3, if $d \in D(\bar{x}, \bar{r}, \bar{\lambda})$, then d cannot be a Fritz John point for $Q(\bar{x}, \bar{r}, \bar{\lambda}, H)$ regardless of the choice of $H \in \Gamma$. Therefore the result follows from Rockafellar [17, Corollary 3.3].

(2) We assume, with no loss of generality, that the norm of $\mathbb{R}^{n+3}$ is given by

$$|(x, \rho, \beta, \lambda)| = \max\{|x|, |\rho|, |\beta|, |\lambda|\}$$

for each $(x, \rho, \beta, \lambda) \in \mathbb{R}^{n+3}$. The fact that D is upper semi-continuous at every point in $X \times \mathrm{cl}\, T \times [0, 1]$ was verified in Proposition 3.3. Thus we only verify the lower semi-continuity of D of $(\bar{x}, \bar{r}, \bar{\lambda})$. Let $\varepsilon > 0$ and $d \in D(\bar{x}, \bar{r}, \bar{\lambda})$ be given. Since g satisfies the CQ condition at $\bar{x}$, we know, from Proposition 3.4 and Lemma 4.2, that $\kappa(\bar{x}, \bar{\rho}, \bar{\lambda}) \neq \varphi(\bar{x})$. Hence, by Lemma 4.2, there is a $d_1 \in \mathbb{R}^n$ such that $|d_1| < \varepsilon/2$ and

$$g(\bar{x}) + g'(\bar{x})(d + d_1) \in \mathrm{int}[C + \kappa(\bar{x}, \bar{\rho}, \bar{\lambda})\, \mathbb{B}],$$

$$\bar{x} + d + d_1 \in \mathrm{ri}\, X,$$

and

$$d + d_1 \in \beta\mathbb{B}.$$

Choose $\varepsilon_0 > 0$ so that

$$\varepsilon_0 < \mathrm{dist}(g(\bar{x}) + g'(\bar{x})(d + d_1)|\, \mathrm{bdry}(C + \kappa(\bar{x}, \bar{r}, \bar{\lambda})\, \mathbb{B}))$$

and

$$\varepsilon_0 < \mathrm{dist}(d + d_1 |\, \mathrm{bdry}(\beta\mathbb{B})).$$

Let $\delta > 0$ be such that $\min\{\varepsilon/2, \varepsilon_0/3\} > \delta$,

$$|g(x) + g'(x)(d + d_1 + d_2) - (g(\bar{x}) + g'(\bar{x})(d + d_1 + d_2))| < \varepsilon_0/3, \quad (9.4)$$

and

$$|\kappa(x, \rho, \lambda) - \kappa(\bar{x}, \bar{\rho}, \bar{\lambda})| < \varepsilon_0/3 \qquad (9.5)$$

whenever $\max\{|x - \bar{x}|, |\rho - \bar{\rho}|, |\lambda - \bar{\lambda}|, |d_2|\} < \delta$ and $(x, \rho, \lambda) \in X \times \mathbb{R}_+ \times [0, 1]$. (Recall that κ is locally Lipschitzian on $X \times \mathbb{R}_+ \times [0, 1]$ by Proposition 3.2.) Now, if $|(x, r, \lambda) - (\bar{x}, \bar{r}, \bar{\lambda})| < \delta$ and $(x, r, \lambda) \in X \times T \times [0, 1]$, then $\mathrm{dist}(x + d + d_1 | X) < \delta$. Hence there is a $d_2 \in \mathbb{R}^n$ such that

$x + d + d_1 + d_2 \in X$ and $|d_2| < \delta$. We have that $|d_1 + d_2| < \varepsilon$ and $d + d_1 + d_2 \in \beta \mathbb{B}$. Hence (9.4) and (9.5) hold so that

$$g(x) + g'(x)(d + d_1 + d_2) + \tfrac{1}{3}\varepsilon_0 \mathbb{B} \subset \text{int}[C + \kappa(\bar{x}, \bar{\rho}, \lambda)\,\mathbb{B}]$$

and

$$C + \kappa(\bar{x}, \bar{\rho}, \lambda)\,\mathbb{B} \subset C + \kappa(x, \rho, \lambda)\,\mathbb{B} + \tfrac{1}{3}\varepsilon_0 \mathbb{B}.$$

Hence, by Radstrom's cancelation lemma,

$$g(x) + g'(x)(d + d_1 + d_2) \subset C + \kappa(x, \rho, \lambda)\,\mathbb{B},$$

whereby D is lower semi-continuous at $(\bar{x}, \bar{r}, \lambda)$. ∎

Theorem 9.3 admits the following simple corollary concerning the continuity of the value function for $Q(\omega)$,

$$q(\omega) := \nabla f(x)^T \tilde{d}(\omega) + \tfrac{1}{2}\tilde{d}(\omega)\,H\tilde{d}(\omega),$$

where $\omega =: (x, r, \lambda, H)$, and the optimal solution vector $\tilde{d}(\omega)$.

COROLLARY 9.6.   Let $\omega_0 = (x_0, r_0, \lambda_0, H_0) \in \Omega$ be such that $(r_0, \lambda_0) \in \text{int}[T \times [0, 1]]$ and g satisfies the CQ condition at x_0. Then there is a neighborhood U_1 of x_0 such that the functions q and $\tilde{d}$ are continuous on the set $(U_1 \times U_2 \times \Gamma) \cap \Omega$, where U_2 is any neighborhood of (r_0, λ_0) satisfying $U_2 \subset \text{int}[T \times [0, 1]]$.

Proof. From the definitions of $q(\omega)$ and $\tilde{d}(\omega)$ it is clear that we need only establish the result for $\tilde{d}(\omega)$. To this end, let $U_1 \subset \mathbb{R}^n$ be the neighborhood of x_0 obtained in Proposition 9.2 and let $U_2 \subset \mathbb{R}^3$ be any neighborhood of (r_0, λ_0). Set $U := U_1 \times U_2 \times \Gamma$ and choose $\omega \in U \cap \Omega$. The result will be established once we have shown that $\tilde{d}$ is continuous at ω.

Let $\{\omega_i\} \subset U \cap \Omega$ be such that $\omega_i \to \omega$. We need to show that $\tilde{d}(\omega_i) \to \tilde{d}(\omega)$. By Theorem 9.3, there is a sequence $\{d_i\} \subset \mathbb{R}^n$ with $d_i \to \tilde{d}(\omega)$ and $d_i \in D(x_i, r_i, \lambda_i)$, where $\omega_i =: (x_i, r_i, \lambda_i, H_i)$, for all $i = 1, 2, \ldots$. Hence

$$\nabla f(x_i)^T \tilde{d}(\omega_i) + \tfrac{1}{2}\tilde{d}(\omega_i)\,H_i \tilde{d}(\omega_i) \leqslant \nabla f(x_i)^T d_i + \tfrac{1}{2}d_i^T H_i d_i \qquad (9.7)$$

for all $i = 1, 2, \ldots$. Now, since $\{\tilde{d}(\omega_i)\}$ is a bounded sequence, it admits a cluster point $\tilde{d}$, where $\tilde{d} \in D(\omega)$ by Proposition 3.3. But then, taking the limit in (9.7) over the appropriate subsequence, we find that

$$\nabla f(x)^T \tilde{d} + \tfrac{1}{2}\tilde{d}^T H\tilde{d} \leqslant \nabla f(x)^T \tilde{d}(\omega) + \tfrac{1}{2}(\omega)^T H\tilde{d}(\omega).$$

Thus, by uniqueness, $\tilde{d} = \tilde{d}(\omega)$. Hence there is only one cluster point of the sequence $\{\tilde{d}(\omega_i)\}$ and it is $\tilde{d}(\omega)$. Therefore $\tilde{d}(\omega_i) \to \tilde{d}(\omega)$. ∎

REFERENCES

1. K. ANSTREICHER, A monotonic projective algorithm for fractional linear programming, preprint, School of Organization and Managment, Yale University, 1986.
2. J. V. BURKE, "Methods for Solving Generalized Inequalities with Application to Nonlinear Programming," Ph.D. thesis, Department of Mathematics, University of Illinois at Urbana–Champaign, 1983.
3. J. V. BURKE AND S-P. HAN, A modified sequential quadratic programming method, submitted for publication.
4. J. V. BURKE, "An Exact Penalization Viewpoint of Constrained Optimization," Argonne National Laboratory, Mathematics and Computer Science Division Report ANL/MCS-TM-95, Argonne, IL, 1987.
5. J. V. BURKE, Second order necessary and sufficient conditions for convex composite N.D.O., *Math. Programming*, in press.
6. P. H. CALAMAI AND J. J. MORÉ, Projected gradient methods for linearly constrained problems, *Math. Programming* **39** (1987), 93–116.
7. F. H. CLARKE, "Optimization and Nonsmooth Analysis," Canadian Mathematical Society Series of Monographs and Advanced Texts, Wiley, New York, 1983.
8. D. GAY, A variant of Karmarkar's linear programming algorithm for problems in standard form, *Math. Programming* **37** (1987), 81–90.
9. S-P. HAN, A globally convergent method for nonlinear programming, *J. Optim. Theory Appl.* **22** (1977), 297–309.
10. N. KARMARKAR, A new polynomial-time algorithm for linear programming, *Combinatorica* **4**(4) (1984), 373–395.
11. J. MAGUREGUI, "Regular Multivalued Functions and Algorithmic Applications, Ph.D. thesis, University of Wisconsin at Madison, 1977.
12. M. J. D. POWELL, A fast algorithm for nonlinearly constrained optimization calculations *in* "Proceedings, 1977 Dundee Biennial Conference on Numerical Analysis," Springer-Verlag, Berlin, 1977.
13. S. M. ROBINSON AND R. R. MEYER, Lower semi-continuity of multivalued linearization mappings, *SIAM J. Control* **11** (1973), 525–533.
14. S. M. ROBINSON, Perturbed Kuhn–Tucker points, and rates of convergence for a class of nonlinear programming algorithms, *Math. Programming* **7** (1974), 1–16.
15. S. M. ROBINSON, Stability theory for systems of inequalities. Part II. Differentiable nonlinear systems, *SIAM J. Numer. Anal.* **13** (1976), 497–513.
16. R. T. ROCKAFELLAR, "Convex Analysis," Princeton Univ. Press, Princeton, NJ, 1970.
17. R. T. ROCKAFELLAR, Lipschitzian properties of multifunctions, *Nonlinear Anal. TMA* **9** (1985), 867–885.
18. R. T. ROCKAFELLAR, Extensions of subgradient calculus with applications to optimization, *Nonlinear Anal. TMA* **9** (1985), 665–698.
19. M. SAHBA, Globally convergent algorithm for nonlinearly constrained optimization problems, *J. Optim. Theory Appl.* **52** (1987), 291–309.
20. K. SCHITTKOWSKI, The nonlinear programming method of Wilson, Han, and Powell with an augmented Lagrangian type line search function. *Numer. Math.* **38** (1981), 83–114.
21. M. J. TODD AND B. P. BURRELL, "An Extension of Karmarkar's Algorithm for Linear Programming Using Dual Variables," Technical Report No. 648, School of Operations Research and Industrial Engineering, Cornell University, 1985.
22. R. B. WILSON, "A Simplicial Algorithm for Concave Programming," Ph.D. dissertation, Graduate School of Business Administration, Harvard University, 1963.