Preprint ANL/MCS-P865-1200, Dec. 2000 (Revised Nov. 2001)
Mathematics and Computer Science Division
Argonne National Laboratory

Stephen J. Wright*

Constraint Identification and Algorithm Stabilization for Degenerate Nonlinear Programs

December 20, 2001

Abstract. In the vicinity of a solution of a nonlinear programming problem at which both strict complementarity and linear independence of the active constraints may fail to hold, we describe a technique for distinguishing weakly active from strongly active constraints. We show that this information can be used to modify the sequential quadratic programming algorithm so that it exhibits superlinear convergence to the solution under assumptions weaker than those made in previous analyses.

Key words. Nonlinear Programming Problems, Degeneracy, Active Constraint Identification, Sequential Quadratic Programming

1. Introduction

Consider the following nonlinear programming problem with inequality constraints:

$$\text{NLP:} \qquad \min_z \ \phi(z) \qquad \text{subject to } g(z) \le 0, \tag{1}$$

where $\phi : \mathsf{R}^n \to \mathsf{R}$ and $g : \mathsf{R}^n \to \mathsf{R}^m$ are twice Lipschitz continuously differentiable functions. Optimality conditions for (1) can be derived from the Lagrangian for (1), which is

$$\mathcal{L}(z, \lambda) = \phi(z) + \lambda^T g(z), \tag{2}$$

where $\lambda \in \mathsf{R}^m$ is the vector of Lagrange multipliers. When a constraint qualification holds at z^* (see discussion below), the first-order necessary conditions for z^* to be a local solution of (1) are that there exists a vector $\lambda^* \in \mathsf{R}^m$ such that

$$\nabla_z \mathcal{L}(z^*, \lambda^*) = 0, \quad g(z^*) \le 0, \quad \lambda^* \ge 0, \quad (\lambda^*)^T g(z^*) = 0. \tag{3}$$

These relations are the well-known Karush-Kuhn-Tucker (KKT) conditions. The set $\mathcal{B}$ of active constraints at z^* is

$$\mathcal{B} = \{i = 1, 2, \ldots, m \,|\, g_i(z^*) = 0\}. \tag{4}$$

Stephen J. Wright: Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439; `wright@mcs.anl.gov`

* Research supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, U.S. Department of Energy, under Contract W-31-109-Eng-38.

It follows immediately from (3) that we can have $\lambda_i^* > 0$ only if $i \in \mathcal{B}$. The *weakly active* constraints are identified by the indices $i \in \mathcal{B}$ for which $\lambda_i^* = 0$ for all λ^* satisfying (3). Conversely, the *strongly active* constraints are those for which $\lambda_i^* > 0$ for at least one multiplier λ^* satisfying (3). The strict complementarity condition holds at z^* if there are no weakly active constraints.

We illustrate these definitions with two simple examples. The two-variable problem

$$\min \tfrac{1}{2}(z_1+1)^2 + \tfrac{1}{2}z_2^2 \text{ subject to } \begin{bmatrix} -z_1 \\ (z_1-1)^2 + z_2^2 \end{bmatrix} \le 0,$$

has a unique solution at $z^* = (0,0)^T$, with Lagrange multipliers

$$\lambda^* = \{(1-2\alpha, \alpha)^T \mid \alpha \in [0, .5]\}.$$

The two constraints are linearly dependent at the solution, but both are strongly active. If we consider the same objective function, but replace the constraints by

$$\begin{bmatrix} -z_1 \\ -z_1 - z_2 \end{bmatrix} \le 0,$$

we find that once again both constraints are active, but the multiplier λ^* now has the unique value $(1,0)^T$. For this modified problem, the first constraint is strongly active and the second is weakly active.

We are interested in degenerate problems, those for which the active constraint gradients at the solution are linearly dependent or the strict complementarity condition fails to hold (or both). The first part of our paper describes a technique for partitioning $\mathcal{B}$ into weakly active and strongly active indices. Section 3 builds on the technique described by Facchinei, Fischer, and Kanzow [5] for identifying $\mathcal{B}$. Our technique requires the solution of a sequence of closely related linear programming subproblems in which the set of strongly active indices is assembled progressively. Solution of one additional linear program yields a Lagrange multiplier estimate λ such that the components λ_i for all strongly active indices i are bounded below by a positive constant.

In the second part of the paper, we use the cited technique to adjust the Lagrange multiplier estimate between iterations of the stabilized sequential quadratic programming (sSQP) algorithm described by Wright [18] and Hager [8]. The resulting technique has the advantage that it converges superlinearly under weaker conditions than considered in these earlier papers. We can drop the assumption of strict complementarity and a "sufficiently interior" starting point made in [18], and we do not need the stronger second-order conditions of [8]. Motivation for the sSQP approach came from work on primal-dual interior-point algorithms described in [20,13]. It is also closely related to the method of multipliers and the "recursive successive quadratic programming" approach of Bartholomew-Biggs [2]. (See Wright [17, Section 6] for a discussion of the similarities.)

Other work on stabilization of the SQP approach to yield superlinear convergence under weakened conditions has been performed by Fischer [6] and Wright [17]. Fischer proposed an algorithm in which an additional quadratic program is solved between iterations of SQP in order to adjust the Lagrange multiplier estimate. He proved superlinear convergence under conditions that are weaker than the standard assumptions but stronger than the ones made in this paper. Wright described superlinear local convergence properties of a class of inexact SQP methods and showed that sSQP and Fischer's method could be expressed as members of this class. This paper also introduced a modification of standard SQP that enforced only a subset of the linearized constraints—those in a "strictly active working set"—and permitted slight violations of the nonenforced constraints yet achieved superlinear convergence under weaker-than-usual conditions.

Bonnans [3] showed that when strict complementarity fails to hold but the active constraint gradients are linearly independent, then the standard SQP algorithm (in which any nonuniqueness in the solution of the SQP subproblem is resolved by taking the solution of minimum norm) converges superlinearly.

Our concern here is with *local* behavior, so we assume availability of a starting point (z^0, λ^0) that is "sufficiently close" to the optimal primal-dual set. We believe, however, that ingredients of the approach proposed here can be embedded in practical algorithms, such as SQP algorithms that include modifications (merit functions and filters) to ensure global convergence. We believe also that this approach could be used to enhance the robustness and convergence rate of other types of algorithms, including augmented Lagrangian and interior-point algorithms, in problems in which there is degeneracy at the solution. We mention one such extension in Section 7.

2. Assumptions, Notation, and Basic Results

We now review the optimality conditions for (1) and outline the assumptions that are used in subsequent sections. These include the second-order sufficient condition we use here, the Mangasarian-Fromovitz constraint qualification, and the definition of weakly-active indices.

Recall the KKT conditions (3). The set of "optimal" Lagrange multipliers λ^* is denoted by $\mathcal{S}_\lambda$, and the primal-dual optimal set is denoted by $\mathcal{S}$. Specifically, we have

$$\mathcal{S}_\lambda = \{\lambda^* \mid \lambda^* \text{ satisfies (3)}\}, \qquad \mathcal{S} = \{z^*\} \times \mathcal{S}_\lambda. \tag{5}$$

An alternative, compact form of the KKT conditions is the following variational inequality formulation:

$$\begin{bmatrix} \nabla\phi(z^*) + \nabla g(z^*)\lambda^* \\ g(z^*) \end{bmatrix} \in \begin{bmatrix} 0 \\ N(\lambda^*) \end{bmatrix}, \tag{6}$$

where $N(\lambda)$ is the set defined by

$$N(\lambda) \stackrel{\text{def}}{=} \begin{cases} \{y \mid y \le 0 \text{ and } y^T\lambda = 0\} & \text{if } \lambda \ge 0, \\ \emptyset & \text{otherwise.} \end{cases} \tag{7}$$

We now introduce notation for subsets of the set $\mathcal{B}$ of active constraint indices at z^*, defined in (4). For any optimal multiplier $\lambda^* \in \mathcal{S}_\lambda$, we define the set $\mathcal{B}_+(\lambda^*)$ to be the "support" of λ^*, that is,

$$\mathcal{B}_+(\lambda^*) = \{i \in \mathcal{B} \mid \; \lambda_i^* > 0\}.$$

We define $\mathcal{B}_+$ (without argument) as

$$\mathcal{B}_+ \stackrel{\text{def}}{=} \cup_{\lambda^* \in \mathcal{S}_\lambda} \mathcal{B}_+(\lambda^*); \tag{8}$$

this set contains the indices of the *strongly active* constraints. Its complement in $\mathcal{B}$ is denoted by $\mathcal{B}_0$, that is,

$$\mathcal{B}_0 \stackrel{\text{def}}{=} \mathcal{B} \backslash \mathcal{B}_+.$$

This set $\mathcal{B}_0$ contains the *weakly active* constraint indices, those indices $i \in \mathcal{B}$ such that $\lambda_i^* = 0$ for all $\lambda^* \in \mathcal{S}_\lambda$. In later sections, we make use of the quantity ϵ_λ defined by

$$\epsilon_\lambda \stackrel{\text{def}}{=} \max_{\lambda^* \in \mathcal{S}_\lambda} \min_{i \in \mathcal{B}_+} \lambda_i^*. \tag{9}$$

Note by the definition of $\mathcal{B}_+$ that $\epsilon_\lambda > 0$.

The Mangasarian-Fromovitz constraint qualification (MFCQ) [12] holds at z^* if there is a vector $\bar{y} \in \mathsf{R}^n$ such that

$$\nabla g_i(z^*)^T \bar{y} < 0 \quad \text{ for all } i \in \mathcal{B}. \tag{10}$$

By defining $\nabla g_{\mathcal{B}}$ to be the $n \times |\mathcal{B}|$ matrix whose rows are $\nabla g_i(\cdot)$, $i \in \mathcal{B}$, we can write this condition alternatively as

$$\nabla g_{\mathcal{B}}(z^*)^T \bar{y} < 0. \tag{11}$$

It is well known that MFCQ is equivalent to nonemptiness and boundedness of the set $\mathcal{S}_\lambda$; see Gauvin [7].

Since $\mathcal{S}_\lambda$ is defined by the linear conditions $\nabla \phi(z^*) + \nabla g(z^*)\lambda^* = 0$ and $\lambda^* \geq 0$, it is closed and convex. Therefore, under MFCQ, it is also compact.

We assume throughout that the following second-order condition is satisfied: there is a scalar $\upsilon > 0$ such that

$$w^T \nabla_{zz} \mathcal{L}(z^*, \lambda^*) w \geq \upsilon \|w\|^2, \quad \text{ for all } \lambda^* \in \mathcal{S}_\lambda, \tag{12}$$

and for all w such that

$$\begin{array}{l} \nabla g_i(z^*)^T w = 0, \text{ for all } i \in \mathcal{B}_+, \\ \nabla g_i(z^*)^T w \leq 0, \text{ for all } i \in \mathcal{B}_0. \end{array} \tag{13}$$

This condition is referred to as Condition 2s.1 in [17, Section 3]. Weaker second-order conditions, stated in terms of a quadratic growth condition of the objective $\phi(z)$ in a feasible neighborhood of z^*, are discussed by Bonnans and Ioffe [4] and Anitescu [1].

Our standing assumption for this paper is as follows.

Assumption 1. *The first-order conditions (3), the MFCQ (11), and the second-order condition (12), (13) are satisfied at z^*. Moreover, the functions ϕ and g are twice Lipschitz continuously differentiable in a neighborhood of z^*.*

In the following result, our claim that z^* is a strict local minimizer means that there exists a neighborhood of z^* such that $f(z^*) \le f(z)$ for all z in this neighborhood with $g(z) \le 0$, and that this inequality is strict if $z \ne z^*$.

Theorem 1. *Suppose that Assumption 1 holds. Then z^* is an isolated stationary point and a strict local minimizer of (1).*

Proof. See Robinson [14, Theorems 2.2 and 2.4].

We define the distance of a vector $x \in \mathbb{R}^r$ to a set $X \in \mathbb{R}^r$ by

$$\mathrm{dist}\,(x, X) = \inf_{\bar{x} \in X} \|\bar{x} - x\|,$$

where here and elsewhere, $\|\cdot\|$ denotes the Euclidean norm unless a subscript specifically indicates otherwise. We use the notation $\delta(\cdot)$ to denote distances from the primal, dual, and primal-dual optimal sets, according to context. Specifically, we define

$$\delta(z) \stackrel{\text{def}}{=} \|z - z^*\|, \quad \delta(\lambda) \stackrel{\text{def}}{=} \mathrm{dist}\,(\lambda, \mathcal{S}_\lambda), \quad \delta(z,\lambda) \stackrel{\text{def}}{=} \mathrm{dist}\,((z,\lambda), \mathcal{S}). \tag{14}$$

We also use $P(\lambda)$ to denote the projection of λ onto $\mathcal{S}_\lambda$; that is, we have $P(\lambda) \in \mathcal{S}_\lambda$ and $\|P(\lambda) - \lambda\| = \mathrm{dist}\,(\lambda, \mathcal{S}_\lambda)$. Note that from (14) we have $\delta(z)^2 + \delta(\lambda)^2 = \delta(z,\lambda)^2$, and therefore

$$\delta(z) \le \delta(z,\lambda), \qquad \delta(\lambda) \le \delta(z,\lambda). \tag{15}$$

Using Assumption 1, we can prove the following result, which gives a practical way to estimate the distance $\delta(z,\lambda)$ of (z,λ) to the primal-dual solution set $\mathcal{S}$. (We use $\min(\lambda, -g(z))$ here to denote the vector whose ith component is $\min(\lambda_i, -g_i(z))$.)

Theorem 2. *Suppose that Assumption 1 holds. Then there are positive constants δ_0, κ_0, and κ_1 such that for all (z,λ) with $\delta(z,\lambda) \le \delta_0$, the quantity $\eta(z,\lambda)$ defined by*

$$\eta(z,\lambda) \stackrel{\text{def}}{=} \left\| \begin{bmatrix} \nabla_z \mathcal{L}(z,\lambda) \\ \min(\lambda, -g(z)) \end{bmatrix} \right\| \tag{16}$$

satisfies

$$\kappa_0 \delta(z,\lambda) \le \eta(z,\lambda) \le \kappa_1 \delta(z,\lambda).$$

See Facchinei, Fischer, and Kanzow [5, Theorem 3.6], Wright [17, Theorem A.1], and Hager and Gowda [9, Lemma 2] for proofs of this result. (The second-order condition is stated in a slightly different fashion in [5] but is equivalent to (12), (13).)

We use order notation in the following (fairly standard) way: If two matrix, vector, or scalar quantities M and A are functions of a common quantity, we write $M = O(\|A\|)$ if there is a constant β such that $\|M\| \le \beta\|A\|$ whenever $\|A\|$ is sufficiently small. We write $M = \Omega(\|A\|)$ if there is a constant β such that $\|M\| \ge \beta^{-1}\|A\|$ whenever $\|A\|$ sufficiently small, and $M = \Theta(\|A\|)$ if both $M = O(\|A\|)$ and $M = \Omega(\|A\|)$. We write $M = o(\|A\|)$ if for all sequences $\{A_k\}$ with $\|A_k\| \to 0$, the corresponding sequence $\{M_k\}$ satisfies $\|M_k\|/\|A_k\| \to 0$. By using this notation, we can rewrite the conclusion of Theorem 2 as follows:

$$\eta(z,\lambda) = \Theta(\delta(z,\lambda)). \tag{17}$$

3. Detecting Active Constraints

We now describe a procedure, named Procedure ID0, for identifying those inequality constraints that are active at the solution, and classifying them according to whether they are weakly active or strongly active. We prove that Procedure ID0 classifies the indices correctly given a point (z,λ) sufficiently close to the primal-dual optimal set $\mathcal{S}$. Finally, we describe some implementation issues for this procedure.

3.1. The Detection Procedure

Facchinei, Fischer, and Kanzow [5] showed that the function $\eta(z,\lambda)$ defined in (17) can be used as the basis of a scheme for identifying the active set $\mathcal{B}$. Following their approach, we choose some scalar $\tau \in (0,1)$ and use the quantity $\eta(z,\lambda)^\tau$ as a threshold for determining whether a constraint is active:

$$\mathcal{A}(z,\lambda) \stackrel{\text{def}}{=} \{i = 1,2,\ldots,m \,|\, g_i(z) \ge -\eta(z,\lambda)^\tau\}. \tag{18}$$

Theorem 3. *Suppose that Assumption 1 holds. Then there exists $\delta_1 \in (0,\delta_0]$ such that for all (z,λ) with $\delta(z,\lambda) \le \delta_1$, we have $\mathcal{A}(z,\lambda) = \mathcal{B}$.*

Proof. The result follows immediately from [5, Definition 2.1, Theorem 2.3] and Theorem 2 above.

A scheme for estimating $\mathcal{B}_+$ (hence, $\mathcal{B}_0$) is described in [5], but it requires the strict MFCQ condition to hold, which implies that $\mathcal{S}_\lambda$ is a singleton. Here we describe a more complicated scheme for estimating $\mathcal{B}_+$ that requires only the conditions of Theorem 3 to hold.

Our scheme is based on linear programming subproblems of the following form. For a given point (z,λ) and a given parameter $\tau \in (0,1)$, we define the "tolerance" χ as follows:

$$\chi(z,\lambda,\tau) = \max\left(\eta(z,\lambda)^\tau, \left\|\nabla\phi(z) + \sum_{i\in\mathcal{A}(z,\lambda)} \lambda_i \nabla g_i(z)\right\|_\infty\right). \tag{19}$$

Then, for given sets $\hat{\mathcal{A}} \subset \mathcal{A}(z,\lambda)$, we solve linear programs of the following form:

$$\max_{\tilde{\lambda}} \textstyle\sum_{i\in\hat{\mathcal{A}}} \tilde{\lambda}_i \text{ subject to} \tag{20a}$$

$$\left\|\nabla\phi(z) + \textstyle\sum_{i\in\mathcal{A}(z,\lambda)} \tilde{\lambda}_i \nabla g_i(z)\right\|_\infty \le \chi(z,\lambda,\tau) \tag{20b}$$

$$\tilde{\lambda}_i \ge 0, \text{ for all } i \in \mathcal{A}(z,\lambda); \quad \tilde{\lambda}_i = 0 \text{ otherwise.} \tag{20c}$$

We make several observations about this problem. First, a linear program is obtained by replacing the constraint (20b) by two-sided bounds. Second, λ itself is a feasible point for the problem provided that it satisfies the conditions (20c); the $\|\cdot\|_\infty$ term in $\chi(z,\lambda,\tau)$ was chosen to ensure this. Third, the objective function involves elements $\tilde{\lambda}_i$ only for indices i in the subset $\hat{\mathcal{A}}$, whereas the $\tilde{\lambda}_i$ are permitted to be nonzero for all $i \in \mathcal{A}(z,\lambda)$. The idea is that $\hat{\mathcal{A}}$ contains those indices that *may* belong to $\mathcal{B}_0$; by the time we solve (20), we have already decided that the other indices $i \in \mathcal{A}(z,\lambda)\backslash\hat{\mathcal{A}}$ probably belong to $\mathcal{B}_+$.

The complete procedure is as follows.

Procedure ID0
Given constants τ and $\hat{\tau}$ satisfying $0 < \hat{\tau} < \tau < 1$, and point (z,λ);
Compute $\xi = \max\left(\eta(z,\lambda)^{\hat{\tau}}, \eta(z,\lambda)^{\tau}, \left\|\nabla\phi(z) + \sum_{i\in\mathcal{A}(z,\lambda)} \lambda_i \nabla g_i(z)\right\|_\infty\right)$;
Evaluate $\eta(z,\lambda)$ from (16), $\mathcal{A}(z,\lambda)$ from (18), and $\chi(z,\lambda,\tau)$ from (19);
Define $\hat{\mathcal{A}}_{\text{init}} = \mathcal{A}(z,\lambda)\backslash\{i \mid \lambda_i \ge \xi\}$;
$\hat{\mathcal{A}} \leftarrow \hat{\mathcal{A}}_{\text{init}}$;
repeat
  if $\hat{\mathcal{A}} = \emptyset$
    stop with $\mathcal{A}_0 = \emptyset$, $\mathcal{A}_+ = \mathcal{A}(z,\lambda)$;
  end(if)
  solve (20) to find $\tilde{\lambda}$;
  set $\mathcal{C} = \{i \in \hat{\mathcal{A}} \mid \tilde{\lambda}_i \ge \xi\}$;
  if $\mathcal{C} = \emptyset$
    stop with $\mathcal{A}_0 = \hat{\mathcal{A}}$, $\mathcal{A}_+ = \mathcal{A}(z,\lambda)\backslash\hat{\mathcal{A}}$;
  else
    set $\hat{\mathcal{A}} \leftarrow \hat{\mathcal{A}}\backslash\mathcal{C}$;
  end(if)
end(repeat)

This procedure terminates finitely; in fact, the number of times that (20) is solved in the "repeat" loop is bounded by the cardinality of $\hat{\mathcal{A}}_{\text{init}}$.

We first prove some elementary results about the quantities χ and ξ.

Lemma 1. *We can choose $\bar{\delta}_1 \in (0,\delta_1]$, where δ_1 is defined in Theorem 3, in such a way that $\chi(z,\lambda,\tau) = \eta(z,\lambda)^\tau$ and $\xi = \eta(z,\lambda)^{\hat{\tau}}$ for all (z,λ) with $\delta(z,\lambda) \le \bar{\delta}_1$.*

Proof. By choosing $\bar{\delta}_1$ small enough that $\eta(z,\lambda) < 1$ whenever $\delta(z,\lambda) \le \bar{\delta}_1$, we have from $\hat{\tau} < \tau$ that $\eta(z,\lambda)^{\hat{\tau}} > \eta(z,\lambda)^{\tau}$.

From Theorem 3, we have that $\mathcal{A}(z,\lambda) = \mathcal{B}$, so using the fact that $\lambda_i = \lambda_i - \lambda_i^* = O(\delta(z,\lambda))$ whenever $\lambda^* \in \mathcal{S}_\lambda$ and $i \notin \mathcal{B}$, and using Theorem 2, we have

$$\begin{aligned}\left\| \nabla\phi(z) + \sum_{i\in\mathcal{A}(z,\lambda)} \lambda_i \nabla g_i(z) \right\|_\infty &= \left\| \nabla\phi(z) + \sum_{i\in\mathcal{B}} \lambda_i \nabla g_i(z) \right\|_\infty \\ &= \left\| \nabla_z \mathcal{L}(z,\lambda) - \sum_{i\notin\mathcal{B}} \lambda_i \nabla g_i(z) \right\|_\infty \\ &= O(\eta(z,\lambda)) + O(\delta(z,\lambda))\end{aligned}$$

Hence, by decreasing $\bar{\delta}_1$ if necessary, and using Theorem 2 and the fact that $\tau \in (0,1)$, we have that

$$\eta(z,\lambda)^{\tau} > \left\| \nabla\phi(z) + \sum_{i\in\mathcal{A}(z,\lambda)} \lambda_i \nabla g_i(z) \right\|_\infty$$

whenever $\delta(z,\lambda) \le \bar{\delta}_1$.

Together, these inequalities yield the desired results.

We prove that Procedure ID0 successfully identifies $\mathcal{B}_+$ (for all $\delta(z,\lambda)$ sufficiently small) in several steps, culminating in Theorem 4. First, we estimate the distance of $(z,\tilde{\lambda})$ to the solution set $\mathcal{S}$, where $\tilde{\lambda}$ is the solution of (20) for some $\hat{\mathcal{A}}$.

Lemma 2. *Suppose that Assumption 1 holds. Then there are positive constants δ_2 and κ_2 such that whenever $\delta(z,\lambda) \le \delta_2$, any feasible point $\tilde{\lambda}$ of (20) at any iteration of Procedure ID0 satisfies*

$$\delta(z,\tilde{\lambda}) \le \kappa_2 \delta(z,\lambda)^{\tau}.$$

Proof. Initially choose $\delta_2 = \bar{\delta}_1$ for $\bar{\delta}_1$ defined in Lemma 1, so that $\mathcal{A}(z,\lambda) = \mathcal{B}$ and $\xi = \eta(z,\lambda)^{\hat{\tau}}$. Hence, we have $\hat{\mathcal{A}} \subset \mathcal{B}$ at all iterations of Procedure ID0.

We now show that by reducing δ_2 if necessary, we can ensure that $\delta(z,\tilde{\lambda}) \le \delta_0$, and hence that Theorem 2 can be applied to $(z,\tilde{\lambda})$. Assume for contradiction that it is not possible to choose such a δ_2. Then there exists a sequence (z^k,λ^k) with $\delta(z^k,\lambda^k) \to 0$ such that for some feasible point $\tilde{\lambda}^k$ of (20) with $(z,\lambda) = (z^k,\lambda^k)$, we have $\delta(z^k,\tilde{\lambda}^k) > \delta_0$. Suppose first that a subsequence of $\{\tilde{\lambda}^k\}$ is bounded. Then without loss of generality, and by taking limits in (20b), (20c), we can define a point $\tilde{\lambda}'$ such that $\tilde{\lambda}^k \to \tilde{\lambda}'$ and the following conditions hold:

$$\tilde{\lambda}_i' \ge 0, \quad \text{for all } i \in \mathcal{B}, \quad \tilde{\lambda}_i' = 0, \quad \text{otherwise}, \quad \nabla\phi(z^*) + \sum_{i\in\mathcal{B}} \tilde{\lambda}_i' \nabla g_i(z^*) = 0.$$

Since these conditions imply that $\tilde{\lambda}' \in \mathcal{S}_\lambda$, we have that

$$\delta_0 \le \delta(z^k, \tilde{\lambda}^k) \to \delta(z^*, \tilde{\lambda}') = 0,$$

giving a contradiction. Therefore the sequence $\{\tilde{\lambda}^k\}$ must be unbounded, so without loss of generality we can define a point $\tilde{\lambda}''$ such that $\tilde{\lambda}^k / \|\tilde{\lambda}^k\| \to \tilde{\lambda}''$ and

$$\|\tilde{\lambda}''\| = 1, \quad \tilde{\lambda}''_i \ge 0 \text{ for all } i \in \mathcal{B}, \quad \tilde{\lambda}''_i = 0 \text{ otherwise}. \tag{21}$$

By dividing (20b) with $(z, \lambda) = (z^k, \lambda^k)$ and $\tilde{\lambda} = \tilde{\lambda}^k$ by $\|\tilde{\lambda}^k\|$ and taking limits, we have also that

$$\sum_{i \in \mathcal{B}} \tilde{\lambda}''_i \nabla g_i(z^*) = 0.$$

However by taking an inner product of this expression with the vector $\bar{y}$ from (10), we obtain a contradiction with the conditions (21). We conclude that it is possible to choose δ_2 sufficiently small to ensure that $\delta(z, \lambda) \le \delta_2 \Rightarrow \delta(z, \tilde{\lambda}) \le \delta_0$.

We now estimate $\eta(z, \tilde{\lambda})$ using the definition (16). We have directly from the constraints (20b) and Lemma 1 that

$$\|\nabla_z \mathcal{L}(z, \tilde{\lambda})\|_\infty \le \eta(z, \lambda)^\tau.$$

For the vector $\min(\tilde{\lambda}, -g(z))$, we have for $i \in \mathcal{B}$ that $g_i(z^*) = 0$ and $\tilde{\lambda}_i \ge 0$, and so

$$i \in \mathcal{B} \;\Rightarrow\; |\min(\tilde{\lambda}_i, -g_i(z))| \le |g_i(z)| = O(\|z - z^*\|) = O(\delta(z, \lambda)).$$

Meanwhile for $i \notin \mathcal{B} = \mathcal{A}(z, \lambda)$, we have $\tilde{\lambda}_i = 0$ and $g_i(z^*) < 0$, and so

$$i \notin \mathcal{B} \;\Rightarrow\; |\min(\tilde{\lambda}_i, -g_i(z))| = \max(0, g_i(z)) \le |g_i(z) - g_i(z^*)| = O(\delta(z, \lambda)).$$

By substituting these estimates into (16), and using the equivalence of $\|\cdot\|_\infty$ and the Euclidean norm and the result of Theorem 2, we have that there is a constant $\bar{\kappa}_2 > 0$ such that

$$\eta(z, \tilde{\lambda}) \le \bar{\kappa}_2 \delta(z, \lambda)^\tau.$$

Using Theorem 2 again, we have

$$\delta(z, \tilde{\lambda}) \le \kappa_0^{-1} \eta(z, \tilde{\lambda}) \le \kappa_0^{-1} \bar{\kappa}_2 \delta(z, \lambda)^\tau, \tag{22}$$

giving the result.

In the next two lemmas and Theorem 4, we show that for $\delta(z, \lambda)$ sufficiently small, Procedure ID0 terminates with $\mathcal{A}_0 = \mathcal{B}_0$ and $\mathcal{A}_+ = \mathcal{B}_+$.

Lemma 3. *Suppose that Assumption 1 holds. Then there is $\delta_3 > 0$ such that whenever $\delta(z, \lambda) \le \delta_3$, Procedure ID0 terminates with $\mathcal{B}_0 \subset \mathcal{A}_0$.*

Proof. Since we know the procedure terminates finitely, we need show only that $\mathcal{B}_0 \subset \hat{\mathcal{A}}$ at all iterations of the procedure. Initially set $\delta_3 = \delta_2 \le \delta$, so that $\mathcal{A}(z,\lambda) = \mathcal{B}$ and the result of Lemma 2 holds. Suppose for contradiction there is an index $j \in \mathcal{B}_0$ such that j either is not included in the initial index set $\hat{\mathcal{A}}_{\rm init}$ or else is deleted from $\hat{\mathcal{A}}$ at some iteration of Procedure ID0.

Suppose first that j is not included in $\hat{\mathcal{A}}_{\rm init}$. Then we must have $\lambda_j > \eta(z,\lambda)^{\hat{\tau}}$, which by Theorem 2 implies that

$$\delta(z,\lambda) \ge |\lambda_j| \ge \eta(z,\lambda)^{\hat{\tau}} \ge \kappa_0^{\hat{\tau}} \delta(z,\lambda)^{\hat{\tau}}. \tag{23}$$

However, by decreasing δ_3 and using $\hat{\tau} \in (0,1)$, we can ensure that (23) does not hold whenever $\delta(z,\lambda) \le \delta_3$. Hence, j is included in $\hat{\mathcal{A}}_{\rm init}$.

Suppose now that $j \in \mathcal{B}_0$ is deleted from $\hat{\mathcal{A}}$ at some subsequent iteration. For this to happen, the subproblem (20) must have a solution $\tilde{\lambda}$ with

$$\tilde{\lambda}_j > \eta(z,\lambda)^{\hat{\tau}} \tag{24}$$

for some $\hat{\mathcal{A}} \subset \mathcal{B}$. Hence from Theorem 2, we have that

$$\delta(z,\tilde{\lambda}) \ge \tilde{\lambda}_j > \eta(z,\lambda)^{\hat{\tau}} \ge \kappa_0^{\hat{\tau}} \delta(z,\lambda)^{\hat{\tau}}. \tag{25}$$

By combining the result of Lemma 2 with (25), we have that

$$\kappa_2 \delta(z,\lambda)^{\tau} \ge \kappa_0^{\hat{\tau}} \delta(z,\lambda)^{\hat{\tau}}.$$

However, this inequality cannot hold when $\delta(z,\lambda)$ is smaller than $(\kappa_0^{\hat{\tau}}\kappa_2^{-1})^{1/(\tau-\hat{\tau})}$. Therefore, by decreasing δ_3 if necessary, we have a contradiction in this case also.

Lemma 4. *Suppose that Assumption 1 holds. Then there is $\delta_4 > 0$ such that whenever $\delta(z,\lambda) \le \delta_4$, Procedure ID0 terminates with $\mathcal{B}_+ \subset \mathcal{A}_+$.*

Proof. Given any $j \in \mathcal{B}_+$, we have for sufficiently small choice of δ_4 that $j \in \mathcal{A}(z,\lambda)$. We prove the result by showing that Procedure ID0 cannot terminate with $j \in \mathcal{A}_0$.

We initially set $\delta_4 = \delta_3$, where δ_3 is the constant from Lemma 3. (We reduce it as necessary, but maintain $\delta_4 > 0$, in the course of the proof.) For contradiction, assume that there is $j \in \mathcal{B}_+$ such that $j \in \hat{\mathcal{A}}$ at all iterations of Procedure ID0, including the iteration on which the procedure terminates and sets $\mathcal{A}_0 = \hat{\mathcal{A}}$. Recalling the definition (9) of ϵ_λ, we use compactness of $\mathcal{S}_\lambda$ to choose $\lambda^* \in \mathcal{S}_\lambda$ such that $\epsilon_\lambda = \min_{i \in \mathcal{B}_+} \lambda_i^*$. In particular, we have

$$\lambda_j^* \ge \epsilon_\lambda > 0$$

for our chosen index j. We claim that, by reducing δ_4 if necessary, we can ensure that λ^* is feasible for (20) whenever $\delta(z,\lambda) \le \delta_4$. Obviously, since $\mathcal{A}(z,\lambda) = \mathcal{B}$ by Theorem 3, λ^* is feasible with respect to (20c). Since $\lambda^* \in \mathcal{S}_\lambda$ and

$$\|z - z^*\| \le \delta(z,\lambda) \le \kappa_0^{-1} \eta(z,\lambda),$$

we have

$$\left\| \nabla \phi(z) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(z) \right\|_\infty = \left\| \nabla \phi(z) - \nabla \phi(z^*) + \sum_{i=1}^{m} \lambda_i^* (\nabla g_i(z) - \nabla g_i(z^*)) \right\|_\infty$$
$$\leq M\|z - z^*\| \leq M\kappa_0^{-1}\eta(z, \lambda), \tag{26}$$

for some constant M that depends on the norms of $\nabla^2 \phi(\cdot)$ and $\nabla^2 g_i(\cdot)$, $i \in \mathcal{B}_+$ in the neighborhood of z^* and on a bound on the set $\mathcal{S}_\lambda$ (which *is* bounded, because of MFCQ). Since $\tau < 1$ and since $\eta(z, \lambda) = \Theta(\delta(z, \lambda))$, we can reduce δ_4 if necessary to ensure that

$$M\kappa_0^{-1}\eta(z, \lambda) < \eta(z, \lambda)^\tau$$

whenever $\delta(z, \lambda) \leq \delta_4$, thereby ensuring that the constraints (20b) are satisfied by λ^*.

Since λ^* is feasible for (20), a lower bound on the optimal objective is

$$\sum_{i \in \hat{\mathcal{A}}} \lambda_i^* \geq \lambda_j^* \geq \epsilon_\lambda.$$

However, since Procedure ID0 terminates with $j \in \hat{\mathcal{A}}$, we must have that $\mathcal{C} = \emptyset$ for the solution $\tilde{\lambda}$ of (20) with this particular choice of $\hat{\mathcal{A}}$. But we can have $\mathcal{C} = \emptyset$ only if $\tilde{\lambda}_i < \eta(z, \lambda)^{\hat{\tau}}$ for all $i \in \hat{\mathcal{A}}$, which means that the optimal objective is no greater than $m\eta(z, \lambda)^{\hat{\tau}}$. But since $\eta(z, \lambda) = \Theta(\delta(z, \lambda))$, we can reduce δ_4 if necessary to ensure that

$$m\eta(z, \lambda)^{\hat{\tau}} < \epsilon_\lambda$$

whenever $\delta(z, \lambda) \leq \delta_4$. This gives a contradiction, so that $\mathcal{A}_0$ (which is set by Procedure ID0 to the final $\hat{\mathcal{A}}$) can contain no indices $j \in \mathcal{B}_+$. Since $\mathcal{B}_+ \subset \mathcal{B} = \mathcal{A}(z, \lambda)$ whenever $\delta(z, \lambda) \leq \delta_4$, we must therefore have $\mathcal{B}_+ \subset \mathcal{A}_+$, as claimed.

By using the quantity δ_4 from Lemma 4, we combine this result with Theorem 3 and Lemma 3 to obtain the following theorem.

Theorem 4. *Suppose that Assumption 1 holds. Then there is* $\delta_4 > 0$ *such that whenever* $\delta(z, \lambda) \leq \delta_4$*, Procedure ID0 terminates with* $\mathcal{A}_+ = \mathcal{B}_+$ *and* $\mathcal{A}_0 = \mathcal{B}_0$.

3.2. Scheme for Finding an Interior Multiplier Estimate

We now describe a scheme for finding a vector $\hat{\lambda}$ that is close to $\mathcal{S}_\lambda$ but not too close to the relative boundary of this set. In other words, the quantity $\min_{i \in \mathcal{B}_+} \hat{\lambda}_i$ is not too far from its maximum achievable value ϵ_λ.

We find $\hat{\lambda}$ by solving a linear programming problem similar to (20) but containing an extra variable to represent $\min_{i \in \mathcal{B}_+} \hat{\lambda}_i$. We state this problem as follows:

$$\max_{\hat{t},\hat{\lambda}} \hat{t} \text{ subject to} \tag{27a}$$

$$\hat{t} \le \hat{\lambda}_i, \quad \text{for all } i \in \mathcal{A}_+, \tag{27b}$$

$$-\eta(z,\lambda)^\tau e \le \nabla\phi(z) + \textstyle\sum_{i\in\mathcal{A}_+} \hat{\lambda}_i \nabla g_i(z) \le \eta(z,\lambda)^\tau e \tag{27c}$$

$$\hat{\lambda}_i \ge 0, \quad \text{for all } i \in \mathcal{A}_+; \quad \hat{\lambda}_i = 0 \text{ otherwise.} \tag{27d}$$

Theorem 5. *Suppose that Assumption 1 holds. Then there is a positive number δ_5 such that (27) is feasible and bounded whenever $\delta(z,\lambda) \le \delta_5$, and its optimal objective is at least ϵ_λ (for ϵ_λ defined in (9)). Moreover, there is a constant $\beta' > 0$ such that $\delta(z,\hat{\lambda}) \le \beta'\delta(z,\lambda)^\tau$.*

Proof. Let $\lambda^* \in \mathcal{S}_\lambda$ be chosen so that $\epsilon_\lambda = \min_{i\in\mathcal{B}_+} \lambda_i^*$. We show first that $(\hat{t},\hat{\lambda}) = (\epsilon_\lambda, \lambda^*)$ is feasible for (27), thereby proving that this linear program is feasible and that the optimum objective value is at least ϵ_λ.

Initially we set $\delta_5 = \delta_4$. By Definition (9), the constraint (27b) is satisfied by $(\hat{t},\hat{\lambda}) = (\epsilon_\lambda, \lambda^*)$. Since $\delta(z,\lambda) \le \delta_5 = \delta_4$, we have from Theorem 4 that $\mathcal{A}_+ = \mathcal{B}_+$, so that (27d) also holds. Satisfaction of (27c) follows from (26), by choice of δ_4. Moreover, it is clear from $\mathcal{A}_+ = \mathcal{B}_+$ that the optimal $(\hat{t},\hat{\lambda})$ will satisfy $\hat{t} = \min_{i\in\mathcal{B}_+} \hat{\lambda}_i$.

We now show that the problem (27) is bounded for $\delta(z,\lambda)$ sufficiently small. Let $\bar{y}$ be the vector in (11), and decrease δ_5 if necessary so that we can choose a number $\zeta > 0$ such that

$$\delta(z,\lambda) \le \delta_5 \quad\Rightarrow\quad \bar{y}^T \nabla g_i(z) \le -\zeta, \quad \text{for all } i \in \mathcal{A}_+ = \mathcal{B}_+. \tag{28}$$

From the constraints (27c) and the triangle inequality, we have that

$$\begin{aligned}
\left\| \sum_{i\in\mathcal{A}_+} \hat{\lambda}_i \bar{y}^T \nabla g_i(z) \right\|_1 &\le \|\bar{y}^T \nabla\phi(z)\|_1 + \left\| \bar{y}^T\nabla\phi(z) + \sum_{i\in\mathcal{A}_+} \hat{\lambda}_i \bar{y}^T \nabla g_i(z) \right\|_1 \\
&\le \|\bar{y}\|_1 \|\nabla\phi(z)\|_\infty + \|\bar{y}\|_1 \left\| \nabla\phi(z) + \sum_{i\in\mathcal{A}_+} \hat{\lambda}_i \nabla g_i(z) \right\|_\infty \\
&\le \|\bar{y}\|_1 \|\nabla\phi(z)\|_\infty + \|\bar{y}\|_1 \eta(z,\lambda)^\tau.
\end{aligned}$$

However, from (28) and $\hat{\lambda}_i \ge 0$, $i \in \mathcal{A}_+$, we have that

$$\left\| \sum_{i\in\mathcal{A}_+} \hat{\lambda}_i \bar{y}^T \nabla g_i(z) \right\|_1 \ge \left\| \hat{\lambda}_{\mathcal{A}_+} \right\|_1 \zeta.$$

By combining these bounds, we obtain that

$$\left\| \hat{\lambda}_{\mathcal{A}_+} \right\|_1 \le \zeta^{-1} \|\bar{y}\|_1 \left[\|\nabla\phi(z)\|_\infty + \eta(z,\lambda)^\tau \right],$$

whenever $\delta(z,\lambda) \le \delta_5$, so that the feasible region for (27) is bounded, as claimed.

To prove our final claim that $\delta(z,\hat{\lambda}) \le \beta' \delta(z,\lambda)^\tau$ for some $\beta' > 0$, we use Theorem 2. We have from (27c) and the cited theorem that

$$\left\| \nabla_z \mathcal{L}(z,\hat{\lambda}) \right\|_\infty \le \eta(z,\lambda)^\tau \le \kappa_1^\tau \delta(z,\lambda)^\tau.$$

For $i \in \mathcal{A}_+ = \mathcal{B}_+$, we have from $\hat{\lambda}_i \ge \epsilon_\lambda$ and $g_i(z^*) = 0$ that

$$\begin{aligned} i \in \mathcal{A}_+ \quad \Rightarrow \quad \left| \min(\hat{\lambda}_i, -g_i(z)) \right| &\le |g_i(z)| \le |g_i(z) - g_i(z^*)| \\ &= O(\|z - z^*\|) = O(\delta(z,\lambda)). \end{aligned}$$

For $i \notin \mathcal{A}_+$, we have $\hat{\lambda}_i = 0$ and $g_i(z^*) \le 0$, and so

$$\begin{aligned} i \notin \mathcal{A}_+ \quad \Rightarrow \quad \left| \min(\hat{\lambda}_i, -g_i(z)) \right| &= \max(0, g_i(z)) \le |g_i(z) - g_i(z^*)| \\ &= O(\|z - z^*\|) = O(\delta(z,\lambda)). \end{aligned}$$

By substituting the last three bounds into (16) and applying Theorem 2, we obtain the result.

3.3. Computational Aspects

Solution of the linear programs (20) is in general less expensive than solution of the quadratic programs or complementarity problems that must be solved at each step of an optimization algorithm with rapid local convergence. Linear programming software is easy to use and readily available. Moreover, given a point (z,λ) with $\delta(z,\lambda)$ small, we can expect $\hat{\mathcal{A}}_{\text{init}}$ not to contain many more indices than the weakly active set $\mathcal{B}_0$, so that few iterations of the "repeat" loop in Procedure ID0 should be needed.

Finally, we note that when more than one iteration of the "repeat" loop is needed in Procedure ID0, the linear programs to be solved at successive iterations differ only in the cost vector in (20a). Therefore, if the dual formulation of (20) is used, the solution of one linear program can typically be obtained at minimal cost from the solution of the previous linear program in the sequence. To clarify this claim, we simplify notation and write (20) as follows:

$$\max c^T \pi \quad \text{subject to } \; b_1 \le A\pi \le b_2, \;\; \pi \ge 0, \tag{29}$$

where $\pi = [\lambda_i]_{i \in \mathcal{A}(z,\lambda)}$, while c, b_1, b_2, and A are defined in obvious ways. In particular, c is a vector with elements 0 and 1, with the 1's in positions corresponding to the index set $\hat{\mathcal{A}}$. The dual of (29) is

$$\max b_1^T y_1 + b_2^T y_2 \quad \text{subject to}$$
$$\left[A^T \; -A^T \; I \right] \begin{bmatrix} y_1 \\ y_2 \\ s \end{bmatrix} = -c, \qquad (y_1, y_2, s) \ge 0.$$

When the set $\hat{\mathcal{A}}$ is changed, some of the 1's in the vector c are replaced by zeros. When only a few such changes are made, and the previous optimal basis is used to hot-start the method, we expect that only a few iterations of the dual simplex method will be needed to recover the solution of the new linear program.

4. SQP and Stabilized SQP

In the best-known form of the SQP algorithm (with exact second-order information), the following inequality constrained subproblem is solved to obtain the step Δz at each iteration:

$$\begin{aligned} \min_{\Delta z} \ & \Delta z^T \nabla \phi(z) + \tfrac{1}{2} \Delta z^T \nabla_{zz} \mathcal{L}(z,\lambda) \Delta z, \\ & \text{subject to } g(z) + \nabla g(z)^T \Delta z \le 0, \end{aligned} \tag{30}$$

where (z,λ) is the current primal-dual iterate. Denoting the Lagrange multipliers for the constraints in (30) by λ^+, we see that the solution Δz satisfies the following KKT conditions (cf. (6)):

$$\begin{bmatrix} \nabla_{zz}\mathcal{L}(z,\lambda)\Delta z + \nabla \phi(z) + \nabla g(z) \lambda^+ \\ g(z) + \nabla g(z)^T \Delta z \end{bmatrix} \in \begin{bmatrix} 0 \\ N(\lambda^+) \end{bmatrix}, \tag{31}$$

where $N(\cdot)$ is defined as in (7).

In the stabilized SQP method, we choose a parameter $\mu \ge 0$ and seek a solution of the following minimax subproblem for $(\Delta z, \lambda^+)$ such that $(\Delta z, \lambda^+ - \lambda)$ is small:

$$\begin{aligned} \min_{\Delta z} \max_{\lambda^+ \ge 0} \ & \Delta z^T \nabla \phi(z) + \tfrac{1}{2} \Delta z^T \nabla_{zz} \mathcal{L}(z,\lambda) \Delta z \\ & + (\lambda^+)^T [g(z) + \nabla g(z)^T \Delta z] - \tfrac{1}{2} \mu \|\lambda^+ - \lambda\|^2. \end{aligned} \tag{32}$$

The parameter μ can depend on an estimate of the distance $\delta(z,\lambda)$ to the primal-dual solution set; for example, $\mu = \eta(z,\lambda)^\sigma$ for some $\sigma \in (0,1)$. We can also write (32) as a linear complementarity problem, corresponding to (31), as follows:

$$\begin{bmatrix} \nabla_{zz}\mathcal{L}(z,\lambda)\Delta z + \nabla \phi(z) + \nabla g(z) \lambda^+ \\ g(z) + \nabla g(z)^T \Delta z - \mu(\lambda^+ - \lambda) \end{bmatrix} \in \begin{bmatrix} 0 \\ N(\lambda^+) \end{bmatrix}. \tag{33}$$

Li and Qi [11] derive a quadratic program in $(\Delta z, \lambda^+)$ that is equivalent to (32) and (33):

$$\begin{aligned} \min_{(\Delta z, \lambda^+)} \ & \Delta z^T \nabla \phi(z) + \tfrac{1}{2} \Delta z^T \nabla_{zz} \mathcal{L}(z,\lambda) \Delta z + \tfrac{1}{2} \mu \|\lambda^+\|^2, \\ & \text{subject to } g(z) + \nabla g(z)^T \Delta z - \mu(\lambda^+ - \lambda) \le 0. \end{aligned} \tag{34}$$

Under conditions stronger than those assumed in this paper, the results of Wright [18] and Hager [8] can be used to show that the iterates generated by (32) (or (33) or (34)) yield superlinear convergence of the sequence (z^k, λ^k) of Q-order $1+\sigma$. Our aim in the next section is to add a strategy for adjusting the multiplier, with a view to obtaining superlinear convergence under a weaker set of conditions.

5. Multiplier Adjustment and Superlinear Convergence

We show in this section that through use of Procedure ID0 and the multiplier adjustment strategy (27), we can devise a stabilized SQP algorithm that converges superlinearly whenever the initial iterate (z^0, λ^0) is sufficiently close to the primal-dual solution set $\mathcal{S}$. Only Assumption 1 is needed for this result.

Key to our analysis is Theorem 1 of Hager [8]. We state this result in Appendix A, using our current notation and making a slight correction to the original statement. Here we state an immediate corollary of Hager's result that applies under our standing assumption.

Corollary 1. *Suppose that Assumption 1 holds, and let $\lambda^* \in \mathcal{S}_\lambda$ be such that $\lambda^*_i > 0$ for all $i \in \mathcal{B}_+$. Then for any sufficiently large positive σ_0, there are positive constants ρ_0, σ_1, $\gamma \geq 1$, and $\bar{\beta}$ such that $\sigma_0 \rho_0 < \sigma_1$, with the following property: For any (z^0, λ^0) with*

$$\|(z^0, \lambda^0) - (z^*, \lambda^*)\| \leq \rho_0, \tag{35}$$

we can generate an iteration sequence $\{(z^k, \lambda^k)\}$, $k = 0, 1, 2, \ldots$, by setting

$$(z^{k+1}, \lambda^{k+1}) = (z^k + \Delta z, \lambda^+),$$

where, at iteration k, $(\Delta z, \lambda^+)$ is the local solution of the sSQP subproblem with

$$(z, \lambda) = (z^k, \lambda^k), \qquad \mu = \mu_k \in [\sigma_0 \|z^k - z^*\|, \sigma_1], \tag{36}$$

that satisfies

$$\left\|(z^k + \Delta z, \lambda^+) - (z^*, \lambda^*)\right\| \leq \gamma \left\|(z^0, \lambda^0) - (z^*, \lambda^*)\right\|. \tag{37}$$

Moreover, we have

$$\delta(z^{k+1}, \lambda^{k+1}) \leq \bar{\beta} \left[\delta(z^k \lambda^k)^2 + \mu_k \delta(\lambda^k)\right]. \tag{38}$$

Recalling our definition (9) of ϵ_λ, we define the following parametrized subset of $\mathcal{S}_\lambda$:

$$\mathcal{S}_\lambda^\nu \stackrel{\text{def}}{=} \{\lambda \in \mathcal{S}_\lambda \mid \min_{i \in \mathcal{B}_+} \lambda_i \geq \nu \epsilon_\lambda\}. \tag{39}$$

It follows easily from the MFCQ assumption and (9) that $\mathcal{S}_\lambda^\nu$ is nonempty, closed, bounded, and therefore compact for any $\nu \in [0, 1]$.

We now show that the particular choice of stabilization parameter $\mu = \eta(z, \lambda)^\sigma$, for some $\sigma \in (0, 1)$, eventually satisfies (36).

Lemma 5. *Suppose the assumptions of Corollary 1 are satisfied, and let λ^* be as defined there. Let σ be any constant in $(0, 1)$. Then there is a quantity $\rho_2 \in (0, \rho_0]$ such that when (z^0, λ^0) satisfies*

$$\|(z^0, \lambda^0) - (z^*, \lambda^*)\| \leq \rho_2, \tag{40}$$

the results of Corollary 1 hold when we set the stabilization parameter at iteration k to the following particular value:

$$\mu = \mu_k = \eta(z^k, \lambda^k)^\sigma. \tag{41}$$

Proof. We prove the result by showing that μ_k defined by (41) satisfies (36) for some choice of ρ_2. For contradiction, suppose that no such choice of ρ_2 is possible, so that for each $\ell = 1, 2, 3, \ldots$, there is a starting point $(z^0_{[\ell]}, \lambda^0_{[\ell]})$ with

$$\left\| (z^0_{[\ell]}, \lambda^0_{[\ell]}) - (z^*, \lambda^*) \right\| \le \ell^{-1} \rho_0 \tag{42}$$

such that the sequence $\left\{ \left(z^k_{[\ell]}, \lambda^k_{[\ell]} \right) \right\}_{k=0,1,2,\ldots}$ generated from this starting point in the manner prescribed by Corollary 1 with $\mu_k = \eta(z^k_{[\ell]}, \lambda^k_{[\ell]})^\sigma$ eventually comes across an index k_ℓ such that this choice of μ_k violates (36), that is, one of the following two conditions holds:

$$\sigma_0 \left\| z^{k_\ell}_{[\ell]} - z^* \right\| > \eta(z^k_{[\ell]}, \lambda^k_{[\ell]})^\sigma, \tag{43a}$$

$$\sigma_1 < \eta(z^k_{[\ell]}, \lambda^k_{[\ell]})^\sigma. \tag{43b}$$

Assume that k_ℓ is the first such index for which the violation (43) occurs. By (37) and (42), we have that

$$\left\| \left(z^{k_\ell}_{[\ell]}, \lambda^{k_\ell}_{[\ell]} \right) - (z^*, \lambda^*) \right\| \le \gamma \left\| \left(z^0_{[\ell]}, \lambda^0_{[\ell]} \right) - (z^*, \lambda^*) \right\| \le \gamma \ell^{-1} \rho_0. \tag{44}$$

Therefore by Theorem 2 and (14), we have for ℓ sufficiently large that

$$\begin{aligned}
\frac{\eta \left(z^{k_\ell}_{[\ell]}, \lambda^{k_\ell}_{[\ell]} \right)^\sigma}{\left\| z^{k_\ell}_{[\ell]} - z^* \right\|} &\ge \frac{\eta \left(z^{k_\ell}_{[\ell]}, \lambda^{k_\ell}_{[\ell]} \right)^\sigma}{\delta \left(z^{k_\ell}_{[\ell]}, \lambda^{k_\ell}_{[\ell]} \right)} \\
&\ge \kappa_0^\sigma \delta \left(z^{k_\ell}_{[\ell]}, \lambda^{k_\ell}_{[\ell]} \right)^{\sigma - 1} \\
&\ge \kappa_0^\sigma \left\| \left(z^{k_\ell}_{[\ell]}, \lambda^{k_\ell}_{[\ell]} \right) - (z^*, \lambda^*) \right\|^{\sigma - 1} \\
&\ge \kappa_0^\sigma \gamma^{\sigma - 1} \rho_0^{\sigma - 1} \ell^{1 - \sigma}.
\end{aligned} \tag{45}$$

Hence, taking limits as $\ell \uparrow \infty$, we have that

$$\frac{\eta \left(z^{k_\ell}_{[\ell]}, \lambda^{k_\ell}_{[\ell]} \right)^\sigma}{\left\| z^{k_\ell}_{[\ell]} - z^* \right\|} \to \infty \quad \text{as } \ell \uparrow \infty.$$

Dividing both sides of (43a) by $\left\| z^{k_\ell}_{[\ell]} - z^* \right\|$, we conclude from finiteness of σ_0 that (43a) is impossible.

By using Theorem 2 again together with (44), we obtain

$$\begin{aligned}
\eta \left(z^{k_\ell}_{[\ell]}, \lambda^{k_\ell}_{[\ell]} \right) &\le \kappa_1 \delta \left(z^{k_\ell}_{[\ell]}, \lambda^{k_\ell}_{[\ell]} \right) \\
&\le \kappa_1 \left\| \left(z^{k_\ell}_{[\ell]}, \lambda^{k_\ell}_{[\ell]} \right) - (z^*, \lambda^*) \right\| \\
&\le \kappa_1 \gamma \rho_0 \ell^{-1},
\end{aligned}$$

and therefore $\eta \left(z^{k_\ell}_{[\ell]}, \lambda^{k_\ell}_{[\ell]} \right)^\sigma \to 0$ as $\ell \uparrow \infty$. Hence, (43b) cannot occur either, and the proof is complete.

We now use a compactness argument to extend Corollary 1 from the single multiplier λ^* in the relative interior of $\mathcal{S}_\lambda$ to the entire set $\mathcal{S}_\lambda^\nu$, for any $\nu \in (0,1]$.

Theorem 6. *Suppose that Assumption 1 holds, and fix $\nu \in (0,1]$. Then there are positive constants $\hat{\delta}$, $\gamma \geq 1$, and β such that the following property holds: Given (z^0, λ^0) with*

$$\mathrm{dist}\left((z^0, \lambda^0), \mathcal{S}_\lambda^\nu\right) \leq \hat{\delta},$$

the iteration sequence $\{(z^k, \lambda^k)\}_{k=0,1,2,\ldots}$ generated in the manner described in Corollary 1, with μ_k, $k = 0,1,2\ldots$ chosen according to (41), satisfies the following relations:

$$\delta(z^{k+1}, \lambda^{k+1}) \leq \beta \delta(z^k, \lambda^k)^{1+\sigma} \tag{46a}$$

$$\lambda_i^k \geq \frac{1}{2}\nu\epsilon_\lambda, \quad \textit{for all } i \in \mathcal{B}_+ \textit{ and all } k = 0,1,2\ldots. \tag{46b}$$

Proof. For each $\lambda^* \in \mathcal{S}_\lambda^\nu$, we use Corollary 1 to obtain positive constants $\sigma_0(\lambda^*)$ (sufficiently large), $\sigma_1(\lambda^*)$, $\gamma(\lambda^*)$, and $\bar{\beta}(\lambda^*)$, using the argument λ^* for each constant to emphasize the dependence on the choice of multiplier λ^*. In the same vein, let $\rho_2(\lambda^*) \in (0, \rho_0(\lambda^*)]$ be the constant from Lemma 5. Now choose $\hat{\delta}(\lambda^*) > 0$ for each $\lambda^* \in \mathcal{S}_\lambda^\nu$ in such a way that

$$0 < \hat{\delta}(\lambda^*) \leq \tfrac{1}{2}\rho_2(\lambda^*), \tag{47a}$$

$$\gamma(\lambda^*)\hat{\delta}(\lambda^*) \leq \tfrac{1}{4}\nu\epsilon_\lambda, \tag{47b}$$

and consider the following open cover of $\mathcal{S}_\lambda^\nu$:

$$\cup_{\lambda^* \in \mathcal{S}_\lambda^\nu} \left\{ \lambda \,|\, \|\lambda - \lambda^*\| < \hat{\delta}(\lambda^*) \right\}. \tag{48}$$

By compactness of $\mathcal{S}_\lambda^\nu$, we can find a finite subcover defined by points $\hat{\lambda}^1, \hat{\lambda}^2, \ldots, \hat{\lambda}^f \in \mathcal{S}_\lambda^\nu$ as follows:

$$\mathcal{S}_\lambda^\nu \subset \mathcal{V} \stackrel{\mathrm{def}}{=} \cup_{j=1,2,\ldots,f} \left\{ \lambda \,|\, \|\lambda - \hat{\lambda}^j\| < \hat{\delta}(\hat{\lambda}^j) \right\}. \tag{49}$$

Note that $\mathcal{V}$ is an open neighborhood of $\mathcal{S}_\lambda^\nu$. We now define

$$\gamma \stackrel{\mathrm{def}}{=} \max_{j=1,2,\ldots,f} \gamma(\hat{\lambda}^j), \quad \bar{\beta} \stackrel{\mathrm{def}}{=} \max_{j=1,2,\ldots,f} \bar{\beta}(\hat{\lambda}^j), \quad \delta \stackrel{\mathrm{def}}{=} \max_{j=1,2,\ldots,f} \hat{\delta}(\hat{\lambda}^j), \tag{50}$$

and choose a quantity $\hat{\delta} > 0$ with the following properties:

$$\hat{\delta} \leq \min_{j=1,2,\ldots,f} \hat{\delta}(\hat{\lambda}^j) \leq \delta, \tag{51a}$$

$$\left\{ \lambda \,|\, \mathrm{dist}(\lambda, \mathcal{S}_\lambda^\nu) \leq \hat{\delta} \right\} \subset \mathcal{V}, \tag{51b}$$

$$\hat{\delta} \leq \frac{\nu\epsilon_\lambda}{4\gamma}, \tag{51c}$$

$$\hat{\delta} \leq 1. \tag{51d}$$

Now consider (z^0, λ^0) with

$$\left\|(z^0, \lambda^0) - (z^*, \lambda^*)\right\| \le \hat{\delta}, \quad \text{for some } \lambda^* \in \mathcal{S}_\lambda^\nu. \tag{52}$$

We have $\text{dist}(\lambda^0, \mathcal{S}_\lambda^\nu) \le \hat{\delta}$, and so $\lambda^0 \in \mathcal{V}$. It follows that for some $j = 1, 2, \dots, f$, we have

$$\|\lambda^0 - \hat{\lambda}^j\| \le \hat{\delta}(\hat{\lambda}^j). \tag{53}$$

Moreover, since $\|z^0 - z^*\| \le \hat{\delta}$, we have from (51a) that

$$\left\|(z^0, \lambda^0) - (z^*, \hat{\lambda}^j)\right\| \le \hat{\delta} + \hat{\delta}(\hat{\lambda}^j) \le 2\hat{\delta}(\hat{\lambda}^j) \le \rho_2(\hat{\lambda}^j), \tag{54}$$

where the final inequality follows from (47a). Application of Corollary 1 and Lemma 5 now ensures that the stabilized SQP sequence starting at (z^0, λ^0) with $\mu = \mu_k$ chosen according to (41) yields a sequence $\{(z^k, \lambda^k)\}_{k=0,1,2,\dots}$ satisfying

$$\begin{aligned} \left\|(z^k, \lambda^k) - (z^*, \hat{\lambda}^j)\right\| &\le \gamma(\hat{\lambda}^j) \left\|(z^0, \lambda^0) - (z^*, \hat{\lambda}^j)\right\| \\ &\le 2\gamma(\hat{\lambda}^j)\hat{\delta}(\hat{\lambda}^j) \le 2\gamma\delta, \end{aligned} \tag{55}$$

where we used (50) to obtain the final inequality.

To prove (46a), we have from Lemma 5, Corollary 1, the bound (15), Theorem 2, the definition (50), and the stabilizing parameter choice (41) that

$$\begin{aligned} \delta(z^{k+1}, \lambda^{k+1}) &\le \bar{\beta}(\hat{\lambda}^j)\left[\delta(z^k, \lambda^k)^2 + \mu_k \delta(\lambda^k)\right] \\ &\le \bar{\beta}\left[\delta(z^k, \lambda^k)^2 + \eta(z^k, \lambda^k)^\sigma \delta(z^k, \lambda^k)\right] \quad \text{from (50) and (41)} \\ &\le \bar{\beta}\left[\delta(z^k, \lambda^k)^2 + \kappa_1^\sigma \delta(z^k, \lambda^k)^{1+\sigma}\right] \quad \text{from Theorem 2} \\ &\le \bar{\beta}\left((2\gamma\delta)^{1-\sigma} + \kappa_1^\sigma\right)\delta(z^k, \lambda^k)^{1+\sigma}, \end{aligned}$$

where in the last line we use $\delta(z^k, \lambda^k) \le \text{dist}((z^k, \lambda^k), \mathcal{S}_\lambda^\nu) \le 2\gamma\delta$. Therefore, the result (46a) follows by setting $\beta = \bar{\beta}\left((2\gamma\delta)^{1-\sigma} + \kappa_1^\sigma\right)$.

Finally, we have from (47b) (with $\lambda^* = \hat{\lambda}^j$) and (55) that

$$\text{dist}\left((z^k, \lambda^k), \mathcal{S}_\lambda^\nu\right) \le 2\gamma(\hat{\lambda}^j)\hat{\delta}(\hat{\lambda}^j) \le \frac{1}{2}\nu\epsilon_\lambda.$$

Therefore, we have

$$i \in \mathcal{B}_+ \;\Rightarrow\; \lambda_i^k \ge \min_{\lambda^* \in \mathcal{S}_\lambda^\nu} \lambda_i^* - \frac{1}{2}\nu\epsilon_\lambda \ge \nu\epsilon_\lambda - \frac{1}{2}\nu\epsilon_\lambda = \frac{1}{2}\nu\epsilon_\lambda,$$

verifying (46b) and completing the proof.

We are now ready to state a stabilized SQP algorithm, in which multiplier adjustment steps (consisting of Procedure ID0 followed by solution of (27)) are applied when the convergence does not appear to be rapid enough.

Algorithm sSQPa
given $\sigma \in (0,1)$, τ and $\hat{\tau}$ with $0 < \hat{\tau} < \tau < 1$, tolerance `tol`;
given initial point (z^0, λ^0) with $\lambda^0 \geq 0$;
$k \leftarrow 0$;
calculate $\mathcal{A}(z^0, \lambda^0)$ from (18);
call Procedure ID0 to obtain $\mathcal{A}_+$, $\mathcal{A}_0$; solve (27) to obtain $\hat{\lambda}^0$;
$\lambda^0 \leftarrow \hat{\lambda}^0$;
repeat
  solve (32) with $(z, \lambda) = (z^k, \lambda^k)$ and $\mu = \mu_k = \eta(z^k, \lambda^k)^\sigma$
    to obtain $(\Delta z, \lambda^+)$;
  if $\eta(z^k + \Delta z, \lambda^+) \leq \eta(z^k, \lambda^k)^{1+\sigma/2}$
    $(z^{k+1}, \lambda^{k+1}) \leftarrow (z^k + \Delta z, \lambda^+)$;
    $k \leftarrow k+1$;
  else
    calculate $\mathcal{A}(z^k, \lambda^k)$ from (18);
    call Procedure ID0 to obtain $\mathcal{A}_+$, $\mathcal{A}_0$; solve (27) to obtain $\hat{\lambda}^k$;
    $\lambda^k \leftarrow \hat{\lambda}^k$;
  end (if)
until $\eta(z^k, \lambda^k) <$ `tol`.

The following result shows that when (z^0, λ^0) is close enough to $\mathcal{S}$, the initial call to Procedure ID0 is the only one needed.

Theorem 7. *Suppose that Assumption 1 holds. Then there is a constant $\bar{\delta} > 0$ such that for any (z^0, λ^0) with $\delta(z^0, \lambda^0) \leq \bar{\delta}$, the "if" condition in Algorithm sSQPa is always satisfied, and the sequence $\delta(z^k, \lambda^k)$ converges superlinearly to zero with Q-order $1 + \sigma$.*

Proof. Our result follows from Theorems 5 and 6. Choose $\nu = 1/2$ in Theorem 6, and let $\hat{\delta}$, γ, and β be as defined there. Using also δ_5 and β' from Theorem 5 and ϵ_λ defined in (9), we choose $\bar{\delta}$ as follows:

$$\bar{\delta} = \min\left(\delta_5, \hat{\delta}, \left(\frac{\epsilon_\lambda}{2\beta'}\right)^{1/\tau}, \left(\frac{\hat{\delta}}{\beta'}\right)^{1/\tau}, \frac{1}{(2\beta)^{1/\sigma}}, \kappa_0\left(\frac{\kappa_0}{\beta\kappa_1}\right)^{2/\sigma}\right). \tag{56}$$

Now let (z^0, λ^0) satisfy $\delta(z^0, \lambda^0) \leq \bar{\delta}$, and let $\hat{\lambda}^0$ be calculated from (27). From Theorem 5 and (56), we have that

$$\delta(z^0, \hat{\lambda}^0) \leq \beta' \delta(z^0, \lambda^0)^\tau \leq \beta' \bar{\delta}^\tau \leq \frac{1}{2}\epsilon_\lambda \tag{57}$$

and

$$\hat{\lambda}_i^0 \geq \epsilon_\lambda, \quad \text{for all } i \in \mathcal{B}_+, \tag{58a}$$
$$\hat{\lambda}_i^0 = 0, \quad \text{for all } i \notin \mathcal{B}_+. \tag{58b}$$

Since $\mathcal{S}_\lambda$ is closed, there is a vector $\hat\lambda^* \in \mathcal{S}_\lambda$ such that

$$\delta(z^0, \hat\lambda^0) = \left\| (z^0, \hat\lambda^0) - (z^*, \hat\lambda^*) \right\|. \tag{59}$$

From (57) and (58a), we have that

$$i \in \mathcal{B}_+ \;\Rightarrow\; \hat\lambda_i^* \geq \hat\lambda_i^0 - \frac{1}{2}\epsilon_\lambda \geq \frac{1}{2}\epsilon_\lambda,$$

so that $\hat\lambda^* \in \mathcal{S}_\lambda^\nu$ for $\nu = 1/2$. We therefore have from (57), (59), and (56) that

$$\mathrm{dist}((z^0, \hat\lambda^0), \mathcal{S}_\lambda^\nu) = \left\| (z^0, \hat\lambda^0) - (z^*, \hat\lambda^*) \right\| \leq \beta' \bar\delta^\tau \leq \hat\delta. \tag{60}$$

From here on, we set $\lambda^0 \leftarrow \hat\lambda^0$, as in Algorithm sSQPa. Because of the last bound, we can apply Theorem 6 to (z^0, λ^0). We use this result to prove the following claims. First,

$$\bar\delta \geq \delta(z^0, \lambda^0) \geq 2\delta(z^1, \lambda^1) \geq 4\delta(z^2, \lambda^2) \geq \cdots. \tag{61}$$

Second,

$$\eta(z^{k+1}, \lambda^{k+1}) \leq \eta(z^k, \lambda^k)^{1+\sigma/2}, \quad \text{for all } k = 0, 1, 2, \ldots. \tag{62}$$

We prove both claims by induction. For $k = 0$ in (61), we have from (60) and $\bar\delta \leq \hat\delta$ in (56) that $\delta(z^0, \lambda^0) \leq \bar\delta$. Assume that the first $k+1$ inequalities in (61) have been verified. From (46a) and (56), we have that

$$\delta(z^{k+1}, \lambda^{k+1}) \leq \beta\delta(z^k, \lambda^k)^{1+\sigma} \leq \beta\bar\delta^\sigma \delta(z^k, \lambda^k) \leq \frac{1}{2}\delta(z^k, \lambda^k),$$

so that the next inequality in the chain is also satisfied. For (62), we have from Theorem 2, (46a), and (61) that

$$\begin{aligned}
\eta(z^{k+1}, \lambda^{k+1}) &\leq \kappa_1 \delta(z^{k+1}, \lambda^{k+1}) \\
&\leq \beta\kappa_1 \delta(z^k, \lambda^k)^{1+\sigma} \\
&\leq \beta\kappa_1 \bar\delta^{\sigma/2} \delta(z^k, \lambda^k)^{1+\sigma/2} \\
&\leq \beta\kappa_1 \bar\delta^{\sigma/2} \kappa_0^{-1-\sigma/2} \eta(z^k, \lambda^k)^{1+\sigma/2} \\
&\leq \eta(z^k, \lambda^k)^{1+\sigma/2},
\end{aligned}$$

where the last bound follows from (56). Hence, (62) is verified, so that the condition in the "if" statement of Algorithm sSQPa is satisfied for all $k = 0, 1, 2, \ldots$. Superlinear convergence with Q-order $1+\sigma$ follows from (46a).

ϵ	$\mathcal{A}_+ = \mathcal{B}_+$ and $\mathcal{A}_0 = \mathcal{B}_0$	$\mathcal{A}(z, \lambda) = \mathcal{B}$
.1	0%	82%
.01	51%	100%
.001	100%	100%

Table 1. Constraint Identification Results for Example 1.

6. Numerical Examples

We illustrate the use of Procedure ID0 in identifying and classifying active constraints with the help of some simple examples, and then describe a simple problem on which the Algorithm sSQPa outperforms the standard SQP approach.

Our tests of Procedure ID0 are similar to those reported by Facchinei, Fischer, and Kanzow [5, Section 4]. For several inequality constrained problems with active constraints at the solution, we generate random points (z, λ) in a neighborhood of $\mathcal{S}$ and find whether Procedure ID0 correctly identifies the active constraints and correctly classifies them as strongly and weakly active. Specifically, for a given positive parameter ϵ, we generate 100 random points by choosing a point (z^*, λ^*) randomly from the set $\mathcal{S}$, and adding a random number chosen from the uniform distribution on $[-\epsilon, \epsilon]$ to each component of (z^*, λ^*). We then monitor the number of trials on which the weakly and strongly active constraints were correctly identified and classified. We also keep track of the number of trials on which the active constraints were identified correctly, but the classification into weakly and strongly active constraints was incorrect.

We use the parameter settings $\tau = 0.7$ and $\hat{\tau} = 0.65$.

Example 1. Three-circle problem.

$$\min z_1 \ \text{ subject to}$$
$$(z_1 - 2)^2 + z_2^2 \le 4, \ \ (z_1 - 4)^2 + z_2^2 \le 16, \ \ z_1^2 + (z_2 - 2)^2 \le 4.$$

The solution is $z^* = (0, 0)^T$, all three constraints are active at the solution, and the optimal Lagrange multiplier set is

$$\mathcal{S}_\lambda = \{(.25 - 2\alpha, \alpha, 0)^T \mid \alpha \in [0, .125]\}.$$

Therefore, we have $\mathcal{B}_+ = \{1, 2\}$ and $\mathcal{B}_0 = \{3\}$.

Results for this example are shown in Table 1. For each of the three tabulated values of ϵ, 100 points were generated as described above. For $\epsilon = .1$, the correct active set was identified in 82 trials, but in none of these cases was the correct classification into $\mathcal{B}_+$ and $\mathcal{B}_0$ obtained. The reason was that the lower bound ξ was large enough that none of the components of $\tilde{\lambda}$ obtained by solving (20) exceeded this value, so all constraints were assigned to $\mathcal{A}_0$. For $\epsilon = .01$, the correct active set was identified in all trials, and on about half of these, the correct classification was also found. For $\epsilon = .001$, the correct classification was made in all trials.

ϵ	$\mathcal{A}_+ = \mathcal{B}_+$ and $\mathcal{A}_0 = \mathcal{B}_0$	$\mathcal{A}(z,\lambda) = \mathcal{B}$
.1	31%	86%
.01	39%	99%
.001	81%	100%
.0001	100%	100%

Table 2. Constraint Identification Results for Example 2.

Example 2. A modification of problem 46 from [10], described in [5, Example 2]. The aim is to minimize

$$(z_1 - z_2)^2 + (z_3 - 1)^2 + (z_4 - 1)^4 + (z_5 - 1)^6$$

subject to

$$\begin{aligned} x_1^2 x_4 + \sin(x_4 - x_5) - 1 &\geq 0, \\ x_2 + x_3^4 x_4^2 - 2 &\geq 0, \\ 1 - x_2 &\geq 0. \end{aligned}$$

The solution is $z^* = (1,1,1,1,1)^T$ and the optimal multiplier is unique: $\lambda^* = (0,0,0)^T$. Therefore, we have $\mathcal{B} = \mathcal{B}_0 = \{1,2,3\}$ and $\mathcal{B}_+ = \emptyset$.

Results are shown in Table 2. The correct active set is identified readily, but the correct classification is obtained only for small values of ϵ. Procedure ID0 tends to misclassify some constraints as being strongly active, because the threshold value ξ is too small. We can obtain better results with choices of $\hat{\tau}$ that yield higher values of ξ; by changing $\hat{\tau}$ from 0.65 to 0.4, we find for $\epsilon = 0.1$ that the proportion of correct classifications jumps from 31% to 64%.

Example 3. A modification of problem 43 from [10], described in [5, Example 3]. The objective function is

$$z_1^2 + z_2^2 + 2z_3^2 + z_4^2 - 5z_1 - 5z_2 - 21z_3 + 7z_4$$

and the constraints are as follows:

$$\begin{aligned} z_1^2 + z_2^2 + z_3^2 + z_4^2 + z_1 - z_2 + z_3 - z_4 - 8 &\leq 0, \\ z_1^2 + 2z_2^2 + z_3^2 + 2z_4^2 - z_1 - z_4 - 10 &\leq 0, \\ 2z_1^2 + z_2^2 + z_3^2 + 2z_1 - z_2 - z_4 - 5 &\leq 0, \\ -z_2^3 - 2z_1^2 - z_4^2 - z_1 + 3z_2 + z_3 - 4z_4 - 7 &\leq 0. \end{aligned}$$

The solution is $z^* = (0,1,2,-1)^T$ and the optimal active sets are $\mathcal{B} = \mathcal{B}_+ = \{1,3,4\}$ and $\mathcal{B}_0 = \emptyset$ (the second constraint is inactive). The optimal multiplier set is

$$\mathcal{S}_\lambda = \{(3-\alpha, 0, \alpha, \alpha - 2)^T \,|\, \alpha \in [2,3]\}.$$

Results are shown in Table 3. This example proves to be fairly easy; in most cases, the three active components of λ are greater than ξ, so Procedure ID0 immediately classifies them all as strongly active without solving any linear programs of the form (20). For the few remaining cases in which one of these

ϵ	$\mathcal{A}_+ = \mathcal{B}_+$ and $\mathcal{A}_0 = \mathcal{B}_0$	$\mathcal{A}(z,\lambda) = \mathcal{B}$
.1	59%	59%
.01	100%	100%
.001	100%	100%

Table 3. Constraint Identification Results for Example 3.

components is close to zero, solution of a single linear program suffices to resolve this component too as being strongly active.

Finally, we compare the performance of Algorithm sSQPa and the standard SQP approach on the following simple problem:

$$\min\ z_1 \ \text{ subject to } \ z_1 \ge 0, \ \ (z_1 - 2)^2 + z_2^2 \le 4. \tag{63}$$

The solution of this problem is $z^* = (0,0)^T$, with both constraints strongly active and

$$\mathcal{S}_\lambda = \{(1 - 4\alpha, \alpha) \,|\, \alpha \in [0, .25]\}. \tag{64}$$

The constraint gradients are linearly dependent, but satisfy MFCQ. The Lagrangian Hessian is

$$\nabla_{zz}\mathcal{L}(z^*, \lambda^*) = 2\lambda_2^* \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

so that it is positive definite on the null space of active constraints for all $\lambda^* \in \mathcal{S}_\lambda$ *except* for the multiplier $\lambda^* = (1,0)^T$ (corresponding to $\alpha = 0$ in the definition (64)).

To compare the local performance of Algorithm sSQPa with the performance of the standard SQP approach (based on the subproblem (30)), we implemented both algorithms in Matlab. For sSQPa, we used the parameters $\tau = 0.7$, $\hat{\tau} = .65$, $\sigma = 0.9$. We terminated the algorithms when $\eta(z,\lambda) < 10^{-12}$. We chose starting points by picking a random point in $\mathcal{S}$, then adding numbers randomly distributed from the interval $[-.01, .01]$ to each component.

The stabilized SQP approach implemented in Algorithm sSQPa has no difficulty solving this problem for all starting points tried. It identifies both constraints as being strongly active, and then chooses a starting multiplier in the relative interior of $\mathcal{S}_\lambda$. Subsequent iterates remain close to this starting point. In all 20 trials we ran, it required just two iterations for convergence. Although the theory of Section 5 does not apply directly to this problem, because the condition (12), (13) is not satisfied for $\lambda^* = (1,0)^T$, the results can be modified to show superlinear convergence from a neighborhood of any subset of $\mathcal{S}$ at which these second-order conditions *are* satisfied.

The standard SQP approach encounters trouble from all starting points tried. It either fails altogether (we declare failure when $\eta(z,\lambda)$ increases significantly on an iteration), or else converges to the point $z^* = (0,0)^T$ and $\lambda^* = (0,1)^T$ at a linear rate of 1/2. Failure occurred 9 times in the 20 trials, while the other 11 trials required between 13 and 15 iterations each. From any starting point close to $\mathcal{S}$, the SQP method is apparently drawn to the troublesome multiplier $(1,0)^T$, and the lack of a second-order condition at this point results in slow convergence.

7. Conclusions

We have presented a technique for identifying the weakly and strongly active inequality constraints at a local solution of a nonlinear programming problem, where the standard assumptions—existence of a strictly complementary solution and linear independence of active constraints gradients—are replaced by weaker assumptions. We have embedded this technique in a stabilized SQP algorithm, resulting in a method that converges superlinearly under the weaker assumptions when started at a point sufficiently close to the (primal-dual) optimal set.

The primal-dual algorithm described by Vicente and Wright [15] can also be improved by using the techniques outlined here. In that paper, strict complementarity is assumed along with MFCQ, and superlinear convergence is proved provided both $\delta(z^0, \lambda^0)$ is sufficiently small and $\lambda_i^0 \geq \gamma$, for all $i \in \mathcal{B} = \mathcal{B}_+$ and some $\gamma > 0$. If we apply the active constraint detection procedure (18) and the subproblem (27) to *any* initial point (z^0, λ^0) with $\delta(z^0, \lambda^0)$ sufficiently small, the same convergence result can be obtained without making the positivity assumption on the components of $\lambda_{\mathcal{B}_+}^0$. (Because of the strict complementarity assumption, Procedure ID0 serves only to verify that $\mathcal{B} = \mathcal{B}_+$.)

Numerous issues remain to be investigated. We believe that degeneracy is an important issue, given the large size of many modern applications of nonlinear programming and their nature as discretizations of continuous problems. Nevertheless, the practical usefulness of constraint identification and stabilization techniques remains to be investigated. The numerical implications should also be investigated, since implementation of these techniques may require solution of ill-conditioned systems of linear equations (see M. H. Wright [16] and S. J. Wright [19]). Embedding of these techniques into globally convergence algorithmic frameworks needs to be examined. We should investigate generalization to equality constraints, possibly involving the use of the "weak MFCQ" condition, which does not require linear independence of the equality constraint gradients.

Acknowledgments

I thank Bill Hager for discussions of his key result, Theorem 8. I also thank Mostafa El-Sayed and Luis Vicente for discussions about the Matlab implementation and about the practical performance of Algorithm sSQPa, and the referee for several useful comments.

A. Hager's Theorem

We restate Theorem 1 of Hager [8], making a slight correction to the original statement concerning the conditions on (z^0, λ^0) and the radius of the neighborhood containing the sequence $\{(z^k, \lambda^k)\}$. No modification to Hager's analysis is needed to prove the following version of this result.

Theorem 8. *Suppose that z^* is a local solution of (1), and that ϕ and g are twice Lipschitz continuously differentiable in a neighborhood of z^*. Let λ^* be some multiplier such that the KKT conditions (3) are satisfied, and define*

$$\bar{\mathcal{B}} \stackrel{\text{def}}{=} \{i \,|\, \lambda_i^* > 0\}.$$

Suppose that there is an $\alpha > 0$ such that

$$w^T \nabla_{zz}\mathcal{L}(z^*, \lambda^*) w \geq \alpha \|w\|^2, \;\; \textit{for all } w \textit{ such that } \nabla g_i(z^*)^T w = 0, \textit{ for all } i \in \bar{\mathcal{B}}.$$

Then for any choice of σ_0 sufficiently large, there are positive constants ρ_0, σ_1, $\gamma \geq 1$, and $\bar{\beta}$ such that $\sigma_0 \rho_0 < \sigma_1$, with the following property: For any (z^0, λ^0) with

$$\|(z^0, \lambda^0) - (z^*, \lambda^*)\| \leq \rho_0,$$

we can generate an iteration sequence $\{(z^k, \lambda^k)\}$, $k = 0, 1, 2, \ldots$, by setting

$$(z^{k+1}, \lambda^{k+1}) = (z^k + \Delta z, \lambda^+),$$

where, at iteration k, $(\Delta z, \lambda^+)$ is the local solution of the sSQP subproblem with

$$(z, \lambda) = (z^k, \lambda^k), \qquad \mu = \mu_k \in [\sigma_0 \|z^k - z^*\|, \sigma_1],$$

that satisfies

$$\left\|(z^k + \Delta z, \lambda^+) - (z^*, \lambda^*)\right\| \leq \gamma \left\|(z^0, \lambda^0) - (z^*, \lambda^*)\right\|.$$

Moreover, we have

$$\delta(z^{k+1}, \lambda^{k+1}) \leq \bar{\beta} \left[\delta(z^k \lambda^k)^2 + \mu_k \delta(\lambda^k)\right].$$

References

1. M. Anitescu, *Degenerate nonlinear programming with a quadratic growth condition*, Preprint ANL/MCS-P761-0699, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., June 1999. To appear in SIAM Journal on Optimization.
2. M. C. Bartholomew-Biggs, *Recursive quadratic programming methods based on the augmented Lagrangian*, Mathematical Programming Study, 31 (1987), pp. 21–41.
3. J. F. Bonnans, *Local study of newton type algorithms for constrained problems*, in Optimization–Fifth French-German Conference, S. Dolecki, ed., no. 1405 in Lecture Notes in Mathematics, Springer-Verlag, 1989, pp. 13–24.
4. J. F. Bonnans and A. Ioffe, *Second-order sufficiency and quadratic growth for nonisolated minima*, Mathematics of Operations Research, 20 (1995), pp. 801–819.
5. F. Facchinei, A. Fischer, and C. Kanzow, *On the accurate identification of active constraints*, SIAM Journal on Optimization, 9 (1998), pp. 14–32.
6. A. Fischer, *Modified Wilson method for nonlinear programs with nonunique multipliers*, Mathematics of Operations Research, 24 (1999), pp. 699–727.
7. J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Mathematical Programming, 12 (1977), pp. 136–138.
8. W. W. Hager, *Stabilized sequential quadratic programming*, Computational Optimization and Applications, 12 (1999), pp. 253–273.
9. W. W. Hager and M. S. Gowda, *Stability in the presence of degeneracy and error estimation*, Mathematical Programming, Series A, 85 (1999), pp. 181–192.

10. W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, no. 187 in Lecture NOtes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1981.
11. D.-H. Li and L. Qi, *A stabilized SQP method via linear equations*, technical report, Mathematics Department, University of New South Wales, 2000.
12. O. L. Mangasarian and S. Fromovitz, *The Fritz-John necessary optimality conditions in the presence of equality and inequality constraints*, Journal of Mathematical Analysis and Applications, 17 (1967), pp. 37–47.
13. D. Ralph and S. J. Wright, *Superlinear convergence of an interior-point method despite dependent constraints*, Mathematics of Operations Research, 25 (2000), pp. 179–194.
14. S. M. Robinson, *Generalized equations and their solutions. part II: Applications to nonlinear programming*, Mathematical Programming Study, 19 (1982), pp. 200–221.
15. L. N. Vicente and S. J. Wright, *Local convergence of a primal-dual methods for degenerate nonlinear programming*, technical report, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., February 2000. To appear in Computational Optimization and Applications.
16. M. H. Wright, *Ill-conditioning and computational error in interior methods for nonlinear programming*, SIAM Journal on Optimization, 9 (1998), pp. 84–111.
17. S. J. Wright, *Modifying SQP for degenerate problems*, Preprint ANL/MCS-P699-1097, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., 1997. Revised June 2000.
18. ———, *Superlinear convergence of a stabilized SQP method to a degenerate solution*, Computational Optimization and Applications, 11 (1998), pp. 253–275.
19. ———, *Effects of finite-precision arithmetic on interior-point methods for nonlinear programming*, SIAM Journal on Optimization, 12 (2001), pp. 36–78.
20. S. J. Wright and D. Ralph, *A superlinear infeasible-interior-point algorithm for monotone nonlinear complementarity problems*, Mathematics of Operations Research, 21 (1996), pp. 815–838.