SIAM J. OPTIM.
Vol. 15, No. 3, pp. 673–696

AN ALGORITHM FOR DEGENERATE NONLINEAR PROGRAMMING WITH RAPID LOCAL CONVERGENCE*

STEPHEN J. WRIGHT†

Abstract. This paper describes and analyzes an algorithmic framework for solving nonlinear programming problems in which strict complementarity conditions and constraint qualifications are not necessarily satisfied at a solution. The framework is constructed from three main algorithmic ingredients. The first is any conventional method for nonlinear programming that produces estimates of the Lagrange multipliers at each iteration; the second is a technique for estimating the set of active constraint indices; the third is a stabilized Lagrange–Newton algorithm with rapid local convergence properties. Results concerning rapid local convergence and global convergence of the proposed framework are proved. The approach improves on existing approaches in that less restrictive assumptions are needed for convergence and/or the computational workload at each iteration is lower.

Key words. nonlinear programming, degeneracy, superlinear convergence

AMS subject classifications. 90C30, 90C55, 65K05, 49M05

DOI. 10.1137/030601235

1. Introduction.

Consider the following nonlinear programming problem with inequality constraints:

$$\text{NLP:} \qquad \min_z \phi(z) \qquad \text{subject to } g(z) \le 0, \tag{1.1}$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable functions. In this paper we describe an algorithmic framework for this problem that converges superlinearly to a solution z^* under mild assumptions on the functions ϕ and g in the neighborhood of z^*. Rapid convergence can be proved even for difficult problems, such as those with complementarity constraints or with equality constraints of the form $h(x) = 0$ that have been split into two inequalities $h(x) \le 0$, $h(x) \ge 0$. The use of less-than-ideal formulation techniques such as the latter is well known in the case of linear programming, for which production software contains presolvers or preprocessors that remove many such infelicities before they can cause difficulties for the underlying algorithm. Since for nonlinear programming it is often difficult to detect and remedy poor formulations at the level of the modeling language or the mathematical formulation, we should aim to design algorithms that perform as well as possible in a variety of difficult circumstances.

Optimality conditions for (1.1) can be derived from the Lagrangian for (1.1), which is

$$\mathcal{L}(z, \lambda) = \phi(z) + \lambda^T g(z), \tag{1.2}$$

where $\lambda \in \mathbb{R}^m$ is the vector of Lagrange multipliers. The Karush–Kuhn–Tucker

*Received by the editors October 16, 2003; accepted for publication (in revised form) June 9, 2004; published electronically April 8, 2005. This research was supported in part by the National Science Foundation under grants MTS-0086559, ACI-0196485, and EIA-0127857.
http://www.siam.org/journals/siopt/15-3/60123.html
†Computer Sciences Department, University of Wisconsin, 1210 W. Dayton Street, Madison, WI 53706 (swright@cs.wisc.edu).

(KKT) conditions are satisfied at a point (z^*, λ^*) if

$$\nabla_z \mathcal{L}(z^*, \lambda^*) = \nabla \phi(z^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(z^*) = 0, \tag{1.3a}$$

$$0 \geq g(z^*) \perp \lambda^* \geq 0, \tag{1.3b}$$

where the notation "$\perp$" signifies that $g(z^*)^T \lambda^* = 0$. We call a point z^* at which these relations are satisfied a *KKT point* or *stationary point*. The set $\mathcal{A}^*$ of active constraints at z^* is defined as follows:

$$\mathcal{A}^* = \{i = 1, 2, \ldots, m \,|\, g_i(z^*) = 0\}. \tag{1.4}$$

In this paper, we develop an algorithmic framework ("Framework INEQ") that exhibits local superlinear convergence under mild assumptions. It dispenses with constraint qualifications altogether and does not require strict complementarity. It assumes only that a KKT point exists and that a certain second-order sufficient condition is satisfied at this point. Framework INEQ incorporates an "outer strategy" whose purpose is to ensure good global convergence properties and an "equality constrained phase" (the "EQ phase") which is activated when the set of active constraints has apparently been identified. We do not dwell on the outer strategy further; techniques involving merit functions, filters, and a variety of algorithms for generating steps can be used for this purpose. Rather, the focus of this paper is on the EQ phase and Framework INEQ. In the EQ phase, a "stabilized Lagrange–Newton" algorithm (Algorithm SLN) is applied to the optimality conditions for the subproblem obtained by enforcing the apparently active constraints as equalities. Algorithm SLN is described in section 4.

The cost of each iteration in the EQ phase of Framework INEQ is essentially the same as in most nonlinear programming algorithms that use second derivatives, except for the solution at some iterations of an ancillary linear program. In particular, computation of the step requires solution of a single linear system whose structure and sparsity are almost identical to the well-known "augmented system" matrix that arises in many interior-point and sequential quadratic programming approaches. Our approach includes various checks (which are inexpensive, apart from the linear program mentioned above) to ensure that the algorithm does not get trapped at a nonoptimal point because of incorrect identification of the active set.

The remainder of the paper is structured as follows. In the remainder of this section, we summarize the previous literature on superlinearly convergent algorithms based on weaker-than-usual assumptions, and compare it to the approach of this paper. In section 2, we give some background on optimality conditions for (1.1) and for the equality constrained nonlinear optimization problem, focusing in particular on the critical cone that is used in the specification of the second-order condition. Section 3 discusses computable estimates of the distance of a given point (z, λ) to the primal-dual solution set for (1.1) in the vicinity of a point satisfying the second-order condition. Algorithm SLN for the equality constrained problem is presented and analyzed in section 4, and we comment on the similarity of the linear systems solved at each iteration of this approach to the corresponding systems for the quadratic penalty and augmented Lagrangian methods. We present Framework INEQ for solving (1.1) in section 5, along with local and global convergence results. Finally, we discuss some consequences of the second-order assumption in section 6, and outline the behavior of the approach on a particular degenerate problem.

1.1. Previous work. Other work on algorithms that retain rapid local convergence properties under weakened conditions has been performed by Fischer [5, 6], Wright [16, 15], Hager [8], and Izmailov and Solodov [11]. Wright [15] proposed a stabilized sequential quadratic programming method ("stabilized SQP") in which a linear-quadratic minimax problem is solved at each iteration. Rapid local convergence is proved for the case in which the Mangasarian–Fromovitz constraint qualification (MFCQ) and strict complementarity hold. This approach also required a starting point that is not too close to the "relative boundary" of the primal-dual solution set. By using a more restrictive second-order condition, Hager [8] dispensed with the constraint qualifications and strict complementarity assumptions. He proved rapid convergence provided that the primal-dual starting point is located in a neighborhood of a particular primal-dual solution satisfying his second-order condition. Fischer [5] 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 nondegeneracy assumptions, but stronger than the ones made in this paper. In [6], Fischer described a more general framework for generalized equations that includes the stabilized SQP approach as a special case. By using assumptions similar to those of Hager [8] (in particular, the stronger second-order condition), he proves local superlinear convergence.

Wright [16] described superlinear local convergence properties of a class of inexact SQP methods and showed that the stabilized SQP approach of [15, 8] and Fischer's method [5] 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 unenforced constraints. Still it achieved superlinear convergence under weaker-than-usual conditions. More recently, Wright [17] described a constraint identification procedure that distinguishes between weakly active constraints (those for which the Lagrange multiplier component is zero, for all Lagrange multiplier vectors satisfying the KKT conditions) and strongly active constraints. Rapid local convergence of a stabilized SQP approach based on this identification technique is proved under MFCQ and a certain second-order condition.

The approach of Izmailov and Solodov [11] is similar to the one described here in that it uses explicit estimates of the active constraint set $\mathcal{A}^*$ and crossover to an equality constrained phase at the final approach to the solution. The second-order assumption of [11] is the same as those used here, although we also consider situations in which these conditions hold uniformly on a compact subset of the set of multipliers satisfying the KKT conditions. The main difference between Framework INEQ and the approach of [11] is that the latter requires computation of a matrix that projects (approximately) onto the kernel of the active constraint Jacobian. This projection can be quite expensive to compute in practice; the singular-value-decomposition approach presented in [11] is not practical for large problems as it yields a dense matrix in general even when the constraint Jacobian is sparse. By contrast, the linear system to be solved at each iteration of our equality constrained phase is quite similar to those that are solved at each iteration of a conventional SQP or primal-dual interior-point method. Another difference is that Framework INEQ includes checks that detect errors in the estimation of the active set. These checks are essential to preventing convergence to a nonoptimal point and in yielding a global convergence result.

We believe that by choosing an appropriate outer strategy, the approach of this paper can be incorporated readily into algorithms (and software) with good global convergence properties, thereby enhancing the performance of such algorithms in problems that exhibit degeneracy at the solution.

2. Assumptions, notation, and basic results.

We now review the optimality conditions for (1.1) and its equality constrained counterpart, and outline the assumptions and notation that are used in subsequent sections. We treat the inequality constrained problem (1.1) in section 2.1 and the equality constrained problem in section 2.2.

2.1. The inequality constrained problem.

Recall the KKT conditions (1.3). 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

$$\mathcal{S}_\lambda \stackrel{\text{def}}{=} \{\lambda^* \mid \lambda^* \text{ satisfies (1.3)}\}, \quad \mathcal{S} \stackrel{\text{def}}{=} \{z^*\} \times \mathcal{S}_\lambda. \tag{2.1}$$

Given the optimal active set $\mathcal{A}^*$, defined in (1.4), we define several related index sets. For any optimal multiplier $\lambda^* \in \mathcal{S}_\lambda$, we define the set $\mathcal{A}_+^*(\lambda^*)$ to be the "support" of λ^*, that is,

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

We define $\mathcal{A}_+^*$ (without argument) as

$$\mathcal{A}_+^* \stackrel{\text{def}}{=} \cup_{\lambda^* \in \mathcal{S}_\lambda} \mathcal{A}_+^*(\lambda^*); \tag{2.2}$$

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

$$\mathcal{A}_0^* \stackrel{\text{def}}{=} \mathcal{A}^* \backslash \mathcal{A}_+^*.$$

This set $\mathcal{A}_0^*$ contains the *weakly active* constraint indices, those indices $i \in \mathcal{A}^*$ such that $\lambda_i^* = 0$ for all $\lambda^* \in \mathcal{S}_\lambda$. It is easy to show, using convexity of $\mathcal{S}_\lambda$, that there exists a particular vector $\hat{\lambda} \in \mathcal{S}_\lambda$ such that

$$\hat{\lambda}_i = 0 \ \text{ for all } i \notin \mathcal{A}_+^*; \quad \hat{\lambda}_i > 0 \ \text{ for all } i \in \mathcal{A}_+^*. \tag{2.3}$$

We define the following second-order sufficient condition at some point $(z^*, \lambda^*) \in \mathcal{S}$: There is a scalar $\upsilon > 0$ such that

$$\begin{aligned} & w^T \nabla_{zz}^2 \mathcal{L}(z^*, \lambda^*) w \geq \upsilon \|w\|^2, \\ & \text{for all } w \text{ such that} \quad \begin{aligned} \nabla g_i(z^*)^T w = 0 \quad & \text{for all } i \in \mathcal{A}_+^*, \\ \nabla g_i(z^*)^T w \leq 0 \quad & \text{for all } i \in \mathcal{A}_0^*. \end{aligned} \end{aligned} \tag{2.4}$$

Condition 2s.1 in [16, Section 3] assumes that (2.4) holds for *all* $\lambda^* \in \mathcal{S}_\lambda$, for a fixed $\upsilon > 0$. In other works (see Hager and Gowda [9] and Fischer [6, Lemma 5]), the condition on $\nabla_{zz}^2 \mathcal{L}(z^*, \lambda^*)$ in (2.4) is assumed to hold at just one multiplier $\lambda^* \in \mathcal{S}_\lambda$, but for all w satisfying the condition

$$\begin{aligned} \nabla g_i(z^*)^T w = 0 \quad & \text{for all } i \in \mathcal{A}_+^*(\lambda^*), \\ \nabla g_i(z^*)^T w \leq 0 \quad & \text{for all } i \in \mathcal{A}^* \backslash \mathcal{A}_+^*(\lambda^*). \end{aligned} \tag{2.5}$$

Since $\mathcal{A}_+^*(\lambda^*) \subset \mathcal{A}_+^*$ for all $\lambda^* \in \mathcal{S}_\lambda$, we might expect that the set of vectors w defined by (2.5) is in general a superset of the set defined in (2.4). In fact, the two sets are identical, as the following simple result shows.

LEMMA 2.1. *The direction sets defined in* (2.4) *and* (2.5) *coincide, for any vector* $\lambda^* \in \mathcal{S}_\lambda$.

Proof. It is clear that the direction set in (2.4) is a subset of (2.5), because $\mathcal{A}_+^*(\lambda^*) \subset \mathcal{A}_+^*$. To prove the reverse inclusion, we use the multiplier $\hat{\lambda} \in \mathcal{S}_\lambda$ defined in (2.3). For w satisfying (2.5), we have from (1.3) that

$$\sum_{i \in \mathcal{A}_+^*(\lambda^*)} \lambda_i^* \nabla g_i(z^*) = -\nabla \phi(z^*) = \sum_{i \in \mathcal{A}_+^*} \hat{\lambda}_i \nabla g_i(z^*).$$

Taking inner products of these expressions with w, and using the first relation in (2.5), we have

$$0 = \sum_{i \in \mathcal{A}_+^*} \hat{\lambda}_i w^T \nabla g_i(z^*).$$

Then, using the fact that $\hat{\lambda}_i > 0$ and $w^T \nabla g_i(z^*) \le 0$ for all $i \in \mathcal{A}_+^*$, it follows that in fact $w^T \nabla g_i(z^*) = 0$ for all $i \in \mathcal{A}_+^*$. We also have from (2.5) that $w^T \nabla g_i(z^*) \le 0$ for all $i \in \mathcal{A}_0^* = \mathcal{A}^* \backslash \mathcal{A}_+^*$. Hence, w satisfies the conditions in (2.4), as claimed. ∎

A less restrictive second-order condition, stated in terms of a quadratic growth condition of the objective $\phi(z)$ in a feasible neighborhood of z^*, is discussed by Bonnans and Ioffe [3] and Anitescu [1].

Our standing assumption for the problem (1.1) in this paper is as follows.

ASSUMPTION 1. *The KKT conditions* (1.3) *hold at* z^* *(that is,* $\mathcal{S} \neq \emptyset$*) and 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 $z \neq z^*$.

THEOREM 2.2. *Suppose that Assumption* 1 *holds and that the second-order condition* (2.4) *is satisfied at some* $\lambda^* \in \mathcal{S}_\lambda$. *Then* z^* *is a strict local minimizer of* (1.1).

Proof. For the proof, see Robinson [14, Theorem 2.2]. ∎

2.2. The equality constrained problem. Consider now the nonlinear programming problem with equality constraints defined by

$$\min_z \phi(z) \qquad \text{subject to } \; h(z) = 0, \tag{2.6}$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are twice continuously differentiable functions. We denote the Lagrangian for (2.6) as follows:

$$\bar{\mathcal{L}}(z, \gamma) \stackrel{\text{def}}{=} \phi(z) + \gamma^T h(z). \tag{2.7}$$

The KKT conditions for (2.6) at z^* are that there exists a vector $\gamma^* \in \mathbb{R}^p$ such that

$$\nabla_z \bar{\mathcal{L}}(z^*, \gamma^*) = \nabla \phi(z^*) + \sum_{i=1}^{p} \gamma_i^* \nabla h_i(z^*) = 0. \tag{2.8}$$

The second-order sufficient condition is satisfied at some multiplier γ^* from (2.8) if there exists $\upsilon > 0$ such that

$$\begin{array}{c} w^T \nabla^2 \bar{\mathcal{L}}(z^*, \gamma^*) w \ge \upsilon \|w\|^2, \\ \text{for all } w \text{ such that} \quad \nabla h_i(z^*)^T w = 0, \;\; i = 1, 2, \dots, p. \end{array} \tag{2.9}$$

Finally, we define

$$\bar{\mathcal{S}}_\gamma \stackrel{\text{def}}{=} \{\gamma^* \,|\, \gamma^* \text{ satisfies (2.8)}\}, \qquad \bar{\mathcal{S}} \stackrel{\text{def}}{=} \{z^*\} \times \bar{\mathcal{S}}_\gamma. \tag{2.10}$$

ASSUMPTION 2. *The KKT conditions* (1.3) *are satisfied at z^* (that is, $\bar{\mathcal{S}} \neq \emptyset$) and the functions ϕ and h are twice Lipschitz continuously differentiable in a neighborhood of z^*.*

Similarly to Theorem 2.2, we have that z^* is a strict local solution of (2.6) when this assumption is satisfied and the second-order condition (2.9) holds for some multiplier γ^* satisfying (2.8).

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

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

where here and elsewhere, $\|\cdot\|$ denotes the Euclidean norm unless a subscript specifically indicates otherwise.

We use 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 $\|M\| \leq \beta\|A\|$ whenever $\|A\|$ is sufficiently small.

For a vector $\lambda \in \mathbb{R}^m$, we denote by $\lambda_{\mathcal{A}}$ the subvector whose components are λ_i, $i \in \mathcal{A}$. We denote by $g_{\mathcal{A}}(z)$ the subvector of the vector function $g(z)$ whose components are $g_i(z)$, $i \in \mathcal{A}$.

3. Distance-to-solution estimates. In this section we describe simple formulae for estimating distance to the solution of (1.1) (or (2.6)) from a given primal-dual point (z, λ) (or (z, γ)). We are interested in two-sided estimates; that is, those that are bounded both above and below by the true distance to the solution. These estimates are used by the methods of sections 4 and 5 both to decide on acceptability of a step and as a parameter in the step computation.

We aim for results that hold in a neighborhood of a compact subset of $\mathcal{S}_\lambda$ (or $\bar{\mathcal{S}}_\gamma$) at which the second-order condition is satisfied.

The following result, which makes use of Assumption 1 and (2.4) at a particular $\lambda^* \in \mathcal{S}_\lambda$, gives a practical estimate of $\text{dist}\,((z, \lambda), \mathcal{S})$. We use $\min(\lambda, -g(z))$ here to denote the vector whose ith component is $\min(\lambda_i, -g_i(z))$.

THEOREM 3.1. *Suppose that Assumption* 1 *holds at z^*, and that the second-order condition* (2.4) *is satisfied at some $\lambda^* \in \mathcal{S}_\lambda$. Then there are positive quantities $\delta_0(\lambda^*)$ and $\beta(\lambda^*)$ such that for all (z, λ) with $\|(z, \lambda) - (z^*, \lambda^*)\| \leq \delta_0(\lambda^*)$, the quantity $\eta(z, \lambda)$ defined by*

$$\eta(z, \lambda) \stackrel{\text{def}}{=} \left\| \begin{bmatrix} \nabla_z \mathcal{L}(z, \lambda) \\ \min(\lambda, -g(z)) \end{bmatrix} \right\|_1 \tag{3.1}$$

satisfies

$$\eta(z, \lambda) \in [1/\beta(\lambda^*), \beta(\lambda^*)]\text{dist}\,((z, \lambda), \mathcal{S}).$$

Proof. One side of the result, namely,

$$\eta(z, \lambda) \leq \beta(\lambda^*)\text{dist}\,((z, \lambda), \mathcal{S}),$$

is relatively trivial; for a proof see Wright [16]. The other inequality, namely,

$$\operatorname{dist}((z,\lambda),\mathcal{S}) \le \beta(\lambda^*)\eta(z,\lambda)$$

has been examined by a number of authors under different assumptions; see, for example, Facchinei, Fischer, and Kanzow [4, Theorem 3.6], Wright [16, Theorem A.1], Hager and Gowda [9], and Fischer [6]. Under the assumptions used here, the result follows from Theorem 2 of Fischer [6], with $\Sigma_0 = \{(z^*,\lambda^*)\}$ in the notation of [6], where we use [9, Lemma 2] to justify [6, Assumption 1], and also Lemma 2.1 above to verify that the critical direction set of [9, Lemma 2] is the same as the direction set defined in (2.4). □

We now extend Theorem 3.1 to the case in which (2.4) holds for all λ^* in a compact subset of $\mathcal{S}_\lambda$.

THEOREM 3.2. *Suppose that Assumption* 1 *holds for the problem* (1.1) *at* z^**, and that the second-order condition* (2.4) *holds for all* $\lambda^* \in \mathcal{V}_\lambda$*, where* $\mathcal{V}_\lambda$ *is a compact subset of* $\mathcal{S}_\lambda$*. Let us define* $\mathcal{V} = \{z^*\} \times \mathcal{V}_\lambda$*. Then there are positive constants* $\delta_1(\mathcal{V}_\lambda)$ *and* $\beta(\mathcal{V}_\lambda)$ *such that for all* (z,λ) *satisfying*

$$\operatorname{dist}((z,\lambda),\mathcal{V}) \le \delta_1(\mathcal{V}_\lambda) \tag{3.2}$$

we have

$$\eta(z,\lambda) \in [1/\beta(\mathcal{V}_\lambda), \beta(\mathcal{V}_\lambda)]\operatorname{dist}((z,\lambda),\mathcal{S}),$$

where $\eta(z,\lambda)$ *is defined in* (3.1).

Proof. From Theorem 3.1, we have for any particular $\lambda^* \in \mathcal{V}_\lambda$ that there are positive constants $\delta_1(\lambda^*)$ and $\beta(\gamma^*)$ such that for any (z,λ) with $\|(z,\lambda)-(z^*,\lambda^*)\|_2 \le \delta_1(\lambda^*)$ we have

$$\eta(z,\lambda) \in [1/\beta(\lambda^*), \beta(\lambda^*)]\operatorname{dist}((z,\lambda),\mathcal{S}).$$

The open balls

$$\mathcal{B}(\lambda^*) \stackrel{\text{def}}{=} \{(z,\lambda) \,|\, \|(z,\lambda)-(z^*,\lambda^*)\|_2 < \delta_1(\lambda^*)\} \qquad \text{for all } \lambda^* \in \mathcal{V}_\lambda$$

form an open cover of $\mathcal{V}$. Since $\mathcal{V}$ is compact, there is a finite subcover, indexed by a finite set of multipliers, say $\{\lambda_1^*, \lambda_2^*, \ldots, \lambda_M^*\}$. We choose $\delta_1(\mathcal{V}_\lambda) > 0$ such that

$$\{(z,\lambda) \,|\, \operatorname{dist}((z,\lambda),\mathcal{V}) \le \delta_1(\mathcal{V}_\lambda)\} \subset \cup_{i=1}^M \mathcal{B}(\lambda_i^*),$$

and set $\beta(\mathcal{V}_\lambda) = \max_{i=1,2,\ldots,M} \beta(\lambda_i^*)$ to obtain the result. □

An active set identification strategy for (1.1) follows immediately from the estimate of Theorem 3.2. Similarly to Facchinei, Fischer, and Kanzow [4, equation (2.5)], we choose a constant $\tau \in (0,1)$ and define

$$\mathcal{A}(z,\lambda) \stackrel{\text{def}}{=} \{i = 1,2,\ldots,m \,|\, g_i(z) \ge -\eta(z,\lambda)^\tau\}. \tag{3.3}$$

Given a set $\mathcal{V}$ defined as in Theorem 3.2, we have, for all (z,λ) with $\operatorname{dist}((z,\lambda),\mathcal{V})$ sufficiently small, that

$$\begin{aligned} i \in \mathcal{A}^* &\;\Rightarrow\; -g_i(z) \le O(\|z-z^*\|) \le \|z-z^*\|^\tau/\beta(\mathcal{V}_\lambda)^\tau \le \eta(z,\lambda)^\tau, \\ i \notin \mathcal{A}^* &\;\Rightarrow\; g_i(z) \le (1/2)g_i(z^*) < -\eta(z,\lambda)^\tau, \end{aligned}$$

so that $\mathcal{A}(z,\lambda) = \mathcal{A}^*$. We formalize this result as follows.

THEOREM 3.3. *Suppose that Assumption* 1 *holds at* z^* *and that the second-order condition* (2.4) *holds for all* $\lambda^* \in \mathcal{V}_\lambda$, *where* $\mathcal{V}_\lambda$ *is a compact subset of* $\mathcal{S}_\lambda$. *Let us define* $\mathcal{V} = \{z^*\} \times \mathcal{V}_\lambda$. *Then there exists* $\delta_2(\mathcal{V}_\lambda) > 0$ *such that for all* (z,λ) *with* $\operatorname{dist}((z,\lambda),\mathcal{V}) \le \delta_2(\mathcal{V}_\lambda)$, *we have* $\mathcal{A}(z,\lambda) = \mathcal{A}^*$.

We now use the results of Hager and Gowda [9] to obtain distance-to-solution estimates for the equality constrained problem (2.6), for multipliers γ in some neighborhood of a compact subset of $\bar{\mathcal{S}}_\gamma$, which we denote by $\bar{\mathcal{V}}_\gamma$. We also introduce the notation $\bar{\mathcal{V}} \stackrel{\text{def}}{=} \{z^*\} \times \bar{\mathcal{V}}_\gamma$ to denote the primal-dual extension of $\bar{\mathcal{V}}_\gamma$.

THEOREM 3.4. *Consider the problem* (2.6), *and suppose that* (2.8) *and* (2.9) *hold at* (z^*,γ^*) *for some fixed* $\upsilon > 0$ *and all* $\gamma^* \in \bar{\mathcal{V}}_\gamma \subset \bar{\mathcal{S}}_\gamma$, *where* $\bar{\mathcal{V}}_\gamma$ *is compact. Then there are positive constants* $\bar{\delta}_1(\bar{\mathcal{V}}_\gamma)$ *and* $\bar{\beta}(\bar{\mathcal{V}}_\gamma)$ *such that for all* (z,γ) *with* $\operatorname{dist}((z,\gamma),\bar{\mathcal{V}}) \le \bar{\delta}_1(\bar{\mathcal{V}}_\gamma)$, *we have*

$$\bar{\eta}(z,\gamma) \stackrel{\text{def}}{=} \left\| \begin{bmatrix} \nabla_z \bar{\mathcal{L}}(z,\gamma) \\ h(z) \end{bmatrix} \right\|_1 \in [1/\bar{\beta}(\bar{\mathcal{V}}_\gamma), \bar{\beta}(\bar{\mathcal{V}}_\gamma)] \operatorname{dist}((z,\gamma),\bar{\mathcal{S}}). \tag{3.4}$$

Proof. From Hager and Gowda [9, Lemma 2, Theorem 3], we have for any particular $\gamma^* \in \bar{\mathcal{V}}_\gamma$ that there are positive constants $\bar{\delta}_1(\gamma^*)$ and $\bar{\beta}(\gamma^*)$ such that for any (z,γ) with $\|(z,\gamma) - (z^*,\gamma^*)\|_2 \le \bar{\delta}_1(\gamma^*)$ we have

$$\bar{\eta}(z,\gamma) \in [1/\bar{\beta}(\gamma^*), \bar{\beta}(\gamma^*)] \operatorname{dist}((z,\gamma),\bar{\mathcal{S}}).$$

The result now follows from the same compactness argument as in the proof of Theorem 3.2. □

4. An algorithm for the equality constrained problem.

In this section, we describe a locally convergent algorithm for the equality constrained problem (2.6), and then discuss its relationship to other methods which require the solution of a similar linear system at each iteration.

4.1. Algorithm SLN and its convergence properties.

We obtain the step from a point (z,γ) by the following process. First, define μ as follows:

$$\mu \stackrel{\text{def}}{=} \left\| \begin{bmatrix} \nabla_z \bar{\mathcal{L}}(z,\gamma) \\ h(z) \end{bmatrix} \right\|_1 = \bar{\eta}(z,\gamma). \tag{4.1}$$

Next, solve the following linear system:

$$\begin{bmatrix} \nabla^2_{zz} \bar{\mathcal{L}}(z,\gamma) & \nabla h(z) \\ -\nabla h(z)^T & \mu I \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta \gamma \end{bmatrix} = \begin{bmatrix} -\nabla_z \bar{\mathcal{L}}(z,\gamma) \\ h(z) \end{bmatrix}. \tag{4.2}$$

We call the approach "stabilized Lagrange-Newton," since it is based on solving the Lagrange-Newton equations that constitute the first-order optimality conditions.

ALGORITHM SLN.

Given starting point (z^0,γ^0)
for $k = 0,1,2,\ldots$
  Define μ_k by setting $(z,\gamma) = (z^k,\gamma^k)$ in (4.1);
  Solve (4.2) with $(z,\gamma) = (z^k,\gamma^k)$ and $\mu = \mu_k$ to obtain $(\Delta z^k, \Delta\gamma^k)$;
  Set $(z^{k+1},\gamma^{k+1}) \leftarrow (z^k,\gamma^k) + (\Delta z^k, \Delta\gamma^k)$;
end (for)

LEMMA 4.1. *Suppose that Assumption 2 holds and that (2.9) holds for constant $\upsilon > 0$ and all $\gamma^* \in \bar{\mathcal{V}}_\gamma$, where $\bar{\mathcal{V}}_\gamma$ is a compact subset of $\bar{\mathcal{S}}_\gamma$. Then there exist positive constants $\delta_3(\bar{\mathcal{V}}_\gamma)$, $\chi_1(\bar{\mathcal{V}}_\gamma)$, and $\chi_2(\bar{\mathcal{V}}_\gamma)$ with $\delta_3(\bar{\mathcal{V}}_\gamma) \le \min(1, \bar{\delta}_1(\bar{\mathcal{V}}_\gamma))$ such that for all (z, γ) with*

$$\text{dist}\,((z,\gamma), \bar{\mathcal{V}}) \le \delta_3(\bar{\mathcal{V}}_\gamma), \tag{4.3}$$

where $\bar{\mathcal{V}} = \{z^*\} \times \bar{\mathcal{V}}_\gamma$, the step $(\Delta z, \Delta\gamma)$ obtained from (4.1), (4.2) satisfies the following relations:

$$\|(\Delta z, \Delta\gamma)\|_1 \le \chi_1(\bar{\mathcal{V}}_\gamma)\bar{\eta}(z,\gamma), \tag{4.4}$$

$$\bar{\eta}(z+\Delta z, \gamma+\Delta\gamma) \le \chi_2(\bar{\mathcal{V}}_\gamma)\bar{\eta}(z,\gamma)^2. \tag{4.5}$$

Proof. We omit a detailed proof of this result, since it follows closely the proofs of Theorems 3.2 and 4.1 in Wright [15]. It is based on a singular value decomposition of the constraint Jacobian $\nabla h(z^*)$, which induces a partitioning of both Δz and $\Delta\gamma$ into two orthogonal components. The uniform positive definiteness of $\nabla^2_{zz}\bar{\mathcal{L}}(z,\gamma)$ on the null space of $\nabla h(z^*)^T$ is crucial to the proof, as is the distance-to-solution result, Theorem 3.4. ∎

THEOREM 4.2. *Suppose that Assumption 2 holds. Let $\hat{\mathcal{V}}_\gamma$ be a compact subset of $\bar{\mathcal{S}}_\gamma$, with $\hat{\mathcal{V}} = \{z^*\} \times \hat{\mathcal{V}}_\gamma$, and define*

$$\bar{\mathcal{V}}_\gamma = \{\gamma \in \bar{\mathcal{S}}_\gamma \,|\, \text{dist}\,(\gamma, \hat{\mathcal{V}}_\gamma) \le \epsilon\} \tag{4.6}$$

for some $\epsilon > 0$. Suppose that (2.9) holds for all $\gamma^* \in \bar{\mathcal{V}}_\gamma$ and fixed $\upsilon > 0$. Then there exists a positive constant $\delta_5(\hat{\mathcal{V}}_\gamma, \epsilon)$ such that if (z^0, γ^0) satisfies

$$\text{dist}\,((z^0,\gamma^0), \hat{\mathcal{V}}) \le \delta_5(\hat{\mathcal{V}}_\gamma, \epsilon), \tag{4.7}$$

then the sequence $\{(z^k, \gamma^k)\}$ generated in Algorithm SLN converges Q-quadratically to a point $(z^*, \gamma^*) \in \bar{\mathcal{S}}$ and satisfies the relations (4.4) and (4.5) at each iteration.

Proof. This proof makes extensive use of the quantities $\bar{\beta}(\bar{\mathcal{V}}_\gamma)$, $\delta_3(\bar{\mathcal{V}}_\gamma)$, $\chi_1(\bar{\mathcal{V}}_\gamma)$, and $\chi_2(\bar{\mathcal{V}}_\gamma)$ defined in Theorem 3.4 and Lemma 4.1. For convenience of notation during this proof, we drop the explicit dependence of the quantities $\delta_3(\bar{\mathcal{V}}_\gamma)$, $\chi_1(\bar{\mathcal{V}}_\gamma)$, $\chi_2(\bar{\mathcal{V}}_\gamma)$, $\bar{\beta}(\bar{\mathcal{V}}_\gamma)$, $\delta_5(\hat{\mathcal{V}}_\gamma, \epsilon)$ on the sets $\bar{\mathcal{V}}_\gamma$ and $\hat{\mathcal{V}}_\gamma$ and the scalar ϵ.

We show that the theorem holds for the following choice of δ_5:

$$\delta_5 \overset{\text{def}}{=} \min\left(\frac{1}{4\chi_2\bar{\beta}}, \frac{\min(\delta_3,\epsilon)}{4}, \frac{\min(\delta_3,\epsilon)}{4\chi_1\bar{\beta}}\right). \tag{4.8}$$

The main part of the proof is to show inductively that the following bounds hold for all $k = 0, 1, 2, \ldots$:

$$\bar{\eta}(z^{k+1}, \gamma^{k+1}) \le (1/2)\bar{\eta}(z^k, \gamma^k), \tag{4.9a}$$

$$\text{dist}\,((z^{k+1}, \gamma^{k+1}), \bar{\mathcal{V}}) \le (1 - 1/2^{k+2})\delta_3, \tag{4.9b}$$

$$\|(\Delta z^k, \Delta\gamma^k)\|_1 \le \min(\delta_3, \epsilon)/2^{k+2}, \tag{4.9c}$$

$$\text{dist}\,(\gamma^{k+1}, \hat{\mathcal{V}}_\gamma) \le (1 - 1/2^{k+1})\epsilon. \tag{4.9d}$$

We first consider $k = 0$. From Theorem 3.4 and (4.8), we have

$$\chi_2\bar{\eta}(z^0,\gamma^0) \le \chi_2\bar{\beta}\text{dist}\,((z^0,\gamma^0), \bar{\mathcal{S}}) \le \chi_2\bar{\beta}\delta_5 \le 1/4.$$

Hence, we have from (4.5) that

$$\bar{\eta}(z^1, \gamma^1) \le \left(\chi_2 \bar{\eta}(z^0, \gamma^0)\right) \bar{\eta}(z^0, \gamma^0) \le (1/4)\bar{\eta}(z^0, \gamma^0),$$

so that (4.9a) holds for $k = 0$. From (4.4) and (4.8), we have

$$\|(\Delta z^0, \Delta\gamma^0)\|_1 \le \chi_1 \bar{\eta}(z^0, \gamma^0) \le \chi_1 \bar{\beta} \mathrm{dist}\,((z^0, \gamma^0), \bar{\mathcal{S}}) \le \chi_1 \bar{\beta} \delta_5 \le (1/4) \min(\delta_3, \epsilon),$$

so (4.9c) holds for $k = 0$. For (4.9b), we have

$$\begin{aligned} &\mathrm{dist}\,((z^1, \gamma^1), \bar{\mathcal{V}}) \le \mathrm{dist}\,((z^0, \gamma^0), \bar{\mathcal{V}}) + \|(\Delta z^0, \Delta\gamma^0)\|_1 \\ &\le \mathrm{dist}\,((z^0, \gamma^0), \hat{\mathcal{V}}) + \|(\Delta z^0, \Delta\gamma^0)\|_1 \le (1/4)\delta_3 + (1/4)\delta_3 < (1 - 1/4)\delta_3. \end{aligned}$$

For (4.9d), we have

$$\mathrm{dist}\,(\gamma^1, \hat{\mathcal{V}}_\gamma) \le \mathrm{dist}\,(\gamma^0, \hat{\mathcal{V}}_\gamma) + \|\Delta\gamma^0\|_1 \le (1/4 + 1/4) \min(\delta_3, \epsilon) \le (1/2)\epsilon,$$

proving that this bound also holds at $k = 0$.

We assume now that (4.9) hold for $0, 1, 2, \ldots, k$, and prove that these estimates continue to hold when k is replaced by $k + 1$. For (4.9c), we have

$$\begin{aligned} &\|(\Delta z^{k+1}, \Delta\gamma^{k+1})\|_1 \\ &\le \chi_1 \bar{\eta}(z^{k+1}, \eta^{k+1}) && \text{by (4.9b) and Lemma 4.1} \\ &\le (1/2^{k+1}) \chi_1 \bar{\eta}(z^0, \gamma^0) && \text{by (4.9a)} \\ &\le (1/2^{k+1}) \chi_1 \bar{\beta} \mathrm{dist}\,((z^0, \gamma^0), \bar{\mathcal{S}}) && \text{by Theorem 3.4} \\ &\le (1/2^{k+1}) \chi_1 \bar{\beta} \delta_5 && \text{by (4.7) and } \hat{V} \subset \bar{\mathcal{S}} \\ &\le (1/2^{k+1})(1/4) \min(\delta_3, \epsilon) && \text{by (4.8)} \\ &= (1/2^{k+3}) \min(\delta_3, \epsilon), \end{aligned}$$

verifying that this bound continues to hold for $k+1$. For (4.9b), we have from (4.9b) at k and (4.9c) at $k + 1$ that

$$\begin{aligned} \mathrm{dist}\,((z^{k+2}, \gamma^{k+2}), \bar{\mathcal{V}}) &\le \mathrm{dist}\,((z^{k+1}, \gamma^{k+1}), \bar{\mathcal{V}}) + \|(\Delta z^{k+1}, \Delta\gamma^{k+1})\|_1 \\ &\le (1 - 1/2^{k+2})\delta_3 + (1/2^{k+3})\delta_3 \\ &= (1 - 1/2^{k+3})\delta_3, \end{aligned}$$

proving this result for $k + 1$. For (4.9a), we have

$$\bar{\eta}(z^{k+1}, \gamma^{k+1}) \le (1/2^{k+1}) \bar{\eta}(z^0, \gamma^0),$$

so that

$$\begin{aligned} &\chi_2 \bar{\eta}(z^{k+1}, \gamma^{k+1}) \\ &\le (1/2^{k+1}) \chi_2 \bar{\eta}(z^0, \gamma^0) && \text{by (4.9a)} \\ &\le (1/2^{k+1}) \chi_2 \bar{\beta} \mathrm{dist}\,((z^0, \gamma^0), \bar{\mathcal{S}}) && \text{by Theorem 3.4} \\ &\le (1/2^{k+1}) \chi_2 \bar{\beta} \mathrm{dist}\,((z^0, \gamma^0), \hat{\mathcal{V}}) && \text{since } \hat{\mathcal{V}} \subset \bar{\mathcal{S}} \\ &\le (1/2^{k+1})(1/4) = 1/2^{k+3} && \text{by (4.7) and (4.8).} \end{aligned}$$

Because we have shown that the bound (4.9b) holds when k is replaced by $k+1$, we have from Lemma 4.1 that

$$\begin{aligned}\bar{\eta}(z^{k+2},\gamma^{k+2}) &\le \left(\chi_2\bar{\eta}(z^{k+1},\gamma^{k+1})\right)\bar{\eta}(z^{k+1},\gamma^{k+1}) \\ &\le (1/2^{k+3})\bar{\eta}(z^{k+1},\gamma^{k+1}),\end{aligned}$$

proving that (4.9a) continues to hold at $k+1$. To prove (4.9d) at $k+1$, we have from (4.9c) at $k+1$ and $\delta_3 \le 1$ that

$$\begin{aligned}&\text{dist}\,(\gamma^{k+2},\hat{\mathcal{V}}_\gamma) \le \text{dist}\,(\gamma^{k+1},\hat{\mathcal{V}}_\gamma) + \|\Delta\gamma^{k+1}\|_1 \\ &\le (1-1/2^{k+1})\min(\delta_3,\epsilon) + (1/2^{k+3})\min(\delta_3,\epsilon) \le (1-1/2^{k+2})\min(\delta_3,\epsilon),\end{aligned}$$

as required.

It follows from (4.9c) that the sequence $\{(z^k,\gamma^k)\}$ is Cauchy and hence convergent. Because of (4.9a), the limit point $((z^*,\gamma^*)$, say) must lie in $\bar{\mathcal{S}}$. Finally, Q-quadratic convergence follows from the following chain of inequalities:

$$\begin{aligned}&\|(z^{k+1},\gamma^{k+1}) - (z^*,\gamma^*)\|_2 \\ &\le \sum_{i=k+1}^{\infty} \|(\Delta z^i,\Delta\gamma^i)\|_1 \\ &\le \sum_{i=k+1}^{\infty} \chi_1\bar{\eta}(z^i,\gamma^i) && \text{by (4.4)} \\ &\le 2\chi_1\bar{\eta}(z^{k+1},\gamma^{k+1}) && \text{by (4.9a)} \\ &\le 2\chi_1\chi_2\bar{\eta}(z^k,\gamma^k)^2 && \text{by (4.5)} \\ &\le 2\chi_1\chi_2\bar{\beta}^2\text{dist}\,((z^k,\gamma^k),\bar{\mathcal{S}})^2 && \text{by Theorem 3.4} \\ &\le 2\chi_1\chi_2\bar{\beta}^2\|(z^k,\gamma^k)-(z^*,\gamma^*)\|^2 && \text{since } (z^*,\gamma^*)\in\bar{\mathcal{S}}. \quad \square\end{aligned}$$

4.2. Discussion. The step (4.2) used by Algorithm SLN is similar to steps that arise in at least two well-known algorithms for the equality constrained problem (2.6), neither of which was designed with the issue of degeneracy in mind. The first example is the quadratic penalty function, which can be defined as

$$P(z;\mu) = \phi(z) + \frac{1}{2\mu}\|h(z)\|_2^2.$$

Algorithms based on this function typically find an approximate minimizer of $P(\cdot;\mu)$, then decrease μ and repeat the process. Under certain assumptions, it is known that the sequence of minimizers approaches a solution z^* as $\mu \downarrow 0$; see [12, Chapter 17].

It is not difficult to show that the Newton step Δz for $P(\cdot;\mu)$ satisfies the following system (in conjunction with another vector $\Delta\gamma$):

$$\begin{bmatrix} \nabla^2_{zz}\bar{\mathcal{L}}(z,h(z)/\mu) & \nabla h(z) \\ -\nabla h(z)^T & \mu I \end{bmatrix}\begin{bmatrix}\Delta z \\ \Delta\gamma\end{bmatrix} = \begin{bmatrix} -\nabla_z\bar{\mathcal{L}}(z,0) \\ h(z)\end{bmatrix}. \tag{4.10}$$

The similarities between the systems (4.2) and (4.10) are evident. Note, however, that the upper left block of the coefficient matrix in (4.10) is guaranteed to be similar to that of (4.2) only when $h(z)/\mu \approx \gamma$—a condition that cannot be expected to hold at many of the iterates visited by the quadratic penalty approach. Hence, the system does not have the nice property of (4.2) in which Assumption 2 guarantees nonsingularity of the coefficient matrix on later iterations.

The augmented Lagrangian for (2.6) is conventionally defined as follows:

$$\hat{\mathcal{L}}(z,\gamma;\mu) \stackrel{\text{def}}{=} \phi(z) + \gamma^T h(z) + \frac{1}{2\mu}\|h(z)\|_2^2.$$

The basic method of multipliers [10, 13, 2] proceeds by finding an approximate minimizer of $\hat{\mathcal{L}}(\cdot,\gamma;\mu)$ for fixed γ and μ, then updating γ by the formula $\gamma \leftarrow \gamma + h(z)/\mu$, then possibly decreasing μ and repeating the process. When Newton's method is used to minimize $\hat{\mathcal{L}}(\cdot,\gamma;\mu)$, the step Δz satisfies the following system:

$$\text{(4.11)} \qquad \begin{bmatrix} \nabla^2_{zz}\bar{\mathcal{L}}(z,\gamma+h(z)/\mu) & \nabla h(z) \\ -\nabla h(z)^T & \mu I \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta\gamma \end{bmatrix} = \begin{bmatrix} -\nabla_z \bar{\mathcal{L}}(z,\gamma) \\ h(z) \end{bmatrix}.$$

The system (4.11) is indeed very similar to (4.2) when $\|h(z)\| \ll \mu$. The method of multipliers differs from Algorithm SLN, however, in that the $\Delta\gamma$ component of the step is not used (since a different formula is used to update γ) and μ is chosen heuristically rather than via the formula (4.1). Indeed, a main motivation of the method of multipliers is to obtain convergence to z^* without driving μ to zero.

We note that steps for the two methods above have traditionally been obtained using unconstrained minimization algorithms in z alone, rather than the systems (4.10) and (4.11) above. Hence, small values of μ lead to ill conditioned subproblems—a problem that still occurs in the formulations (4.10) and (4.11) above in the degenerate case. Nevertheless, our results in this section, and our observations about numerical implementations in section 6, suggest that local convergence of both approaches could be improved by crossing over to Algorithm SLN on later iterations.

5. Extension to equality constrained problems. We now describe an algorithmic framework for the inequality constrained problem (1.1) that is based on the active constraint identification technique of section 3 and Algorithm SLN of section 4, along with an "outer strategy" that could be derived from any globally convergent algorithm. This framework, which we call Framework INEQ, enters an "EQ phase" when there is reason to believe that it has identified the active constraints correctly. In the EQ phase, it applies Algorithm SLN to the problem obtained by fixing these constraints as equalities and ignoring the others. Checks are applied at each iteration of the EQ phase to ensure that the active constraint identification was in fact correct, and that the algorithm is not converging to a point that does not solve (or, at least, is not a KKT point for) the original problem (1.1). When false convergence of this type is detected, or when the convergence is not sufficiently rapid, Framework INEQ exits the EQ phase and restores the values of (z,λ) that were in force on entry to this phase. (Restoration is necessary because components of λ are allowed to become negative during the EQ phase, making them potentially unsuitable for use as multiplier estimates for the original problem.) Outside of the EQ phase, the iterates (z^k,λ^k) are generated by the outer strategy.

We denote by EQ($\mathcal{A}$) the equality-constrained problem obtained by enforcing a subset $\mathcal{A}$ of the components of g in (1.1) as equalities, that is,

$$\text{(5.1)} \qquad \text{EQ}(\mathcal{A}): \qquad \min_z \phi(z) \qquad \text{subject to } g_i(z) = 0 \text{ for all } i \in \mathcal{A}.$$

The following result, also noted by Izmailov and Solodov [11, Proposition 3.2], shows that the solution of the inequality constrained problem satisfying the second-order sufficient condition is a solution of the problem EQ($\mathcal{A}^*$) for the optimal active set $\mathcal{A}^*$ and also satisfies the second-order sufficient condition for this reduced problem.

THEOREM 5.1. *Suppose that Assumption* 1 *holds at* z^**, and that for some multiplier* $\lambda^* \in \mathcal{S}_\lambda$*, the conditions* (2.4) *are satisfied. Then the equality constrained problem* (5.1) *with* $\mathcal{A} = \mathcal{A}^*$ *has local solution* z^* *satisfying Assumption* 2*, and the KKT conditions* (2.8) *and the second-order condition* (2.9) *are satisfied by the multiplier* $\lambda^*_{\mathcal{A}^*}$.

The next result is a kind of converse of Theorem 5.1. We omit its proof, which is also simple.

THEOREM 5.2. *Suppose for some* $\mathcal{A} \subset \{1, 2, \dots, m\}$ *that* z^* *is a point satisfying Assumption* 2 *for the problem EQ($\mathcal{A}$). Suppose too that there is a multiplier vector* $\lambda^*_{\mathcal{A}}$ *satisfying* (2.8) *such that* $\lambda^*_{\mathcal{A}} \geq 0$*, and in addition that* $g_i(z^*) \leq 0$ *for all* $i \notin \mathcal{A}$*. Then* z^* *satisfies the KKT conditions* (1.3) *for the problem* (1.1)*.*

An important role in detecting incorrect identification of the active constraints is played by the following test:

kkt-test$(\bar{z}, \bar{\lambda}_{\mathcal{A}}, \mathcal{A}, B)$: There exists $\lambda_{\mathcal{A}}$ with $\lambda_{\mathcal{A}} \geq 0$ and $\|\lambda_{\mathcal{A}}\|_1 \leq B$ such that

$$\left\| \nabla\phi(\bar{z}) + \sum_{i \in \mathcal{A}} \lambda_i \nabla g_i(\bar{z}) \right\|_1 \leq \bar{\eta}(\bar{z}, \bar{\lambda}_{\mathcal{A}})^\sigma, \tag{5.2}$$

where $\sigma \in (.5, 1)$ is a specified constant. Note that the test can be applied by solving the linear program

$$\min_{\lambda_{\mathcal{A}} \geq 0,\, \|\lambda_{\mathcal{A}}\|_1 \leq B} \left\| \nabla\phi(\bar{z}) + \sum_{i \in \mathcal{A}} \lambda_i \nabla g_i(\bar{z}) \right\|_1,$$

though it may not be necessary to iterate all the way to optimality.

We are now ready to specify the framework.

Framework INEQ.

O: Given (z^0, λ^0) with $\lambda^0 \geq 0$, $\tau_{\text{EQ}} \in (0, .5]$, and $\sigma \in (.5, 1)$; Set $k \leftarrow 0$;

1: (* Estimate active constraints *)
Evaluate $\eta(z^k, \lambda^k)$; Set $\mathcal{A} \leftarrow \mathcal{A}(z^k, \lambda^k)$;
if $\eta(z^k, \lambda^k) = 0$
  STOP;

2: (* Test for entry into EQ Phase *)
if $\bar{\eta}(z^k, \lambda^k_{\mathcal{A}}) \leq \tau_{\text{EQ}}$
  $z^{\text{store}} \leftarrow z^k$; $\lambda^{\text{store}} \leftarrow \lambda^k$;
  go to **3**;
else
  go to **6**;

3: (* EQ Phase *)
Given $(z^k, \lambda^k_{\mathcal{A}})$, compute $(\Delta z^k, \Delta\lambda^k_{\mathcal{A}})$ for problem EQ($\mathcal{A}$) from (4.1), (4.2);

4: (* Test validity of EQ Phase step *)
if $\|(\Delta z^k, \Delta\lambda^k_{\mathcal{A}})\|_1 > \bar{\eta}(z^k, \lambda^k_{\mathcal{A}})^\sigma$
  go to **5**;

if $\bar\eta(z^k + \Delta z^k, \lambda^k_{\mathcal{A}} + \Delta\lambda^k_{\mathcal{A}}) > \bar\eta(z^k, \lambda^k_{\mathcal{A}})^{1+\sigma}$
    go to **5**;
if $g_i(z^k + \Delta z^k) > 0$ for any $i \notin \mathcal{A}$
    go to **5**;
if $\lambda^k_i + \Delta\lambda^k_i < 0$ for some $i \in \mathcal{A}$ and
    kkt-test$(z^k + \Delta z^k, \lambda^k_{\mathcal{A}} + \Delta\lambda^k_{\mathcal{A}}, \mathcal{A}, \|\lambda^k_{\mathcal{A}}\|_1 + 2)$ is false
    go to **5**;
$(z^{k+1}, \lambda^{k+1}_{\mathcal{A}}) \leftarrow (z^k + \Delta z^k, \lambda^k_{\mathcal{A}} + \Delta\lambda^k_{\mathcal{A}})$;
$k \leftarrow k + 1$;
if $\bar\eta(z^k, \lambda^k_{\mathcal{A}}) = 0$
    STOP;
go to **3**;

5: (* Drop out of EQ phase *)
$\tau_{\text{EQ}} \leftarrow \tau_{\text{EQ}}/2$; $z^k \leftarrow z^{\text{store}}$; $\lambda^k \leftarrow \lambda^{\text{store}}$;

6: (* Apply outer strategy *)
Compute (z^{k+1}, λ^{k+1}) using the "outer" strategy;
$k \leftarrow k + 1$; go to **1**;

The main computational costs in each iteration of the EQ phase are (i) the solution of the single linear system (4.2), which has similar structure and sparsity to systems that appear in standard methods for nonlinear programming, and (ii) the solution of the feasibility problem in (5.2). In fact, the latter calculation may not be needed at every step of the EQ phase, since it is performed only if the new iteration survives the earlier tests in step 4 and if some components of the Lagrange multiplier estimates have become negative. In addition, Framework INEQ can be modified easily to perform this test at only a subset of the EQ-phase iterations, reducing the computational requirements further.

Our first convergence result shows that if the iterates enter the EQ phase and remain there, any limit point must be a KKT point for (1.1).

THEOREM 5.3. *Suppose that Framework INEQ does not terminate finitely and that the iterates (z^k, λ^k) remain in the EQ phase for all $k \geq K$, where K is some index. Then the sequence $\{z^k\}$ converges R-superlinearly to a limit $\hat z$ which is a KKT point for* (1.1).

Proof. Because of the second test in step 4, $\bar\eta(z^k, \lambda^k_{\mathcal{A}})$ approaches zero geometrically. Using the first test in step 4, we have $\|(\Delta z^k, \Delta\lambda^k_{\mathcal{A}})\|_1 \leq \bar\eta(z^k, \lambda^k_{\mathcal{A}})^\sigma$, so the sequence $\{(z^k, \lambda^k_{\mathcal{A}})\}$ is Cauchy and hence convergent, say to $(\hat z, \hat\lambda_{\mathcal{A}})$. Since, from the third test in step 4, we have $g_i(z^k) \leq 0$ for all $i \notin \mathcal{A}$, then $g_i(\hat z) \leq 0$ for all $i \notin \mathcal{A}$. Since $|g_i(z^k)| \leq \bar\eta(z^k, \lambda^k_{\mathcal{A}}) \downarrow 0$ for all $i \in \mathcal{A}$, we have $g_i(\hat z) = 0$ for $i \in \mathcal{A}$. Therefore, $\hat z$ is feasible. Further, by taking limits, we have that $\nabla\phi(\hat z) + \sum_{i\in\mathcal{A}} \hat\lambda_i \nabla g_i(\hat z) = 0$.

Because of the fourth test in step 4, we have either that $\lambda^{k+1}_{\mathcal{A}} \geq 0$ for infinitely many $k \geq K$ or that kkt-test$(z^{k+1}, \lambda^{k+1}_{\mathcal{A}}, \mathcal{A}, \|\lambda^k_{\mathcal{A}}\|_1 + 2)$ is true for infinitely many $k \geq K$ (or possibly both). In the former case, we have that $\hat\lambda_{\mathcal{A}} \geq 0$, so by extending $\hat\lambda_{\mathcal{A}}$ to a full multiplier vector $\hat\lambda$ by adding zero components at $i \notin \mathcal{A}$, we conclude that $(\hat z, \hat\lambda)$ satisfies the KKT conditions (1.3). In the latter case, let $\tilde\lambda^{k+1}_{\mathcal{A}}$ be the vector of multipliers that makes kkt-test$(z^{k+1}, \lambda^{k+1}_{\mathcal{A}}, \mathcal{A}, \|\lambda^k_{\mathcal{A}}\|_1 + 2)$ true. Since the sequence of such vectors is bounded, there is a subsequential limit $\tilde\lambda_{\mathcal{A}}$ such that $\tilde\lambda_{\mathcal{A}} \geq 0$ and $\nabla\phi(\hat z) + \sum_{i\in\mathcal{A}} \tilde\lambda_i \nabla g_i(\hat z) = 0$. By extending $\tilde\lambda_{\mathcal{A}}$ to a full multiplier vector $\tilde\lambda$ by adding zero components at $i \notin \mathcal{A}$, we conclude that $(\hat{z}, \tilde{\lambda})$ satisfies the KKT conditions (1.3). In either case, we conclude that $\hat{z}$ is a KKT point for (1.1).

To verify the rate of convergence, note that for all k sufficiently large, we have from the first and second tests in step 4 that

$$\begin{aligned}
\|(z^k, \lambda^k_{\mathcal{A}}) - (z^*, \hat{\lambda}_{\mathcal{A}})\|_1 &\le \sum_{j=k}^{\infty} \|(\Delta z^j, \Delta\lambda^j_{\mathcal{A}})\|_1 \\
&\le \sum_{j=k}^{\infty} \bar{\eta}(z^j, \lambda^j_{\mathcal{A}})^{\sigma} \\
&\le \sum_{j=k}^{\infty} (\bar{\eta}(z^k, \lambda^k_{\mathcal{A}})^{\sigma})^{(1+\sigma)^{j-k}} \\
&\le \sum_{j=k}^{\infty} (\bar{\eta}(z^k, \lambda^k_{\mathcal{A}})^{\sigma})^{1+(j-k)\sigma} \\
&= \bar{\eta}(z^k, \lambda^k_{\mathcal{A}})^{\sigma} \sum_{j=k}^{\infty} \bar{\eta}(z^k, \lambda^k_{\mathcal{A}})^{(j-k)\sigma^2} \\
&\le \bar{\eta}(z^k, \lambda^k_{\mathcal{A}})^{\sigma} / (1 - \bar{\eta}(z^k, \lambda^k_{\mathcal{A}})^{\sigma^2}) \\
&\le 10\bar{\eta}(z^k, \lambda^k_{\mathcal{A}})^{\sigma},
\end{aligned}$$

where the final inequality follows from

$$\bar{\eta}(z^k, \lambda^k_{\mathcal{A}})^{\sigma^2} \le \tau_{\mathrm{EQ}}^{\sigma^2} \le (1/2)^{1/4} < .9.$$

Since $\{\bar{\eta}(z^k, \lambda^k_{\mathcal{A}})^{\sigma}\}$ converges Q-superlinearly to zero with Q-order $(1+\sigma)$, we have that the sequence $\{z^k\}$ converges R-superlinearly to $\hat{z}$ with R-order $1+\sigma$, as claimed. ∎

Our next result shows that if the outer strategy is steering the iterates toward the solution set $\mathcal{S}$ for (1.1), and under appropriate assumptions on the boundedness of the multiplier estimates and the second-order condition, the algorithm will eventually enter and remain inside the EQ phase, and will converge superlinearly to a KKT point for (1.1). Note that we cannot be certain (without additional assumptions on the problem) that the stationary limit point lies in $\mathcal{S}$; the limit of $\{z^k\}$ may be different from z^*. Hence, this result combines elements of both local and global convergence analysis.

THEOREM 5.4. *Suppose that Assumption* 1 *holds for* (1.1) *at* z^**, that all iterates* (z^k, λ^k) *generated by the outer strategy have* $\|\lambda^k\|_1 \le c$ *for some* $c > 0$*, and that the second-order condition* (2.4) *is satisfied for all* $\lambda^* \in \mathcal{S}_\lambda$ *with* $\|\lambda^*\|_1 \le c$*. Suppose too that the subsequence of iterates* (z^k, λ^k) *generated by the outer strategy is either finite or else satisfies* $\operatorname{dist}((z^k, \lambda^k), \mathcal{S}) \to 0$*. Then Framework INEQ either terminates finitely, with the final point* z^k *satisfying* (1.3) *for some* λ^**, or else it eventually enters the EQ phase and remains there, and the sequence* $\{z^k\}$ *converges R-superlinearly to a KKT point* $\bar{z}$ *satisfying* (1.3) *with R-order* $1+\sigma$*.*

Proof. It is not difficult to show that if Framework INEQ terminates finitely (in either step 1 or step 4), it does so at a point z^k that satisfies the KKT conditions (1.3). (If the termination occurs in step 4, however, the final multiplier $\lambda^k_{\mathcal{A}}$ may not be appropriate for (1.3) as some of its components may be negative. However, the kkt-test (5.2) ensures the existence of an appropriate multiplier vector.) For the remainder of the proof, we assume that finite termination does not occur.

For the main part of the proof, we assume for contradiction that Framework INEQ never stays permanently in the EQ phase.

We define the set $\mathcal{V}_\lambda$ in Theorem 3.2 as

$$\mathcal{V}_\lambda = \{\lambda^* \in \mathcal{S}_\lambda \mid \|\lambda^*\|_1 \le c+1\},$$

and $\hat{\mathcal{V}}_\gamma$ in Theorem 4.2, for the problem EQ($\mathcal{A}^*$), as

$$\hat{\mathcal{V}}_\gamma = \{\lambda^*_{\mathcal{A}^*} \in \bar{\mathcal{S}}_\gamma \mid \|\lambda^*_{\mathcal{A}^*}\|_1 \le c+1\}. \tag{5.3}$$

Accordingly, we define $\mathcal{V} = \{z^*\} \times \mathcal{V}_\lambda$ and $\hat{\mathcal{V}} = \{z^*\} \times \hat{\mathcal{V}}_\gamma$. Since $\operatorname{dist}((z^k, \lambda^k), \mathcal{S}) \to 0$ on the outer iterations, and since $\|\lambda^k\|_1 \le c$, the projection of (z^k, λ^k) onto $\mathcal{S}$ will eventually lie in $\mathcal{V}$, so by Theorem 3.2, there is $\beta > 0$ such that

$$\eta(z^k, \lambda^k) \in [1/\beta, \beta] \operatorname{dist}((z^k, \lambda^k), \mathcal{S})$$

for all outer strategy iterations k sufficiently large. (For simplicity, as in the proof of Theorem 4.2, we drop the explicit dependence of constants such as β, $\bar{\beta}$, $\bar{\delta}_1$, δ_5, etc. on the sets $\mathcal{V}_\lambda$, $\hat{\mathcal{V}}_\gamma$, $\bar{\mathcal{V}}_\gamma$, and the scalar ϵ.) It follows from Theorem 3.3 that step 1 sets $\mathcal{A}(z^k, \lambda^k) = \mathcal{A}^*$ at all outer iterations k sufficiently large. Since step 1 is executed infinitely often, we can therefore assume, as we do in the remainder of the proof, that $\mathcal{A} = \mathcal{A}^*$ for *all* k sufficiently large (both for outer-strategy iterations and iterations in the EQ phase).

Consider now the EQ problem EQ($\mathcal{A}^*$). Because of Assumption 1, there is $\upsilon > 0$ such that the second-order condition (2.9) is satisfied for problem EQ($\mathcal{A}^*$) for all $\lambda^*_{\mathcal{A}^*} \in \hat{\mathcal{V}}_\gamma$. By choosing ϵ in Theorem 4.2 sufficiently small, we have that (2.9) is in fact satisfied for all $\lambda^*_{\mathcal{A}^*} \in \bar{\mathcal{V}}_\gamma$, where $\bar{\mathcal{V}}_\gamma$ is defined in (4.6), possibly with a reduced (but positive) value of υ. Defining $\bar{\mathcal{V}} = \{z^*\} \times \bar{\mathcal{V}}_\gamma$ as before, we have from $\|\lambda^k_{\mathcal{A}^*}\|_1 \le \|\lambda^k\|_1 \le c$ that for all outer-strategy iterations k sufficiently large, the following relations hold:

$$\begin{aligned}
&\operatorname{dist}((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{V}}) \\
&\le \operatorname{dist}((z^k, \lambda^k_{\mathcal{A}^*}), \hat{\mathcal{V}}) && \text{since } \hat{\mathcal{V}} \subset \bar{\mathcal{V}} \\
&\le \left[\operatorname{dist}((z^k, \lambda^k_{\mathcal{A}^*}), \hat{\mathcal{V}})^2 + \|\lambda^k_{(\mathcal{A}^*)^c}\|_2^2\right]^{1/2} \\
&\le \operatorname{dist}((z^k, \lambda^k), \mathcal{V}) && \text{since } \lambda^* \in \mathcal{V}_\lambda \Rightarrow \lambda^*_{\mathcal{A}^*} \in \hat{\mathcal{V}}_\gamma \\
&= \operatorname{dist}((z^k, \lambda^k), \mathcal{S}) \to 0,
\end{aligned} \tag{5.4}$$

where $\lambda^k_{(\mathcal{A}^*)^c}$ is the vector whose components are λ^k_i, $i \notin \mathcal{A}^*$, and the last equality follows from the fact that the projection (z^*, λ^{k*}) of (z^k, λ^k) onto the set $\mathcal{S}$ must eventually have $\|\lambda^{k*}\|_1 \le c+1$, and must therefore also lie in the set $\mathcal{V}$. Hence, the assumptions of Theorem 3.4, Lemma 4.1, and Theorem 4.2 are satisfied by EQ($\mathcal{A}^*$) for all outer-strategy iterations k sufficiently large. In particular, we have from Theorem 3.4 that there is a constant $\bar{\beta} > 0$ such that

$$\bar{\eta}(z^k, \lambda^k_{\mathcal{A}^*}) \in [1/\bar{\beta}, \bar{\beta}] \operatorname{dist}((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}}),$$

for all outer-strategy iterations k sufficiently large.

Given that Framework INEQ never remains permanently in the EQ phase, we claim now that it enters this phase infinitely often. Otherwise, after the final departure from the EQ phase, τ_{EQ} will be at a fixed nonzero value for the remainder of the iterations. However, in this case, some subsequent iteration k will have

$$\bar{\eta}(z^k, \lambda^k_{\mathcal{A}^*}) \le \bar{\beta}\,\text{dist}\,((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}}) \le \bar{\beta}\,\text{dist}\,((z^k, \lambda^k), \mathcal{S}) \to 0,$$

so that the threshold test in step 2 will be satisfied at some subsequent k, activating the EQ phase once again.

Since the algorithm enters the EQ phase infinitely often, it eventually enters with

$$\text{dist}\,((z^k, \lambda^k_{\mathcal{A}^*}), \hat{\mathcal{V}}) \le \text{dist}\,((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}}) \le \text{dist}\,((z^k, \lambda^k), \mathcal{S}) \le \hat{\delta} \tag{5.5}$$

for any choice of small positive $\hat{\delta}$. We now derive conditions on $\hat{\delta}$ that ensure that every EQ phase iteration $j \ge k$ survives the four tests in step 4, so that all subsequent iterates remain in the EQ phase.

First, we ensure that the conditions of Theorems 3.4 and 4.2 are satisfied by setting

$$\hat{\delta} \le \delta_5. \tag{5.6}$$

This condition ensures that the estimates (4.4) and (4.5) hold at all subsequent iterations of the EQ phase—and indeed that (4.9) are satisfied as well. In particular, we have from (4.9b), the definition of δ_3 in Lemma 4.1, and Theorem 3.4 that the estimate (3.4) holds for all iterations $j \ge k$ in the EQ phase. Note too from (4.9a) that $\bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})$ decreases monotonically at all subsequent iterations $j \ge k$ in the EQ phase.

To handle the first test in step 4, we require that

$$\chi_1 \bar{\beta}^{1-\sigma} \hat{\delta}^{1-\sigma} \le 1. \tag{5.7}$$

Then for all iterates $j \ge k$ in the EQ phase, we have from (3.4) and (4.9a) that

$$\begin{aligned}
&\|(\Delta z^j, \Delta\lambda^j_{\mathcal{A}^*})\|_1 \\
&\le \chi_1 \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*}) && \text{from (4.4) in Lemma 4.1} \\
&= \left[\chi_1 \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})^{1-\sigma}\right] \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})^{\sigma} \\
&\le \left[\chi_1 \bar{\eta}(z^k, \lambda^k_{\mathcal{A}^*})^{1-\sigma}\right] \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})^{\sigma} && \text{by (4.9a)} \\
&\le \left[\chi_1 \bar{\beta}^{1-\sigma} \text{dist}\,((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}})^{1-\sigma}\right] \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})^{\sigma} && \text{by Theorem 3.4} \\
&\le \left[\chi_1 \bar{\beta}^{1-\sigma} \hat{\delta}^{1-\sigma}\right] \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})^{\sigma} && \text{by (5.5)} \\
&\le \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})^{\sigma} && \text{by (5.7).}
\end{aligned}$$

Hence, all iterates $j \ge k$ in the EQ phase survive the first test in step 4 when (5.7) holds.

To handle the second conditional statement in step 4, we require that

$$\hat{\delta} \le \chi_2^{-1/(1-\sigma)} / \bar{\beta}. \tag{5.8}$$

We then have from Theorem 3.4 and (5.5) that

$$\bar{\eta}(z^k, \lambda^k_{\mathcal{A}^*}) \le \bar{\beta}\,\text{dist}\,((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}}) \le \bar{\beta}\hat{\delta} \le \chi_2^{-1/(1-\sigma)}.$$

Hence, at all iterates $j \geq k$ in the EQ phase, we have from the bound above together with Lemma 4.1 and (4.9a) that

$$\begin{aligned}\bar{\eta}(z^{j+1}, \lambda^{j+1}_{\mathcal{A}^*}) &\leq \chi_2 \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})^2 \\ &= \left[\chi_2 \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})^{1-\sigma}\right] \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})^{1+\sigma} \\ &\leq \left[\chi_2 \bar{\eta}(z^k, \lambda^k_{\mathcal{A}^*})^{1-\sigma}\right] \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})^{1+\sigma} \\ &\leq \bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*})^{1+\sigma},\end{aligned}$$

ensuring that all iterates $j \geq k$ in the EQ phase survive the second test.

For the third condition in step 4, we first define $\delta_6 > 0$ small enough that $g_i(z) < 0$ for all $i \notin \mathcal{A}^*$ and all z with $\|z - z^*\| \leq \delta_6$. By requiring

$$\hat{\delta} \leq \delta_6 / \bar{\beta}^2, \tag{5.9}$$

we have from Theorem 3.4 and (4.9a) that for all iterations $j \geq k$ in the EQ phase,

$$\begin{aligned}&\|z^j - z^*\| \leq \operatorname{dist}\left((z^j, \lambda^j_{\mathcal{A}^*}), \bar{\mathcal{S}}\right) \\ &\leq \bar{\beta}\bar{\eta}(z^j, \lambda^j_{\mathcal{A}^*}) \leq \bar{\beta}\bar{\eta}(z^k, \lambda^k_{\mathcal{A}^*}) \leq \bar{\beta}^2 \operatorname{dist}\left((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}}\right) \leq \bar{\beta}^2 \hat{\delta} \leq \delta_6,\end{aligned}$$

ensuring that each subsequent iteration in the EQ phase also survives the third test in step 4.

We turn now to the fourth test in step 4. We claim that the condition in this test will hold if $\hat{\delta}$ satisfies

$$n^{1/2} \bar{\beta}^{3-\sigma} \hat{\delta}^{1-\sigma} < 1. \tag{5.10}$$

As a preliminary, we show that there exists a vector $\lambda^*_{\mathcal{A}^*} \in \hat{\mathcal{V}}_\gamma$ with $e^T \lambda^*_{\mathcal{A}^*} \leq \|\lambda^j_{\mathcal{A}^*}\|_1 + 2$, for all $j \geq k$ in the EQ phase, and which therefore satisfies the first two conditions in (5.2) for $B = \|\lambda^j_{\mathcal{A}^*}\|_1 + 2$, as defined in step 4. Defining λ^* by projecting (z^k, λ^k) onto $\mathcal{S}$, we have

$$\begin{aligned}\|\lambda^*_{\mathcal{A}^*}\|_1 &\leq \|\lambda^k_{\mathcal{A}^*}\|_1 + n^{1/2} \|\lambda^k_{\mathcal{A}^*} - \lambda^*_{\mathcal{A}^*}\|_2 \\ &\leq \|\lambda^k_{\mathcal{A}^*}\|_1 + n^{1/2} \operatorname{dist}\left((z^k, \lambda^k), \mathcal{S}\right) \\ &\leq c + n^{1/2} \hat{\delta} \leq c + 1,\end{aligned}$$

where the final inequality follows from (5.10), since $n^{1/2}\hat{\delta} \leq (n^{1/2}\hat{\delta})(\bar{\beta}^{3-\sigma}\hat{\delta}^{-\sigma}) = n^{1/2}\bar{\beta}^{3-\sigma}\hat{\delta}^{1-\sigma} < 1$. Hence, $\lambda^*_{\mathcal{A}^*} \in \hat{\mathcal{V}}_\gamma \subset \bar{\mathcal{V}}_\gamma$, where $\hat{\mathcal{V}}_\gamma$ is defined in (5.3) and $\bar{\mathcal{V}}_\gamma$ is defined in (4.6). We have from (4.9b) and the definition of δ_3 in Lemma 4.1 that

$$\|z^j - z^*\| \leq \operatorname{dist}\left((z^j, \lambda^*_{\mathcal{A}^*}), \bar{\mathcal{V}}\right) = \|z^j - z^*\| \leq \delta_3 \leq \bar{\delta}_1, \tag{5.11}$$

so the estimate (3.4) holds for $(z^j, \lambda^*_{\mathcal{A}^*})$ for all iterations $j \geq k$ in the EQ phase.

By applying (4.9c), (5.10), and the fact that $\delta_3 \leq 1$ (see Lemma 4.1), we have that

$$\begin{aligned}\|\lambda^*_{\mathcal{A}^*}\|_1 &\leq \|\lambda^j_{\mathcal{A}^*}\|_1 + \|\lambda^k_{\mathcal{A}^*} - \lambda^*_{\mathcal{A}^*}\|_1 + \sum_{i=k}^{j-1} \|\Delta\lambda^i_{\mathcal{A}^*}\|_1 \\ &\leq \|\lambda^j_{\mathcal{A}^*}\|_1 + n^{1/2}\hat{\delta} + \min(\delta_3, \epsilon) \\ &\leq \|\lambda^j_{\mathcal{A}^*}\|_1 + 1 + 1,\end{aligned}$$

showing that the inequality $\|\lambda^*_{\mathcal{A}^*}\|_1 \le B = \|\lambda^j_{\mathcal{A}^*}\|_1 + 2$ is satisfied for all iterations $j \ge k$ in the EQ phase.

Turning now to kkt-test$(z^j + \Delta z^j, \lambda^j_{\mathcal{A}^*} + \Delta\lambda^j_{\mathcal{A}^*}, \mathcal{A}, B)$ with $B = \|\lambda^j_{\mathcal{A}^*}\|_1 + 2$, we have for all iterates $j \ge k$ in the EQ phase that

$$\begin{aligned}
&\min_{\lambda_{\mathcal{A}^*}\ge 0,\, \|\lambda_{\mathcal{A}^*}\|_1 \le B} \left\| \nabla\phi(z^j) + \sum_{i\in\mathcal{A}^*} \lambda_i \nabla g_i(z^j) \right\|_1 \\
&\le \min_{\lambda_{\mathcal{A}^*}\ge 0,\, \|\lambda_{\mathcal{A}^*}\|_1 \le B} n^{1/2} \left\| \nabla\phi(z^j) + \sum_{i\in\mathcal{A}^*} \lambda_i \nabla g_i(z^j) \right\|_2 \\
&\le \min_{\lambda_{\mathcal{A}^*}\ge 0,\, \|\lambda_{\mathcal{A}^*}\|_1 \le B} n^{1/2} \bar\eta(z^j, \lambda_{\mathcal{A}^*}) \\
&\le n^{1/2}\bar\eta(z^j, \lambda^*_{\mathcal{A}^*}) && \text{for } \lambda^*_{\mathcal{A}^*} \text{ defined above} \\
&\le n^{1/2}\bar\beta \operatorname{dist}((z^j, \lambda^*_{\mathcal{A}^*}), \bar{\mathcal{S}}) && \text{by (5.11) and Theorem 3.4} \\
&\le n^{1/2}\bar\beta \|z^j - z^*\| && \text{since } \lambda^*_{\mathcal{A}^*} \in \bar{\mathcal{S}} \\
&\le n^{1/2}\bar\beta \operatorname{dist}((z^j, \lambda^j_{\mathcal{A}^*}), \bar{\mathcal{S}}), \\
&\le n^{1/2}\bar\beta^2 \bar\eta(z^j, \lambda^j_{\mathcal{A}^*}) && \text{by Theorem 3.4} \\
&\le \left[n^{1/2}\bar\beta^2 \bar\eta(z^j, \lambda^j_{\mathcal{A}^*})^{1-\sigma}\right] \bar\eta(z^j, \lambda^j_{\mathcal{A}^*})^\sigma \\
&\le \left[n^{1/2}\bar\beta^2 \bar\eta(z^k, \lambda^k_{\mathcal{A}^*})^{1-\sigma}\right] \bar\eta(z^j, \lambda^j_{\mathcal{A}^*})^\sigma && \text{by (4.9a)} \\
&\le \left[n^{1/2}\bar\beta^{3-\sigma} \operatorname{dist}((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}})^{1-\sigma}\right] \bar\eta(z^j, \lambda^j_{\mathcal{A}^*})^\sigma && \text{by Theorem 3.4} \\
&\le \left[n^{1/2}\bar\beta^{3-\sigma} \hat\delta^{1-\sigma}\right] \bar\eta(z^j, \lambda^j_{\mathcal{A}^*})^\sigma && \text{by (5.5)} \\
&< \bar\eta(z^j, \lambda^j_{\mathcal{A}^*})^\sigma && \text{by (5.10).}
\end{aligned}$$

Therefore, we have that

$$\min_{\lambda_{\mathcal{A}^*}\ge 0,\, \|\lambda_{\mathcal{A}^*}\|_1 \le B} \left\| \nabla\phi(z^j) + \sum_{i\in\mathcal{A}^*} \lambda_i \nabla g_i(z^j) \right\|_1 \le \bar\eta(z^j, \lambda^j_{\mathcal{A}^*})^\sigma$$

for all $j \ge k$ in the EQ phase, so iterate j survives the fourth test in step 4.

We conclude that when the EQ phase is entered at some iteration k such that $\mathcal{A}(z^k, \lambda^k) = \mathcal{A}^*$ and (5.5) holds, where $\hat\delta$ satisfies the conditions (5.6), (5.7), (5.8), (5.9), (5.10), that all subsequent iterates remain in the EQ phase, contradicting our assertion that the iterates never eventually remain inside the EQ phase.

Therefore, unless Framework INEQ terminates finitely, the iterates eventually enter and remain permanently in the EQ phase for some fixed set $\mathcal{A}$. The final claims of the proof are an immediate consequence of Theorem 5.3. ∎

Our final result is a more typical local result that assumes a starting point (z^0, λ^0) close to a solution $(z^*, \lambda^*) \in \mathcal{S}$ at which (2.4) is satisfied.

THEOREM 5.5. *Suppose that Assumption* 1 *holds for* (1.1) *at some* $\lambda^* \in \mathcal{S}_\lambda$ *and that the second-order condition* (2.4) *holds at* λ^*. *Then if the starting point* (z^0, λ^0) *is sufficiently close to* (z^*, λ^*), *Framework INEQ will enter the EQ phase on the first iteration with* $\mathcal{A} = \mathcal{A}^*$, *and subsequently* $\operatorname{dist}((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}})$ *converges Q-quadratically to zero and* $\{z^k\}$ *converges R-quadratically to* z^*.

Proof. We start by setting $\mathcal{V}_\lambda = \{\lambda^*\}$ and $\hat{\mathcal{V}}_\gamma = \{\lambda^*_{\mathcal{A}^*}\}$ for purposes of applying the results of sections 3 and 4. By choosing (z^0, λ^0) sufficiently close to (z^*, λ^*), we can ensure both that the active set $\mathcal{A}^*$ is identified correctly in step 1 and that the EQ phase is entered in step 2.

It remains to show that all subsequent iterates remain in the EQ phase, that is, that all iterations pass the tests in step 4. The logic here is essentially identical to the proof of Theorem 5.4, so we do not repeat it here. However, by forcing (z^0, λ^0) to be sufficiently close to (z^*, λ^*) we can prove a stronger result concerning the limit and the rate of convergence. We have from Theorem 3.4 that $\operatorname{dist}((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}}) \le \bar{\beta}\bar{\eta}(z^k, \lambda^k_{\mathcal{A}^*})$ for all k, so that in fact $\operatorname{dist}((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}}) \to 0$. Hence, we can apply Lemma 4.1 to obtain quadratic convergence of $\bar{\eta}(z^k, \lambda^k_{\mathcal{A}^*})$ to zero, and by applying Theorem 3.4 again we have that $\operatorname{dist}((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}})$ also converges Q-quadratically to zero. The R-quadratic convergence claim follows from $\|z^k - z^*\| \le \operatorname{dist}((z^k, \lambda^k_{\mathcal{A}^*}), \bar{\mathcal{S}})$. □

In this result, the second-order condition and boundedness of the multiplier estimates for the original problem (1.1) are needed only to ensure that the identification of the active constraint set $\mathcal{A}^*$ in step 1 is immediately correct. For fast convergence of the EQ phase, we need only the (weaker) second-order condition for the problem EQ($\mathcal{A}^*$).

Theorem 5.5 is similar to the local convergence results in Hager [8], Fischer [6, Theorem 8], and Wright [17, Theorem 7] the main difference being that the second-order condition assumed here is less restrictive. It uses similar assumptions to the result in Izmailov and Solodov [11], which dealt with a different algorithm.

6. Discussion. Several interesting types of nonlinear optimization problems satisfy conditions such as those listed above. A simple case is the one in which the equality constrained problem

$$\min_z \phi(z) \qquad \text{subject to } \ h(z) = 0 \tag{6.1}$$

is reformulated by splitting the equality into two inequalities as follows:

$$\min_z \phi(z) \qquad \text{subject to } \ h(z) \ge 0, \ \ h(z) \le 0. \tag{6.2}$$

Even when z^* is a solution of the original problem at which $\nabla h(z^*)$ has full rank (yielding a unique multiplier γ^*), the reformulation (6.2) satisfies no constraint qualifications. The KKT conditions for (6.2) are that there exist γ^*_+ and γ^*_- such that

$$h(z^*) = 0, \qquad \nabla\phi(z^*) - \nabla h(z^*)\gamma^*_+ + \nabla h(z^*)\gamma^*_-, \qquad \gamma^*_+ \ge 0, \ \ \gamma^*_- \ge 0.$$

Clearly, the multipliers for this formulation are neither unique nor bounded; any vectors γ^*_+, γ^*_- such that

$$\gamma^* = \gamma^*_- - \gamma^*_+, \qquad \gamma^*_+ \ge 0, \quad \gamma^*_- \ge 0, \tag{6.3}$$

will suffice. The second-order sufficient condition (2.9) for (6.1) is that

$$\begin{gathered} w^T \left(\nabla^2\phi(z^*) + \textstyle\sum_{i=1}^m \gamma^*_i \nabla^2 h_i(z^*)\right) w \ge v\|w\|^2 \\ \text{for all } w \text{ such that } \quad \nabla h_i(z^*)^T w = 0, \ \ i = 1, 2, \dots, p. \end{gathered} \tag{6.4}$$

Since $\mathcal{A}^*_0 = \emptyset$, we have following (2.4) that the second-order condition for (6.2) is that

$$\begin{gathered} w^T \left(\nabla^2\phi(z^*) - \textstyle\sum_{i=1}^m (\gamma^*_+)_i \nabla^2 h_i(z^*) + \sum_{i=1}^m (\gamma^*_-)_i \nabla^2 h_i(z^*)\right) w \ge v\|w\|^2 \\ \text{for all } w \text{ such that } \quad \nabla h_i(z^*)^T w = 0, \ \ i = 1, 2, \dots, p. \end{gathered} \tag{6.5}$$

Obviously, the set of directions in (6.5) is identical to the set in (6.4). It is easy to see that the relation (6.4) holds for *all* γ_+^*, γ_-^* satisfying (6.3). Hence, the second-order assumption in Theorem 5.4 holds for any choice of the multiplier bound c. Therefore, for any "reasonable" outer strategy which ensures boundedness of the multipliers and convergence of $(z^k, \gamma_+^k, \gamma_-^k)$ to the solution set for (6.2), Framework INEQ will converge superlinearly. The stabilization term in the step calculation (4.1), (4.2) overcomes the potentially woeful consequences of the inferior formulation (6.2).

More general examples of a similar type, in which some of the inequality constraints in (1.1) are fixed linear combinations of others, can also be constructed.

Another class of interest is mathematical programs with equilibrium constraints (MPECs) which, when formulated as nonlinear programs in the usual ways, fail to satisfy constraint qualifications at any feasible point. The second-order condition (2.4) corresponds in this case to the second-order condition for the relaxed nonlinear programming formulation (Fletcher et al. [7, eq. (3.2)]). (These are slightly more restrictive than the second-order condition for the MPEC itself.) Izmailov and Solodov [11, section 4] give examples in which these conditions are satisfied locally for some multipliers and show convergence of their algorithm in these cases. Theorem 5.5 implies that Framework INEQ has similar rapid local convergence properties in these cases.

We illustrate the performance of the constraint identification and convergence properties of our approach using the following MPEC in two variables:

$$\min z_2 \ \text{ subject to } \ z_1 \geq 0, \ \ z_2 \geq 0, \ \ z_1 z_2 \leq 0, \ \ z_2^2 \geq 1, \tag{6.6}$$

whose solution is $z^* = (0, 1)$ with optimal objective 1 and active set $\mathcal{A}^* = \{1, 3, 4\}$ (where we assume that the constraints are ordered as written). The matrix of active constraint gradients is obviously rank deficient and the optimal multiplier set for this problem is

$$\mathcal{S}_\lambda = \{(\alpha, 0, \alpha, 0.5) \,|\, \alpha \geq 0\},$$

so that $\mathcal{A}_+^* = \{1, 3, 4\}$ and $\mathcal{A}_0^* = \emptyset$. The second-order conditions for this problem are satisfied vacuously, since the direction set in (2.4) is simply the zero vector.

This problem has the property that standard SQP may fail to produce a valid step for some z arbitrarily close to z^*, since the linearized constraints are inconsistent at $z = (\epsilon, 1 - \epsilon)$ for all small positive ϵ [7, section 2.2]. We consider the performance of one step of Framework INEQ from such a point, along with a Lagrange multiplier estimate that is a similar distance from $\mathcal{S}_\lambda$. We set

$$z^0 = (\epsilon, 1 - \epsilon), \quad \lambda^0 = (1, 0, 1, .5) + \epsilon v, \tag{6.7}$$

where $v \in \mathbb{R}^4$ is a vector whose elements are drawn from a uniform distribution over $[-1, 1]$. Obviously, we have that $\operatorname{dist}((z^0, \lambda^0), \mathcal{S}) = O(\epsilon)$ and it is easy to show that $\eta(z^0, \lambda^0) = O(\epsilon)$ as well (in accordance with Theorem 3.2). For all ϵ sufficiently small, we have that $\mathcal{A}(z^0, \lambda^0) = \{1, 3, 4\} = \mathcal{A}^*$ and also that $\bar{\eta}(z^0, \lambda^0_{\mathcal{A}^*}) = O(\epsilon)$. Hence, the test in step 2 of Framework INEQ is satisfied, and the method enters the EQ phase.

Table 6.1 shows the effect of the first step taken in the EQ phase for various values of ϵ. (The results were obtained from a Matlab implementation.) The second column shows that the estimate of $\mathcal{A}$ is precise for all values tried (though it is usually inaccurate for $\epsilon \geq 1/4$). The third and fourth columns show the values of $\bar{\eta}(z^k, \lambda^k_{\mathcal{A}^*})$ for $k = 0$ and $k = 1$. (We use the notation $3.9(-1)$ to denote 3.9×10^{-1}, and so on.) The final column gives an estimate of the Q-order of convergence of $\bar{\eta}$ to 0 over this step. The Q-quadratic convergence of Theorem 5.5 is apparent.

TABLE 6.1
One step of framework INEQ from starting point (6.7).

ϵ	$\mathcal{A}$	$\bar\eta_0$	$\bar\eta_1$	$\log\bar\eta_1/\log\bar\eta_0$
2^{-3}	$\{1,3,4\}$	3.9(−1)	3.9(−2)	3.5
2^{-5}	$\{1,3,4\}$	1.1(−1)	1.2(−3)	3.0
2^{-10}	$\{1,3,4\}$	3.0(−3)	2.1(−6)	2.2
2^{-15}	$\{1,3,4\}$	1.0(−4)	1.4(−9)	2.2
2^{-20}	$\{1,3,4\}$	3.1(−6)	1.4(−12)	2.1
2^{-30}	$\{1,3,4\}$	2.7(−9)	7.1(−19)	2.1
2^{-40}	$\{1,3,4\}$	2.5(−12)	4.2(−25)	2.1

The linear system (4.1), (4.2) that is solved to obtain the step tabulated in Table 6.1 is evidently ill conditioned for small ϵ. (We see this by noting that $\nabla h(z) \in \mathbb{R}^{2\times 3}$ has column rank 2, while the 3×3 diagonal block below it has $\mu = O(\epsilon)$ and hence is approaching zero.) It is interesting to ask, therefore, if the computed version of the step $(\Delta z, \Delta\gamma)$ will differ from the exact version so markedly as to destroy the rapid convergence properties of the algorithm. This issue has essentially been analyzed in [15]. The analysis takes into account both the errors in evaluating the matrix and right-hand side of (4.2) and the roundoff errors that occur in factorizing and solving this system using a numerically stable scheme. The evaluation errors may be particularly significant in the right-hand side of (4.2), since while all components of this vector are theoretically of size $O(\bar\eta)$, they are generally computed by manipulating intermediate quantities of size $O(1)$, so may contain errors of absolute size similar to $\mathbf{u}$, where $\mathbf{u}$ is unit roundoff.

The analysis in [15], translated appropriately into the current setting, is based on a singular value decomposition of $\nabla h(z^*)$, namely,

$$\nabla h(z^*) = \left[\begin{array}{cc} \hat U & \hat V \end{array}\right] \left[\begin{array}{cc} S & 0 \\ 0 & 0 \end{array}\right] \left[\begin{array}{c} U^T \\ V^T \end{array}\right], \tag{6.8}$$

where S is positive diagonal, and $\hat U$, $\hat V$, U, and V are orthonormal matrices. While the exact step $(\Delta z, \Delta\gamma)$ is of size $O(\bar\eta)$ (Lemma 4.1), it is shown in [15] that the step $(\widehat{\Delta z}, \widehat{\Delta\gamma})$ computed under the assumptions above satisfies

$$(\widehat{\Delta z} - \Delta z, U^T(\widehat{\Delta\gamma} - \Delta\gamma)) = O(\bar\eta^2) + \delta_{\mathbf{u}},$$

where $\delta_{\mathbf{u}}$ represents a vector whose entries are bounded by a modest multiple of $\mathbf{u}$. The errors introduced into these components are as modest as we could expect, given that the right-hand side contains absolute errors similar to $\mathbf{u}$. In the remaining component of the step, however, the error is more severe; we have

$$V^T(\widehat{\Delta\gamma} - \Delta\gamma) = O(\bar\eta + \bar\eta^{-1}\delta_{\mathbf{u}}).$$

In this subspace, the ill conditioning of the matrix combines with the evaluation and roundoff errors to produce large differences between true and computed steps. However, as noted in [15], the large error in this space has little effect on the convergence behavior of the algorithm. Roughly speaking, we have

$$\bar\eta(z + \widehat{\Delta z}, \gamma + \widehat{\Delta\gamma}) = O(\bar\eta(z,\gamma)^2 + \delta_{\mathbf{u}}),$$

TABLE 6.2
One step of framework INEQ from starting point (6.7), with simulated roundoff errors.

ϵ	$\bar{\eta}_0$	$\bar{\eta}_1$	$\log \bar{\eta}_1 / \log \bar{\eta}_0$	$(\widehat{\Delta\lambda}_1, \widehat{\Delta\lambda}_3)$
2^{-5}	9.4(−2)	2.9(−3)	2.5	$10^{-2}(.53, .57)$
2^{-10}	2.8(−3)	3.3(−6)	2.1	$10^{-3}(.12, .23)$
2^{-15}	9.9(−5)	1.5(−9)	2.2	$10^{-4}(-.006, .1)$
2^{-20}	3.0(−6)	2.1(−12)	2.1	$10^{-6}(.06, .84)$
2^{-30}	2.7(−9)	2.1(−12)	1.4	$10^{-4}(-.48, -.48)$
2^{-40}	2.7(−12)	1.0(−12)	1.0	$(-.33, -.33)$

so that we can expect to see finite-precision arithmetic interfering with the quadratic convergence of $\bar{\eta}$ only when $\bar{\eta}$ drops below $\sqrt{\mathbf{u}}$.

Our example (6.6) is too small and simple to exhibit the errors discussed in the previous paragraph, but we can simulate the effects of such errors by introducing arbitrary perturbations into (4.2) and tracking the effects of these perturbations on the computed step. For this "simulated roundoff" experiment, we set $\mathbf{u} = 10^{-12}$ and add perturbations of size $\mathbf{u}(\|A\|_2/2)\rho$ (where A is the coefficient matrix in (4.2) and ρ is a scalar drawn from the uniform distribution on $[-1, 1]$) to all components of the matrix and right-hand side in (4.2). For our example, we have

$$\nabla h(z^*) = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}, \qquad V = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix},$$

so we expect to see errors in the computed values of the Lagrange multiplier step of the following form:

$$\begin{bmatrix} \Delta\lambda_1 - \widehat{\Delta\lambda}_1 \\ \Delta\lambda_3 - \widehat{\Delta\lambda}_3 \\ \Delta\lambda_4 - \widehat{\Delta\lambda}_4 \end{bmatrix} \approx \beta \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \approx - \begin{bmatrix} \widehat{\Delta\lambda}_1 \\ \widehat{\Delta\lambda}_3 \\ \widehat{\Delta\lambda}_4 \end{bmatrix}, \tag{6.9}$$

where $|\beta| = O(\bar{\eta} + \bar{\eta}^{-1}\mathbf{u})$.

Table 6.2 shows the behavior of the first step of Framework INEQ, computed as in Table 6.1 but with the introduction of simulated roundoff errors as described above. Interference of these errors with the theoretical convergence behavior is evident only in the last two lines of the table, when $\bar{\eta}_0$ drops below about 10^{-6}. The final column verifies that, as predicted in (6.9), the size of the computed multipliers essentially tracks ϵ for $\epsilon \geq \sqrt{\mathbf{u}}$, but grows in the range space of V like $\bar{\eta}^{-1}\mathbf{u}$ for smaller values of $\bar{\eta}$.

Acknowledgments. I am grateful to the editor Sven Leyffer, to the referees, and to Mihai Anitescu, Mikhail Solodov, and Alexey Izmailov for helpful comments on the first version of the paper.

REFERENCES

[1] M. ANITESCU, *Degenerate nonlinear programming with a quadratic growth condition*, SIAM J. Optim., 10 (2000), pp. 1116–1135.

[2] D. P. BERTSEKAS, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[3] J. F. Bonnans and A. Ioffe, *Second-order sufficiency and quadratic growth for nonisolated minima*, Math. Oper. Res., 20 (1995), pp. 801–819.

[4] F. Facchinei, A. Fischer, and C. Kanzow, *On the accurate identification of active constraints*, SIAM J. Optim., 9 (1998), pp. 14–32.

[5] A. Fischer, *Modified Wilson method for nonlinear programs with nonunique multipliers*, Math. Oper. Res., 24 (1999), pp. 699–727.

[6] A. Fischer, *Local behavior of an iterative framework for generalized equations with nonisolated solutions*, Math. Program. Ser. A, 94 (2002), pp. 91–124.

[7] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes, *Local Convergence of SQP Methods for Mathematical Programs with Equilibrium Constraints*, Numerical Analysis Report NA/209, Department of Mathematics, University of Dundee, May 2002.

[8] W. W. Hager, *Stabilized sequential quadratic programming*, Comput. Optim. Appl., 12 (1999), pp. 253–273.

[9] W. W. Hager and M. S. Gowda, *Stability in the presence of degeneracy and error estimation*, Math. Program. Ser. A, 85 (1999), pp. 181–192.

[10] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl., 4 (1969), pp. 303–320.

[11] A. F. Izmailov and M. V. Solodov, *Newton-type methods for optimization problems without constraint qualifications*, SIAM J. Optim., 15 (2004), pp. 210–228.

[12] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.

[13] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298.

[14] S. M. Robinson, *Generalized equations and their solutions. Part* II: *Applications to nonlinear programming*, Math. Program. Stud., 19 (1982), pp. 200–221.

[15] S. J. Wright, *Superlinear convergence of a stabilized SQP method to a degenerate solution*, Comput. Optim. Appl., 11 (1998), pp. 253–275.

[16] S. J. Wright, *Modifying SQP for degenerate problems*, SIAM J. Optim., 13 (2002), pp. 470–497.

[17] S. J. Wright, *Constraint identification and algorithm stabilization for degenerate nonlinear programs*, Math. Program. Ser. B, 95 (2003), pp. 137–160.