PREPRINT ANL/MCS-P643-0297,
MATHEMATICS AND COMPUTER SCIENCE DIVISION, ARGONNE NATIONAL LABORATORY

# SUPERLINEAR CONVERGENCE OF A STABILIZED SQP METHOD TO A DEGENERATE SOLUTION

STEPHEN J. WRIGHT*

**Abstract.** We describe a slight modification of the well-known sequential quadratic programming method for nonlinear programming that attains superlinear convergence to a primal-dual solution even when the Jacobian of the active constraints is rank deficient at the solution. We show that rapid convergence occurs even in the presence of the roundoff errors that are introduced when the algorithm is implemented in floating-point arithmetic.

**AMS(MOS) subject classifications.** 90C33, 90C30, 49M45

## 1. Introduction.

We describe a slight modification of the well-known sequential quadratic programming (SQP) algorithm for the nonlinear programming problem

$$\min \phi(z) \qquad \text{subject to } g(z) \leq 0, \tag{1}$$

where $\phi : \mathsf{R}^n \to \mathsf{R}$ and $g : \mathsf{R}^n \to \mathsf{R}^m$ are smooth functions. Our particular interest is in the case in which the active constraints at the solution $z^*$—those for which $g_i(z^*) = 0$—have linearly dependent gradients. This property can interfere with the superlinear convergence rate of SQP, as we demonstrate with a simple example. The advantage of our *stabilized SQP* algorithm is that superlinear convergence is still attainable when this property holds.

The Lagrangian for (1) is

$$\mathcal{L}(z,\lambda) = \phi(z) + \sum_{i=1}^{m} \lambda_i g_i(z) = \phi(z) + \lambda^T g(z), \tag{2}$$

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

$$\mathcal{L}_z(z^*,\lambda^*) = 0, \quad g(z^*) \leq 0, \quad \lambda^* \geq 0, \quad (\lambda^*)^T g(z^*) = 0. \tag{3}$$

These relations are the well-known Karush-Kuhn-Tucker (KKT) conditions. The *active set* at $z^*$ is defined by

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

while its complement is

$$\mathcal{N} = \{1, 2, \ldots, m\} \backslash \mathcal{B}. \tag{5}$$

We partition the Lagrange multiplier vector $\lambda$ and the function $g(\cdot)$ according to $\mathcal{B} \cup \mathcal{N}$ as

$$g_{\mathcal{B}}(\cdot) = [g_i(\cdot)]_{i \in \mathcal{B}}, \quad g_{\mathcal{N}}(\cdot) = [g_i(\cdot)]_{i \in \mathcal{N}}, \quad \lambda_{\mathcal{B}} = [\lambda_i]_{i \in \mathcal{B}}, \quad \lambda_{\mathcal{N}} = [\lambda_i]_{i \in \mathcal{N}}.$$

* Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, U.S.A. This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Computational and Technology Research, U.S. Department of Energy, under Contract W-31-109-Eng-38.

For notational convenience, we often omit transpose notation and write $\lambda = (\lambda_{\mathcal{B}}, \lambda_{\mathcal{N}})$ and $(z, \lambda) = (z, \lambda_{\mathcal{B}}, \lambda_{\mathcal{N}})$.

In this article, we assume that for at least one of the points $(z^*, \lambda^*)$ satisfying the first-order conditions (3), we have

$$\lambda^*_{\mathcal{B}} > 0; \tag{6}$$

that is, *strict complementarity* holds. Since $g_i(z^*) < 0$ for $i \in \mathcal{N}$ (by the definition (4)), we have from (3) that $\lambda^*_{\mathcal{N}} = 0$ for all $\lambda^*$ that satisfy (3). We also assume the following second-order sufficient condition: For any $\lambda^*$ such that $(z^*, \lambda^*)$ satisfies the KKT conditions (3), the two-sided projection of the Lagrangian Hessian $\mathcal{L}_{zz}(z^*, \lambda^*)$ onto $\ker Dg_{\mathcal{B}}(z^*)$ is positive definite. That is, there is a $\sigma > 0$ such that

$$\begin{gathered} w^T \mathcal{L}_{zz}(z^*, \lambda^*) w \geq \sigma \|w\|^2, \\ \text{for all } \lambda^* \text{ such that } (z^*, \lambda^*) \text{ satisfies (3), and all } w \in \ker Dg_{\mathcal{B}}(z^*). \end{gathered} \tag{7}$$

Finally, we assume that the Mangasarian-Fromovitz [8] constraint qualification (MFCQ) holds at $z^*$. That is,

$$Dg_{\mathcal{B}}(z^*) \bar{y} < 0 \quad \text{for some } \bar{y} \in \mathrm{R}^n. \tag{8}$$

This assumption is weaker than the linear independence constraint qualification, which assumes that $Dg_{\mathcal{B}}(z^*)$ has full row rank and which is frequently used in the local convergence analysis of algorithms for nonlinear programming.

We use $\mathcal{S}$ to denote the primal-dual solution set whose $z$-component is $z^*$, that is,

$$\mathcal{S} = \{(z^*, \lambda^*) \,|\, \lambda^* \text{ satisfies (3)}\}. \tag{9}$$

Since $\mathcal{L}_z(z^*, \lambda)$ is linear in $\lambda$, it follows immediately that $\mathcal{S}$ is convex. We use $\mathcal{S}_\lambda$ to denote the set of optimal Lagrange multipliers for $z^*$, that is,

$$\mathcal{S}_\lambda = \{\lambda^* \,|\, (z^*, \lambda^*) \in \mathcal{S}\}. \tag{10}$$

In the best-known form of the SQP algorithm, the step $\Delta z$ is obtained by solving the following subproblem:

$$\min_{\Delta z} \Delta z^T D\phi(z) + \tfrac{1}{2} \Delta z^T \mathcal{L}_{zz}(z, \lambda) \Delta z, \quad \text{subject to } g(z) + Dg(z)\Delta z \leq 0, \tag{11}$$

where $(z, \lambda)$ is the current primal-dual iterate. Denoting the Lagrange multipliers for the constraints in (11) by $\lambda^+$, we see that the solution $\Delta z$ satisfies the following KKT conditions (cf. (3)):

$$\begin{aligned} \mathcal{L}_{zz}(z, \lambda)\Delta z + D\phi(z) + Dg(z)^T \lambda^+ &= 0, && \text{(12a)} \\ g(z) + Dg(z)\Delta z &\leq 0, && \text{(12b)} \\ \lambda^+ &\geq 0, && \text{(12c)} \\ (\lambda^+)^T [g(z) + Dg(z)\Delta z] &= 0. && \text{(12d)} \end{aligned}$$

We can derive the subproblem (11) and the KKT conditions (12) from the following min-max problem involving the Lagrangian of (11):

$$\min_{\Delta z} \max_{\lambda^+ \geq 0} \Delta z^T D\phi(z) + \tfrac{1}{2} \Delta z^T \mathcal{L}_{zz}(z, \lambda) \Delta z + (\lambda^+)^T \left(g(z) + Dg(z)\Delta z\right).$$

Most practical implementations of SQP perform a line search either along the primal space in the direction $\Delta z$ or in the primal-dual space in the direction $(\Delta z, \lambda^+ - \lambda)$, with the aim of improving the value of some merit function. When $(z, \lambda)$ is sufficiently close to the primal-dual solution set $\mathcal{S}$, these globalization strategies should allow unit steps to be taken to yield rapid convergence; that is,

$$(z, \lambda) \leftarrow (z + \Delta z, \lambda^+).$$

Under the conditions discussed above, the subproblem (11) has a local solution $\Delta z$ when $(z, \lambda)$ is sufficiently close to $\mathcal{S}$. (For a proof of this statement, see Theorem 4.3 and Section 5 of Robinson [11].) However, the Lagrange multiplier $\lambda^+$ for the linear constraints in (11) may not be uniquely determined by (11) because of rank deficiency in the Jacobian $Dg_{\mathcal{B}}(z)$, so that the Hessian $\mathcal{L}_{zz}$ may not be uniquely defined at the *next* iteration of SQP. More important, SQP may no longer yield a "quadratic" decrease in distance to the solution set $\mathcal{S}$, even from points that are arbitrarily close to this set. We illustrate this fact with the following example.

**Example:** Consider the problem

$$\min z_1 \quad \text{subject to} \quad \begin{array}{rcl} (z_1-2)^2 + z_2^2 & \leq & 4, \\ (z_1-4)^2 + z_2^2 & \leq & 16, \end{array} \tag{13}$$

which has a unique minimizer at $z^* = 0$ at which both constraints are active and MFCQ is satisfied. The optimal multiplier set is defined by

$$\mathcal{S}_\lambda = \{(1/4 - 2\alpha, \alpha) \,|\, 0 \leq \alpha \leq 1/8\}.$$

Given a primal-dual point $(z, \lambda)$, the quantities needed to define the SQP subproblem (11) are as follows:

$$D\phi(z) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \mathcal{L}_{zz}(z,\lambda) = 2(\lambda_1 + \lambda_2) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$
$$g(z) = \begin{bmatrix} (z_1-2)^2 + z_2^2 - 4 \\ (z_1-4)^2 + z_2^2 - 16 \end{bmatrix}, \qquad Dg(z) = 2\begin{bmatrix} (z_1-2) & z_2 \\ (z_1-4) & z_2 \end{bmatrix}.$$

Suppose we apply SQP from the point $z = (\epsilon, \epsilon)$, $\lambda \geq 0$, where $\epsilon$ is small and positive. Note that $\|z - z^*\| = \sqrt{2}\epsilon$. It can be shown that the solution $\Delta z$ of (11) satisfies the following linear system:

$$\begin{bmatrix} 2(\lambda_1+\lambda_2) & 0 & 2(\epsilon-4) \\ 0 & 2(\lambda_1+\lambda_2) & 2\epsilon \\ -2(\epsilon-4) & -2\epsilon & 0 \end{bmatrix} \begin{bmatrix} \Delta z_1 \\ \Delta z_2 \\ \lambda_2^+ \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 2\epsilon^2 - 8\epsilon \end{bmatrix},$$

where $\lambda_2^+$ is the Lagrange multiplier for the second linearized constraint in (11). (The first constraint is inactive.) By solving this system we obtain

$$\begin{aligned} \Delta z_1 &= -\epsilon \frac{8 - \epsilon(2 - 1/4(\lambda_1+\lambda_2)) + O(\epsilon^2)}{8 - 2\epsilon + O(\epsilon^2)} = -\epsilon + O(\epsilon^2), \\ \lambda_2^+ &= \frac{-1}{2(\epsilon-4)} + O(\epsilon) = 1/8 + O(\epsilon), \\ \Delta z_2 &= -\frac{\epsilon\lambda_2^+}{\lambda_1+\lambda_2} = -\frac{\epsilon}{8(\lambda_1+\lambda_2)} + O(\epsilon^2). \end{aligned}$$

If $\lambda = (1/4, 0)$—an optimal multiplier for (13)—we have

$$\Delta z = (-\epsilon, -\epsilon/2) + O(\epsilon^2),$$

and therefore

$$\|(z + \Delta z) - z^*\| = \|(O(\epsilon^2), \epsilon/2 + O(\epsilon^2))\| = \epsilon/2 + O(\epsilon^2),$$

giving just a "linear" decrease in the distance to the primal optimum $z^*$ on this iteration, even for $\epsilon$ arbitrarily small. The new primal-dual iterate $(z + \Delta z, \lambda^+)$ is also only linearly closer to $\mathcal{S}$ than is $(z, \lambda)$. In fact, a linear decrease is obtained when $\lambda$ is *any* optimal multiplier for (13), unless it happens to be close to the extreme point $(0, 1/8)$ of $\mathcal{S}_\lambda$.

We now describe a stabilized variant of SQP for which a quadratic improvement in the error is guaranteed whenever $(z, \lambda)$ is sufficiently close to a certain large subset of the the primal-dual solution set $\mathcal{S}$—a subset that encompasses most of the relative interior of $\mathcal{S}$. Iterates of the stabilized SQP algorithm are obtained by solving the following min-max problem for $(\Delta z, \lambda^+)$:

$$\begin{aligned} \min_{\Delta z} \max_{\lambda^+ \geq 0} \; & \Delta z^T D\phi(z) + \tfrac{1}{2}\Delta z^T \mathcal{L}_{zz}(z, \lambda)\Delta z \\ & + (\lambda^+)^T \left(g(z) + Dg(z)\Delta z\right) - \tfrac{1}{2}\mu\|\lambda^+ - \lambda\|^2, \end{aligned} \tag{14}$$

where the parameter $\mu$ is defined as

$$\mu = \mu(z, \lambda) \stackrel{\text{def}}{=} \left\|(\mathcal{L}_z(z, \lambda), g(z)_+, \lambda^T g(z))\right\|. \tag{15}$$

Note that (14) differs from the standard min-max formulation only in the inclusion of the proximal penalty term $\frac{1}{2}\mu\|\lambda^+ - \lambda\|^2$. The optimality conditions for a candidate solution $(\Delta z, \lambda^+)$ of (14) are likewise similar to (12), namely,

$$\mathcal{L}_{zz}(z, \lambda)\Delta z + D\phi(z) + Dg(z)^T\lambda^+ = 0, \tag{16a}$$

$$g(z) + Dg(z)\Delta z - \mu(\lambda^+ - \lambda) \leq 0, \tag{16b}$$

$$\lambda^+ \geq 0, \tag{16c}$$

$$(\lambda^+)^T \left[g(z) + Dg(z)\Delta z - \mu(\lambda^+ - \lambda)\right] = 0. \tag{16d}$$

We show later that for $(z, \lambda)$ sufficiently close to some strictly complementary primal-dual solution $(z^*, \lambda^*)$, there is a unique solution $(\Delta z, \lambda^+)$ of (16) for which $\|(\Delta z, \lambda^+ - \lambda)\| = O(\mu)$. This solution satisfies the following linear system:

$$\begin{aligned} \begin{bmatrix} \mathcal{L}_{zz}(z, \lambda) & Dg_{\mathcal{B}}(z)^T \\ -Dg_{\mathcal{B}}(z) & \mu I \end{bmatrix} \begin{bmatrix} \Delta z \\ \lambda_{\mathcal{B}}^+ - \lambda_{\mathcal{B}} \end{bmatrix} &= \begin{bmatrix} -D\phi(z) - Dg_{\mathcal{B}}(z)^T\lambda_{\mathcal{B}} \\ g_{\mathcal{B}}(z) \end{bmatrix}, \\ \lambda_{\mathcal{N}}^+ &= 0. \end{aligned} \tag{17}$$

In Section 3, we show that the norm of the stabilized SQP step is small; in fact, $\|(\Delta z, \lambda^+ - \lambda)\|$ approaches zero at the same rate as $\mu$, while other possible solutions $(\Delta z, \lambda^+)$ to (14), if they exist, cannot satisfy this estimate. These observations hold even if the active constraint Jacobian $Dg_{\mathcal{B}}(\cdot)$ is rank deficient at or near the solution $z^*$. We show in Section 4 that a full step along the direction produces a "quadratic" decrease in $\mu$ and in the distance to the solution set and that local quadratic convergence follows as a consequence. Our analysis has much in common with the analysis of Ralph and Wright [9, 10], who deal with an interior-point algorithm rather than an SQP-based algorithm. Our focus on the SQP algorithm has practical significance because of the popularity of this method and because full rank of the active constraint Jacobian is frequently violated (or nearly so) on large-scale problems.

We consider, too, the effects of finite-precision floating-point arithmetic on the step obtained from (17). While the errors in some components of the stabilized SQP steps may grow quite large near the solution, rapid convergence is still attained. Our analysis of the floating-point case follows easily from the exact analysis because we use the tools of linear algebra in proving our results.

Our conclusions extend to the case in which equality constraints are explicitly present, as we outline in Section 5.

**2. Assumptions, Preliminary Results, and Notation.** Throughout the remainder of the article, we make the following assumption.

ASSUMPTION 1. *The vector $z^*$ is a local solution of (1) and the functions $\phi(\cdot)$ and $g(\cdot)$ are twice Lipschitz continuously differentiable in an open neighborhood of $z^*$. The first-order conditions (3) and the second-order condition (7) are satisfied at $z^*$, and the strict complementarity condition (6) holds for some vector $\lambda^*$ for which (3) are satisfied.*

Assumption 1 and the MFCQ (8) lead to two preliminary results that have appeared in previous work, as noted below.

LEMMA 2.1. (Gauvin [5]) *Suppose that Assumption 1 holds. Then $\mathcal{S}_\lambda$ is bounded if and only if the MFCQ (8) is satisfied.*

A proof of the following result can be found in Bertsekas [1, Proposition 3.3.2], for example.

LEMMA 2.2. *If Assumption 1 holds, then $z^*$ is a locally unique solution of (1).*

Given the definition (9) of $\mathcal{S}$ and Lemma 2.2, we define the constant $\xi$ as follows:

$$\xi = \max_{\lambda^* \in \mathcal{S}_\lambda} \min_{i \in \mathcal{B}} \lambda_i^*. \tag{18}$$

Assumption 1 implies that $\xi$ is positive, while Lemma 2.1 implies that it is bounded. For each $\gamma \in (0,1)$, we define $\mathcal{N}^\gamma(\epsilon)$ by

$$\begin{aligned}\mathcal{N}^\gamma(\epsilon) = \{(z,\lambda) \,|\, &\|(z,\lambda) - (z^*,\lambda^*)\| \le \epsilon \\ &\text{for some } \lambda^* \in \mathcal{S}_\lambda \text{ with } \lambda^*_{\mathcal{B}} \ge \gamma \xi e, \text{ and } \lambda \ge 0\}.\end{aligned} \tag{19}$$

From Lemma 2.1, the set $\mathcal{S}$ is compact, so by the definition (18), the set $\{(z^*,\lambda^*) \in \mathcal{S} \,|\, \lambda^*_{\mathcal{B}} \ge \gamma\xi e\}$ is nonempty and compact also for each $\gamma \in (0,1)$.

In the remainder of the article, we assume that the following collection of assumptions is satisfied.

> *Standing Assumptions:* We assume that Assumption 1 holds and that the MFCQ (8) is satisfied. We assume, too, that $\gamma$ used to define (19) is a fixed constant in the range $(0,1)$.

We conclude this section with some items of notation to be used in subsequent sections. We define integers $\bar{m}$ and $\breve{m}$ by

$$\bar{m} = |\mathcal{B}|, \qquad \breve{m} = \operatorname{rank} Dg_{\mathcal{B}}(z^*), \tag{20}$$

so that $0 \le \breve{m} \le \bar{m} \le m$. Since Lemma 2.1 implies boundedness of $\mathcal{S}_\lambda$, there is a constant $\bar{\sigma} > 0$ such that

$$\|\mathcal{L}_{zz}(z^*,\lambda^*)\| \le \bar{\sigma}, \qquad \text{for all } \lambda^* \in \mathcal{S}_\lambda. \tag{21}$$

For convenience, we assume that the problem is scaled so that $\bar{\sigma}$ and $\|Dg(z^*)\|$ are not too much different from 1.

The distance to a set $\mathcal{T}$ is defined as

$$\operatorname{dist}(w, \mathcal{T}) = \inf \{ \|w - w^*\| \mid w^* \in \mathcal{T} \}.$$

We use $P(\cdot)$ to denote projection onto the set of optimal Lagrange multipliers, that is,

$$P(\lambda) = \arg \min_{\lambda^* \in \mathcal{S}_\lambda} \|\lambda^* - \lambda\|.$$

Given two continuous functions $\psi_1(\cdot)$ and $\psi_2(\cdot)$ that map some Euclidean space to $[0, \infty)$, we say that $\psi_1(x) = O(\psi_2(x))$ if there is a constant $\epsilon > 0$ and a moderate constant $C_\epsilon > 0$ such that

$$\psi_2(x) \in [0, \epsilon] \quad \Rightarrow \quad \psi_1(x) \le C_\epsilon \psi_2(x).$$

We say that $\psi_1(x) = \Omega(\psi_2(x))$ if both $\psi_1(x) = O(\psi_2(x))$ and $\psi_2(x) = O(\psi_1(x))$.

We use $\|\cdot\|$ to denote the Euclidean norm of a matrix or vector, and $\kappa(M) = \|M\| \, \|M^{-1}\|$ to denote the condition number of a nonsingular matrix with respect to this norm.

Finally, we mention that when functions such as $g_\mathcal{B}$, $\mathcal{L}_z$, and $\mathcal{L}_{zz}$ appear without specific arguments, the arguments are understood to be the current points $z$ or $(z, \lambda)$, as appropriate.

**3. Step Size Estimates.** In this section, we show that the step calculated from any point $(z, \lambda) \in \mathcal{N}^\gamma(\epsilon)$ via (17) in exact arithmetic satisfies the estimate

$$(\Delta z, \Delta \lambda_\mathcal{B}) \stackrel{\text{def}}{=} (\Delta z, \lambda_\mathcal{B}^+ - \lambda_\mathcal{B}) = O(\mu), \tag{22}$$

while any other local solution of (16) cannot satisfy this estimate. We also discuss the effect of finite-precision floating-point arithmetic on this estimate.

Our first result shows that $\mu$ defined in (15) is closely related to the distance from the current point to the solution set $\mathcal{S}$.

LEMMA 3.1. *Suppose that the standing assumptions hold. Then there is a constant $\epsilon > 0$ such that for all $(z, \lambda) \in \mathcal{N}^\gamma(\epsilon)$ we have*

$$\operatorname{dist}((z, \lambda), \mathcal{S}) = \Omega(\mu). \tag{23}$$

*Proof.* We show first that $\mu = O(\operatorname{dist}((z, \lambda), \mathcal{S}))$. Let $(z^*, P(\lambda))$ be the projection of $(z, \lambda)$ onto the (compact) solution set $\mathcal{S}$, so that

$$\|(z, \lambda) - (z^*, P(\lambda))\| = \operatorname{dist}((z, \lambda), \mathcal{S}).$$

By the optimality condition (3) and Assumption 1, we have

$$\|\mathcal{L}_z(z, \lambda)\| = \|\mathcal{L}_z(z, \lambda) - \mathcal{L}_z(z^*, P(\lambda))\| = O(\operatorname{dist}((z, \lambda), \mathcal{S})). \tag{24}$$

Similarly, we have

$$\|g(z)_+\| = \|g(z)_+ - g(z^*)_+\| \le \|g(z) - g(z^*)\| = O(\operatorname{dist}((z, \lambda), \mathcal{S})). \tag{25}$$

By boundedness of $\mathcal{S}$, we have that $P(\lambda)$ is bounded by a constant independent of $\lambda$. Hence, we can write

$$\begin{aligned}
&\lambda^T g(z) \\
&\quad = \lambda^T g(z) - P(\lambda)^T g(z^*) \\
&\quad = [\lambda - P(\lambda)]^T [g(z) - g(z^*)] + P(\lambda)^T [g(z) - g(z^*)] + [\lambda - P(\lambda)]^T g(z^*) \\
&\quad = O(\operatorname{dist}((z,\lambda),\mathcal{S})). \qquad (26)
\end{aligned}$$

We obtain the result by substituting (24), (25), and (26) into (15).

The proof of reverse estimate—$\operatorname{dist}((z,\lambda),\mathcal{S}) = O(\mu)$—follows the proof of [9, Lemma 5.5] closely. The condition $(z,\lambda) \in \mathcal{N}^{\gamma}(\epsilon)$ is important in deriving this estimate; that is, the $\lambda$ component should be close to a "sufficiently strictly complementary" point in $\mathcal{S}_\lambda$ rather than near some extreme point of this set. We omit the details and refer the reader to the earlier paper. $\square$

To analyze the step $(\Delta z, \Delta\lambda_{\mathcal{B}})$, we decompose it to conform with the singular value decomposition (svd) of the optimal Jacobian $Dg_{\mathcal{B}}(z^*)$. From the definitions (4) and (20), this matrix has dimensions $\bar{m} \times n$ and rank $\breve{m}$. We write its svd as

$$Dg_{\mathcal{B}}(z^*) = \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{U}^T \\ \hat{V}^T \end{bmatrix}, \qquad (27)$$

where

$$\begin{gathered}
S \text{ is diagonal with diagonal elements } \sigma_1 \geq \cdots \geq \sigma_{\breve{m}} > 0; \\
\begin{bmatrix} U & V \end{bmatrix} \text{ and } \begin{bmatrix} \hat{U} & \hat{V} \end{bmatrix} \text{ are orthogonal;} \\
U \text{ is } \bar{m} \times \breve{m},\ V \text{ is } \bar{m} \times (\bar{m} - \breve{m}),\ \hat{U} \text{ is } n \times \breve{m}, \text{ and } \hat{V} \text{ is } n \times (n - \breve{m}).
\end{gathered}$$

Note, in particular, that the columns of $\hat{V}$ constitute an orthonormal basis for $\ker Dg_{\mathcal{B}}(z^*)$.

**3.1. Exact Arithmetic.** We first analyze the step $(\Delta z, \Delta\lambda_{\mathcal{B}})$ obtained by solving (17), assuming exact arithmetic. We decompose this step as

$$\Delta z = \hat{U} y_{\hat{U}} + \hat{V} y_{\hat{V}}, \qquad (28a)$$

$$\Delta\lambda_{\mathcal{B}} = \lambda_{\mathcal{B}}^+ - \lambda_{\mathcal{B}} = U w_U + V w_V. \qquad (28b)$$

By substituting (28) into (17) and premultiplying the blocks of this system by the matrices $\hat{U}^T$, $\hat{V}^T$, $U^T$, and $V^T$, we obtain

$$\begin{bmatrix}
\hat{U}^T \mathcal{L}_{zz} \hat{U} & \hat{U}^T \mathcal{L}_{zz} \hat{V} & \hat{U}^T (Dg_{\mathcal{B}})^T U & \hat{U}^T (Dg_{\mathcal{B}})^T V \\
\hat{V}^T \mathcal{L}_{zz} \hat{U} & \hat{V}^T \mathcal{L}_{zz} \hat{V} & \hat{V}^T (Dg_{\mathcal{B}})^T U & \hat{V}^T (Dg_{\mathcal{B}})^T V \\
-U^T (Dg_{\mathcal{B}}) \hat{U} & -U^T (Dg_{\mathcal{B}}) \hat{V} & \mu I & 0 \\
-V^T (Dg_{\mathcal{B}}) \hat{U} & -V^T (Dg_{\mathcal{B}}) \hat{V} & 0 & \mu I
\end{bmatrix}
\begin{bmatrix} y_{\hat{U}} \\ y_{\hat{V}} \\ w_U \\ w_V \end{bmatrix}
=
\begin{bmatrix} r_{\hat{U}} \\ r_{\hat{V}} \\ r_U \\ r_V \end{bmatrix}, \qquad (29)$$

where the right-hand side is

$$\begin{bmatrix} r_{\hat{U}} \\ r_{\hat{V}} \\ r_U \\ r_V \end{bmatrix}
=
\begin{bmatrix} -\hat{U}^T (D\phi + Dg_{\mathcal{B}}^T \lambda_{\mathcal{B}}) \\ -\hat{V}^T (D\phi + Dg_{\mathcal{B}}^T \lambda_{\mathcal{B}}) \\ U^T g_{\mathcal{B}} \\ V^T g_{\mathcal{B}} \end{bmatrix}. \qquad (30)$$

Since

$$Dg_{\mathcal{B}}(z) - Dg_{\mathcal{B}}(z^*) = O(\|z - z^*\|) = O(\mu),$$

by Assumption 1 and Lemma 3.1, we have from (27) that

$$\begin{aligned}
U^T Dg_{\mathcal{B}}(z)\hat{U} &= S + O(\mu), && \text{(31a)}\\
U^T Dg_{\mathcal{B}}(z)\hat{V} &= O(\mu), && \text{(31b)}\\
V^T Dg_{\mathcal{B}}(z)\hat{U} &= O(\mu), && \text{(31c)}\\
V^T Dg_{\mathcal{B}}(z)\hat{V} &= O(\mu). && \text{(31d)}
\end{aligned}$$

Meanwhile, by the second-order condition (7) and orthonormality of $\hat{V}$, we have that the matrix $\hat{V}^T \mathcal{L}_{zz}(z^*, \lambda^*)\hat{V}$ satisfies

$$v^T \hat{V}^T \mathcal{L}_{zz}(z^*, \lambda^*)\hat{V} v \geq \sigma \|v\|^2, \tag{32}$$

for all $v \in \mathsf{R}^{n-\tilde{m}}$, all $\lambda^* \in \mathcal{S}_\lambda$, and some $\sigma > 0$ independent of $v$ and $\lambda^*$.

By substituting the estimates (31) into (29), we obtain

$$\left[\begin{array}{cccc} \hat{U}^T\mathcal{L}_{zz}\hat{U} & \hat{U}^T\mathcal{L}_{zz}\hat{V} & S + O(\mu) & O(\mu) \\ \hat{V}^T\mathcal{L}_{zz}\hat{U} & \hat{V}^T\mathcal{L}_{zz}\hat{V} & O(\mu) & O(\mu) \\ -S + O(\mu) & O(\mu) & \mu I & 0 \\ O(\mu) & O(\mu) & 0 & \mu I \end{array}\right] \left[\begin{array}{c} y_{\hat{U}} \\ y_{\hat{V}} \\ w_U \\ w_V \end{array}\right] = \left[\begin{array}{c} r_{\hat{U}} \\ r_{\hat{V}} \\ r_U \\ r_V \end{array}\right]. \tag{33}$$

By eliminating $w_V$ from this system and rearranging the resulting block $3 \times 3$ coefficient matrix, we obtain

$$w_V = \mu^{-1} r_V + O(\|y_{\hat{U}}\|) + O(\|y_{\hat{V}}\|) \tag{34}$$

and

$$\left[\begin{array}{ccc} S + O(\mu) & \hat{U}^T\mathcal{L}_{zz}\hat{V} + O(\mu) & \hat{U}^T\mathcal{L}_{zz}\hat{U} + O(\mu) \\ O(\mu) & \hat{V}^T\mathcal{L}_{zz}\hat{V} + O(\mu) & \hat{V}^T\mathcal{L}_{zz}\hat{U} + O(\mu) \\ \mu I & O(\mu) & -S + O(\mu) \end{array}\right] \left[\begin{array}{c} w_U \\ y_{\hat{V}} \\ y_{\hat{U}} \end{array}\right] = \left[\begin{array}{c} r_{\hat{U}} + O(\|r_V\|) \\ r_{\hat{V}} + O(\|r_V\|) \\ r_U \end{array}\right]. \tag{35}$$

(Note that elimination of $w_V$ has introduced $O(\mu)$ perturbations into the $\mathcal{L}_{zz}$ blocks.) The coefficient matrix in (35) is an $O(\mu)$ perturbation of the block upper triangular matrix $M(z, \lambda)$ defined by

$$M(z, \lambda) = \left[\begin{array}{ccc} S & \hat{U}^T\mathcal{L}_{zz}\hat{V} & \hat{U}^T\mathcal{L}_{zz}\hat{U} \\ 0 & \hat{V}^T\mathcal{L}_{zz}\hat{V} & \hat{V}^T\mathcal{L}_{zz}\hat{U} \\ 0 & 0 & -S \end{array}\right], \tag{36}$$

whose condition number can be bounded independently of $(z, \lambda)$ for all $(z, \lambda) \in \mathcal{N}^\gamma(\epsilon)$ for $\epsilon$ sufficiently small. To verify this claim, note first that the above-diagonal blocks are bounded in norm by $\|\mathcal{L}_{zz}\|$ which, by (21), Assumption 1, and Lemma 3.1, satisfies the bound

$$\begin{aligned}
\|\mathcal{L}_{zz}(z, \lambda)\| &\leq \|\mathcal{L}_{zz}(z^*, P(\lambda))\| + O\left(\|(z, \lambda) - (z^*, P(\lambda))\|\right) \\
&\leq \bar{\sigma} + O(\mu) \\
&\leq 2\bar{\sigma}, && \text{(37)}
\end{aligned}$$

for a sufficiently small choice of $\epsilon$. By decreasing $\epsilon$ further if necessary, we have by applying (32), Assumption 1, and Lemma 3.1 that

$$v^T \hat{V}^T \mathcal{L}_{zz}(z,\lambda)\hat{V}v = v^T \hat{V}^T \mathcal{L}_{zz}(z^*, P(\lambda))\hat{V}v + O(\mu)\|v\|^2 \geq (\sigma/2)\|v\|^2,$$

for all $v \in \mathsf{R}^{n\times(n-\check{m})}$. Hence, using (37) again, we obtain

$$\kappa(\hat{V}^T \mathcal{L}_{zz}\hat{V}) = \|\hat{V}^T \mathcal{L}_{zz}\hat{V}\| \, \|[\hat{V}^T \mathcal{L}_{zz}\hat{V}]^{-1}\| \leq 4(\bar{\sigma}/\sigma). \tag{38}$$

The other diagonal blocks in (36) are also well conditioned, since by the definition of $S$ we have

$$\kappa(S) = \|S\| \, \|S^{-1}\| = \sigma_1/\sigma_{\check{m}}. \tag{39}$$

Since

$$M(z,\lambda)^{-1} = \begin{bmatrix} S^{-1} & M_{12} & M_{13} \\ 0 & [\hat{V}^T \mathcal{L}_{zz}\hat{V}]^{-1} & M_{23} \\ 0 & 0 & -S^{-1} \end{bmatrix},$$

where

$$\begin{aligned} M_{12} &= -S^{-1}(\hat{U}^T \mathcal{L}_{zz}\hat{V})(\hat{V}^T \mathcal{L}_{zz}\hat{V})^{-1}, \\ M_{23} &= (\hat{V}^T \mathcal{L}_{zz}\hat{V})^{-1}(\hat{V}^T \mathcal{L}_{zz}\hat{U})S^{-1}, \\ M_{13} &= S^{-1}\left[-(\hat{U}^T \mathcal{L}_{zz}\hat{V})M_{23} + (\hat{U}^T \mathcal{L}_{zz}\hat{U})S^{-1}\right], \end{aligned}$$

it is easy to see from (37), (38), and (39) that, for a sufficiently small choice of $\epsilon$, the quantities $\|M(z,\lambda)^{-1}\|$ and $\kappa(M(z,\lambda))$ are bounded for all $(z,\lambda) \in \mathcal{N}^\gamma(\epsilon)$. That is, we can define a constant $C_M$ such that

$$\|M(z,\lambda)^{-1}\| \leq C_M, \quad \kappa(M(z,\lambda)) \leq C_M, \quad \text{for all } (z,\lambda) \in \mathcal{N}^\gamma(\epsilon).$$

Using (36), we rewrite (35) as

$$[M(z,\lambda) + O(\mu)] \begin{bmatrix} w_U \\ y_{\hat{V}} \\ y_{\hat{U}} \end{bmatrix} = \begin{bmatrix} r_{\hat{U}} + O(\|r_V\|) \\ r_{\hat{U}} + O(\|r_V\|) \\ r_U \end{bmatrix}. \tag{40}$$

By decreasing $\epsilon$ if necessary, we can ensure nonsingularity of the coefficient matrix and in fact that

$$\left\| [M(z,\lambda) + O(\mu)]^{-1} \right\| \leq 2\|M(z,\lambda)^{-1}\| \leq 2C_M.$$

Hence, we have immediately from (40) that

$$\|(w_U, y_{\hat{V}}, y_{\hat{U}})\| = O\left(\|(r_{\hat{U}}, r_{\hat{V}}, r_U, r_V)\|\right). \tag{41}$$

It follows by substitution in (34) that

$$\|w_V\| = O(\mu^{-1})\|r_V\| + O\left(\|(r_{\hat{U}}, r_{\hat{V}}, r_U)\|\right). \tag{42}$$

We obtain the estimate (22) for $(\Delta z, \Delta\lambda_{\mathcal{B}})$ by using the definition of the right-hand side in (30). We have

$$\|(r_{\hat{U}}, r_{\hat{V}})\| = \|D\phi(z) + Dg_{\mathcal{B}}(z)^T \lambda_{\mathcal{B}}\| \leq \|\mathcal{L}_z(z,\lambda)\| + \|Dg_{\mathcal{N}}(z)\|\|\lambda_{\mathcal{N}}\|.$$

From the definition (15), we have directly that $\|\mathcal{L}_z(z,\lambda)\| \le \mu$. By compactness of $\mathcal{N}^\gamma(\epsilon)$ and smoothness of $Dg(\cdot)$, we have that $\|Dg_{\mathcal{N}}(z)\|$ is bounded above, while for $(z,\lambda) \in \mathcal{N}^\gamma(\epsilon)$ we have from Lemma 3.1 that $\|\lambda_{\mathcal{N}}\| \le \mathrm{dist}\,((z,\lambda),\mathcal{S}) = O(\mu)$. We conclude that

$$\|(r_{\hat{U}}, r_{\hat{V}})\| = O(\mu). \tag{43}$$

Since $g_{\mathcal{B}}(z^*) = 0$, we have from Assumption 1 and Lemma 3.1 that

$$\|r_U\| = \|U^T g_{\mathcal{B}}(z)\| \le \|g_{\mathcal{B}}(z) - g_{\mathcal{B}}(z^*)\| = O(\|z - z^*\|) = O(\mu). \tag{44}$$

The remaining estimate is slightly more refined. By Assumption 1 and Lemma 3.1, we have

$$\begin{aligned} r_V &= V^T g_{\mathcal{B}}(z) \\ &= V^T \left[g_{\mathcal{B}}(z^*) + Dg_{\mathcal{B}}(z^*)(z - z^*)\right] + O(\|z - z^*\|^2) \\ &= O(\mu^2), \end{aligned} \tag{45}$$

since $g_{\mathcal{B}}(z^*) = 0$ and $V^T Dg_{\mathcal{B}}(z^*) = 0$ by (27). Substitution of (43), (44), and (45) into (41) and (42) yields

$$\|(y_{\hat{U}}, y_{\hat{V}}, w_U, w_V)\| = O(\mu).$$

The estimate (22) follows immediately from (28).

We summarize this result as a theorem.

THEOREM 3.2. *Suppose that the standing assumptions hold. Then there is a constant $\epsilon > 0$ such that for all $(z,\lambda) \in \mathcal{N}^\gamma(\epsilon)$, the solution $(\Delta z, \lambda^+)$ of (17) satisfies the estimate*

$$\|(\Delta z, \lambda^+ - \lambda)\| \le C' \mu, \tag{46}$$

*for some $C' > 0$ that depends on $\gamma$ and $\epsilon$ but not on $\mu$ or $(z,\lambda)$.*

*Proof.* We have shown above that (22) holds. It remains only to examine the $\mathcal{N}$-components of $\lambda^+$ and $\lambda$. Since $\lambda^*_{\mathcal{N}} = 0$ for all $\lambda^* \in \mathcal{S}_\lambda$, we have

$$\|\lambda_{\mathcal{N}} - \lambda^+_{\mathcal{N}}\| = \|\lambda_{\mathcal{N}}\| \le \mathrm{dist}\,((z,\lambda),\mathcal{S}) = O(\mu).$$

□

It is not difficult to show that the solution $(\Delta z, \lambda^+)$ of (17) is the only local solution of (16) satisfying the estimate (46). Let $(\widehat{\Delta z}, \hat{\lambda}^+)$ be any solution of (16) that satisfies (46). If there is an index $i$ for which $i \notin \mathcal{N}$ but $\hat{\lambda}^+_i = 0$, we have by the definition (19) and $(z,\lambda) \in \mathcal{N}^\gamma(\epsilon)$ that $\lambda_i \ge \gamma\xi - \epsilon$. Hence, by choosing $\epsilon$ smaller than $\gamma\xi/2$, we have that

$$\|\lambda - \hat{\lambda}^+\| \ge |\lambda_i - \hat{\lambda}_i| \ge \gamma\xi/2,$$

for this particular index $i$. Hence, each $i \notin \mathcal{N}$ must have $\hat{\lambda}^+_i > 0$.

If, on the other hand, there is an index $i \in \mathcal{N}$ for which $\hat{\lambda}^+_i > 0$, by complementarity and equation (12b) we have

$$g_i(z) + Dg_i(z)\widehat{\Delta z} - \mu(\hat{\lambda}^+_i - \lambda_i) = 0.$$

If the estimate $\|(\widehat{\Delta z}, \hat{\lambda}_i^+ - \lambda_i)\| = O(\mu)$ holds (as assumed), this equation yields $g_i(z) = O(\mu)$. However, Lemma 3.1 and $g_i(z^*) < 0$ imply that $g_i(z)$ is bounded away from zero for $\mu$ sufficiently small, giving a contradiction.

We conclude from these two cases that $i \in \mathcal{N}$ if and only if $\hat{\lambda}_i^+ = 0$, so that the index partition for the solution $(\widehat{\Delta z}, \hat{\lambda}^+)$ is simply $\mathcal{B} \cup \mathcal{N}$. Since the coefficient matrix in (17) is nonsingular, we must have $(\widehat{\Delta z}, \hat{\lambda}^+) = (\Delta z, \lambda^+)$.

**3.2. Finite-Precision Arithmetic.** We now examine the effect of finite-precision floating-point arithmetic on the step calculated from (17). In our discussion of floating-point arithmetic, we use $\mathbf{u}$ to denote unit roundoff, which we define by the following statement: When $x$ and $y$ are any two floating-point numbers, op denotes $+, -, \times$, or $/$, and $\mathit{fl}(z)$ denotes the floating-point approximation of any real number $z$, we have

$$\mathit{fl}(x \operatorname{op} y) = (x \operatorname{op} y)(1+\delta), \qquad |\delta| \le \mathbf{u}.$$

We also introduce the unfamiliar notation $\delta_{\mathbf{u}}$ to represent a scalar quantity that is a modest multiple of $\mathbf{u}$.

In solving (17), errors due to floating-point arithmetic arise from two sources:

a. Error incurred during the evaluation of the components of the matrix and the right-hand side, and
b. Error incurred during factorization of the matrix in (17) and the subsequent triangular substitutions.

We consider (a) first. The quantities on the right-hand side—$D\phi(z) + Dg_{\mathcal{B}}(z)^T\lambda_{\mathcal{B}}$ and $g_{\mathcal{B}}(z)$—are $O(\mu)$ in exact arithmetic. However, they are typically evaluated by adding and subtracting quantities whose size is independent of $\mu$, and hence they contain evaluation errors of size $\delta_{\mathbf{u}}$. Of all floating-point errors, these are the most significant in their effect on the accuracy of the step and on the algorithm's convergence behavior. We assume that

$$\mu \gg \mathbf{u}, \tag{47}$$

since, if not, the perturbed right-hand side may bear no relation to the exact version, so we could not expect any similarity between the perturbed solution and its exact counterpart.

Evaluation errors may also appear in the blocks of the coefficient matrix in (17), except for the $(2,2)$ block $\mu I$. Since we have assumed for convenience that $\|\mathcal{L}_{zz}(z^*, \lambda^*)\|$ and $Dg_{\mathcal{B}}(z^*)$ are not too different from 1, these errors can be accounted for by introducing perturbations of size $\delta_{\mathbf{u}}$ into the $(1,1)$, $(1,2)$, and $(2,1)$ blocks of the matrix in (17).

In assessing the errors that arise from cause (b), we assume that the matrix is factored by a stable procedure, one in which the elements of the submatrices that arise during the factorization are not too large relative to the norm of the original matrix. Suitable algorithms could include Gaussian elimination with pivoting or a Bunch-Parlett or Bunch-Kaufman algorithm applied to a symmetric indefinite reformulation of the problem. (See, for example, Fourer and Mehrotra [4] and Wright [12] for a discussion of algorithms for symmetric indefinite matrices and their stability properties.) A standard backward error analysis applied to the general square system $Mx = r$ shows that the approximate solution $\hat{x}$ computed by stable factorization and triangular substitutions satisfies

$$(M + E_M)\hat{x} = r, \qquad \text{where } \|E_M\| \le \delta_{\mathbf{u}}\|M\|$$

(see for example Golub and Van Loan [6, Chapter 3]). The effects of the errors of type (b) can be accounted for by introducing perturbations of size $\delta_{\mathbf{u}}$ into all the elements of the matrix in (17).

Collating these errors, we find that the computed approximation $(\tilde{\Delta z}, \tilde{\Delta\lambda_{\mathcal{B}}})$ to the step $(\Delta z, \Delta\lambda_{\mathcal{B}})$ satisfies the equation

$$\left\{ \begin{bmatrix} \mathcal{L}_{zz}(z,\lambda) & Dg_{\mathcal{B}}(z)^T \\ -Dg_{\mathcal{B}}(z) & \mu I \end{bmatrix} + \bar{E} \right\} \begin{bmatrix} \tilde{\Delta z} \\ \tilde{\Delta\lambda_{\mathcal{B}}} \end{bmatrix} = \begin{bmatrix} -\mathcal{L}_z(z,\lambda) \\ g_{\mathcal{B}}(z) \end{bmatrix} + \bar{e}, \tag{48}$$

where

$$\|\bar{E}\| \le \delta_{\mathbf{u}}, \qquad \|\bar{e}\| \le \delta_{\mathbf{u}}. \tag{49}$$

By decomposing the approximate step as

$$\tilde{\Delta z} = \hat{U}\tilde{y}_{\hat{U}} + \hat{V}\tilde{y}_{\hat{V}}, \qquad \tilde{\Delta\lambda_{\mathcal{B}}} = U\tilde{w}_U + V\tilde{w}_V, \tag{50}$$

and partitioning the system as in (33), we obtain

$$\begin{aligned} &\left\{ \begin{bmatrix} \hat{U}^T\mathcal{L}_{zz}\hat{U} & \hat{U}^T\mathcal{L}_{zz}\hat{V} & S+O(\mu) & O(\mu) \\ \hat{V}^T\mathcal{L}_{zz}\hat{U} & \hat{V}^T\mathcal{L}_{zz}\hat{V} & O(\mu) & O(\mu) \\ -S+O(\mu) & O(\mu) & \mu I & 0 \\ O(\mu) & O(\mu) & 0 & \mu I \end{bmatrix} + E \right\} \begin{bmatrix} \tilde{y}_{\hat{U}} \\ \tilde{y}_{\hat{V}} \\ \tilde{w}_U \\ \tilde{w}_V \end{bmatrix} \\ &\qquad = \begin{bmatrix} r_{\hat{U}} + e_{\hat{U}} \\ r_{\hat{V}} + e_{\hat{V}} \\ r_U + e_U \\ r_V + e_V \end{bmatrix}, \end{aligned} \tag{51}$$

where the norms of $E$, $e_{\hat{U}}$, $e_{\hat{V}}$, $e_U$, and $e_V$ all have size $\delta_{\mathbf{u}}$. From the last block row we obtain

$$\tilde{w}_V = [\mu I + E_{VV}]^{-1}\left[r_V + e_V + (O(\mu)+\delta_{\mathbf{u}})(\|\tilde{y}_{\hat{U}}\| + \|\tilde{y}_{\hat{V}}\|) + \delta_{\mathbf{u}}\|\tilde{w}_U\|\right], \tag{52}$$

where $E_{VV}$ is the lower right block of $E$. We use the condition (47) and the estimate $\|E\| = \delta_{\mathbf{u}}$ to deduce that

$$[\mu I + E_{VV}]^{-1} = \left[I + \mu^{-1}E_{VV}\right]^{-1}\mu^{-1} = O(\mu^{-1}). \tag{53}$$

Hence, from (52) we have

$$\|\tilde{w}_V\| = O(\mu^{-1})\|r_V + e_V\| + O(1)\left[\|\tilde{y}_{\hat{U}}\| + \|\tilde{y}_{\hat{V}}\|\right] + O(\mu^{-1}\delta_{\mathbf{u}})\|\tilde{w}_U\|. \tag{54}$$

By eliminating $\tilde{w}_V$ from (51) and using the estimate (53), we obtain

$$\begin{aligned} &\left\{ \begin{bmatrix} S & \hat{U}^T\mathcal{L}_{zz}\hat{V} & \hat{U}^T\mathcal{L}_{zz}\hat{U} \\ 0 & \hat{V}^T\mathcal{L}_{zz}\hat{V} & \hat{V}^T\mathcal{L}_{zz}\hat{U} \\ 0 & 0 & -S \end{bmatrix} + \tilde{E} \right\} \begin{bmatrix} \tilde{w}_U \\ \tilde{y}_{\hat{V}} \\ \tilde{y}_{\hat{U}} \end{bmatrix} \\ &\qquad = \begin{bmatrix} r_{\hat{U}} + e_{\hat{U}} \\ r_{\hat{V}} + e_{\hat{V}} \\ r_U + e_U \end{bmatrix} + \begin{bmatrix} O(\|r_V + e_V\|) \\ O(\|r_V + e_V\|) \\ O(\mu^{-1}\delta_{\mathbf{u}}\|r_V + e_V\|) \end{bmatrix}, \end{aligned} \tag{55}$$

where

$$\|\tilde{E}\| \le O(\mu) + \delta_{\mathbf{u}}.$$

As before, we choose the neighborhood radius $\epsilon$ small enough to ensure that the error-free part of the coefficient matrix in (55) dominates any $O(\mu)$ perturbations. We assume, too, that $\mathbf{u}$ is small enough that perturbations of size $\delta_{\mathbf{u}}$ are also dominated by the error-free part. By applying the logic that follows equation (35), we obtain as in (41) that $\|(\tilde{w}_U, \tilde{y}_{\hat{V}}, \tilde{y}_{\hat{U}})\|$ is of the order of the right-hand side in (55), that is,

$$\begin{aligned} \|(\tilde{w}_U, \tilde{y}_{\hat{V}}, \tilde{y}_{\hat{U}})\| &= O\left(\|(r_{\hat{U}} + e_{\hat{U}}, r_{\hat{V}} + e_{\hat{V}}, r_U + e_U, r_V + e_V)\|\right) \\ &= O(\mu) + \delta_{\mathbf{u}}. \end{aligned} \tag{56}$$

By using some simple manipulation involving (35) and (55), we can show further that

$$\|(\tilde{w}_U, \tilde{y}_{\hat{V}}, \tilde{y}_{\hat{U}}) - (w_U, y_{\hat{V}}, y_{\hat{U}})\| = O(\mu^2) + \delta_{\mathbf{u}}, \tag{57}$$

so that the relative accuracy of the computed step components $(\tilde{w}_U, \tilde{y}_{\hat{V}}, \tilde{y}_{\hat{U}})$ remains high. (Note that these estimates hold even when $\mu$ is similar in size to $\mathbf{u}$, provided that the $\tilde{E}$ term is small enough in the sense described above.)

Returning to the component $\tilde{w}_V$, we have from (54), (45), (49), and (56) that

$$\|\tilde{w}_V\| = O(\mu + \mu^{-1}\delta_{\mathbf{u}}). \tag{58}$$

Note that if we allow $\mu \approx \mathbf{u}$, the matrix in (55) may be singular and the norm of $\tilde{w}_V$ can be arbitrarily large.

From the estimates (56) and (58), we conclude that finite-precision arithmetic has little effect on the $(w_U, y_{\hat{V}}, y_{\hat{U}})$ step components, while it has a potentially significant effect on the $w_V$ component. In fact, for $\mu < \sqrt{\mathbf{u}}$, the estimate (58) indicates that the error in $\tilde{w}_V$ can dominate the "exact" contribution. Fortunately, as we see below, the potentially large error in this component has little effect on the local convergence properties of the algorithm.

By substituting the estimates (56) and (58) into (50), we obtain

$$\|\tilde{\Delta z}\| = O(\mu) + \delta_{\mathbf{u}}, \qquad \|\Delta\tilde{\lambda}_{\mathcal{B}}\| = O(\mu + \mu^{-1}\delta_{\mathbf{u}}). \tag{59}$$

**4. Local Convergence.** We now examine the effect of the exact and inexact steps on the decrease in the optimality measure $\mu$. We show that the exact step yields a "quadratic" decrease in $\mu$, indicating a quadratic rate of convergence of the iterates $(z, \lambda)$ to the primal-dual solution set $\mathcal{S}$. In finite precision, this convergence behavior is affected less severely than one might expect, but we show that reduction of $\mu$ below the level of $\mathbf{u}$ cannot be achieved in general and is, in any case, undesirable.

Since we are interested in the asymptotic behavior, we assume that $\mu$ (and hence $\mathrm{dist}\,((z,\lambda),\mathcal{S})$) is small enough that the new iterate is obtained from (17), even when floating-point arithmetic is used.

First, we analyze the case of exact steps. Using the definition (2) of the Lagrangian $\mathcal{L}$, Taylor's theorem, Assumption 1, and $\lambda^+_{\mathcal{N}} = 0$, we find that

$$\begin{aligned} &\mathcal{L}_z(z + \Delta z, \lambda^+) \\ &\quad = D\phi(z + \Delta z) + Dg(z + \Delta z)^T \lambda^+ \\ &\quad = D\phi(z + \Delta z) + Dg_{\mathcal{B}}(z + \Delta z)^T \lambda^+_{\mathcal{B}} \\ &\quad = D\phi(z) + D^2\phi(z)\Delta z + O(\|\Delta z\|^2) + Dg_{\mathcal{B}}(z)^T \lambda^+_{\mathcal{B}} \\ &\qquad\qquad + \sum_{i \in \mathcal{B}} \lambda^+_i D^2 g_i(z)\Delta z + O(\|\lambda_{\mathcal{B}} + \Delta\lambda_{\mathcal{B}}\|\,\|\Delta z\|^2) \\ &\quad = \mathcal{L}_{zz}(z,\lambda)\Delta z + Dg_{\mathcal{B}}(z)^T \lambda^+_{\mathcal{B}} + D\phi(z) \\ &\qquad\qquad + \sum_{i \in \mathcal{B}} \Delta\lambda_i D^2 g_i(z)\Delta z + O(\|\Delta z\|^2). \end{aligned}$$

From the first block row of (17), the first three terms in this expression sum to zero, so we have from (22) that

$$\mathcal{L}_z(z+\Delta z, \lambda^+) = O(\|\Delta\lambda_{\mathcal{B}}\|\,\|\Delta z\|) + O(\|\Delta z\|^2) = O(\mu^2). \tag{60}$$

For the $\mathcal{B}$ components of $g(\cdot)$, we have from the second block row in (17) and (22) that

$$\begin{aligned} g_{\mathcal{B}}(z+\Delta z) &= g_{\mathcal{B}}(z) + Dg_{\mathcal{B}}(z)\Delta z + O(\|\Delta z\|^2) \\ &= \mu\Delta\lambda_{\mathcal{B}} + O(\|\Delta z\|^2) \\ &= O(\mu^2). \end{aligned} \tag{61}$$

For the $\mathcal{N}$ components, since $g_{\mathcal{N}}(z)$ is negative and bounded away from zero for $(z,\lambda) \in \mathcal{N}^\gamma(\epsilon)$, we have

$$g_{\mathcal{N}}(z+\Delta z) < 0,$$

provided that $\epsilon$ is chosen to be sufficiently small. By combining the last two expressions we obtain

$$\|g(z+\Delta z)_+\| = \|g_{\mathcal{B}}(z+\Delta z)_+\| \le \|g_{\mathcal{B}}(z+\Delta z)\| = O(\mu^2). \tag{62}$$

For the third component in the definition (15) of $\mu$, we obtain from (17), (22), $\lambda_{\mathcal{N}}^+ = 0$, and boundedness of $\mathcal{S}_\lambda$ that

$$\begin{aligned} (\lambda^+)^T g(z+\Delta z) &= (\lambda_{\mathcal{B}} + \Delta\lambda_{\mathcal{B}})^T g_{\mathcal{B}}(z+\Delta z) \\ &= (\lambda_{\mathcal{B}}^+)^T \left(g_{\mathcal{B}}(z) + Dg_{\mathcal{B}}(z)\Delta z + O(\|\Delta z\|^2)\right) \\ &= (\lambda_{\mathcal{B}}^+)^T \left(\mu\Delta\lambda_{\mathcal{B}} + O(\|\Delta z\|^2)\right) \\ &= O(\mu\|\Delta\lambda_{\mathcal{B}}\|) + O(\|\Delta z\|^2) \\ &= O(\mu^2). \end{aligned} \tag{63}$$

By combining the estimates above, we have the following result concerning the decrease in $\mu$ produced by a unit step.

THEOREM 4.1. *Suppose that the standing assumptions hold. Then there is a constant $\epsilon > 0$ such that for all $(z,\lambda) \in \mathcal{N}^\gamma(\epsilon)$, the solution $(\Delta z, \lambda^+)$ of (17) yields*

$$\mu_+ \stackrel{\text{def}}{=} \mu(z+\Delta z, \lambda^+) \le C\mu^2, \tag{64}$$

*for some $C > 0$ that depends on $\gamma$ and $\epsilon$ but not on $\mu$ or $(z,\lambda)$.*

*Proof.* We obtain (64) by substituting (60), (62), and (63) into the definition (15). ☐

The local convergence result follows as a simple corollary of Theorems 3.2 and 4.1.

COROLLARY 4.2. *Suppose that the standing assumptions hold, and let $\epsilon$ be small enough that Theorems 3.2 and 4.1 both apply. Then if $(z^0,\lambda^0)$ is a point that satisfies*

$$(z^0,\lambda^0) \in \mathcal{N}^\gamma(\epsilon/2), \quad C'\mu_0 \le \epsilon/4, \quad C\mu_0 \le 1/2,$$

*the stabilized SQP method with unit steps converges Q-quadratically to a point $(z^*,\lambda^*) \in \mathcal{S}$.*

*Proof.* The first step of the stabilized SQP satisfies

$$\|(\Delta z^0, \lambda^1 - \lambda^0)\| \le C'\mu_0 \le \epsilon/4,$$

so from the definition of $\mathcal{N}^\gamma(\epsilon)$ (19), we have

$$(z^1, \lambda^1) \in \mathcal{N}^\gamma(\epsilon/2 + \epsilon/4) = \mathcal{N}^\gamma(3\epsilon/4). \tag{65}$$

Theorem 4.1 applies, and therefore $\mu_1 \le C\mu_0^2 \le \mu_0/2$ from (64). Because of (65), $(z^1, \lambda^1)$ satisfies the hypotheses of Theorem 3.2, so we have

$$\|(\Delta z^1, \lambda^2 - \lambda^1)\| \le C'\mu_1 \le C'\mu_0/2 \le \epsilon/8,$$

and therefore

$$(z^2, \lambda^2) \in \mathcal{N}^\gamma(7\epsilon/8) \quad \text{and} \quad \mu_2 \le C\mu_1^2 \le \mu_1/2.$$

By repeating this argument, we find that $(z^k, \lambda^k) \in \mathcal{N}^\gamma(\epsilon)$ for all $k$. Moreover, $\{\mu_k\}$ converges Q-quadratically to zero, and therefore, by Lemma 3.1, $\{\operatorname{dist}((z^k, \lambda^k), \mathcal{S})\}$ also converges Q-quadratically to zero. For any indices $k$ and $l$ with $l > k$, we have

$$\|(z^k, \lambda^k) - (z^l, \lambda^l)\| \le \sum_{j=k}^{l-1} \|(\Delta z^j, \lambda^{j+1} - \lambda^j)\| \le C' \sum_{j=k}^{l-1} \mu_j \le 2C'\mu_k \to 0, \tag{66}$$

as $k$ and $l$ approach $\infty$, so the sequence $\{(z^k, \lambda^k)\}$ is Cauchy. Hence this sequence has a single limit point $(z^*, \lambda^*) \in \mathcal{S}$. We prove Q-quadratic convergence to this point by using Lemma 3.1 and an argument like (66). We have

$$\begin{aligned} &\|(z^{k+1}, \lambda^{k+1}) - (z^*, \lambda^*)\| \\ &\quad \le 2C'\mu_{k+1} \le 2C'C\mu_k^2 = O(\operatorname{dist}((z^k, \lambda^k), \mathcal{S})^2) = O(\|(z^k, \lambda^k) - (z^*, \lambda^*)\|^2). \end{aligned}$$

$\square$

We turn now to the case of inexact arithmetic and continue to assume that $\mu$ and $\mathbf{u}$ satisfy $\mu \gg \mathbf{u}$ so that the estimates (59) apply. As in (48), we have that

$$\begin{aligned} \mathcal{L}_z(z + \tilde{\Delta z}, \lambda + \tilde{\Delta\lambda}) &= \mathcal{L}_z(z, \lambda) + Dg_{\mathcal{B}}(z)^T \tilde{\Delta\lambda}_{\mathcal{B}} + \mathcal{L}_{zz}(z, \lambda)\tilde{\Delta z} \\ &\quad + O(\|\tilde{\Delta\lambda}_{\mathcal{B}}\| \, \|\tilde{\Delta z}\|) + O(\|\tilde{\Delta z}\|^2). \end{aligned} \tag{67}$$

From (48), (49), and (59), and the estimate $\mu \gg \mathbf{u}$, we obtain

$$\begin{aligned} &\left\| \mathcal{L}_z(z, \lambda) + Dg_{\mathcal{B}}(z)^T \tilde{\Delta\lambda}_{\mathcal{B}} + \mathcal{L}_{zz}(z, \lambda)\tilde{\Delta z} \right\| \\ &\le \delta_{\mathbf{u}} \|\tilde{\Delta\lambda}_{\mathcal{B}}\| + \delta_{\mathbf{u}} \|\tilde{\Delta z}\| + \delta_{\mathbf{u}} \\ &= \delta_{\mathbf{u}}^2 + O(\mu)\delta_{\mathbf{u}} + O(\mu^{-1})\delta_{\mathbf{u}}^2 + \delta_{\mathbf{u}} \\ &= \delta_{\mathbf{u}}. \end{aligned}$$

Hence, by substituting into (67) and using (59) again, we obtain

$$\|\mathcal{L}_z(z + \tilde{\Delta z}, \lambda + \tilde{\Delta\lambda})\| = \delta_{\mathbf{u}} + O(\mu^{-1})\delta_{\mathbf{u}}^2 + O(\mu^2) = \delta_{\mathbf{u}} + O(\mu^2). \tag{68}$$

For $g_{\mathcal{B}}(\cdot)$, we have as in (61), by using (48), (49), and (59), that

$$\begin{aligned} \|g_{\mathcal{B}}(z + \tilde{\Delta z})\| &= \|g_{\mathcal{B}}(z) + Dg_{\mathcal{B}}(z)\tilde{\Delta z}\| + O(\|\tilde{\Delta z}\|^2) \\ &\le \mu\|\tilde{\Delta\lambda}_{\mathcal{B}}\| + \delta_{\mathbf{u}}\|\tilde{\Delta z}\| + \delta_{\mathbf{u}}\|\tilde{\Delta\lambda}_{\mathcal{B}}\| + O(\|\tilde{\Delta z}\|^2) \\ &= O(\mu^2) + \delta_{\mathbf{u}}. \end{aligned}$$

Since $\|\tilde{\Delta z}\|$ is small, we have by the usual argument that $g_{\mathcal{N}}(z+\tilde{\Delta z}) < 0$. Hence, as in (62), we obtain

$$\|g(z+\tilde{\Delta z})_+\| = O(\mu^2) + \delta_{\mathbf{u}}. \tag{69}$$

For the third component of $\mu$, we have as in (63), using (48), (49), and (59), that

$$\begin{aligned}(\lambda + \tilde{\Delta\lambda})^T g(z+\tilde{\Delta z}) \\ &= (\lambda_{\mathcal{B}} + \Delta\tilde{\lambda}_{\mathcal{B}})^T \left[\mu\Delta\tilde{\lambda}_{\mathcal{B}} + \delta_{\mathbf{u}}\|\tilde{\Delta z}\| + \delta_{\mathbf{u}}\|\Delta\tilde{\lambda}_{\mathcal{B}}\| + \delta_{\mathbf{u}}\right] \\ &= O(\mu^2) + \delta_{\mathbf{u}}.\end{aligned} \tag{70}$$

By combining (68), (69), and (70), we obtain from the definition (15) that

$$\mu(z+\tilde{\Delta z}, \lambda + \tilde{\Delta\lambda}) = O(\mu^2) + \delta_{\mathbf{u}}.$$

This expression suggests that the effect of finite-precision arithmetic on the convergence of the algorithm is not really evident until $\mu$ reaches the level of $\sqrt{\mathbf{u}}$. Below this threshold, the algorithm will continue to run and to converge rapidly until $\mu$ reaches the level of $\mathbf{u}$. (Typically, just one or two iterations suffice to reduce $\mu$ from $\sqrt{\mathbf{u}}$ to $\mathbf{u}$.) However, we cannot in general reduce $\mu$ below the level of $\mathbf{u}$. We would not want to do so in any case because, by our assumptions on evaluation error, even an exact solution $(z^*, \lambda^*)$ may yield a $\mu$ value of size $\delta_{\mathbf{u}}$. Our pleasing (and slightly surprising) conclusion is that rapid local convergence to a nearly exact solution occurs even though the stabilized SQP step contains large errors.

**5. Equality Constraints.** The algorithm can be modified easily to handle the case in which equality constraints are present explicitly in the formulation; that is, we have

$$\min \phi(z) \qquad \text{subject to } g(z) \le 0,\ h(z) = 0,$$

where $h : \mathsf{R}^n \to \mathsf{R}^p$ is also smooth. The Lagrangian $\mathcal{L}$ is redefined as

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

where $\eta \in \mathsf{R}^p$ is the vector of Lagrange multipliers for the equality constraints. The extension of the KKT conditions to this case is well known, while the second-order condition now requires positive definiteness of the Hessian $\mathcal{L}_{zz}(z^*, \lambda^*, \eta^*)$ on the subspace

$$\ker \begin{bmatrix} Dg_{\mathcal{B}}(z^*) \\ Dh(z^*) \end{bmatrix}.$$

The appropriate extension of the Mangasarian-Fromovitz condition (8) is that $Dh(^*)$ has full row rank and that

$$Dh(z^*)\bar{y} = 0, \quad Dg_{\mathcal{B}}(z^*)\bar{y} < 0, \quad \text{for some } \bar{y} \in \mathsf{R}^n.$$

To extend the algorithm, we first redefine $\mu$ as

$$\mu = \mu(z, \lambda) \stackrel{\text{def}}{=} \left\|(\mathcal{L}_z(z,\lambda), g(z)_+, h(z), \lambda^T g(z))\right\|. \tag{71}$$

The min-max subproblem (14) becomes

$$\min_{\Delta z} \max_{\lambda^+ \geq 0,\, \eta^+} \Delta z^T D\phi(z) + \tfrac{1}{2}\Delta z^T \mathcal{L}_{zz}(z,\lambda)\Delta z$$
$$+(\eta^+)^T\left(h(z) + Dh(z)\Delta z\right) + (\lambda^+)^T\left(g(z) + Dg(z)\Delta z\right) - \tfrac{1}{2}\mu\|\lambda^+ - \lambda\|^2. \tag{72}$$

(No stabilization with respect to the Lagrange multipliers of the equality constraints is needed.) Analogously to (17), we can show that for $(z,\lambda,\eta)$ sufficiently close to a strictly complementary primal-dual solution $(z^*,\lambda^*,\eta^*)$, the solution of (72) satisfies the following system:

$$\begin{bmatrix} \mathcal{L}_{zz}(z,\lambda,\eta) & Dg_{\mathcal{B}}(z)^T & Dh(z)^T \\ -Dg_{\mathcal{B}}(z) & \mu I & 0 \\ -Dh(z) & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta z \\ \lambda_{\mathcal{B}}^+ - \lambda_{\mathcal{B}} \\ \eta^+ - \eta \end{bmatrix}$$
$$= \begin{bmatrix} -D\phi(z) - Dg_{\mathcal{B}}(z)^T\lambda_{\mathcal{B}} - Dh(z)^T\eta \\ g_{\mathcal{B}}(z) \\ h(z) \end{bmatrix}, \qquad \lambda_{\mathcal{N}}^+ = 0. \tag{73}$$

The proof of Lemma 3.1 can be extended to show that $\mu$ in (71) remains a good measure of the distance to the solution set, for points that are sufficiently strictly complementary. To prove that the solution of (73) satisfies the estimate

$$(\Delta z, \lambda^+ - \lambda, \eta^+ - \eta) = O(\mu), \tag{74}$$

we extend the analysis of Section 3.1 by redefining the svd in (27) as

$$\begin{bmatrix} Dg_{\mathcal{B}}(z^*) \\ Dh(z^*) \end{bmatrix} = \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{U}^T \\ \hat{V}^T \end{bmatrix}.$$

Because we assume full row rank of $Dh(z^*)$, we can show that $V$ has the form

$$V = \begin{bmatrix} V_1 \\ 0 \end{bmatrix}, \qquad V_1 \in \mathrm{R}^{\bar{m}\times(\bar{m}+p-\check{m})},$$

that is, the last $p$ rows of $V$ are zero. In rewriting the system (73) analogously to (29), then, we obtain a block $\mu I$ in the diagonal position corresponding to the $w_V$ component—an important feature in proving the estimate $\|w_V\| = O(\mu)$. Estimates for the other components, and hence for the full step (74), proceed as in Section 3.1. It is also easy to show, as in Section 4, that a unit step produces a quadratic reduction in $\mu$. As before, finite-precision computations have little effect on the convergence behavior.

**6. Discussion.** Our superlinear convergence result rests on two assumptions that deserve further comment. The first is the MFCQ, which is used to obtain boundedness of the optimal multiplier set (Lemma 2.1) and therefore to show that $\mu$ is a valid estimate of $\operatorname{dist}((z,\lambda),\mathcal{S})$. We can replace MFCQ with an assumption that the algorithm eventually generates an iterate $(z^k,\lambda^k)$ that lies inside a neighborhood of some bounded subset of $\mathcal{S}$, where the radius $\epsilon$ of the neighborhood depends on the bound on this subset. A step estimate like Theorem 3.2 and a local capture and convergence result like Corollary 4.2 could then be proved. Some kind of constraint qualification is still required, however, because without one the KKT conditions are not even necessary, as demonstrated by the following well-known example:

$$\min z_1 \qquad \text{subject to } z_2 \leq z_1^3,\; z_1 \geq 0.$$

The second assumption concerns the requirement that the starting guess for $\lambda$ is sufficiently strictly complementary, that is, not too close to the boundary of the set $\mathcal{S}_\lambda$. Without this assumption, it is not possible to prove, at least using techniques like those in Lemma 3.1, that $\mu$ is a measure of the distance to optimality $\operatorname{dist}((z, \lambda), \mathcal{S})$. Nor is it possible to apply the linear-algebra-based analysis of Section 3 without significant modification, since it is no longer true that all indices in $\mathcal{B}$ are active at the solution of the stabilized SQP subproblem (14).

We can accommodate this restriction with the help of an active set identification technique such as that of Facchinei, Fischer, and Kanzow [3]. Our algorithm can periodically make an estimate of the active constraint set and then solve a subproblem to modify the current value of $\lambda$ to make it "more strictly complementary" without increasing $\mu$ by too much. If $\operatorname{dist}((z, \lambda), \mathcal{S})$ is small enough that $\mathcal{B}$ is identified correctly, a single $\lambda$ adjustment should suffice. Subsequent iterates are capured within a neighborhood such as $\mathcal{N}^\gamma(\epsilon)$ and the superlinear convergence follows.

Still, it would be more satisfactory to know that the algorithm exhibited the desired behavior without this identification/adjustment step. Hager [7] describes the behavior of a variant of stabilized SQP in which the coefficient $\mu$ in the stabilization term is replaced by a value that is bounded below by a sufficiently large multiple of $\operatorname{dist}((z, \lambda), \mathcal{S})$. His analysis, which rests partly on perturbation results developed by Dontchev and Hager [2], does not require the starting $\lambda$ to be sufficiently far from the boundary of $\mathcal{S}_\lambda$. Using a suitably extended second-order sufficient condition (which requires essentially that the subspace on which $\mathcal{L}_{zz}(z^*, \lambda^*)$ is required to be positive definite is larger when $\lambda^*$ is at a boundary of $\mathcal{S}_\lambda$), Hager proves that superlinear convergence of stabilized SQP is still attainable.

**Acknowledgments.** I am grateful to Bill Hager for helpful discussions and pointers during the preparation of this work.

REFERENCES

[1] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Mass., 1995.

[2] A. L. Dontchev and W. W. Hager, *Lipschitzian stability for state constrained nonlinear optimal control*, technical report, Department of Mathematics, University of Florida, Gainesvile, Fl., 1996. To appear in SIAM Journal on Control and Optimization.

[3] F. Facchinei, A. Fischer, and C. Kanzow, *On the accurate identification of active constraints*, tech. rep., Universita' di Roma "La Sapienza", 1996.

[4] R. Fourer and S. Mehrotra, *Solving symmetric indefinite systems in an interior-point method for linear programming*, Mathematical Programming, 62 (1993), pp. 15–39.

[5] J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Mathematical Programming, 12 (1977), pp. 136–138.

[6] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 2nd ed., 1989.

[7] W. W. Hager, *Convergence of Wright's stabilized SQP algorithm*, tech. rep., Department of Mathematics, University of Florida, Gainesville, Fl., January 1997.

[8] 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.

[9] D. Ralph and S. J. Wright, *Superlinear convergence of an interior-point method for monotone variational inequalities*, Preprint MCS-P556–0196, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., January 1996. To appear in Proceedings of International Conference on Complementarity Problems: Engineering and Economic Applications and Computational Methods, SIAM, 1997.

[10] ———, *Superlinear local convergence of an interior-point method to a degenerate solution*, Preprint ANL.MCS-P622–1196, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., November 1996.

[11] S. M. Robinson, *Local epi-continuity and local optimization*, Mathematical Programming, 19 (1987), pp. 208–222.

[12] S. J. Wright, *Stability of augmented system factorizations in interior-point methods*, Preprint MCS–P446–0694, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., June 1994. Revised, July 1995. To appear in SIAM Journal of Matrix Analysis and Applications.