OPTIMIZATION TECHNICAL REPORT 02-05, AUGUST 2002, COMPUTER SCIENCES DEPT, UNIV. OF WISCONSIN
TEXAS-WISCONSIN MODELING AND CONTROL CONSORTIUM REPORT TWMCC-2002-01

A FEASIBLE TRUST-REGION SEQUENTIAL QUADRATIC PROGRAMMING ALGORITHM

STEPHEN J. WRIGHT* AND MATTHEW J. TENNY†

Abstract. We describe an algorithm for smooth nonlinear constrained optimization problems in which a sequence of feasible iterates are generated by solving a trust-region sequential quadratic programming (SQP) subproblem at each iteration, and perturbing the resulting step to retain feasibility of each iterate. By retaining feasibility, we avoid several complications of other trust-region-SQP approaches: The objective function can be used as a merit function, and the SQP subproblems are feasible for all choices of the trust-region radius. We analyze the global convergence properties under different assumptions on the approximate Hessian. By making additional assumptions (including assumptions on the feasibility perturbation technique), we can prove quadratic convergence to points satisfying second-order sufficient conditions.

1. Introduction. We consider the general smooth constrained optimization problem:

$$\min f(z) \text{ subject to } c(z) = 0, \ \ d(z) \le 0, \tag{1.1}$$

where $z \in \mathsf{R}^n$, $f : \mathsf{R}^n \to \mathsf{R}$, $c : \mathsf{R}^n \to \mathsf{R}^m$, and $d : \mathsf{R}^n \to \mathsf{R}^r$ are smooth (twice continuously differentiable) functions. We denote the set of feasible points for (1.1) by $\mathcal{F}$.

At a feasible point z, let H be an $n \times n$ symmetric matrix. The basic SQP approach obtains a step Δz by solving the following subproblem

$$\min_{\Delta z} \ m(\Delta z) \stackrel{\text{def}}{=} \nabla f(z)^T \Delta z + \tfrac{1}{2} \Delta z^T H \Delta z \text{ subject to} \tag{1.2a}$$

$$c(z) + \nabla c(z)^T \Delta z = 0, \ \ d(z) + \nabla d(z)^T \Delta z \le 0. \tag{1.2b}$$

The matrix H is chosen as some approximation to the Hessian of the Lagrangian, possibly obtained by a quasi-Newton technique, or possibly a "partial Hessian" computed in some application-dependent way from some of the objective and constraint functions and Lagrange multiplier estimates. The function $m(\cdot)$ is the quadratic model for the change in f around the current point z.

Although the basic approach (1.2) often works well in the vicinity of a solution to (1.1), trust-region or line-search devices must be added to improve its robustness and global convergence behavior. In this paper, we consider a trust region of the form

$$\|D \Delta z\|_p \le \Delta, \tag{1.3}$$

where D is uniformly bounded above and $p \in [1, \infty]$. The choice $p = \infty$ makes (1.2), (1.3) a quadratic program, since we can then restate the trust-region constraint as $-\Delta e \le Dz \le \Delta e$, where $e = (1, 1, \ldots, 1)^T$. The choice $p = 2$ produces the quadratic constraint $\Delta z^T D^T D \Delta z \le \Delta^2$, and since z is feasible for (1.1), we can show that the solution Δz of (1.2), (1.3) is identical to the solution of (1.2) alone, with H replaced by $H + \gamma D^T D$ for some $\gamma \ge 0$. For generality, we develop the convergence theory for our method to apply to any choice of p, making frequent use of the equivalence between $\| \cdot \|_p$ and $\| \cdot \|_2$.

*Computer Sciences Department, 1210 W. Dayton Street, University of Wisconsin, Madison, WI 53706, U.S.A.
†Chemical Engineering Department, University of Wisconsin, Madison, WI 53706, U.S.A.

By allowing D to have zero eigenvalues, the constraint (1.3) generally allows Δz to be unrestricted by the trust region in certain directions. We assume, however, that the combination of (1.3) and (1.2b) ensures that the all components of the step are controlled by the trust region; see Assumption 1 below.

When the iterate z is not feasible for the original problem (1.1), we cannot in general simply add the restriction (1.3) to the constraints in the subproblem (1.2), since the resulting subproblem will be infeasible for small Δ. Practical trust-region methods such as those due to Celis-Dennis-Tapia [3] and Byrd-Omojokun [11] do not insist on satisfaction of the constraints (1.2b) by the step Δz, but rather achieve some reduction in the infeasibility, while staying within the trust region (1.3) and reducing the objective in the subproblem (1.2a).

Another issue that arises in the practical SQP methods is the use of a merit or penalty function to measure the worth of each point z. Typically this function is some combination of the objective $f(z)$ and the violations of the constraints, that is, $|c_i(z)|$, $i = 1, 2, \ldots, m$ and $d_i^+(z)$, $i = 1, 2, \ldots, r$. The merit function may also depend on estimates of the Lagrange multipliers for the constraints in (1.1). It is sometimes difficult to choose weighting parameters in these merit functions appropriately, in a way that drives the iterates to a solution (or at least a first-order stationary point) of (1.1).

In this paper, we propose an algorithm called Algorithm FP-SQP (for feasibility perturbed SQP), in which all iterates z^k are feasible; that is, $z^k \in \mathcal{F}$ for all k. We obtain a step by solving a problem of the form (1.2) at a feasible point $z \in \mathcal{F}$ with a trust-region constraint (1.3). We then find a perturbation $\widetilde{\Delta z}$ of the step Δz that satisfies two crucial properties: First, *feasibility*:

$$z + \widetilde{\Delta z} \in \mathcal{F}. \tag{1.4}$$

Second, *asymptotic exactness*: There is a continuous monotonically increasing function $\phi : \mathsf{R}^+ \to \mathsf{R}^+$ with $\phi(0) = 0$ such that

$$\|\Delta z - \widetilde{\Delta z}\|_2 \leq \phi(\|\Delta z\|_2)\, \|\Delta z\|_2. \tag{1.5}$$

These conditions on $\widetilde{\Delta z}$ suffice to prove good global convergence properties for the algorithm. Additional assumptions on the feasibility perturbation technique can be made to obtain fast local convergence; see Section 4.

The effectiveness of our method depends on its being able to calculate efficiently a perturbed step $\widetilde{\Delta z}$ with the properties (1.4) and (1.5). This task is not difficult for certain structured problems such as those arising in optimal control. Additionally, in the special case in which the constraints c and d are linear, we can simply set $\widetilde{\Delta z} = \Delta z$. When some constraints are nonlinear, $\widetilde{\Delta z}$ can be obtained from the projection of $z + \Delta z$ onto the feasible set $\mathcal{F}$. For general problems, this projection is nontrivial to compute, but for problems with structured constraints, it may be inexpensive.

By maintaining feasible iterates, our method gains several advantages. First, the trust region restriction (1.3) can be added to the SQP problem (1.2) without concern as to whether it will yield an infeasible subproblem. There is no need for a composite-step approach such as those mentioned above [3, 11]. Second, the objective function f can itself be used as a merit function in deciding whether to take a step. Third, if the algorithm is terminated early, we will be able to use the latest iterate z^k as a *feasible* suboptimal point, which in many applications is far preferable to an infeasible suboptimum.

The advantages of the previous paragraph are of course shared by other feasible SQP methods. For problems with only inequality constraints, the FSQP approach described in Lawrence and Tits [9] (based on an earlier version of Panier and Tits [12] and also using ideas from Birge, Qi, and Wei [2]) calculates the main search direction via a modified SQP subproblem that includes a parameter for "tilting" the search direction toward the interior of the feasible set. A second subproblem is solved to obtain a second-order correction, and an "arc search" is performed along these two directions to find a new iterate that satisfies feasibility as well as a sufficient decrease condition in the objective f. Our algorithm differs in that it uses a trust region rather than arc searches to attain global convergence, and also in that it is less specific than in [9] about the details of calculating the step. In this sense, Algorithm FP-SQP represents an algorithmic *framework* rather than a specific algorithm.

Heinkenschloss [7] considers projected SQP methods for problems with equality constraints in addition to bounds on a subset of the variables. He specifically targets optimal control problems with bounds on the controls—a similar set of problems to those we discuss in a companion manuscript [14]. The linearized equality constraints are used to express the free variables in terms of the bounded variables, and a projected Newton direction (see [1]) is constructed for the bounded variables. The step is computed by performing a line search along this direction with projection onto the bound constraints. Besides using a line search rather than a trust region, this method contrasts with ours in that feasibility is not enforced with respect to the equality constraints, so that an augmented Lagrangian merit function must be used to determine an acceptable step length.

Other related work includes the feasible algorithm for problems with convex constraints discussed in Conn, Gould, and Toint [4]. At each iteration, this algorithm seeks an approximate minimizer of the model function over the intersection of the trust region with the original feasible set. The algorithm is targeted to problems in which the constraint set is simple (especially bound-constrained problems with ∞-norm trust regions, for which the intersection is defined by componentwise bounds). Aside from not requiring convexity, our method could be viewed as a particular instance of Algorithm 12.2.1 of [4, p. 452], in which the model function approximates the Lagrangian and the trial step is a perturbed SQP step. It may then be possible to apply the analysis of [4], once we show that the step generated in this fashion satisfies the assumptions in [4], at least for sufficiently small values of the trust-region radius. It appears nontrivial, however, to put our algorithm firmly into the framework of [4], and to extend the latter algorithm to handle a class of problems (featuring nonconvexity) which its designers did not have in mind. Therefore, we present an analysis that was developed independently of that of [4]. We note that several features of the analysis in [4, Chapter 12] are similar to ours; for instance, χ on [4, p. 452] is similar to $C(z, 1)$ defined below in (3.1), except that minimization in χ is taken over the original feasible set rather than over its linearized approximation, as in (3.1). Others aspects of the analysis in [4] and this paper are different; for instance, the generalized Cauchy point in [4, Section 12.2.1] is defined in a much more complex fashion with respect to the projected-gradient path, rather than along the straight line as in Lemma 3.3 below.

The remainder of the paper is structured as follows. The algorithm is specified in Section 2, and in Section 2.1 we show that it is possible to find a feasible perturbation of the SQP step that satisfies our requirements. We present global convergence results in Section 3. After some basic lemmas in Section 3.1, we describe conditions under which the algorithm has at least one stationary accumulation point in Section 3.2. In particular, we assume in this section that the approximate Hessian H_k in (1.2) satisfies the bound $\|H_k\|_2 \le \sigma_0 + \sigma_1 k$ for some constant σ_0 and σ_1—a type of bound often satisfied by quasi-Newton update formulae. In Section 3.3, we make the stronger assumption that $\|H_k\|$ is uniformly bounded, and prove the stronger result that *all* limit points of the algorithm are stationary. Under stronger assumptions on the limit point z^* and the feasibility projection technique, we prove fast local convergence in Section 4. Some final comments appear in Section 5.

A companion report of Tenny, Wright, and Rawlings [14] describes application of the algorithm to nonlinear optimization problems arising in model predictive control.

1.1. Optimality Results and Notation. The Lagrangian function for (1.1) is

$$\mathcal{L}(z, \mu, \lambda) \stackrel{\text{def}}{=} f(z) + \mu^T c(z) + \lambda^T d(z), \tag{1.6}$$

where $\mu \in \mathsf{R}^m$ and $\lambda \in \mathsf{R}^r$ are Lagrange multipliers for the constraints. The Karush-Kuhn-Tucker conditions for (1.1) are as follows:

$$\begin{aligned} \nabla_z \mathcal{L}(z, \mu, \lambda) = \nabla f(z) + \nabla c(z)\mu + \nabla d(z)\lambda &= 0, && (1.7\text{a}) \\ c(z) &= 0, && (1.7\text{b}) \\ d(z) \le 0 \perp \lambda &\ge 0, && (1.7\text{c}) \end{aligned}$$

where $\perp$ indicates that $\lambda^T d(z) = 0$. For a feasible point z, we denote the *active set* $\mathcal{A}(z)$ as follows:

$$\mathcal{A}(z) \stackrel{\text{def}}{=} \{i = 1, 2, \ldots, r \,|\, d_i(z) = 0\}. \tag{1.8}$$

To ensure that the tangent cone to the constraint set at a feasible point z adequately captures the geometry of the feasible set near z, we need a constraint qualification. In the global convergence analysis of Section 3, we use the Mangasarian-Fromovitz constraint qualification (MFCQ), which requires that

$$\begin{aligned} &\nabla c(z) \text{ has full column rank; and} && (1.9\text{a}) \\ &\text{there exists a vector } v \in \mathsf{R}^n \text{ such that} \\ &\nabla c(z)^T v = 0 \text{ and } v^T \nabla d_i(z) < 0 \text{ for all } i \in \mathcal{A}(z). && (1.9\text{b}) \end{aligned}$$

A more stringent constraint qualification, used in the local convergence analysis of Section 4, is the linear independence constraint qualification (LICQ), which requires that

$$\{\nabla c_i(z),\ i = 1, 2, \ldots, m\} \cup \{\nabla d_i(z),\ i \in \mathcal{A}(z)\} \text{ is linearly independent.} \tag{1.10}$$

If z is a stationary point for (1.1), and when a constraint qualification such as (1.9) or (1.10) is satisfied at z, then there exist vectors μ and λ such that (1.7) is satisfied by the triplet (z, μ, λ).

We say that the *strict complementarity* condition is satisfied at a stationary point z if there are Lagrange multiplier vectors μ and λ such that the triple (z, μ, λ) satisfies the KKT conditions and

$$\lambda - d(z) > 0. \tag{1.11}$$

That is, $\lambda_i > 0$ for all $i \in \mathcal{A}(z)$.

We use $\mathcal{B}(z,t)$ to denote the open ball (in the Euclidean norm) of radius t about z. When the subscript on the norm $\|\cdot\|$ is omitted, we assume that the Euclidean norm is signified.

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

2. The Algorithm. We now specify the algorithm. We assume only that the perturbed step $\widetilde{\Delta z}$ satisfies (1.4) and (1.5), without specifying how it is calculated. As with all trust-region algorithms, a critical role is played by the ratio of actual to predicted decrease, which is defined for a given SQP step Δz^k and its perturbed counterpart $\widetilde{\Delta z}^k$ as follows:

$$\rho_k = \frac{f(z^k) - f(z^k + \widetilde{\Delta z}^k)}{-m_k(\Delta z^k)}. \tag{2.1}$$

The algorithm is specified as follows.

Algorithm 2.1 (FP-SQP).

Given starting point z_0*, trust-region upper bound* $\bar{\Delta} \ge 1$*, initial radius* $\Delta_0 \in (0, \bar{\Delta}]$*,* $\eta \in [0, 1/4)$*, and* $p \in [1, \infty]$*;*

for $k = 0, 1, 2, \cdots$
  Obtain Δz^k *by solving (1.2), (1.3);*
  Seek $\widetilde{\Delta z}^k$ *with the properties (1.4) and (1.5);*
  if *no such* $\widetilde{\Delta z}^k$ *is found;*
    $\Delta_{k+1} \leftarrow (1/2)\|D_k \Delta z^k\|_p$*;*
    $z^{k+1} \leftarrow z^k$*;* $D_{k+1} \leftarrow D_k$*;*
  else
    Calculate ρ_k *using (2.1);*
    if $\rho_k < 1/4$
      $\Delta_{k+1} \leftarrow (1/2)\|D_k \Delta z^k\|_p$*;*
    else **if** $\rho_k > 3/4$ *and* $\|D_k \Delta z^k\|_p = \Delta_k$
        $\Delta_{k+1} \leftarrow \min(2\Delta_k, \bar{\Delta})$*;*
      else
        $\Delta_{k+1} \leftarrow \Delta_k$*;*
    if $\rho_k > \eta$
      $z^{k+1} \leftarrow z^k + \widetilde{\Delta z}^k$*;*
      choose new scaling matrix D_{k+1}*;*
    else
      $z^{k+1} \leftarrow z^k$*;* $D_{k+1} \leftarrow D_k$*;*

end (for).

We now state some assumptions that are used in the subsequent analysis. First, we define the level set L_0 as follows:

$$L_0 \stackrel{\text{def}}{=} \{z \,|\, c(z) = 0,\ d(z) \le 0,\ f(z) \le f(z_0)\} \subset \mathcal{F}.$$

Our assumption on the trust-region bound (1.3) is as follows:

Assumption 1. *There is a constant δ such that for all points $z \in L_0$ and all scaling matrices D used by the algorithm, we have for any Δz satisfying the constraints*

$$c(z) + \nabla c(z)^T \Delta z = 0, \quad d(z) + \nabla d(z)^T \Delta z \le 0$$

that

$$\delta^{-1} \|\Delta z\|_2 \le \|D \Delta z\|_p \le \delta \|\Delta z\|_2. \tag{2.2}$$

In this assumption, the constant that relates $\|\cdot\|_2$ with the equivalent norms $\|\cdot\|_p$ for all p between 1 and ∞ is absorbed into δ.

Note that for unconstrained problems (in which c and d are vacuous), this assumption is satisfied when all scaling matrices D used by the algorithm are uniformly positive definite. Another special case of relevance to optimal control problems occurs when the constraints have the form

$$c(u, v) = 0, \quad c : \mathsf{R}^{n-m} \times \mathsf{R}^m \to \mathsf{R}^m, \tag{2.3}$$

(that is $u \in \mathsf{R}^{n-m}$ and $v \in \mathsf{R}^m$), and the trust-region constraint is imposed only on the u variables; that is,

$$\|D_u \Delta u\|_p \le \Delta, \tag{2.4}$$

where D_u is a diagonal matrix with positive diagonal elements. The linearized constraints (1.2b) then have the form

$$\nabla_u c(u, v)^T \Delta u + \nabla_v c(u, v)^T \Delta v = 0, \tag{2.5}$$

which if $\nabla_v c(u, v)$ is invertible leads to

$$\Delta v = -\left(\nabla_v c(u, v)\right)^{-T} \nabla_u c(u, v)^T \Delta u.$$

If we assume that $\nabla_v c(u, v)$ is invertible for all points (u, v) in the region of interest, with $\|(\nabla_v c(u, v))^{-1}\|$ bounded, we can define a constant $\hat{\delta} > 0$ such that $\|\Delta v\|_p \le \hat{\delta} \|\Delta u\|_p$. We then have

$$\|(\Delta u, \Delta v)\|_p \le (1 + \hat{\delta}) \|\Delta u\|_p \le (1 + \hat{\delta}) D_{\min}^{-1} \|D_u \Delta u\|_p = (1 + \hat{\delta}) D_{\min}^{-1} \|D(\Delta u, \Delta v)\|_p,$$

where we define $D_{\min}$ to be a lower bound on the diagonals of D_u, and $D = \mathrm{diag}(D_u, 0)$. On the other hand, we have

$$\|D(\Delta u, \Delta v)\|_p = \|D_u \Delta u\|_\infty \le D_{\max} \|\Delta u\|_p \le D_{\max} \|(\Delta u, \Delta v)\|_p,$$

where $D_{\max}$ is an upper bound on the diagonals of D_u. It follows from the last two expressions that Assumption 1 is satisfied in this situation.

Second, for some results we make an assumption on the boundedness of the level set for problem (1.1) that is defined by the initial point z_0.

Assumption 2. *The level set L_0 is bounded, and the functions f, c, and d in (1.1) are twice continuously differentiable in an open neighborhood $\mathcal{N}(L_0)$ of this set.*

Note that L_0 is certainly closed, so that if Assumption 2 holds, it is also compact.

2.1. Algorithm FP-SQP is Well Defined. We show first that the algorithm is well defined, in the sense that given a feasible point z_k, a step $\widetilde{\Delta z}_k$ satisfying (1.4) and (1.5) can be found for all sufficiently small Δ_k, under certain assumptions.

We note first that whenever $z = z_k$ is feasible, the subproblem (1.2), (1.3) has a solution. This fact follows from nonemptiness, closedness, and boundedness of the feasible set for the subproblem. To show that there exists $\widetilde{\Delta z}_k$ satisfying (1.4) and (1.5), we make use of the following assumption.

ASSUMPTION 3. *For every point $\hat{z} \in L_0$, there are positive quantities ζ and $\hat{\Delta}_3$ such that for all $z \in \mathrm{cl}(\mathcal{B}(\hat{z}, \hat{\Delta}_3))$ we have*

$$\min_{v \in \mathcal{F}} \|v - z\| \leq \zeta \left(\|c(z)\| + \|[d(z)]_+\| \right), \tag{2.6}$$

where $[d(z)]_+ = [\max(d_i(z), 0)]_{i=1}^r$.
(Recall our convention that $\|\cdot\|$ denotes $\|\cdot\|_2$.)

This assumption requires the constraint system to be regular enough near each feasible point that a bound like that of Hoffmann [8] for systems of linear equalities and inequalities is satisfied. Assumption 3 is essentially the same as Assumption C of Lucidi, Sciandrone, and Tseng [10]. A result of Robinson [13, Corollary 1] shows that Assumption 3 is satisfied whenever MFCQ is satisfied at all points in L_0. The following result shows that a bound similar to (2.6) also holds *locally*, in the vicinity of a feasible point satisfying MFCQ.

LEMMA 2.1. *Let $\hat{z}$ be a feasible point for (1.1) at which MFCQ is satisfied. Then there exist positive quantities ζ and $\hat{R}_1$ such that for all $z \in \mathrm{cl}(\mathcal{B}(\hat{z}, \hat{R}_1))$, the bound (2.6) is satisfied.*

Proof. We first choose $\bar{R}_1$ small enough that for all $\tilde{z} \in \mathrm{cl}(\mathcal{B}(\hat{z}, \bar{R}_1)) \cap \mathcal{F}$, we have that $\mathcal{A}(\tilde{z}) \subset \mathcal{A}(\hat{z})$, where $\mathcal{A}(\cdot)$ is defined by (1.8). Let v be a vector satisfying (1.9) at $z = \hat{z}$, and assume without loss of generality that $\|v\|_2 = 1$. Because $\nabla c(\hat{z})$ has full column rank, we have by decreasing $\bar{R}_1$ if necessary that for any $\tilde{z} \in \mathrm{cl}(\mathcal{B}(\hat{z}, \hat{R}_1))$, $\nabla c(\tilde{z})$ also has full column rank. Moreover, using the full rank of $\nabla c(\tilde{z})$, we can find a perturbation $\tilde{v}$ of v satisfying $\|\tilde{v} - v\| = O(\|\tilde{z} - \hat{z}\|)$ and (after possibly decreasing $\bar{R}_1$ again) $\|\tilde{v}\| \geq 0.5$, such that

$$\nabla c(\tilde{z})^T \tilde{v} = 0 \text{ and } \tilde{v}^T \nabla d_i(\tilde{z}) < 0 \text{ for all } i \in \mathcal{A}(\hat{z}) \supset \mathcal{A}(\tilde{z}), \text{ all } \tilde{z} \in \mathrm{cl}(\mathcal{B}(\hat{z}, \hat{R}_1)) \cap \mathcal{F}.$$

Hence, the MFCQ condition is satisfied for all $\tilde{z} \in \mathrm{cl}(\mathcal{B}(\hat{z}, \bar{R}_1)) \cap \mathcal{F}$.

We now appeal to Corollary 1 of Robinson [13]. From this result, we have that there is $\zeta > 0$ (depending on $\hat{z}$ but not on $\tilde{z}$) and an open neighborhood $M(\tilde{z})$ of each $\tilde{z} \in \mathrm{cl}(\mathcal{B}(\hat{z}, \bar{R}_1)) \cap \mathcal{F}$ such that (2.6) holds for all $z \in M(\tilde{z})$. Since

$$\hat{M}(\hat{z}) \stackrel{\mathrm{def}}{=} \cup_{\tilde{z}} \{ M(\tilde{z}) \,|\, \tilde{z} \in \mathrm{cl}(\mathcal{B}(\hat{z}, \bar{R}_1)) \cap \mathcal{F} \}$$

is an open neighborhood of the compact set $\mathrm{cl}(\mathcal{B}(\hat{z}, \bar{R}_1)) \cap \mathcal{F}$, we can define $\hat{R}_1 \leq \bar{R}_1$ small enough that $\mathrm{cl}(\mathcal{B}(\hat{z}, \hat{R}_1)) \subset \hat{M}(\hat{z})$. Thus, since (2.6) holds for all $z \in M(\tilde{z})$, our proof is complete. ∎

We observe that when Assumption 1 is satisfied, the solution Δz of (1.2), (1.3) is well defined at each $z \in \mathcal{F}$. The feasible set for this problem is compact and lies in the closed ball $\mathrm{cl}\,(\mathcal{B}(0, \delta\Delta))$. Using the other assumptions, we can show that the algorithm is well defined, in the sense that a perturbed vector $\widetilde{\Delta z}$ satisfying the properties (1.4) and (1.5) can also be found.

THEOREM 2.2. *Suppose that Assumptions 1, 2, and 3 are satisfied. Then there is a positive constant* $\Delta_{\rm def}$ *such that for any* $z \in L_0$*, there is a step* $\widetilde{\Delta z}$ *that satisfies the properties (1.4) and (1.5), where* Δz *is the solution of (1.2), (1.3) with* $\Delta \le \Delta_{\rm def}$*.*

Proof. We show that the result holds for the function $\phi(t) = \sqrt{t}$ in (1.5).

We first choose $\hat{\Delta}_0$ small enough that $\mathcal{B}(z, \delta\hat{\Delta}_0) \subset \mathcal{N}(L_0)$ for *all* $z \in \mathcal{L}_0$, where $\mathcal{N}(L_0)$ is defined as in Assumption 2. Thus, for $\Delta \le \hat{\Delta}_0$ and Δz solving (1.2), (1.3), we have for all $\alpha \in [0,1]$ that

$$\|\alpha\Delta z\|_2 \le \|\Delta z\|_2 \le \delta\|D\Delta z\|_p \le \delta\hat{\Delta}_0, \tag{2.7}$$

so that $z + \alpha\Delta z \in \mathcal{N}(L_0)$.

Given any $\hat{z} \in L_0$, we seek a positive constant $\hat{\Delta}$ such that for all $z \in \mathrm{cl}(\mathcal{B}(\hat{z}, \hat{\Delta}/2)) \cap \mathcal{F}$, and all $\Delta \le \hat{\Delta}/2$, there is a step $\widetilde{\Delta z}$ that satisfies the properties (1.4) and (1.5).

We choose initially $\hat{\Delta} = \hat{\Delta}_0$, and assume that Δz satisfies $\|\Delta z\| \le \Delta \le \hat{\Delta}/2$. From feasibility of z, (2.7), and (1.2b), and the fact that c and d are twice continuously differentiable in $\mathcal{N}(L_0)$, we have that

$$c(z + \Delta z) = c(z) + \nabla c(z)^T \Delta z + O(\|\Delta z\|^2) = O(\|\Delta z\|^2)$$

and

$$[d(z + \Delta z)]_+ = \left[d(z) + \nabla d(z)^T \Delta z + O(\|\Delta z\|^2)\right]_+ = O(\|\Delta z\|^2).$$

We now set $\hat{\Delta} \leftarrow \min(\hat{\Delta}, \hat{\Delta}_3)$ and apply Assumption 3. Since

$$\|(z + \Delta z) - \hat{z}\| \le \|z - \hat{z}\| + \|\Delta z\| \le \hat{\Delta}/2 + \hat{\Delta}/2 \le \hat{\Delta} \le \hat{\Delta}_3,$$

we have from Assumption 3 and the estimates above that

$$\min_{v \in \mathcal{F}} \|v - (z + \Delta z)\| \le \zeta\left(\|c(z + \Delta z)\| + \|[d(z + \Delta z)]_+\|\right) = O(\zeta\|\Delta z\|^2), \tag{2.8}$$

where ζ may depend on $\hat{z}$. Since $v = z$ is feasible for (2.8), we have that any solution of this projection problem satisfies $\|v - (z + \Delta z)\| \le \|\Delta z\|$. Hence, the minimization on the left-hand side of (2.8) may be restricted to the nonempty compact set $\mathrm{cl}(\mathcal{B}(z + \Delta z, \|\Delta z\|)) \cap \mathcal{F}$, so the minimum is attained. If we use the minimizer v to define $\widetilde{\Delta z} = v - z$, then from (2.8) we have

$$\|\widetilde{\Delta z} - \Delta z\| = O(\zeta\|\Delta z\|^2).$$

Therefore, by decreasing $\hat{\Delta}$ if necessary, we find that (1.5) is satisfied for our choice $\phi(t) = \sqrt{t}$.

Since $\mathcal{B}(\hat{z}, \hat{\Delta}/2)$, $\hat{z} \in L_0$ forms an open cover of L_0, and since L_0 is compact, we can define a finite subcover. By defining $\Delta_{\rm def}$ to be the minimum of the $\hat{\Delta}/2$ over the subcover, we have that $\Delta_{\rm def}$ is positive and has the desired property. $\square$

3. Global Convergence. In this section, we prove convergence to stationary points of (1.1). Our results are of two types. We show first that at least one limit point is stationary in Section 3.2, and then under a stronger assumption that all limit points are stationary in Section 3.3.

We start with some technical results.

3.1. Technical Results. The first result concerns the solution of a linear programming variant of the SQP subproblem (1.2), (1.3). Its proof appears in the Appendix.

LEMMA 3.1. *Let f, c, and d be as defined in (1.1), and let $C(z,\tau)$ denote the* negative *of the optimal value function of the following problem, for some $z \in \mathcal{F}$ and $\tau > 0$:*

$$\text{CLP}(z,\tau): \qquad \min_{w} \nabla f(z)^T w \text{ subject to} \tag{3.1a}$$

$$c(z) + \nabla c(z)^T w = 0, \ \ d(z) + \nabla d(z)^T w \leq 0, \ \ w^T w \leq \tau^2. \tag{3.1b}$$

Let $\bar{z} \in \mathcal{F}$, and suppose that the MFCQ conditions (1.9) are satisfied at $\bar{z}$. Then $C(\bar{z},1) \geq 0$, with $C(\bar{z},1) = 0$ if and only if the KKT conditions (1.7) are satisfied at $\bar{z}$; that is, if $\bar{z}$ is a stationary point for (1.1).

When (1.7) are not *satisfied at $\bar{z}$, there exist positive quantities R_2 and ϵ such that for any $z \in \mathcal{B}(\bar{z}, R_2) \cap \mathcal{F}$, we have $C(z,1) \geq \epsilon$.*

An immediate consequence of this result is that for any subsequence $\{z^k\}_{k\in\mathcal{K}}$ such that $z^k \to \bar{z}$ and $C(z^k,1) \to 0$, where $\bar{z}$ satisfies the MFCQ conditions, we must have that $\bar{z}$ is stationary for (1.1). As a consequence, if all limit points of the algorithm satisfy MFCQ but none are stationary, we have that there is a constant $\epsilon > 0$ such that $C(z^k,1) \geq \epsilon$ for all k.

Note that $C(z,\tau)$ is a convex function of $\tau > 0$. In particular, if $w(z,\tau)$ is the optimum in $\text{CLP}(z,\tau)$, the point $\alpha w(z,\tau)$ is feasible in $\text{CLP}(z,\alpha\tau)$ for all $\alpha \in [0,1]$, so that

$$C(z,\alpha\tau) \geq \alpha C(z,\tau), \qquad \text{for all } \tau > 0, \text{ all } \alpha \in [0,1]. \tag{3.2}$$

For convenience, we restate the the subproblem (1.2), (1.3) at an arbitrary feasible point z as follows:

$$\min_{\Delta z} m(\Delta z) \stackrel{\text{def}}{=} \nabla f(z)^T \Delta z + \tfrac{1}{2}\Delta z^T H \Delta z \text{ subject to} \tag{3.3a}$$

$$c(z) + \nabla c(z)^T \Delta z = 0, \ \ d(z) + \nabla d(z)^T \Delta z \leq 0, \tag{3.3b}$$

$$\|D\Delta z\|_p \leq \Delta, \tag{3.3c}$$

where D satisfies Assumption 1. Consider now the following problem, obtained by omitting the quadratic term from (3.3a):

$$\min_{\Delta z^{\text{L}}} \nabla f(z)^T \Delta z^{\text{L}} \text{ subject to} \tag{3.4a}$$

$$c(z) + \nabla c(z)^T \Delta z^{\text{L}} = 0, \ \ d(z) + \nabla d(z)^T \Delta z^{\text{L}} \leq 0, \tag{3.4b}$$

$$\|D\Delta z^{\text{L}}\|_p \leq \Delta. \tag{3.4c}$$

Denote the *negative* of the optimal value function for this problem by $V(z,D,\Delta)$. Referring to (3.1) and Assumption 1, we see that the feasible region for $\text{CLP}(z,\delta^{-1}\Delta)$ is contained in the feasible region for (3.4), and the objectives are the same. Hence for $\Delta \in (0,1]$, we have from (3.2) that

$$V(z,D,\Delta) \geq C(z,\delta^{-1}\Delta) \geq \delta^{-1}C(z,1)\Delta.$$

For $\Delta > 1$, on the other hand, we have

$$V(z,D,\Delta) \geq C(z,\delta^{-1}\Delta) \geq \delta^{-1}C(z,1).$$

Hence, by combining these observations, we obtain that

$$V(z, D, \Delta) \geq \delta^{-1} C(z, 1) \min(1, \Delta). \tag{3.5}$$

The following result follows immediately from this observation together with Lemma 3.1.

LEMMA 3.2. *Suppose that Assumption 1 holds. Let $\bar{z}$ be a feasible point for (1.1) and that the MFCQ conditions (1.9) but not the KKT conditions (1.7) are satisfied at $\bar{z}$. Then there exist positive quantities R_2 and ϵ such that for any $z \in \mathcal{B}(\bar{z}, R_2) \cap \mathcal{F}$ and any $\Delta > 0$, we have*

$$V(z, D, \Delta) \geq \delta^{-1} C(z, 1) \min(1, \Delta) \geq \delta^{-1} \epsilon \min(1, \Delta). \tag{3.6}$$

If Assumption 1 holds, we have that

$$\|\Delta z^{\mathrm{L}}\|_2 \leq \delta \|D \Delta z^{\mathrm{L}}\|_p \leq \delta \Delta. \tag{3.7}$$

Hence, since Δz is optimal for (3.3), and since Δz^{L} that solves (3.4) is feasible for this problem we have

$$\begin{aligned}
& m(\Delta z) \leq m(\Delta z^{\mathrm{L}}) \\
& = (\Delta z^{\mathrm{L}})^T \nabla f(z) + \tfrac{1}{2} (\Delta z^{\mathrm{L}})^T H (\Delta z^{\mathrm{L}}) \\
& \leq -V(z, D, \Delta) + \tfrac{1}{2} \delta^2 \|H\| \Delta^2 \\
& \leq -\delta^{-1} \min(1, \Delta) C(z, 1) + \tfrac{1}{2} \delta^2 \|H\| \Delta^2
\end{aligned} \tag{3.8}$$

where the last inequality follows from (3.5).

We now define the Cauchy point for problem (3.3) as

$$\Delta z^{\mathrm{C}} = \alpha^{\mathrm{C}} \Delta z^{\mathrm{L}}, \tag{3.9}$$

where

$$\alpha^{\mathrm{C}} = \arg \min_{\alpha \in [0,1]} \alpha \nabla f(z)^T \Delta z^{\mathrm{L}} + \tfrac{1}{2} \alpha^2 (\Delta z^{\mathrm{L}})^T H \Delta z^{\mathrm{L}}. \tag{3.10}$$

We show that Δz^{C} has the following property:

$$m(\Delta z^{\mathrm{C}}) \leq -\tfrac{1}{2} C(z, 1) \min\left[\delta^{-1}, \delta^{-1}\Delta, (\delta^4 \bar{\Delta}^2 \|H\|_2)^{-1} C(z, 1)\right]. \tag{3.11}$$

We prove (3.11) by considering two cases. First, when $(\Delta z^{\mathrm{L}})^T H \Delta z^{\mathrm{L}} \leq 0$, we have $\alpha^{\mathrm{C}} = 1$ in (3.10) and hence $\Delta z^{\mathrm{C}} = \Delta z^{\mathrm{L}}$. As in (3.6), we have

$$m(\Delta z^{\mathrm{C}}) = m(\Delta z^{\mathrm{L}}) \leq -V(z, D, \Delta) \leq -\delta^{-1} C(z, 1) \min(1, \Delta),$$

so the result (3.11) holds in this case. In the alternative case $(\Delta z^{\mathrm{L}})^T H \Delta z^{\mathrm{L}} > 0$, we have

$$\alpha = \min\left(1, \frac{-\nabla f(z)^T \Delta z^{\mathrm{L}}}{(\Delta z^{\mathrm{L}})^T H \Delta z^{\mathrm{L}}}\right). \tag{3.12}$$

If the minimum is achieved at 1, we have from $(\Delta z^{\mathrm{L}})^T H \Delta z^{\mathrm{L}} \leq -\nabla f(z)^T \Delta z^{\mathrm{L}}$ and Lemma 3.2 that

$$m(\Delta z^{\mathrm{C}}) = m(\Delta z^{\mathrm{L}}) \leq \tfrac{1}{2} \nabla f(z)^T \Delta z^{\mathrm{L}} \leq -\tfrac{1}{2} \delta^{-1} C(z, 1) \min(1, \Delta), \tag{3.13}$$

and therefore again (3.11) is satisfied. If the min in (3.12) is achieved at $-\nabla f(z)^T \Delta z^{\mathrm{L}}/(\Delta z^{\mathrm{L}})^T H \Delta z^{\mathrm{L}}$, we have from Lemma 3.2 that

$$m(\Delta z^{\mathrm{C}}) = m(\alpha \Delta z^{\mathrm{L}}) = -\tfrac{1}{2} \frac{(\nabla f(z)^T \Delta z^{\mathrm{L}})^2}{(\Delta z^{\mathrm{L}})^T H \Delta z^{\mathrm{L}}} \leq -\tfrac{1}{2} \frac{\delta^{-2} \min(1, \Delta^2) C(z,1)^2}{\|H\|_2 \|\Delta z^{\mathrm{L}}\|_2^2}. \tag{3.14}$$

Because of (3.7), we have from (3.14) that

$$\begin{aligned} m(\Delta z^{\mathrm{C}}) &\leq -\tfrac{1}{2} \frac{\delta^{-2} \min(1, \Delta^2) C(z,1)^2}{\delta^2 \Delta^2 \|H\|_2} \\ &= -\tfrac{1}{2} (\delta^4 \|H\|_2)^{-1} \min(1, \Delta^{-2}) C(z,1)^2 \leq -\tfrac{1}{2} (\delta^4 \bar{\Delta}^2 \|H\|_2)^{-1} C(z,1)^2, \end{aligned}$$

which again implies that (3.11) is satisfied.

Since Δz^{C} is feasible for (3.3), we have proved the following lemma.

LEMMA 3.3. *Suppose that z is a feasible point and that Assumption 1 holds. Suppose that Δz^C is obtained from (3.4), (3.9), and (3.10). Then the decrease in the model function m obtained by the point Δz^C satisfies the bound (3.11), and therefore the solution Δz of (3.3) satisfies the similar bound*

$$m(\Delta z) \leq -\tfrac{1}{2} C(z,1) \min \left[\delta^{-1}, \delta^{-1}\Delta, (\delta^4 \bar{\Delta}^2 \|H\|_2)^{-1} C(z,1) \right], \tag{3.15}$$

where $C(z,1)$ is the negative of the optimal value function of CLP$(z,1)$ defined in (3.1).

Note that this lemma holds even when we assume only that Δz is feasible for (3.3) and satisfies $m(\Delta z) \leq m(\Delta z^{\mathrm{C}})$. This relaxation is significant since, when H is indefinite, the complexity of finding a solution of (3.3) is greater than the complexity of computing Δz^{C}.

3.2. Result I: At Least One Stationary Limit Point. We now discuss convergence of the sequence of iterates generated by the algorithm under the assumptions of Section 2, and the additional assumption that the Hessians H_k of (1.2) are bounded as follows:

$$\|H_k\|_2 \leq \sigma_0 + \sigma_1 k, \quad k = 0, 1, 2 \ldots. \tag{3.16}$$

The style of analysis follows that of a number of earlier works on convergence of trust-region algorithms for unconstrained, possibly nonsmooth problems; for example Yuan [16], Wright [15]. However, many modifications are needed to adapt the algorithms to constrained problems and to the algorithm of Section 2.

We first prove a key lemma as a preliminary to the global convergence result of this section. It finds a lower bound on the trust-region radii in the case that the algorithm has no stationary limit points.

LEMMA 3.4. *Suppose that Assumptions 1, 2, and 3 are satisfied, and that all limit points of the algorithm satisfy MFCQ and none are stationary; that is, there is $\epsilon > 0$ such that*

$$C(z^k, 1) \geq \epsilon, \quad k = 0, 1, 2, \ldots.$$

Then there is a constant $T > 0$ such that

$$\Delta_k \geq T/N_k, \quad k = 0, 1, 2, \ldots, \tag{3.17}$$

where

$$N_k \stackrel{\text{def}}{=} 1 + \max_{i=0,1,\ldots,k} \|H_k\|_2.$$

Proof. For $\Delta_k \geq 1$, the claim (3.17) obviously holds with $T = 1$. Hence, we assume for the remainder of the proof that $\Delta_k \in (0, 1]$.

From Lemma 3.3, we have

$$\begin{aligned} -m_k(\Delta z^k) &\geq \tfrac{1}{2}\epsilon \min\left[\delta^{-1}\Delta_k, (\delta^4 \bar{\Delta}^2 \|H_k\|_2)^{-1}\epsilon\right] \\ &\geq \tfrac{1}{2}\epsilon \min\left[\delta^{-1}\Delta_k, (\delta^4 \bar{\Delta}^2 N_k)^{-1}\epsilon\right]. \end{aligned} \tag{3.18}$$

We define the constants $\bar{\sigma}$ and γ as follows:

$$\bar{\sigma} = \sup\{\|\nabla^2 f(z)\|_2 \,|\, z \in \mathcal{N}(L_0)\}, \qquad \gamma = \sup\{\|\nabla f(z)\|_2 \,|\, z \in \mathcal{F}\}, \tag{3.19}$$

where $\mathcal{N}(L_0)$ is the neighborhood defined in Assumption 2. Suppose now that T is chosen to satisfy the following conditions:

$$T \leq 1, \tag{3.20a}$$
$$\{z \,|\, \mathrm{dist}(z, L_0) \leq 2\delta T\} \subset \mathcal{N}(L_0), \tag{3.20b}$$
$$2T \leq \epsilon/(\delta^3, \bar{\Delta}^2), \tag{3.20c}$$
$$\phi(2\delta T) \leq 1, \tag{3.20d}$$
$$(\gamma + 2\bar{\sigma}\delta)\phi(2\delta T)\delta^2 \leq (1/48)\epsilon, \tag{3.20e}$$
$$2\bar{\sigma}\delta^3 T \leq (1/48)\epsilon, \tag{3.20f}$$
$$\delta^3 T \leq (1/48)\epsilon, \tag{3.20g}$$

where $\phi(\cdot)$ is defined in (1.5).

For any k with

$$\|\Delta z^k\| \leq 2\delta T, \tag{3.21}$$

we have from Taylor's Theorem and the definition of m_k that

$$\begin{aligned} & f(z^k) - f(z^k + \widetilde{\Delta z}^k) + m_k(\Delta z^k) \\ &= -\nabla f(z^k)^T \widetilde{\Delta z}^k - \tfrac{1}{2}(\widetilde{\Delta z}^k)^T \nabla^2 f(z^k_\theta) \widetilde{\Delta z}^k + \nabla f(z^k)^T \Delta z^k + \tfrac{1}{2}(\Delta z^k)^T H_k \Delta z^k \\ &= \left[\nabla f(z^k) + \nabla^2 f(z^k_\theta)\Delta z^k\right]^T (\Delta z^k - \widetilde{\Delta z}^k) \\ &\quad - \tfrac{1}{2}(\widetilde{\Delta z}^k - \Delta z^k)^T \nabla^2 f(z^k_\theta)(\widetilde{\Delta z}^k - \Delta z^k) - \tfrac{1}{2}(\Delta z^k)^T(\nabla^2 f(z^k_\theta) - H_k)\Delta z^k, \end{aligned} \tag{3.22}$$

where z^k_θ lies on the line segment between z^k and $z^k + \widetilde{\Delta z}^k$. If k is an index satisfying (3.21), we have from feasibility of both z^k and $z^k + \widetilde{\Delta z}^k$ that

$$\begin{aligned} \mathrm{dist}(z^k_\theta, L_0) &\leq \tfrac{1}{2}\|\widetilde{\Delta z}^k\|_2 \\ &\leq \tfrac{1}{2}\left(\|\Delta z^k\|_2 + \|\Delta z^k - \widetilde{\Delta z}^k\|_2\right) \\ &\leq \tfrac{1}{2}\left(\|\Delta z^k\|_2 + \phi(\|\Delta z^k\|_2)\|\Delta z^k\|_2\right) \\ &\leq \tfrac{1}{2}(2\delta T + \phi(2\delta T)2\delta T) \leq 2\delta T, \end{aligned}$$

$\|\nabla^2 f(z_\theta^k)\|_2 \le \bar{\sigma}$. For k satisfying (3.21), we have from (3.22) that

$$\begin{aligned}
&\left| f(z^k) - f(z^k + \widetilde{\Delta z}^k) + m_k(\Delta z^k) \right| \\
&\le \left(\|\nabla f(z^k)\|_2 + \|\nabla^2 f(z_\theta^k)\|_2 \|\Delta z^k\|_2 \right) \|\Delta z^k - \widetilde{\Delta z}^k\|_2 \\
&\qquad + \tfrac{1}{2}\|\nabla^2 f(z_\theta^k)\|_2 \|\widetilde{\Delta z}^k - \Delta z^k\|_2^2 + \tfrac{1}{2}(\|\nabla^2 f(z_\theta^k)\|_2 + \|H_k\|_2)\|\Delta z^k\|_2^2 \\
&\le (\gamma + 2\bar{\sigma}\delta T)\|\Delta z^k - \widetilde{\Delta z}^k\|_2 + \tfrac{1}{2}\bar{\sigma}\|\Delta z^k - \widetilde{\Delta z}^k\|_2^2 + \tfrac{1}{2}(\bar{\sigma} + N_k)\|\Delta z^k\|_2^2. \quad (3.23)
\end{aligned}$$

Now using (1.5) and Assumption 1,we have for indices k satisfying (3.21) that

$$\begin{aligned}
&\left| f(z^k) - f(z^k + \widetilde{\Delta z}^k) + m_k(\Delta z^k) \right| \\
&\le (\gamma + 2\bar{\sigma}\delta T)\phi(\|\Delta z^k\|_2)\|\Delta z^k\|_2 + \tfrac{1}{2}\bar{\sigma}\phi(\|\Delta z^k\|_2)^2\|\Delta z^k\|_2^2 + \tfrac{1}{2}(\bar{\sigma} + N_k)\|\Delta z^k\|_2^2 \\
&\le \left[(\gamma + 2\bar{\sigma}\delta T)\phi(2\delta T) + \tfrac{1}{2}\bar{\sigma}\phi(2\delta T)^2 2\delta T + \bar{\sigma}\delta T + \tfrac{1}{2}N_k\|\Delta z^k\|_2 \right] \|\Delta z^k\|_2 \\
&\le \left[(\gamma + 2\bar{\sigma}\delta T)\phi(2\delta T) + \bar{\sigma}\delta T + \bar{\sigma}\delta T + \tfrac{1}{2}N_k\|\Delta z^k\|_2 \right] \|\Delta z^k\|_2 \\
&\le \left[\frac{1}{48}\frac{\epsilon}{\delta^2} + \frac{1}{48}\frac{\epsilon}{\delta^2} + \frac{1}{2}N_k\|\Delta z^k\|_2 \right] \|\Delta z^k\|_2, \quad (3.24)
\end{aligned}$$

where we used (3.20) to derive the various inequalities.

Now suppose that (3.17) is not satisfied for all k and for our choice of T, and suppose that l is the first index at which it is violated, that is,

$$\Delta_l < T/N_l. \quad (3.25)$$

We exclude the case $l = 0$ (by decreasing T further, if necessary), and consider the index $l-1$. Since $\Delta_k \ge (1/2)\|D_{k-1}\Delta z^{k-1}\|_p$ for all k, and since $N_l \ge 1$, we have

$$\|\Delta z^{l-1}\|_2 \le \delta\|D_{l-1}\Delta z^{l-1}\|_p \le 2\delta\Delta_l < 2\delta T, \quad (3.26)$$

so that $l-1$ satisfies (3.21). Hence, the bound (3.24) applies with $k = l-1$, and we have

$$\begin{aligned}
&\left| f(z^{l-1}) - f(z^{l-1} + \widetilde{\Delta z}^{l-1}) + m_{l-1}(\Delta z^{l-1}) \right| \quad (3.27) \\
&\le \left[\frac{1}{24}\frac{\epsilon}{\delta^2} + \frac{1}{2}N_{l-1}\|\Delta z^{l-1}\|_2 \right] \|\Delta z^{l-1}\|_2.
\end{aligned}$$

Since $N_{l-1} \le N_l$, we have from (3.26) that

$$N_{l-1}\|\Delta z^{l-1}\|_2 \le 2\delta N_l \Delta_l < 2\delta T. \quad (3.28)$$

Therefore by using (3.27) and (3.20g), we obtain

$$\begin{aligned}
&\left| f(z^{l-1}) - f(z^{l-1} + \widetilde{\Delta z}^{l-1}) + m_{l-1}(\Delta z^{l-1}) \right| \\
&\le \left(\frac{1}{24}\frac{\epsilon}{\delta^2} + \delta T \right) \|\Delta z^{l-1}\|_2 \le \frac{1}{16}\frac{\epsilon}{\delta^2}\|\Delta z^{l-1}\|_2. \quad (3.29)
\end{aligned}$$

Returning to (3.18), we have for $k = l-1$ that

$$\delta^{-1}\Delta_{l-1} \ge \delta^{-1}\|D_{l-1}\Delta z^{l-1}\|_p \ge \delta^{-2}\|\Delta z^{l-1}\|_2,$$

and using (3.28) and (3.20c), we have

$$\frac{\epsilon}{\delta^4 \bar{\Delta}^2 N_{l-1}} \geq \frac{\epsilon}{\delta^4 \bar{\Delta}^2} \frac{\|\Delta z^{l-1}\|_2}{2\delta T} \geq \delta^{-2} \|\Delta z^{l-1}\|_2.$$

Hence, from (3.18) and the last two inequalities, we have

$$-m_{l-1}(\Delta z^{l-1}) \geq \frac{1}{2} \frac{\epsilon}{\delta^2} \|\Delta z^{l-1}\|_2. \tag{3.30}$$

By comparing (3.29) and (3.30), we have from (2.1) that

$$\begin{aligned}
\rho_{l-1} &= \frac{f(z^{l-1}) - f(z^{l-1} + \widetilde{\Delta z}^{l-1})}{-m_{l-1}(\Delta z^{l-1})} \\
&\geq 1 - \frac{\left| f(z^{l-1}) - f(z^{l-1} + \widetilde{\Delta z}^{l-1}) + m_{l-1}(\Delta z^{l-1}) \right|}{-m_{l-1}(\Delta z^{l-1})} \\
&\geq 1 - \frac{1}{8} = \frac{7}{8}.
\end{aligned}$$

Hence, by the workings of the algorithm, we have $\Delta_l \geq \Delta_{l-1}$. But since $N_{l-1} \leq N_l$, we have $N_{l-1}\Delta_{l-1} \leq N_l \Delta_l$, so that $\Delta_{l-1} \leq T/N_{l-1}$, which contradicts the definition of l as the first index that violates (3.17). We conclude that no such l exists, and hence that (3.17) holds. □

The following technical lemma, attributed to M. J. D. Powell, is proved in Yuan [16, Lemma 3.4].

LEMMA 3.5. *Suppose $\{\Delta_k\}$ and $\{N_k\}$ are two sequences such that $\Delta_k \geq T/N_k$ for all k and some constant $T > 0$. Let $\mathcal{K} \subset \{0, 1, 2, \dots\}$ be defined such that*

$$\Delta_{k+1} \leq \tau_0 \Delta_k \text{ if } k \in \mathcal{K}, \tag{3.31a}$$

$$\Delta_{k+1} \leq \tau_1 \Delta_k \text{ if } k \notin \mathcal{K}, \tag{3.31b}$$

$$N_{k+1} \geq N_k \text{ for all } k, \tag{3.31c}$$

$$\sum_{k \in \mathcal{K}} \min(\Delta_k, 1/N_k) < \infty, \tag{3.31d}$$

where τ_0 and τ_1 are constants satisfying $0 < \tau_1 < 1 < \tau_0$. Then

$$\sum_{k=0}^{\infty} 1/N_k < \infty. \tag{3.32}$$

Our main global convergence result for this section is as follows.

THEOREM 3.6. *Suppose that Assumptions 1, 2, and 3 are satisfied, and assume that all limit points of the algorithm satisfy MFCQ. Suppose in addition that the approximate Hessians H_k satisfy (3.16); that is, $\|H_k\|_2 \leq \sigma_0 + k\sigma_1$, for some nonnegative constants σ_0 and σ_1. Then the algorithm has a stationary limit point.*

Proof. Since all iterates belong to the compact level set L_0, the algorithm has at least one limit point. If none are stationary, then the conditions of Lemma 3.4 are satisfied. We apply Lemma 3.5, choosing $\mathcal{K}$ to be the subsequence of $\{0, 1, 2 \dots\}$ at which the trust-region radius is *not* reduced. We can then set $\tau_0 = 2$, $\tau_1 = 0.5$, and define N_k as in Lemma 3.4. At the iterates $k \in \mathcal{K}$, the algorithm takes a step, and we have $\rho_k \geq \eta$. By using (3.18), we then have

$$\begin{aligned} f(z^k) - f(z^k + \widetilde{\Delta z}^k) &\geq -\eta m_k(\Delta z^k) \\ &\geq \tfrac{1}{2}\eta\epsilon \min(\delta^{-1}, \delta^{-1}\Delta_k, \delta^{-4}\bar{\Delta}^{-2}\epsilon/N_k) \\ &\geq \tfrac{1}{2}\eta\epsilon \min(\delta^{-1}, \delta^{-4}\bar{\Delta}^{-2}\epsilon)\min(\Delta_k, 1/N_k). \end{aligned}$$

By summing both sides of this inequality over $k \in \mathcal{K}$, and using the fact that $f(z^k)$ is bounded below (by $f(\bar{z})$, where $\bar{z}$ is a limit point), we have that condition (3.31d) is satisfied. Then the conclusion (3.32) holds. However this inequality contradicts (3.16), so we conclude that the algorithm must have a stationary limit point. ∎

3.3. Result II: All Limit Points are Stationary. In this section, we replace the bound (3.16) on the Hessians H_k by a uniform bound

$$\|H_k\|_2 \leq \sigma, \tag{3.33}$$

for some constant σ, and obtain a stronger global convergence result; namely, that all limit points are stationary.

As a preliminary to the main result of this section, we show that for any limit point $\bar{z}$ of Algorithm FP-SQP at which MFCQ but not KKT conditions are satisfied, there is a subsequence $\mathcal{K}$ with $z^k \to_{k\in\mathcal{K}} \bar{z}$ and $\Delta_k \to_{k\in\mathcal{K}} 0$.

LEMMA 3.7. *Suppose that Assumptions 1, 2, and 3 are satisfied, and that the Hessians H_k satisfy the bound (3.33) for some $\sigma > 0$. Suppose that $\bar{z}$ is a limit point of the sequence $\{z^k\}$ such that the MFCQ condition (1.9) holds but the KKT conditions (1.7) are not satisfied at $\bar{z}$. Then there exists an (infinite) subsequence $\mathcal{K}$ such that*

$$\lim_{k\in\mathcal{K}} z^k = \bar{z}, \tag{3.34}$$

and

$$\lim_{k\in\mathcal{K}} \Delta_k = 0. \tag{3.35}$$

Proof. We first define ϵ at the point $\bar{z}$ as in Lemma 3.2, and define

$$R = \min(R_2, \hat{R}), \tag{3.36}$$

where $\hat{R}$ is defined as in the proof of Theorem 2.2 and R_2 is defined as in Lemma 3.2. Then for all $z \in \mathcal{B}(\bar{z}, R) \cap \mathcal{F}$, we have $C(z,1) \geq \epsilon$. Hence, from Lemma 3.3, the solution Δz of the trust-region subproblem at (3.3) with $\Delta \in (0,1]$ satisfies

$$\begin{aligned} m(\Delta z) &\leq -\tfrac{1}{2}C(z,1)\min\left[\delta^{-1}, \delta^{-1}\Delta, (\delta^4\bar{\Delta}^2\|H\|_2)^{-1}C(z,1)\right] \\ &\leq -\tfrac{1}{2}\epsilon\min\left[\delta^{-1}, \delta^{-1}\Delta, (\delta^4\bar{\Delta}^2\sigma)^{-1}\epsilon\right], \end{aligned} \tag{3.37}$$

where we used the bound (3.33) to obtain the second inequality.

Because $\bar{z}$ is a limit point, we can certainly choose a subsequence $\mathcal{K}$ satisfying (3.34). By deleting the elements of this subsequence that lie outside the ball $\mathcal{B}(\bar{z}, R)$, we have from (3.37) that

$$m_k(\Delta z^k) \leq -\tfrac{1}{2}\epsilon\min\left[\delta^{-1}, \delta^{-1}\Delta_k, (\delta^4\bar{\Delta}^2\sigma)^{-1}\epsilon\right], \quad \text{for all } k \in \mathcal{K}. \tag{3.38}$$

We prove the result (3.35) by modifying $\mathcal{K}$ and taking further subsequences as necessary. Consider first the case in which $\{z^k\}_{k\in\mathcal{K}}$ takes on only a finite number of distinct values. We then must have that $z^k = \bar{z}$ for all $k \in \mathcal{K}$ sufficiently large. Now, fix $k \in \mathcal{K}$ large enough that this property is satisfied, and suppose for contradiction that some subsequent iterate in the full sequence $\{z^k\}$ is different from $\bar{z}$. If $\bar{k} \geq k$ is some iterate such that

$$f(z^{\bar{k}}) < f(z^k) = f(\bar{z}),$$

we have by monotonicity of $\{f(z^l)\}$ (for the full sequence of function values) that

$$f(z^l) \leq f(z^{\bar{k}}) < f(\bar{z})$$

for *all* $l > \bar{k}$. Hence the function values in the tail of the full sequence are bounded away from $f(\bar{z})$, so it is not possible to choose a subsequence $\mathcal{K}$ with the property (3.34). Therefore, we have that $z^l = \bar{z}$ for *all* $l \geq k$, so that all steps generated by Algorithm FP-SQP after iteration k fail the acceptance condition. We then have that

$$\Delta_{l+1} = \tfrac{1}{2}\|D_l \Delta z^l\|_p \leq \tfrac{1}{2}\Delta_l, \ \text{ for all } l \geq k,$$

so that $\Delta_l \to 0$ as $l \to \infty$ (for the full sequence). Hence, in particular, (3.35) holds.

We consider now the second case, in which $\{z^k\}_{k\in\mathcal{K}}$ takes on an infinite number of distinct values. Without loss of generality, we can assume that *all* elements of this subsequence are distinct (by dropping the repeated elements if necessary). Moreover, we can assume that $z^{k+1} \neq z^k$ for all $k \in \mathcal{K}$, by replacing k if necessary by the largest index $\bar{k}$ such that $\bar{k} \geq k$ and $z^{\bar{k}} = z^k$. Thus, we have that the sufficient decrease condition $\rho_k \geq \eta$ is satisfied at all $k \in \mathcal{K}$. Therefore from (2.1) and (3.37), and the easily demonstrated fact that $f(z^l) \geq f(\bar{z})$ for *all* $l = 0, 1, 2, \ldots$, we have

$$\begin{aligned} f(z^k) - f(\bar{z}) &\geq f(z^k) - f(z^{k+1}) \\ &\geq -\eta m_k(\Delta z^k) \\ &\geq \tfrac{1}{2}\eta\epsilon \min\left[\delta^{-1}, \delta^{-1}\Delta_k, (\delta^4\bar{\Delta}^2\sigma)^{-1}\epsilon\right] \geq 0. \end{aligned}$$

Since $f(z^k) \to_{k\in\mathcal{K}} f(\bar{z})$, we have from this chain of inequalities that (3.35) is satisfied in this case too. Hence, we have demonstrated (3.35). ☐

We now prove the main global convergence result of this section.

THEOREM 3.8. *Suppose that Assumptions 1, 2, and 3 are satisfied, and that the Hessian approximations H_k satisfy (3.33). Then the algorithm cannot have a limit point $\bar{z}$ at which the MFCQ condition (1.9) holds but the KKT conditions (1.7) are not satisfied.*

Proof. Let R, ϵ, σ be defined as in the proof of Lemma 3.7. Suppose for contradiction that there exists a subsequence $\{z^k\}_{k\in\mathcal{K}}$ such that $z^k \to_{k\in\mathcal{K}} \bar{z}$. From Lemma 3.7, without loss of generality, we can assume that (3.35) holds. The inequality (3.38) also holds for the subsequence $\mathcal{K}$.

Let $\bar{\sigma}$ and γ be defined as in (3.19). We now decrease R if necessary (while maintaining positivity), and define $\Delta_\phi > 0$ in such a way that the following conditions hold:

$$\phi(\Delta_\phi) \leq 1, \tag{3.39a}$$

$$\gamma\phi(\Delta_\phi) \leq \frac{1}{16}\frac{\epsilon}{\delta^2}, \tag{3.39b}$$

$$\mathcal{B}(\bar{z}, R + \Delta_\phi) \cap \mathcal{F} \subset \mathcal{N}(L_0), \tag{3.39c}$$

$$\Delta_\phi \leq \Delta_{\text{def}}, \tag{3.39d}$$

where $\Delta_{\rm def}$ is defined in Theorem 2.2. Note in particular from the latter theorem that $\widetilde{\Delta z}$ satisfying (1.4) and (1.5) exists whenever $\|\Delta z\|_2 \le \Delta_\phi$.

Given R and Δ_ϕ, we can now define $\tilde\Delta > 0$ small enough to satisfy the following properties:

$$\tilde\Delta \le 1, \tag{3.40a}$$

$$(2\bar\sigma + \tfrac{1}{2}\sigma)\delta\tilde\Delta \le \frac{1}{16}\frac{\epsilon}{\delta^2}, \tag{3.40b}$$

$$\tilde\Delta \le \frac{\Delta_\phi}{2\delta}, \tag{3.40c}$$

$$\tilde\Delta \le \frac{\epsilon}{\delta^3\bar\Delta^2\sigma}. \tag{3.40d}$$

We then define $\hat\epsilon > 0$ as follows:

$$\hat\epsilon = \tfrac{1}{2}\eta\epsilon \min\left(\delta^{-1}\tilde\Delta, \tfrac{1}{4}R/\delta^2, (\delta^4\bar\Delta^2\sigma)^{-1}\epsilon\right). \tag{3.41}$$

Finally, we define an index $q \in \mathcal{K}$ sufficient large that

$$\|z^q - \bar z\|_2 < R/2, \tag{3.42a}$$

$$f(z^q) - f(\bar z) \le \hat\epsilon/2. \tag{3.42b}$$

(Existence of such an index q follows immediately from $z^k \to_{\mathcal{K}} \bar z$.)

Consider the neighborhood

$$\mathrm{cl}\,(\mathcal{B}(z^q, R/2)) \cap \mathcal{F}, \tag{3.43}$$

which is contained in $\mathcal{B}(\bar z, R) \cap \mathcal{F}$ because of (3.42a). We consider two cases.

Case I: All remaining iterates $z^{q+1}, z^{q+2}, \ldots$ of the full sequence remain inside the neighborhood (3.43). If

$$\|D_k \Delta z^k\|_p \le \tilde\Delta, \quad \text{for any } k = q, q+1, q+2, \ldots, \tag{3.44}$$

we have from (1.5), monotonicity of ϕ, and the definition (3.40c) that

$$\begin{aligned} &\|\Delta z^k - \widetilde{\Delta z}^k\|_2 \le \phi(\|\Delta z^k\|_2)\|\Delta z^k\|_2 \\ &\quad\le \phi(\delta\|D_k\Delta z^k\|_p)\|\Delta z^k\|_2 \le \phi(\delta\tilde\Delta)\|\Delta z^k\|_2 \le \phi(\Delta_\phi)\|\Delta z^k\|_2 = \|\Delta z^k\|_2 \end{aligned} \tag{3.45}$$

so that

$$\|\widetilde{\Delta z}^k\|_2 \le 2\|\Delta z^k\|_2 \le 2\delta\|D_k\Delta z^k\|_p \le 2\delta\tilde\Delta \le \Delta_\phi. \tag{3.46}$$

We now show that whenever (3.44) occurs, the ratio ρ_k defined by (2.1) is at least $3/4$, so that the trust-region radius Δ_{k+1} for the next iteration is no smaller than the one for this iteration, Δ_k. As in the proof of Lemma 3.4, the relation (3.22) holds, with z_θ^k satisfying

$$\mathrm{dist}(z_\theta^k, L_0) \le \tfrac{1}{2}\|\widetilde{\Delta z}^k\|_2 \le \tfrac{1}{2}\Delta_\phi.$$

Hence, from (3.19) and (3.39c), we have $\|\nabla^2 f(z_\theta^k)\|_2 \le \bar\sigma$. Similarly to (3.23), we have

$$\begin{aligned} &\left| f(z^k) - f(z^k + \widetilde{\Delta z}^k) + m_k(\Delta z^k)\right| \\ &\le \left(\|\nabla f(z^k)\|_2 + \|\nabla^2 f(z_\theta^k)\|_2\|\Delta z^k\|_2\right)\|\Delta z^k - \widetilde{\Delta z}^k\|_2 \\ &\qquad + \tfrac{1}{2}\|\nabla^2 f(z_\theta^k)\|_2\|\widetilde{\Delta z}^k - \Delta z^k\|_2^2 + \tfrac{1}{2}(\|\nabla^2 f(z_\theta^k)\|_2 + \|H_k\|_2)\|\Delta z^k\|_2^2 \\ &\le (\gamma + \bar\sigma\delta\tilde\Delta)\phi(\|\Delta z^k\|_2)\|\Delta z^k\|_2 + \tfrac{1}{2}\bar\sigma\phi(\|\Delta z^k\|_2)^2\|\Delta z^k\|_2^2 + \tfrac{1}{2}(\bar\sigma + \sigma)\|\Delta z^k\|_2^2, \end{aligned}$$

where we used (3.19) and $\|\Delta z^k\|_2 \le \delta\tilde{\Delta}$ from (3.46) in deriving the second inequality. Now using (3.46) again, together with monotonicity of ϕ, (3.39a), (3.39b), and (3.40b), we have

$$\begin{aligned}
&\left| f(z^k) - f(z^k + \widetilde{\Delta z}^k) + m_k(\Delta z^k) \right| \\
&\le (\gamma + \bar{\sigma}\delta\tilde{\Delta})\phi(\Delta_\phi)\|\Delta z^k\|_2 + \left[\tfrac{1}{2}\bar{\sigma}\phi(\Delta_\phi)^2\delta\tilde{\Delta} + \tfrac{1}{2}(\bar{\sigma}+\sigma)\delta\tilde{\Delta}\right]\|\Delta z^k\|_2 \\
&\le \left[\gamma\phi(\Delta_\phi) + \left(\bar{\sigma}\delta\tilde{\Delta} + \tfrac{1}{2}\bar{\sigma}\delta\tilde{\Delta} + \tfrac{1}{2}(\bar{\sigma}+\sigma)\delta\tilde{\Delta}\right)\right]\|\Delta z^k\|_2 \\
&= \left[\gamma\phi(\Delta_\phi) + (2\bar{\sigma} + \tfrac{1}{2}\sigma)\delta\tilde{\Delta}\right]\|\Delta z^k\|_2 \\
&\le \left(\frac{1}{16}\frac{\epsilon}{\delta^2} + \frac{1}{16}\frac{\epsilon}{\delta^2}\right)\|\Delta z^k\|_2 = \frac{1}{8}\frac{\epsilon}{\delta^2}\|\Delta z^k\|_2. \qquad (3.47)
\end{aligned}$$

Meanwhile, from (3.38) and since $z^k \in \mathcal{B}(\bar{z}, R) \cap \mathcal{F}$, we have

$$-m_k(\Delta z^k) \ge \tfrac{1}{2}\epsilon \min(\delta^{-1}, \delta^{-1}\Delta_k, (\delta^4\bar{\Delta}^2\sigma)^{-1}\epsilon). \qquad (3.48)$$

Now from Assumption 1 we have

$$\Delta_k \ge \|D_k\Delta z^k\|_p \ge \delta^{-1}\|\Delta z^k\|_2,$$

while from (3.40a) and (3.44), we have

$$1 \ge \tilde{\Delta} \ge \|D_k\Delta z^k\|_p \ge \delta^{-1}\|\Delta z^k\|_2.$$

From (3.40d) and Assumption 1, we have

$$\epsilon \ge \delta^3\bar{\Delta}^2\sigma\tilde{\Delta} \ge \delta^3\bar{\Delta}^2\sigma\|D_k\Delta z^k\|_p \ge \delta^2\bar{\Delta}^2\sigma\|\Delta z^k\|_2.$$

By substituting these last three expressions into (3.48), we obtain

$$-m_k(\Delta z^k) \ge \tfrac{1}{2}\frac{\epsilon}{\delta^2}\|\Delta z^k\|_2. \qquad (3.49)$$

We then have from (2.1), and using (3.47) and (3.49), that

$$\begin{aligned}
\rho_k &= \frac{f(z^k) - f(z^k + \widetilde{\Delta z}^k)}{-m_k(\Delta z^k)} \\
&\ge 1 - \frac{|f(z^k) - f(z^k + \widetilde{\Delta z}^k) + m_k(\Delta z^k)|}{-m_k(\Delta z^k)} \\
&\ge 3/4.
\end{aligned}$$

It follows that the algorithm sets

$$\Delta_{k+1} \ge \Delta_k \qquad (3.50)$$

for all k satisfying (3.44). For $k = q, q+1, q+2, \ldots$ *not* satisfying (3.44), Algorithm FP-SQP indicates that we may have reduction of the trust-region radius to

$$\Delta_{k+1} = (1/2)\|D_k\Delta z^k\|_p \ge (1/2)\tilde{\Delta}. \qquad (3.51)$$

By considering both cases, we conclude that

$$\Delta_k \geq \min(\Delta_q, (1/2)\tilde{\Delta}), \quad \text{for all } k = q, q+1, q+2, \ldots,$$

which contradicts (3.35). Hence, Case I cannot occur.

We now consider the alternative.

Case II: Some subsequent iterate $z^{q+1}, z^{q+2}, \ldots$ leaves the neighborhood (3.43). If z^l is the first iterate outside this neighborhood, note that all iterates z^k, $k = q, q+1, q+2, \ldots, l-1$ lie inside the set $\mathcal{B}(\bar{z}, R) \cap \mathcal{F}$, within which (3.37) applies. By summing over the "successful" iterates in this span, we have the following:

$$\begin{aligned}
& f(z^q) - f(z^l) \\
&= \sum_{\substack{k=q \\ z^k \neq z^{k+1}}}^{l-1} f(z^k) - f(z^{k+1}) \\
&\geq \sum_{\substack{k=q \\ z^k \neq z^{k+1}}}^{l-1} -\eta m_k(\Delta z^k) \qquad \text{by (2.1) and Algorithm FP-SQP} \\
&\geq \eta \sum_{\substack{k=q \\ z^k \neq z^{k+1}}}^{l-1} \tfrac{1}{2}\epsilon \min\left[\delta^{-1}, \delta^{-1}\Delta_k, (\delta^4 \bar{\Delta}^2 \sigma)^{-1}\epsilon\right] \qquad \text{by (3.37)} \\
&\geq \tfrac{1}{2}\eta\epsilon \min\left[\delta^{-1}, \delta^{-1} \sum_{\substack{k=q \\ z^k \neq z^{k+1}}}^{l-1} \Delta_k, (\delta^4 \bar{\Delta}^2 \sigma)^{-1}\epsilon\right].
\end{aligned} \tag{3.52}$$

If $\Delta_k \geq \tilde{\Delta}$ for any $k = q, q+1, \ldots, l-1$ with $z^k \neq z^{k+1}$, we have from (3.52) that

$$f(z^q) - f(z^l) \geq \tfrac{1}{2}\eta\epsilon \min\left[\delta^{-1}, \delta^{-1}\tilde{\Delta}, (\delta^4 \bar{\Delta}^2 \sigma)^{-1}\epsilon\right]. \tag{3.53}$$

Otherwise, since $\Delta_k \leq \tilde{\Delta}$ for all $k = q, q+1, \ldots, l-1$ with $z^k \neq z^{k+1}$, we have from Assumption 1 and (3.46) that

$$\Delta_k \geq \|D_k \Delta z^k\|_p \geq \delta^{-1}\|\Delta z^k\|_2 \geq \tfrac{1}{2}\delta^{-1}\|\widetilde{\Delta z}^k\|_2.$$

In this case, (3.52) becomes

$$f(z^q) - f(z^l) \geq \tfrac{1}{2}\eta\epsilon \min\left[\delta^{-1}, \sum_{\substack{k=q \\ z^k \neq z^{k+1}}}^{l-1} \tfrac{1}{2}\delta^{-2}\|\widetilde{\Delta z}^k\|_2, (\delta^4 \bar{\Delta}^2 \sigma)^{-1}\epsilon\right]. \tag{3.54}$$

However, because z^l lies outside the neighborhood (3.43) we have that

$$R/2 \leq \|z^q - z^l\|_2 \leq \sum_{\substack{k=q \\ z^k \neq z^{k+1}}}^{l-1} \|\widetilde{\Delta z}^k\|_2.$$

In this case, (3.54) becomes

$$f(z^q) - f(z^l) \geq \tfrac{1}{2}\eta\epsilon \min\left[\delta^{-1}, \tfrac{1}{4}\delta^{-2}R, (\delta^4\bar{\Delta}^2\sigma)^{-1}\epsilon\right]. \tag{3.55}$$

Hence, in either (3.53) or (3.55), and using (3.40a), we have

$$f(z^q) - f(z^l) \geq \hat{\epsilon},$$

for $\hat{\epsilon}$ defined in (3.41). But since $f(z^l) \geq f(\bar{z})$ (since $\bar{z}$ is a limit point of the full sequence), this inequality contradicts (3.42b). Hence, Case II cannot occur either, and the proof is complete. ∎

4. Local Convergence. We now examine local convergence behavior of the algorithm to a point z^* satisfying second-order sufficient conditions for optimality, under the assumption that $z^k \to z^*$. We do not attempt to obtain the most general possible superlinear convergence result, but rather make the kind of assumptions that are typically made in the local convergence analysis of SQP methods in which second derivatives of the objective and constraint functions are available. We also make additional assumptions on the feasibility perturbation process that is used to recover $\widetilde{\Delta z}^k$ from Δz^k. Ultimately, we show that the trust region becomes inactive, and that z^k converges Q-quadratically to z^*.

We assume a priori that z^* satisfies the first-order necessary conditions for optimality, and define the active set $\mathcal{A}^*$ as follows:

$$\mathcal{A}^* \stackrel{\text{def}}{=} \mathcal{A}(z^*), \tag{4.1}$$

where $\mathcal{A}(\cdot)$ is defined in (1.8). In this section, we use the following subvector notation:

$$d_{\mathcal{I}}(z) \stackrel{\text{def}}{=} [d_i(z)]_{i\in\mathcal{I}}, \quad \text{where } \mathcal{I} \subset \{1, 2, \ldots, r\}.$$

ASSUMPTION 4.

(a) The linear independence constraint qualification (LICQ) (1.10) is satisfied at z^*

(b) Strict complementarity holds; that is, for the (unique) multipliers (μ^*, λ^*) *satisfying the KKT conditions (1.7) at* $z = z^*$*, we have* $\lambda_i^* > 0$ *for all* $i \in \mathcal{A}^*$.

(c) Second-order sufficient conditions are satisfied at z^**; that is, there is* $\alpha > 0$ *such that*

$$v^T \nabla^2_{zz}\mathcal{L}(z^*, \mu^*, \lambda^*)v \geq \alpha\|v\|^2, \;\; \forall v \text{ s.t. } \nabla c(z^*)^T v = 0, \;\; \nabla d_{\mathcal{A}^*}(z^*)^T v = 0,$$

where the Lagrangian function $\mathcal{L}$ *is defined in (1.6).*

Besides these additional assumptions on the nature of the limit point z^*, we make additional assumptions on the algorithm itself. As mentioned above, we start by assuming that $z^k \to z^*$. We further assume that estimates $\mathcal{W}_k$ of the active set $\mathcal{A}^*$ and estimates (μ^k, λ^k) of the optimal Lagrange multipliers (μ^*, λ^*) are calculated at each iteration k, and that these estimates are asymptotically exact. It is known (see, for example, Facchinei, Fischer, and Kanzow [5]) that an asymptotically exact estimate $\mathcal{W}_k$ of $\mathcal{A}^*$ is available, given that $(z^k, \mu^k, \lambda^k) \to (z^*, \mu^*, \lambda^*)$, under weaker conditions than assumed here. On the other hand, it is also known that given an asymptotically exact $\mathcal{W}_k$, we can use a least-squares procedure to compute an asymptotically exact estimate (μ^k, λ^k) of (μ^*, λ^*). However, the *simultaneous* estimation of $\mathcal{W}_k$ and (μ^k, λ^k) is less straightforward. We anticipate, however, that a procedure that works well in practice would be relatively easy to implement, especially under the LICQ and strict complementarity assumptions. Given an initial guess of $\mathcal{W}_k$, such a procedure would alternate between a least-squares estimate of (μ^k, λ^k) and an active-set identification procedure like those in [5], until the estimate of $\mathcal{W}_k$ settles down. We note that the multipliers for the linearized constraints in the subproblem (1.2), (1.3) (denoted in the analysis below by $\bar{\mu}^k$ and $\bar{\lambda}^k$) do not necessarily satisfy the asymptotic exactness condition, unless it is known a priori that the trust region is inactive for all k sufficiently large. Fletcher and Sainz de la Maza [6] have analyzed the behavior of these multipliers in the context of a sequential linear programming algorithm and show that, under certain assumptions, (μ^*, λ^*) is a limit point of the sequence $\{(\bar{\mu}^k, \bar{\lambda}^k)\}$.

We summarize the algorithmic assumptions as follows.

ASSUMPTION 5.

(a) $z^k \to z^*$.

(b) $\mathcal{W}_k = \mathcal{A}^*$ *for all* k *sufficiently large, where* $\mathcal{W}_k$ *is the estimate of the optimal active set.*

(c) $(\mu^k, \lambda^k) \to (\mu^*, \lambda^*)$.

(d) In addition to (1.4), the perturbed step $\widetilde{\Delta z}^k$ *satisfies*

$$d_i(z^k + \widetilde{\Delta z}^k) = d_i(z^k) + \nabla d_i(z^k)^T \Delta z^k, \quad \forall i \in \mathcal{W}_k \tag{4.2}$$

and

$$\|\Delta z^k - \widetilde{\Delta z}^k\| = O(\|\Delta z\|^2). \tag{4.3}$$

The condition (4.2) is an explicit form of "second-order correction," a family of techniques that are often needed to ensure fast local convergence of SQP algorithms. Note that (4.3) is a stronger form of the asymptotic exactness condition (1.5). Note too that Assumption 5(a) implies that $\Delta z^k \to 0$, since

$$\|\Delta z^k\| = \|z^{k+1} - z^k\| \le \|z^{k+1} - z^*\| + \|z^k - z^*\| \to 0. \tag{4.4}$$

Similarly, we have because of Assumption 5(d) that

$$\widetilde{\Delta z}^k = \Delta z^k + O(\|\Delta z^k\|^2) \to 0. \tag{4.5}$$

We start with a technical result to show that the various requirements on the perturbed step $\widetilde{\Delta z}^k$ are consistent.

LEMMA 4.1. *Suppose that the functions* f*,* c*, and* d *are twice continuously differentiable in a neighborhood of* z^* *and that Assumption 4 and Assumptions 5(a),(b) hold. Then for all sufficiently large* k*, there exists* $\widetilde{\Delta z}^k$ *such that (1.4), (4.2), and (4.3) are all satisfied.*

Proof. Assume first that k is chosen large enough that $\mathcal{W}_k = \mathcal{A}^*$. We prove the result constructively, generating $\widetilde{\Delta z}^k$ as the solution of the following problem:

$$\min_w \tfrac{1}{2}\|w - \Delta z^k\|_2^2 \ \text{ s.t.} \tag{4.6a}$$

$$c(z^k + w) = 0, \tag{4.6b}$$

$$d_i(z^k + w) = d_i(z^k) + \nabla d_i(z^k)^T \Delta z^k, \quad \forall i \in \mathcal{W}_k. \tag{4.6c}$$

When the right hand sides of (4.6b), (4.6c) are replaced by $c(z^k + \Delta z^k)$ and $d_i(z^k + \Delta z^k)$, respectively, the solution is $w = \Delta z^k$. These modified right-hand sides represent only an $O(\|\Delta z^k\|^2)$ perturbation of the right-hand sides in (4.6b), (4.6c), because from (1.2b) we have

$$\begin{aligned}&c_i(z^k + \Delta z^k)\\ &= c_i(z^k) + \nabla c_i(z^k)^T \Delta z^k + \tfrac{1}{2}(\Delta z^k)^T \nabla^2 c_i(z^k + \theta_i^c \Delta z^k)^T \Delta z^k\\ &= \tfrac{1}{2}(\Delta z^k)^T \nabla^2 c_i(z^k + \theta_i^c \Delta z^k)^T \Delta z^k,\end{aligned}$$

for all $i = 1, 2, \ldots, m$ and some $\theta_i^c \in (0,1)$; and

$$d_i(z^k + \Delta z^k) - \left[d_i(z^k) + \nabla d_i(z^k)^T \Delta z^k\right] = \tfrac{1}{2}(\Delta z^k)^T \nabla^2 d_i(z^k + \theta_i^d \Delta z^k)^T \Delta z^k,$$

for all $i \in \mathcal{W}_k = \mathcal{A}^*$ and some $\theta_i^d \in (0,1)$. Note that the Jacobian of the constraints (4.6b), (4.6c) has full row rank at $z^k + \Delta z^k$, because of Assumptions 4(a) 5(a), and (4.4). Hence, the Jacobian matrix of the KKT conditions for the problem (4.6) (which is a "square" system of nonlinear equations) is nonsingular at $z^k + \Delta z^k$, and a straightforward application of the implicit function theorem to this system yields that the solution $w = \widetilde{\Delta z}^k$ of (4.6) satisfies the property (4.3) for all k sufficiently large. We note too that because of (1.2b), we have

$$d_i(z^k + \widetilde{\Delta z}^k) = d_i(z^k) + \nabla d_i(z^k)^T \Delta z^k \le 0, \quad \forall i \in \mathcal{A}^*,$$

while for $i \notin \mathcal{A}^*$ we have from $d_i(z^*) < 0$, Assumption 5(a), and (4.5) that $d_i(z^k + \Delta z^k) \le (1/2) d_i(z^*) < 0$ for all k sufficiently large. For the equality constraints we have immediately from (4.6b) that $c(z^k + \widetilde{\Delta z}^k) = 0$. Hence $z^k + \widetilde{\Delta z}^k \in \mathcal{F}$, so condition (1.4) is also satisfied. □

We assume the Hessian matrix H_k in the subproblem (1.2), (1.3) at $z = z^k$ is the Hessian of the Lagrangian $\mathcal{L}$ evaluated at this point, with appropriate estimates of the multipliers μ^k and λ^k; that is,

$$H_k = \nabla^2_{zz} \mathcal{L}(z^k, \mu^k, \lambda^k) = \nabla^2 f(z^k) + \sum_{i=1}^m \mu_i^k \nabla^2 c_i(z^k) + \sum_{i=1}^r \lambda_i^k \nabla^2 d_i(z^k). \tag{4.7}$$

We show now that with this choice of H_k, the ratio ρ_k is close to 1 for all k sufficiently large, and hence that $\Delta_{k+1} \ge \Delta_k$ for all k sufficiently large. We prove the result specifically for the Euclidean-norm trust region; a minor generalization yields the proof for general $p \in [1, \infty]$.

Lemma 4.2. *Suppose that the functions f, c, and d are twice continuously differentiable in a neighborhood of z^*, that $p = 2$ in (1.3), that Assumptions 1, 4, and 5 hold, and that H_k is defined by (4.7). Then we have $\rho_k \to 1$, where ρ_k is defined by (2.1).*

Proof. From (2.1) we have

$$\rho_k = 1 + \frac{f(z^k) - f(z^k + \widetilde{\Delta z}^k) + m_k(\Delta z^k)}{-m_k(\Delta z^k)}. \tag{4.8}$$

we prove the result by showing that the numerator of the final term in this expression is $o(\|\Delta z^k\|^2)$, while the denominator is $\Omega(\|\Delta z^k\|^2)$.

We assume initially that k is large enough that $\mathcal{W}_k = \mathcal{A}^*$. We work first with the numerator in (4.8). By elementary manipulation, using Taylor's theorem and the definition of $m_k(\cdot)$, we have for some $\theta^f \in (0,1)$ that

$$
\begin{aligned}
& f(z^k) - f(z^k + \widetilde{\Delta z}^k) + m_k(\Delta z^k) \\
&= -\nabla f(z^k)^T \widetilde{\Delta z}^k - \tfrac{1}{2}(\widetilde{\Delta z}^k)^T \nabla^2 f(z^k + \theta^f \widetilde{\Delta z}^k)\widetilde{\Delta z}^k + \nabla f(z^k)^T \Delta z^k + \tfrac{1}{2}(\Delta z^k)^T H_k \Delta z^k \\
&= \left(\nabla f(z^k) + H_k \widetilde{\Delta z}^k\right)^T (\Delta z^k - \widetilde{\Delta z}^k) + \tfrac{1}{2}(\widetilde{\Delta z}^k)^T \left(H_k - \nabla^2 f(z^k + \theta^f \widetilde{\Delta z}^k)\right) \widetilde{\Delta z}^k \\
&\qquad + O(\|\Delta z^k - \widetilde{\Delta z}^k\|^2) \\
&= \nabla f(z^k)^T (\Delta z^k - \widetilde{\Delta z}^k) + \tfrac{1}{2}(\widetilde{\Delta z}^k)^T \left(H_k - \nabla^2 f(z^k)\right) \widetilde{\Delta z}^k + O(\|\Delta z^k\|^3),
\end{aligned}
\tag{4.9}
$$

where we used (4.3) and boundedness of H_k to derive the final equality. Now from (1.2b), we have for all $i = 1, 2, \ldots, m$ and some $\theta_i^c \in (0,1)$ that

$$
\begin{aligned}
0 &= c_i(z^k + \widetilde{\Delta z}^k) \\
&= c_i(z^k) + \nabla c_i(z^k)^T \widetilde{\Delta z}^k + \tfrac{1}{2}(\widetilde{\Delta z}^k)^T \nabla^2 c_i(z^k + \theta_i^c \widetilde{\Delta z}^k)\widetilde{\Delta z}^k \\
&= \nabla c_i(z^k)^T (\widetilde{\Delta z}^k - \Delta z^k) + \tfrac{1}{2}(\widetilde{\Delta z}^k)^T \nabla^2 c_i(z^k) \widetilde{\Delta z}^k + O(\|\Delta z^k\|^3).
\end{aligned}
\tag{4.10}
$$

From (4.2), we have for all $i \in \mathcal{A}^*$ that

$$
\begin{aligned}
0 &= d_i(z^k + \widetilde{\Delta z}^k) - d_i(z^k) - \nabla d_i(z^k)^T \Delta z^k \\
&= \nabla d_i(z^k)^T (\widetilde{\Delta z}^k - \Delta z^k) + \tfrac{1}{2}(\widetilde{\Delta z}^k)^T \nabla^2 d_i(z^k) \widetilde{\Delta z}^k + O(\|\Delta z^k\|^3).
\end{aligned}
\tag{4.11}
$$

For $i \notin \mathcal{A}^*$, we have from $\lambda_i^k \to \lambda_i^* = 0$ and (4.3) that

$$
\lambda_i^k \nabla d_i(z^k)^T (\widetilde{\Delta z}^k - \Delta z^k) + \tfrac{1}{2}\lambda_i^k (\widetilde{\Delta z}^k)^T \nabla^2 d_i(z^k) \widetilde{\Delta z}^k = o(\|\Delta z^k\|^2).
\tag{4.12}
$$

We now multiply equations (4.10) and (4.11) by their corresponding Lagrange multipliers (μ_i^k and λ_i^k, respectively), and subtract them together with (4.12) from (4.9), to obtain

$$
\begin{aligned}
& f(z^k) - f(z^k + \widetilde{\Delta z}^k) + m_k(\Delta z^k) \\
&= \left(\nabla f(z^k) + \nabla c(z^k)\mu^k + \nabla d(z^k)\lambda^k\right)^T (\Delta z^k - \widetilde{\Delta z}^k) \\
&\qquad + \tfrac{1}{2}(\widetilde{\Delta z}^k)^T \left[H_k - \nabla^2 f(z^k) - \sum_{i=1}^m \mu_i^k \nabla^2 c_i(z^k) - \sum_{i=1}^r \lambda_i^k \nabla^2 d_i(z^k)\right] \widetilde{\Delta z}^k \\
&\qquad + o(\|\widetilde{\Delta z}^k\|^2) \\
&= O(\|(z^k, \mu^k, \lambda^k) - (z^*, \mu^*, \lambda^*)\|)\|\Delta z^k\|^2 + o(\|\widetilde{\Delta z}^k\|^2) \\
&= o(\|\widetilde{\Delta z}^k\|^2),
\end{aligned}
\tag{4.13}
$$

where we used the KKT condition (1.7a) at $(z, \mu, \lambda) = (z^*, \mu^*, \lambda^*)$ and the definition (4.7) to derive the second equality, and Assumption 5(a),(c) to derive the third equality. Hence we have shown that the numerator of the last term in (4.8) is $o(\|\widetilde{\Delta z}^k\|^2)$.

In the remainder of the proof we use the following shorthand notation for the Hessian of the Lagrangian:

$$(\nabla^2_{zz}\mathcal{L})_k = \nabla^2_{zz}\mathcal{L}(z^k,\mu^k,\lambda^k); \quad (\nabla^2_{zz}\mathcal{L})_* = \nabla^2_{zz}\mathcal{L}(z^*,\mu^*,\lambda^*). \tag{4.14a}$$

Given $p=2$ in (1.3), we see that the KKT conditions for Δz^k to be a solution of (1.2), (1.3) at $z = z^k$ are that there exist multipliers $\bar{\mu}^k$ and $\bar{\lambda}^k$ such that

$$\nabla f(z^k) + (\nabla^2_{zz}\mathcal{L})_k \Delta z^k + \nabla c(z^k)\bar{\mu}^k + \nabla d(z^k)\bar{\lambda}^k + \gamma_k D_k^T D_k \Delta z^k = 0, \tag{4.15a}$$

$$c(z^k) + \nabla c(z^k)^T \Delta z^k = 0, \tag{4.15b}$$

$$d(z^k) + \nabla d(z^k)^T \Delta z^k \le 0 \perp \bar{\lambda}^k \ge 0, \tag{4.15c}$$

$$\|D_k \Delta z^k\|_2^2 - \Delta_k^2 \le 0 \perp \gamma_k \ge 0, \tag{4.15d}$$

where γ_k is the Lagrange multiplier for the trust-region constraint $\|D_k\Delta z^k\|_2^2 \le \Delta_k^2$. From (4.15c), we have

$$(\bar{\lambda}^k)^T \nabla d(z^k)^T \Delta z^k = -(\bar{\lambda}^k)^T d(z^k). \tag{4.16}$$

We turn now to the denominator in (4.8), and show that it has size $\Omega(\|\Delta z^k\|^2)$ for all k sufficiently large. From the definition of $m_k(\cdot)$, (4.7), and (4.14a), we have

$$\begin{aligned} -m_k(\Delta z^k) &= -\nabla f(z^k)^T \Delta z^k - \tfrac{1}{2}(\Delta z^k)^T (\nabla^2_{zz}\mathcal{L})_k \Delta z^k \\ &= -(\Delta z^k)^T \left(\nabla f(z^k) + (\nabla^2_{zz}\mathcal{L})_k \Delta z^k\right) + \tfrac{1}{2}(\Delta z^k)^T (\nabla^2_{zz}\mathcal{L})_k \Delta z^k. \end{aligned}$$

By substituting from (4.15a), then using (4.15b) and (4.16), we obtain

$$\begin{aligned} -m_k(\Delta z^k) &= (\Delta z^k)^T \left(\nabla c(z^k)\bar{\mu}^k + \nabla d(z^k)\bar{\lambda}^k + \gamma_k D_k^T D_k \Delta z^k\right) \\ &\quad + \tfrac{1}{2}(\Delta z^k)^T (\nabla^2_{zz}\mathcal{L})_k \Delta z^k \\ &= -d(z^k)^T \bar{\lambda}^k + \gamma_k \|D_k \Delta z^k\|_2^2 + \tfrac{1}{2}(\Delta z^k)^T (\nabla^2_{zz}\mathcal{L})_k \Delta z^k. \end{aligned}$$

By using Assumption 1, we then obtain

$$-m_k(\Delta z^k) \ge -d(z^k)^T \bar{\lambda}^k + \gamma_k \delta^{-2} \|\Delta z^k\|_2^2 + \tfrac{1}{2}(\Delta z^k)^T (\nabla^2_{zz}\mathcal{L})_k \Delta z^k. \tag{4.17}$$

We now define the constant $\bar{\gamma}$ as follows:

$$\bar{\gamma} \stackrel{\text{def}}{=} \max\left(2\delta^2 \left\|(\nabla^2_{zz}\mathcal{L})_*\right\|_2, 1\right). \tag{4.18}$$

Then for all k sufficiently large, we have

$$\left\|(\nabla^2_{zz}\mathcal{L})_k\right\|_2 \le 2\left\|(\nabla^2_{zz}\mathcal{L})_*\right\|_2 \le \delta^{-2}\bar{\gamma}. \tag{4.19}$$

We derive the estimate for $-m_k(\Delta z^k)$ from (4.17) by considering two cases. In the first case, we assume that $\gamma_k \ge \bar{\gamma}$. We then have from (4.17), using $\bar{\lambda}^k \ge 0$ and $d(z^k) \le 0$, that the following bound holds for k large enough to satisfy (4.19):

$$\begin{aligned} -m_k(\Delta z^k) &\ge \gamma_k \delta^{-2}\|\Delta z^k\|_2^2 + \tfrac{1}{2}(\Delta z^k)^T (\nabla^2_{zz}\mathcal{L})_k \Delta z^k \\ &\ge \bar{\gamma}\delta^{-2}\|\Delta z^k\|_2^2 - \tfrac{1}{2}\|\Delta z^k\|^2 \left\|(\nabla^2_{zz}\mathcal{L})_k\right\|_2 \\ &\ge \tfrac{1}{2}\bar{\gamma}\delta^{-2}\|\Delta z^k\|_2^2, \end{aligned} \tag{4.20}$$

so we see that the estimate $-m_k(\Delta z^k) = \Omega(\|\Delta z^k\|^2)$ is satisfied in this case .

In the second case of $\gamma_k \le \bar\gamma$, a little more analysis is needed. We show first that

$$\lim_{k\to\infty, \gamma_k \le \bar\gamma} (\bar\mu^k, \bar\lambda^k) = (\mu^*, \lambda^*).$$

For $i \notin \mathcal{W}_k = \mathcal{A}^*$, we have from $d_i(z^*) < 0$, Assumption 5(a), and (4.4) that $d_i(z^k) + \nabla d_i(z^k)^T \Delta z^k < 0$. Hence, from (4.15c), we have $\bar\lambda_i^k = 0$ for all $i \notin \mathcal{A}^*$. By rearranging (4.15a), we therefore have

$$\nabla c(z^k)\bar\mu^k + \nabla d_{\mathcal{A}^*}(z^k)\bar\lambda^k_{\mathcal{A}^*} = -\nabla f(z^k) - (\nabla^2_{zz}\mathcal{L})_k \Delta z^k - \gamma_k D_k^T D_k \Delta z^k.$$

By comparing this expression with the KKT condition for z^*, namely,

$$\nabla c(z^*)\mu^* + \nabla d_{\mathcal{A}^*}(z^*)\lambda^*_{\mathcal{A}^*} = -\nabla f(z^*),$$

and using the LICQ (Assumption 4(a)) and $\gamma_k \le \bar\gamma$, we obtain

$$\|(\bar\mu^k, \bar\lambda^k_{\mathcal{A}^*}) - (\mu^*, \lambda^*_{\mathcal{A}^*})\| = O(\|z^k - z^*\|) + (1+\bar\gamma) O(\|\Delta z^k\|) \to 0.$$

Hence, by strict complementarity (Assumption 4(b)), we can identify a constant $\bar\lambda_{\min} > 0$ such that

$$\bar\lambda_i^k \ge \bar\lambda_{\min}, \quad \forall i \in \mathcal{A}^*, \quad \forall k \, \text{sufficiently large with } \gamma_k \le \bar\gamma. \tag{4.21}$$

Therefore, by the complementarity condition (4.15c), we have that

$$\nabla d_{\mathcal{A}^*}(z^k)^T \Delta z^k = -d_{\mathcal{A}^*}(z^k).$$

Using this expression together with (4.15b), we deduce that

$$\begin{aligned}
\begin{bmatrix} \nabla c(z^*)^T \\ \nabla d_{\mathcal{A}^*}(z^*)^T \end{bmatrix} \Delta z^k &= \begin{bmatrix} (\nabla c(z^*) - \nabla c(z^k))^T \Delta z^k \\ d_{\mathcal{A}^*}(z^k) + (\nabla d_{\mathcal{A}^*}(z^*) - \nabla d_{\mathcal{A}^*}(z^k))^T \Delta z^k \end{bmatrix} \\
&= O(\|d_{\mathcal{A}^*}(z^k)\|) + O(\|z^k - z^*\| \|\Delta z^k\|).
\end{aligned} \tag{4.22}$$

By full row rank of the coefficient matrix on the left-hand side of (4.22), we have that there exists a vector s^k with

$$\begin{bmatrix} \nabla c(z^*)^T \\ \nabla d_{\mathcal{A}^*}(z^*)^T \end{bmatrix} s^k = \begin{bmatrix} \nabla c(z^*)^T \\ \nabla d_{\mathcal{A}^*}(z^*)^T \end{bmatrix} \Delta z^k, \tag{4.23a}$$

$$\|s^k\| = O(\|d_{\mathcal{A}^*}(z^k)\|) + O(\|z^k - z^*\| \|\Delta z^k\|). \tag{4.23b}$$

Since the vector $\Delta z^k - s^k$ satisfies the conditions on v in the second-order sufficient conditions (Assumptions 4(c)), we have

$$(\Delta z^k - s^k)^T (\nabla^2_{zz}\mathcal{L})_* (\Delta z^k - s^k) \ge \alpha \|\Delta z^k - s^k\|_2^2,$$

so that for k sufficiently large, we have by Assumption 5(a),(c) that

$$(\Delta z^k - s^k)^T (\nabla^2_{zz}\mathcal{L})_k (\Delta z^k - s^k) \ge \tfrac{1}{2}\alpha \|\Delta z^k - s^k\|_2^2.$$

By using this inequality together with (4.23b) and Assumption 5(a), we obtain

$$\begin{aligned}
&(\Delta z^k)^T (\nabla^2_{zz}\mathcal{L})_k \Delta z^k \\
&= (\Delta z^k - s^k)^T (\nabla^2_{zz}\mathcal{L})_k (\Delta z^k - s^k) + O(\|s^k\| \|\Delta z^k - s^k\|) + O(\|s^k\|^2) \\
&\ge \tfrac{1}{2}\alpha \|\Delta z^k - s^k\|_2^2 + O(\|\Delta z^k\| \|s^k\|) + O(\|s^k\|^2) \\
&= \tfrac{1}{2}\alpha \|\Delta z^k\|_2^2 + O(\|\Delta z^k\| \|s^k\|) + O(\|s^k\|^2) \\
&= \tfrac{1}{2}\alpha \|\Delta z^k\|_2^2 + O(\|d_{\mathcal{A}^*}(z^k)\| \|\Delta z^k\|) + O(\|d_{\mathcal{A}^*}(z^k)\|^2) + o(\|\Delta z^k\|^2) \\
&\ge \tfrac{1}{4}\alpha \|\Delta z^k\|_2^2 + O(\|d_{\mathcal{A}^*}(z^k)\| \|\Delta z^k\|) + O(\|d_{\mathcal{A}^*}(z^k)\|^2),
\end{aligned} \tag{4.24}$$

for all k sufficiently large with $\gamma_k \le \bar{\gamma}$. Because of (4.21), we have

$$-(\bar{\lambda}^k)^T d(z^k) \ge \sum_{i \in \mathcal{A}^*} \bar{\lambda}_i^k (-d_i(z^k)) \ge \bar{\lambda}_{\min} \| d_{\mathcal{A}^*}(z^k) \|_1. \tag{4.25}$$

By substituting (4.24) and (4.25) into (4.17), and dropping the second term on the right-hand side of (4.17) (which is positive in any case), we obtain

$$\begin{aligned}
-m_k(\Delta z^k) &\ge \bar{\lambda}_{\min} \| d_{\mathcal{A}^*}(z^k) \|_1 + \tfrac{1}{2} (\Delta z^k)^T (\nabla_{zz}^2 \mathcal{L})_k \Delta z^k \\
&\ge \bar{\lambda}_{\min} \| d_{\mathcal{A}^*}(z^k) \|_1 + (1/8)\alpha \| \Delta z^k \|_2^2 + O(\| d_{\mathcal{A}^*}(z^k) \| \| \Delta z^k \|) + +O(\| d_{\mathcal{A}^*}(z^k) \|^2) \\
&\ge (1/8)\alpha \| \Delta z^k \|_2^2,
\end{aligned} \tag{4.26}$$

since the term $\bar{\lambda}_{\min} \| d_{\mathcal{A}^*}(z^k) \|_1$ eventually dominates the remainder terms.

We conclude from (4.20) and (4.26) that for all k sufficiently large, we have $-m_k(\Delta z^k) = \Omega(\| \Delta z^k \|^2)$. By combining this estimate with (4.13) and (4.8), we obtain that

$$\rho_k = 1 + \frac{o(\| D z^k \|^2}{\Omega(\| \Delta z^k \|^2)} \to 1,$$

as claimed. □

We are now ready for the quadratic convergence result.

THEOREM 4.3. *Suppose that the functions f, c, and d are twice continuously differentiable in a neighborhood of z^*, that $p = 2$ in (1.3), that Assumptions 1, 4, and 5 hold, and that H_k is defined by (4.7). Then Algorithm FP-SQP converges Q-quadratically to z^*.*

Proof. because of Lemma 4.2, we have that Δ_k is bounded below, while because of (4.4) the standard SQP step satisfies $\Delta z^k \to 0$. Hence from (4.5), the algorithm takes the step $\widetilde{\Delta z}^k$ at iteration k, where $\| \widetilde{\Delta z}^k \|_2 < \Delta_k$ for all k sufficiently large. From standard local convergence theory for SQP in the neighborhood of a point z^* satisfying the properties in Assumption 4, we have that

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

and therefore

$$\| \Delta z^k \| = \| z^k - z^* \| + O(\| z^k - z^* \|^2).$$

By using these results together with (4.3), we have

$$\begin{aligned}
\| z^{k+1} - z^* \| &= \| z^k + \widetilde{\Delta z}^k - z^* \| \\
&\le \| \widetilde{\Delta z}^k - \Delta z^k \| + \| z^k + \Delta z^k - z^* \| \\
&= O(\| \Delta z^k \|^2) + O(\| z^k - z^* \|^2) \\
&= O(\| z^k - z^* \|^2),
\end{aligned}$$

as claimed. □

5. Conclusions. We have described a simple feasibility perturbed trust-region SQP algorithm for nonlinear programming with good global and local convergence properties. As discussed above, we believe that the feasibility perturbation often can be carried out efficiently when the constraints are separable or otherwise structured.

The companion report [14] describes application of the algorithm to optimal control problems with constraints on the inputs (controls).

We believe that a result concerning global convergence to points satisfying second-order necessary conditions can also be proved. When the assumptions used in Section 3 are satisfied, z^* is a stationary limit point of the sequence $\{z^k\}$ at which LICQ and strict complementarity are satisfied, asymptotically exact estimates of (μ^k, λ^k) and $\mathcal{W}_k$ are available on the convergent subsequence $\mathcal{K}$, H_k is chosen as in (4.7), and Assumption 5(d) is satisfied, then the following second-order necessary condition holds at z^*:

$$v^T \nabla^2_{zz} \mathcal{L}(z^*, \mu^*, \lambda^*) v \geq 0, \ \ \forall v \text{ s.t. } \nabla c(z^*)^T v = 0, \ \ \nabla d_{\mathcal{A}^*}(z^*)^T v = 0,$$

We omit a full analysis, which uses many of the same techniques as in Sections 3 and 4.

Acknowledgments. We thank Paul Tseng for suggesting Assumption 3. We gratefully acknowledge the financial support of the industrial members of the Texas-Wisconsin Modeling and Control Consortium. In addition, research of the second author was supported in part by NSF grants #CTS-0105360, while research of the first author was supported in part by NSF grants #MTS-0086559, #ACI-0196485, and #EIA-0127857.

Appendix A. Value Function of a Parametrized Linear Program.

Here we prove Lemma 3.1.

Proof. Note first that $C(z, 1) \geq 0$ for any feasible z, since $w = 0$ is feasible for (3.1).

We have $C(z, 1) = 0$ if and only if $w = 0$ is a solution of the problem (3.1). The bound $w^T w \leq 1$ is inactive at $w = 0$, and the optimality conditions for (3.1) are then identical to the KKT conditions (1.7) for (1.1). Hence, $C(z, 1) = 0$ if and only if z satisfies the KKT conditions.

Suppose now that $\bar{z} \in \mathcal{F}$ satisfies MFCQ (1.9) but not the KKT conditions (1.7). Suppose for contradiction that there exists a sequence $\{z^l\}$ with $z^l \to \bar{z}$, $z_l \in \mathcal{F}$ such that

$$0 \leq C(z^l, 1) \leq l^{-1}, \ \ l = 1, 2, 3, \ldots.$$

The optimality conditions for the solution w^l of (3.1) at $z = z^l$ are that there exist multipliers $\mu^l \in \mathsf{R}^m$, $\lambda^l \in \mathsf{R}^r$, and $\beta^l \in \mathsf{R}$ such that:

$$\nabla f(z^l) + \nabla c(z^l)\mu^l + \nabla d(z^l)\lambda^l + 2\beta^l w^l = 0, \tag{A.1a}$$
$$c(z^l) + \nabla c(z^l)^T w^l = 0, \tag{A.1b}$$
$$d(z^l) + \nabla d(z^l)^T w^l \leq 0 \perp \lambda^l \geq 0, \tag{A.1c}$$
$$(w^l)^T w^l - 1 \leq 0 \perp \beta^l \geq 0. \tag{A.1d}$$

From these relations, and using the fact that $z^l \in \mathcal{F}$, we have that

$$\begin{aligned} C(z^l, 1) &= -\nabla f(z^l)^T w^l \\ &= (w^l)^T \nabla c(z^l) \mu^l + (w^l)^T \nabla d(z^l) \lambda^l + 2\beta^l (w^l)^T w^l \\ &= -d(z^l)^T \lambda^l + 2\beta^l \geq 0 \end{aligned} \tag{A.2}$$

By taking limits as $l \to \infty$, and since both terms in (A.2) are nonnegative, we have that

$$\beta^l \to 0, \quad d(z^l)^T \lambda^l \to 0. \tag{A.3}$$

Consider first the case in which there is a subsequence $\mathcal{K}$ of multipliers from (A.1), that is, $\{\mu^l, \lambda^l\}_{l\in\mathcal{K}}$ is bounded. By compactness, and taking a further subsequence of $\mathcal{K}$ if necessary, we can identify $\bar{\mu}$ and $\bar{\lambda} \geq 0$ such that

$$(\mu^l, \lambda^l)_{l\in\mathcal{K}} \to (\bar{\mu}, \bar{\lambda}). \tag{A.4}$$

Then by taking limits in (A.1a), and using (A.3) and (1.6), we have that

$$\nabla_z \mathcal{L}(\bar{z}, \bar{\mu}, \bar{\lambda}) = 0, \;\; d(\bar{z})^T \bar{\lambda} = 0. \tag{A.5}$$

By using these relations together with feasibility of $\bar{z}$, we see that $\bar{z}$ is a KKT point, which is a contradiction.

Hence, the sequence $\{\mu^l, \lambda^l\}$ must have no convergent subsequence. By taking another subsequence $\mathcal{K}$, we can identify a vector $(\hat{\mu}, \hat{\lambda})$ with $\|(\hat{\mu}, \hat{\lambda})\|_2 = 1$ and $\hat{\lambda} > 0$ such that

$$\frac{(\mu^l, \lambda^l)}{\|(\mu^l, \lambda^l)\|_2} \;\to_{l\in\mathcal{K}}\; (\hat{\mu}, \hat{\lambda}).$$

By dividing both sides of (A.1a) by $\|(\mu^l, \lambda^l)\|_2$ and using (A.3), we obtain

$$\nabla c(\bar{z})\hat{\mu} + \nabla d(\bar{z})\hat{\lambda} = 0, \;\; d(\bar{z})^T \hat{\lambda} = 0, \;\; \hat{\lambda} \geq 0. \tag{A.6}$$

It is easy to show that (A.6) together with the MFCQ (1.9) implies that $(\hat{\mu}, \hat{\lambda}) = 0$, which contradicts $\|(\hat{\mu}, \hat{\lambda})\|_2 = 1$.

Therefore, we obtain a contradiction, so that no sequence $\{z^l\}$ with the claimed properties exists, and therefore $C(z, 1)$ is bounded away from zero in a neighborhood of $\bar{z}$. $\square$

REFERENCES

[1] D. P. Bertsekas. Projected Newton methods for optimization problems with simple constraints. *SIAM Journal on Control and Optimization*, 20:221–246, 1982.

[2] J. R. Birge, L. Qi, and Z. Wei. A varaint of the Topkis-Veinott method for solving inequality constrained optimization problems. *J. Appl. Math. Optim.*, 41:309–330, 2000.

[3] M.R. Celis, J.E. Dennis, and R.A. Tapia. A trust region strategy for nonlinear equality constrained optimization. In P.T. Boggs, R.H. Byrd, and R.B. Schnabel, editors, *Numerical Optimization*, pages 71–82. SIAM, 1985.

[4] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. MPS-SIAM Series on Optimization. SIAM, 2000.

[5] F. Facchinei, A. Fischer, and C. Kanzow. On the accurate identification of active constraints. *SIAM Journal on Optimization*, 9(1):14–32, 1998.

[6] R. Fletcher and E. Sainz de la Maza. Nonlinear programming and nonsmooth optimization by successive linear programming. *mprog*, 43:235–256, 1989.

[7] M. Heinkenschloss. Projected sequential quadratic programming methods. *SIAM Journal on Optimization*, 6(2):373–417, 1996.

[8] A. J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49:263–265, 1952.

[9] C. T. Lawrence and A. L. Tits. A computationally efficient feasible sequential quadratic programming algorithm. *siopt*, 11(4):1092–1118, 2001.

[10] S. Lucidi, M. Sciandrone, and P. Tseng. Objective-derivative-free methods for constrained optimization. Technical report, Department of Mathematics, University of Washington, November 1999.

[11] E. O. Omojokun. *Trust-region algorithms for optimization wiht nonlinear equality and inequality constraints*. PhD thesis, University of Colorado, Boulder, 1989.

[12] E. R. Panier and A. L. Tits. On combining feasibility, descent and superlinear convergence in inequality constrained optimization. *Mathematical Programming*, 59:261–276, 1993.

[13] S. M. Robinson. Stability theory for systems of inequalities, Part II: Differentiable nonlinear systems. *SIAM Journal on Numerical Analysis*, 13:497–513, 1976.

[14] M. J. Tenny, S. J. Wright, and J. B. Rawlings. Nonlinear model predictive control via feasibility-perturbed sequential quadratic programming. Optimization Technical Report 02-06, University of Wisconsin-Madison, Computer Sciences Departments, August 2002. Also Texas-Wisconsin Modeling and Control Consortium Report TWMCC-2002-02.

[15] S. J. Wright. Convergence of an inexact algorithm for composite nonsmooth optimization. *IMA Journal of Numerical Analysis*, 10:299–321, 1990.

[16] Y. Yuan. Conditions for convergence of a trust-region method for nonsmooth optimization. *Mathematical Programming*, 31:269–285, 1985.