SIAM J. OPTIM.
Vol. 24, No. 2, pp. 839–872

A SEQUENTIAL QUADRATIC OPTIMIZATION ALGORITHM WITH RAPID INFEASIBILITY DETECTION*

JAMES V. BURKE†, FRANK E. CURTIS‡, AND HAO WANG‡

Abstract. We present a sequential quadratic optimization (SQO) algorithm for nonlinear constrained optimization. The method attains all of the strong global and fast local convergence guarantees of classical SQO methods, but has the important additional feature that fast local convergence is guaranteed when the algorithm is employed to solve infeasible instances. A two-phase strategy, carefully constructed parameter updates, and a line search are employed to promote such convergence. The first phase subproblem determines the reduction that can be obtained in a local model of an infeasibility measure when the objective function is ignored. The second phase subproblem then seeks to minimize a local model of the objective while ensuring that the resulting search direction attains a reduction in the local model of the infeasibility measure that is proportional to that attained in the first phase. The subproblem formulations and parameter updates ensure that, near an optimal solution, the algorithm reduces to a classical SQO method for constrained optimization, and, near an infeasible stationary point, the algorithm reduces to a (perturbed) SQO method for minimizing constraint violation. Global and local convergence guarantees for the algorithm are proved under reasonable assumptions and numerical results are presented for a large set of test problems.

Key words. nonlinear optimization, sequential quadratic optimization, infeasibility detection, line search methods, exact penalization, superlinear convergence

AMS subject classifications. 49M05, 49M15, 49M37, 65K05, 65K10, 90C26, 90C30, 90C55

DOI. 10.1137/120880045

1. Introduction.

Sequential quadratic optimization (SQO) methods are known to be extremely efficient when applied to solve nonlinear constrained optimization problems [24, 34, 37]. Indeed, it has long been known [3, 4, 5] that with an appropriate globalization mechanism, SQO methods can guarantee global convergence from remote starting points to *feasible* optimal solutions, or to *infeasible* stationary points if the constraints are incompatible. One of the main additional strengths of SQO is that in the neighborhood of a solution point satisfying common assumptions and an appropriate constraint qualification, fast local convergence to *feasible* optimal solutions can be attained [35].

Despite these important and well-known properties of SQO methods, there is an important feature that many contemporary SQO methods lack, and it is for this reason that the algorithm in this paper has been designed, analyzed, and tested. Specifically, in addition to possessing the convergence guarantees mentioned in the previous paragraph, we have proved that the algorithm proposed in this paper yields *fast local convergence when applied to solve infeasible problem instances*. The rapid detection of infeasibility is an important issue in nonlinear optimization as many contemporary methods either fail or require an excessive number of iterations and/or function evaluations before being able to detect that a given problem instance is infeasible [32]. As a result, modelers are forced to wait an unacceptable amount of time, only to be

*Received by the editors June 6, 2012; accepted for publication (in revised form) March 17, 2014; published electronically June 5, 2014.

http://www.siam.org/journals/siopt/24-2/88004.html

†Department of Mathematics, University of Washington, Seattle, WA 98195 (jvburke@uw.edu).

‡Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA 18018 (frank.e.curtis@gmail.com, haw309@lehigh.edu). The second author was supported in part by National Science Foundation grant DMS-1016291.

told eventually (if at all) that model and/or data inconsistencies are present. Rapid infeasibility detection is also important in areas including branch-and-bound methods for nonlinear mixed-integer and parametric optimization, as algorithms for solving such problems often require the solution of a number of nonlinear subproblems. Slow infeasibility detection by such algorithms can create huge bottlenecks.

There are two main novel features of our algorithm. Most importantly, it is an algorithm that possesses global and local superlinear convergence guarantees for feasible *and* infeasible problems *without* having to resort to feasibility restoration. This feature, in that a single approach is employed for solving both feasible and infeasible problems, means that the algorithm avoids many of the inefficiencies that may arise when contemporary methods are employed to solve problems with incompatible constraints. The second novel feature of our algorithm is that it is able to attain these strong convergence properties with at most two quadratic optimization (QO) subproblem solves per iteration. This is in contrast to recently proposed methods that provide rapid infeasibility detection, but only at a much higher per-iteration cost.

In the following section, we compare and contrast our approach with recently proposed SQO methods, focusing on properties of those methods related to infeasibility detection. We then present our algorithm in section 3 and analyze its global and local convergence properties in section 4. Our numerical experiments in section 5 illustrate that an implementation of our algorithm yields solid results when applied to a large set of test problems. Finally, concluding remarks are the subject of section 6.

We remark at the outset that we analyze the local convergence properties of our algorithm under assumptions that are classically common for analyzing that of SQO methods. We explain that our algorithm can be backed by similarly strong convergence guarantees under more general settings (see our discussion in section 4.3), but have made the conscience decision to use these common assumptions to avoid unnecessary distractions in the analysis. Overall, the main purpose of this paper is to focus on the novelties of our algorithm—which include the unique formulations of our subproblems, our use of separate multiplier estimates for the optimization and a corresponding feasibility problem, and our unique combination of updates for the penalty parameter—which provide our algorithm with global and fast local convergence guarantees on both feasible and infeasible problem instances.

2. Literature review.

Our algorithm is designed to act as an SQO method for solving an optimization problem when the problem is feasible, and otherwise, it is designed to act as a perturbed SQO method [15] for a problem to minimize constraint violation. In this respect, our method has features in common with those in the class of penalty-SQO methods [19] where search directions are computed by minimizing a quadratic model of the objective combined with a penalty on the violation of the linearized constraints. In such algorithms, if the penalty parameter is driven to an extreme value, then the algorithm transitions to solely minimizing constraint violation. We believe that this approach is reasonable, though there are two main disadvantages of the manner in which penalty-SQO methods are often implemented. One disadvantage is that the penalty parameter takes on all of the responsibility for driving constraint violation minimization. This leads to a common criticism of penalty methods, which is that the performance of the algorithm is too highly dependent on the penalty parameter updating scheme. The second disadvantage is that, if the penalty parameter is not driven to its extreme value sufficiently quickly, then convergence, especially for infeasible problems, can be slow. These disadvantages motivate us to design a method that reduces to a classical SQO approach for feasible problems, where updates for the penalty parameter lead to rapid convergence in infeasible cases.

The immediate predecessor of our work is the penalty-SQO method proposed in [6]. In particular, the approach in [6] is also proved to yield fast local convergence guarantees for infeasible problems. That method does, however, have certain practical disadvantages. The most significant of these is that, particularly in infeasible cases, the method may require the solution of numerous QO subproblems per iteration. Indeed, near an infeasible stationary point, at least three QO subproblems must be solved. The first will reveal that for the current penalty parameter value it is not possible to compute a linearly feasible step, the second then gauges the progress toward linearized feasibility that can be made locally, and the third may produce the actual search direction. (In fact, if the conditions necessary for global convergence are not satisfied after the third QO subproblem solve, then even more QO subproblem solves are needed until the conditions are satisfied.) In contrast, the algorithm proposed in this paper solves *at most two* QO subproblems per iteration. It also relies less on the penalty parameter for driving constraint violation minimization, and involves separate multiplier estimates for the optimization and feasibility problems. This last feature of our algorithm—that of having two separate multiplier estimates—is quite unique for an optimization algorithm. However, we believe that it is natural as the optimization algorithm must implicitly decide which of two problems to solve: the given optimization problem or a problem to minimize constraint violation.

Our algorithm is a multiphase active-set method that has similarities with other such methods that have been proposed over the last few decades. For instance, the method in [6] borrows the idea, proposed in [11] and later incorporated into the line-search method in [10], of "steering" the algorithm with the penalty parameter. Consequently, that method at least suffers from the same disadvantages as the method in [6] when it comes to infeasibility detection. More commonly, multiphase SQO methods have taken the approach of solving a first-phase inequality-constrained subproblem—typically a linear optimization (LO) subproblem—to estimate an optimal active set, and then solving a second-phase equality-constrained subproblem to promote fast convergence; see, e.g., [8, 9, 13, 17, 18, 21]. A method of this type that solves two QO subproblems is that in [29], though again the second-phase subproblem in that method is equality-constrained as it only involves linearizations of constraints predicted to be active at an optimal solution. Our algorithm differs from these in that we do no active-set prediction, and rather solve up to two inequality-constrained subproblems. The methods in [22, 23] involve the solution of up to three subproblems per iteration: one to compute a "predictor" step, one to compute a "Cauchy" step, and one to compute an "accelerator" step. In fact, various subproblems are proposed for the "accelerator" step, including both equality-constrained and inequality-constrained alternatives. Our algorithm differs from these in that ours is a line search method, whereas they are trust region methods, and our first-phase subproblem computes a pure feasibility step rather than one influenced by a local model of the objective. This latter feature makes our method similar to those in [3, 4], though again our work is unique in that we ensure rapid infeasibility detection, which is not provided by any of the aforementioned methods besides that in [6]. Finally, we mention that multiphase strategies have also been employed in interior-point techniques; see, e.g., [7, 30].

3. Algorithm description.

We present our algorithm in the context of the generic nonlinear constrained optimization problem

$$\begin{array}{llll} \underset{x}{\text{minimize}} & (\underset{x}{\min}) & f(x) \\ \text{subject to} & (\text{s.t.}) & c^{\mathcal{E}}(x) = 0, \quad c^{\mathcal{I}}(x) \leq 0, \end{array} \tag{3.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $c^{\mathcal{E}} : \mathbb{R}^n \to \mathbb{R}^{m^{\mathcal{E}}}$, and $c^{\mathcal{I}} : \mathbb{R}^n \to \mathbb{R}^{m^{\mathcal{I}}}$ are twice-continuously differentiable. If the constraints of (3.1) are infeasible, then the algorithm is designed to return an infeasibility certificate in the form of a minimizer of the ℓ_1 infeasibility measure of the constraints; i.e., in such cases it is designed to solve

$$\min_x \; v(x), \quad \text{where} \;\; v(x) := \|c^{\mathcal{E}}(x)\|_1 + \|[c^{\mathcal{I}}(x)]^+\|_1. \tag{3.2}$$

Here, for a vector c, we define $[c]^+ := \max\{c, 0\}$ and, for future reference, define $[c]^- := \max\{-c, 0\}$ (both componentwise). The priority is to locate a stationary point for (3.1), but in all cases the algorithm is at least guaranteed to find a stationary point for (3.2), i.e., a stationary point for v. We say a point x is stationary for v if $0 \in \partial v(x)$, where $\partial v(x)$ is the Clarke subdifferential of v at x [2, 14] (see [3] for a complete review of first-order theory for potentially infeasible problems).

Each iteration of our algorithm consists of solving at most two QO subproblems, updating a penalty parameter, and performing a line search on an exact penalty function. In this regard, the method is broadly similar to that proposed in [3]; however, the algorithm contains numerous refinements included to ensure rapid local convergence in both feasible and infeasible cases. In this section, we present the details of each step of the algorithm. Of particular importance is the integration of our penalty parameter updates around the QO solves as this parameter is critical for driving fast local convergence for infeasible instances. A complete description of our algorithm is presented at the end of this section.

We begin by describing the conditions under which our algorithm terminates finitely. In short, the algorithm continues iterating unless a stationary point for problem (3.1) has been found. We define such stationary points according to first-order optimality conditions for problems (3.1) and (3.2), all of which can be presented by utilizing the Fritz John (FJ) function for (3.1), namely

$$\mathcal{F}(x, \rho, \lambda) := \rho f(x) + {\lambda^{\mathcal{E}}}^T c^{\mathcal{E}}(x) + {\lambda^{\mathcal{I}}}^T c^{\mathcal{I}}(x).$$

Here, $\rho \in \mathbb{R}$ is an objective multiplier and λ, with $\lambda^{\mathcal{E}} \in \mathbb{R}^{m^{\mathcal{E}}}$ and $\lambda^{\mathcal{I}} \in \mathbb{R}^{m^{\mathcal{I}}}$, are constraint multipliers. For future reference, we note that ρ also plays the role of the penalty parameter in the ℓ_1 exact penalty function

$$\phi(x, \rho) := \rho f(x) + v(x). \tag{3.3}$$

Our algorithm updates ρ and seeks stationary points for (3.1) through decreases in ϕ.

One possibility for finite termination is that the algorithm locates a first-order optimal point for (3.1). First-order optimality conditions for problem (3.1) are

$$\begin{aligned} \nabla_x \mathcal{F}(x, \rho, \lambda) = \rho \nabla f(x) + \nabla c^{\mathcal{E}}(x)\lambda^{\mathcal{E}} + \nabla c^{\mathcal{I}}(x)\lambda^{\mathcal{I}} &= 0, \\ c^{\mathcal{E}}(x) = 0, \quad c^{\mathcal{I}}(x) &\le 0, \\ \lambda^{\mathcal{I}} \ge 0, \quad \lambda^{\mathcal{I}} \cdot c^{\mathcal{I}}(x) &= 0. \end{aligned} \tag{3.4}$$

Here, $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is the gradient of f, $[\nabla c^{\mathcal{E}}]^T : \mathbb{R}^n \to \mathbb{R}^{m^{\mathcal{E}} \times n}$ is the Jacobian of $c^{\mathcal{E}}$ (and similarly for $[\nabla c^{\mathcal{I}}]^T$), and for vectors a and b we denote their componentwise (i.e., Hadamard or Schur) product by $a \cdot b$, a vector with entries $(a \cdot b)^i = a^i b^i$. If (x_*, ρ_*, λ_*) with $(\rho_*, \lambda_*) \neq 0$ satisfies (3.4), then we call (x_*, ρ_*, λ_*) stationary for (3.1); in particular, it is an FJ point [26]. Of particular interest are those FJ points with $\rho_* > 0$ as these correspond to Karush–Kuhn–Tucker (KKT) points for (3.1) [27, 28].

The other possibility for finite termination is that the algorithm locates a stationary point for (3.2) that is infeasible for problem (3.1). Hereafter, defining e as a vector of ones (whose size is determined by the context), first-order optimality conditions for problem (3.2) are

$$\begin{aligned} \nabla_x \mathcal{F}(x,0,\lambda) = \nabla c^{\mathcal{E}}(x)\lambda^{\mathcal{E}} + \nabla c^{\mathcal{I}}(x)\lambda^{\mathcal{I}} = 0, \\ -e \le \lambda^{\mathcal{E}} \le e, \quad 0 \le \lambda^{\mathcal{I}} \le e, \\ (e+\lambda^{\mathcal{E}}) \cdot [c^{\mathcal{E}}(x)]^- = 0, \quad (e-\lambda^{\mathcal{E}}) \cdot [c^{\mathcal{E}}(x)]^+ = 0, \\ \lambda^{\mathcal{I}} \cdot [c^{\mathcal{I}}(x)]^- = 0, \quad (e-\lambda^{\mathcal{I}}) \cdot [c^{\mathcal{I}}(x)]^+ = 0. \end{aligned} \tag{3.5}$$

If (x_*, λ_*) satisfies (3.5) and $v(x_*) > 0$, then we call (x_*, λ_*) stationary for (3.1); in particular, it is an infeasible stationary point. Despite the fact that such a point is infeasible for (3.1), it is deemed stationary as first-order information indicates that no further improvement in minimizing constraint violation locally is possible.

We now describe our technique for computing a search direction and multiplier estimates, which involves the solution of the QO subproblems (3.7) and (3.9) below. Once the details of these subproblems have been specified, we will describe an updating strategy for the penalty parameter that is integrated around these QO solves.

At the beginning of iteration k, the algorithm assumes an iterate of the form

$$(x_k, \rho_k, \overline{\lambda}_k, \widehat{\lambda}_k) \text{ with } \rho_k > 0, \ -e \le \overline{\lambda}_k^{\mathcal{E}} \le e, \ 0 \le \overline{\lambda}_k^{\mathcal{I}} \le e, \text{ and } \widehat{\lambda}_k^{\mathcal{I}} \ge 0. \tag{3.6}$$

As all stationary points for (3.1) are necessarily stationary for the constraint violation measure v, we initiate computation in iteration k by seeking to measure the possible improvement in minimizing the following linearized model of v at x_k:

$$l(d;x_k) := \|c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T d\|_1 + \|[c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T d]^+\|_1.$$

Specifically, defining $H(x,\rho,\lambda)$ as an approximation for the Hessian of $\mathcal{F}$ at (x,ρ,λ), we solve the following QO subproblem whose solution we denote as $(\overline{d}_k, \overline{r}_k, \overline{s}_k, \overline{t}_k)$:

$$\begin{aligned} \min_{(d,r,s,t)} \ & e^T(r+s) + e^T t + \tfrac{1}{2} d^T H(x_k, 0, \overline{\lambda}_k) d \\ \text{s.t. } & \begin{cases} c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T d = r - s, \\ c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T d \le t, \\ (r,s,t) \ge 0. \end{cases} \end{aligned} \tag{3.7}$$

As shown in Lemma 4.3 in section 4.1, this subproblem is always feasible and, if $H(x_k, 0, \overline{\lambda}_k)$ is positive definite, then the solution component $\overline{d}_k$ is unique. In addition, $\overline{d}_k$ yields a nonnegative reduction in $l(\cdot; x_k)$, i.e.,

$$\Delta l(\overline{d}_k; x_k) := l(0;x_k) - l(\overline{d}_k; x_k) \ge 0, \tag{3.8}$$

where equality holds if and only if x_k is stationary for v.

Upon solving subproblem (3.7) and setting

$$\overline{\lambda}_{k+1} \text{ with } -e \le \overline{\lambda}_{k+1}^{\mathcal{E}} \le e \text{ and } 0 \le \overline{\lambda}_{k+1}^{\mathcal{I}} \le e$$

as the optimal multipliers for the linearized equality and inequality constraints in (3.7), we check for termination at an infeasible stationary point. Specifically, we consider the constraint violation measure v and the following residual for (3.5):

$$\begin{aligned}\mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1}) := \max\{&\|\nabla_x \mathcal{F}(x_k, 0, \overline{\lambda}_{k+1})\|_\infty,\\ &\|(e - \lambda^{\mathcal{E}}_{k+1}) \cdot [c^{\mathcal{E}}(x_k)]^+\|_\infty, \|(e + \lambda^{\mathcal{E}}_{k+1}) \cdot [c^{\mathcal{E}}(x_k)]^-\|_\infty,\\ &\|(e - \lambda^{\mathcal{I}}_{k+1}) \cdot [c^{\mathcal{I}}(x_k)]^+\|_\infty, \|\lambda^{\mathcal{I}}_{k+1} \cdot [c^{\mathcal{I}}(x_k)]^-\|_\infty\}.\end{aligned}$$

If $\mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1}) = 0$ and $v(x_k) > 0$, then $(x_k, \overline{\lambda}_{k+1})$ is an infeasible stationary point. Otherwise, as shown in Lemma 4.3 in section 4.1, it follows that either $v(x_k) = 0$ or $\overline{d}_k$ is a direction of strict descent for v from x_k.

Having measured, in a particular sense, the possible improvement in minimizing constraint violation by solving the QO subproblem (3.7), the algorithm solves a second QO subproblem that seeks optimality. Denoting $\mathcal{E}_k$ and $\mathcal{I}_k$ as the sets of constraints that are *linearly satisfied* at the solution of (3.7) (i.e., that have $\overline{r}^i_k = \overline{s}^i_k = 0$ for $i \in \mathcal{E}$ or $\overline{t}^i_k = 0$ for $i \in \mathcal{I}$, respectively), we require that the computed direction maintains this set of linearly satisfied constraints. The other *linearly violated* constraints in $\mathcal{E}^c_k \cup \mathcal{I}^c_k$ (where $\mathcal{E}^c_k := \mathcal{E} \setminus \mathcal{E}_k$ and $\mathcal{I}^c_k := \mathcal{I} \setminus \mathcal{I}_k$) remain relaxed with slack variables whose values are penalized in the subproblem objective. The value of the penalty parameter employed at this stage is the value for ρ_k immediately prior to this second phase subproblem, which for future notational convenience we denote as $\widehat{\rho}_k$. Overall, we solve the following regularized QO subproblem whose solution we denote as $(\widehat{d}_k, \widehat{r}_k^{\mathcal{E}^c_k}, \widehat{s}_k^{\mathcal{E}^c_k}, \widehat{t}_k^{\mathcal{I}^c_k})$:

(3.9)
$$\begin{aligned}\min_{(d, r^{\mathcal{E}^c_k}, s^{\mathcal{E}^c_k}, t^{\mathcal{I}^c_k})} \quad & \widehat{\rho}_k \nabla f(x_k)^T d + e^T(r^{\mathcal{E}^c_k} + s^{\mathcal{E}^c_k}) + e^T t^{\mathcal{I}^c_k} + \tfrac{1}{2} d^T H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k) d\\ \text{s.t.} \quad & \begin{cases} c^{\mathcal{E}_k}(x_k) + \nabla c^{\mathcal{E}_k}(x_k)^T d = 0, & c^{\mathcal{E}^c_k}(x_k) + \nabla c^{\mathcal{E}^c_k}(x_k)^T d = r^{\mathcal{E}^c_k} - s^{\mathcal{E}^c_k},\\ c^{\mathcal{I}_k}(x_k) + \nabla c^{\mathcal{I}_k}(x_k)^T d \le 0, & c^{\mathcal{I}^c_k}(x_k) + \nabla c^{\mathcal{I}^c_k}(x_k)^T d \le t^{\mathcal{I}^c_k},\\ & (r^{\mathcal{E}^c_k}, s^{\mathcal{E}^c_k}, t^{\mathcal{I}^c_k}) \ge 0.\end{cases}\end{aligned}$$

Upon solving (3.9) and setting

$$\widehat{\lambda}_{k+1} \quad \text{with} \quad -e \le \widehat{\lambda}^{\mathcal{E}^c_k}_{k+1} \le e, \ 0 \le \widehat{\lambda}^{\mathcal{I}^c_k}_{k+1} \le e, \quad \text{and} \quad \widehat{\lambda}^{\mathcal{I}_k}_{k+1} \ge 0$$

as the optimal multipliers for the linearized equality and inequality constraints in (3.9), it is again appropriate to check for finite termination of the algorithm, this time with respect to the optimality conditions for (3.1). Given $(x_k, \rho_k, \widehat{\lambda}_{k+1})$ we consider the violation measure v and the following residual corresponding to (3.4):

$$\mathcal{R}_{opt}(x_k, \rho_k, \widehat{\lambda}_{k+1}) := \max\{\|\nabla_x \mathcal{F}(x_k, \rho_k, \widehat{\lambda}_{k+1})\|_\infty, \|\widehat{\lambda}^{\mathcal{I}}_{k+1} \cdot c^{\mathcal{I}}(x_k)\|_\infty\}.$$

We prove in Lemma 4.5 in section 4.1 that if the algorithm reaches this stage, then ρ_k is strictly positive. Thus, if $\mathcal{R}_{opt}(x_k, \rho_k, \widehat{\lambda}_{k+1}) = 0$ and $v(x_k) = 0$, then $(x_k, \rho_k, \widehat{\lambda}_{k+1})$ is a KKT point for (3.1).

If the algorithm has not terminated finitely due to this last check of optimality, then the search direction d_k is chosen as a convex combination of the directions obtained from subproblems (3.7) and (3.9). Given a constant $\beta \in (0, 1)$, our criterion for the selection of the weights in this combination is

$$\Delta l(d_k; x_k) \ge \beta \Delta l(\overline{d}_k; x_k). \tag{3.10}$$

For $w \in [0,1]$, the reduction in $l(\cdot;x_k)$ obtained by

$$d(w) := w\overline{d}_k + (1-w)\widehat{d}_k \tag{3.11}$$

is a piecewise linear function of w. If $\Delta l(\overline{d}_k;x_k) = 0$, then by the formulation of (3.9), we have $\Delta l(\widehat{d}_k;x_k) = 0$ and so (3.10) is satisfied by $w = 0$. Otherwise, if $\Delta l(\overline{d}_k;x_k) > 0$, then since $\Delta l(d(1);x_k) = \Delta l(\overline{d}_k;x_k) > \beta\Delta l(\overline{d}_k;x_k)$, there exists a threshold $\underline{w} \in [0,1)$ such that (3.10) holds for all $w \geq \underline{w}$. We define w_k as the smallest value in $[0,1)$ such that (3.10) holds and set the search direction as $d_k \leftarrow d(w_k)$.

We have presented our techniques for computing the primal search direction d_k as well as new multiplier estimates $\overline{\lambda}_{k+1}$ and $\widehat{\lambda}_{k+1}$. Within this discussion, we have accounted for finite termination of the algorithm and highlighted certain consequences of our step computation procedure (e.g., (3.8) and (3.10)) that will be critical in our convergence analysis. All that remains in the specification of our algorithm is our updating strategy for the penalty parameter and the conditions of our line search, which we now present. Note that with respect to ρ, an update is considered twice in a given iteration. The first time an update is considered is between the two QO subproblem solves, as it is at this point in the algorithm where the solution of (3.7) may trigger aggressive action toward infeasibility detection. The second time an update is considered is after the solution of (3.9). The update considered at that time is representative of typical contemporary updating strategies, used to ensure a well-defined line search and global convergence of the algorithm.

Prior to solving the second subproblem (3.9) (and before fixing $\widehat{\rho}_k$), we potentially modify ρ_k and $\widehat{\lambda}_k$ (computed in iteration $k-1$) to reduce the weight of the objective f and promote fast infeasibility detection. (Note that ρ_k and $\widehat{\lambda}_k$ will both influence the objective of (3.9).) If the current iterate is infeasible and the reduction in linearized feasibility obtained by $\overline{d}_k$ is small compared to the level of nonlinear infeasibility, then there is evidence that the algorithm is converging to an infeasible stationary point. In such cases, we consider modifying ρ_k before solving subproblem (3.9) so that the rest of the iteration places a higher emphasis on reducing constraint violation. A corresponding modification to $\widehat{\lambda}_k$ is also necessary to guarantee fast infeasibility detection (see Theorem 4.25). Defining constants $\theta \in (0,1)$, $\kappa_\rho > 0$, and $\kappa_\lambda > 0$, if

$$v(x_k) > 0 \quad \text{and} \quad \Delta l(\overline{d}_k;x_k) \leq \theta v(x_k), \tag{3.12}$$

then we set ρ_k by

$$\rho_k \leftarrow \min\{\rho_k, \kappa_\rho \mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1})^2\} \tag{3.13}$$

and modify $\widehat{\lambda}_k$ so that

$$\|\widehat{\lambda}_k - \overline{\lambda}_k\| \leq \kappa_\lambda \mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1})^2. \tag{3.14}$$

Otherwise, we maintain the current ρ_k and $\widehat{\lambda}_k$. For satisfying (3.14), a simple approach is to set $\widehat{\lambda}_k \leftarrow \alpha_\lambda\widehat{\lambda}_k + (1-\alpha_\lambda)\overline{\lambda}_k$ where α_λ is the largest value in $[0,1]$ such that (3.14) is satisfied. (This is the approach taken in our implementation described in section 5.)

Upon solving (3.9) and assuming the algorithm does not immediately terminate, we turn to a second update for ρ and our line search. For these purposes, we employ the ℓ_1 exact penalty function ϕ (recall (3.3)). At x_k, a linear model of $\phi(\cdot,\rho)$ is

$$m(d;x_k,\rho) := \rho(f(x_k) + \nabla f(x_k)^T d) + l(d;x_k)$$

and the corresponding reduction in this model yielded by the search direction d_k is

$$\Delta m(d_k; x_k, \rho) := m(0; x_k, \rho) - m(d_k; x_k, \rho) = -\rho \nabla f(x_k)^T d_k + \Delta l(d_k; x_k). \tag{3.15}$$

Prior to the line search, the new penalty parameter ρ_{k+1} is set so that its reciprocal is larger than the largest multiplier (derived from (3.9)) and that the reduction $\Delta m(d_k; x_k, \rho_{k+1})$ is at least proportional to $\Delta l(d_k; x_k)$. That is, we set ρ_{k+1} so that

$$\rho_{k+1} \|\widehat{\lambda}_{k+1}\|_\infty \le 1 \tag{3.16}$$

and, for a given constant $\epsilon \in (0, 1)$, we have

$$\Delta m(d_k; x_k, \rho_{k+1}) \ge \epsilon \Delta l(d_k; x_k). \tag{3.17}$$

Given constants $\delta \in (0, 1)$ and $\omega \in (0, 1)$, (3.16) and (3.17) can be achieved by setting

$$\rho_k \leftarrow \min\left\{\delta\rho_k, \frac{(1-\epsilon)}{\|\widehat{\lambda}_{k+1}\|_\infty}\right\} \quad \text{if} \quad \rho_k \|\widehat{\lambda}_{k+1}\|_\infty > 1 \tag{3.18}$$

followed by

$$\rho_k \leftarrow \begin{cases} \delta\rho_k & \text{if } \Delta m(d_k; x_k, \rho_k) \ge \epsilon \Delta l(d_k; x_k) \text{ and } w_k \ge \omega, \\ \min\{\delta\rho_k, \zeta_k\} & \text{if } \Delta m(d_k; x_k, \rho_k) < \epsilon \Delta l(d_k; x_k), \end{cases} \tag{3.19}$$

where

$$\zeta_k := \frac{(1-\epsilon)\Delta l(d_k; x_k)}{\nabla f(x_k)^T d_k + \frac{1}{2} d_k^T H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k) d_k},$$

and then setting $\rho_{k+1} \leftarrow \rho_k$. (Note that when $\Delta m(d_k; x_k, \rho_k) \ge \epsilon \Delta l(d_k; x_k)$ and $w_k \ge \omega$ in (3.19), it is not necessary to reduce ρ_k in order to satisfy (3.17); this condition would have been satisfied without invoking (3.19). However, since $w_k \ge \omega$, it follows from (3.11) that the search direction is dominated by the $\overline{d}_k$ component, which indicates that a reduction in the penalty parameter is appropriate. This feature of the update (3.19) is important in our convergence analysis. Moreover, we prove in Lemma 4.5(b) that even after invoking (3.19) in this case, the condition (3.17) will be satisfied.) Once ρ_{k+1} has been set in this manner, we perform a backtracking line search along d_k to determine α_k such that, for $\eta \in (0, 1)$, we have

$$\phi(x_k + \alpha_k d_k, \rho_{k+1}) - \phi(x_k, \rho_{k+1}) \le -\eta \alpha_k \Delta m(d_k; x_k, \rho_{k+1}). \tag{3.20}$$

Our proposed algorithm, hereafter nicknamed SQuID, is presented as Algorithm 1. We claim that the algorithmic framework of SQuID is globally convergent for choices of subproblems other than (3.7). For instance, a linear subproblem with a trust region would be appropriate for determining the best local improvement in linearized feasibility; see, e.g., [3, 4]. Under certain common assumptions, this choice should also allow for rapid local convergence for *feasible* problem instances. We present SQuID as solving two QO subproblems per iteration, however, as this choice also allows for rapid local convergence for *infeasible* instances, the main focus of this paper. In particular, in the neighborhood of an infeasible stationary point satisfying the assumptions of section 4.3, it can be seen that as $\rho_k \to 0$ and $\widehat{\lambda}_k \to \overline{\lambda}_k$, subproblem (3.9) produces

SQO-like steps for the minimization of constraint violation, thus causing rapid convergence toward stationary points for v. That being said, efficient implementations of SQuID may avoid two QO solves per iteration. For example, at (nearly) feasible points, one may consider skipping subproblem (3.7) entirely, as we do in our implementation described in section 5. For the purposes of this paper, however, we analyze the behavior of SQuID as it has been presented.

Algorithm 1 Sequential quadratic optimizer with rapid infeasibility detection (SQuID).

1. Choose $\beta \in (0,1)$, $\theta \in (0,1)$, $\kappa_\rho > 0$, $\kappa_\lambda > 0$, $\epsilon \in (0,1)$, $\omega \in (0,1)$, $\delta \in (0,1)$, $\eta \in (0,1)$, and $\gamma \in (0,1)$. Set $k \leftarrow 0$ and choose $(x_k, \rho_k, \overline{\lambda}_k, \widehat{\lambda}_k)$ satisfying (3.6).
2. Compute $(\overline{d}_k, \overline{r}_k, \overline{s}_k, \overline{t}_k, \overline{\lambda}_{k+1})$ as the optimal primal-dual solution for (3.7).
3. If $\mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1}) = 0$ and $v(x_k) > 0$, then terminate; $(x_k, \overline{\lambda}_{k+1})$ is an infeasible stationary point for problem (3.1).
4. If (3.12) holds, then set ρ_k by (3.13) and $\widehat{\lambda}_k$ so that (3.14) holds. Set $\widehat{\rho}_k \leftarrow \rho_k$.
5. Compute $(\widehat{d}_k, \widehat{r}_k^{\mathcal{E}_k^c}, \widehat{s}_k^{\mathcal{E}_k^c}, \widehat{t}_k^{\mathcal{I}_k^c}, \widehat{\lambda}_{k+1})$ as the optimal primal-dual solution for (3.9).
6. If $\mathcal{R}_{opt}(x_k, \rho_k, \widehat{\lambda}_{k+1}) = 0$ and $v(x_k) = 0$, then terminate; $(x_k, \rho_k, \widehat{\lambda}_{k+1})$ is a KKT point for problem (3.1).
7. Set d_k by (3.11) where w_k is the smallest value in $[0,1)$ such that (3.10) holds.
8. Update ρ_k by (3.18), then by (3.19), and finally set $\rho_{k+1} \leftarrow \rho_k$.
9. Let α_k be the largest value in $\{\gamma^0, \gamma^1, \gamma^2, \dots\}$ such that (3.20) holds.
10. Set $x_{k+1} \leftarrow x_k + \alpha_k d_k$ and $k \leftarrow k+1$ and go to step 1.

4. Convergence analysis.

The convergence properties of SQuID are the subject of this section. We prove the well-posedness of the algorithm along with global and local convergence results for feasible and infeasible problem instances. A few of the earlier results in this section are well-known in (nonsmooth) composite function theory, so for the sake of brevity we provide only citations for proofs.

4.1. Well-posedness.

We prove that SQuID is well-posed in that each iteration is well-defined and, if the overall algorithm does not terminate finitely, then an infinite sequence of iterates will be produced. This can be guaranteed under the following assumption. (Note that for simplicity here and in section 4.2, we assume that subproblems (3.7) and (3.9) are convex. See section 4.3 for a discussion of how this assumption can be relaxed without sacrificing local superlinear convergence guarantees.)

ASSUMPTION 4.1. *The following hold true for the iterates generated by SQuID:*

(a) *The problem functions f, $c^{\mathcal{E}}$, and $c^{\mathcal{I}}$ are continuously differentiable in an open convex set containing $\{x_k\}$ and $\{x_k + d_k\}$.*

(b) *For all k, $H(x_k, 0, \overline{\lambda}_k)$ and $H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)$ are positive definite.*

Our first lemma reveals that $-\Delta l(d; x_k)$ and $-\Delta m(d; x_k, \rho)$, respectively, play the roles of surrogates for the directional derivatives of v and $\phi(\cdot.\rho)$ from x_k along the direction d. For a proof, see [2, Lemma 2.3]. We use the lemma to show that as long as a search direction d_k yields a strictly positive reduction in $l(\cdot, x_k)$ $(m(\cdot; x_k, \rho))$, then it is a direction of strict decrease for v $(\phi(\cdot, \rho))$.

LEMMA 4.2. *The reductions in $l(\cdot; x_k)$ and $m(\cdot; x_k, \rho)$ produced by d satisfy*

$$Dv(d; x_k) \le -\Delta l(d; x_k) \quad \textit{and} \quad D\phi(d; x_k, \rho) \le -\Delta m(d; x_k, \rho), \tag{4.1}$$

where $Dv(d; x_k)$ and $D\phi(d; x_k, \rho)$ represent the directional derivatives of v and $\phi(\cdot, \rho)$ at x_k corresponding to a step d, respectively.

The next lemma enumerates relevant properties of subproblem (3.7) related to the well-posedness of SQuID. It states that as long as x_k is not stationary for v, the solution component $\overline{d}_k$ will be a descent direction for v from x_k. These properties are well-known; see, e.g., [2, Theorem 3.6].

LEMMA 4.3. *Suppose Assumption* 4.1 *holds. Then, during iteration k of SQuID we have the following:*

(a) *Subproblem* (3.7) *is feasible and the solution component $\overline{d}_k$ is unique.*

(b) $\Delta l(\overline{d}_k; x_k) \geq 0$ *where equality holds if and only if $\overline{d}_k = 0$.*

(c) $\overline{d}_k = 0$ *if and only if x_k is stationary for v.*

(d) $\overline{d}_k = 0$ *if and only if $(x_k, \overline{\lambda}_{k+1})$ satisfies* (3.5).

Properties of subproblem (3.9) related to the well-posedness of SQuID are enumerated in the next lemma.

LEMMA 4.4. *Suppose Assumption* 4.1 *holds. Then, during iteration k of SQuID we have the following:*

(a) *Subproblem* (3.9) *is feasible and the solution component $\widehat{d}_k$ is unique.*

(b) *With $\rho_k > 0$ and $v(x_k) = 0$, step* 1 *yields $\widehat{d}_k = 0$ if and only if $(x_k, \rho_k, \widehat{\lambda}_{k+1})$ is a KKT point for* (3.1).

Proof. By straightforward verification of the constraint function values, it follows that $\overline{d}_k$ is feasible for subproblem (3.9). Moreover, as $H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)$ is positive definite under Assumption 4.1, the objective of (3.9) is strictly convex and bounded below over the feasible set of the subproblem. Together, these statements imply that subproblem (3.9) is feasible and that the solution component $\widehat{d}_k$ is unique. This proves part (a).

For part (b), the solution $(\widehat{d}_k, \widehat{r}_k, \widehat{s}_k, \widehat{t}_k, \widehat{\lambda}_{k+1})$ satisfies the KKT conditions

$$\rho_k \nabla f(x_k) + H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)\widehat{d}_k + \nabla c^{\mathcal{E}}(x_k)\widehat{\lambda}^{\mathcal{E}}_{k+1} + \nabla c^{\mathcal{I}}(x_k)\widehat{\lambda}^{\mathcal{I}}_{k+1} = 0, \tag{4.2a}$$

$$c^{\mathcal{E}_k}(x_k) + \nabla c^{\mathcal{E}_k}(x_k)^T\widehat{d}_k = 0, \quad c^{\mathcal{E}^c_k}(x_k) + \nabla c^{\mathcal{E}^c_k}(x_k)^T\widehat{d}_k = \widehat{r}_k^{\mathcal{E}^c_k} - \widehat{s}_k^{\mathcal{E}^c_k}, \tag{4.2b}$$

$$c^{\mathcal{I}_k}(x_k) + \nabla c^{\mathcal{I}_k}(x_k)^T\widehat{d}_k \leq 0, \quad c^{\mathcal{I}^c_k}(x_k) + \nabla c^{\mathcal{I}^c_k}(x_k)^T\widehat{d}_k \leq \widehat{t}_k^{\mathcal{I}^c_k}, \tag{4.2c}$$

$$\widehat{\lambda}^{\mathcal{I}_k}_{k+1} \cdot (c^{\mathcal{I}_k}(x_k) + \nabla c^{\mathcal{I}_k}(x_k)^T\widehat{d}_k) = 0, \tag{4.2d}$$

$$\widehat{\lambda}^{\mathcal{I}^c_k}_{k+1} \cdot (c^{\mathcal{I}^c_k}(x_k) + \nabla c^{\mathcal{I}^c_k}(x_k)^T\widehat{d}_k - t_k^{\mathcal{I}^c_k}) = 0, \tag{4.2e}$$

$$(e - \widehat{\lambda}^{\mathcal{E}^c_k}_{k+1}) \cdot \widehat{r}_k^{\mathcal{E}^c_k} = 0, \quad (e + \widehat{\lambda}^{\mathcal{E}^c_k}_{k+1}) \cdot \widehat{s}_k^{\mathcal{E}^c_k} = 0, \quad (e - \widehat{\lambda}^{\mathcal{I}^c_k}_{k+1}) \cdot \widehat{t}_k^{\mathcal{I}^c_k} = 0, \tag{4.2f}$$

$$-e \leq \widehat{\lambda}^{\mathcal{E}^c_k}_{k+1} \leq e, \ 0 \leq \widehat{\lambda}^{\mathcal{I}^c_k}_{k+1} \leq e, \ \text{and} \ \widehat{\lambda}^{\mathcal{I}_k}_{k+1} \geq 0, \tag{4.2g}$$

from which it is easily shown that

$$\widehat{r}_k^{\mathcal{E}^c_k} = [c^{\mathcal{E}^c_k}(x_k) + \nabla c^{\mathcal{E}^c_k}(x_k)^T\widehat{d}_k]^+, \quad \widehat{s}_k^{\mathcal{E}^c_k} = [c^{\mathcal{E}^c_k}(x_k) + \nabla c^{\mathcal{E}^c_k}(x_k)^T\widehat{d}_k]^-,$$
$$\text{and} \quad \widehat{t}_k^{\mathcal{I}^c_k} = [c^{\mathcal{I}^c_k}(x_k) + \nabla c^{\mathcal{I}^c_k}(x_k)^T\widehat{d}_k]^+.$$

Since we assume $v(x_k) = 0$, it follows that $(\overline{d}_k, \overline{r}_k, \overline{s}_k, \overline{t}_k) = 0$ is optimal for (3.7), which means $\mathcal{E}_k = \mathcal{E}$ and $\mathcal{I}_k = \mathcal{I}$. The optimality conditions (4.2) thus reduce to

$$\rho_k \nabla f(x_k) + H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)\widehat{d}_k + \nabla c^{\mathcal{E}}(x_k)\widehat{\lambda}^{\mathcal{E}}_{k+1} + \nabla c^{\mathcal{I}}(x_k)\widehat{\lambda}^{\mathcal{I}}_{k+1} = 0, \tag{4.3a}$$

$$c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T\widehat{d}_k = 0, \quad c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T\widehat{d}_k \leq 0, \tag{4.3b}$$

$$\widehat{\lambda}^{\mathcal{I}}_{k+1} \geq 0, \quad \widehat{\lambda}^{\mathcal{I}}_{k+1} \cdot (c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T\widehat{d}_k) = 0. \tag{4.3c}$$

Since we assume $\rho_k > 0$, by comparing the elements of $\mathcal{R}_{opt}(x_k, \rho_k, \widehat{\lambda}_{k+1})$ with those of (4.3), it follows that $\widehat{d}_k = 0$ if and only if $(x_k, \rho_k, \widehat{\lambda}_{k+1})$ is a KKT point for (3.1). ∎

The next lemma shows that the updates for the penalty parameter in steps 1 and 1 are well-defined and that the latter update guarantees that $\Delta m(d_k; x_k, \rho_{k+1})$ is nonnegative. This can then be used to show, as we do in the lemma, that the line search in step 1 will terminate finitely with a positive step-size $\alpha_k > 0$.

LEMMA 4.5. *Suppose Assumption* 4.1 *holds. Then, during iteration k of SQuID we have the following:*

(a) *If at the beginning of iteration k we have $\rho_k > 0$, then, after step* 1, $\rho_k > 0$.

(b) *If at the beginning of step* 1 *we have $\rho_k > 0$, then, after step* 1, $\rho_{k+1} > 0$ *and*

$$\Delta m(d_k; x_k, \rho_{k+1}) \geq \epsilon \Delta l(d_k; x_k) \geq \beta \epsilon \Delta l(\overline{d}_k; x_k) \geq 0. \tag{4.4}$$

(c) *The line search in step* 1 *terminates with $\alpha_k > 0$.*

Proof. If at step 1 we have $\mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1}) = 0$, then we must have $v(x_k) = 0$ or else SQuID would have terminated in step 1. Thus, since (3.12) does not hold, step 1 will maintain the current $\rho_k > 0$. On the other hand, if at step 1 we have $\mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1}) > 0$, then either ρ_k will be maintained at its current positive value or (3.13) will set $\rho_k > 0$. This proves part (a).

For part (b), first consider (3.18). If $\|\widehat{\lambda}_{k+1}\|_\infty = 0$, then $\rho_k \|\widehat{\lambda}_{k+1}\|_\infty < 1$, meaning that (3.18) will not trigger a reduction in ρ_k. On the other hand, if $\|\widehat{\lambda}_{k+1}\|_\infty > 0$, then (3.18) will only ever yield $\rho_k > 0$. Thus, after applying (3.18), we have $\rho_k > 0$. Now consider (3.19). We have $\Delta l(d_k; x_k) \geq \beta \Delta l(\overline{d}_k; x_k) \geq 0$ due to (3.8) and (3.10), so all that remains is to show that $\rho_{k+1} > 0$ and $\Delta m(d_k; x_k, \rho_{k+1}) \geq \epsilon \Delta l(d_k; x_k)$. There are two cases to consider: $\Delta l(d_k; x_k) = 0$ and $\Delta l(d_k; x_k) > 0$. If $\Delta l(d_k; x_k) = 0$, then according to (3.10) and Lemma 4.3 we must have $\overline{d}_k = 0$. Moreover, if $v(x_k) \neq 0$, then Lemma 4.3 implies that the algorithm would have terminated in step 1, so since we are in step 1, we must have $v(x_k) = 0$, $\mathcal{E}_k = \mathcal{E}$, and $\mathcal{I}_k = \mathcal{I}$. It follows that in step 1 we obtain $d_k = \widehat{d}_k$ (i.e., $w_k = 0$) satisfying $\nabla f(x_k)^T d_k \leq 0$. Observing (3.15), we find that $\Delta m(d_k; x_k, \rho_k) \geq \epsilon \Delta l(d_k; x_k)$ and $w_k < \omega$, so a reduction in ρ_k is not triggered by (3.19), the algorithm sets $\rho_{k+1} \leftarrow \rho_k$, and (4.4) is satisfied. Now consider when $\Delta l(d_k; x_k) > 0$. If $\Delta m(d_k; x_k, \rho_k) \geq \epsilon \Delta l(d_k; x_k)$ and $w_k < \omega$, then there is nothing left to prove as the algorithm sets $\rho_{k+1} \leftarrow \rho_k$ and (4.4) holds. If $\Delta m(d_k; x_k, \rho_k) \geq \epsilon \Delta l(d_k; x_k)$, but $w_k \geq \omega$, then

$$\Delta m(d_k; x_k, \rho_k) \geq \epsilon \Delta l(d_k; x_k) \implies \rho_k \nabla f(x_k)^T d_k \leq (1-\epsilon) \Delta l(d_k; x_k).$$

Then, since $\Delta l(d_k; x_k) > 0$, it follows that prior to the update (3.19) we have

$$\delta \rho_k \nabla f(x_k)^T d_k \leq (1-\epsilon) \Delta l(d_k; x_k) \implies \Delta m(d_k; x_k, \delta \rho_k) \geq \epsilon \Delta l(d_k; x_k).$$

As a result, after the update (3.19), we again have that $\rho_{k+1} > 0$ and (4.4) holds. Finally, if $\Delta m(d_k; x_k, \rho_k) < \epsilon \Delta l(d_k; x_k)$, then by (3.15) we must have $\nabla f(x_k)^T d_k > 0$. In such cases, after ρ_k is updated by (3.19) (to a positive value since $\zeta_k > 0$), we have

$$\begin{aligned} \Delta m(d_k; x_k, \rho_k) &= -\rho_k \nabla f(x_k)^T d_k + \Delta l(d_k; x_k) \\ &\geq -\frac{(1-\epsilon)\Delta l(d_k; x_k)}{\nabla f(x_k)^T d_k + \frac{1}{2} d_k^T H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k) d_k} \nabla f(x_k)^T d_k + \Delta l(d_k; x_k) \\ &\geq (\epsilon - 1)\Delta l(d_k; x_k) + \Delta l(d_k; x_k) \\ &= \epsilon \Delta l(d_k; x_k), \end{aligned}$$

completing the proof of part (b) of the lemma.

Finally, for part (c), we first claim that $\Delta m(d_k; x_k, \rho_{k+1}) > 0$ in step 1. Indeed, by part (b), the model reduction satisfies $\Delta m(d_k; x_k, \rho_{k+1}) = 0$ only if $\Delta l(\overline{d}_k; x_k) = 0$. However, by Lemma 4.3 and the formulation of (3.9), this occurs if and only if x_k is stationary for v. If $v(x_k) > 0$, then SQuID would have terminated in step 1; thus, we may assume $v(x_k) = 0$. Moreover, if $d_k = 0$, then by Lemma 4.4, SQuID would have terminated in step 1; thus, we may assume $d_k \neq 0$. Since under these conditions the point $(d, r, s, t) = (0, 0, 0, 0)$ is feasible for (3.9) and yields an objective value of 0 for that subproblem, we must have $\nabla f(x_k)^T d_k < 0$, meaning that $\Delta m(d_k; x_k, \rho_{k+1}) = -\rho_{k+1} \nabla f(x_k)^T d_k > 0$. Overall, we have shown that if the algorithm enters step 1, then $\Delta m(d_k; x_k, \rho_{k+1}) > 0$. This fact and Lemma 4.2 reveal that d_k is a direction of strict descent for $\phi(\cdot, \rho_{k+1})$ from x_k, implying that the backtracking line search will terminate with a positive step-size $\alpha_k > 0$. ∎

Our main theorem in this subsection summarizes the well-posedness of SQuID.

THEOREM 4.6. *Suppose Assumption* 4.1 *holds. Then, one of the following holds:*

(a) *SQuID terminates in step* 1 *with* $(x_k, \overline{\lambda}_{k+1})$ *satisfying*

$$\mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1}) = 0 \quad and \quad v(x_k) > 0;$$

(b) *SQuID terminates in step* 1 *with* $(x_k, \rho_k, \widehat{\lambda}_{k+1})$ *satisfying*

$$\rho_k > 0, \quad \mathcal{R}_{opt}(x_k, \rho_k, \widehat{\lambda}_{k+1}) = 0, \quad and \quad v(x_k) = 0;$$

(c) *SQuID generates an infinite sequence* $\{(x_k, \rho_k, \overline{\lambda}_k, \widehat{\lambda}_k)\}$ *where, for all* k,

$$\rho_k > 0,\ -e \le \overline{\lambda}_k^{\mathcal{E}} \le e,\ 0 \le \overline{\lambda}_k^{\mathcal{I}} \le e,\ -e \le \widehat{\lambda}_k^{\mathcal{E}_k^c} \le e,\ 0 \le \widehat{\lambda}_k^{\mathcal{I}_k^c} \le e,\ and\ \widehat{\lambda}_k^{\mathcal{I}_k} \ge 0.$$

Proof. By Lemmas 4.3, 4.4, and 4.5, each iteration of SQuID terminates finitely. If SQuID itself does not terminate finitely in step 1 or 1, then steps 1 and 1 and the optimality conditions for subproblems (3.7) and (3.9) yield the bounds in statement (c). Moreover, by Lemma 4.5(a)–(b), it follows that an infinite number of SQuID iterates yields $\{\rho_k\} > 0$. ∎

4.2. Global convergence. We now prove properties related to the global convergence of SQuID under the assumption that an infinite sequence of iterates is generated; i.e., we focus on the situation described in Theorem 4.6(c). These properties require a slight strengthening of our assumptions from section 4.1. (As Assumption 4.7 is stronger than Assumption 4.1, it follows that all results from section 4.1 still hold.)

ASSUMPTION 4.7. *The following hold true for the iterates generated by SQuID:*

(a) *The problem functions* f, $c^{\mathcal{E}}$, $c^{\mathcal{I}}$ *and their first derivatives are bounded and Lipschitz continuous in an open convex set containing* $\{x_k\}$ *and* $\{x_k + d_k\}$.

(b) *There exist constants* $\overline{\mu} \ge \underline{\mu} > 0$ *such that, for all* k *and* $d \in \mathbb{R}^n$,

$$\underline{\mu}\|d\|^2 \le d^T H(x_k, 0, \overline{\lambda}_k) d \le \overline{\mu}\|d\|^2 \quad and \quad \underline{\mu}\|d\|^2 \le d^T H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k) d \le \overline{\mu}\|d\|^2.$$

Of particular interest at the end of this section is the behavior of SQuID in the vicinity of points satisfying the Mangasarian–Fromovitz constraint qualification (MFCQ) for problem (3.1). We define this well-known qualification for convenience.

DEFINITION 4.8. *A point* x *satisfies the MFCQ for problem* (3.1) *if* $v(x) = 0$, $\nabla c^{\mathcal{E}}(x)$ *has full column rank, and there exists* $d \in \mathbb{R}^n$ *such that*

$$c^{\mathcal{E}}(x) + \nabla c^{\mathcal{E}}(x)^T d = 0 \quad and \quad c^{\mathcal{I}}(x) + \nabla c^{\mathcal{I}}(x)^T d < 0.$$

In this and the following subsection, at x_k, let the sets of positive, zero, and negative-valued equality constraints be defined, respectively, as

$$\mathcal{P}_k := \{i \in \mathcal{E} : c^i(x_k) > 0\},\ \mathcal{Z}_k := \{i \in \mathcal{E} : c^i(x_k) = 0\}, \text{ and } \mathcal{N}_k := \{i \in \mathcal{E} : c^i(x_k) < 0\}.$$

Similarly, let the sets of violated, active, and strictly satisfied inequality constraints, respectively, be

$$\mathcal{V}_k := \{i \in \mathcal{I} : c^i(x_k) > 0\},\ \mathcal{A}_k := \{i \in \mathcal{I} : c^i(x_k) = 0\}, \text{ and } \mathcal{S}_k := \{i \in \mathcal{I} : c^i(x_k) < 0\}.$$

We similarly define the sets $\mathcal{P}_*$, $\mathcal{Z}_*$, $\mathcal{N}_*$, $\mathcal{V}_*$, $\mathcal{A}_*$, and $\mathcal{S}_*$ when referring to those index sets corresponding to a point of interest x_*.

The following lemma shows that the norms of the search directions are bounded. This result can also be seen to follow if one applies [2, Lemma 3.4].

LEMMA 4.9. *Suppose Assumption* 4.7 *holds. Then, the sequences* $\{\|\overline{d}_k\|\}$ *and* $\{\|\widehat{d}_k\|\}$ *are bounded above, so the sequence* $\{\|d_k\|\}$ *is bounded above.*

Proof. Under Assumption 4.7, there exists $\tau > 0$ such that $v(x_k) \le \tau$ for any k. In order to derive a contradiction to the statement in the lemma, suppose that $\{\|\overline{d}_k\|\}$ is not bounded. Then, there exists an iteration k yielding $\|\overline{d}_k\|^2 > 2\tau/\underline{\mu}$. The objective value of subproblem (3.7) corresponding to this $\overline{d}_k$ satisfies

$$l(\overline{d}_k; x_k) + \tfrac{1}{2}\overline{d}_k^T H(x_k, 0, \overline{\lambda}_k)\overline{d}_k \ge \tfrac{1}{2}\underline{\mu}\|\overline{d}_k\|^2 > \tau \ge v(x_k).$$

However, this is a contradiction as $v(x_k)$ is the objective value corresponding to $(d, r, s, t) = (0, [c^{\mathcal{E}}(x_k)]^+, [c^{\mathcal{E}}(x_k)]^-, [c^{\mathcal{I}}(x_k)]^+)$, which is also feasible for this subproblem. Thus, $\|\overline{d}_k\|^2 \le 2\tau/\underline{\mu}$ for all k, so $\{\|\overline{d}_k\|\}$ is bounded.

Now suppose, in order to derive a different contradiction, that for some k the optimal solution for (3.9) yields $\underline{\mu}\|\widehat{d}_k\| > 8\rho_0\|\nabla f(x_k)\|$ and $\underline{\mu}\|\widehat{d}_k\|^2 > 2\overline{\mu}\|\overline{d}_k\|^2$. Then, under Assumption 4.7, we find

$$\begin{aligned}
&- \rho_k \nabla f(x_k)^T \widehat{d}_k + \rho_k \nabla f(x_k)^T \overline{d}_k + \tfrac{1}{2}\overline{d}_k^T H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)\overline{d}_k \\
&\le \rho_0\|\nabla f(x_k)\|\|\widehat{d}_k\| + \rho_0\|\nabla f(x_k)\|\|\overline{d}_k\| + \tfrac{1}{2}\overline{\mu}\|\overline{d}_k\|^2 \\
&< \tfrac{1}{8}\underline{\mu}\|\widehat{d}_k\|^2 + \tfrac{1}{8}\underline{\mu}\sqrt{\frac{\underline{\mu}}{2\overline{\mu}}}\|\widehat{d}_k\|^2 + \tfrac{1}{4}\underline{\mu}\|\widehat{d}_k\|^2 \\
&\le \tfrac{1}{2}\underline{\mu}\|\widehat{d}_k\|^2 \\
&\le \tfrac{1}{2}\widehat{d}_k^T H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)\widehat{d}_k.
\end{aligned}$$

Since $(\overline{d}_k, \overline{r}_k, \overline{s}_k, \overline{t}_k)$ is feasible for (3.9) and the above inequality implies

$$\rho_k \nabla f(x_k)^T \widehat{d}_k + \tfrac{1}{2}\widehat{d}_k^T H(x_k, \rho_k, \widehat{\lambda}_k)\widehat{d}_k > \rho_k \nabla f(x_k)^T \overline{d}_k + \tfrac{1}{2}\overline{d}_k^T H(x_k, \rho_k, \widehat{\lambda}_k)\overline{d}_k,$$

it follows that $(\widehat{d}_k, \widehat{r}_k, \widehat{s}_k, \widehat{t}_k)$ cannot be the optimal solution for (3.9), a contradiction. Thus, for all k,

$$\|\widehat{d}_k\| \le \max\left\{8\rho_0\|\nabla f(x_k)\|/\underline{\mu},\ \sqrt{2\overline{\mu}/\underline{\mu}}\|\overline{d}_k\|\right\}$$

and since $\{\|\overline{d}_k\|\}$ and $\{\nabla f(x_k)\}$ are bounded by the above paragraph and Assumption 4.7, respectively, it follows that $\{\|\widehat{d}_k\|\}$ is also bounded.

The boundedness of $\{\|d_k\|\}$ follows from the above inequality results and the fact that d_k is chosen as a convex combination of $\overline{d}_k$ and $\widehat{d}_k$ for all k. ∎

We also have the following lemma, providing a lower bound for α_k for each k.

Lemma 4.10. *Suppose Assumption* 4.7 *holds. Then, for all* k*, the stepsize satisfies* $\alpha_k \geq c\Delta m(d_k; x_k, \rho_{k+1})$ *for some constant* $c > 0$ *independent of* k*.*

Proof. Under Assumption 4.7, applying Taylor's theorem and Lemma 4.2, we have that for all positive α that are sufficiently small, there exists $\tau > 0$ such that

$$\phi(x_k + \alpha d_k, \rho_{k+1}) - \phi(x_k, \rho_{k+1}) \leq -\alpha\Delta m(d_k; x_k, \rho_{k+1}) + \tau\alpha^2\|d_k\|^2.$$

Thus, for any $\alpha \in [0, (1-\eta)\Delta m(d_k; x_k, \rho_{k+1})/(\tau\|d_k\|^2)]$, we have

$$-\alpha\Delta m(d_k; x_k, \rho_{k+1}) + \tau\alpha^2\|d_k\|^2 \leq -\alpha\eta\Delta m(d_k; x_k, \rho_{k+1}),$$

meaning that the sufficient decrease condition (3.20) holds. During the line search, the stepsize is multiplied by γ until (3.20) holds, so we know by the above inequality that the backtracking procedure terminates with

$$\alpha_k \geq \gamma(1-\eta)\Delta m(d_k; x_k, \rho_{k+1})/(\tau\|d_k\|^2).$$

The result follows from this inequality since, by Lemma 4.9, $\{\|d_k\|\}$ is bounded. ∎

We now prove that, in the limit, the reductions in the models of the constraint violation measure and the penalty function vanish. For this purpose, it will be convenient to work with the shifted penalty function

$$\varphi(x, \rho) := \rho(f(x) - \underline{f}) + v(x) \geq 0, \tag{4.5}$$

where $\underline{f}$ is the infimum of f over the smallest convex set containing $\{x_k\}$. The existence of $\underline{f}$ follows from Assumption 4.7. The function φ possesses a useful monotonicity property proved in the following lemma.

Lemma 4.11. *Suppose Assumption* 4.7 *holds. Then, for all* k*,*

$$\varphi(x_{k+1}, \rho_{k+2}) \leq \varphi(x_k, \rho_{k+1}) - \eta\alpha_k\Delta m(d_k; x_k, \rho_{k+1}),$$

so, by Lemmas 4.5 *and* 4.10*,* $\{\varphi(x_k, \rho_{k+1})\}$ *is monotonically decreasing.*

Proof. By the line search condition (3.20), we have

$$\varphi(x_{k+1}, \rho_{k+1}) \leq \varphi(x_k, \rho_{k+1}) - \eta\alpha_k\Delta m(d_k; x_k, \rho_{k+1}),$$

which implies

$$\varphi(x_{k+1}, \rho_{k+2}) \leq \varphi(x_k, \rho_{k+1}) - (\rho_{k+1} - \rho_{k+2})(f(x_{k+1}) - \underline{f}) - \eta\alpha_k\Delta m(d_k; x_k, \rho_{k+1}).$$

The result then follows from this inequality, the fact that $\{\rho_k\}$ is monotonically decreasing, and since $f(x_{k+1}) \geq \underline{f}$ for all k. ∎

We now show that the model reductions vanish in the limit.

Lemma 4.12. *Suppose Assumption* 4.7 *holds. Then, the following limits hold:*

$$0 = \lim_{k\to\infty} \Delta m(d_k; x_k, \rho_{k+1}) = \lim_{k\to\infty} \Delta l(d_k; x_k) = \lim_{k\to\infty} \Delta l(\overline{d}_k; x_k) = \lim_{k\to\infty} \Delta l(\widehat{d}_k; x_k).$$

Proof. In order to derive a contradiction, suppose that $\Delta m(d_k; x_k, \rho_{k+1})$ does not converge to 0. Then, by Lemma 4.5, there exists $\tau > 0$ and an infinite subsequence of iterates K such that $\Delta m(d_k; x_k, \rho_{k+1}) \geq \tau$ for all $k \in K$. By Lemmas 4.10 and 4.11, this would imply that $\varphi(x_k, \rho_{k+1}) \to -\infty$, which is impossible since $\{\varphi(x_k, \rho_{k+1})\}$ is bounded below by 0. Hence, $\Delta m(d_k; x_k, \rho_{k+1}) \to 0$. The other limits follow by Lemma 4.5(b), the fact that d_k is a convex combination of $\overline{d}_k$ and $\widehat{d}_k$ for all k, and the convexity of $\Delta l(\cdot; x_k)$ for all k. ∎

We now show that the primal solution components for the subproblems vanish in the limit, and thus the primal search directions vanish in the limit.

LEMMA 4.13. *Suppose Assumption* 4.7 *holds. Then, the following limits hold:*

$$0 = \lim_{k\to\infty} \overline{d}_k = \lim_{k\to\infty} \widehat{d}_k = \lim_{k\to\infty} d_k.$$

Proof. First, we prove by contradiction that $\overline{d}_k \to 0$. Suppose there exists $\tau > 0$ and an infinite subsequence of iterates K such that $\|\overline{d}_k\| \geq \tau$ for all $k \in K$. By Lemma 4.12, there exists $k' \geq 0$ such that for all $k \geq k'$ we have $\Delta l(\overline{d}_k; x_k) \leq \underline{\mu}\tau^2/4$. (Recall that $\underline{\mu}$ is defined in Assumption 4.7.) Hence, we have that for some $\overline{k} \in K$ such that $k \geq k'$, the optimal objective value of (3.7) satisfies

$$v(x_k) - \Delta l(\overline{d}_k; x_k) + \tfrac{1}{2}\overline{d}_k H(x_k, 0, \overline{\lambda}_k)\overline{d}_k \geq v(x_k) - \tfrac{1}{4}\underline{\mu}\tau^2 + \tfrac{1}{2}\underline{\mu}\tau^2 > v(x_k).$$

This is a contradiction as $v(x_k)$ is the objective value corresponding to $(d, r, s, t) = (0, [c^{\mathcal{E}}(x_k)]^+, [c^{\mathcal{E}}(x_k)]^-, [c^{\mathcal{I}}(x_k)]^+)$, which is also feasible. Thus, $\overline{d}_k \to 0$.

Now we prove that $\widehat{d}_k \to 0$. To do this, we first prove that

$$\lim_{k\to\infty} \rho_k \nabla f(x_k)^T \widehat{d}_k = 0. \tag{4.6}$$

Indeed, (4.6) clearly holds if $\rho_k \to 0$ since $\{\nabla f(x_k)\}$ and $\{\widehat{d}_k\}$ are bounded by Assumption 4.7 and Lemma 4.9, respectively. Otherwise, if $\rho_k \nrightarrow 0$, then due to update (3.19) we must have $\Delta m(d_k; x_k, \rho_k) \geq \epsilon \Delta l(d_k; x_k)$ and $w_k < \omega$ for all sufficiently large k. Hence, by Lemma 4.12, (3.15), (3.11), the fact that $\{\rho_k\}$ is monotonically decreasing, the boundedness of $\{\nabla f(x_k)\}$ under Assumption 4.7, and $\overline{d}_k \to 0$, we have

$$\begin{aligned} 0 &= \lim_{k\to\infty} (\Delta l(d_k; x_k) - \Delta m(d_k; x_k, \rho_{k+1})) \\ &= \lim_{k\to\infty} \rho_{k+1} \nabla f(x_k)^T d_k \\ &= \lim_{k\to\infty} \rho_{k+1} \nabla f(x_k)^T (w_k \overline{d}_k + (1 - w_k)\widehat{d}_k) \\ &= \lim_{k\to\infty} \rho_{k+1} (1 - w_k) \nabla f(x_k)^T \widehat{d}_k. \end{aligned} \tag{4.7}$$

Since $(1 - w_k) > (1 - \omega) > 0$ for all sufficiently large k, and since $(\rho_{k+1} - \rho_k) \to 0$ follows from the facts that $\{\rho_k\}$ is monotonically decreasing and bounded below by zero, the limit (4.7) implies (4.6).

We may now use (4.6) to prove by contradiction that $\widehat{d}_k \to 0$. Suppose there exists $\tau > 0$ and an infinite subsequence of iterations K such that $\|\widehat{d}_k\| \geq \tau$ for all $k \in K$. By (4.6), there exists $k' \geq 0$ such that for all $k \geq k'$ we have $\rho_k \nabla f(x_k)^T \widehat{d}_k \geq -\underline{\mu}\tau^2/4$. Moreover, since $\overline{d}_k \to 0$, $\{\rho_k\}$ is monotonically decreasing, and $\{\nabla f(x_k)\}$ and $\{H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)\}$ are bounded under Assumption 4.7, there exists $k'' \geq 0$ such that for all $k \geq k''$ we have

$$\rho_k \nabla f(x_k)^T \overline{d}_k < \tfrac{1}{16}\underline{\mu}\tau^2 \quad \text{and} \quad \tfrac{1}{2}\overline{d}_k^T H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)\overline{d}_k < \tfrac{1}{16}\underline{\mu}\tau^2. \tag{4.8}$$

Therefore, for $k \in K$ with $k \geq \max\{k', k''\}$, the above inequality and Assumption 4.7(b) imply that the optimal objective value of (3.9) satisfies

$$\rho_k \nabla f(x_k)^T \widehat{d}_k + \tfrac{1}{2}\widehat{d}_k^T H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)\widehat{d}_k \geq \tfrac{1}{4}\underline{\mu}\tau^2 > \rho_k \nabla f(x_k)^T \overline{d}_k + \frac{1}{2}\overline{d}_k^T H(x_k, \widehat{\rho}_k, \lambda_k)\overline{d}_k.$$

This contradicts the fact that $\widehat{d}_k$ is an optimal solution component of (3.9) since $(\overline{d}_k, \overline{r}_k, \overline{s}_k, \overline{t}_k)$ is feasible for (3.9) and the above inequality implies that it yields a lower objective value than $(\widehat{d}_k, \widehat{r}_k, \widehat{s}_k, \widehat{t}_k)$. Hence, $\widehat{d}_k \to 0$.

The remainder of the result, namely that $d_k \to 0$, follows from the above discussion and the fact that d_k is a convex combination of $\overline{d}_k$ and $\widehat{d}_k$ for all k. ∎

We now present our first theorem of this subsection, which states that all limit points of a sequence generated by SQuID are first-order optimal for problem (3.2).

THEOREM 4.14. *Suppose Assumption* 4.7 *holds. Then, the following limit holds:*

$$\lim_{k\to\infty} \mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1}) = 0. \tag{4.9}$$

Therefore, all limit points of $\{(x_k, \overline{\lambda}_{k+1})\}$ *are first-order optimal for problem* (3.2).

Proof. Necessary and sufficient conditions for the optimality of $(\overline{d}_k, \overline{\lambda}_{k+1})$ with respect to (3.7) are

$$H(x_k, 0, \overline{\lambda}_k)\overline{d}_k + \nabla c^{\mathcal{E}}(x_k)\overline{\lambda}^{\mathcal{E}}_{k+1} + \nabla c^{\mathcal{I}}(x_k)\overline{\lambda}^{\mathcal{I}}_{k+1} = 0, \tag{4.10a}$$

$$-e \leq \overline{\lambda}^{\mathcal{E}}_{k+1} \leq e,\ 0 \leq \overline{\lambda}^{\mathcal{I}}_{k+1} \leq e, \tag{4.10b}$$

$$(e - \overline{\lambda}^{\mathcal{E}}_{k+1}) \cdot [c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T\overline{d}_k]^+ = 0, \tag{4.10c}$$

$$(e + \overline{\lambda}^{\mathcal{E}}_{k+1}) \cdot [c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T\overline{d}_k]^- = 0, \tag{4.10d}$$

$$(e - \overline{\lambda}^{\mathcal{I}}_{k+1}) \cdot [c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T\overline{d}_k]^+ = 0, \tag{4.10e}$$

$$\overline{\lambda}^{\mathcal{I}}_{k+1} \cdot [c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T\overline{d}_k]^- = 0, \tag{4.10f}$$

where we have eliminated

$$\overline{r}_k = [c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T\overline{d}_k]^+, \quad \overline{s}_k = [c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T\overline{d}_k]^-,$$
$$\text{and} \quad \overline{t}_k = [c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T\overline{d}_k]^+.$$

By Lemma 4.13, we have $\overline{d}_k \to 0$. Thus, as $\{H(x_k, 0, \overline{\lambda}_k)\}$, $\{\nabla c^{\mathcal{E}}(x_k)\}$, and $\{\nabla c^{\mathcal{I}}(x_k)\}$ are bounded under Assumption 4.7 and $\{\overline{\lambda}_{k+1}\}$ is bounded by (4.10b), it follows from (4.10) that $\mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1}) \to 0$. ∎

We now prove that if the penalty parameter remains bounded away from zero, then all feasible limit points of the iterate sequence correspond to KKT points.

THEOREM 4.15. *Suppose Assumption* 4.7 *holds. Then, if* $\rho_k \to \rho_*$ *for some constant* $\rho_* > 0$ *and* $v(x_k) \to 0$*, the following limit holds:*

$$\lim_{k\to\infty} \mathcal{R}_{opt}(x_k, \rho_k, \widehat{\lambda}_{k+1}) = 0.$$

Thus, every limit point (x_*, ρ_*, λ_*) *of* $\{(x_k, \rho_k, \widehat{\lambda}_{k+1})\}$ *with* $\rho_* > 0$ *and* $v(x_*) = 0$ *is a KKT point for problem* (3.1).

Proof. It follows from (4.2a) and Lemma 4.13 that under Assumption 4.7 we have

$$\nabla_x \mathcal{F}(x_k, \rho_k, \widehat{\lambda}_{k+1}) = -H(x_k, \rho_k, \widehat{\lambda}_k)d_k \to 0. \tag{4.11}$$

Thus, it only remains to show that $\widehat{\lambda}_k^{\mathcal{I}} \cdot c^{\mathcal{I}}(x_k) \to 0$ when $v(x_k) \to 0$. By Lemma 4.12 and the fact that

$$\Delta l(\widehat{d}_k; x_k) = v(x_k) - e^T(\widehat{r}_k^{\mathcal{E}_k^c} + \widehat{s}_k^{\mathcal{E}_k^c}) - e^T\widehat{t}_k^{\mathcal{I}_k^c} \quad \text{with} \quad (\widehat{r}_k^{\mathcal{E}_k^c}, \widehat{s}_k^{\mathcal{E}_k^c}, \widehat{t}_k^{\mathcal{I}_k^c}) \geq 0,$$

we have $\lim_{k\to\infty} \|\widehat{r}_k^{\mathcal{E}_k^c}\|_1 = \lim_{k\to\infty} \|\widehat{s}_k^{\mathcal{E}_k^c}\|_1 = \lim_{k\to\infty} \|\widehat{t}_k^{\mathcal{I}_k^c}\|_1 = 0$. If $\|\widehat{\lambda}_{k+1}\|_\infty$ is unbounded, then $\rho_k \to 0$ by (3.18), contradicting the conditions of the theorem. Hence, it follows from Lemma 4.13 that under Assumption 4.7 we have $\widehat{d}_k^T \nabla c^{\mathcal{I}}(x_k) \widehat{\lambda}_{k+1}^{\mathcal{I}} \to 0$. Consequently, from (4.2d) and (4.2e), we have

$$\begin{aligned} c^{\mathcal{I}_k^c}(x_k) \cdot \widehat{\lambda}_{k+1}^{\mathcal{I}_k^c} &= (\widehat{t}_k^{\mathcal{I}_k^c} - \nabla c^{\mathcal{I}_k^c}(x_k)^T \widehat{d}_k)^T \widehat{\lambda}_{k+1}^{\mathcal{I}_k^c} \to 0 \\ \text{and} \quad c^{\mathcal{I}_k}(x_k) \cdot \widehat{\lambda}_{k+1}^{\mathcal{I}_k} &= -\nabla c^{\mathcal{I}_k}(x_k)^T \widehat{d}_k^T \widehat{\lambda}_{k+1}^{\mathcal{I}_k} \to 0. \end{aligned} \tag{4.12}$$

The result follows from these limits and (4.11). ∎

We conclude this subsection with a theorem describing properties of limit points of SQuID whenever the penalty parameter vanishes.

THEOREM 4.16. *Suppose Assumption* 4.7 *holds. Moreover, suppose* $\rho_k \to 0$ *and let* K_ρ *be the subsequence of iterations during which* ρ_k *is decreased by* (3.13), (3.18), *and/or* (3.19). *Then, the following hold true:*

(a) *Either all limit points of* $\{x_k\}$ *are feasible for* (3.1) *or all are infeasible.*

(b) *If all limit points of* $\{x_k\}$ *are feasible, then all limit points of* $\{x_k\}_{k\in K_\rho}$ *correspond to FJ points for problem* (3.1) *where the MFCQ fails.*

Proof. For part (a), in order to derive a contradiction, suppose there exist infinite subsequences K_* and $K_\times$ such that $\{x_k\}_{k\in K_*} \to x_*$ with $v(x_*) = 0$ and $\{x_k\}_{k\in K_\times} \to x_\times$ with $v(x_\times) = \tau > 0$. Under Assumption 4.7 and since $\rho_k \to 0$, there exists $k_* \geq 0$ such that for all $k \in K_*$ with $k \geq k_*$ we have $\rho_{k+1}(f(x_k) - \underline{f}) < \tau/4$ and $v(x_k) < \tau/4$, meaning that $\varphi(x_k, \rho_{k+1}) < \tau/2$. (Recall that $\underline{f}$ has been defined as the infimum of f over the smallest convex set containing $\{x_k\}$.) On the other hand, we know that $\rho_{k+1}(f(x_k) - \underline{f}) \geq 0$ for all $k \geq 0$ and there exists $k_\times \geq 0$ such that for all $k \in K_\times$ with $k \geq k_\times$ we have $v(x_k) \geq \tau/2$, meaning that $\varphi(x_k, \rho_{k+1}) \geq \tau/2$. This is a contradiction since by Lemma 4.11 $\{\varphi(x_k, \rho_{k+1})\}$ is monotonically decreasing. Thus, the set of limit points of $\{x_k\}$ cannot include feasible and infeasible points at the same time.

For part (b), consider a subsequence $K_* \subseteq K_\rho$ such that $\{x_k\}_{k\in K_*} \to x_*$ for some limit point x_*. Let $K_1 \subseteq K_*$ be the subsequence of iterations during which ρ_k is decreased by (3.13), and let $K_2 \subseteq K_*$ be the subsequence of iterations during which it is decreased by (3.18) and/or (3.19). Since $K_1 \cup K_2 = K_*$ and K_* is infinite, it follows that K_1 or K_2 is infinite, or both. We complete the proof by considering two cases depending on the size of the index set K_2. In each case, our goal will be to show that a set of multipliers produced by SQuID have a nonzero limit point λ_* such that $(x_*, 0, \lambda_*)$ is a FJ point for problem (3.1). We then complete the proof by showing that the MFCQ fails at such limit points.

Case 1. Suppose K_2 is finite, meaning that for all sufficiently large k the algorithm does not decrease ρ_k in (3.18) nor in (3.19). Since $\{\overline{\lambda}_{k+1}\}_{k\in K_1}$ is bounded by (4.10b), it follows that this subsequence has a limit point. If all limit points of $\{\overline{\lambda}_{k+1}\}_{k\in K_1}$ are zero, then for all sufficiently large k we have $-e < \overline{\lambda}_{k+1}^{\mathcal{E}} < e$ and $0 \leq \overline{\lambda}_{k+1}^{\mathcal{I}} < e$. By (4.10c), (4.10d), and (4.10e), this implies

$$c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T \overline{d}_k = 0 \quad \text{and} \quad c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T \overline{d}_k \leq 0,$$

meaning that $\Delta l(\overline{d}_k; x_k) = v(x_k)$ for all such k. However, this result implies that for all such k the algorithm does not decrease ρ_k by (3.13), implying that K_1 is also finite, a contradiction. Therefore, if K_2 is finite, then K_1 is infinite and there exists a nonzero limit point $\overline{\lambda}_*$ of $\{\overline{\lambda}_{k+1}\}_{k\in K_1}$.

Consider a subsequence $K_\lambda \subseteq K_1$ such that $\{(x_k, \overline{\lambda}_{k+1})\}_{k\in K_\lambda} \to (x_*, \overline{\lambda}_*)$. By Theorem 4.14, we have

$$\mathcal{R}_{opt}(x_*, 0, \overline{\lambda}_*) = \lim_{\substack{k\in K_\lambda \\ k\to\infty}} \mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1}) = 0,$$

meaning that $(x_*, 0, \overline{\lambda}_*)$ is a FJ point for problem (3.1).

Case 2. Suppose K_2 is infinite. We first prove that $\|\widehat{\lambda}_{k+1}\|_\infty > 1-\epsilon$ for all sufficiently large $k \in K_2$. By contradiction, suppose there exists an infinite subsequence $K_\epsilon \subseteq K_2$ such that $\|\widehat{\lambda}_{k+1}\|_\infty \le 1-\epsilon$ for all $k\in K_\epsilon$. We will show that ρ_k will not be updated by (3.18) nor by (3.19), contradicting the fact that $k \in K_2$. Since $\rho_k \to 0$, we know that for all sufficiently large $k \in K_\epsilon$ we have $\rho_k\|\widehat{\lambda}_{k+1}\|_\infty < 1$, implying that ρ_k is not reduced by (3.18). Now consider (3.19). By (4.2f), we find that for $k \in K_\epsilon$ we obtain $\widehat{r}_k^{\mathcal{E}_k^c} = \widehat{s}_k^{\mathcal{E}_k^c} = 0$ and $\widehat{t}_k^{\mathcal{I}_k^c} = 0$. This implies that $\Delta l(\widehat{d}_k; x_k) = \Delta l(\overline{d}_k; x_k) = v(x_k)$, so we obtain $w_k = 0 < \omega$, $d_k = \widehat{d}_k$, and $\Delta l(d_k, x_k) = v(x_k)$. We then find that when the algorithm encounters (3.19), we have (temporarily using $\widehat{H}_k$ to denote $H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)$)

$$\begin{aligned}\Delta m(d_k; x_k, \rho_k) - d_k^T\widehat{H}_k d_k &= -\rho_k\nabla f(x_k)^T d_k + \Delta l(d_k, x_k) - d_k^T\widehat{H}_k d_k \\ &= d_k^T\nabla c^{\mathcal{E}}(x_k)\widehat{\lambda}_{k+1}^{\mathcal{E}} + d_k^T\nabla c^{\mathcal{I}}(x_k)\widehat{\lambda}_{k+1}^{\mathcal{I}} + \Delta l(d_k; x_k) \\ &= \left(\|c^{\mathcal{E}}(x_k)\|_1 + \|[c^{\mathcal{I}}(x_k)]^+\|_1\right) \\ &\qquad - c^{\mathcal{E}}(x_k)^T\widehat{\lambda}_{k+1}^{\mathcal{E}} - c^{\mathcal{I}}(x_k)^T\widehat{\lambda}_{k+1}^{\mathcal{I}} \\ &= \left(\|c^{\mathcal{E}}(x_k)\|_1 - c^{\mathcal{E}}(x_k)^T\widehat{\lambda}_{k+1}^{\mathcal{E}}\right) \\ &\qquad + \left(\|[c^{\mathcal{I}}(x_k)]^+\|_1 - c^{\mathcal{I}}(x_k)^T\widehat{\lambda}_{k+1}^{\mathcal{I}}\right).\end{aligned} \tag{4.13}$$

Here, the first equality follows by the definition of $\Delta m(d_k; x_k, \rho_k)$ and the second follows by (4.2a). Then, since (4.2g) implies $\widehat{\lambda}_{k+1}^{\mathcal{I}} \ge 0$, we find that for all $i\in\mathcal{I}$ either $\widehat{\lambda}_{k+1}^i = 0$ or, by (4.2e), $\widehat{\lambda}_{k+1}^i > 0$ and $\nabla c^i(x_k)^T\widehat{d}_k = -c^i(x_k)$. Consequently, we have

$$d_k^T\nabla c^{\mathcal{I}}(x_k)\widehat{\lambda}_{k+1}^{\mathcal{I}} = -c^{\mathcal{I}}(x_k)^T\widehat{\lambda}_{k+1}^{\mathcal{I}}. \tag{4.14}$$

This, along with (4.2b), (4.2c), and the definition of $\Delta l(d_k; x_k)$, yields the third and fourth equalities above, the latter being a rearrangement of the former. Since $\widehat{\lambda}_{k+1}^{\mathcal{E}} \le \|\widehat{\lambda}_{k+1}^{\mathcal{E}}\|_\infty e$, we have $c^{\mathcal{E}}(x_k)^T\widehat{\lambda}_{k+1}^{\mathcal{E}} \le \|\widehat{\lambda}_{k+1}^{\mathcal{E}}\|_\infty\|c^{\mathcal{E}}(x_k)\|_1$, and as $0 \le \widehat{\lambda}_{k+1}^{\mathcal{I}} \le \|\widehat{\lambda}_{k+1}^{\mathcal{I}}\|_\infty e$,

$$c^{\mathcal{I}}(x_k)^T\widehat{\lambda}_{k+1}^{\mathcal{I}} \le [c^{\mathcal{I}}(x_k)]^{+T}\widehat{\lambda}_{k+1}^{\mathcal{I}} \le \|[c^{\mathcal{I}}(x_k)]^+\|_1\|\widehat{\lambda}_{k+1}\|_\infty.$$

Consequently, we have from (4.13) that

$$\begin{aligned}\Delta m(d_k; x_k, \rho_k) &\ge d_k^T\widehat{H}_k d_k + (1-\|\widehat{\lambda}_{k+1}\|_\infty)\|c^{\mathcal{E}}(x_k)\|_1 + (1-\|\widehat{\lambda}_{k+1}\|_\infty)\|[c^{\mathcal{I}}(x_k)]^+\|_1 \\ &= d_k^T\widehat{H}_k d_k + (1-\|\widehat{\lambda}_{k+1}\|_\infty)\Delta l(d_k; x_k) \\ &\ge \epsilon\Delta l(d_k; x_k),\end{aligned} \tag{4.15}$$

meaning that ρ_k will not be reduced by (3.19). Overall, we have contradicted the fact that $k\in K_2$. Hence, we have shown that for large $k \in K_2$, we have $\|\widehat{\lambda}_{k+1}\|_\infty > 1-\epsilon$.

Now let $\widetilde{\rho}_k = \widehat{\rho}_k/\|\widehat{\lambda}_{k+1}\|$ and $\widetilde{\lambda}_{k+1} = \widehat{\lambda}_{k+1}/\|\widehat{\lambda}_{k+1}\|_\infty$ be defined for all $k \in K_2$ such that $\|\widehat{\lambda}_{k+1}\|_\infty > 1-\epsilon$. Since there is an infinite number of such k, it follows that $\widetilde{\rho}_k \to 0$ and there exists a nonzero limit point $\widetilde{\lambda}_*$ of $\{\widetilde{\lambda}_{k+1}\}_{k\in K_2}$. Consider an infinite subsequence $K_\lambda \subseteq K_2$ such that $\{(x_k, \widetilde{\lambda}_{k+1})\}_{k\in K_\lambda} \to (x_*, \widetilde{\lambda}_*)$. By (4.2a), we find that for $k \in K_\lambda$,

$$\nabla_x \mathcal{F}(x_k, \widetilde{\rho}_k, \widetilde{\lambda}_{k+1}) = \widetilde{\rho}_k \nabla f(x_k)^T + \nabla c^{\mathcal{E}}(x_k)\widetilde{\lambda}^{\mathcal{E}}_{k+1} + \nabla c^{\mathcal{I}}(x_k)\widetilde{\lambda}^{\mathcal{I}}_{k+1} = \widehat{H}_k d_k/\|\widehat{\lambda}_{k+1}\|_\infty.$$

Since $d_k \to 0$ by Lemma 4.13 and $\|\widehat{\lambda}_{k+1}\|_\infty$ is bounded below for sufficient large $k \in K_\lambda$, we have that under Assumption 4.7

$$\nabla_x \mathcal{F}(x_*, 0, \lambda_*) = \lim_{\substack{k\in K_\lambda\\ k\to\infty}} \nabla_x \mathcal{F}(x_k, \widetilde{\rho}_k, \widetilde{\lambda}_{k+1}) = 0.$$

Moreover, since $\|\widetilde{\lambda}_{k+1}\|_\infty$ is bounded, as in (4.12), we have

$$c^{\mathcal{I}}(x_*)^T \widetilde{\lambda}^{\mathcal{I}}_* = \lim_{\substack{k\in K_\lambda\\ k\to\infty}} c^{\mathcal{I}}(x_k)^T \widehat{\lambda}^{\mathcal{I}}_{k+1}/\|\widehat{\lambda}_{k+1}\|_\infty = 0.$$

Overall, we have shown that $(x_*, 0, \widetilde{\lambda}_*)$ is a FJ point for problem (3.1).

Let $(x_*, 0, \lambda_*)$ be a FJ point as described above where $\lambda_* = \overline{\lambda}_*$ if we are in Case 1 and $\lambda_* = \widetilde{\lambda}_*$ if we are in Case 2. Then, from dual feasibility in (3.4) we have

$$\nabla_x \mathcal{F}(x_*, 0, \lambda_*) = \nabla c^{\mathcal{I}}(x_*)\lambda^{\mathcal{I}}_* + \nabla c^{\mathcal{E}}(x_*)\lambda^{\mathcal{E}}_* = 0. \tag{4.16}$$

Moreover, from the complementarity conditions in (3.4), we have

$$\nabla c^{\mathcal{A}_*}(x_*)\lambda^{\mathcal{A}_*}_* + \nabla c^{\mathcal{E}}(x_*)\lambda^{\mathcal{E}}_* = 0. \tag{4.17}$$

In order to derive a contradiction, suppose that the MFCQ holds at x_*. Since the MFCQ holds and $v(x_*) = 0$, there exists a vector u such that $\nabla c^{\mathcal{A}_*}(x_*)^T u < 0$ and $\nabla c^{\mathcal{E}}(x_*)^T u = 0$. By (4.17), we then have

$$0 = u^T \nabla c^{\mathcal{A}_*}(x_*)\lambda^{\mathcal{A}_*}_* + u^T \nabla c^{\mathcal{E}}(x_*)\lambda^{\mathcal{E}}_* = u^T \nabla c^{\mathcal{A}_*}(x_*)\lambda^{\mathcal{A}_*}_*. \tag{4.18}$$

Since $\nabla c^{\mathcal{A}_*}(x_*)^T u < 0$ and $\lambda^{\mathcal{A}_*}_* \geq 0$, (4.18) implies $\lambda^{\mathcal{A}_*}_* = 0$. Thus, from (4.17) and the fact that under the MFCQ the columns of $\nabla c^{\mathcal{E}}(x_*)$ are linearly independent, we have $\lambda^{\mathcal{E}}_* = 0$. Overall, we have shown that $\lambda_* = 0$, which contradicts the fact that $(x_*, 0, \lambda_*)$ is a FJ point for problem (3.1). Hence, MFCQ fails at x_*. ∎

We end our global convergence theory with a corollary that summarizes the results of the previous theorems. It also provides a stronger result in a special case when the primal iterates are bounded. This occurs, e.g., when the sublevel sets of the shifted penalty function $\varphi(\cdot, \rho)$ (recall (4.5)) are bounded for all ρ in the closure of $\{\rho_k\}$.

COROLLARY 4.17. *Suppose Assumption* 4.7 *holds and let K_ρ be defined as in Theorem* 4.16. *Then, one of the following situations occurs:*

(i) *$\rho_k \to \rho_*$ for some constant $\rho_* > 0$ and each limit point of $\{x_k\}$ either corresponds to a KKT point or an infeasible stationary point for problem* (3.1)*;*
(ii) *$\rho_k \to 0$ and all limit points of $\{x_k\}$ are infeasible stationary points for* (3.1)*;*
(iii) *$\rho_k \to 0$, all limit points of $\{x_k\}$ are feasible for* (3.1)*, and all limit points of $\{x_k\}_{k\in K_\rho}$ correspond to FJ points for* (3.1) *where the MFCQ fails.*

Consequently, if $\{x_k\}$ *is bounded and all limit points of this sequence are feasible for* (3.1) *and satisfy the MFCQ, then* $\rho_k \to \rho_*$ *for some constant* $\rho_* > 0$ *and all limit points of* $\{x_k\}$ *are KKT points for problem* (3.1).

Proof. The fact that one of the situations (i)–(iii) occurs follows from Theorems 4.14–4.16 and the fact that $\{\rho_k\}$ is monotonically decreasing and bounded below by zero. All that remains is to prove the last sentence of the corollary. In order to derive a contradiction, suppose that under the stated conditions we have $\rho_k \to 0$. Then, since $\{x_k\}$ is bounded, it follows that the sequence $\{x_k\}_{k\in K_\rho}$ has at least one limit point. However, by Theorem 4.16, it follows that such a limit point violates the MFCQ, which in turn contradicts the stated conditions. Hence, $\rho_k \to \rho_*$ for some constant $\rho_* > 0$ and $v(x_k) \to 0$, so the result follows from Theorem 4.15. ☐

4.3. Local convergence. We consider the local convergence of SQuID in the neighborhood of first-order optimal points satisfying certain common assumptions, delineated below. For the most part, our assumptions in this subsection represent a strengthening of the assumptions in section 4.2. However, we loosen our assumptions on the quadratic terms in subproblems (3.7) and (3.9) as in this subsection we only require that they are positive definite in the null space of the Jacobian of the constraints that are active at a given first-order optimal point.

First, for a given x_*, we will use the following assumption.

ASSUMPTION 4.18. *The problem functions* f, $c^{\mathcal{E}}$, *and* $c^{\mathcal{I}}$ *and their first and second derivatives are bounded and Lipschitz continuous in an open convex set containing* x_*.

Second, we make the following assumption concerning a given stationary point x_* of (3.2). As such a point may be feasible or infeasible for (3.1), we make this assumption throughout our local analysis.

ASSUMPTION 4.19. *Let* x_* *be a first-order optimal point for* (3.2) *such that there exists* $\overline{\lambda}_*$ *with* $(x_*, \overline{\lambda}_*)$ *satisfying* (3.5). *Then, Assumption* 4.18 *holds at* x_* *and*

(a) $\nabla c^{\mathcal{Z}_* \cup \mathcal{A}_*}(x_*)^T$ *has full row rank.*

(b) $-e < \overline{\lambda}_*^{\mathcal{Z}_*} < e$ *and* $0 < \overline{\lambda}_*^{\mathcal{A}_*} < e$.

(c) $d^T H(x_*, 0, \overline{\lambda}_*) d > 0$ *for all* $d \neq 0$ *such that* $\nabla c^{\mathcal{Z}_* \cup \mathcal{A}_*}(x_*)^T d = 0$.

Moreover, the following hold true for the iterates generated by SQuID:

(d) $x_k \to x_*$.

(e) *For all large* k, $H(x_k, 0, \overline{\lambda}_k)$ *and* $H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)$ *are the exact Hessian of* $\mathcal{F}$ *at* $(x_k, 0, \overline{\lambda}_k)$ *and* $(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)$, *respectively.*

(f) *For all large* k, $\alpha_k = 1$.

Finally, if $x_k \to x_*$, where x_* is a KKT point for (3.1), we make the following assumption. (While we state Assumption 4.20 now, we will not use it until section 4.3.2.)

ASSUMPTION 4.20. *Let* x_* *be a first-order optimal point for* (3.1) *such that Assumption* 4.19 *holds and there are* $\rho_* > 0$ *and* $\widehat{\lambda}_*$ *with* $(x_*, \rho_*, \widehat{\lambda}_*)$ *satisfying* (3.4). *Then,*

(a) $\rho_k \to \rho_*$.

(b) $\widehat{\lambda}_*^{\mathcal{A}_*} + c^{\mathcal{A}_*}(x_*) > 0$.

(c) $d^T H(x_*, \rho_*, \widehat{\lambda}_*) d > 0$ *for all* $d \neq 0$ *such that* $\nabla c^{\mathcal{E}_* \cup \mathcal{A}_*}(x_*)^T d = 0$.

The assumptions above may be viewed as strong when one considers the fact that local superlinear convergence guarantees for SQO methods have been provided in more general settings. Our algorithm is able to achieve such convergence in such settings, but accounting for more general conditions would only add unnecessary complications to the analysis and detract attention away from our central focus, i.e., the novel feature of attaining superlinear convergence for both feasible and infeasible problem instances with a single algorithm. In particular, consider Assumption 4.19(e) and (f). The former of these assumptions is strong since, if an exact Hessian is indefinite, the algorithm must ensure that of all of the local minimizers of the corresponding QO subproblem, the subproblem solver computes one satisfying certain conditions (implicit in Lemma 3.23 later on). This is challenging as nonconvex quadratic optimization is known to be NP-hard [33]. On the other hand, assuming only that the Hessian is positive definite in the null space of the active constraint Jacobian, the algorithm could ensure that the QO subproblem has a unique solution by modifying the Hessian in appropriate ways so that fast local convergence is still possible. For example, this can be achieved by augmenting the Hessian with $\sigma \nabla c^{\mathcal{Z}_* \cup \mathcal{A}_*}(x_k) \nabla c^{\mathcal{Z}_* \cup \mathcal{A}_*}(x_k)^T$ for a sufficiently large $\sigma > 0$ and then applying the characterization result for superlinear convergence found in [1]. As for Assumption 4.19(f), the primary practical concern is the Maratos effect [31], which makes this assumption inappropriate in many cases. However, we may assume that a watchdog mechanism [12] or a second-order correction [19] is employed to ensure that unit steplengths are accepted by the line search for large k. We leave it a subject of future research to see how many of the assumptions above (in addition to Assumptions 4.19(e) and (f)) can be relaxed while still ensuring the convergence guarantees below, potentially with minor algorithmic variations.

4.3.1. Local convergence to an infeasible stationary point. Suppose Assumption 4.19 holds where x_* is an infeasible stationary point for (3.1). We show that, in such cases, SQuID converges quadratically to $(x_*, \overline{\lambda}_*)$. Some of our analysis for this case follows that in [6], though we provide proofs for completeness.

A critical component of our local convergence analysis in this subsection is to show that there is an inherent relationship between problem (3.2) and the following:

$$
\begin{aligned}
&\min_{(x, r^{\mathcal{P}_*}, s^{\mathcal{N}_*}, t^{\mathcal{V}_*})} \ \rho f(x) + e^T r^{\mathcal{P}_*} + e^T s^{\mathcal{N}_*} + e^T t^{\mathcal{V}_*} \\
&\text{s.t.} \begin{cases} c^{\mathcal{P}_*}(x) = r^{\mathcal{P}_*}, \quad c^{\mathcal{Z}_*}(x) = 0, \quad -c^{\mathcal{N}_*}(x) = s^{\mathcal{N}_*}, \\ c^{\mathcal{V}_*}(x) \le t^{\mathcal{V}_*}, \quad c^{\mathcal{A}_* \cup \mathcal{S}_*}(x) \le 0, \\ (r^{\mathcal{P}_*}, s^{\mathcal{N}_*}, t^{\mathcal{V}_*}) \ge 0. \end{cases}
\end{aligned} \tag{4.19}
$$

In particular, in our first two lemmas, we establish that solutions of (4.19) converge to that of (3.2) as $\rho \to 0$.

The following lemma shows that x_* corresponds to a solution of (4.19) for $\rho = 0$.

LEMMA 4.21. *Suppose Assumption* 4.19 *holds and* $v(x_*) > 0$. *Then,* x_* *and*

$$(r_*^{\mathcal{P}_*}, s_*^{\mathcal{N}_*}, t_*^{\mathcal{V}_*}) = (c^{\mathcal{P}_*}(x_*), -c^{\mathcal{N}_*}(x_*), c^{\mathcal{V}_*}(x_*))$$

correspond to a first-order optimal point for (4.19) *for* $\rho = 0$. *Moreover, the corresponding dual solution is the unique* $\overline{\lambda}_*$ *such that* $(x_*, \overline{\lambda}_*)$ *satisfies* (3.5).

Proof. First-order optimality conditions for (4.19) are the following:

$$\rho\nabla f(x) + \nabla c^{\mathcal{E}}(x)\lambda^{\mathcal{E}} + \nabla c^{\mathcal{I}}(x)\lambda^{\mathcal{I}} = 0, \tag{4.20a}$$
$$c^{\mathcal{P}_*}(x) = r^{\mathcal{P}_*}, \quad c^{\mathcal{Z}_*}(x) = 0, \quad -c^{\mathcal{N}_*}(x) = s^{\mathcal{N}_*}, \tag{4.20b}$$
$$c^{\mathcal{V}_*}(x) \le 0, \quad c^{\mathcal{A}_*\cup\mathcal{S}_*}(x) \le 0, \tag{4.20c}$$
$$(r^{\mathcal{P}_*}, s^{\mathcal{N}_*}, t^{\mathcal{V}_*}) \ge 0, \tag{4.20d}$$
$$\lambda^{\mathcal{A}_*\cup\mathcal{S}_*} \cdot c^{\mathcal{A}_*\cup\mathcal{S}_*}(x) = 0, \tag{4.20e}$$
$$\lambda^{\mathcal{V}_*} \cdot (c^{\mathcal{V}_*}(x) - t^{\mathcal{V}_*}) = 0, \tag{4.20f}$$
$$(e - \lambda^{\mathcal{P}_*}) \cdot r^{\mathcal{P}_*} = 0, \quad (e + \lambda^{\mathcal{N}_*}) \cdot s^{\mathcal{N}_*} = 0, \quad (e - \lambda^{\mathcal{V}_*}) \cdot t^{\mathcal{V}_*} = 0, \tag{4.20g}$$
$$\lambda^{\mathcal{P}_*} \le e, \ \lambda^{\mathcal{N}_*} \ge -e, \ \lambda^{\mathcal{A}_*\cup\mathcal{S}_*} \ge 0, \ 0 \le \lambda^{\mathcal{V}_*} \le 0. \tag{4.20h}$$

If x_* is an infeasible stationary point, then by definition there exists $\overline{\lambda}_* \neq 0$ such that $(x_*, \overline{\lambda}_*)$ satisfies (3.5). Then, with $r^{\mathcal{P}_*}$, $s^{\mathcal{N}_*}$, and $t^{\mathcal{V}_*}$ chosen as in the statement of this lemma, it is easily verified that $(x_*, r^{\mathcal{P}_*}, s^{\mathcal{N}_*}, t^{\mathcal{V}_*}, \overline{\lambda}_*)$ satisfies (4.20) for $\rho = 0$. Moreover, from (4.20e) and (4.20g), we find $\overline{\lambda}_*^{\mathcal{S}_*} = 0$, $\overline{\lambda}_*^{\mathcal{P}_*} = e$, $\overline{\lambda}_*^{\mathcal{N}_*} = -e$, and $\overline{\lambda}^{\mathcal{V}_*} = e$. These equations and (4.20a) imply that we have

$$\nabla c^{\mathcal{Z}_*\cup\mathcal{A}_*}(x_*)\overline{\lambda}_*^{\mathcal{Z}_*\cup\mathcal{A}_*} = -\nabla c^{\mathcal{P}_*\cup\mathcal{V}_*}(x_*)e + \nabla c^{\mathcal{N}_*}(x_*)e. \tag{4.21}$$

Under Assumption 4.19(a), $\overline{\lambda}_*^{\mathcal{Z}_*\cup\mathcal{A}_*}$ in (4.21) is unique. Thus, $\overline{\lambda}_*$ is unique. ◻

We now show that for sufficiently small $\rho > 0$, the solution of problem (4.19) shares critical properties with that of problem (3.2). This result is formalized in our next lemma, which makes use of the following nonlinear system of equations:

$$\begin{aligned} 0 &= F(x, \rho, \lambda^{\mathcal{Z}_*\cup\mathcal{A}_*}) \\ &:= \begin{bmatrix} \rho\nabla f(x) + \nabla c^{\mathcal{Z}_*\cup\mathcal{A}_*}(x)\lambda^{\mathcal{Z}_*\cup\mathcal{A}_*} + \nabla c^{\mathcal{P}_*\cup\mathcal{V}_*}(x)e - \nabla c^{\mathcal{N}_*}(x)e \\ c^{\mathcal{Z}_*\cup\mathcal{A}_*}(x) \end{bmatrix}. \end{aligned} \tag{4.22}$$

By differentiating F with respect to $(x, \lambda^{\mathcal{Z}_*\cup\mathcal{A}_*})$, we obtain

$$F'(x, \rho, \lambda^{\mathcal{Z}_*\cup\mathcal{A}_*}) := \frac{\partial F(x, \rho, \lambda^{\mathcal{Z}_*\cup\mathcal{A}_*})}{\partial(x, \lambda^{\mathcal{Z}_*\cup\mathcal{A}_*})} = \begin{bmatrix} G(x, \rho, \lambda^{\mathcal{Z}_*\cup\mathcal{A}_*}) & \nabla c^{\mathcal{Z}_*\cup\mathcal{A}_*}(x) \\ \nabla c^{\mathcal{Z}_*\cup\mathcal{A}_*}(x)^T & 0 \end{bmatrix} \tag{4.23}$$

where

$$G(x, \rho, \lambda^{\mathcal{Z}_*\cup\mathcal{A}_*}) := \rho\nabla^2 f(x) + \sum_{i\in\mathcal{P}_*\cup\mathcal{V}_*} \nabla^2 c^i(x) + \sum_{i\in\mathcal{Z}_*\cup\mathcal{A}_*} \lambda^i \nabla^2 c^i(x) - \sum_{i\in\mathcal{N}_*} \nabla^2 c^i(x).$$

LEMMA 4.22. *Suppose Assumption* 4.19 *holds and* $v(x_*) > 0$. *Then, for all* ρ *sufficiently small, problem* (4.19) *has a solution* $(x_\rho, r_\rho^{\mathcal{P}_*}, s_\rho^{\mathcal{N}_*}, t_\rho^{\mathcal{V}_*})$ *where* x_ρ *yields the same sets of positive, zero, and negative-valued equality constraints and violated, active, and strictly satisfied inequality constraints as* x_*. *Moreover, for such* ρ, *the corresponding dual variables satisfy* $\overline{\lambda}_\rho^{\mathcal{P}_*} = e$, $-e < \overline{\lambda}_\rho^{\mathcal{Z}_*} < e$, $\overline{\lambda}_\rho^{\mathcal{N}_*} = -e$, $\overline{\lambda}_\rho^{\mathcal{V}_*} = e$, $0 < \overline{\lambda}_\rho^{\mathcal{A}_*} < e$, *and* $\overline{\lambda}_\rho^{\mathcal{S}_*} = 0$, *and we have*

$$\left\| \begin{bmatrix} x_\rho - x_* \\ \overline{\lambda}_\rho - \overline{\lambda}_* \end{bmatrix} \right\| = O(\rho). \tag{4.24}$$

Proof. Under Assumption 4.18, F in (4.22) is a continuously differentiable mapping about $(x_*, 0, \overline{\lambda}_*^{\mathcal{Z}_*\cup\mathcal{A}_*})$, and under Assumption 4.19(a) and (c), the matrix F' in (4.23) is nonsingular at $(x_*, 0, \overline{\lambda}_*^{\mathcal{Z}_*\cup\mathcal{A}_*})$. Thus, by the implicit function theorem [36, Theorem 9.28], there exist open sets $\mathcal{B}_x \subset \mathbb{R}^n$, $\mathcal{B}_\rho \subset \mathbb{R}$, and $\mathcal{B}_\lambda \subset \mathbb{R}^{|\mathcal{Z}_*\cup\mathcal{A}_*|}$ containing x_*, 0, and $\overline{\lambda}_*^{\mathcal{Z}_*\cup\mathcal{A}_*}$, respectively, and continuously differentiable functions $x(\rho) : \mathcal{B}_\rho \to \mathcal{B}_x$ and $\overline{\lambda}^{\mathcal{Z}_*\cup\mathcal{A}_*}(\rho) : \mathcal{B}_\rho \to \mathcal{B}_\lambda$ such that

$$x(0) = x_*, \quad \overline{\lambda}^{\mathcal{Z}_*\cup\mathcal{A}_*}(0) = \overline{\lambda}_*^{\mathcal{Z}_*\cup\mathcal{A}_*}, \quad \text{and} \quad F(x(\rho), \rho, \overline{\lambda}^{\mathcal{Z}_*\cup\mathcal{A}_*}(\rho)) = 0 \quad \text{for all} \quad \rho \in \mathcal{B}_\rho.$$

By the second equation in (4.22) and since $x(\rho)$ varies continuously with ρ, we have

$$c^{\mathcal{Z}_*\cup\mathcal{A}_*}(x(\rho)) = 0, \quad c^{\mathcal{P}_*\cup\mathcal{V}_*}(x(\rho)) > 0, \quad \text{and} \quad c^{\mathcal{N}_*\cup\mathcal{S}_*}(x(\rho)) < 0 \tag{4.25}$$

for ρ sufficiently small. Similarly, since $-e < \overline{\lambda}_*^{\mathcal{Z}_*} < e$ and $0 < \overline{\lambda}_*^{\mathcal{A}_*} < e$ under Assumption 4.19(b), the fact that $\overline{\lambda}^{\mathcal{Z}_*\cup\mathcal{A}_*}(\rho)$ varies continuously with ρ implies that $-e < \overline{\lambda}^{\mathcal{Z}_*}(\rho) < e$ and $0 < \overline{\lambda}^{\mathcal{A}_*}(\rho) < e$ for ρ sufficiently small. If we define

$$\overline{\lambda}^{\mathcal{P}_*\cup\mathcal{V}_*}(\rho) := e, \quad \overline{\lambda}^{\mathcal{N}_*}(\rho) := -e, \quad \text{and} \quad \overline{\lambda}^{\mathcal{S}_*}(\rho) := 0$$

along with

$$r^{\mathcal{P}^*}(\rho) := [c^{\mathcal{P}_*}(x(\rho))]^+, \quad s^{\mathcal{N}_*}(\rho) := [c^{\mathcal{N}_*}(x(\rho))]^-, \quad \text{and} \quad t^{\mathcal{V}_*}(\rho) := [c^{\mathcal{V}_*}(x(\rho))]^+,$$

it follows that $(x(\rho), r^{\mathcal{P}_*}(\rho), s^{\mathcal{N}_*}(\rho), t^{\mathcal{V}_*}(\rho), \overline{\lambda}(\rho))$ satisfies (4.20), and is, therefore, a first-order optimal point for (4.19) for sufficiently small ρ. Hence, by (4.25), we have that $x_\rho = x(\rho)$ for ρ sufficiently small has the same sets of positive, zero, and negative-valued equality and violated, active, and strictly satisfied inequality constraints as x_*.

All that remains is to establish (4.24). From the differentiability of $x_\rho = x(\rho)$ and $\overline{\lambda}_\rho^{\mathcal{Z}_*\cup\mathcal{A}_*} = \overline{\lambda}^{\mathcal{Z}_*\cup\mathcal{A}_*}(\rho)$ and their derivatives given by the implicit function theorem, we have for ρ sufficiently small that

$$\begin{bmatrix} x_\rho \\ \overline{\lambda}_\rho^{\mathcal{Z}_*\cup\mathcal{A}_*} \end{bmatrix} = \begin{bmatrix} x_* \\ \overline{\lambda}_*^{\mathcal{Z}_*\cup\mathcal{A}_*} \end{bmatrix} - F'_{x,\lambda^{\mathcal{Z}_*\cup\mathcal{A}_*}}(x_*, 0, \overline{\lambda}_*^{\mathcal{Z}_*\cup\mathcal{A}_*})^{-1} F'_\rho(x_*, 0, \overline{\lambda}_*^{\mathcal{Z}_*\cup\mathcal{A}_*})\rho + o(\rho).$$

Hence, under Assumption 4.19, (4.24) is satisfied. ∎

We now turn back to the iterates produced by SQuID. In particular, as in the previous lemma, we show that in a neighborhood of an infeasible stationary point, subproblems (3.7) and (3.9) will suggest the optimal partition of the index sets $\mathcal{E}$ and $\mathcal{I}$. This result is reminiscent of the well-known result in [35].

LEMMA 4.23. *Suppose Assumption* 4.19 *holds and* $v(x_*) > 0$. *Then, for all* $\widehat{\rho}_k$ *sufficiently small and for all* $(x_k, \overline{\lambda}_k)$ *and* $(x_k, \widehat{\lambda}_k)$ *each sufficiently close to* $(x_*, \overline{\lambda}_*)$ *we have the following*:

(a) *There are local solutions for* (3.7) *and* (3.9) *such that* $\overline{d}_k$ *and* $\widehat{d}_k$ *yield the same sets of positive, zero, and negative-valued equality and violated, active, and strictly satisfied inequality constraints as* x_*. *Moreover, with* $(\rho, H) = (0, H(x_k, 0, \overline{\lambda}_k))$ *and* $(\rho, H) = (\widehat{\rho}_k, H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k))$, *respectively, the optimal solutions for* (3.7) *and* (3.9) *satisfy*

$$\begin{bmatrix} H & \nabla c^{\mathcal{Z}_*\cup\mathcal{A}_*}(x_k) \\ \nabla c^{\mathcal{Z}_*\cup\mathcal{A}_*}(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \lambda^{\mathcal{Z}_*\cup\mathcal{A}_*} \end{bmatrix} = -\begin{bmatrix} \rho\nabla f(x_k) + \nabla c^{\mathcal{P}_*\cup\mathcal{V}_*}(x_k) - \nabla c^{\mathcal{N}_*}(x_k) \\ c^{\mathcal{Z}_*\cup\mathcal{A}_*}(x_k) \end{bmatrix} \tag{4.26}$$

and

$$\lambda^{\mathcal{P}_*\cup\mathcal{V}_*} = e,\ -e < \lambda^{\mathcal{Z}_*} < e,\ \lambda^{\mathcal{N}_*} = -e,\ 0 < \lambda^{\mathcal{A}_*} < e,\ \textit{and}\ \lambda^{\mathcal{S}_*} = 0. \tag{4.27}$$

(b) *The update* (3.13) *is triggered infinitely often, yielding* $(\rho_k, \widehat{\rho}_k) \to 0$.

Proof. For part (a), consider subproblem (3.7), meaning that we let $(\rho, H) = (0, H(x_k, 0, \overline{\lambda}_k))$ in (4.26). With $\overline{d}_k = 0$, (4.10) reduces to (3.5). Thus, (4.10) is solved at $(x_*, \overline{\lambda}_*)$ by $(d, \lambda) = (0, \overline{\lambda}_*)$. By (4.10c)–(4.10f), we have $\overline{\lambda}_*^{\mathcal{P}_*\cup\mathcal{V}_*} = e$, $\overline{\lambda}_*^{\mathcal{N}_*} = -e$, and $\overline{\lambda}_*^{\mathcal{S}_*} = 0$. Hence, by (4.10a) and the definitions of $\mathcal{Z}_*$ and $\mathcal{A}_*$, the linear system (4.26) is satisfied at $(x_*, \overline{\lambda}_*)$ by $(d, \lambda^{\mathcal{Z}_*\cup\mathcal{A}_*}) = (0, \overline{\lambda}^{\mathcal{Z}_*\cup\mathcal{A}_*})$. Under Assumption 4.19(a) and (c), the matrix in (4.26) is nonsingular at $(x_*, \overline{\lambda}_*)$, and hence the solution of (4.26) varies continuously in a neighborhood of $(x_*, \overline{\lambda}_*)$. In addition, under Assumption 4.19(c), it follows that $H = H(x_k, 0, \overline{\lambda}_k)$ in (4.26) is positive definite on the null space of $\nabla c^{\mathcal{Z}_*\cup\mathcal{A}_*}(x_k)^T$ in a neighborhood of $(x_*, \overline{\lambda}_*)$.

It follows from the conclusions in the previous paragraph that for all $(x_k, \overline{\lambda}_k)$ sufficiently close to $(x_*, \overline{\lambda}_*)$, the solution $(\overline{d}_k, \overline{\lambda}_{k+1}^{\mathcal{Z}_*\cup\mathcal{A}_*})$ to (4.26) is sufficiently close to $(0, \overline{\lambda}_*^{\mathcal{Z}_*\cup\mathcal{A}_*})$ such that it satisfies

$$\begin{aligned} -e < \overline{\lambda}_{k+1}^{\mathcal{Z}_*} < e,\quad 0 < \overline{\lambda}_{k+1}^{\mathcal{A}_*} < e,& \\ c^{\mathcal{P}_*\cup\mathcal{V}_*}(x_k) + \nabla c^{\mathcal{P}_*\cup\mathcal{V}_*}(x_k)^T \overline{d}_k > 0,& \\ \text{and}\quad c^{\mathcal{N}_*\cup\mathcal{S}_*}(x_k) + \nabla c^{\mathcal{N}_*\cup\mathcal{S}_*}(x_k)^T \overline{d}_k < 0.& \end{aligned}$$

By construction, such a solution $(\overline{d}_k, \overline{\lambda}_{k+1}^{\mathcal{Z}_*\cup\mathcal{A}_*})$ satisfies (4.26) and, therefore, satisfies (4.10) together with $\overline{\lambda}_{k+1}^{\mathcal{P}_*\cup\mathcal{V}_*} = e$, $\overline{\lambda}_{k+1}^{\mathcal{N}_*} = -e$, and $\overline{\lambda}_{k+1}^{\mathcal{S}_*} = 0$. Therefore, $(\overline{d}_k, \overline{\lambda}_{k+1})$ is a KKT point of subproblem (3.7), and, as revealed above, it identifies the same sets of positive, zero, and negative-valued equality and violated, active, and strictly satisfied inequality constraints as x_*.

The proof of the result corresponding to subproblem (3.9) is similar. Indeed, from the discussion above, we find that for ρ_k (and hence $\widehat{\rho}_k$) sufficiently small and $(x_k, \overline{\lambda}_k)$ sufficiently close to $(x_*, \overline{\lambda}_*)$, the algorithm will set $\mathcal{E}_k = \mathcal{Z}_*$, $\mathcal{E}_k^c = \mathcal{P}_*\cup\mathcal{N}_*$, $\mathcal{I}_k = \mathcal{A}_*\cup\mathcal{S}_*$ and $\mathcal{I}_k^c = \mathcal{V}_*$. The remainder of the proof follows as above with $H(x_k, 0, \overline{\lambda}_k)$, (4.10), and $(\overline{d}_k, \overline{\lambda}_{k+1})$ replaced by $H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)$, (4.2), and $(\widehat{d}_k, \widehat{\lambda}_{k+1})$, respectively.

Now we prove part (b). We first argue that (3.12) holds for all sufficiently large k so that ρ_k is set by (3.13) infinitely many times. Then, we show that this yields $\rho_k \to 0$. As x_k approaches x_*, we have that $v(x_k) > \frac{1}{2}v(x_*) > 0$ for all large k. On the other hand, in a neighborhood of x_*, the constraint functions $c^{\mathcal{E}}$ and $c^{\mathcal{I}}$ are bounded under Assumption 4.18. Thus, by the definition of $\Delta l(\overline{d}_k; x_k)$ and since for all $(x_k, \overline{\lambda}_k)$ sufficiently close to $(x_*, \overline{\lambda}_*)$, the solution $(\overline{d}_k, \overline{\lambda}_{k+1})$ to (4.26) is sufficiently close to $(0, \overline{\lambda}_*)$, we have that $\Delta l(\overline{d}_k; x_k) \le \frac{\theta}{2}v(x_*) < \theta v(x_k)$ for sufficiently large k. Overall, this implies that (3.12) holds for such k. Hence, (3.13) is triggered infinitely many times. Finally, to see that (3.13) drives $\rho_k \to 0$, it suffices to see that $(\overline{d}_k, \overline{\lambda}_{k+1}) \to (0, \overline{\lambda}_*)$, (4.10a), and (4.10c)–(4.10f) yield $\mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1}) \to 0$. ∎

Lemma 4.23 can be used to show that near $(x_*, \overline{\lambda}_*)$, the solutions of system (4.26) with $(\rho, H) = \big(0, H(x_k, 0, \overline{\lambda}_k)\big)$ and $(\rho, H) = \big(\widehat{\rho}_k, H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)\big)$ correspond to Newton steps for $F(x, 0, \lambda^{\mathcal{Z}_*\cup\mathcal{A}_*}) = 0$ and $F(x, \widehat{\rho}_k, \lambda^{\mathcal{Z}_*\cup\mathcal{A}_*}) = 0$, respectively. We formalize this property in the following lemma.

LEMMA 4.24. *Suppose Assumption* 4.19 *holds and* $v(x_*) > 0$. *Then we have the following:*

(a) *If* $(x_k, \overline{\lambda}_k)$ *is sufficiently close to* $(x_*, \overline{\lambda}_*)$ *and* $(\overline{d}_k, \overline{\lambda}_{k+1})$ *generated by subproblem* (3.7) *is obtained via* (4.26) *with* $(\rho, H) = (0, H(x_k, 0, \overline{\lambda}_k))$*, then*

$$\left\| \begin{bmatrix} x_k + \overline{d}_k - x_* \\ \overline{\lambda}_{k+1} - \overline{\lambda}_* \end{bmatrix} \right\| \leq \overline{C} \left\| \begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix} \right\|^2 \tag{4.28}$$

for some constant $\overline{C} > 0$ *independent of* k.

(b) *If* $(x_k, \widehat{\lambda}_k)$ *is sufficiently close to* $(x_*, \overline{\lambda}_*)$ *and* $(\widehat{d}_k, \widehat{\lambda}_{k+1})$ *generated by subproblem* (3.9) *is obtained via* (4.26) *with* $(\rho, H) = (\widehat{\rho}_k, H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k))$*, then, with* (x_ρ, λ_ρ) *defined as in Lemma 4.22, we have*

$$\left\| \begin{bmatrix} x_k + \widehat{d}_k - x_\rho \\ \widehat{\lambda}_{k+1} - \lambda_\rho \end{bmatrix} \right\| \leq \widehat{C} \left\| \begin{bmatrix} x_k - x_\rho \\ \widehat{\lambda}_k - \lambda_\rho \end{bmatrix} \right\|^2 \tag{4.29}$$

for some constant $\widehat{C} > 0$ *independent of* k.

Proof. For both parts (a) and (b), our proof follows that of [6, Lemma 3.5].

For part (a), by Lemma 4.23(a), if $(x_k, \overline{\lambda}_k)$ is sufficiently close to $(x_*, \overline{\lambda}_*)$, then $(\overline{d}_k, \overline{\lambda}_{k+1})$ generated by subproblem (3.7) can be obtained via (4.27) and (4.26) with $(\rho, H) = (0, H(x_k, 0, \overline{\lambda}_k))$. Therefore, since $H(x_k, 0, \overline{\lambda}_k) = G(x_k, 0, \overline{\lambda}_k^{\mathcal{Z}_* \cup \mathcal{A}_*})$ in such cases, (4.26) constitutes a Newton iteration for $F(x, 0, \lambda^{\mathcal{Z}_* \cup \mathcal{A}_*}) = 0$ at $(x_k, 0, \overline{\lambda}_k)$. We can now apply standard Newton analysis. By Assumption 4.18 we have that F is continuously differentiable and F' is Lipschitz continuous in a neighborhood of $(x_*, 0, \overline{\lambda}_*)$. By Assumption 4.19(a) and (c), the matrix F' is nonsingular at $(x_*, 0, \overline{\lambda}_*^{\mathcal{Z}_* \cup \mathcal{A}_*})$, so its inverse exists and is bounded in a neighborhood of $(x_*, 0, \overline{\lambda}_*^{\mathcal{Z}_* \cup \mathcal{A}_*})$. By [16, Theorem 5.2.1], if $(x_k, \overline{\lambda}_k)$ is sufficiently close to $(x_*, \overline{\lambda}_*)$, then we have that (4.28) holds true.

For part (b), by Lemma 4.23(a), if $(x_k, \widehat{\lambda}_k)$ is sufficiently close to $(x_*, \overline{\lambda}_*)$, then $(\widehat{d}_k, \widehat{\lambda}_{k+1})$ generated by subproblem (3.9) can be obtained via (4.27) and (4.26) with $(\rho, H) = (\widehat{\rho}_k, H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k))$. Therefore, since $H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k) = G(x_k, \widehat{\rho}_k, \widehat{\lambda}_k^{\mathcal{Z}_* \cup \mathcal{A}_*})$ in such cases, system (4.26) constitutes a Newton iteration for $F(x, \rho, \lambda^{\mathcal{Z}_* \cup \mathcal{A}_*}) = 0$ at $(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)$. By Assumption 4.18 we have that F is continuously differentiable and F' is Lipschitz continuous in a neighborhood of $(x_*, \rho, \overline{\lambda}_*^{\mathcal{Z}_* \cup \mathcal{A}_*})$. Moreover, since ρ is bounded, the Lipschitz constant κ_1 for F' in a neighborhood of $(x_*, \rho, \overline{\lambda}_*^{\mathcal{Z}_* \cup \mathcal{A}_*})$ is independent of ρ. By Assumption 4.19(a) and (c), the matrix F' is nonsingular at $(x_*, 0, \overline{\lambda}_*^{\mathcal{Z}_* \cup \mathcal{A}_*})$, and hence its inverse exists and is bounded in norm by a constant κ_2 in a neighborhood of that point. By [16, Theorem 5.2.1],

$$\text{if} \quad \left\| \begin{bmatrix} x_k - x_\rho \\ \widehat{\lambda}_k - \lambda_\rho \end{bmatrix} \right\| \leq \frac{1}{\kappa_1 \kappa_2}, \quad \text{then} \quad \left\| \begin{bmatrix} x_k + \widehat{d}_k - x_\rho \\ \widehat{\lambda}_{k+1} - \lambda_\rho \end{bmatrix} \right\| \leq \kappa_1 \kappa_2 \left\| \begin{bmatrix} x_k - x_\rho \\ \widehat{\lambda}_k - \lambda_\rho \end{bmatrix} \right\|^2.$$

This can be achieved if ρ is sufficiently small such that (x_ρ, λ_ρ) and $(x_k, \widehat{\lambda}_k)$ satisfy

$$\left\| \begin{bmatrix} x_\rho - x_* \\ \lambda_\rho - \overline{\lambda}_* \end{bmatrix} \right\| \leq \frac{1}{4\kappa_1 \kappa_2} \quad \text{and} \quad \left\| \begin{bmatrix} x_k - x_* \\ \widehat{\lambda}_k - \overline{\lambda}_* \end{bmatrix} \right\| \leq \frac{1}{4\kappa_1 \kappa_2}. \qquad \square$$

We are now ready to prove our main theorem concerning the local convergence of SQuID in the neighborhood of infeasible stationary points. The theorem shows that the convergence rate is dependent on how fast ρ is decreased and $\widehat{\lambda}_k$ approaches $\overline{\lambda}_k$.

THEOREM 4.25. *Suppose Assumption* 4.19 *holds and* $v(x_*) > 0$*. Then, if* $(x_k, \overline{\lambda}_k)$ *and* $(x_k, \widehat{\lambda}_k)$ *are each sufficiently close to* $(x_*, \overline{\lambda}_*)$*,* $(\overline{d}_k, \overline{\lambda}_{k+1})$ *is obtained via* (4.27) *and* (4.26) *with* $(\rho, H) = (0, H(x_k, 0, \overline{\lambda}_k))$, *and* $(\widehat{d}_k, \widehat{\lambda}_{k+1})$ *is obtained via* (4.27) *and* (4.26) *with* $(\rho, H) = (\widehat{\rho}_k, H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k))$, *then*

$$\left\| \begin{bmatrix} x_{k+1} - x_* \\ \overline{\lambda}_{k+1} - \overline{\lambda}_* \end{bmatrix} \right\| \leq C \left\| \begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix} \right\|^2 + O(\|\widehat{\lambda}_k - \overline{\lambda}_k\|) + O(\rho) \tag{4.30}$$

for some constant $C > 0$ *independent of* k. *Consequently, as* (3.13) *and* (3.14) *yield*

$$\rho_k = O\left(\left\| \begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix} \right\|^2 \right) \quad \textit{and} \quad \|\widehat{\lambda}_k - \overline{\lambda}_k\| = O\left(\left\| \begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix} \right\|^2 \right),$$

$\{(x_k, \overline{\lambda}_k)\}$ *converges to* $(x_*, \overline{\lambda}_*)$ *quadratically. If* (3.13) *and* (3.14) *merely yielded*

$$\rho_k = o\left(\left\| \begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix} \right\| \right) \quad \textit{and} \quad \|\widehat{\lambda}_k - \overline{\lambda}_k\| = o\left(\left\| \begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix} \right\| \right),$$

then $\{(x_k, \overline{\lambda}_k)\}$ *would converge to* $(x_*, \overline{\lambda}_*)$ *superlinearly.*

Proof. For a given $\rho > 0$, let (x_ρ, λ_ρ) be defined as in Lemma 4.22. Under Assumption 4.19(f), $x_{k+1} = x_k + w_k \overline{d}_k + (1 - w_k)\widehat{d}_k$ for all k. Thus,

$$\begin{aligned}
\left\| \begin{bmatrix} x_{k+1} - x_* \\ \overline{\lambda}_{k+1} - \overline{\lambda}_* \end{bmatrix} \right\| &\leq w_k \left\| \begin{bmatrix} x_k + \overline{d}_k - x_* \\ \overline{\lambda}_{k+1} - \overline{\lambda}_* \end{bmatrix} \right\| + (1 - w_k) \left\| \begin{bmatrix} x_k + \widehat{d}_k - x_* \\ \overline{\lambda}_{k+1} - \overline{\lambda}_* \end{bmatrix} \right\| \\
&\leq w_k \overline{C} \left\| \begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix} \right\|^2 \\
&\quad + (1 - w_k) \left(\left\| \begin{bmatrix} x_k + \widehat{d}_k - x_\rho + x_\rho - x_* \\ 0 \end{bmatrix} \right\| + \left\| \begin{bmatrix} 0 \\ \overline{\lambda}_{k+1} - \overline{\lambda}_* \end{bmatrix} \right\| \right) \\
&\leq w_k \overline{C} \left\| \begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix} \right\|^2 \\
&\quad + (1 - w_k) \left(\left\| \begin{bmatrix} x_k + \widehat{d}_k - x_\rho + x_\rho - x_* \\ \widehat{\lambda}_{k+1} - \lambda_\rho + \lambda_\rho - \overline{\lambda}_* \end{bmatrix} \right\| + \left\| \begin{bmatrix} x_k + \overline{d}_k - x_* \\ \overline{\lambda}_{k+1} - \overline{\lambda}_* \end{bmatrix} \right\| \right) \\
&\leq \overline{C} \left\| \begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix} \right\|^2 \\
&\quad + (1 - w_k) \left(\left\| \begin{bmatrix} x_k + \widehat{d}_k - x_\rho \\ \widehat{\lambda}_{k+1} - \lambda_\rho \end{bmatrix} \right\| + \left\| \begin{bmatrix} x_\rho - x_* \\ \lambda_\rho - \overline{\lambda}_* \end{bmatrix} \right\| \right) \\
&\leq \overline{C} \left\| \begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix} \right\|^2 + \widehat{C} \left\| \begin{bmatrix} x_k - x_\rho \\ \widehat{\lambda}_k - \lambda_\rho \end{bmatrix} \right\|^2 + O(\rho). \qquad (4.31)
\end{aligned}$$

Here, the second and fourth inequalities follow from Lemma 4.24(a), the third holds as we have simply augmented the latter two vector norms, and the last follows from

Lemmas 4.22 and 4.24(b). By applying Lemma 4.22, we also have that

$$\begin{aligned}
\left\|\begin{bmatrix} x_k - x_\rho \\ \widehat{\lambda}_k - \lambda_\rho \end{bmatrix}\right\|^2 &\le \left\|\begin{bmatrix} x_k - x_* \\ \widehat{\lambda}_k - \overline{\lambda}_* \end{bmatrix}\right\|^2 + 2\left\|\begin{bmatrix} x_k - x_* \\ \widehat{\lambda}_k - \overline{\lambda}_* \end{bmatrix}\right\| \left\|\begin{bmatrix} x_\rho - x_* \\ \lambda_\rho - \overline{\lambda}_* \end{bmatrix}\right\| + \left\|\begin{bmatrix} x_\rho - x_* \\ \lambda_\rho - \overline{\lambda}_* \end{bmatrix}\right\|^2 \\
&\le \left\|\begin{bmatrix} x_k - x_* \\ \widehat{\lambda}_k - \overline{\lambda}_* \end{bmatrix}\right\|^2 + O(\rho) \\
&\le \left\|\begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix}\right\|^2 + 2\|\widehat{\lambda}_k - \overline{\lambda}_k\| \left\|\begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix}\right\| + \|\widehat{\lambda}_k - \overline{\lambda}_k\|^2 + O(\rho) \\
&\le \left\|\begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix}\right\|^2 + O(\|\widehat{\lambda}_k - \overline{\lambda}_k\|) + O(\rho). \qquad (4.32)
\end{aligned}$$

By (4.31) and (4.32), we obtain

$$\left\|\begin{bmatrix} x_{k+1} - x_* \\ \overline{\lambda}_{k+1} - \overline{\lambda}_* \end{bmatrix}\right\| \le (\overline{C} + \widehat{C}) \left\|\begin{bmatrix} x_k - x_* \\ \overline{\lambda}_k - \overline{\lambda}_* \end{bmatrix}\right\|^2 + O(\|\widehat{\lambda}_k - \overline{\lambda}_k\|) + O(\rho). \qquad (4.33)$$

Letting $C := \overline{C} + \widehat{C}$, we have shown (4.30). ∎

4.3.2. Local convergence to a KKT point. We now consider the local convergence of SQuID in the neighborhood of a KKT point for (3.1) satisfying Assumption 4.20. Our first result shows that in the neighborhood of a solution point, subproblem (3.7) yields a linearly feasible search direction, the penalty parameter remains constant, and the multipliers are not modified outside of the QO solves.

LEMMA 4.26. *Suppose Assumption* 4.20 *holds. Then, for all sufficiently large k with* $\|(x_k, \overline{\lambda}_k) - (x_*, \overline{\lambda}_*)\|$ *and* $\|(x_k, \widehat{\lambda}_k) - (x_*, \widehat{\lambda}_*)\|$ *each sufficiently small we have the following:*

(a) *A solution of* (3.7) *has* $(\overline{r}_k, \overline{s}_k, \overline{t}_k) = 0$, *yielding* $\mathcal{E}_k = \mathcal{E}$ *and* $\mathcal{I}_k = \mathcal{I}$.

(b) ρ_k *is not decreased by* (3.13), (3.18), *or* (3.19), *and the multipliers* $\widehat{\lambda}_k$ *are not modified by* (3.14).

Proof. The proof of part (a) is similar to the proof of Lemma 4.23(a). That is, under Assumption 4.20 (which means that Assumption 4.19 holds), a solution of (3.7) with $(x_k, \overline{\lambda}_k)$ sufficiently close to $(x_*, \overline{\lambda}_*)$ has $\overline{d}_k$ yielding the same sets of positive, zero, and negative-valued equality and violated, active, and strictly satisfied inequality constraints as x_*. In this case, $\mathcal{Z}_* = \mathcal{E}$ and $\mathcal{S}_* \cup \mathcal{A}_* = \mathcal{I}$, so $(\overline{r}_k, \overline{s}_k, \overline{t}_k) = 0$.

Now consider part (b). If x_k is feasible, then $v(x_k) = 0$ and (3.12) is violated. On the other hand, if x_k is infeasible, then we have $\Delta l(\overline{d}_k, x_k) = v(x_k)$ by part (a), which implies (3.12) is violated again. Overall, these conclusions imply that (3.13) and (3.14) are both not triggered. As for (3.18) and (3.19), every time either of these updates is triggered, ρ_k is at least reduced by a fraction of its current value. Therefore, if either of these updates is triggered an infinite number of times, then we would have $\rho_k \to 0$. However, under Assumption 4.20 we have $\rho_k \to \rho_* > 0$, so for all sufficiently large k, ρ_k is not decreased by either update. ∎

Our second result is similar to Lemma 4.23; again, recall [35].

LEMMA 4.27. *Suppose Assumption* 4.20 *holds. Then, for all sufficiently large k with* $\|(x_k, \overline{\lambda}_k) - (x_*, \overline{\lambda}_*)\|$ *and* $\|(x_k, \widehat{\lambda}_k) - (x_*, \widehat{\lambda}_*)\|$ *each sufficiently small, there is a local solution for* (3.9) *such that* $\widehat{d}_k$ *yields the same sets of active and strictly satisfied inequality constraints as* x_*. *Moreover,* $(\widehat{d}_k, \widehat{\lambda}_{k+1})$ *satisfies*

$$\begin{bmatrix} H(x_k, \rho_*, \widehat{\lambda}_k) & \nabla c^{\mathcal{E} \cup \mathcal{A}_*}(x_k) \\ \nabla c^{\mathcal{E} \cup \mathcal{A}_*}(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} \widehat{d}_k \\ \widehat{\lambda}_{k+1}^{\mathcal{E} \cup \mathcal{A}_*} \end{bmatrix} = -\begin{bmatrix} \rho_* \nabla f(x_k) \\ c^{\mathcal{E} \cup \mathcal{A}_*}(x_k) \end{bmatrix} \qquad (4.34)$$

and

$$\widehat{\lambda}_{k+1}^{\mathcal{A}_*} > 0 \quad \textit{and} \quad \widehat{\lambda}_{k+1}^{\mathcal{S}_*} = 0. \tag{4.35}$$

Proof. By Lemma 4.26, we have $\mathcal{E}_k = \mathcal{E}$ and $\mathcal{I}_k = \mathcal{I}$ under the conditions of this lemma. Thus, with $\widehat{d}_k = 0$, the optimality conditions (4.2) reduce to (3.4), so (4.2) is solved at $(x_*, \widehat{\lambda}_*)$ by $(d, \lambda) = (0, \widehat{\lambda}_*)$. By (4.2d), $\widehat{\lambda}_*^{\mathcal{S}_*} = 0$. Hence, by (4.2a) and the definition of $\mathcal{A}_*$, the linear system (4.34) is satisfied at $(x_*, \widehat{\lambda}_*)$ by $(d, \lambda^{\mathcal{E}\cup\mathcal{A}_*}) = (0, \widehat{\lambda}_*^{\mathcal{E}\cup\mathcal{A}_*})$. Under Assumptions 4.19(a) and 4.20(c), the matrix in (4.34) is nonsingular at $(x_*, \widehat{\lambda}_*)$, and hence the solution of (4.34) varies continuously in a neighborhood of $(x_*, \widehat{\lambda}_*)$. In addition, under Assumption 4.20(c), $H(x_k, \rho_*, \widehat{\lambda}_k)$ in (4.34) is positive definite on the null space of $\nabla c^{\mathcal{E}\cup\mathcal{A}_*}(x_k)^T$ in a neighborhood of $(x_*, \widehat{\lambda}_*)$.

It follows from the conclusions in the previous paragraph that for all $(x_k, \widehat{\lambda}_k)$ sufficiently close to $(x_*, \widehat{\lambda}_*)$, the solution $(\widehat{d}_k, \widehat{\lambda}_{k+1}^{\mathcal{E}\cup\mathcal{A}_*})$ to (4.34) is sufficiently close to $(0, \widehat{\lambda}_*^{\mathcal{E}\cup\mathcal{A}_*})$ such that it satisfies

$$\widehat{\lambda}_{k+1}^{\mathcal{A}_*} > 0 \quad \text{and} \quad c^{\mathcal{S}_*}(x_k) + \nabla c^{\mathcal{S}_*}(x_k)^T \widehat{d}_k < 0.$$

By construction, such a solution also satisfies (4.2) together with $\widehat{\lambda}_{k+1}^{\mathcal{S}_*} = 0$. Therefore, $(\widehat{d}_k, \widehat{\lambda}_{k+1})$ is a KKT point of subproblem (3.9), and, as revealed above, it identifies the same sets of active and strictly satisfied inequality constraints as x_*. ∎

We are now prepared to prove our main theorem concerning the local convergence of SQuID in the neighborhood of KKT points for (3.1).

THEOREM 4.28. *Suppose Assumption* 4.20 *holds. Then, for all large k with* $\|(x_k, \overline{\lambda}_k) - (x_*, \overline{\lambda}_*)\|$ *and* $\|(x_k, \widehat{\lambda}_k) - (x_*, \widehat{\lambda}_*)\|$ *each sufficiently small,* $(\widehat{d}_k, \widehat{\lambda}_{k+1})$ *is obtained via* (4.34)*,* $d_k \leftarrow \widehat{d}_k$*, and*

$$\left\| \begin{bmatrix} x_{k+1} - x_* \\ \widehat{\lambda}_{k+1} - \widehat{\lambda}_* \end{bmatrix} \right\| \leq C \left\| \begin{bmatrix} x_k - x_* \\ \widehat{\lambda}_k - \widehat{\lambda}_* \end{bmatrix} \right\|^2 \tag{4.36}$$

for some constant $C > 0$ *independent of* k*.*

Proof. By Lemma 4.27, under the conditions of the theorem, $(\widehat{d}_k, \widehat{\lambda}_{k+1})$ generated by subproblem (3.9) can be obtained via (4.34) and (4.35). This implies that d_k is a linearly feasible direction, so $w_k \leftarrow 0$ and $d_k \leftarrow \widehat{d}_k$. Therefore, since $H(x_k, \rho_*, \widehat{\lambda}_k) = G(x_k, \rho_*, \widehat{\lambda}_k^{\mathcal{E}\cup\mathcal{A}_*})$ in such cases, (4.34) (with $\widehat{d}$ interchanged with d_k) constitutes a Newton iteration applied to the nonlinear system $F(x, \rho_*, \lambda^{\mathcal{E}\cup\mathcal{A}_*}) = 0$ at $(x_k, \rho_*, \widehat{\lambda}_k)$. We can now apply standard Newton analysis. Under Assumption 4.18 we have that F is continuously differentiable and F' is Lipschitz continuous in a neighborhood of $(x_*, \rho_*, \widehat{\lambda}_*^{\mathcal{E}\cup\mathcal{A}_*})$. Moreover, under Assumptions 4.19(a) and 4.20(c), the matrix F' is nonsingular at $(x_*, \rho_*, \widehat{\lambda}_*^{\mathcal{E}\cup\mathcal{A}_*})$, so its inverse exists and is bounded in norm in a neighborhood of $(x_*, \rho_*, \widehat{\lambda}_*^{\mathcal{E}\cup\mathcal{A}_*})$. By [16, Theorem 5.2.1], if $(x_k, \widehat{\lambda}_k)$ is sufficiently close to $(x_*, \widehat{\lambda}_*)$, then we have that (4.36) holds true. ∎

5. Numerical experiments. In this section, we summarize the performance of SQuID as it was employed to solve collections of feasible and infeasible problem instances. Our code is a prototype MATLAB implementation of Algorithm 1.

Mention of a few specifications of our implementation are appropriate before we present our numerical results. First, in order to avoid numerical issues caused by poor scaling of the problem functions, each function was scaled so that the ℓ_∞-norm of its gradient at the initial point was no larger than a given constant $g_{\max} > 0$. Moreover, our termination conditions are defined to take into account the magnitudes of the quantities involved in the computation of the optimality and feasibility errors. Specifically, we terminate and declare that an optimal solution has been found if

$$\mathcal{R}_{opt}(x_k, \rho_k, \widehat{\lambda}_{k+1}) \leq \gamma \max\{\chi_{opt,k}, 1\} \quad \text{and} \quad v_{inf}(x_k) \leq \gamma \max\{v_{inf}(x_0), 1\}, \tag{5.1}$$

where $\gamma > 0$ is a given constant,

$$\chi_{opt,k} := \max\{\rho_k, \|\nabla f(x_k)\|_\infty, \|\nabla c^{\mathcal{E}}(x_k)\|_\infty, \|\nabla c^{\mathcal{I}}(x_k)\|_\infty, \|\widehat{\lambda}^{\mathcal{E}}_{k+1}\|_\infty, \|\widehat{\lambda}^{\mathcal{I}}_{k+1}\|_\infty\},$$
$$\text{and} \quad v_{inf}(x_k) := \max\{\|c^{\mathcal{E}}(x_k)\|_\infty, \|\max\{c^{\mathcal{I}}(x_k), 0\}\|_\infty\}.$$

We terminate and declare that an infeasible stationary point has been found if

$$\mathcal{R}_{inf}(x_k, \overline{\lambda}_{k+1}) \leq \gamma \max\{\chi_{inf,k}, 1\},\ v_{inf}(x_k) > \gamma \max\{v_{inf}(x_0), 1\},\ \text{and}\ \rho_k \leq \overline{\rho}, \tag{5.2}$$

where $\overline{\rho} > 0$ is a given constant and

$$\chi_{inf,k} := \max\{\|\nabla c^{\mathcal{E}}(x_k)\|_\infty, \|\nabla c^{\mathcal{I}}(x_k)\|_\infty, \|\overline{\lambda}^{\mathcal{E}}_{k+1}\|_\infty, \|\overline{\lambda}^{\mathcal{I}}_{k+1}\|_\infty\}.$$

Despite the fact that Theorem 4.14 implies that we do not necessarily need $\rho_k \to 0$ when converging to an infeasible stationary point, we only terminate and declare infeasibility when ρ_k is sufficiently small, as specified in (5.2). This may lead to extra iterations being performed before infeasibility is declared, but aids the algorithm in avoiding declarations of infeasibility when applied to problem instances that are actually feasible. Since ρ_k is decreased rapidly in the neighborhood of an infeasible stationary point due to (3.12), the additional cost is worthwhile. We also take into account the scaling of the problem functions when considering whether a given point is sufficiently feasible so that subproblem (3.7) may be skipped. Specifically, if $v_{inf}(x_k) \leq \overline{\gamma} \max\{v_{inf}(x_0), 1\}$ for some $\overline{\gamma} > 0$, then we save computational expense by approximating the solution of subproblem (3.7) with $\overline{d}_k \leftarrow 0$ and $\overline{\lambda}_{k+1} \leftarrow \overline{\lambda}_k$.

Our implementation requires that subproblems (3.7) and (3.9) are convex, so we modify $H(x_k, 0, \overline{\lambda}_k)$ and $H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)$, if necessary, to make them positive definite. We do this by iteratively adding multiples of the identity matrix until the smallest computed eigenvalue is sufficiently positive. Specifically, if one of these matrices needs to be modified at iteration k, then with some $\xi > 1$ and an initial increment μ_k, we add $\mu_k I, \xi\mu_k I, \xi^2\mu_k I, \ldots$ until the smallest eigenvalue of the matrix is larger than a positive parameter $\mu_{\min}$. We then set $\mu_{k+1} \leftarrow \max\{\mu_{\min}, \psi\mu_k\}$ for some $\psi \in (0, 1)$ to help save the computational expense of computing eigenvalues and modifying the matrix in the following iteration. If a matrix does not need to be modified during iteration k, then we reset $\mu_{k+1} \leftarrow \mu_{\min}$ for the following iteration. (We maintain different increments, μ^0_k and μ^ρ_k, for $H(x_k, 0, \overline{\lambda}_k)$ and $H(x_k, \widehat{\rho}_k, \widehat{\lambda}_k)$, respectively.) Of course, these modifications may slow the local convergence rate of the algorithm in the neighborhood of optimal solutions or infeasible stationary points that may fail to satisfy a strict second-order sufficiency condition, but they allow a prototype implementation such as ours to be well-defined when applied to nonconvex problems.

For computing the weight w_k required in (3.11) for iteration k, we initialize $w_k \leftarrow 0$ and check if (3.10) holds for $d_k \leftarrow \widehat{d}_k$. If it does, then the algorithm continues with these values for the weight and step, and otherwise we apply a bisection method to attempt to find the smallest root w_k of

$$\Psi(w) = \Delta l(w\overline{d}_k + (1-w)\widehat{d}_k; x_k) - \beta \Delta l(\overline{d}_k; x_k).$$

Since when $\overline{d}_k \neq 0$ we have $\Psi(1) > 0$ and $\Psi(0) < 0$, the bisection method is well-defined and there exists $w_k \in (0,1)$ such that $\Psi(w_k) = 0$. (Note that if $\overline{d}_k = 0$, then both $\overline{d}_k$ and $\widehat{d}_k$ will be linearly feasible, and so (3.10) is satisfied with $w_k \leftarrow 0$.) We terminate the bisection method when the width of the current interval is less than 10^{-8}. This and our choice of $\omega \leftarrow (1-10^{-18})$ ensure that we always compute $w_k < \omega$, effectively making this threshold value inconsequential for our numerical experiments.

As final notes on the particulars of our implementation, we remark that (3.7) and (3.9) are solved using MATLAB's built-in `quadprog` routine. Also, the parameter values used are those provided in Table 1.

Table 1
Input parameters for a prototype MATLAB implemenation of Algorithm 1.

Parameter	ρ_0	β	θ	κ_ρ	κ_λ	ϵ	δ	η
Value	10^{-1}	10^{-2}	10^{-1}	10	10	10^{-2}	5×10^{-1}	10^{-8}
Parameter	$g_{\max}$	γ	$\overline{\rho}$	$\overline{\gamma}$	ξ	μ_0	ψ	$\mu_{\min}$
Value	10^2	10^{-6}	10^{-8}	10^{-8}	2	10^{-4}	10^{-1}	10^{-4}

We tested our implementation on 123 of the Hock–Schittkowski problems [25] available as AMPL models [20].[1] (Problems `hs068` and `hs069` were excluded from the original set of 125 problems as the required external function was not compiled.) The original versions of all of these problems are feasible, but we created a corresponding set of infeasible problems by adding the incompatible constraints $x_1 \leq 0$ and $x_1 \geq 1$, where x_1 is the first variable in the problem statement.

Termination results for our implementation applied to these problems are shown in Tables 2 and 3, which contain statistics for the feasible and infeasible problems, respectively. In Table 2, the "Succeed" column reveals the number and percentage of problems for which a point satisfying (5.1) was obtained, and the "Infeasible" column reveals those statistics for problems for which a point satisfying (5.2) was obtained. Similarly, the "Succeed" column in Table 3 reveals the number and percentage of problems for which a point satisfying (5.2) was obtained, and the "Feasible" column reveals those statistics for problems for which a point satisfying (5.1) was obtained. In Tables 2 and 3, a termination result in the latter of these two columns represents a situation where the algorithm failed to solve the problem correctly. Any time the algorithm fails to terminate within 10^3 iterations, the algorithm is deemed to "Fail." (Problem `hs112x` was excluded in the set of feasible problems due to a function evaluation error that occurred during the run.)

Table 2
Performance statistics of SQuID on feasible problems.

Problem type	Succeed	Fail	Infeasible	Total
Feasible	110 (90.16%)	11 (9.02%)	1 (0.82%)	122

From Tables 2 and 3, one can see that our code consistently attained a success rate of at least 90%, which is strong for a prototype implementation. In fact, for

[1] http://orfe.princeton.edu/~rvdb/ampl/nlmodels/cute/

TABLE 3
Performance statistics of SQuID on infeasible problems.

Problem type	Succeed	Fail	Feasible	Total
Infeasible	111 (90.24%)	12 (9.76%)	0 (0.0%)	123

most of the failures and for the feasible problem that was reported to be infeasible, we found the problems to be very nonconvex. This led to excessive modifications of the Hessian matrices, and in many cases search directions that were poorly scaled. The results may be improved with a more sophisticated Hessian modification strategy and/or the incorporation of second-order correction steps.

We conclude our discussion of this set of numerical experiments by illustrating the local convergence behavior of SQuID on these sets of test problems. For those instances that are successfully solved within the iteration limit, we store the logarithms of $\mathcal{R}_{opt}$ and $\mathcal{R}_{inf}$ for the last ten iterations for the feasible and infeasible problem instances and plot them in Figures 1 and 2, respectively. In the plots, T represents the last iteration for each run. (If a given problem is solved in fewer than ten iterations, then its corresponding plot begins in the middle of the graph.) In Figures 1 and 2, one can see that most of the curves turn significantly downward on the right-hand side of the graph. The curves with a slope less than -1 over the last iterations indicate local superlinear convergence, and the curves with slope less than -2 indicate quadratic convergence. One finds that many of the curves possess slopes of this type, providing empirical evidence for the convergence results in section 4.

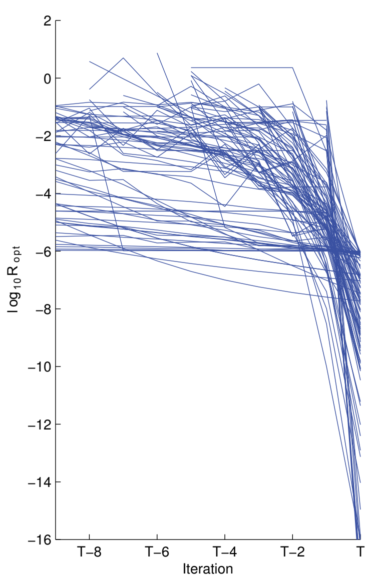

FIG. 1. $\log_{10} \mathcal{R}_{opt}$ *for the last ten iterations of SQuID applied to feasible instances.*

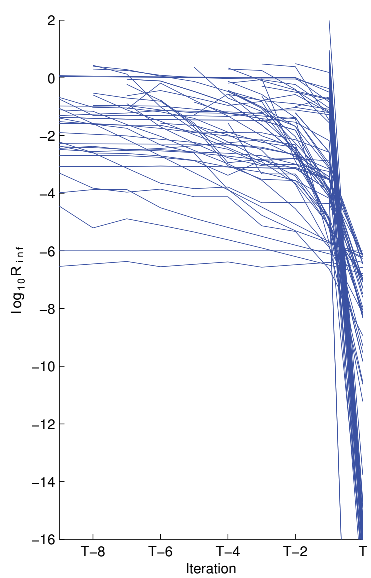

FIG. 2. $\log_{10} \mathcal{R}_{inf}$ *for the last ten iterations of SQuID applied to infeasible instances.*

We close this section with a comparison between SQuID and the algorithm proposed in [6] when applied to solve the infeasible problems presented in [6]. As previously mentioned in section 2, the algorithm in [6] represents an immediate predecessor of SQuID. That algorithm also possesses superlinear convergence guarantees, but, unlike SQuID, suffers from the disadvantage that more than two QO subproblem solves may be required in each iteration. After modifying the input parameters in our implementation of SQuID so that they match those used in [6]—e.g., in [6], the initial penalty parameter was set to 1—we obtained the results presented in Table 4. (Here, the "Iter." columns indicate the numbers of (nonlinear) iterations performed and the "QOs" columns indicate the number of QO subproblems solved prior to termination.) It is clear in these results that both algorithms detect infeasibility (or locate an optimal solution in the case of problem "batch") in few iterations, but SQuID typically requires fewer QO solves. (The only exception is the problem "robot." This problem is nonconvex and the performance of both algorithms varies depending on the input parameters that affect the modifications of the Hessian approximations to make them positive definite.) These results provide evidence for our claim that SQuID yields consistent improvement over the algorithm in [6]. That is, SQuID possesses similar theoretical convergence guarantees, but yields better practical performance by limiting the number of QO subproblem solves per iteration.

Table 4
Performance measures for test problems in [6].

	unique		robot		isolated		batch		batch1		nactive	
Alg.	Iter.	QOs	Iter.	QOs	Iter.	QOs	Iter.	QOs	Iter.	QOs	Iter.	QOs
SQuID	9	19	27	55	9	19	10	22	15	31	7	15
Ref [6]	9	24	13	34	7	20	11	28	15	40	6	17

6. Conclusion. In this paper, we have proposed, analyzed, and tested a SQO method that possesses global and fast local convergence guarantees for both feasible and infeasible problem instances. Novelties of the algorithm are its unique two-phase approach and carefully designed updating strategy for the penalty parameter. The subproblems in each phase and the penalty parameter update are designed to strike a balance between moving toward feasibility and optimality in each iteration. Near an optimal point satisfying common assumptions, the penalty parameter remains constant and the algorithm reduces to a classical SQO method, yielding fast local convergence. Similarly, near an infeasible stationary point, the penalty parameter is reduced sufficiently and quickly to yield fast infeasibility detection.

The convergence properties that we have proved for our algorithm were illustrated empirically on test sets of feasible and infeasible problems. We remark, however, that there remain various practical issues that one faces when considering an implementation of SQuID. As with any SQO method, the primary concern is the efficiency of the QO subproblem solver. This is especially the case when one wishes to use exact second-order derivative information and the resulting Hessian matrices are not positive definite. We have employed a Hessian modification strategy in our numerical experiments, but as for any SQO method that employs such a strategy, these modifications are cumbersome in large-scale settings and may inhibit superlinear convergence. We leave it a subject of future research to investigate ways in which inexactness can be incorporated into the subproblem solves and negative curvature can be handled, knowing that the algorithm and analysis presented in this paper provides a strong backbone for rapid infeasibility detection when such additional features are developed.

REFERENCES

[1] P. T. Boggs, J. W. Tolle, and P. Wang, *On the local convergence of quasi-Newton methods for constrained optimization*, SIAM J. Control Optim., 20 (1982), pp. 161–171.
[2] J. V. Burke, *Descent methods for composite nondifferentiable optimization problems*, Math. Programming, 33 (1985), pp. 260–279.
[3] J. V. Burke, *A sequential quadratic programming algorithm for potentially infeasible mathematical programs*, J. Math. Anal. Appl., 139 (1989), pp. 319–351.
[4] J. V. Burke, *A robust trust region method for constrained nonlinear programming problems*, SIAM J. Optim., 2 (1992), pp. 325–347.
[5] J. V. Burke and S.-P. Han, *A robust sequential quadratic programming method*, Math. Programming, 43 (1989), pp. 277–303.
[6] R. H. Byrd, F. E. Curtis, and J. Nocedal, *Infeasibility detection and SQP methods for nonlinear optimization*, SIAM J. Optim., 20 (2010), pp. 2281–2299.
[7] R. H. Byrd, J. C. Gilbert, and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming*, Math. Program., 89 (2000), pp. 149–185.
[8] R. H. Byrd, N. I. M. Gould, J. Nocedal, and R. A. Waltz, *An algorithm for nonlinear optimization using linear programming and equality constrained subproblems*, Math. Program., 100 (2004), pp. 27–48.
[9] R. H. Byrd, N. I. M. Gould, J. Nocedal, and R. A. Waltz, *On the convergence of successive linear-quadratic programming algorithms*, SIAM J. Optim., 16 (2005), pp. 471–489.
[10] R. H. Byrd, G. López-Calva, and J. Nocedal, *A line search exact penalty method using steering rules*, Math. Program., 133 (2012), pp. 39–73.
[11] R. H. Byrd, J. Nocedal, and R. A. Waltz, *Steering exact penalty methods for nonlinear programming*, Optim. Methods Softw., 23 (2008), pp. 197–213.
[12] R. M. Chamberlain, M. J. D. Powell, C. Lemaréchal, and H. C. Pedersen, *The watchdog technique for forcing convergence in algorithms for constrained optimization*, Math. Programming Stud., 16 (1982), pp. 1–17.
[13] C. M. Chin and R. Fletcher, *On the global convergence of an SLP-filter algorithm that takes EQP steps*, Mathematical Programming, 96 (2003), pp. 161–177.
[14] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Canad. Math. Soc. Ser. Monogr. Adv. Texts, John Wiley & Sons, New York, 1983.
[15] D. Cores and R. A. Tapia, *Perturbation lemma for the Newton method with application to the SQP Newton method*, J. Optim. Theory Appl., 97 (1998), pp. 271–280.
[16] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Classics Appl. Math. 16, SIAM, Philadelphia, 1996.
[17] F. Facchinei and S. Lucidi, *Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problems*, J. Optim. Theory Appl., 85 (1995), pp. 265–289.
[18] F. Fletcher and E. Sainz de la Maza, *Nonlinear programming and nonsmooth optimization by successive linear programming*, Math. Programming, 43 (1989), pp. 235–256.
[19] R. Fletcher, *Practical Methods of Optimization*, 2nd ed., John Wiley and Sons, Chichester, UK, 1987.
[20] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL: A Modeling Language for Mathematical Programming*, Brooks/Cole, Pacific Grove, CA, 2002.
[21] M. P. Friedlander, N. I. M. Gould, and S. Leyffer, *A Filter Active-set Trust-region Method*, Technical report, Preprint ANL/MCS-P1456-097. Argonne National Laboratory, Lemont, IL, 2007.
[22] N. I. M. Gould and D. P. Robinson, *A second derivative SQP method: Global convergence*, SIAM J. Optim., 20 (2010), pp. 2023–2048.
[23] N. I. M. Gould and D. P. Robinson, *A second derivative SQP method: Local convergence and practical issues*, SIAM J. Optim., 20 (2010), pp. 2049–2079.
[24] S. P. Han, *A globally convergent method for nonlinear programming*, J. Optim. Theory Appl., 22 (1977), pp. 297–309.
[25] W. Hock and K. Schittkowski, *Test examples for nonlinear programming codes*, J. Optim. Theory Appl., 30 (1980), pp. 127–129.
[26] F. John, *Extremum Problems with Inequalities as Side Conditions*, in Studies and Essays, Courant Anniversary Volume, K. O. Friedrichs, O. E. Neugebauer, and J. J. Stoker, eds., Wiley (Interscience), New York, 1948, pp. 187–204.
[27] W. Karush, *Minima of Functions of Several Variables with Inequalities as Side Constraints*, Master's thesis, Department of Mathematics, University of Chicago, Chicago, IL, 1939.

[28] H. W. Kuhn and A. W. Tucker, *Nonlinear Programming*, in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman, ed., University of California Press, Berkeley, CA, 1951, pp. 481–492.

[29] J. L. Morales, J. Nocedal, and Y Wu, *A sequential quadratic programming algorithm with an additional equality constrained phase*, IMA J. Numer. Anal., 32 (2012), pp. 553–579.

[30] X. W. Liu and Y. Yuan, *A null-space primal-dual interior-point algorithm for nonlinear optimization with nice convergence properties*, Math. Program., 125 (2010), pp. 163–193.

[31] N. Maratos, *Exact Penalty Function Algorithms for Finite Dimensional and Control Optimization Problems*, Ph.D. thesis, University of London, London, England, 1978.

[32] J. Nocedal, F. Öztoprak, and R. A. Waltz, *An interior point method for nonlinear programming with infeasibility detection capabilities*, Optim. Methods Softw., 29 (2014), pp. 837–854.

[33] P. M. Pardalos and S. A. Vavasis, *Quadratic programming with one negative eigenvalue is NP-hard*, J. Global Optim., 1 (1991), pp. 15–22.

[34] M. J. D. Powell, *A Fast Algorithm for Nonlinearly Constrained Optimization Calculations*, in Numerical Analysis, Lecture Notes in Math. 630, Springer, Berlin, 1978, pp. 144–157.

[35] S. M. Robinson, *A quadratically-convergent algorithm for general nonlinear programming problems*, Math. Programming, 3 (1972), pp. 145–156.

[36] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw-Hill New York, 1976.

[37] R. B. Wilson, *A Simplicial Algorithm for Concave Programming*, Ph.D. thesis, Graduate School of Business Administration, Harvard University, Cambridge, MA, 1963.