SIAM J. OPTIM.
Vol. 24, No. 1, pp. 175–209

A FILTER METHOD WITH UNIFIED STEP COMPUTATION FOR NONLINEAR OPTIMIZATION*

NICHOLAS I. M. GOULD†, YUELING LOH‡, AND DANIEL P. ROBINSON‡

Abstract. We present a filter line search method for solving general nonlinear and nonconvex optimization problems. The method is of the filter variety but uses a robust (always feasible) subproblem based on an exact penalty function to compute a search direction. This contrasts traditional filter methods that use a (separate) restoration phase designed to reduce infeasibility until a feasible subproblem is obtained. Therefore, an advantage of our approach is that every trial step is computed from subproblems that value reducing both the constraint violation *and* the objective function. Moreover, our step computation involves subproblems that are computationally tractable and utilize second derivative information when it is available. The formulation of each subproblem and the choice of weighting parameter is crucial for obtaining an efficient, robust, and practical method. Our strategy is based on steering methods designed for exact penalty functions but is fortified with a trial step convexification scheme that ensures that a single quadratic optimization problem is solved per iteration. Moreover, we use local feasibility estimates that emerge during the steering process to define a new and improved margin (envelope) of the filter. Under common assumptions, we show that the iterates converge to a local first-order solution of the optimization problem from an arbitrary starting point.

Key words. filter, restoration phase, large-scale, sequential quadratic programming, nonlinear programming

AMS subject classifications. 49M05, 49M15, 65K05, 65K10, 65K15

DOI. 10.1137/130920599

1. Introduction.

This paper considers the general nonlinear optimization problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \ f(x) \ \ \text{subject to} \ \ c(x) \geq 0, \tag{1.1}$$

where both the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint function $c : \mathbb{R}^n \to \mathbb{R}^m$ are assumed to be twice continuously differentiable. We seek a first-order KKT point (x, y) that satisfies

$$F_{\text{KKT}}(x,y) := \begin{pmatrix} g(x) - J(x)^T y \\ \min\left[c(x), y\right] \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{1.2}$$

where $g(x) := \nabla f(x) \in \mathbb{R}^n$ is the gradient of the objective function, $J(x) := \nabla c(x) \in \mathbb{R}^{m\times n}$ is the Jacobian of the constraint function, y is the Lagrange multiplier vector, and the minimum is taken componentwise. Our algorithm may easily handle constraints with general lower/upper bounds and handle equality constraints directly, i.e., it does not replace them with pairs of inequality constraints. Problems of this type arise naturally in many areas, including optimal control [2, 3, 6, 25, 32], resource

*Received by the editors May 10, 2013; accepted for publication (in revised form) October 28, 2013; published electronically January 28, 2014. This research was supported by National Science Foundation grant DMS-1217153 and EPSRC grant EP/I013067/1.

http://www.siam.org/journals/siopt/24-1/92059.html

†Scientific Computing Department, Rutherford Appleton Laboratory, Chilton, OX11-0QX, UK (nick.gould@stfc.ac.uk).

‡Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218 (yueling.loh@jhu.edu, daniel.p.robinson@gmail.com).

allocation [1, 31], solution of equilibrium models [20, 38], and structural engineering [5, 34].

Popular methods for solving (1.1) can broadly be characterized as interior-point or active-set methods. Interior-point algorithms [44, 46, 47] offer polynomial-time complexity bounds in many cases and readily scale up to problems involving millions of variables. Their main disadvantage is the inability to use effectively a good initial estimate of a solution. In fact, many interior-point methods immediately move the initial guess into the strict interior of the feasible region. It is from this interior location that future iterates are forced to remain and it justifies the name "interior-point" methods; more modern "infeasible" interior-point methods avoid this defect to some degree.

Active-set methods [10, 16, 27, 28, 29, 30, 37, 39] complement interior-point methods since they naturally utilize information derived from a good estimate of a solution. In fact, if the optimal active set (the set containing those constraints satisfied as equalities at a solution) was known in advance, then problem (1.1) could be solved as an equality constrained problem and its combinatorial nature would be eliminated. It is precisely this property that makes active-set methods widely used to solve the previously mentioned class of problems. The main weakness of active-set algorithms is that each subproblem typically requires the solution of a linear or quadratic program, which is often expensive when compared to interior-point methods that require a single linear system solve per iteration.

In this paper we describe an active-set method that generates a sequence of iterates from the solutions of subproblems defined by local models of the nonlinear problem functions. The subproblems are always feasible since they are based on an exact penalty function. To ensure that these models result in productive steps, we use steering techniques [11] to adaptively adjust the weighting (penalty) parameter. In contrast to original steering methods, we use a step convexification procedure similar to [7, 18] to avoid solving multiple quadratic programs during each iteration.

To provide convergence guarantees, we must include a mechanism for determining when one point is "better" than another. A merit (penalty) function or a filter is among the most common tools used for this purpose. A merit function combines the objective function and a measure of constraint violation into a single function, whereby their individual perceived importance is determined by a weighting parameter. The quality of competing points is then measured by comparing their respective merit function values. A potential weakness is that the quality of iterates depends on the value of the weighting parameter, which can make step acceptance sensitive to its value. In part, filter methods surfaced to mitigate this parameter dependence. In the context of nonlinear optimization, filter methods were introduced by Fletcher, Leyffer, and Toint [22, 23] and have since been rather popular [13, 14, 15, 21, 24, 44]. A filter views problem (1.1) as a multicriterion optimization problem consisting of minimizing the objective function and minimizing some measure of the constraint violation, with certain preference given to the latter. Roughly, a trial iterate is then considered acceptable if it has a smaller value of *either* the objective function or the constraint violation compared to the previously encountered points. Consequently, it is often the case that filter methods accept more iterates and perform better. It should be mentioned, however, that every known provably convergent filter method has a weak dependence between these two criteria that is embedded in the step acceptance criteria. In fact, this observation partly motivated the work on *flexible* penalty methods by Curtis and Nocedal [19]. They describe how a *single* element filter is essentially equivalent to the union of points acceptable to the ℓ_1-penalty function defined over an interval of weighting parameter values.

A great disadvantage of filter methods is that they (traditionally) require the use of a restoration phase. A restoration phase is (typically) entered when the subproblem used to compute trial steps is infeasible; some algorithms, e.g., [46], enter the restoration phase for additional reasons. When this situation occurs, the restoration phase is triggered and a sequence of iterates focused on reducing the constraint violation is computed until the desired subproblem becomes feasible. During this phase, the objective function is essentially *ignored*, which is highly undesirable from both a practical and a computational perspective.

Our active-set method is globalized by using a filter but never needs to enter a (traditional) restoration phase. This is accomplished by using subproblems that are always feasible and, in certain instances, allow for the acceptance of iterates that decrease both the exact penalty function and the constraint violation. In essence, we replace an undesirable restoration phase with an attractive penalty phase. Thus, we combine ideas from both filter and penalty methods to formulate a robust and effective method; we believe this further builds upon the basic observations in [19].

This paper contains three main contributions. First, we present a filter method that avoids a traditional and highly undesirable restoration phase. To this end, we utilize subproblems based on exact penalty functions that are *always* feasible and formed from models of *both* the objective function and constraint violation. Second, our method incorporates second derivative information without requiring global minimizers of nonconvex constrained subproblems (cf. [21]). Our step computation is most similar to [28, 30], which was described in the context of line search and trust-region penalty methods. Third, we use local feasibility estimates that emerge during the steering step computation to define a new and improved margin (envelope) of the filter. This allows us to define an adaptive and practical margin.

Our work is not the only method designed to resolve weaknesses in traditional filter methods. Chen and Goldfarb [12] presented an interior point method that uses two penalty functions to determine step acceptance: a piecewise linear penalty function whose break points are essentially elements in the filter, and the ℓ_2-penalty function. Under this scheme, a trial step is accepted if it provides sufficient reduction for either penalty function.

The remainder of this paper is organized as follows. In section 2 we describe the algorithm in detail, and in section 3 we prove that it is well-posed. We provide convergence results in section 4 and conclude with final remarks in section 5.

2. A filter sequential quadratic programming method. In this section we describe our new filter sequential quadratic programming method, FiSQP. The algorithm is iterative and relies on computing trial steps from carefully constructed subproblems. These subproblems and the resulting trial steps are explained in sections 2.1–2.6. In section 2.7 we introduce the filter construct and related terminology; we emphasize that acceptability to the filter is only a necessary condition for accepting a trial iterate. A full statement and description of the algorithm are given in section 2.8.

Our step computation is based on the ℓ_1-penalty function

$$\phi(x;\sigma) := f(x) + \sigma v(x), \tag{2.1}$$

where σ is a positive weighting parameter and the constraint violation at x is defined by

$$v(x) := \left\| [c(x)]^- \right\|_1 \quad \text{with} \quad [y]^- := \max(-y, 0) \tag{2.2}$$

and where the maximum is taken componentwise.

We use linear and quadratic model approximations of the objective function f given by

$$\begin{aligned} \ell^f(s;x) &:= f(x) + g(x)^T s \quad \text{and} \\ q^f(s;x,M) &:= \ell^f(s;x) + \tfrac{1}{2}s^T M s = f(x) + g(x)^T s + \tfrac{1}{2}s^T M s \end{aligned} \tag{2.3}$$

for a given symmetric matrix $M \in \mathbb{R}^{n\times n}$ and use a piecewise-linear approximation to the constraint violation function v given by

$$\ell^v(s;x) := \left\| [c(x) + J(x)s]^- \right\|_1$$

to form the following linear and quadratic models of ϕ:

$$\ell^\phi(s;x,\sigma) := \ell^f(s;x) + \sigma \ell^v(s;x) = f(x) + g(x)^T s + \sigma \left\| [c(x) + J(x)s]^- \right\|_1 \tag{2.4}$$

$$\begin{aligned} q^\phi(s;x,M,\sigma) &:= q^f(s;x,M) + \sigma \ell^v(s;x) \\ &= f(x) + g(x)^T s + \tfrac{1}{2}s^T M s + \sigma \left\| [c(x) + J(x)s]^- \right\|_1 . \end{aligned} \tag{2.5}$$

Using these models we may predict the change in v with the function

$$\Delta\ell^v(s;x) := \ell^v(0;x) - \ell^v(s;x) = \left\| [c(x)]^- \right\|_1 - \left\| [c(x) + J(x)s]^- \right\|_1 , \tag{2.6}$$

the change in f with the functions

$$\Delta\ell^f(s;x) := \ell^f(0;x) - \ell^f(s;x) = -g(x)^T s \quad \text{and} \tag{2.7a}$$

$$\begin{aligned} \Delta q^f(s;x,M) &:= q^f(0;x,M) - q^f(s;x,M) \\ &= \Delta\ell^f(s;x) - \tfrac{1}{2}s^T M s = -g(x)^T s - \tfrac{1}{2}s^T M s, \end{aligned} \tag{2.7b}$$

and the change in the penalty function ϕ with the functions

$$\begin{aligned} \Delta\ell^\phi(s;x,\sigma) &:= \ell^\phi(0;x,\sigma) - \ell^\phi(s;x,\sigma) = \Delta\ell^f(s;x) + \sigma\Delta\ell^v(s;x) \\ &= -g(x)^T s + \sigma \left(\left\| [c(x)]^- \right\|_1 - \left\| [c(x) + J(x)s]^- \right\|_1 \right) \end{aligned} \tag{2.8}$$

and

$$\begin{aligned} \Delta q^\phi(s;x,M,\sigma) &:= q^\phi(0;x,M,\sigma) - q^\phi(s;x,M,\sigma) = \Delta\ell^\phi(s;x,\sigma) - \tfrac{1}{2}s^T M s \\ &= -g(x)^T s - \tfrac{1}{2}s^T M s + \sigma \left(\left\| [c(x)]^- \right\|_1 - \left\| [c(x) + J(x)s]^- \right\|_1 \right) . \end{aligned} \tag{2.9}$$

For the remainder of this section, let (x_k, y_k) denote the current estimate of a solution to (1.1).

2.1. The steering step s_k^s. In order to strike a proper balance between reducing the objective function and the constraint violation, we compute a *steering* step s_k^s as a solution to the linear program

$$\underset{(s,r)\in\mathbb{R}^{n+m}}{\text{minimize}} \; e^T r \quad \text{subject to} \quad c_k + J_k s + r \ge 0, \;\; r \ge 0, \;\; \|s\|_\infty \le \delta_k, \tag{2.10}$$

where $c_k = c(x_k)$, $J_k = J(x_k)$, $\delta_k \in [\delta_{\min}, \delta_{\max}]$, and $0 < \delta_{\min} \le \delta_{\max} < \infty$. Problem (2.10) is equivalent to the nonsmooth problem

$$\underset{s\in\mathbb{R}^n}{\text{minimize}} \; \ell^v(s;x_k) \quad \text{subject to} \quad \|s\|_\infty \le \delta_k \tag{2.11}$$

since s solves (2.11) if and only if (s, r) solves (2.10), where $r = \max(-(c_k + J_k s), 0)$. Since $\ell^v(0; x_k) = v(x_k)$, ℓ^v is a convex function, and $s = 0$ is feasible for (2.11), it follows from (2.6) that $\Delta\ell^v(s_k^s; x_k) \geq 0$. The quantity $\Delta\ell^v(s_k^s; x_k)$ is the best local improvement in linearized constraint feasibility for steps of size δ_k.

All methods for nonconvex optimization may converge to an infeasible point that is a local minimizer of the constraint violation as measured by v. Points of this type are known as infeasible stationary points, which we now define by utilizing the steering subproblem.

DEFINITION 2.1 (infeasible stationary point). *The vector x^I is an infeasible stationary point if $v(x^I) > 0$ and $\Delta\ell^v(s^I; x^I) = 0$, where $s^I = \operatorname{argmin}_{s\in\mathbb{R}^n} \ell^v(s; x^I)$ subject to $\|s\|_\infty \leq \delta$ for some $\delta > 0$.*

2.2. The predictor step s_k^p. The *predictor* step is computed as the *unique* solution to one of the following strictly convex minimization problems:

$$
s_k^p = \begin{cases} \underset{s\in\mathbb{R}^n}{\operatorname{argmin}} \;\; f_k + g_k^T s + \tfrac{1}{2} s^T B_k s \;\; \text{subject to} \;\; c_k + J_k s \geq 0 & \text{(2.12a)} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \Delta\ell^v(s_k^s; x_k) = v(x_k), & \\ \underset{s\in\mathbb{R}^n}{\operatorname{argmin}} \;\; q^\phi(s; x_k, B_k, \sigma_k), \qquad\qquad \text{otherwise}, & \text{(2.12b)} \end{cases}
$$

where $\sigma_k > 0$ is the kth value of the penalty parameter, $f_k = f(x_k)$, $g_k = \nabla f(x_k)$, $c_k = c(x_k)$, $J_k = \nabla c(x_k)$, B_k is a positive-definite matrix that we are free to choose such that $B_k \approx \nabla_{xx}^2 L(x_k, y_k)$, and the Lagrangian L is defined by $L(x, y) = f(x) - c(x)^T y$. Analogous to the steering subproblem, the nonsmooth minimization problem (2.12b) is equivalent to the smooth problem

$$
\underset{s\in\mathbb{R}^n, r\in\mathbb{R}^n}{\text{minimize}} \;\; f_k + g_k^T s + \tfrac{1}{2} s^T B_k s + \sigma_k e^T r \;\; \text{subject to} \;\; c_k + J_k s + r \geq 0, \;\; r \geq 0, \tag{2.13}
$$

which is the problem solved in practice. We use y_k^{P} to denote the Lagrange multiplier vector for the constraint $c_k + J_k s \geq 0$ in (2.12a) and $c_k + J_k s + r \geq 0$ in (2.13) (equivalently (2.12b)). Possible choices for the positive-define matrix include (i) the trivial choice $B_k = I$, (ii) a scaled diagonal matrix based on the Barzilai–Borwein [4] method, (iii) quasi-Newton updates such as BFGS [40] and L-BFGS [33], and (iv) modified Cholesky factorizations [26, 43]. Moreover, relaxing the positive-definition assumption on B_k may also be possible provided conditions such as those used in [18] are enforced to ensure that sufficient descent directions are computed. Since this relaxation would introduce unnecessary complications into our algorithm and require the use of an indefinite QP solver, we will assume throughout that B_k is a positive-definite matrix.

The next result shows how convergence to KKT points may be deduced from the predictor problem.

LEMMA 2.2. *Suppose that x_* satisfies*

$$
\begin{aligned} & v(x_*) = 0 \quad \textit{and} \\ & \quad 0 = \underset{s\in\mathbb{R}^n}{\operatorname{argmin}}\, f(x_*) + g(x_*)^T s + \tfrac{1}{2} s^T B s \;\; \text{subject to} \;\; c(x_*) + J(x_*) s \geq 0 \end{aligned} \tag{2.14}
$$

for some positive definite matrix B, and let y_* denote the associated Lagrange multiplier vector. Then, it follows that (x_*, y_*) is a KKT point for problem (1.1) *as defined by* (1.2).

Proof. Since B is positive definite, $s = 0$ is the unique solution to the optimization problem in (2.14). It then follows from the first-order necessary optimality conditions at $s = 0$ that

$$g(x_*) = J(x_*)^T y_* \quad \text{and} \quad \min\big(c(x_*), y_*\big) = 0,$$

where y_* is the Lagrange multiplier for the constraint $c(x_*) + J(x_*)s \geq 0$. It now follows from Definition 1.2 that (x_*, y_*) is a KKT point for problem (1.1). ∎

2.3. The search direction s_k. The steering direction s_k^s provides a measure of local progress in infeasibility. Since we desire a search direction s_k that makes progress toward feasibility, we define

$$s_k := (1 - \tau_k)s_k^s + \tau_k s_k^p, \tag{2.15}$$

where τ_k is the largest number on $[0, 1]$ such that

$$\Delta\ell^v(s_k; x_k) \geq \eta_v \Delta\ell^v(s_k^s; x_k) \geq 0 \text{ for some } \eta_v \in (0, 1). \tag{2.16}$$

The next lemma shows that $\tau_k > 0$ when x_k is not an infeasible stationary point. This is important since the step s_k then has a significant contribution from s_k^p, which was computed from a subproblem that modeled *both* objective and constraint functions; this contrasts traditional filter methods when restoration is entered since the subproblem formulations then focus solely on the constraint violation.

LEMMA 2.3. *If x_k is not an infeasible stationary point as given by Definition* 2.1*, then $\tau_k > 0$.*

Proof. If $v(x_k) = 0$, then $\Delta\ell^v(s_k^s; x_k) = 0$. It then follows from (2.12a) that $c_k + J_k s_k^p \geq 0$, which in turn implies that $\Delta\ell^v(s_k^p; x_k) = 0$. Thus, the choice $\tau_k = 1$ satisfies (2.15) and (2.16).

Now suppose that $v(x_k) > 0$ and define

$$s(\tau) = (1 - \tau)s_k^s + \tau s_k^p$$

so that $\lim_{\tau\downarrow 0} s(\tau) = s_k^s$. It then follows from continuity of $\Delta\ell^v(\cdot\,; x_k)$ and the fact that $\Delta\ell^v(s_k^s; x_k) > 0$ since x_k is not an infeasible stationary point by assumption that

$$\lim_{\tau\downarrow 0} \Delta\ell^v\big(s(\tau); x_k\big) = \Delta\ell^v(s_k^s; x_k) > 0.$$

Therefore, there exists $\tau' > 0$ such that

$$\big|\Delta\ell^v\big(s(\tau); x_k\big) - \Delta\ell^v(s_k^s; x_k)\big| < (1 - \eta_v)\Delta\ell^v(s_k^s; x_k) \quad \text{for all } \tau \in [0, \tau']$$

since $\eta_v \in (0, 1)$ in (2.16) and $\Delta\ell^v(s_k^s; x_k) > 0$. However, this implies that

$$\Delta\ell^v\big(s(\tau); x_k\big) \geq \eta_v \Delta\ell^v(s_k^s; x_k) \quad \text{for all } \tau \in [0, \tau'],$$

which guarantees that $t_k \geq \tau' > 0$. ∎

We now proceed to show that if $\Delta\ell^v(s_k^s; x_k) > 0$, then s_k is a descent direction for $v(\cdot)$. We require the definition of the directional derivative of a function.

DEFINITION 2.4. *The directional derivative of a function $h(\cdot)$ in the direction d and at the point x is defined (when it exists) as*

$$[D_d h](x) := \lim_{t\downarrow 0} \frac{h(x + td) - h(x)}{t}.$$

We now show that the directional derivative is bounded by the negative of the change in its model.

LEMMA 2.5. *At any point x and for any direction d, it follows that*

$$[D_d v](x) \leq -\Delta\ell^v(d; x),$$

where the function $D_d v$ is the directional derivative of v in the direction d.

Proof. Since ℓ^v is a convex function and $\ell^v(0;x)$ is finite, it follows from [42, Theorem 23.1] that

$$\frac{\ell^v(td; x) - \ell^v(0; x)}{t}$$

is monotonically nondecreasing with t, $[D_d\ell^v](0;x)$ exists, and

$$[D_d\ell^v](0; x) = \inf_{t>0} \frac{\ell^v(td; x) - \ell^v(0; x)}{t}. \tag{2.17}$$

It then follows from [9, Lemma 3.1], (2.17), and the definition of $\Delta\ell^v$ that

$$[D_d v](x) = [D_d\ell^v](0; x) \leq \ell^v(d; x) - \ell^v(0; x) = -\Delta\ell^v(d; x),$$

which is the desired result. $\square$

Thus the search direction s_k is a descent direction for v when our infeasibility measure is positive.

LEMMA 2.6. *If $\Delta\ell^v(s_k^s; x_k) > 0$, the direction s_k is a descent direction for v at the point x_k, i.e.,*

$$[D_{s_k} v](x_k) \leq -\Delta\ell^v(s_k; x_k) \leq -\eta_v\Delta\ell^v(s_k^s; x_k) < 0, \quad \textit{where } \eta_v \textit{ is defined in } (2.16).$$

Proof. It follows directly from Lemma 2.5, (2.16), and $\Delta\ell^v(s_k^s; x_k) > 0$ that

$$[D_{s_k} v](x_k) \leq -\Delta\ell^v(s_k; x_k) \leq -\eta_v\Delta\ell^v(s_k^s; x_k) < 0,$$

which implies that s_k is a descent direction for v at the point x_k. $\square$

We now consider the case when our infeasibility measure is zero.

LEMMA 2.7. *If $\Delta\ell^v(s_k^s; x_k) = 0$, then one of the following must occur:*

(i) *$v(x_k) > 0$ and x_k is an infeasible stationary point or*

(ii) *$v(x_k) = 0$ and $\Delta\ell^\phi(s_k; x_k, \sigma) \geq \frac{1}{2}{s_k^p}^T B_k s_k^p$ for all $0 < \sigma < \infty$.*

Proof. If $v(x_k) > 0$, then by Definition 2.1, x_k is an infeasible stationary point which is part (i). Now, suppose that $v(x_k) = 0$. As in the proof of Lemma 2.3, it follows that

$$\Delta\ell^v(s_k^p; x_k) = 0, \quad \tau_k = 1, \quad \text{and} \quad s_k = s_k^p. \tag{2.18}$$

We may then use the definition of s_k^p in (2.12a), (2.18), and (2.7b) to conclude that

$$0 \leq \Delta q^\phi(s_k^p; x_k, B_k, \sigma_k) = \Delta q^f(s_k^p; x_k, B_k) = \Delta\ell^f(s_k^p; x_k) - \tfrac{1}{2}{s_k^p}^T B_k s_k^p,$$

which yields $\Delta\ell^f(s_k^p; x_k) \geq \frac{1}{2}{s_k^p}^T B_k s_k^p$. Combining this with (2.8) and (2.18), we have that

$$\begin{aligned}\Delta\ell^\phi(s_k; x_k, \sigma) &= \Delta\ell^f(s_k; x_k) + \sigma\Delta\ell^v(s_k; x_k) \\ &= \Delta\ell^f(s_k; x_k) = \Delta\ell^f(s_k^p; x_k) \geq \tfrac{1}{2}{s_k^p}^T B_k s_k^p \text{ for all finite } \sigma,\end{aligned}$$

which completes the proof. $\square$

2.4. Updating the weighting parameter. By design, the trial step s_k is a descent direction for v when local improvement in feasibility is possible. Since the weighting parameter provides a balance between reducing the objective function and the constraint violation, it makes sense to adjust the weighting parameter so that s_k is also a descent direction for ϕ. This is accomplished by defining

$$\sigma_{k+1}=\begin{cases}\sigma_k & \text{if } \Delta\ell^\phi(s_k;x_k,\sigma_k)\ge\sigma_k\eta_\sigma\Delta\ell^v(s_k^s;x_k),\\ \max\left\{\sigma_k+\sigma_{\text{inc}},\ \frac{-\Delta\ell^f(s_k;x_k)}{\Delta\ell^v(s_k;x_k)-\eta_\sigma\Delta\ell^v(s_k^s;x_k)}\right\} & \text{otherwise}\end{cases} \tag{2.19}$$

for some $\sigma_{\text{inc}}>0$ and η_σ satisfying $0<\eta_\sigma<\eta_v<1$, where η_v is defined in (2.16).

LEMMA 2.8. *If x_k is not an infeasible stationary point, then the parameter update* (2.19) *is well defined and ensures that*

$$\Delta\ell^\phi(s_k;x_k,\sigma_{k+1})\ge\sigma_{k+1}\eta_\sigma\Delta\ell^v(s_k^s;x_k)\ge 0 \quad \textit{for all } k\ge 0. \tag{2.20}$$

Proof. If $\Delta\ell^\phi(s_k;x_k,\sigma_k)\ge\sigma_k\eta_\sigma\Delta\ell^v(s_k^s;x_k)$, then the desired result immediately follows from the update $\sigma_{k+1}=\sigma_k$. Thus, for the remainder of the proof we assume that

$$\Delta\ell^\phi(s_k;x_k,\sigma_k)<\sigma_k\eta_\sigma\Delta\ell^v(s_k^s;x_k). \tag{2.21}$$

Suppose, for a contradiction, that $\Delta\ell^v(s_k^s;x_k)=0$. Since x_k is not an infeasible stationary point by assumption, it follows that $v(x_k)=0$. Then, it follows from Lemma 2.7 and the fact that B_k is positive definite by assumption that $\Delta\ell^\phi(s_k;x_k,\sigma_k)\ge\frac{1}{2}s_k^{pT}B_ks_k^p\ge 0$, which contradicts (2.21) since $\Delta\ell^v(s_k^s;x_k)=0$. Thus, we conclude that $\Delta\ell^v(s_k^s;x_k)>0$. Combining this with the choice $0<\eta_\sigma<\eta_v<1$ in (2.19) and (2.16) we conclude that $\Delta\ell^v(s_k;x_k)\ge\eta_v\Delta\ell^v(s_k^s;x_k)>\eta_\sigma\Delta\ell^v(s_k^s;x_k)>0$, and thus

$$\eta_\sigma\Delta\ell^v(s_k^s;x_k)-\Delta\ell^v(s_k;x_k)<0. \tag{2.22}$$

It then follows from (2.8), (2.21), (2.22), and the fact that $\sigma_k>0$ that

$$\Delta\ell^f(s_k;x_k)=\Delta\ell^\phi(s_k;x_k,\sigma_k)-\sigma_k\Delta\ell^v(s_k;x_k)<\sigma_k\big[\eta_\sigma\Delta\ell^v(s_k^s;x_k)-\Delta\ell^v(s_k;x_k)\big]<0. \tag{2.23}$$

Inequalities (2.22) and (2.23) imply that the penalty parameter update (2.19) is well-defined and positive.

It now follows from (2.19) that

$$\sigma_{k+1}\ge\frac{-\Delta\ell^f(s_k;x_k)}{\Delta\ell^v(s_k;x_k)-\eta_\sigma\Delta\ell^v(s_k^s;x_k)},$$

which may then be combined with (2.22) to yield

$$\sigma_{k+1}\eta_\sigma\Delta\ell^v(s_k^s;x_k)\le\Delta\ell^f(s_k;x_k)+\sigma_{k+1}\Delta\ell^v(s_k;x_k)=\Delta\ell^\phi(s_k;x_k,\sigma_{k+1}),$$

which is the desired result (2.20). $\square$

The next result will allow us to show that s_k is a descent direction for ϕ under certain assumptions.

LEMMA 2.9. *For any given value of the penalty parameter σ, point x, direction d, and positive-definite matrix B, it follows that*

$$[D_d\phi](x;\sigma)\le-\Delta\ell^\phi(d;x,\sigma)\le-\Delta q^\phi(d;x,B,\sigma).$$

Proof. Linearity of the directional derivative, (2.7a), Lemma 2.5, (2.8), (2.9), and the fact that B_k is positive definite by choice imply that

$$\begin{aligned}
&[D_d\phi](x;\sigma)\\
&\quad= [D_d f](x) + \sigma[D_d v](x) = -g(x)^T d + \sigma[D_d v](x) \le -\Delta\ell^f(d;x) - \sigma\Delta\ell^v(d;x)\\
&\quad= -\Delta\ell^\phi(d;x,\sigma) = -\Delta q^\phi(d;x,B,\sigma) - \tfrac{1}{2}d^T Bd \le -\Delta q^\phi(d;x,B,\sigma),
\end{aligned}$$

which is the desired result. ∎

In most situations, we may now show that s_k is a descent direction for the penalty function.

LEMMA 2.10. *If x_k is neither an infeasible stationary point nor a KKT point for problem* (1.1)*, then the direction s_k is a descent direction for $\phi(x;\sigma_{k+1})$ at the point x_k, i.e.,*

$$[D_{s_k}\phi](x_k;\sigma_{k+1}) \le -\Delta\ell^\phi(s_k;x_k,\sigma_{k+1}) < 0.$$

Proof. If $\Delta\ell^v(s_k^s;x_k) > 0$, then x_k cannot be an infeasible stationary point, and it follows from Lemma 2.9, Lemma 2.8, and (2.20) that $[D_{s_k}\phi](x_k;\sigma_{k+1}) \le -\Delta\ell^\phi(s_k;x_k,\sigma_{k+1}) < 0$, which is the desired result. Conversely, if $\Delta\ell^v(s_k^s;x_k) = 0$, then $v(x_k) = 0$ since x_k is not an infeasible stationary point by assumption. It now follows from Lemma 2.9, $v(x_k) = 0$, Lemma 2.7, the fact that B_k is positive definite, and $s_k^p \ne 0$ since x_k is not a KKT point for problem (1.1) by assumption (see Lemma 2.2), that $[D_{s_k}\phi](x_k;\sigma_{k+1}) \le -\Delta\ell^\phi(s_k;x_k,\sigma_{k+1}) \le -\frac{1}{2}{s_k^p}^T B_k s_k^p < 0$, which completes the proof. ∎

2.5. The accelerator step s_k^a. To improve performance, we compute an additional "acceleration" step; here we consider a single (simple) possibility, but other variants may be used [28].

Under common assumptions, the predictor step s_k^p will ultimately correctly identify those constraints that are active at a local solution of (1.1) [41]. A prediction based on s_k^p is formulated by

$$\mathcal{A}_k := \{i : [c_k + J_k s_k^p]_i = 0\}. \tag{2.24}$$

It is then natural to compute an *accelerator* step s_k^a as the solution to

$$\underset{s\in\mathbb{R}^n}{\text{minimize}}\ q^f(s_k^p + s;x_k,H_k)\ \text{ subject to }\ [J_k s]_{\mathcal{A}_k} = 0,\ \ \|s\|_2 \le \delta_k^a, \tag{2.25}$$

where $\delta_k^a > 0$ is the trust-region radius, H_k is the exact second derivative of the Lagrangian $\nabla_{xx}^2 L(x_k,y_k)$, and y_k is a suitable Lagrange multiplier vector such as those from the predictor subproblem. (In fact, our global convergence analysis allows for any symmetric bounded sequence $\{H_k\}$, but here for concreteness we simply use $H_k = \nabla_{xx}^2 L(x_k,y_k)$.) We note that subproblem (2.25) may be solved, for example, with the projected GLTR algorithm. (See [17, section 7.5.4] and the notes at the end that describe how to cope with the affine constraints $[J_k s]_{\mathcal{A}_k} = 0$.) It can be shown that if $c_k + J_k s \ge 0$ is feasible, σ_k is sufficiently large, and x_k is "close enough" to a solution of (1.1) that satisfies certain second-order sufficient optimality conditions, then $s_k^p + s_k^a$ is the solution to

$$\underset{s\in\mathbb{R}^n}{\text{minimize}}\ q^f(s;x_k,H_k)\ \text{ subject to }\ c_k + J_k s \ge 0, \tag{2.26}$$

which is the traditional SQP subproblem. However, our method of step computation is robust, whereas the generally nonconvex subproblem (2.26) introduces many points of contention such as multiple solutions, unboundedness, and inconsistent constraints.

2.6. The Cauchy steps s_k^{cf} and $s_k^{c\phi}$. Since the matrix B_k is positive definite by construction and the exact second derivative matrix H_k is generally an indefinite matrix, they may differ dramatically. To account for this when assessing overall step acceptance, we define and use a *Cauchy-f* step s_k^{cf} and a *Cauchy-ϕ* step $s_k^{c\phi}$ as follows.

Given the search direction s_k, we define the Cauchy-f step as

$$s_k^{cf} := \alpha_k^f s_k, \quad \text{where} \quad \alpha_k^f := \operatorname*{argmin}_{0\le\alpha\le 1} \; q^f(\alpha s_k; x_k, H_k). \tag{2.27}$$

Similarly, we define the Cauchy-ϕ step as

$$s_k^{c\phi} := \alpha_k^\phi s_k, \quad \text{where} \quad \alpha_k^\phi := \operatorname*{argmin}_{0\le\alpha\le 1} \; q^\phi(\alpha s_k; x_k, H_k, \sigma_{k+1}). \tag{2.28}$$

The step size α_k^ϕ may be found efficiently by examining the piecewise quadratic function $q^\phi(\alpha s_k; x_k, H_k, \sigma_{k+1})$ segment by segment between each derivative discontinuity.

2.7. The filter. The global convergence proof for our method is driven by maintaining/updating a filter $\mathcal{F}_k$ during each iteration. A filter is defined as follows, where $\mathbb{R}^+$ denotes the positive real numbers.

DEFINITION 2.11 (filter). *A* filter *is any* finite *set of points in* $\mathbb{R}^+ \times \mathbb{R}$.

The initial filter is defined to be $\mathcal{F}_0 = \emptyset$ and then sequentially updated in a manner that guarantees that $\mathcal{F}_k \subseteq \{(v_j, f_j) : 0 \le j < k\}$. The decision to add certain ordered pairs to the filter depends on the concept of trial points being acceptable to the filter, which we now define.

DEFINITION 2.12 (acceptable to $\mathcal{F}_k$). *We say that the point x is acceptable to $\mathcal{F}_k$ if its associated ordered pair $\big(v(x), f(x)\big)$ satisfies*

$$\begin{aligned} v(x) &\le \max\left\{v_i - \alpha_i\eta_v\Delta\ell^v(s_i^s; x_i),\, \beta v_i\right\} \quad \text{or} \\ f(x) &\le f_i - \gamma\min\left\{v_i - \alpha_i\eta_v\Delta\ell^v(s_i^s; x_i),\, \beta v_i\right\} \end{aligned} \tag{2.29}$$

for all $0 \le i < k$ such that $(v_i, f_i) \in \mathcal{F}_k$ and some constants $\{\eta_v, \beta, \gamma\} \subset (0,1)$.

The first inequality in (2.29) ensures that the constraint violation has been sufficiently reduced. We note that previous filter methods have not used the first quantity in the max on the right-hand side. Our improved condition takes advantage of the information supplied by the steering steps s_k^s. Previous filter methods may easily have requested a decrease in the constraint violation that was unreasonable. In these circumstances, the trust-region radius would be decreased until the subproblem became infeasible and then a feasibility restoration phase would be entered. Our modified definition provides a practical target constraint violation based on local information derived from the steering step s_k^s. The second inequality in (2.29) guarantees that the objective function is sufficiently smaller at the point x than at points x_i whose ordered pair is in the current filter $\mathcal{F}_k$. These two conditions provide a so-called *margin* around the elements of the filter.

Note that Definition 2.12 does not require and does not imply that the current vector x_k is in $\mathcal{F}_k$ when determining acceptability. During our search for an improved estimate of a solution to (1.1), it often does not make sense to accept a new point unless it is acceptable to the current filter *and* better than the current point x_k. This leads to the following definition.

DEFINITION 2.13 (acceptable to $\mathcal{F}_k$ augmented by x_k). *We say that x is acceptable to $\mathcal{F}_k$ augmented by x_k if x is acceptable to $\mathcal{F}_k$ as given by Definition* 2.12 *and* (2.29) *holds with $i = k$.*

In the next section we present our main filter SQP method. Each iteration requires the search for a new point that must satisfy a subset of specified conditions. We stress that the updated point x_{k+1} is not necessarily acceptable to $\mathcal{F}_k$. Moreover, the vector x_{k+1} being acceptable to $\mathcal{F}_k$ (possibly augmented by x_k) is a necessary, but not sufficient, condition for adding the ordered pair (v_{k+1}, f_{k+1}) to the filter $\mathcal{F}_k$. Details of how we update $\mathcal{F}_k$ are described in the next section.

2.8. The algorithm. Our method is formally stated as Algorithm 1. Every iteration begins by computing the set of trial steps $\{s_k^a, s_k\}$ as described in sections 2.5 and 2.3. Once these trial steps are computed, we seek a step length α_k such that for some $\hat{s}_k \in \{s_k^a, s_k\}$ the step $x_k + \alpha_k \hat{s}_k$ satisfies one of four possible sets of conditions. Which sets of conditions we seek to satisfy depends on whether the algorithm is in filter mode (roughly a traditional filter strategy) or penalty mode (our alternative to a traditional restoration phase). We now discuss these two modes in detail.

In filter mode, we perform a backtracking line search until we find a pair $(\alpha_k, \hat{s}_k)$ with $\hat{s}_k \in \{s_k, s_k^a\}$ that forms a v-pair or an o-pair, or a pair (α_k, s_k) that forms a b-pair. We discuss these in turn.

A v-pair is defined as follows.

DEFINITION 2.14 (v-pair). *The pair (α, s) constitutes a v-pair if $x_k + \alpha s$ is acceptable to $\mathcal{F}_k$ augmented by x_k and*

$$\Delta\ell^f(s_k; x_k) < \gamma_v \Delta\ell^v(s_k; x_k) \ \textit{ for some } \gamma_v \in (0,1). \tag{2.30}$$

A v-pair $(\alpha_k, \hat{s}_k)$ earns its name since the step $x_k + \alpha_k \hat{s}_k$ is acceptable to the current filter augmented by x_k, but the step s_k did not predict sufficient decrease in f as measured by (2.30); we say that k is a v-iterate since the focus of the iteration is on reducing the constraint violation v. In this case, we choose to add the pair (v_k, f_k) to the filter $\mathcal{F}_k$.

An o-pair is characterized as follows.

DEFINITION 2.15 (o-pair). *The pair (α, s) constitutes an o-pair if $x_k + \alpha s$ is acceptable to $\mathcal{F}_k$,*

$$\Delta\ell^f(s_k; x_k) \geq \gamma_v \Delta\ell^v(s_k; x_k) \ \textit{ and} \tag{2.31a}$$

$$f(x_k + \alpha s) \leq f(x_k) - \gamma_f \alpha \rho_k^f, \tag{2.31b}$$

where $\gamma_v \in (0,1)$ is the same constant used to define a v-pair, $\gamma_f \in (0,1)$, and

$$\rho_k^f := \min\left[\Delta\ell^f(s_k; x_k),\ \Delta q^f(s_k^{cf}; x_k, H_k)\right]. \tag{2.32}$$

An o-pair $(\alpha_k, \hat{s}_k)$ is so designated since $x_k + \alpha_k \hat{s}_k$ is acceptable to the filter, s_k predicts decrease in the objective function as measured by (2.31a), and a sufficient decrease in the objective is realized as given by (2.31b); we say that k is an o-iterate since progress has been made on decreasing the objective function. In this case we do not add any new entries to the filter.

The definitions of v- and o-pairs are natural in light of the mechanism of the filter and are similar in spirit to conditions used by previous methods [13, 15, 21, 22, 24, 46]. As for these methods, these two sets of conditions are not sufficient for ensuring convergence since previously added filter entries may prevent (block) additional progress. In this situation, other filter algorithms typically decrease the trust-region radius or perform backtracking until a restoration phase is triggered. To prevent this undesirable situation we introduce the following definition of a b-pair.

ALGORITHM 1. Filter sequential quadratic programming algorithm.

1: Input an initial primal-dual pair (x_0, y_0).
2: Choose $\{\eta_v, \eta_\sigma, \eta_\phi, \sigma_{\text{inc}}, \beta, \gamma, \gamma_v, \gamma_f, \gamma_\phi, \xi\} \subset (0,1)$ and $0 < \delta_{\min} \le \delta_{\max} < \infty$.
3: Set $k \leftarrow 0$, $\mathcal{F}_0 \leftarrow \emptyset$, $\mathcal{P}$-*mode* $\leftarrow$ **false**, and choose $\sigma_0 > 0$ and $\delta_0 \in [\delta_{\min}, \delta_{\max}]$.
4: **loop**
5: Compute s_k^s as a solution of (2.10), and then calculate $\Delta\ell^v(s_k^s; x_k)$ from (2.6).
6: **if** $\Delta\ell^v(s_k^s; x_k) = 0$ and $v(x_k) > 0$, **then**
7: **return** with the infeasible stationary point x_k for problem (1.1).
8: Choose $B_k \succ 0$ and compute s_k^p as the solution of (2.12) with multiplier y_k^p.
9: **if** $\Delta q^\phi(s_k^p; x_k, \sigma_k) = v(x_k) = 0$, **then**
10: **return** with the KKT point (x_k, y_k^p) for problem (1.1).
11: Compute $s_k = (1-\tau_k)s_k^s + \tau_k s_k^p$ from (2.15) such that (2.16) is satisfied.
12: Compute the new weight σ_{k+1} from (2.19).
13: Choose $\delta_k^a > 0$ and then compute s_k^a as an (approximate) solution of (2.25).
14: Compute $s_k^{c\phi}$ from (2.28) and then calculate $\Delta q^\phi(s_k^{c\phi}; x_k, H_k, \sigma_{k+1})$ from (2.9).
15: **if** $\mathcal{P}$-*mode* **then**
16: **for** $j = 0, 1, 2, \ldots$ **do**
17: Set $\alpha_k \leftarrow \xi^j$.
18: **for** $\hat{s}_k \in \{s_k^a, s_k\}$ **do**
19: **if** $(\alpha_k, \hat{s}_k)$ is a p-pair **then**
20: Set $\mathcal{F}_{k+1} \leftarrow \mathcal{F}_k$ and go to line 21. $\triangleright$ p-iterate
21: **if** $x_k + \alpha_k \hat{s}_k$ is acceptable to $\mathcal{F}_k$ **then**
22: Set $\mathcal{P}$-*mode* $\leftarrow$ **false**.
23: **else**
24: Compute s_k^{cf} from (2.27) and then calculate $\Delta q^f(s_k^{cf}; x_k, H_k)$ from (2.7b).
25: **for** $j = 0, 1, 2, \ldots$ **do**
26: Set $\alpha_k \leftarrow \xi^j$.
27: **for** $\hat{s}_k \in \{s_k^a, s_k\}$ **do**
28: **if** $(\alpha_k, \hat{s}_k)$ is a v-pair **then**
29: Set $\mathcal{F}_{k+1} \leftarrow \mathcal{F}_k \cup \{(v_k, f_k)\}$ and go to line 34. $\triangleright$ v-iterate
30: **if** $(\alpha_k, \hat{s}_k)$ is an o-pair **then**
31: Set $\mathcal{F}_{k+1} \leftarrow \mathcal{F}_k$ and go to line 34. $\triangleright$ o-iterate
32: **if** (α_k, s_k) is a b-pair **then**
33: Set $\mathcal{F}_{k+1} \leftarrow \mathcal{F}_k \cup \{(v_k, f_k)\}$, $\mathcal{P}$-*mode* $\leftarrow$ **true**, and go to line 34. $\triangleright$ b-iterate
34: **if** (2.36) is satisfied **then**
35: Set $\sigma_{k+1} \leftarrow \sigma_{k+1} + \sigma_{\text{inc}}$.
36: Set $x_{k+1} \leftarrow x_k + \alpha_k \hat{s}_k$, $y_{k+1} \leftarrow y_k^{\text{p}}$, $\delta_{k+1} \in [\delta_{\min}, \delta_{\max}]$, and $k \leftarrow k+1$.

DEFINITION 2.16 (b-pair). *The pair (α, s) constitutes a b-pair if*

$$v(x_k + \alpha s) < v(x_k) \tag{2.33}$$

and

$$\phi(x_k + \alpha s; \sigma_{k+1}) \le \phi(x_k; \sigma_{k+1}) - \gamma_\phi \alpha \rho_k^\phi \quad \textit{for some } \gamma_\phi \in (0,1), \tag{2.34}$$

where

$$\rho_k^\phi := \min\Big[\, \Delta\ell^\phi(s_k; x_k, \sigma_{k+1}),\ \Delta q^\phi(s_k^{c\phi}; x_k, H_k, \sigma_{k+1}) \Big]. \tag{2.35}$$

An iterate $x_k+\alpha_k s_k$ associated with a b-pair (α_k, s_k) decreases both the constraint violation and the penalty function; we say that k is a b-iterate since the conditions that define a b-pair suggest that one or more filter entries are blocking a productive step. In this case, we choose to accept the trial point, add (v_k, f_k) to the filter, and enter what we will refer to as a penalty mode. We view penalty mode as an alternative to a traditional restoration phase. Moreover, since steps are always tested for acceptability based on the filter criteria, i.e., o- and v-pairs, before checking for decrease in the constraint violation and penalty function as stipulated by b-pairs, we give clear preference to staying in filter mode.

In penalty mode, we calculate a new iterate by perform a backtracking line search until we find a pair $(\alpha_k, \hat{s}_k)$ for some $\hat{s}_k \in \{s_k^a, s_k\}$ that satisfies the following conditions that define a p-pair.

Definition 2.17 (p-pair). *The pair* (α, s) *constitutes a p-pair if* (2.34) *is satisfied.*

If $(\alpha_k, \hat{s}_k)$ is a p-pair, then $\phi(x_k+\alpha_k\hat{s}_k; \sigma_{k+1})$ is sufficiently smaller than $\phi(x_k; \sigma_{k+1})$; we say that k is a p-iterate since the penalty function has been decreased. In addition, if $x_k + \alpha_k\hat{s}_k$ is acceptable to the current filter, we return to filter mode; otherwise, we remain in penalty mode.

Finally, after a new trial step is computed, we choose to increase the penalty parameter if

$$\Delta q^{\phi}(s_k; x_k, B_k, \sigma_{k+1}) < \eta_\phi \Delta q^{\phi}(s_k^p; x_k, B_k, \sigma_{k+1}) \quad \text{for some } \eta_\phi \in (0,1), \tag{2.36}$$

since this indicates that τ_k is very small and the search direction s_k does not adequately reflect the decrease predicted by s_k^p in the penalty function.

For future reference we define the following index sets based on the different types of pairs:

$$\begin{aligned} \mathcal{S}_v &= \{k :\ k \text{ is a v-iterate}\}, & \mathcal{S}_o &= \{k :\ k \text{ is an o-iterate}\}, \\ \mathcal{S}_p &= \{k :\ k \text{ is a p-iterate}\}, & \mathcal{S}_b &= \{k :\ k \text{ is an b-iterate}\}. \end{aligned}$$

We complete this section by summarizing the computational complexity of each iteration of Algorithm 1, which requires the calculation of multiple directions. Specifically, each iteration requires the solution of a linear program to obtain the steering step (see (2.10)), a strictly convex quadratic program to get the predictor step (see (2.12)), a single matrix-vector multiplication to solve the one-dimensional optimization problem for the Cauchy-f step (see (2.27)), a one-dimensional search along a piecewise linear path to obtain the Cauchy-ϕ step (see (2.28)), and an approximate solution to an equality-constrained quadratic problem for the an accelerator step (see (2.25)). Therefore, the predominant computational cost for each iteration is the calculation of the steering and predictor steps.

3. Well-posedness. In this section we verify that every step of the method is well-posed under the following assumption, which we do not explicitly state for each result.

Assumption 3.1. The functions f and c are both differentiable with Lipschitz continuous derivatives in the neighborhood of the point x_k.

We begin by observing that the steering problem (2.11) is convex and always feasible, and the objective function is bounded below by zero, i.e., it is well-defined. If $v(x_k) > 0$ and $\Delta\ell^v(s_k^s; x_k) = 0$, then x_k is an infeasible stationary point and we exit in line 7 of Algorithm 1. Otherwise, x_k is not an infeasible stationary point and we proceed to compute a predictor step from problem (2.12), which we now argue is well-defined. This is obvious when $\Delta\ell^v(s_k^s;x_k) \neq v(x_k)$ since then the strictly convex problem (2.12b) is always feasible. On the other hand, if $\Delta\ell^v(s_k^s;x_k) = v(x_k)$, then it follows that $||[c(x_k) + J(x_k)s_k^s]^-||_1 = 0$, which implies that $c_k + J_k s_k^s \geq 0$. Thus, $s = s_k^s$ is feasible for (2.12a), and the predictor problem is well-defined. Lemma 2.3 shows that $\tau_k > 0$ and Lemma 2.8 shows that the update to the weighting parameter is well-defined. The accelerator problem (2.25) does not cause difficulties since by construction it is feasible, has bounded solutions, and may be solved (approximately) as noted in section 2.5. It is also easy to see that both Cauchy step problems (2.27) and (2.28) are well-defined.

We now proceed to show that the line search terminates finitely. To this end, we first show that feasible iterates are never added to the filter.

LEMMA 3.1. *Algorithm* 1 *ensures that if* (v_k, f_k) *is added to the filter, then* $v_k > 0$.

Proof. For a proof by contradiction, suppose that $v(x_k) = 0$. It follows from $v(x_k) = 0$ and the fact that ℓ^v is a convex function that $\Delta\ell^v(s_k^s, x_k) = 0$, and we may then use (2.12a), (2.15), (2.16), and the fact that x_k is not a KKT point for (1.1) (otherwise we would already have exited on line 10 of Algorithm 1) to show that

$$\tau_k = 1, \quad s_k = s_k^p \neq 0, \quad \text{and} \quad \Delta\ell^v(s_k;x_k) = \Delta\ell^v(s_k^p;x_k) = 0. \tag{3.1}$$

It then follows from (3.1), (2.8), Lemma 2.7, $v(x_k) = 0$, and the fact that $B_k \succ 0$ that

$$\Delta\ell^f(s_k;x_k) = \Delta\ell^f(s_k^p;x_k) = \Delta\ell^\phi(s_k^p;x_k,\sigma_{k+1}) = \Delta\ell^\phi(s_k;x_k,\sigma_{k+1}) \geq \tfrac{1}{2}s_k^{pT} B_k s_k^p > 0. \tag{3.2}$$

Since (v_k, f_k) was added to the filter, it follows from the construction of Algorithm 1 that either $(\alpha_k, \hat{s}_k)$ is a v-iterate or (α_k, s_k) is a b-pair, which implies that at least one of $v(x_k + \alpha_k s_k) < v(x_k)$ or $\Delta\ell^f(s_k;x_k) < \gamma_v \Delta\ell^v(s_k;x_k)$ holds, among other requirements. However, since $v(x_k + \alpha_k s_k) < v(x_k) = 0$ is not possible, we conclude that $\Delta\ell^f(s_k;x_k) < \gamma_v\Delta\ell^v(s_k;x_k) = 0$, where we have also used (3.1); this contradicts (3.2) and proves the result. ∎

The next two results show that our line search procedure terminates any time $\mathcal{P}$-*mode* has the value false at the beginning of the kth iteration. We first consider the case when x_k is feasible.

LEMMA 3.2. *If* $\mathcal{P}$-*mode* = **false** *at the beginning of the* k*th iteration,* $v(x_k) = 0$, *and* x_k *is not a first-order solution to problem* (1.1)*, then the pair* (α, s_k) *is an o-pair for all* $\alpha > 0$ *sufficiently small. Moreover,* $k \in \mathcal{S}_o$.

Proof. As in the proof of Lemma 3.1, it follows that $v(x_k) = \Delta\ell^v(s_k^s;x_k) = 0$. This may be combined with the fact that x_k is assumed to not be a first-order solution to (1.1), (2.12a), (2.15), (2.16), Lemma 2.7, the fact that B_k is positive definite, and the definition of $\Delta\ell^\phi$ to conclude that

$$\begin{aligned} &s_k = s_k^p \neq 0, \quad c_k + J_k s_k \geq 0, \quad \text{and} \\ &\Delta\ell^f(s_k;x_k) = \Delta\ell^\phi(s_k;x_k,\sigma_{k+1}) > 0 = \gamma_v\Delta\ell^v(s_k;x_k). \end{aligned} \tag{3.3}$$

Next, $v(x_k) = 0$ and (3.3) imply that $c_k + \alpha J_k s_k \geq 0$ for all $\alpha \in [0,1]$. Combining this fact with Taylor's theorem, Assumption 3.1, and (3.3) yields

$$v(x_k + \alpha s_k) = \left\|[c(x_k + \alpha s_k)]^-\right\|_1 = \left\|[c_k + \alpha J_k s_k + O(\alpha^2)]^-\right\|_1 \leq O(\alpha^2) \ \text{ for } \alpha \in [0,1]. \tag{3.4}$$

Since Lemma 3.1 implies that $v_i > 0$ for all $(v_i, f_i) \in \mathcal{F}_k$, we may conclude from (3.4) that

$$v(x_k + \alpha s_k) \le \min_{(v_i,f_i)\in\mathcal{F}_k} \beta v_i \quad \text{for all } \alpha > 0 \text{ sufficiently small,}$$

where $\beta \in (0,1)$ is defined in (2.29), so that

$$x_k + \alpha s_k \ \text{ is acceptable to the filter for all } \alpha > 0 \text{ sufficiently small.} \tag{3.5}$$

Next, Taylor's theorem, Assumption 3.1, the definition of $\Delta\ell^f$, and (3.3) imply that

$$\begin{aligned} f(x_k + \alpha s_k) &= f_k + \alpha g_k^T s_k + O(\alpha^2) = f_k - \alpha\Delta\ell^f(s_k; x_k) + O(\alpha^2) \\ &\le f_k - \gamma_f \alpha \Delta\ell^f(s_k; x_k) \ \text{ for all } \alpha > 0 \text{ sufficiently small,} \end{aligned} \tag{3.6}$$

where $\gamma_f \in (0,1)$ is defined in (2.31b). It follows from (3.3), (3.5), and (3.6) that (α, s_k) is an o-pair for all $\alpha > 0$ sufficiently small, which proves the first result of this lemma.

We just proved that the **for** loop on line 25 in Algorithm 1 always terminates. Moreover, it cannot terminate on line 28 since (3.3) holds. Also, it can never terminate as a result of the **if** on line 32 since $v(x_k + \alpha s_k) < v(x_k) = 0$ is impossible for all α. Therefore, the line search must terminate with an o-pair $(\alpha_k, \hat{s}_k)$, which implies that $k \in \mathcal{S}_o$. ∎

We now consider the case when x_k is infeasible.

LEMMA 3.3. *If $\mathcal{P}$-mode* = **false** *at the beginning of iteration k, $v(x_k) > 0$, and x_k is not an infeasible stationary point, then (α, s_k) is a b-pair for all $\alpha > 0$ sufficiently small.*

Proof. It follows from the assumptions of this lemma and Lemma 2.10 that

$$[D_{s_k}\phi](x_k; \sigma_{k+1}) \le -\Delta\ell^\phi(s_k; x_k, \sigma_{k+1}) < 0 \tag{3.7}$$

so that the direction s_k is a strict descent direction for ϕ at x_k with penalty parameter σ_{k+1}. Using the definition of the directional derivative, (3.7), $\gamma_\phi \in (0,1)$ defined in (2.34), and (2.35) we conclude that

$$\begin{aligned} \phi(x_k + \alpha s_k; \sigma_{k+1}) &\le \phi(x_k; \sigma_{k+1}) + \alpha\gamma_\phi [D_{s_k}\phi](x_k; \sigma_{k+1}) \\ &\le \phi(x_k; \sigma_{k+1}) - \alpha\gamma_\phi \Delta\ell^\phi(s_k; x_k, \sigma_{k+1}) \\ &\le \phi(x_k; \sigma_{k+1}) - \alpha\gamma_\phi \rho_k^\phi \quad \text{for all } \alpha > 0 \text{ sufficiently small.} \end{aligned} \tag{3.8}$$

Since $v(x_k) \neq 0$ and x_k is not an infeasible stationary point, we know that $\Delta\ell^v(s_k^s; x_k) > 0$. Lemma 2.6 then implies that

$$[D_{s_k}v](x_k) \le -\Delta\ell^v(s_k; x_k) \le -\eta_v \Delta\ell^v(s_k^s; x_k) < 0$$

so that s_k is a descent direction for v at x_k. An argument similar to the one that lead to (3.8) yields

$$v(x_k + \alpha s_k) < v(x_k) \quad \text{for all } \alpha > 0 \text{ sufficiently small.} \tag{3.9}$$

It follows from (3.8) and (3.9) that (α, s_k) is a b-pair for all $\alpha > 0$ sufficiently small. ∎

The next lemma considers the case when $\mathcal{P}$-*mode* is true at the beginning of the kth iteration and shows that successful trial iterates may be obtained through backtracking as performed in Algorithm 1.

LEMMA 3.4. *If $\mathcal{P}$-mode* = **true** *at the beginning of the kth iteration and x_k is neither an infeasible stationary point nor a first-order solution to problem* (1.1)*, then* (α, s_k) *is a p-pair for all $\alpha > 0$ sufficiently small.*

Proof. The proof follows exactly as in the first part of Lemma 3.3. ∎

We now combine these results to prove that Algorithm 1 is well-posed.

THEOREM 3.5. *Algorithm* 1 *is well-posed.*

Proof. As described in the first paragraph of section 3, every subproblem and step computation is well defined, and Lemma 2.8 ensures that the update to the weighting parameter is well defined.

All that remains is to prove that the line search terminates. First, if $\mathcal{P}$-*mode* has the value false at the beginning of iteration x_k and $v(x_k) = 0$, then Lemma 3.2 guarantees finite termination and that $k \in \mathcal{S}_o$. Second, if $\mathcal{P}$-*mode* has the value false and $v(x_k) > 0$, then Lemma 3.3 ensures that the backtracking line search will terminate finitely. Finally, suppose that $\mathcal{P}$-*mode* has the value true at the beginning of iteration k. It then follows from Lemma 3.4 that the backtracking terminates finitely. ∎

4. Global convergence.

In this section we prove that limit points of the iterates generated by Algorithm 1 have desirable properties. To this end, we use the following common assumptions.

Assumption 4.1. The iterates $\{x_k\}$ lie in an open, bounded, convex set $\mathcal{X}$.

Assumption 4.2. The problem functions $f(x)$ and $c(x)$ are twice continuously differentiable on $\mathcal{X}$.

Assumption 4.3. The matrices B_k are uniformly positive definite and bounded, i.e., there exists values $0 < \lambda_{\min} < \lambda_{\max} < \infty$ such that $\lambda_{\min} \|s\|_2^2 \le s^T B_k s \le \lambda_{\max} \|s\|_2^2$ for all $s \in \mathbb{R}^n$ and all B_k.

Assumption 4.4. The matrices H_k are uniformly bounded, i.e., $\|H_k\|_2 \le \mu_{\max}$ for some $\mu_{\max} \ge 1$.

For clarity and motivational purposes, we immediately state our main convergence theorem, which makes use of the Mangasarian–Fromovitz constraint qualification (MFCQ) [35].

THEOREM 4.1. *If Assumptions* 4.1–4.4 *hold, then one of the following must occur:*

(i) *Algorithm* 1 *terminates finitely with either a first-order KKT point or an infeasible stationary point in lines* 10 *or* 7*, respectively, for problem* (1.1).

(ii) *Algorithm* 1 *generates infinitely many iterations* $\{x_k\}$, $\sigma_k = \bar{\sigma} < \infty$ *for all k sufficiently large, and there exists a limit point x_* of $\{x_k\}$ that is either a first-order KKT point or an infeasible stationary point for problem* (1.1).

(iii) *Algorithm* 1 *generates infinitely many iterations* $\{x_k\}$, $\lim_{k\to\infty} \sigma_k = \infty$*, and there exists a limit point x_* of $\{x_k\}$ that is either an infeasible stationary point or a feasible point at which the MFCQ fails.*

Proof. The result follows from the following analysis that considers the various cases that can occur. In particular, it follows from Theorems 4.11, 4.14, 4.18, and 4.21 and the construction of Algorithm 1. ∎

We now present a sequence of lemmas that will be useful in the convergence analysis. The first result is adapted from [9, Theorem 3.6] and provides a bound on the trial step s_k.

LEMMA 4.2. *If Assumptions* 4.1–4.3 *hold and x_k and s_k are generated by Algorithm* 1*, then*

$$\|s_k\|_2 \le \max\left\{1, \frac{2}{\lambda_{min}}\big[\, \|g_k\|_2 + \sigma_k v(x_k)\big], \sqrt{n}\delta_{max}\right\}. \tag{4.1}$$

Furthermore, if $\{\sigma_k\}$ *is bounded, then there exists a constant* $M_s > 0$ *such that* $\|s_k\|_2 \le M_s$ *for all* k.

Proof. First, we claim that the predictor step s_k^p must satisfy

$$\|s_k^p\|_2 \le \max\left\{1, \frac{2}{\lambda_{\min}}\left[\,\|g_k\|_2 + \sigma_k v(x_k)\right]\right\}, \tag{4.2}$$

which can be seen as follows. Suppose that (4.2) is not satisfied so that

$$\|s_k^p\|_2 > 1 \quad \text{and} \quad \tfrac{1}{2}\lambda_{\min}\|s_k^p\|_2 > \|g_k\|_2 + \sigma_k v(x_k). \tag{4.3}$$

It then follows from the definitions of Δq^ϕ and ℓ^v, the Cauchy–Schwarz inequality, Assumption 4.3, and (4.3) that

$$\begin{aligned}\Delta q^\phi(s_k^p; x_k, B_k, \sigma_k) &= -g_k^T s_k^p - \tfrac{1}{2}s_k^{p\,T} B_k s_k^p + \sigma_k\big(\ell^v(0;x_k) - \ell^v(s_k^p;x_k)\big)\\ &\le \|g_k\|_2\,\|s_k^p\|_2 - \tfrac{1}{2}\lambda_{\min}\|s_k^p\|_2^2 + \sigma_k v(x_k)\\ &\le \|g_k\|_2\,\|s_k^p\|_2 - \tfrac{1}{2}\lambda_{\min}\|s_k^p\|_2^2 + \|s_k^p\|_2\,\sigma_k v(x_k)\\ &= \|s_k^p\|_2\Big(\,\|g_k\|_2 - \tfrac{1}{2}\lambda_{\min}\|s_k^p\|_2 + \sigma_k v(x_k)\Big) < 0,\end{aligned}$$

which contradicts the fact that s_k^p is the unique global minimizer to the strictly convex predictor problem. Thus, (4.2) must hold and when combined with (2.15), the use of the triangle inequality, and the use of the trust-region radius $\delta_k \in [\delta_{\min}, \delta_{\max}]$ in the steering problem implies that

$$\|s_k\|_2 \le \max\left\{1, \frac{2}{\lambda_{\min}}\left[\,\|g_k\|_2 + \sigma_k v(x_k)\right], \sqrt{n}\delta_{\max}\right\}, \tag{4.4}$$

which proves (4.1). Since g_k and $v(x_k)$ are uniformly bounded as a result of Assumptions 4.1 and 4.2, it is clear that if $\{\sigma_k\}$ is bounded, then there exists $M_s < \infty$ such that $\|s_k\|_2 \le M_s$ for all k. ∎

The following result provides a relationship between the predicted change in the linear model and the change achieved in the line search process for both the objective function and the constraint violation.

LEMMA 4.3 (equivalent to [45, Lemma 3]). *Suppose that Assumptions* 4.1 *and* 4.2 *hold. Then, there exist constants* $\{C_f, C_v\} > 0$ *such that for all* k *and* $\alpha \in (0,1]$, *we have*

$$f(x_k + \alpha s) \le f(x_k) - \alpha\Delta\ell^f(s;x_k) + \alpha^2 C_f\|s\|_2^2 \tag{4.5}$$

and

$$v(x_k + \alpha s) \le v(x_k) - \alpha\Delta\ell^v(s;x_k) + \alpha^2 C_v\|s\|_2^2. \tag{4.6}$$

Proof. Inequality (4.5) is a direct result of Taylor's theorem and Assumption 4.2.

For (4.6), it follows from the integral mean-value theorem, Assumptions 4.1 and 4.2, and the implied Lipschitz continuity of $J(x)$, the triangle inequality, and the convexity of ℓ^v, that for some constant Lipschitz constant C,

$$
\begin{aligned}
&v(x_k + \alpha s)\\
&\quad = \left\| [c(x_k + \alpha s)]^- \right\|_1 = \left\| \left[c(x_k) + \alpha J_k s + \alpha \int_0^1 [J(x_k + \theta\alpha s) - J(x_k)] s d\theta \right]^- \right\|_1\\
&\quad \le \left\| [c(x_k) + \alpha J_k s]^- \right\|_1 + \alpha^2 \sqrt{n} C \left\| s \right\|_2^2\\
&\quad \le (1-\alpha) \left\| [c(x_k)]^- \right\|_1 + \alpha \left\| [c(x_k) + J_k s]^- \right\|_1 + \alpha^2 \sqrt{n} C \left\| s \right\|_2^2\\
&\quad = v(x_k) - \alpha \Delta\ell^v(s; x_k) + \alpha^2 \sqrt{n} C \left\| s \right\|_2^2 \quad \text{for all } \alpha \in (0,1].
\end{aligned}
$$

This proves (4.6) by defining $C_v := \sqrt{n}C$. □

The next two lemmas provide a relationship between the predicted linear decrease in the objective function and the quantity ρ_k^f defined by (2.32).

LEMMA 4.4. *If Assumption* 4.4 *holds and* $\Delta\ell^f(s_k; x_k) \ge 0$, *then*

$$
\Delta q^f(s_k^{cf}; x_k, H_k) \ge \tfrac{1}{2} \Delta\ell^f(s_k; x_k) \min\left\{ \frac{\Delta\ell^f(s_k; x_k)}{\mu_{max} \left\| s_k \right\|_2^2}, 1 \right\}. \tag{4.7}
$$

Proof. If $\Delta\ell^f(s_k; x_k) = 0$, then the result follows immediately from the definition of s_k^{cf} in (2.27).

Now, suppose that $\Delta\ell^f(s_k; x_k) > 0$. It follows from (2.27) and the definition of Δq^f that

$$
\Delta q^f(s_k^{cf}; x_k, H_k) \ge \Delta q^f(\alpha s_k; x_k, H_k) = -\alpha g_k^T s_k - \tfrac{1}{2}\alpha^2 s_k^T H_k s_k \;\text{ for all } 0 \le \alpha \le 1.
$$

The right-hand side of the previous equation may be written as

$$
q(\alpha) = a\alpha^2 + b\alpha, \;\text{ where }\; a = -\tfrac{1}{2} s_k^T H_k s_k \;\text{ and }\; b = \Delta\ell^f(s_k; x_k) = -g_k^T s_k > 0.
$$

We wish to maximize q on the interval $[0,1]$ so we differentiate $q(\alpha)$ with respect to α and set the result to zero to obtain a stationary point at $-\frac{b}{2a}$. Now, consider three cases.

Case 1 ($a < 0$ and $-\frac{b}{2a} \le 1$). The maximum of $q(\alpha)$ on the interval $[0,1]$ is achieved at $\alpha = -\frac{b}{2a}$. Note that $\alpha > 0$, since $b = \Delta\ell^f(s_k; x_k) > 0$ by assumption. Then, we have

$$
q\left(-\frac{b}{2a}\right) = a\frac{b^2}{4a^2} - b\frac{b}{2a} = -\frac{b^2}{4a}.
$$

It follows from the definition of a and b, the Cauchy–Schwarz inequality, and Assumption 4.4 that

$$
q\left(-\frac{b}{2a}\right) = \frac{\Delta\ell^f(s_k; x_k)^2}{2 s_k^T H_k s_k} \ge \frac{\Delta\ell^f(s_k; x_k)^2}{2 \left\| H_k \right\|_2 \left\| s_k \right\|_2^2} \ge \frac{\Delta\ell^f(s_k; x_k)^2}{2\mu_{\max} \left\| s_k \right\|_2^2}.
$$

Case 2 ($a < 0$ and $-\frac{b}{2a} > 1$). The maximum of $q(\alpha)$ on the interval $[0,1]$ is achieved at $\alpha = 1$, where

$$
q(1) = a + b > -\tfrac{1}{2}b + b = \tfrac{1}{2}b = \tfrac{1}{2}\Delta\ell^f(s_k; x_k).
$$

Case 3 ($a \ge 0$). The maximum of $q(\alpha)$ on the interval $[0,1]$ is achieved at $\alpha = 1$ so that

$$
q(1) = a + b > b > \tfrac{1}{2}b = \tfrac{1}{2}\Delta\ell^f(s_k; x_k).
$$

Finally, combining all three cases and defining $\alpha' = \arg\max_{\alpha\in[0,1]} q(\alpha)$, it follows that

$$\Delta q^f(s_k^{cf}; x_k, H_k) = q(\alpha') \geq \min\left\{\frac{\Delta\ell^f(s_k;x_k)^2}{2\mu_{\max}\|s_k\|_2^2},\ \tfrac{1}{2}\Delta\ell^f(s_k;x_k)\right\}$$
$$= \tfrac{1}{2}\Delta\ell^f(s_k;x_k)\min\left\{\frac{\Delta\ell^f(s_k;x_k)}{\mu_{\max}\|s_k\|_2^2},\ 1\right\}$$

as desired. ▯

LEMMA 4.5. *Suppose that Assumptions* 4.1–4.4 *are satisfied and that* $\{\sigma_k\}$ *is bounded. Then, there exists a constant* $C_\rho > 0$ *such that whenever* $\Delta\ell^f(s_k;x_k) \geq 0$, *it follows that*

$$\rho_k^f \geq \min\left[C_\rho \Delta\ell^f(s_k;x_k)^2,\ \tfrac{1}{2}\Delta\ell^f(s_k;x_k)\right]. \tag{4.8}$$

Proof. It follows from (2.32), Lemma 4.4, Lemma 4.2, and the assumption $\Delta\ell^f(s_k;x_k) \geq 0$ that

$$\begin{aligned}
\rho_k^f &= \min\left[\Delta\ell^f(s_k;x_k),\ \Delta q^f(s_k^{cf};x_k,H_k)\right]\\
&\geq \min\left[\Delta\ell^f(s_k;x_k),\ \min\left\{\frac{\Delta\ell^f(s_k;x_k)^2}{2\mu_{\max}\|s_k\|_2^2},\ \tfrac{1}{2}\Delta\ell^f(s_k;x_k)\right\}\right]\\
&\geq \min\left[\frac{\Delta\ell^f(s_k;x_k)^2}{2\mu_{\max}\|s_k\|_2^2},\ \tfrac{1}{2}\Delta\ell^f(s_k;x_k)\right] \geq \min\left[\frac{\Delta\ell^f(s_k;x_k)^2}{2\mu_{\max}M_s^2},\ \tfrac{1}{2}\Delta\ell^f(s_k;x_k)\right],
\end{aligned}$$

where $\{M_s, \mu_{\max}\} \subset (0,\infty)$ are defined in (4.1) and Assumption 4.4, respectively. The result now follows by defining $C_\rho := 1/(2\mu_{\max}M_s^2)$. ▯

The next two results provide a relationship between the predicted linear change in the penalty function and the quantity ρ_k^ϕ defined by (2.35).

LEMMA 4.6. *If Assumption* 4.4 *holds and* x_k *is not an infeasible stationary point, then*

$$\Delta q^\phi(s_k^{c\phi}; x_k, H_k, \sigma_{k+1}) \geq \tfrac{1}{2}\Delta\ell^\phi(s_k;x_k,\sigma_{k+1})\min\left\{\frac{\Delta\ell^\phi(s_k;x_k,\sigma_{k+1})}{\mu_{max}\|s_k\|_2^2},\ 1\right\}. \tag{4.9}$$

Proof. Since x_k is not an infeasible stationary point by assumption, it follows from Lemma 2.8 that $\Delta\ell^\phi(s_k;x_k,\sigma_{k+1}) \geq 0$. If $\Delta\ell^\phi(s_k;x_k,\sigma_{k+1}) = 0$, then the result follows immediately. Therefore, for the remainder of the proof we assume that $\Delta\ell^\phi(s_k;x_k,\sigma_{k+1}) > 0$.

It follows from (2.28), the convexity of $\ell^v(\cdot)$, and simple algebra that

$$\begin{aligned}
&\Delta q^\phi(s_k^{c\phi};x_k,H_k,\sigma_{k+1})\\
&\quad\geq \Delta q^\phi(\alpha s_k;x_k,H_k,\sigma_{k+1})\\
&\quad= -\alpha g_k^T s_k - \tfrac{1}{2}\alpha^2 s_k^T H_k s_k + \sigma_{k+1}\left(\left\|[c_k]^-\right\|_1 - \left\|[c_k+\alpha J_k s_k]^-\right\|_1\right)\\
&\quad\geq -\alpha g_k^T s_k - \tfrac{1}{2}\alpha^2 s_k^T H_k s_k + \sigma_{k+1}\left(\left\|[c_k]^-\right\|_1 - \alpha\left\|[c_k+J_k s_k]^-\right\|_1 - (1-\alpha)\left\|[c_k]^-\right\|_1\right)\\
&\quad= -\alpha g_k^T s_k - \tfrac{1}{2}\alpha^2 s_k^T H_k s_k + \alpha\sigma_{k+1}\left(\left\|[c_k]^-\right\|_1 - \left\|[c_k+J_k s_k]^-\right\|_1\right)\\
&\quad= \alpha\Delta\ell^\phi(s_k;x_k,\sigma_{k+1}) - \tfrac{1}{2}\alpha^2 s_k^T H_k s_k \ \text{ for all } \alpha\in[0,1].
\end{aligned}$$

The right-hand side of the equation is a quadratic function of α:

$$q(\alpha) = a\alpha^2 + b\alpha, \quad \text{where } a = -\tfrac{1}{2}s_k^T H_k s_k \text{ and } b = \Delta\ell^\phi(s_k; x_k, \sigma_{k+1}) > 0.$$

Analysis similar to that used in the proof of Lemma 4.4 yields

$$\Delta q^\phi(s_k^{c\phi}; x_k, H_k, \sigma_{k+1}) \geq \min\left\{ \frac{\Delta\ell^\phi(s_k; x_k, \sigma_{k+1})^2}{2\mu_{\max}\|s_k\|_2^2}, \tfrac{1}{2}\Delta\ell^\phi(s_k; x_k, \sigma_{k+1}) \right\}, \tag{4.10}$$

where $\mu_{\max}$ is from Assumption 4.4, as desired. ∎

LEMMA 4.7. *Suppose that Assumptions* 4.1–4.4 *are satisfied, that Algorithm* 1 *never encounters an infeasible stationary point, and that* $\{\sigma_k\}$ *is bounded. Then, there exists a constant* $C_\rho \in (0, \infty)$ *such that*

$$\rho_k^\phi \geq \min\left[C_\rho \Delta\ell^\phi(s_k; x_k, \sigma_{k+1})^2, \tfrac{1}{2}\Delta\ell^\phi(s_k; x_k, \sigma_{k+1})\right] \quad \textit{for all } k \geq 0.$$

Proof. The proof follows exactly as in Lemma 4.5. ∎

4.1. Convergence analysis under bounded weighting parameter. In this section we study Algorithm 1 under the assumption that the weighting parameter stays bounded. It follows from this assumption and Lemma 4.2 that there exists some k' and $\bar\sigma < \infty$ such that

$$\|s_k\|_2 \leq M_s < \infty \quad \text{and} \quad \sigma_k = \bar\sigma < \infty \quad \text{for all } k \geq k'. \tag{4.11}$$

Part (iii) of Theorem 4.1 implies that this scenario is guaranteed to occur, for example, when all limit points are neither infeasible stationary points nor feasible points at which the MFCQ fails.

We begin by showing that the line search step length is bounded away from zero in certain situations.

LEMMA 4.8. *If Assumptions* 4.1–4.3 *and* (4.11) *hold and* $\epsilon > 0$*, then the following hold:*

(i) *There exists a constant* $\alpha_P > 0$ *such that* $\alpha_k \geq \alpha_P > 0$ *for all* $k \in \mathcal{K}_P$*, where*

$$\mathcal{K}_P = \{k \in \mathcal{S}_p : k \geq k' \textit{ and } \Delta\ell^\phi(s_k; x_k, \bar\sigma) \geq \epsilon\}.$$

(ii) *There exists a constant* $\alpha_F > 0$ *such that* $\alpha_k \geq \alpha_F > 0$ *for all* $k \in \mathcal{K}_F$*, where*

$$\mathcal{K}_F = \{k \in \mathcal{S}_v \cup \mathcal{S}_o \cup \mathcal{S}_b : k \geq k' \textit{ and } \Delta\ell^v(s_k^s; x_k) \geq \epsilon\}.$$

(iii) *There exists a constant* $\alpha_f > 0$ *such that* $(\alpha, s) = (\alpha, s_k)$ *satisfies* (2.31b) *for all* $0 < \alpha \leq \alpha_f$ *and all* $k \in \mathcal{K}_f$*, where*

$$\mathcal{K}_f = \{k \geq k' : \Delta\ell^f(s_k; x_k) \geq \epsilon\}.$$

Proof. From [8, Lemma 3.4], there exists some positive constant C_ϕ such that

$$\left|\phi(x_k + \alpha s_k; \bar\sigma) - \ell^\phi(\alpha s_k; x_k, \bar\sigma)\right| \leq C_\phi \|\alpha s_k\|_2^2 \quad \text{for all } k \geq k' \text{ and } \alpha \in [0, 1]. \tag{4.12}$$

We first prove part (i). Suppose that α satisfies

$$0 \leq \alpha \leq \frac{(1 - \gamma_\phi)\epsilon}{C_\phi M_s^2}, \tag{4.13}$$

where $\gamma_\phi \in (0,1)$ is set in Algorithm 1 and M_s is defined in (4.11). We then use $\phi(x_k;\bar\sigma) = \ell^\phi(0;x_k,\bar\sigma)$, the convexity of $\ell^\phi(\,\cdot\,;x_k,\bar\sigma)$, (4.12), $\Delta\ell^\phi(s_k;x_k,\bar\sigma) \ge \epsilon$ for $k \in \mathcal{K}_P$, (4.11), (4.13), and (2.35) to conclude that

$$\begin{aligned}
&\phi(x_k;\bar\sigma) - \phi(x_k+\alpha s_k;\bar\sigma)\\
&\quad = \left[\ell^\phi(0;x_k,\bar\sigma) - \ell^\phi(\alpha s_k;x_k,\bar\sigma)\right] - \left[\phi(x_k+\alpha s_k;\bar\sigma) - \ell^\phi(\alpha s_k;x_k,\bar\sigma)\right]\\
&\quad \ge \left[\ell^\phi(0;x_k,\bar\sigma) - \alpha\ell^\phi(s_k;x_k,\bar\sigma) - (1-\alpha)\ell^\phi(0;x_k,\bar\sigma)\right] - C_\phi\alpha^2\|s_k\|_2^2\\
&\quad = \alpha\left[\ell^\phi(0;x_k,\bar\sigma) - \ell^\phi(s_k;x_k,\bar\sigma)\right] - C_\phi\alpha^2\|s_k\|_2^2\\
&\quad = \gamma_\phi\alpha\Delta\ell^\phi(s_k;x_k,\bar\sigma) + (1-\gamma_\phi)\alpha\Delta\ell^\phi(s_k;x_k,\bar\sigma) - C_\phi\alpha^2\|s_k\|_2^2\\
&\quad \ge \gamma_\phi\alpha\Delta\ell^\phi(s_k;x_k,\bar\sigma) + (1-\gamma_\phi)\alpha\epsilon - C_\phi\alpha^2\|s_k\|_2^2\\
&\quad \ge \gamma_\phi\alpha\Delta\ell^\phi(s_k;x_k,\bar\sigma) \ge \gamma_\phi\alpha\rho_k^\phi \;\text{ for all } k \in \mathcal{K}_P,
\end{aligned}$$

which with (2.34) implies that (α, s_k) is a p-pair. Thus, Algorithm 1 must select an α_k that satisfies

$$\alpha_k \ge \min\left\{\frac{\xi(1-\gamma_\phi)\epsilon}{C_\phi M_s^2}, 1\right\} =: \alpha_{\mathrm{P}}, \tag{4.14}$$

where $\xi \in (0,1)$ is the backtracking parameter in Algorithm 1, which completes the proof of part (i).

We now prove part (ii). It follows from (2.20) that

$$\Delta\ell^\phi(s_k;x_k,\bar\sigma) \ge \bar\sigma\eta_\sigma\Delta\ell^v(s_k^s;x_k) \ge \bar\sigma\eta_\sigma\epsilon \;\text{ for } k \in \mathcal{K}_F. \tag{4.15}$$

If α satisfies

$$\alpha < \min\left[\frac{(1-\gamma_\phi)\bar\sigma\eta_\sigma\epsilon}{C_\phi M_s^2}, \frac{\eta_v\epsilon}{C_v M_s^2}\right], \tag{4.16}$$

where C_v is defined in (4.6) and η_v is defined in (2.16), then we may use (4.15) and proceed as in the proof of part (i) to conclude that (2.34) holds. Moreover, we have from Lemma 4.3, (2.16), (4.16), $\Delta\ell^v(s_k^s;x_k) \ge \epsilon$ for $k \in \mathcal{K}_F$, and (4.11) that

$$\begin{aligned}
&v(x_k+\alpha s_k) - v(x_k)\\
&\quad \le -\alpha\Delta\ell^v(s_k;x_k) + \alpha^2 C_v\|s_k\|_2^2 < -\alpha\eta_v\Delta\ell^v(s_k^s;x_k) + \alpha\frac{\eta_v\epsilon}{C_v M_s^2}C_v\|s_k\|_2^2\\
&\quad \le -\alpha\eta_v\epsilon + \alpha\eta_v\epsilon = 0 \;\text{ for all } k \in \mathcal{K}_F,
\end{aligned} \tag{4.17}$$

where the strict inequality holds since $s_k \ne 0$ as a result of (4.15). Combining (4.17) with (2.34) implies that (α, s_k) is a b-pair. Thus, we conclude from the structure of Algorithm 1 that

$$\alpha_k \ge \min\left\{\frac{\xi(1-\gamma_\phi)\bar\sigma\eta_\sigma\epsilon}{C_\phi M_s^2}, \frac{\xi\eta_v\epsilon}{C_v M_s^2}, 1\right\} =: \alpha_{\mathrm{F}} > 0 \;\text{ for all } k \in \mathcal{K}_F, \tag{4.18}$$

where $\xi \in (0,1)$ is the backtracking parameter used in Algorithm 1.

Part (iii) is a standard result used in continuous unconstrained optimization that follows since $\Delta\ell^f(s_k;x_k) \ge \epsilon$ is equivalent to $g(x_k)^T s_k \le -\epsilon < 0$ and s_k is uniformly bounded by (4.11). ∎

The next lemma justifies the three cases that we consider when analyzing Algorithm 1.

LEMMA 4.9. *If Algorithm* 1 *does not terminate finitely, then one of the following scenarios occurs:*

Case 1. $k \in \mathcal{S}_p$ *for all* k *sufficiently large,*

Case 2. $k \in \mathcal{S}_o$ *for all* k *sufficiently large, or*

Case 3. $|\mathcal{S}_v \cup \mathcal{S}_b| = \infty$.

Proof. We proceed by contradiction and assume that none of the cases occurs. In particular, since Case 3 does not hold it follows that $k \in \mathcal{S}_p \cup \mathcal{S}_o$ for all k sufficiently large. Combining this with the fact that Cases 1 and 2 do not hold implies that the iterates must oscillate between p- and o-iterates. However, this is not possible since there is no mechanism in Algorithm 1 that allows for iterate $k+1$ to be a p-iterate if iterate k is an o-iterate. ◻

We now analyze Algorithm 1 for each of the three possible scenarios stated in the previous result.

Case 1: $k \in \mathcal{S}_p$ for all k sufficiently large. In this case, there exists k'' such that

$$k \in \mathcal{S}_p \quad \text{for all } k \geq k'' \geq k', \tag{4.19}$$

where k' is defined in (4.11). We first show that our measure of feasibility converges to zero.

LEMMA 4.10. *If Assumptions* 4.1–4.4*,* (4.11)*, and* (4.19) *hold, then* $\lim_{k\to\infty} \Delta\ell^v(s_k^s; x_k) = 0$.

Proof. For a proof by contradiction, suppose that there exists an infinite subsequence

$$\mathcal{S}'' := \{k \geq k'' : \Delta\ell^v(s_k^s; x_k) \geq \epsilon''\}$$

for some constant $\epsilon'' > 0$. It follows from (4.19), (2.20), (4.11), and the definition of $\mathcal{S}''$ that

$$\Delta\ell^\phi(s_k; x_k, \bar\sigma) \geq \bar\sigma\eta_\sigma\Delta\ell^v(s_k^s; x_k) \geq \bar\sigma\eta_\sigma\epsilon'' =: \epsilon > 0 \ \text{ for all } k \in \mathcal{S}'', \tag{4.20}$$

which implies with (4.19) that

$$\mathcal{S}'' \subseteq \mathcal{K}_P := \{k \in \mathcal{S}_p : k \geq k' \ \text{ and } \ \Delta\ell^\phi(s_k; x_k, \bar\sigma) \geq \epsilon > 0\}.$$

Combining $\mathcal{K}_P$ with Lemma 4.8 implies the existence of a positive α_{P} such that $\alpha_k \geq \alpha_{\mathrm{P}} > 0$ for all $k \in \mathcal{S}''$, which used with the definitions of $\mathcal{S}''$ and $\mathcal{S}_p$, (4.19), Lemma 4.7, and (4.20) yields

$$\begin{aligned} \phi(x_k; \bar\sigma) - \phi(x_k + \alpha_k\hat s_k; \bar\sigma) &\geq \gamma_\phi\alpha_k\rho_k^\phi \\ &\geq \gamma_\phi\alpha_k \min\left[C_\rho\Delta\ell^\phi(s_k; x_k, \bar\sigma)^2,\ \tfrac12\Delta\ell^\phi(s_k; x_k, \bar\sigma)\right] \\ &\geq \gamma_\phi\alpha_{\mathrm{P}} \min\left[C_\rho\epsilon^2,\ \tfrac12\epsilon\right] > 0 \ \text{ for all } k \in \mathcal{S}''. \end{aligned} \tag{4.21}$$

Now, for $k'' \leq k \in \mathcal{S}_p \setminus \mathcal{S}''$, it follows from (2.34), (2.20), and (2.28) that

$$\phi(x_k; \bar\sigma) - \phi(x_k + \alpha_k\hat s_k; \bar\sigma) \geq \gamma_\phi\alpha_k \min\left[\Delta\ell^\phi(s_k; x_k, \bar\sigma),\ \Delta q^\phi(s_k^{c\phi}; x_k, H_k, \bar\sigma)\right] \geq 0. \tag{4.22}$$

It is now easy to see from (4.21), (4.22), and (4.19) that $\lim_{k\to\infty} \phi(x_k; \bar\sigma) = -\infty$, which contradicts Assumptions 4.1 and 4.2. Thus, we conclude that $\lim_{k\to\infty} \Delta\ell^v(s_k^s; x_k) = 0$. ◻

We now show that all limit points are infeasible stationary points for problem (1.1).

THEOREM 4.11. *Suppose that Assumptions* 4.1–4.4, (4.11), *and* (4.19) *hold. If* x_* *is any limit point of the sequence* $\{x_k\}$ *generated by Algorithm* 1, *then* x_* *is an infeasible stationary point for problem* (1.1).

Proof. Let $v_{\min} := \min\{v_j : (v_j, f_j) \in \mathcal{F}_{k''}\} \equiv \min\{v_j : (v_j, f_j) \in \mathcal{F}_k \text{ and } k \geq k''\}$, where the second equality holds since by assumption $k \in \mathcal{S}_p$ for all $k \geq k''$ and the filter is never expanded when $k \in \mathcal{S}_p$. It follows from Lemma 3.1 that $v_{\min} > 0$. But then if there was a feasible limit point x_*, there must be iterates x_k, $k > k''$ that are arbitrarily close to feasibility and thus ultimately one such that x_k is acceptable to $\mathcal{F}_k$. Thus line 21 of Algorithm 1 implies that there will be an iterate $k > k''$ for which $k \notin \mathcal{S}_p$, which contradicts (4.19). Thus, all limit points are infeasible. It follows from this fact, Lemma 4.10, and Lemma 2.1 that all limit points are infeasible stationary points. ∎

Importantly, the previous result shows that our algorithm remains in penalty mode for all k sufficiently large only when all limit points are infeasible stationary points.

Case 2: $k \in \mathcal{S}_o$ for all k sufficiently large. In this case, there exists k'' such that

$$k \in \mathcal{S}_o \quad \text{for all } k \geq k'' \geq k', \tag{4.23}$$

where k' is defined in (4.11). We begin by showing that our feasibility measure converges to zero.

LEMMA 4.12. *If Assumptions* 4.1–4.4, (4.11), *and* (4.23) *hold, then* $\lim_{k\to\infty} \Delta\ell^v(s_k^s; x_k) = 0$.

Proof. For a contradiction, suppose that there exists $\epsilon'' > 0$ and an infinite subsequence

$$\mathcal{S}'' := \{k \geq k'' : \Delta\ell^v(s_k^s; x_k) \geq \epsilon''\} \subseteq \mathcal{S}_o,$$

where we have used k'' defined in (4.23). It then follows from the definition of $\mathcal{S}_o$, the o-pair $(\alpha_k, \hat{s}_k)$ selected in Algorithm 1, (2.31b), (4.11), Lemma 4.5, (2.31a), (2.16), and part (ii) of Lemma 4.8 that

$$\begin{aligned}
f(x_k) - f(x_k + \alpha_k \hat{s}_k) &\geq \gamma_f \alpha_k \rho_k^f \\
&\geq \gamma_f \alpha_k \min\left\{C_\rho \Delta\ell^f(s_k; x_k)^2,\ \tfrac{1}{2}\Delta\ell^f(s_k; x_k)\right\} \\
&\geq \gamma_f \alpha_k \min\left\{C_\rho\big[\gamma_v \Delta\ell^v(s_k; x_k)\big]^2,\ \tfrac{1}{2}\gamma_v \Delta\ell^v(s_k; x_k)\right\} \\
&\geq \gamma_f \alpha_k \min\left\{C_\rho\big[\gamma_v \eta_v \Delta\ell^v(s_k^s; x_k)\big]^2,\ \tfrac{1}{2}\gamma_v \eta_v \Delta\ell^v(s_k^s; x_k)\right\} \\
&\geq \gamma_f \alpha_{\text{F}} \min\left\{C_\rho\big[\gamma_v \eta_v \epsilon''\big]^2,\ \tfrac{1}{2}\gamma_v \eta_v \epsilon''\right\} \quad \text{for all } k \in \mathcal{S}''
\end{aligned}$$

for some $\alpha_F > 0$. Similarly, for $k'' \leq k \in \mathcal{S}_o \setminus \mathcal{S}''$, it follows from (2.31), (2.16), and (2.27) that

$$f(x_k) - f(x_k + \alpha_k \hat{s}_k) \geq \gamma_f \alpha_k \rho_k^f \geq \gamma_f \alpha_k \min\left\{\gamma_v \Delta\ell^v(s_k; x_k),\ \Delta q^f(s_k^{cf}; x_k, H_k)\right\} \geq 0.$$

Combining the two previous inequalities with the definition of k'' yields $\lim_{k\to\infty} f(x_k) = -\infty$, which contradicts the fact that f is bounded as a consequence of Assumptions 4.1 and 4.2. This proves the result. ∎

We now show that feasible limit points are also first-order solutions of the penalty function.

LEMMA 4.13. *Suppose that Assumptions* 4.1–4.4, (4.11), *and* (4.23) *hold. If* $x_* = \lim_{k\in\mathcal{S}} x_k$ *for some subsequence* $\mathcal{S}$ *and* $v(x_*) = 0$, *then* $\lim_{k\in\mathcal{S}} \Delta q^\phi(s_k^p; x_k, B_k, \bar\sigma) = 0$.

Proof. Suppose that there exists a constant $\epsilon'' > 0$ and an infinite subsequence

$$\mathcal{S}'' := \{k \in \mathcal{S} : k \ge k'' : \Delta q^\phi(s_k^p; x_k, B_k, \bar\sigma) \ge \epsilon''\},$$

where k'' is defined in (4.23). It follows from line 34 of Algorithm 1, (4.11), and (4.23) that

$$\Delta q^\phi(s_k; x_k, B_k, \bar\sigma) \ge \eta_\phi \Delta q^\phi(s_k^p; x_k, B_k, \bar\sigma) \ge \eta_\phi\epsilon'' \quad \text{for } k \in \mathcal{S}''. \tag{4.24}$$

From (2.6) and (2.16), we know that $v(x_k) \ge \Delta\ell^v(s_k; x_k) \ge \eta_v \Delta\ell^v(s_k^s; x_k) \ge 0$ for all k, which may be combined with $\lim_{k\in\mathcal{S}} v(x_k) = v(x_*) = 0$ (holds by assumption) to conclude that

$$\Delta\ell^v(s_k; x_k) \le \frac{\eta_\phi\epsilon''}{\bar\sigma + \gamma_v} \quad \text{for } k \in \mathcal{S} \text{ sufficiently large,} \tag{4.25}$$

where $\gamma_v \in (0,1)$ is defined in (2.31a). It follows from (2.8), (2.9), (4.24), $B_k \succ 0$, and (4.25) that

$$\begin{aligned} \Delta\ell^f(s_k; x_k) &\\ &\ge \tfrac{1}{2} s_k^T B_k s_k - \bar\sigma\Delta\ell^v(s_k; x_k) + \eta_\phi\epsilon'' > \eta_\phi\epsilon'' - \bar\sigma\Delta\ell^v(s_k; x_k) \\ &\ge \eta_\phi\epsilon'' - \bar\sigma\frac{\eta_\phi\epsilon''}{\bar\sigma+\gamma_v} = \frac{\gamma_v\eta_\phi\epsilon''}{\bar\sigma+\gamma_v} =: \epsilon^f > 0 \quad \text{for } k \in \mathcal{S}'' \text{ sufficiently large.} \end{aligned} \tag{4.26}$$

Combining this with part (iii) of Lemma 4.8, we know that there exists some $\alpha_f > 0$ such that (α, s_k) satisfies (2.31b) for all $k \in \mathcal{S}''$ sufficiently large and $\alpha \in (0, \alpha_f]$, since by assumption $\mathcal{S}_o = \mathcal{S}_v \cup \mathcal{S}_o \cup \mathcal{S}_b$ for $k \ge k''$.

Next, we define

$$\Phi_k := \min_{(v_i,f_i)\in\mathcal{F}_k} \Big\{ \max\big[\, v_i - \alpha_i\eta_v\Delta\ell^v(s_i^s; x_i),\ \beta v_i \,\big] \Big\} > 0, \tag{4.27}$$

where $\mathcal{F}_k$ is the kth filter. The fact that $\Phi_k > 0$ follows since $v_i > 0$ for all $(v_i, f_i) \in \mathcal{F}_k$ as a consequence of Lemma 3.1. Moreover, it follows from (4.23) that $\mathcal{F}_k \equiv \mathcal{F}_{k''}$ for all $k \in \mathcal{S}''$ so that $\Phi_k \equiv \Phi_{k''} > 0$ for all $k \in \mathcal{S}''$. Now, pick $\epsilon^v > 0$ such that $\Phi_{k''} - C_v M_s^2 \le \epsilon^v < \Phi_{k''}$ and consider α such that

$$0 < \alpha \le \frac{\Phi_{k''} - \epsilon^v}{C_v M_s^2} \le 1. \tag{4.28}$$

It then follows from Lemma 4.3, the fact that $\lim_{x\in\mathcal{S}} v(x_k) = 0$, (2.16), (4.23), and (4.28) that

$$\begin{aligned} v(x_k + \alpha s_k) &\\ &\le v(x_k) - \alpha\Delta\ell^v(s_k; x_k) + \alpha^2 C_v \|s_k\|_2^2 \le \epsilon^v + \alpha^2 C_v M_s^2 \\ &\le \epsilon^v + \alpha C_v M_s^2 \le \epsilon^v + \frac{\Phi_{k''} - \epsilon^v}{C_v M_s^2} C_v M_s^2 = \Phi_{k''} \quad \text{for all } k \in \mathcal{S}'' \text{ sufficiently large.} \end{aligned}$$

Thus, $x_k + \alpha s_k$ is acceptable to $\mathcal{F}_k \equiv \mathcal{F}_{k''}$ for all α satisfying (4.28) and $k \in \mathcal{S}''$ sufficiently large.

Combining the above, (4.23), and the structure of Algorithm 1, we conclude that

$$\alpha_k \geq \min\left\{\xi \frac{\Phi_{k''} - \epsilon^v}{C_v M_s^2}, \xi\alpha_f, 1\right\} =: \alpha_{\min} > 0 \text{ for all } k \in \mathcal{S}'' \text{ sufficiently large,} \tag{4.29}$$

where $\xi \in (0,1)$ is the backtracking parameter used in Algorithm 1. It then follows from (4.23), (2.31b), (2.16), Lemma 4.5, (2.31a), (4.29), and (4.26) that

$$\begin{aligned} &f(x_k) - f(x_k + \alpha_k \hat{s}_k) \\ &\quad \geq \gamma_f \alpha_k \rho_k^f \geq \gamma_f \alpha_k \min\left[C_\rho \Delta\ell^f(s_k; x_k)^2,\ \tfrac{1}{2}\Delta\ell^f(s_k; x_k)\right] \\ &\quad \geq \gamma_f \alpha_{\min} \min\left[C_\rho \left(\epsilon^f\right)^2,\ \tfrac{1}{2}\epsilon^f\right] > 0 \text{ for all } k \in \mathcal{S}'' \text{ sufficiently large.} \end{aligned} \tag{4.30}$$

However, for *all* $k \in \mathcal{S}_o$, it follows from (2.31b), (2.31a), (2.16), and (2.27) that $f(x_k) - f(x_k + \alpha_k \hat{s}_k) \geq 0$. This observation combined with (4.30) implies that $\lim_{k\to\infty} f(x_k) = -\infty$, which contradicts the fact that f is bounded as a consequence of Assumptions 4.1 and 4.2. This completes the proof. □

We now show that limit points are either infeasible stationary points or KKT points for problem (1.1).

THEOREM 4.14. *Suppose that Assumptions* 4.1–4.4, (4.11)*, and* (4.23) *hold. If x_* is a limit point of $\{x_k\}$, then either*

(i) *x_* is an infeasible stationary point or*

(ii) *x_* is a KKT point for problem* (1.1).

Proof. Suppose that $\lim_{k\in\mathcal{S}} x_k = x_*$ for some subsequence $\mathcal{S}$. It follows from Lemma 4.12 that $\lim_{k\to\infty} \Delta\ell^v(s_k^s; x_k) = 0$, so that if $v(x_*) > 0$, then x_* is an infeasible stationary point (see Definition 2.1). Otherwise, we have that $v(x_*) = 0$. In this case, it follows from Lemma 4.13 and (4.11) that $\lim_{k\in\mathcal{S}} \Delta q^\phi(s_k^p; x_k, B_k, \bar{\sigma}) = 0$. It follows from this fact, $v(x_*) = 0$, and Lemma 2.2 that x_* is a KKT point for problem (1.1). □

Case 3: $|\mathcal{S}_v \cup \mathcal{S}_b| = \infty$. The next result shows that if $\mathcal{P}$-*mode* = false at the beginning of the kth iteration, then x_k is acceptable to the filter $\mathcal{F}_k$.

LEMMA 4.15. *If $\mathcal{P}$-mode* = false *at the beginning of iteration k, then x_k is acceptable to $\mathcal{F}_k$.*

Proof. The result immediately follows from the construction of Algorithm 1 and consideration of the possible outcomes associated with iteration $k-1$. □

We first show that the feasibility measure converges to zero along $\mathcal{S}_v \cup \mathcal{S}_b$.

LEMMA 4.16. *If Assumptions* 4.1–4.3 *hold and $|\mathcal{S}_v \cup \mathcal{S}_b| = \infty$, then* $\lim_{k\in\mathcal{S}_v\cup\mathcal{S}_b} \Delta\ell^v(s_k^s; x_k) = 0$.

Proof. To reach a contradiction, suppose that we have the infinite subsequence

$$\mathcal{S} := \{k \in \mathcal{S}_v \cup \mathcal{S}_b : \Delta\ell^v(s_k^s; x_k) \geq \epsilon\}$$

for some constant $\epsilon > 0$. It follows from the definition of $\mathcal{S}$, Lemma 4.15, and (2.29) that

$$\begin{aligned} &v_k \leq \max\left\{v_j - \alpha_j \eta_v \Delta\ell^v(s_j^s; x_j),\ \beta v_j\right\} \quad \text{or} \\ &f_k \leq f_j - \gamma \min\left\{v_j - \alpha_j \eta_v \Delta\ell^v(s_j^s; x_j),\ \beta v_j\right\} \end{aligned} \tag{4.31}$$

for $k \in \mathcal{S}$ and $(v_j, f_j) \in \mathcal{F}_k$; note that by construction $(v_k, f_k) \in \mathcal{F}_{k+1}$ for all $k \in \mathcal{S}$. Moreover, it follows from the definitions of $\Delta\ell^v$ and $\mathcal{S}$ that $v_k \geq \Delta\ell^v(s_k^s; x_k) \geq \epsilon$ for $k \in \mathcal{S}$. Using Assumptions 4.1 and 4.2 we have a subsequence $\mathcal{S}' \subseteq \mathcal{S}$ so that

$$\lim_{k\in\mathcal{S}'} \Delta\ell^v(s_k^s;x_k)=\theta_\ell \quad \text{and} \quad \lim_{k\in\mathcal{S}'} v_k=\theta_v \quad \text{for constants } \theta_v\ge\theta_\ell\ge\epsilon>0.$$

For *any* $\epsilon_\ell\in(0,\theta_\ell)$ and $\epsilon_v\in(0,\theta_v)$, it follows that

$$|\Delta\ell^v(s_k^s;x_k)-\theta_\ell|<\epsilon_\ell \quad \text{and} \quad |v_k-\theta_v|<\epsilon_v \;\text{ for all } k\in\mathcal{S}'\subseteq\mathcal{S} \text{ sufficiently large.} \tag{4.32}$$

Using (4.32), the definitions of ϵ_ℓ, η_v, $\Delta\ell^v$ and $\mathcal{S}$, $\alpha_k\in(0,1]$, $\mathcal{S}'\subseteq\mathcal{S}$, and part (ii) of Lemma 4.8 gives

$$0\le v_k-\alpha_k\eta_v\Delta\ell^v(s_k^s;x_k)<v_k-\alpha_{\mathrm{F}}\eta_v(\theta_\ell-\epsilon_\ell)\le\beta_2 v_k \;\text{ for all } k\in\mathcal{S}' \text{ sufficiently large} \tag{4.33}$$

and some $\alpha_{\mathrm{F}}>0$, where

$$\beta_2:=\frac{(\theta_v+\epsilon_v)-\alpha_{\mathrm{F}}\eta_v(\theta_\ell-\epsilon_\ell)}{(\theta_v+\epsilon_v)}\in(0,1)$$

and β_2 may be forced to lie in $(0,1)$ by choosing ϵ_v sufficiently close to zero and ϵ_ℓ sufficiently close to θ_ℓ. Now define $\beta^*:=\max\{\beta_2,\beta\}\in(0,1)$,

$$\epsilon^*=\min\left\{\frac{1-\beta^*}{1+\beta^*}\theta_v,\ \epsilon_v\right\}>0,$$

and the subsequence $\mathcal{S}''=\{k\in\mathcal{S}':|v_k-\theta_v|<\epsilon^*\}$ so that

$$\frac{2\beta^*}{1+\beta^*}\theta_v<v_k<\frac{2}{1+\beta^*}\theta_v \;\text{ for all } k\in\mathcal{S}''\subseteq\mathcal{S}' \text{ sufficiently large.} \tag{4.34}$$

Given $k\in\mathcal{S}''$, define $k^+\in\mathcal{S}''$ to be the successor to k in $\mathcal{S}''$. It then follows from (4.34), the definition of β^*, and (4.33) that

$$\begin{aligned} v_{k^+}>\frac{2\beta^*}{1+\beta^*}\theta_v>\beta^*v_k&=\max\{\beta_2,\beta\}v_k\\ &\ge\max\left\{v_k-\alpha_k\eta_v\Delta\ell^v(s_k^s;x_k),\ \beta v_k\right\} \;\text{ for all } k\in\mathcal{S}''. \end{aligned}$$

Since $\mathcal{S}''\subseteq\mathcal{S}'\subseteq\mathcal{S}$, it follows from the previous inequality, the definition of k^+, the fact that $(v_k,f_k)\in\mathcal{F}_{k^+}$, (4.31), the definition of $\Delta\ell^v(s_k^s;x_k)$, $\alpha_k\in(0,1]$, $\eta_v\in(0,1)$, $\beta\in(0,1)$, $\gamma\in(0,1)$, $\theta_v>\epsilon_v\ge\epsilon^*$, and the definition of $\mathcal{S}''$ that

$$\begin{aligned} f_k-f_{k^+}&\ge\gamma\min\left\{v_k-\alpha_k\eta_v\Delta\ell^v(s_k^s;x_k),\ \beta v_k\right\}\\ &=\gamma\min\left\{(1-\alpha_k\eta_v)v_k+\alpha_k\eta_v\left\|[c(x_k)+J(x_k)s_k^s]^-\right\|_1,\ \beta v_k\right\}\\ &\ge\gamma\min\{1-\alpha_k\eta_v,\beta\}v_k\ge\gamma\min\{1-\eta_v,\beta\}(\theta_v-\epsilon^*)>0 \;\text{ for all } k \text{ in } \mathcal{S}''. \end{aligned}$$

Summing over $k\in\mathcal{S}''$, we deduce that $\lim_{k\in\mathcal{S}''}f(x_k)=-\infty$, which contradicts Assumptions 4.1 and 4.2. ∎

We now prove that our optimality measure for ϕ converges to zero along a certain subsequence.

LEMMA 4.17. *Suppose that Assumptions* 4.1–4.4 *and* (4.11) *hold and that* $|\mathcal{S}_v\cup\mathcal{S}_b|=\infty$.

(i) *If* $|\mathcal{S}_v|=\infty$ *and* $\lim_{k\in\mathcal{S}_v}x_k=x_*$ *for some* x_* *satisfying* $v(x_*)=0$, *then*

$$\lim_{k\in\mathcal{S}_v}\Delta q^\phi(s_k^p;x_k,B_k,\bar\sigma)=0.$$

(ii) *If* $|\mathcal{S}_v| < \infty$ *and* $\lim_{k\in\mathcal{S}_b} x_k = x_*$ *for some* x_* *satisfying* $v(x_*) = 0$*, then*

$$\liminf_{k\in\mathcal{S}_b} \Delta q^\phi(s_k^p; x_k, B_k, \bar{\sigma}) = 0.$$

Proof. We first prove part (i). To obtain a contradiction, suppose that there exists the subsequence

$$\mathcal{S}' := \{k \in \mathcal{S}_v : k \geq k' \text{ and } \Delta q^\phi(s_k^p; x_k, B_k, \bar{\sigma}) \geq \epsilon'\}$$

for some constant $\epsilon' > 0$ and k' defined in (4.11). It then follows from line 34 of Algorithm 1 that

$$\Delta q^\phi(s_k; x_k, B_k, \bar{\sigma}) \geq \eta_\phi \Delta q^\phi(s_k^p; x_k, B_k, \bar{\sigma}) \geq \eta_\phi \epsilon' \text{ for } k \in \mathcal{S}'. \tag{4.35}$$

Then, since $v(x_*) = 0$ by assumption, we may use (4.35) (analogous to (4.24)) and follow the same steps that led to (4.26) to show that

$$\Delta\ell^f(s_k; x_k) \geq \epsilon^f \geq \gamma_v \Delta\ell^v(s_k; x_k) \text{ for } k \in \mathcal{S}' \text{ sufficiently large and some } \epsilon^f > 0,$$

where the second inequality follows from $\lim_{k\in\mathcal{S}_v} x_k = x_*$, $v(x_*) = 0$, and the definition of $\Delta\ell^v$. Thus, (2.30) does not hold, which implies that $k \notin \mathcal{S}_v$. This is a contradiction and proves part (i).

We now prove part (ii), where $|\mathcal{S}_v| < \infty = |\mathcal{S}_b|$. To obtain a contradiction, suppose that

$$\Delta q^\phi(s_k^p; x_k, B_k, \bar{\sigma}) \geq \epsilon' \text{ for } k \in \mathcal{S}_b \text{ sufficiently large}$$

and some constant $\epsilon' > 0$. It then follows from line 34 of Algorithm 1 that

$$\Delta q^\phi(s_k; x_k, B_k, \bar{\sigma}) \geq \eta_\phi \Delta q^\phi(s_k^p; x_k, B_k, \bar{\sigma}) \geq \eta_\phi \epsilon' \text{ for } k \in \mathcal{S}_b \text{ sufficiently large.} \tag{4.36}$$

Since (4.36) is analogous to (4.35), we may again conclude as above that

$$\Delta\ell^f(s_k; x_k) \geq \epsilon^f \geq \gamma_v \Delta\ell^v(s_k; x_k) \text{ for } k \in \mathcal{S}_b \text{ sufficiently large and some } \epsilon^f > 0. \tag{4.37}$$

Using (4.37), $\lim_{k\in\mathcal{S}_b} v(x_k) = v(x_*) = 0$, and part (iii) of Lemma 4.8, we may conclude that there exists $\alpha_f > 0$ such that (α, s_k) satisfies (2.31b) for all $\alpha \in (0, \alpha_f]$ and $k \in \mathcal{S}_b$ sufficiently large. Now, if $\alpha_k \to 0$ along some subsequence $\mathcal{S}_b' \subseteq \mathcal{S}_b$, then it follows from the previous sentence and (4.37) that (α_k, s_k) satisfies (2.31a) and (2.31b) for all $k \in \mathcal{S}_b$ sufficiently large. We now show that $x_k + \alpha_k s_k$ is also acceptable to the filter $\mathcal{F}_k$ for all $k \in \mathcal{S}_b'$ sufficiently large.

To this end, let $(v_i, f_i) \in \mathcal{F}_k$ for some $k \in \mathcal{S}_b'$. It then follows from Lemma 4.15 that either $v_k \leq \max\{v_i - \alpha_i\eta_v\Delta\ell^v(s_i^s; x_i),\, \beta v_i\}$ or $f_k \leq f_i - \gamma\min\{v_i - \alpha_i\eta_v\Delta\ell^v(s_i^s; x_i),\, \beta v_i\}$. In this first case, it follows from the definition of a b-pair that $v(x_k + \alpha_k s_k) \leq v_k \leq \max\{v_i - \alpha_i\eta_v\Delta\ell^v(s_i^s; x_i),\, \beta v_i\}$ for all $k \in \mathcal{S}_b$. In the second case, we have from the fact that (2.31b) holds for $k \in \mathcal{S}_b'$ sufficiently large (recall that $\alpha_k \to 0$ on $\mathcal{S}_b'$), (4.37), and Lemma 4.5 that $f(x_k + \alpha_k s_k) \leq f_k \leq f_i - \gamma\min\{v_i - \alpha_i\eta_v\Delta\ell^v(s_i^s; x_i),\, \beta v_i\}$. Thus, in either case we have that $\big(v(x_k + \alpha_k s_k), f(x_k + \alpha_k s_k)\big)$ is acceptable to the single element filter $\{(v_i, f_i)\}$ for all $k \in \mathcal{S}_b'$ sufficiently large. Since this element (v_i, f_i) of the filter $\mathcal{F}_k$ was arbitrary, we may conclude that $\big(v(x_k + \alpha_k s_k), f(x_k + \alpha_k s_k)\big)$ is, in fact, acceptable to the filter $\mathcal{F}_k$ for all $k \in \mathcal{S}_b'$ sufficiently large.

To summarize, we have shown that (α_k, s_k) is an o-pair for $k \in \mathcal{S}_b'$ sufficiently large. This is a contradiction since Algorithm 1 would have labeled such an iterate as an o-iterate, not a b-iterate. Thus, there exists α_b such that $\alpha_k \geq \alpha_b > 0$ for all $k \in \mathcal{S}_b$ sufficiently large. Combining this with (2.1), (2.34), $v(\cdot) \geq 0$, Lemma 4.7, (2.8), (2.16), and (4.37) gives

(4.38)

$$\begin{aligned} &f(x_k) - f(x_k + \alpha_k \hat{s}_k) \\ &\quad = \phi(x_k;\bar\sigma) - \phi(x_k + \alpha_k\hat{s}_k;\bar\sigma) - \bar\sigma\big(v(x_k) - v(x_k+\alpha_k\hat{s})\big) \geq \gamma_\phi\alpha_k\rho_k^\phi - \bar\sigma v(x_k) \\ &\quad \geq \gamma_\phi\alpha_b \min\left[C_\rho \Delta\ell^\phi(s_k;x_k,\bar\sigma)^2,\ \tfrac12\Delta\ell^\phi(s_k;x_k,\bar\sigma)\right] - \bar\sigma v(x_k) \\ &\quad = \gamma_\phi\alpha_b\min\Big[C_\rho\big(\Delta\ell^f(s_k;x_k)^2 + 2\bar\sigma\Delta\ell^f(s_k;x_k)\Delta\ell^v(s_k;x_k) + \bar\sigma^2\Delta\ell^v(s_k;x_k)^2\big), \\ &\qquad\qquad \tfrac12\big(\Delta\ell^f(s_k;x_k) + \bar\sigma\Delta\ell^v(s_k;x_k)\big)\Big] - \bar\sigma v(x_k) \\ &\quad \geq \gamma_\phi\alpha_b\min\left[C_\rho\Delta\ell^f(s_k;x_k)^2,\ \tfrac12\Delta\ell^f(s_k;x_k)\right] - \bar\sigma v(x_k) \\ &\quad \geq \gamma_\phi\alpha_b\min\Big[C_\rho\big(\epsilon^f\big)^2,\ \tfrac12\epsilon^f\Big] - \bar\sigma v(x_k) \ \text{ for } k\in\mathcal{S}_b \text{ sufficiently large.} \end{aligned}$$

Since $|\mathcal{S}_b| = \infty > |\mathcal{S}_v|$, we may define k^+ as the first iteration greater than k such that $k^+ \in \mathcal{S}_b \cup \mathcal{S}_o$. It then follows from the construction of Algorithm 1 and $|\mathcal{S}_v| < \infty$ that

if $k \in \mathcal{S}_b$ is sufficiently large, then $k^+ \in \mathcal{S}_b \cup \mathcal{S}_o$ and $l \in \mathcal{S}_p$ for all $k < l < k^+$.

Using (2.34), $\alpha_i \geq 0$, (2.20), and (2.28) we conclude that

$$\begin{aligned} \phi(x_{k+1};\bar\sigma) - \phi(x_{k^+};\bar\sigma) &= \sum_{i=k+1}^{k^+-1} \phi(x_i;\bar\sigma) - \phi(x_i + \alpha_i\hat{s}_i;\bar\sigma) \geq \sum_{i=k+1}^{k^+-1}\gamma_\phi\alpha_i\rho_i^\phi \\ &= \sum_{i=k+1}^{k^+-1}\gamma_\phi\alpha_i\min\Big[\,\Delta\ell^\phi(s_i;x_i,\bar\sigma),\ \Delta q^\phi(s_i^{c\phi};x_i,H_i,\bar\sigma)\Big] \\ &\geq 0 \ \text{ for } k\in\mathcal{S}_b \text{ sufficiently large,} \end{aligned}$$

which may be combined with (2.1), $v(\cdot) \geq 0$, and (2.33) to conclude that

$$\begin{aligned} f(x_{k+1}) - f(x_{k^+}) &\geq \bar\sigma\big(v(x_{k^+}) - v(x_{k+1})\big) \\ &\geq -\bar\sigma v(x_{k+1}) > -\bar\sigma v(x_k) \ \text{ for } k\in\mathcal{S}_b \text{ sufficiently large.} \end{aligned} \tag{4.39}$$

It then follows from (4.38) and (4.39) that

$$\begin{aligned} f(x_k) - f(x_{k^+}) &= \big(f(x_k) - f(x_k + \alpha_k\hat{s}_k)\big) + \big(f(x_{k+1}) - f(x_{k^+})\big) \\ &> \gamma_\phi\alpha_b\min\Big[C_\rho\big(\epsilon^f\big)^2,\ \tfrac12\epsilon^f\Big] - 2\bar\sigma v(x_k) \ \text{ for } k\in\mathcal{S}_b \text{ sufficiently large.} \end{aligned} \tag{4.40}$$

Next, since $\lim_{k\in\mathcal{S}_b} v(x_k) = 0$ we know that

$$v(x_k) \leq \frac{1}{4\bar\sigma}\gamma_\phi\alpha_b\min\Big[C_\rho\big(\epsilon^f\big)^2,\ \tfrac12\epsilon^f\Big] \quad \text{for } k\in\mathcal{S}_b \text{ sufficiently large,}$$

which may be combined with (4.40) to deduce that

(4.41)

$$f(x_k) - f(x_{k^+}) > \tfrac12\gamma_\phi\alpha_b\min\Big[C_\rho\big(\epsilon^f\big)^2,\ \tfrac12\epsilon^f\Big] =: \epsilon^v > 0 \ \text{ for } k\in\mathcal{S}_b \text{ sufficiently large.}$$

If we define $\hat{k}^+$ to be the first b-iteration greater than k (thus, $\hat{k}^+ \geq k^+$), it follows from (4.41), the fact that Algorithm 1 does not allow further p-iterations until it has its next b-iteration, and the fact that the objective f is decreased during o-iterations that $f(x_k) - f(x_{\hat{k}^+}) > \epsilon^v$ for $k \in \mathcal{S}_b$ sufficiently large. Since $|\mathcal{S}_b| = \infty$, this implies that $\lim_{k \in \mathcal{S}_b} f(x_k) = -\infty$, which contradicts the fact that f is bounded as a consequence of Assumptions 4.1 and 4.2. This proves the result. ∎

We now show that limit points of $\{x_k\}_{\mathcal{S}_v \cup \mathcal{S}_b}$ are infeasible stationary or KKT point for problem (1.1).

THEOREM 4.18. *Suppose that the Assumptions* 4.1–4.4, (4.11), *and* $|\mathcal{S}_v \cup \mathcal{S}_b| = \infty$ *hold. Then, there exists a limit point* x_* *of* $\{x_k\}_{\mathcal{S}_v \cup \mathcal{S}_b}$ *such that either*

(i) x_* *is a KKT point of problem* (1.1) *or*

(ii) x_* *is an infeasible stationary point.*

Proof. From Assumptions 4.1 and 4.2 we know that there exists a limit point x_* of $\{x_k\}_{\mathcal{S}_v \cup \mathcal{S}_b}$. First, if $v(x_*) > 0$, then it follows from Lemma 4.16 and Lemma 2.1 that x_* is an infeasible stationary point, which is part (ii) of this theorem. Second, if $v(x_*) = 0$ and $|\mathcal{S}_v| = \infty$, then it follows from part (i) of Lemma 4.17 and Lemma 2.2 that x_* is a KKT point of problem (1.1). This is case (i) of this theorem. Finally, if $v(x_*) = 0$ and $|\mathcal{S}_v| < \infty$ (so that $|\mathcal{S}_b| = \infty$), then it follows from part (ii) of Lemma 4.17 and Lemma 2.2 that x_* is a KKT point of problem (1.1), which once again is case (i) of the theorem. ∎

4.2. Convergence analysis under unbounded weighting parameter. We now consider the situation when the weighting parameter increases without bound, i.e., that

$$\lim_{k \to \infty} \sigma_k = \infty. \tag{4.42}$$

Our analysis begins with the following lemma, which is similar to [9, Lemma 3.8].

LEMMA 4.19. *Suppose that Assumptions* 4.1–4.4 *are satisfied,* (4.42) *holds,* x_* *is a limit point of* $\{x_k\}$ *satisfying* $v(x_*) > 0$*, and* $\Delta \ell^v(s_*^s; x_*) > 0$*, where* s_*^s *is the solution to*

$$\underset{(s,r) \in \mathbb{R}^{n+m}}{\text{minimize}} \ e^T r \quad \text{subject to} \quad c(x_*) + J(x_*)s + r \geq 0, \ \ r \geq 0, \ \ \|s\|_\infty \leq \delta,$$

for some $\delta \in [\delta_{min}, \delta_{max}]$*. Then, along any subsequence* $\{x_k\}_{k \in \mathcal{K}}$ *that converges to* x_**, the weighting parameter is updated only a finite number of times.*

Proof. We begin by defining

$$s_k^\phi(\sigma) := \underset{s \in \mathbb{R}^n}{\text{argmin}} \ q^\phi(s; x_k, B_k, \sigma) \tag{4.43}$$

and

$$\mu := \mu(\sigma) := \left(1 - \frac{\eta_\sigma}{\eta_v}\right) \sigma < \sigma, \tag{4.44}$$

where we used the fact that $0 < 1 - \eta_\sigma/\eta_v < 1$ holds since $0 < \eta_\sigma < \eta_v < 1$ is defined in (2.19).

Using the fact that $\Delta q^\phi(s_k^\phi(\mu); x_k, B_k, \mu) \geq 0$, (2.9), and the definition of $\mu = \mu(\sigma)$, we can see that

$$\begin{aligned}
&\Delta q^{\phi}(s_k^{\phi}(\mu); x_k, B_k, \mu) \\
&\quad = \Delta q^{f}(s_k^{\phi}(\mu); x_k, B_k) + \left(1 - \frac{\eta_\sigma}{\eta_v}\right)\sigma \Delta\ell^{v}(s_k^{\phi}(\mu); x_k) \\
&\quad = \Delta q^{\phi}(s_k^{\phi}(\mu); x_k, B_k, \sigma) - \frac{\eta_\sigma}{\eta_v}\sigma \Delta\ell^{v}(s_k^{\phi}(\mu); x_k) \geq 0 \text{ for } \mu = \mu(\sigma) \text{ and all } \sigma > 0,
\end{aligned}$$

which implies that

$$\Delta q^{\phi}(s_k^{\phi}(\mu); x_k, B_k, \sigma) \geq \frac{\eta_\sigma}{\eta_v}\sigma \Delta\ell^{v}(s_k^{\phi}(\mu); x_k) \text{ for } \mu = \mu(\sigma) \text{ and all } \sigma > 0. \tag{4.45}$$

Since $\Delta\ell^v(s_*^s; x_*) > 0$ and $\lim_{k\in\mathcal{K}} x_k = x_*$ by assumption, it follows from [17, Theorem 3.2.8] that there exists $\epsilon \in (0, 1)$ and k' such that

$$\Delta\ell^{v}(s_k^{s}; x_k) > \epsilon \text{ for all } k' \leq k \in \mathcal{K}. \tag{4.46}$$

Moreover, since the Newton step $-B_k^{-1} g_k$ minimizes $q^f(s; x_k, B_k)$, it follows from Assumption 4.3 that

$$\begin{aligned}
q^{f}(s_k^{\phi}(\sigma); x_k, B_k) &\geq q^{f}(-B_k^{-1} g_k; x_k, B_k) = f_k - \tfrac{1}{2} g_k{}^T B_k^{-1} g_k \\
&\geq f_k - \frac{\|g_k\|_2^2}{2\lambda_{\min}} \text{ for all } \sigma > 0.
\end{aligned} \tag{4.47}$$

Next, it follows from (2.10), the choice $\delta_k \in [\delta_{\min}, \delta_{\max}]$, norm inequalities, and Assumption 4.3 that

$$q^{f}(s_k^{s}; x_k, B_k) \leq f_k + \|g_k\|_2 \, \delta_{\max} + \tfrac{1}{2}\lambda_{\max}\delta_{\max}^2. \tag{4.48}$$

Then, (4.47), (4.48), and Assumptions 4.1 and 4.2 imply the existence of a constant $C_{\mathrm{qf}} > 0$ such that

$$q^{f}(s_k^{s}; x_k, B_k) - q^{f}(s_k^{\phi}(\sigma); x_k, B_k) \leq C_{\mathrm{qf}} \text{ for all } \sigma > 0. \tag{4.49}$$

We now define

$$\sigma_{\mathrm{crit}} := \frac{C_{\mathrm{qf}}}{\epsilon(1-\eta_v)\left(1 - \frac{\eta_\sigma}{\eta_v}\right)} > \mu\left(\sigma_{\mathrm{crit}}\right) = \frac{C_{\mathrm{qf}}}{\epsilon(1-\eta_v)} > 0, \tag{4.50}$$

and the associated infinite subsequence

$$\mathcal{S}' = \{k \in \mathcal{K} : k \geq k' \text{ and } \sigma_k \geq \sigma_{\mathrm{crit}}\}. \tag{4.51}$$

It follows from the fact that $\Delta q^{\phi}(s_k^{\phi}(\sigma); x_k, B_k, \sigma) \geq \Delta q^{\phi}(s_k^{s}; x_k, B_k, \sigma)$ (by the definition of $s_k^{\phi}(\sigma)$), (2.9), (4.49), (4.46), and (4.50) that

$$\begin{aligned}
\Delta\ell^{v}(s_k^{\phi}(\sigma); x_k) &\geq \Delta\ell^{v}(s_k^{s}; x_k) - \frac{1}{\sigma}\left(q^{f}(s_k^{s}; x_k, B_k) - q^{f}(s_k^{\phi}(\sigma); x_k, B_k)\right) \\
&\geq \Delta\ell^{v}(s_k^{s}; x_k) - \frac{1}{\sigma} C_{\mathrm{qf}} = \Delta\ell^{v}(s_k^{s}; x_k)\left(1 - \frac{C_{\mathrm{qf}}}{\sigma \Delta\ell^{v}(s_k^{s}; x_k)}\right) \\
&\geq \eta_v \Delta\ell^{v}(s_k^{s}; x_k) \text{ for } \sigma \geq \mu(\sigma_{\mathrm{crit}}) \text{ and } k' \leq k \in \mathcal{K}.
\end{aligned} \tag{4.52}$$

We may now use the definition of $\mathcal{S}'$, (4.52), the fact that $s_k^\phi(\sigma_k) \equiv s_k^p$, (2.15), and (2.16) to show that

$$\tau_k = 1, \ \ s_k = s_k^p, \ \text{ and } \ \Delta q^\phi(s_k; x_k, B_k, \sigma_{k+1}) \geq \eta_\phi \Delta q^\phi(s_k^p; x_k, B_k, \sigma_{k+1}) \ \text{ for } k \in \mathcal{S}' \tag{4.53}$$

since $\eta_\phi \in (0,1)$ in Algorithm 1. Next, it follows from (4.53), (2.9), $B_k \succ 0$, the fact that $s_k^p \equiv s_k^\phi(\sigma_k)$ and s_k^p minimizes $q^\phi(s; x_k, B_k, \sigma_k)$, (4.45), the fact that $\mu(\sigma_k) \geq \mu(\sigma_{\text{crit}})$ for $k \in \mathcal{S}'$, and (4.52) that

$$\begin{aligned} \Delta \ell^\phi(s_k; x_k, \sigma_k) &\geq \Delta q^\phi(s_k^p; x_k, B_k, \sigma_k) \geq \Delta q^\phi(s_k^\phi(\mu(\sigma_k)); x_k, B_k, \sigma_k) \\ &\geq \frac{\eta_\sigma}{\eta_v} \sigma_k \Delta \ell^v(s_k^\phi(\mu(\sigma_k)); x_k) \\ &\geq \frac{\eta_\sigma}{\eta_v} \sigma_k \big(\eta_v \Delta \ell^v(s_k^s; x_k)\big) = \sigma_k \eta_\sigma \Delta \ell^v(s_k^s; x_k) \ \text{ for } k \in \mathcal{S}'. \end{aligned} \tag{4.54}$$

We now conclude from (4.53), (4.54), (2.19), and the fact that the weighting parameter is only increased in lines 12 and 34 of Algorithm 1 that σ_k is increased a finite number of times on $\mathcal{K}$. ◻

We now consider feasible limit points at which the MFCQ [35] holds.

LEMMA 4.20. *Suppose that Assumptions* 4.1–4.4 *are satisfied,* (4.42) *holds, x_* is a limit point of $\{x_k\}$ at which $v(x_*) = 0$, and the MFCQ holds. Then, the following hold for all x_k sufficiently close to x_* and σ_k sufficiently large:* (i) $\Delta \ell^v(s_k^p; x_k) = v(x_k)$*;* (ii) $s_k = s_k^p$*; and* (iii) *σ_k is not increased during iteration k.*

Proof. We may use [9, Lemmas 3.12 and 3.13] since the proofs only used the properties of the MFCQ, the continuity of the problem functions f and g, and the convexity of their penalty and steering subproblems. Their subproblem [9, equations 2.7(a–d)] is equivalent to our predictor subproblem (2.12) and both methods minimize the same quadratic model of the penalty function. A small difference is that our predictor subproblem is designed so that if $\ell^v(s_k^s; x_k) = 0$, then $\ell^v(s_k^p; x_k) = 0$ as well; they satisfy this requirement by increasing their penalty parameter in Step 4a of [9, equation 2.11] and re-solving for a new step. Their steering subproblem [9, equations 2.9(a–e)] is equivalent to (2.10).

The assumptions of this lemma and [9, Lemma 3.12] imply the existence of $r > 0$ and $k' \geq 0$ so that

$$\ell^v(s_k^p, x_k) = v(x_k) \ \text{ for all } k \in \mathcal{S}', \tag{4.55}$$

where $\mathcal{S}' := \{k : \|x_k - x_*\| \leq r \ \text{ and } \ k \geq k'\}$, which proves part (i). The inequality $\Delta \ell^v(s_k^s; x_k) \geq 0$, (4.55), and the definition of $\Delta \ell^v$ imply

$$\Delta \ell^v(s_k^p; x_k) \geq v(x_k) - \ell^v(s_k^s; x_k) = \Delta \ell^v(s_k^s; x_k) \geq \eta_v \Delta \ell^v(s_k^s; x_k) \ \text{ for } k \in \mathcal{S}',$$

where $\eta_v \in (0,1)$ is defined in (2.16). Thus, we conclude from (2.16) that $\tau_k = 1$ and $s_k = s_k^p$ for $k \in \mathcal{S}'$, which proves part (ii). Finally, it follows from [9, Lemma 3.13] and the assumptions of this lemma that

$$\Delta q^\phi(s_k^p; x_k, B_k, \sigma_k) \geq \sigma_k \eta_\sigma v(x_k) \geq \sigma_k \eta_\sigma \Delta \ell^v(s_k^s; x_k) \ \text{ for } k \in \mathcal{S}', \tag{4.56}$$

where the last inequality follows from the definition of $\Delta \ell^v$. It then follows from part (ii) of this lemma, (2.9), $B_k \succ 0$, and (4.56) that

$$\Delta \ell^\phi(s_k; x_k, \sigma_k) = \Delta \ell^\phi(s_k^p; x_k, \sigma_k) \geq \Delta q^\phi(s_k^p; x_k, B_k, \sigma_k) \geq \sigma_k \eta_\sigma \Delta \ell^v(s_k^s; x_k) \ \text{ for } k \in \mathcal{S}'.$$

We may conclude from this inequality, (2.19), and the fact that σ_k will not be increased on line 34 as a result of part (ii) of this lemma that $\sigma_{k+1} = \sigma_k$ for $k \in \mathcal{S}'$, which proves part (iii). ∎

THEOREM 4.21. *If Assumptions* 4.1–4.4 *and* (4.42) *hold, there is a limit point* x_* *such that either*

(i) x_* *is an infeasible stationary point or*

(ii) x_* *is feasible, but the MFCQ does not hold.*

Proof. Let $\mathcal{D}$ to be the infinite index set consisting of the iterations for which the weighting parameter is increased. Then, let x_* be a limit point of $\{x_k\}_{k\in\mathcal{D}}$, which must exist as a consequence of Assumptions 4.1 and 4.2. First, suppose that $v(x_*) > 0$. It then follows from Lemma 4.19 that if $\Delta\ell^v(s_*^s; x_*) > 0$ (s_*^s is defined in Lemma 4.19), then the weighting parameter is updated only a finite number of times along $\mathcal{D}$, which is a contradiction. Therefore, we deduce that $\Delta\ell^v(s_*^s; x_*) = 0$ and consequently that x_* is an infeasible stationary point. Second, suppose that $v(x_*) = 0$. It then follows from Lemma 4.20 that if the MFCQ holds at x_*, then σ_k will be increased only a finite number of times along $\mathcal{D}$. This is a contradiction and, therefore, the MFCQ does not hold at x_*. ∎

5. Conclusions. In this paper, we presented a new filter line search method that replaced the traditional restoration phase with a penalty mode that systematically decreased an exact penalty function. Importantly, we solved a *single* strictly convex quadratic program subproblem during each iteration that was *always* feasible. Each search direction was defined as a convex combination of a steering step (a solution of a linear program) that represented the best local improvement in constraint violation and a predictor step that reduced our strictly convex quadratic model of the exact penalty function. We also allowed for the computation of an accelerator step defined as a solution to a simple equality constrained quadratic program (plus trust-region constraint) to promote fast local convergence. In this manner, the trial step always incorporated information from both the objective function and constraint violation. To further promote step acceptance, we utilized second-order information in the computation of Cauchy steps that provided realistic measurements of the decrease one might expect from the nonlinear problem functions. By using local feasibility estimates that emerged during the steering process, we defined a new and improved margin (envelope) of the filter. This new definition encouraged the acceptance of steps that made reasonable progress but might be considered inadmissible by a traditional filter. Under standard assumptions, we proved global convergence of our algorithm.

The fact that every subproblem of our method is feasible has an interesting (favorable) consequence when compared to previous SQP filter methods. Those methods trigger a restoration phase in multiple situations, the most common being when the traditional SQP subproblem is infeasible. In this case, the primary role of the restoration phase is to obtain a new feasible subproblem. This undesirable situation is not encountered in our method since all subproblems are feasible. Our method still may enter a penalty phase, but only when overwhelming evidence indicates that previously added filter entries are blocking progress. We believe this feature of our method is far more attractive and practical in comparison to previous filter methods.

Local convergence issues have not been considered in this paper and are currently under investigation. It is evident that our method—like filter methods and SQP methods based on exact penalty functions—may experience the Maratos effect [36], i.e., reject the unit step (the traditional SQP step) when the current iterate is arbitrarily close to a minimizer. We remain optimistic, however, that superlinear convergence may be established if we make common second-order optimality assumptions on the minimizer and include either a second-order correction strategy or a nonmonotone approach [17, section 15.3.2.3].

Simple and straightforward modifications to our method allow for the solution of problems defined by a mixture of inequality and equality constraints. For instance, the definition of the constraint violation would be augmented to represent an ℓ_1 measure of infeasibility for both inequality and equality constraints. Each key subproblem must also be modified. For example, (2.13) would additionally include the linearized equality constraints augmented by a pair of nonnegative elastic variables. Otherwise, the algorithm remains unchanged.

We are currently implementing our algorithm. Once a robust and well-tested code is obtained, we will investigate many interesting practical questions related to our method. These questions include (i) computing the frequency with which our method enters a penalty mode relative to a traditional restoration phase; (ii) evaluating the benefits and possible disadvantages of our penalty mode versus a traditional restoration phase on test problems that trigger such phases; (iii) studying the stability and quality of the iterations typically generated by a traditional restoration phase and our penalty mode; and (iv) investigating the practical benefits of defining our filter margin adaptively based on local estimates. However, such tests should only be performed after a polished implementation is obtained so that reliable conclusions may be obtained.

Acknowledgment. The authors are grateful to the referees for constructive comments that significantly improved the presentation.

REFERENCES

[1] G. Alarcón, C. Torres, and L. Gómez, *Global optimization of gas allocation to a group of wells in artificial lift using nonlinear constrained programming*, J. Energy Resources Technology, 124 (2002), p. 262.

[2] A. Barclay, P. E. Gill, and J. B. Rosen, *SQP methods in optimal control*, in 12th Conference on Variational Calculus, Optimal Control and Applications, R. Bulirsch, L. Bittner, W. Schmidt, and K. Heier, eds., Internat. Ser. Numer. Math. 124, Birkhäuser, Basel, 1998, pp. 207–222.

[3] A. Barclay, P. E. Gill, and J. B. Rosen, *SQP methods in optimal control*, in Variational Calculus, Optimal Control and Applications, R. Bulirsch, L. Bittner, W. H. Schmidt, and K. Heier, eds., Internat. Ser. Numer. Math. 124, Birkhäuser, Basel, 1998, pp. 207–222.

[4] J. Barzilai and J. M. Borwein, *Two-point step size gradient methods*, IMA J. Numer. Anal., 8 (1988), pp. 141–148.

[5] M. Bazaraa, H. Sherali, and C. Shetty, *Nonlinear Programming: Theory and Algorithms*, John Wiley and Sons, New York, 2006.

[6] A. Bryson and Y. Ho, *Applied Optimal Control*, Taylor and Francis, New York, 1975.

[7] J. V. Burke, F. E. Curtis, and H. Wang, *A Sequential Quadratic Optimization Algorithm with Fast Infeasibility Detection*, Tech. report 12T-012, COR@L Laboratory, Department of ISE, Lehigh University, 2012.

[8] 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 (2006), pp. 471–489.

[9] R. H. Byrd, G. Lopez-Calva, and J. Nocedal, *A line search exact penalty method using steering rules*, Math. Program., 133 (2012), pp. 39–73.

[10] R. H. Byrd, J. Nocedal, and R. A. Waltz, *KNITRO*, http://www.ziena.com/knitro/kindex.htm.

[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] L. Chen and D. Goldfarb, *An interior-point piecewise linear penalty method for nonlinear programming*, Math. Program., 128 (2011), pp. 73–122.

[13] C. M. Chin and R. Fletcher, *On the global convergence of an SLP-filter algorithm that takes EQP steps*, Math. Program., 96 (2003), pp. 161–177.

[14] C. M. Chin, A. H. A. Rashid, and K. M. Nor, *A combined filter line search and trust region method for nonlinear programming*, WSEAS Trans. Math., 5 (2006), pp. 656–662.

[15] C. M. Chin, A. H. A. Rashid, and K. M. Nor, *Global and local convergence of a filter line search method for nonlinear programming*, Optim. Methods Softw., 22 (2007), pp. 365–390.

[16] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *LANCELOT: A Fortran package for large-scale nonlinear optimization (Release A)*, Lecture Notes in Comput. Math. 17, Springer-Verlag, New York, 1992.

[17] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *Trust-Region Methods*, MOS-SIAM Ser. Optim., SIAM, Philadelphia, 2000.

[18] F. E. Curtis, T. Johnson, D. P. Robinson, and A. Wachter, *An Inexact Sequential Quadratic Optimization Algorithm for Large-Scale Optimization*, Technical report TR-1-13, Johns Hopkins University, Baltimore, MD, 2013.

[19] F. E. Curtis and J. Nocedal, *Flexible penalty functions for nonlinear constrained optimization*, IMA J. Numer. Anal., 28 (2008), pp. 749–769.

[20] S. Flåm and A. Antipin, *Equilibrium programming using proximal-like algorithms*, Math. Program., 78 (1996), pp. 29–41.

[21] R. Fletcher, N. I. M. Gould, S. Leyffer, Ph. L. Toint, and A. Wächter, *Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming*, SIAM J. Optim., 13 (2002), pp. 635–659.

[22] R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function*, Math. Program., 91 (2002), pp. 239–269.

[23] R. Fletcher, S. Leyffer, and Ph. L. Toint, *On the Global Convergence of an SLP-Filter Algorithm*, Report 98/13, Department of Mathematics, University of Namur, Belgium, 1998.

[24] R. Fletcher, S. Leyffer, and Ph. L. Toint, *On the global convergence of a filter-SQP algorithm*, SIAM J. Optim., 13 (2002), pp. 44–59.

[25] P. E. Gill, L. O. Jay, M. W. Leonard, L. R. Petzold, and V. Sharma, *An SQP method for the optimal control of large-scale dynamical systems*, J. Comput. Appl. Math., 120 (2000), pp. 197–213.

[26] P. E. Gill and W. Murray, *Newton-type methods for unconstrained and linearly constrained optimization*, Math. Program., 7 (1974), pp. 311–350.

[27] P. E. Gill, W. Murray, and M. A. Saunders, *SNOPT: An SQP algorithm for large-scale constrained optimization*, SIAM Rev., 47 (2005), pp. 99–131.

[28] N. I. M. Gould and D. P. Robinson, *A second derivative SQP method: Global convergence*, SIAM J. Optim., 20 (2010), pp. 2023–2048.

[29] 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.

[30] N. I. M. Gould and D. P. Robinson, *A second derivative SQP method with a "trust-region-free" predictor step*, IMA J. Numer. Anal., 32 (2012), pp. 580–601.

[31] T. Johansen, T. Fossen, and S. Berge, *Constrained nonlinear control allocation with singularity avoidance using sequential quadratic programming*, IEEE Trans. Control Systems Technology, 12 (2004), pp. 211–216.

[32] D. Kraft, *On converting optimal control problems into nonlinear programming problems*, in Computational Mathematical Programming, K. Schittkowski, ed., NATO ASI Ser. F Comput. Systems Sci. 15, Springer-Verlag, New York, 1985, pp. 261–280.

[33] D. C. Liu and J. Nocedal, *On the limited memory BFGS method for large scale optimization*, Math. Program., 45 (1989), pp. 503–528.

[34] P. Liu and A. Der Kiureghian, *Optimization algorithms for structural reliability*, Struct. Safety, 9 (1991), pp. 161–177.

[35] O. L. Mangasarian and S. Fromovitz, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl., 17 (1967), pp. 37–47.

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

[37] J. Morales, J. Nocedal, and Y. Wu, *A sequential quadratic programming algorithm with an additional equality constrained phase*, IMA J. Numer. Anal., 32 (2011), pp. 553–579.

[38] F. Murphy, H. Sherali, and A. Soyster, *A mathematical programming approach for determining oligopolistic market equilibrium*, Math. Program., 24 (1982), pp. 92–106.

[39] B. A. Murtagh and M. A. Saunders, *MINOS* 5.5 *User's Guide*, Report SOL 83-20R, Department of Operations Research, Stanford University, Stanford, CA, 1998.

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

[41] S. M. Robinson, *Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms*, Math. Program., 7 (1974), pp. 1–16.

[42] R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[43] R. B. Schnabel and E. Eskow, *A new modified Cholesky factorization*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 1136–1158.

[44] M. Ulbrich, S. Ulbrich, and L. Vicente, *A globally convergent primal-dual interior-point filter method for nonlinear programming*, Math. Program., 100 (2004), pp. 379–410.

[45] A. Wächter and L. T. Biegler, *Line search filter methods for nonlinear programming: Motivation and global convergence*, SIAM J. Optim., 16 (2006), pp. 1–31.

[46] A. Wächter, L. T. Biegler, Y.-D. Lang, and A. Raghunathan, *IPOPT: An Interior Point Algorithm for Large-Scale Nonlinear Optimization*, http://www.coin-or.org (2002).

[47] R. Waltz, J. Morales, J. Nocedal, and D. Orban, *An interior algorithm for nonlinear optimization that combines line search and trust region steps*, Math. Program., 107 (2006), pp. 391–408.