Line Search Filter Methods for Nonlinear Programming: Motivation and Global Convergence

Andreas Wächter* and Lorenz T. Biegler†

September 16, 2004

Abstract

Line search methods are proposed for nonlinear programming using Fletcher and Leyffer's filter method, which replaces the traditional merit function. Their global convergence properties are analyzed. The presented framework is applied to active set SQP and barrier interior point algorithms. Under mild assumptions it is shown that every limit point of the sequence of iterates generated by the algorithm is feasible, and that there exists at least one limit point that is a stationary point for the problem under consideration. A new alternative filter approach employing the Lagrangian function instead of the objective function with identical global convergence properties is briefly discussed.

Keywords: nonlinear programming – nonconvex constrained optimization – filter method – line search – SQP – interior point – barrier method – global convergence

1 Introduction

Recently, Fletcher and Leyffer [9] proposed filter methods, offering an alternative to merit functions, as a tool to guarantee global convergence in algorithms for nonlinear programming (NLP). The underlying concept is that trial points are accepted if they improve the objective function *or* improve the constraint violation instead of a combination of those two measures defined by a merit function. The practical results reported for the filter trust region sequential quadratic programming (SQP) method in [9] are encouraging, and subsequently global convergence results for related algorithms were established by Fletcher et al. [7, 10]. Other researchers also proposed global convergence results for different trust region based filter methods, such as for an interior point approach (Ulbrich et al. [21]), a bundle method for non-smooth optimization (Fletcher and Leyffer [8]), and a pattern search algorithm for derivative-free optimization (Audet and Dennis [1]).

In this paper we propose and analyze a filter method framework based on line search which can be applied to active set SQP methods as well as barrier interior point methods. The motivation given by Fletcher and Leyffer [9] for the development of the filter method is to avoid the necessity to determine a suitable value of the penalty parameter in the merit function. In addition, in the context of a line search method, the filter approach offers another important advantage regarding robustness. It has been known for some time that line search methods can converge to "spurious solutions," infeasible points that are not even critical points for a measure of infeasibility, if the gradients of the constraints become linearly dependent at non-feasible points. In [19], Powell gives an example for this behavior. More recently, Wächter and Biegler [25] demonstrated another global

*IBM T.J. Watson Research Center, Yorktown Heights, NY; E-mail: andreasw@watson.ibm.com
†Carnegie Mellon University, Pittsburgh, PA; E-mail: lb01@andrew.cmu.edu

convergence problem for many line search interior point methods on a simple well-posed example. Here, the affected methods generate search directions that point outside of the region $\mathcal{I}$ defined by the inequality constraints because they are forced to satisfy the linearization of the equality constraints. Consequently, an increasingly smaller fraction of the proposed step can be taken, and the iterates eventually converge to an infeasible point at the boundary of $\mathcal{I}$, which once again is not even a stationary point for any norm of the constraint violation (see also Marazzi and Nocedal [15] for a detailed discussion of "feasibility control"). Using a filter approach within a line search algorithm helps to overcome these problems. If the trial step size becomes too small in order to guarantee sufficient progress toward a solution of the problem, the proposed filter method reverts to a feasibility restoration phase, whose goal is to deliver a new acceptable iterate by decreasing the constraint violation, or to converge to a local minimizer of infeasibility if this is not possible. In this way, the filter line search procedure detects problematic cases automatically, so that global convergence problems described above cannot occur if a suitable algorithm for the restoration phase is used.

This paper is organized as follows. For easy comprehension of the derivation and analysis of the proposed line search filter method, the main part of the paper considers the particular case of solving nonlinear optimization problems without inequality constraints. At the end of the paper it is shown how the presented techniques can be applied to general NLPs using active set SQP methods and a barrier approach.

In Section 2 we motivate and state the algorithm for the solution of the equality constrained problem. The method is motivated by the trust region SQP method proposed by Fletcher et al. [7]. An important difference, however, lies in the condition that determines when to switch between certain sufficient decrease criteria. The proposed rule is more general and allows us to show fast local convergence of the proposed line search filter method in the companion paper [26]. We then show in Section 3 that every limit point of the sequence of iterates generated by the algorithm is feasible, and that there is at least one limit point that satisfies the first order optimality conditions for the problem.

In Section 4.1 we propose an alternative measure for the filter acceptance criteria. Here, a trial point is accepted if it reduces the infeasibility or the value of the Lagrangian function (instead of the objective function). The global convergence results still hold for this modification. Having presented the line search filter framework on the simple case of problems with equality constraints only, we show in Section 4.2 how it can be applied to SQP methods handling inequality constraints, preserving the same global convergence properties. Finally, Section 4.3 shows how the presented line search filter method can be applied in a barrier interior point framework.

1.1 Notation

We denote the i-th component of a vector $v \in \mathbb{R}^n$ by $v^{(i)}$, and the i-th unit coordinate vector is called e_i in the text. Norms $\|\cdot\|$ denote a fixed vector norm and its compatible matrix norm unless otherwise noted. For brevity, we use the convention $(x, \lambda) = (x^T, \lambda^T)^T$ for vectors x, λ. For a matrix A, we denote by $\sigma_{\min}(A)$ the smallest singular value of A, and for a symmetric, positive definite matrix A we call the smallest eigenvalue $\lambda_{\min}(A)$. Given two vectors $v, w \in \mathbb{R}^n$, we define the convex segment $[v, w] := \{v + t(w - v) : t \in [0, 1]\}$. Finally, we denote by $O(t_k)$ a sequence $\{v_k\}$ satisfying $\|v_k\| \leq \beta\ t_k$ for some constant $\beta > 0$ independent of k, and by $o(t_k)$ a sequence $\{v_k\}$ satisfying $\|v_k\| \leq \beta_k t_k$ for some positive sequence $\{\beta_k\}$ with $\lim_k \beta_k = 0$.

2 A Line Search Filter Approach

For simplicity, we first describe and analyze the line search filter method for NLPs with equality constraints only, i.e. we assume that the problem to be solved is stated as

$$\min_{x\in\mathbb{R}^n} \quad f(x) \tag{1a}$$

$$\text{subject to} \quad c(x) = 0 \tag{1b}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the equality constraints $c : \mathbb{R}^n \to \mathbb{R}^m$ with $m < n$ are sufficiently smooth. We show later, how this approach can be used in an active set SQP (Section 4.2) and an interior point (Section 4.3) framework in order to tackle general NLPs.

The Karush-Kuhn-Tucker (KKT) conditions for the NLP (1) are

$$g(x) + A(x)\lambda = 0 \tag{2a}$$

$$c(x) = 0, \tag{2b}$$

where we denote with $A(x) := \nabla c(x)$ the transpose of the Jacobian of the constraints c, and with $g(x) := \nabla f(x)$ the gradient of the objective function. The vector λ corresponds to the Lagrange multipliers for the equality constraints (1b). Under certain constraint qualifications, such as linear independence of the constraint gradients, the KKT conditions are the first order optimality conditions for (1) (see e.g. [17]).

Given an initial estimate x_0, the line search algorithm proposed in this section generates a sequence of improved estimates x_k of the solution for the NLP (1). For this purpose in each iteration k a search direction d_k is computed from the linearization at x_k of the KKT conditions (2),

$$\begin{bmatrix} H_k & A_k \\ A_k^T & 0 \end{bmatrix} \begin{pmatrix} d_k \\ \lambda_k^+ \end{pmatrix} = -\begin{pmatrix} g_k \\ c_k \end{pmatrix}. \tag{3}$$

Here, $A_k := A(x_k)$, $g_k := g(x_k)$, and $c_k := c(x_k)$. The symmetric matrix H_k denotes the Hessian $\nabla^2_{xx}\mathcal{L}(x_k, \lambda_k)$ of the Lagrangian

$$\mathcal{L}(x, \lambda) := f(x) + c(x)^T\lambda \tag{4}$$

of the NLP (1), or an approximation to this Hessian. The vector λ_k is some estimate of the optimal multipliers corresponding to the equality constraints (1b), and λ_k^+ in (3) can be used to determine a new estimate λ_{k+1} for the next iteration. As is common for most line search methods, we assume that the projection of the Hessian approximation H_k onto the null space of the constraint Jacobian is uniformly positive definite.

After a search direction d_k has been computed, a step size $\alpha_k \in (0, 1]$ is determined in order to obtain the next iterate

$$x_{k+1} := x_k + \alpha_k d_k. \tag{5}$$

We want to guarantee that ideally the sequence $\{x_k\}$ of iterates converges to a solution of the NLP (1). In this paper we consider a backtracking line search procedure, where a decreasing sequence of step sizes $\alpha_{k,l} \in (0, 1]$ $(l = 0, 1, 2, \ldots)$ is tried until some acceptance criterion is satisfied. Traditionally, a trial step size $\alpha_{k,l}$ is accepted if the corresponding trial point

$$x_k(\alpha_{k,l}) := x_k + \alpha_{k,l} d_k \tag{6}$$

provides sufficient reduction of a *merit function*, such as the exact penalty function [14]

$$\phi_\rho(x) = f(x) + \rho\,\theta(x) \tag{7}$$

where we define the infeasibility measure $\theta(x)$ by

$$\theta(x) = \|c(x)\| .$$

Under certain regularity assumptions it can be shown that a feasible strict local minimum of the exact penalty function coincides with a local solution of the NLP (1) if the value of the *penalty parameter* $\rho > 0$ is chosen sufficiently large [14].

In order to avoid the determination of an appropriate value of the penalty parameter ρ, Fletcher and Leyffer [9] propose the concept of a *filter method* in the context of a trust region SQP algorithm. In the remainder of this section we describe how this concept can be applied to the line search framework outlined above.

The underlying idea is to interpret the NLP (1) as a bi-objective optimization problem with two goals: minimizing the constraint violation $\theta(x)$ and minimizing the objective function $f(x)$. A certain emphasis is placed on the first measure, since a point has to be feasible in order to be an optimal solution of the NLP. Here, we do not require that a trial point $x_k(\alpha_{k,l})$ provides progress in a merit function such as (7), which combines these two goals as a linear combination into one single measure. Instead, following Fletcher and Leyffer's original idea, the trial point $x_k(\alpha_{k,l})$ is accepted if it improves feasibility, i.e. if $\theta(x_k(\alpha_{k,l})) < \theta(x_k)$, *or* if it improves the objective function, i.e. if $f(x_k(\alpha_{k,l})) < f(x_k)$. Note, that this criterion is less demanding than the enforcement of decrease in the penalty function (7) and might in general allow larger steps.

Of course, this simple concept is not sufficient to guarantee global convergence. Several precautions have to be added as we outline in the following; these are closely related to those proposed in [7]. The overall line search filter algorithm is formally stated in Section 2.4.

2.1 Sufficient Reduction

Line search methods that use a merit function ensure *sufficient* progress toward the solution. For example, they may do so by enforcing an Armijo condition for the exact penalty function (7) (see e.g. [17]). Here, we borrow the idea from [7, 10] and replace this condition by requiring that the next iterate provides at least as much progress in one of the measures θ or f that corresponds to a small fraction of the current constraint violation, $\theta(x_k)$. More precisely, for fixed constants $\gamma_\theta, \gamma_f \in (0, 1)$, we say that a trial step size $\alpha_{k,l}$ provides sufficient reduction with respect to the current iterate x_k, if

$$\begin{aligned} & \theta(x_k(\alpha_{k,l})) && \leq && (1-\gamma_\theta)\theta(x_k) && \text{(8a)} \\ \text{or} \quad & f(x_k(\alpha_{k,l})) && \leq && f(x_k) - \gamma_f \theta(x_k). && \text{(8b)} \end{aligned}$$

In a practical implementation, the constants γ_θ, γ_f typically are chosen to be small. However, relying solely on this criterion would allow the acceptance of a sequence $\{x_k\}$ that always provides *sufficient reduction* of the constraint violation (8a) alone, and not the objective function. This could result in convergence to a feasible, but non-optimal point. In order to prevent this, we change to a different sufficient reduction criterion whenever for the current trial step size $\alpha_{k,l}$ the *f-type switching condition*

$$m_k(\alpha_{k,l}) < 0 \qquad \text{and} \qquad [-m_k(\alpha_{k,l})]^{s_f}[\alpha_{k,l}]^{1-s_f} > \delta\,[\theta(x_k)]^{s_\theta} \tag{9}$$

holds with fixed constants $\delta > 0, s_\theta > 1, s_f \geq 1$, where

$$m_k(\alpha) := \alpha g_k^T d_k \tag{10}$$

is the linear model of the objective function f in the direction d_k. We choose to formulate the f-type switching condition (9) in terms of a general model $m_k(\alpha)$ as it allows us later, in Section 4.1, to define the algorithm for an alternative measure that replaces "$f(x)$".

If the condition (9) holds, the step d_k is a descent direction for the objective function. Then, instead of insisting on (8), we require that $\alpha_{k,l}$ satisfies the Armijo-type condition

$$f(x_k(\alpha_{k,l})) \leq f(x_k) + \eta_f m_k(\alpha_{k,l}). \tag{11}$$

Here, $\eta_f \in (0, \frac{1}{2})$ is a fixed constant. It is possible that for several trial step sizes $\alpha_{k,l}$ with $l = 1, \ldots, \tilde{l}$ condition (9), but not (11) is satisfied. In this case we note that for smaller step sizes the f-type switching condition (9) may no longer be valid, so that the method reverts to the acceptance criterion (8).

The second part of the switching condition (9) deserves some discussion. It ensures that the progress for the objective function enforced by the Armijo-condition (11) is sufficiently large compared to the current constraint violation. In this way, the decrease in the objective function from (11) cannot be arbitrarily small at points remote from the feasible region. Note that if we choose $s_f = 1$, condition (9) simplifies to "$-m_k(\alpha_{k,l}) > \delta[\theta(x_k)]^{s_\theta}$" and relates the progress predicted by the linear model of f for the step size $\alpha_{k,l}$ to a power of the constraint violation. This is identical to the condition used in filter trust region methods proposed in [7], except that a quadratic model is used there. However, the analysis presented below allows for larger and maybe less intuitive values of s_f. In particular, we might choose $s_f > 2s_\theta$, as required for the local convergence analysis in the companion paper [26]. This choice of s_f makes it possible to show that, close to a local solution, the condition (9) only holds true, if a full step, possibly improved by a second order correction step, satisfies (11) and is accepted.

In accordance with previous publications on filter methods (e.g. [7, 10]), we call $\alpha_{k,l}$ an "f-step-size," if it satisfies the f-type switching condition (9), indicating that then decrease of the objective function is required. Similarly, if an f-step-size $\alpha_{k,l}$ is accepted as the final step size α_k in iteration k, we refer to k as an "f-type iteration."

2.2 Filter as Taboo-Region

Beside requiring sufficient decrease with respect to the current iterate, the filter line search algorithm also needs to avoid cycling. For example, cycling may occur between two points that alternatingly improve one of the measures, θ and f, and worsen the other one. For this purpose, Fletcher and Leyffer [9] define a "taboo region" in the half-plane $\{(\theta, f) \in \mathbb{R}^2 : \theta \geq 0\}$. They maintain a list of $(\theta(x_p), f(x_p))$-pairs (called *filter*) corresponding to (some of) the previous iterates x_p and require that a point, in order to be accepted, has to improve at least one of the two measures compared to those previous iterates. In other words, a trial step $x_k(\alpha_{k,l})$ can only be accepted, if

$$\begin{array}{rrcl} & \theta(x_k(\alpha_{k,l})) & < & \theta(x_p) \\ \text{or} & f(x_k(\alpha_{k,l})) & < & f(x_p) \end{array}$$

for all $(\theta(x_p), f(x_p))$ in the current filter.

In contrast to the notation in [7, 9], for the sake of a simplified notation we define the filter in this paper not as a list but as a *set* $\mathcal{F}_k \subseteq [0, \infty) \times \mathbb{R}$ containing *all* (θ, f)-pairs that are "prohibited" in iteration k. We say, that a trial point $x_k(\alpha_{k,l})$ is *acceptable to the filter* if its (θ, f)-pair does not lie in the taboo-region, i.e. if

$$\Big(\theta(x_k(\alpha_{k,l})), f(x_k(\alpha_{k,l}))\Big) \notin \mathcal{F}_k. \tag{12}$$

During the optimization we make sure that the current iterate x_k is always acceptable to the current filter $\mathcal{F}_k$.

At the beginning of the optimization, the filter is initialized to be empty, $\mathcal{F}_0 := \emptyset$, or — if one wants to impose an explicit upper bound on the constraint violation — as $\mathcal{F}_0 := \{(\theta, f) \in \mathbb{R}^2 : \theta \geq \theta_{\max}\}$ for some $\theta_{\max} > \theta(x_0)$. Throughout the optimization the filter is then augmented in some iterations after the new iterate x_{k+1} has been accepted. For this, the updating formula

$$\mathcal{F}_{k+1} := \mathcal{F}_k \cup \Big\{(\theta, f) \in \mathbb{R}^2 : \theta \geq (1-\gamma_\theta)\theta(x_k) \quad \text{and} \quad f \geq f(x_k) - \gamma_f \theta(x_k)\Big\} \tag{13}$$

is used (see also [7]). If the filter is not augmented, it remains unchanged, i.e. $\mathcal{F}_{k+1} := \mathcal{F}_k$. Note, that then $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ for all k. This ensures that all later iterates will have to provide sufficient reduction with respect to x_k as defined by criterion (8), if the filter has been augmented in iteration k. Note, that for a practical implementation it is sufficient to store the "corner entries"

$$\Big((1-\gamma_\theta)\theta(x_k),\ f(x_k) - \gamma_f \theta(x_k)\Big). \tag{14}$$

It remains to decide which iterations should augment the filter. In order to keep the filter approach less conservative, we do not want to augment the filter in every iteration. In addition, as we see in the discussion of the next safeguard below, it is important for the proposed method that we never include feasible points in the filter. The following rule from [7] is motivated by these considerations.

We always augment the filter if the current iteration is not an f-type iteration, i.e. if for the accepted trial step size α_k the f-type switching condition (9) does not hold. Otherwise, the Armijo-condition (11) must be satisfied, and the value of the objective function is strictly decreased. To see that this indeed prevents cycling let us assume for a moment that the algorithm generates a cycle of length l

$$x_K, x_{K+1}, \ldots, x_{K+l-1}, x_{K+l} = x_K, x_{K+l+1} = x_{K+1}, \ldots \tag{15}$$

Since a point x_k can never be reached again if the filter is augmented in iteration k, the existence of a cycle would imply that the filter is not augmented for all $k \geq K$. However, this would imply that $f(x_k)$ is a strictly decreasing sequence for $k \geq K$, giving a contradiction, so that (15) cannot be a cycle.

2.3 Feasibility Restoration Phase

If the linear system (3) is consistent, d_k satisfies the linearization of the constraints and we have $\theta(x_k(\alpha_{k,l})) < \theta(x_k)$ whenever $\alpha_{k,l} > 0$ is sufficiently small. It is not guaranteed, however, that there exists a trial step size $\alpha_{k,l} > 0$ that indeed provides *sufficient* reduction as defined by criterion (8).

In this situation, where no admissible step size can be found, the method switches to a *feasibility restoration phase*, whose purpose is to find a new iterate x_{k+1} that satisfies (8) and is also acceptable to the current filter by trying to decrease the constraint violation. In this paper, we do not specify the particular procedure for this feasibility restoration phase. It could be any iterative algorithm with the goal of finding a less infeasible point, and different methods could even be used at different stages of the optimization procedure. For example, a nonlinear optimization algorithm might be applied to minimize θ, possibly ignoring the objective function. If the feasibility restoration phase terminates successfully by delivering a new admissible iterate, the filter is augmented according to (13) to avoid cycling back to the problematic point x_k.

Since a feasible iterate is never included in the filter (see Lemma 4 below), it is reasonable to assume that a suitable feasibility restoration phase algorithm is either able to find a new acceptable iterate satisfying (8), or converges to a local minimizer (or at least a stationary point) for some measure of infeasibility. The latter case may be important information for the user, as it indicates that the problem seems (at least locally) infeasible. This is of course no guarantee that the problem possesses no feasible point; proving infeasibility is as difficult as finding a global minimizer and beyond the capabilities of methods for finding local solutions like those discussed in this paper. However, we believe that it is a desirable practical feature of a nonlinear optimization code to return at least a local minimizer of the constraint violation if the method fails to find a solution of the optimization problem, instead of terminating at a less informative and possibly random point.

In order to detect the situation where no admissible step size can be found and the restoration phase has to be invoked, we propose the following rule. Consider the case when the current trial step size $\alpha_{k,l}$ is still large enough so that the f-type switching condition (9) holds for some $\alpha \leq \alpha_{k,l}$. In this case, we do not switch to the feasibility restoration phase, since there is still the chance that a shorter step length might be accepted by the Armijo condition (11). Therefore, we can see from the f-type switching condition (9) and the definition of m_k (10) that we do not want to revert to the feasibility restoration phase if $g_k^T d_k < 0$ and

$$\alpha_{k,l} > \frac{\delta[\theta(x_k)]^{s_\theta}}{[-g_k^T d_k]^{s_f}}. \tag{16}$$

However, if the f-type switching condition (9) is not satisfied for the current trial step size $\alpha_{k,l}$ and all shorter trial step sizes, then the decision whether to switch to the feasibility restoration phase is based on the linear approximations

$$\tilde{\theta}(x_k + \alpha d_k) = \theta(x_k) - \alpha\theta(x_k) \tag{17a}$$

$$\tilde{f}(x_k + \alpha d_k) = f(x_k) + \alpha g_k^T d_k. \tag{17b}$$

(Note that indeed $\tilde{\theta}(x_k+\alpha d_k) = \theta(x_k+\alpha d_k)+O(\alpha^2)$, since $A_k^T d_k+c(x_k) = 0$ from (3)). Substituting (17a) into the sufficient decrease condition for the infeasibility measure (8a) indicates that (8a) may not be satisfied for step sizes satisfying $\alpha_{k,l} \leq \gamma_\theta$. Similarly, in case $g_k^T d_k < 0$, the sufficient decrease criterion for the objective function (8b) may not be satisfied for step sizes satisfying

$$\alpha_{k,l} \leq \frac{\gamma_f \theta(x_k)}{-g_k^T d_k}.$$

We can summarize this in the following formula for a minimal trial step size

$$\alpha_k^{\min} := \gamma_\alpha \cdot \begin{cases} \min\left\{\gamma_\theta, \frac{\gamma_f\theta(x_k)}{-g_k^T d_k}, \frac{\delta[\theta(x_k)]^{s_\theta}}{[-g_k^T d_k]^{s_f}}\right\} & \text{if } g_k^T d_k < 0 \\ \gamma_\theta & \text{otherwise} \end{cases} \tag{18}$$

and switch to the feasibility restoration phase when $\alpha_{k,l}$ becomes smaller than $\alpha_k^{\min}$. Here, $\gamma_\alpha \in (0,1]$ is a safety-factor that might be useful in a practical implementation in order to compensate for the neglected higher order terms in the linearization (17) and to avoid invoking the feasibility restoration phase unnecessarily.

It is possible, however, to employ more sophisticated rules to decide when to switch to the feasibility restoration phase while still maintaining the convergence properties. These rules could, for example, be based on higher order approximations of θ and/or f. We only need to ensure that the algorithm does not switch to the feasibility restoration phase as long as (9) holds for a step size $\alpha \leq \alpha_{k,l}$ where $\alpha_{k,l}$ is the current trial step size, and that the backtracking line search procedure is finite, i.e. it eventually either delivers a new iterate x_{k+1} or reverts to the feasibility restoration phase.

The proposed method also allows to switch to the feasibility restoration phase in any iteration, in which the infeasibility $\theta(x_k)$ does not become arbitrarily small. For example, this might be necessary, when the Jacobian of the constraints A_k^T is (nearly) rank-deficient, so that the linear system (3) is (nearly) singular and no search direction can be computed. For the purpose of the analysis we assume that the algorithm is able to detect a situation in which the singular values of A_k become arbitrarily small and switch to the restoration phase in that case, even if the linear system can be solved numerically (see Assumption (G4) below). The search direction from (3) might still be used to generate the next iterate x_{k+1} using (5), as long as $x_{k+1} \notin \mathcal{F}_k$ and (8) can be satisfied. Even though we could consider this a non-f-type iteration, we formally treat this case as if the restoration phase is called. (Note that the iterate x_{k+1} returned from the restoration phase does not necessarily have to satisfy (8a) if (8b) holds instead.)

2.4 The Algorithm

We are now ready to formally state the overall algorithm for solving the equality constrained NLP (1).

Algorithm I

Given: Starting point x_0; constants $\theta_{\max} \in (\theta(x_0), \infty]$; $\gamma_\theta, \gamma_f \in (0,1)$; $\delta > 0$; $\gamma_\alpha \in (0,1]$; $s_\theta > 1$; $s_f \geq 1$; $\eta_f \in (0, \frac{1}{2})$; $0 < \tau_1 \leq \tau_2 < 1$.

1. *Initialize.* Initialize the filter $\mathcal{F}_0 := \{(\theta, f) \in \mathbb{R}^2 : \theta \geq \theta_{\max}\}$ and the iteration counter $k \leftarrow 0$.

2. *Check convergence.* Stop if x_k is a stationary point of the NLP (1), i.e. if it satisfies the KKT conditions (2) for some $\lambda \in \mathbb{R}^m$.

3. *Compute search direction.* Compute the search direction d_k from the linear system (3). If this system is detected to be too ill-conditioned (see the assumptions in the next section), go to the feasibility restoration phase in Step 8.

4. *Backtracking line search.*

   4.1. *Initialize line search.* Set $\alpha_{k,0} = 1$ and $l \leftarrow 0$.

   4.2. *Compute new trial point.* If the trial step size becomes too small, i.e. $\alpha_{k,l} < \alpha_k^{\min}$ with $\alpha_k^{\min}$ defined by (18), go to the feasibility restoration phase in Step 8. Otherwise, compute the new trial point $x_k(\alpha_{k,l}) = x_k + \alpha_{k,l} d_k$.

   4.3. *Check acceptability to the filter.* If $x_k(\alpha_{k,l}) \in \mathcal{F}_k$, reject the trial step size and go to Step 4.5.

   4.4. *Check sufficient decrease with respect to current iterate.*

      4.4.1. *Case I: $\alpha_{k,l}$ is an f-step-size (i.e. (9) holds):* If the Armijo condition (11) for the objective function holds, accept the trial step and go to Step 5.
      Otherwise, go to Step 4.5.

      4.4.2. *Case II: $\alpha_{k,l}$ is not an f-step-size (i.e. (9) is not satisfied):* If (8) holds, accept the trial step and go to Step 5.
      Otherwise, go to Step 4.5.

   4.5. *Choose new trial step size.* Choose $\alpha_{k,l+1} \in [\tau_1 \alpha_{k,l}, \tau_2 \alpha_{k,l}]$, set $l \leftarrow l+1$, and go back to Step 4.2.

5. *Accept trial point.* Set $\alpha_k := \alpha_{k,l}$ and $x_{k+1} := x_k(\alpha_k)$.

6. *Augment filter if necessary.* If k is not an f-type iteration, augment the filter using (13); otherwise leave the filter unchanged, i.e. set $\mathcal{F}_{k+1} := \mathcal{F}_k$.

   (Note, that Step 4.3 and Step 4.4.2 ensure, that $(\theta(x_{k+1}), f(x_{k+1})) \notin \mathcal{F}_{k+1}$.)

7. *Continue with next iteration.* Increase the iteration counter $k \leftarrow k+1$ and go back to Step 2.

8. *Feasibility restoration phase.* Compute a new iterate x_{k+1} by decreasing the infeasibility measure θ, so that x_{k+1} satisfies the sufficient decrease conditions (8) and is acceptable to the filter, i.e. $(\theta(x_{k+1}), f(x_{k+1})) \notin \mathcal{F}_k$. Augment the filter using (13) (for x_k) and continue with the regular iteration in Step 7.

2.5 Remarks

Remark 1 *From Step 4.5 it is clear that $\lim_l \alpha_{k,l} = 0$. In the case that $\theta(x_k) > 0$ it can be seen from (18) that $\alpha_k^{\min} > 0$. Therefore, the algorithm either accepts a new iterate in Step 4.4, or switches to the feasibility restoration phase. If on the other hand $\theta(x_k) = 0$ and the algorithm does not stop in Step 2 at a KKT point, then the positive definiteness of H_k on the null space of A_k^T implies that $g_k^T d_k < 0$ (see e.g. Lemma 4 below). In that case, $\alpha_k^{\min} = 0$, and the Armijo condition (11) is satisfied for a sufficiently small step size $\alpha_{k,l}$, i.e. a new iterate is accepted in Step 4.4.1. Overall, we see that the inner loop in Step 4 always terminates after a finite number of trial steps, and the algorithm is well-defined.*

Remark 2 *The algorithm generates an infinite sequence $\{x_k\}$ of iterates, unless it encounters a KKT point and terminates in Step 2, or if the feasibility restoration phase in Step 8 is not able to return a new iterate. In the latter case, the restoration phase algorithm converges to a stationary point for the constraint violation, assuming that a suitable method is used.*

Remark 3 *The mechanisms of the filter ensure that $(\theta(x_k), f(x_k)) \notin \mathcal{F}_k$ for all k. Furthermore, the initialization of the filter in Step 1 and the update rule (13) imply that for all k the filter has the following property:*

$$(\bar{\theta}, \bar{f}) \notin \mathcal{F}_k \quad \Longrightarrow \quad (\theta, f) \notin \mathcal{F}_k \text{ if } \theta \leq \bar{\theta} \text{ and } f \leq \bar{f}. \tag{19}$$

Remark 4 *For practical purposes, it might not be efficient to restrict the step size by enforcing an Armijo-type decrease (11) in the objective function, if the current constraint violation is not small. It is possible to change the algorithm so that the step acceptance criterion is always (8), unless the f-type switching condition (9) holds* and *$\theta(x_k) \leq \theta_{\text{sml}}$ for some fixed $\theta_{\text{sml}} > 0$, in which case the Armijo condition (11) has to be satisfied. In this modified method, the filter is augmented (using (13)), whenever (9) or (11) does not hold. The global convergence properties are not affected by this modification.*

Remark 5 *The proposed method has many similarities with the trust region filter SQP method proposed and analyzed in [7]. However, we discuss a more general f-type switching rule (9) in order to be able to show fast local convergence in the companion paper [26]. Further differences result from the fact, that the proposed method follows a line search approach, so that in contrast to [7] the actual step taken does not necessarily satisfy the linearization of the constraints, i.e. we might have $A_k^T(x_k - x_{k+1}) \neq c(x_k)$ in some iterations. As a related consequence, the condition when to switch to the feasibility restoration phase in Step 4.2 could not be chosen to be the detection of* *infeasibility of the trust region QP, but has to be defined by means of a minimal step size (18). Due to these differences, the global convergence analysis presented in [7] does not apply to the proposed line search filter method.*

3 Global Convergence

3.1 Assumptions

In the remainder of this paper we denote the set of indices of those iterations, in which the filter has been augmented, by $\mathcal{A} \subseteq \mathbb{N}$; i.e.

$$\mathcal{F}_k \subsetneqq \mathcal{F}_{k+1} \qquad \Longleftrightarrow \qquad k \in \mathcal{A}.$$

The set $\mathcal{R} \subseteq \mathbb{N}$ is defined as the set of all iteration indices in which the feasibility restoration phase is invoked. Since Step 8 makes sure that the filter is augmented in every iteration in which the restoration phase is invoked, we have $\mathcal{R} \subseteq \mathcal{A}$. We denote with $\mathcal{R}_{\text{inc}} \subseteq \mathcal{R}$ the set of those iteration counters, in which the restoration phase is invoked from Step 3.

Let us now state the assumptions necessary for the global convergence analysis of Algorithm I. We first state these assumptions in technical terms, and discuss their practical relevance afterwards.

Assumptions G. *Let $\{x_k\}$ be the sequence generated by Algorithm I, where we assume that the feasibility restoration phase in Step 8 always terminates successfully and that the algorithm does not stop in Step 2 at a KKT point.*

(G1) *There exists an open set $\mathcal{C} \subseteq \mathbb{R}^n$ with $[x_k, x_k + d_k] \subseteq \mathcal{C}$ for all $k \notin \mathcal{R}_{\text{inc}}$, so that f and c are differentiable on $\mathcal{C}$, and their function values, as well as their first derivatives, are bounded and Lipschitz-continuous over $\mathcal{C}$.*

(G2) *The matrices H_k approximating the Hessian of the Lagrangian in (3) are uniformly bounded for all $k \notin \mathcal{R}_{\text{inc}}$.*

(G3) *The Hessian approximations H_k are uniformly positive definite on the null space of the Jacobian A_k^T. In other words, there exists a constant $M_H > 0$, so that for all $k \notin \mathcal{R}_{\text{inc}}$*

$$\lambda_{\min}\left(Z_k^T H_k Z_k\right) \geq M_H, \tag{20}$$

where the columns of $Z_k \in \mathbb{R}^{n\times(n-m)}$ form an orthonormal basis matrix of the null space of A_k^T.

(G4) *There exists a constant $M_A > 0$, so that for all $k \notin \mathcal{R}_{\text{inc}}$ we have*

$$\sigma_{\min}(A_k) \geq M_A. \tag{21}$$

(G5) *The iterates, for which the restoration phase is invoked from Step 3 (for example, because (20) or (21) are violated), are not arbitrarily close to the feasible region. In other words, there exists a constant $\theta_{\text{inc}} > 0$, so that $k \notin \mathcal{R}_{\text{inc}}$ whenever $\theta(x_k) \leq \theta_{\text{inc}}$*

Assumptions (G1) and (G2) merely establish smoothness and boundedness of the problem data. As we see later in Lemma 2, Assumption (G3) ensures a certain descent property and it is similar to common assumptions on the reduced Hessian in SQP line search methods (see e.g. [17]). To guarantee this requirement in a practical implementation, one could compute a QR-factorization of A_k to obtain matrices $Y_k \in \mathbb{R}^{n\times m}$ and $Z_k \in \mathbb{R}^{n\times(n-m)}$ so that the columns of $[Y_k\ Z_k]$ form an orthonormal basis of $\mathbb{R}^n$, and the columns of Z_k are a basis of the null space of A_k^T (see e.g. [11]). Then, the overall search direction can be decomposed into two orthogonal components,

$$d_k = q_k + p_k, \qquad \text{where} \tag{22a}$$
$$q_k := Y_k\bar{q}_k \quad \text{and} \quad p_k := Z_k\bar{p}_k, \tag{22b}$$

with

$$\bar{q}_k := -\left[A_k^T Y_k\right]^{-1} c_k \tag{23a}$$
$$\bar{p}_k := -\left[Z_k^T H_k Z_k\right]^{-1} Z_k^T\left(g_k + H_k q_k\right) \tag{23b}$$

(see e.g. [17]). The eigenvalues for the reduced Hessian in (23b) (the term in square brackets) could be monitored and modified if necessary. However, this procedure is prohibitive for large-scale problems, and in those cases one instead might employ heuristics to ensure at least positive definiteness of the reduced Hessian, for example, by monitoring and possibly modifying the inertia of the iteration matrix in (3) (see e.g. [23]). Note, on the other hand, that (20) holds in the neighborhood of a local solution x_* satisfying the sufficient second order optimality conditions (see e.g. [17]), if H_k approaches the exact Hessian of the Lagrangian of the NLP (1). Then, close to x_*, no eigenvalue correction is necessary and fast local convergence can be expected, assuming that full steps are taken close to x_*. See the companion paper [26] for a local convergence analysis of the presented method.

In the description of the algorithm in Section 2.4 we did not specify when precisely the method switches in Step 3 to the feasibility restoration phase, since there might be several practical implementations compatible with Assumptions G. For completeness, one possible option is outlined next. By monitoring and possibly modifying the eigenvalues of the reduced Hessian it is possible to make sure that (20) is valid in *every* iteration. Similarly, we can guarantee that the *entire* sequence $\{H_k\}$ is uniformly bounded. Let us further make the assumption (on the problem statement) that the gradients of the constraints are uniformly linearly independent for all iterates x_k close to the feasible region, i.e. there exist constants $b_1, b_2 > 0$, so that

$$\theta(x_k) \le b_1 \qquad \Longrightarrow \qquad \sigma_{\min}(A_k) \ge b_2.$$

Then, if we decide in Step 3 to invoke the feasibility restoration phase whenever $\sigma_{\min}(A_k) \le b_3\theta(x_k)$ for some fixed constant $b_3 > 0$, then Assumptions G hold (with $M_A = \min\{b_2, b_1 b_3\}$ and $\theta_{\text{inc}} = \frac{M_A}{2b_3}$).

3.2 Preliminary Results

Similar to the analysis in [7], we make use of a *first order criticality measure* $\chi(x_k) \in [0,\infty]$ with the property that, if a subsequence $\{x_{k_i}\}$ of iterates with $\chi(x_{k_i}) \to 0$ converges to a feasible limit point x_*, then x_* corresponds to a KKT solution. In the case of the Algorithm I this means that there exist λ_*, so that the KKT conditions (2) are satisfied for (x_*, λ_*).

For the convergence analysis of the filter method we define the criticality measure for iterations $k \notin \mathcal{R}_{\text{inc}}$ as

$$\chi(x_k) := \|\bar{p}_k\|_2\,, \tag{24}$$

with $\bar{p}_k$ from (23b). Note that this definition is unique, since p_k in (22a) is unique due to the orthogonality of Y_k and Z_k, and since $\|\bar{p}_k\|_2 = \|p_k\|_2$ due to the orthonormality of Z_k. For completeness, we define $\chi(x_k) := \infty$ for $k \in \mathcal{R}_{\text{inc}}$.

In order to see that $\chi(x_k)$ defined in this way is indeed a criticality measure under Assumptions G, let us consider a subsequence of iterates $\{x_{k_i}\}$ with $\lim_i \chi(x_{k_i}) = 0$ and $\lim_i x_{k_i} = x_*$ for some feasible limit point x_*. Since $\chi(x_{k_i}) = \infty$ if $k_i \in \mathcal{R}_{\rm inc}$, we then have $k_i \notin \mathcal{R}_{\rm inc}$ for i sufficiently large. Furthermore, from Assumption (G4) and (23a) we have $\lim_i \bar{q}_{k_i} = 0$, and then from $\lim_i \chi(x_{k_i}) = 0$, (24), (23b), and Assumption (G3) we have that $\lim_{i\to\infty} \|Z_{k_i}^T g_{k_i}\| = 0$, which is a well-known optimality measure (see e.g. [17]).

Before we begin the global convergence analysis, let us state some preliminary results.

Lemma 1 *Suppose Assumptions G hold. Then there exist constants M_d, M_λ, $M_m > 0$, such that*

$$\|d_k\| \le M_d, \qquad \|\lambda_k^+\| \le M_\lambda, \qquad |m_k(\alpha)| \le M_m \alpha \tag{25}$$

for all $k \notin \mathcal{R}_{\rm inc}$ and $\alpha \in (0,1]$.

Proof. From (G1) we have that the right hand side of (3) is uniformly bounded. Additionally, Assumptions (G2), (G3), and (G4) guarantee that the inverse of the matrix in (3) exists and is uniformly bounded for all $k \notin \mathcal{R}_{\rm inc}$. Consequently, the solution of (3), (d_k, λ_k^+), is uniformly bounded, and therefore also $m_k(\alpha)/\alpha = g_k^T d_k$. □

The following result shows that the search direction is a direction of sufficient descent for the objective function at points that are sufficiently close to feasible and non-optimal.

Lemma 2 *Suppose Assumptions G hold. If $\{x_{k_i}\}$ is a subsequence of iterates for which $\chi(x_{k_i}) \ge \epsilon$ with a constant $\epsilon > 0$ independent of i then there exist constants $\epsilon_1, \epsilon_2 > 0$, such that*

$$\theta(x_{k_i}) \le \epsilon_1 \qquad \Longrightarrow \qquad m_{k_i}(\alpha) \le -\epsilon_2 \alpha.$$

for all i and $\alpha \in (0,1]$.

Proof. Consider a subset $\{x_{k_i}\}$ of iterates with $\chi(x_{k_i}) = \|\bar{p}_{k_i}\|_2 \ge \epsilon$. Then, by Assumption (G5), for all x_{k_i} with $\theta(x_{k_i}) \le \theta_{\rm inc}$ we have $k_i \notin \mathcal{R}_{\rm inc}$. Furthermore, with $q_{k_i} = O(\|c(x_{k_i})\|)$ (from (23a) and Assumption (G4)) it follows that for $k_i \notin \mathcal{R}_{\rm inc}$

$$\begin{aligned}
m_{k_i}(\alpha)/\alpha \quad &= \quad g_{k_i}^T d_{k_i} \overset{(22)}{=} g_{k_i}^T Z_{k_i} \bar{p}_{k_i} + g_{k_i}^T q_{k_i} && \text{(26a)}\\
&\overset{(23b)}{=} \quad -\bar{p}_{k_i}^T \left[Z_{k_i}^T H_{k_i} Z_{k_i}\right] \bar{p}_{k_i} - \bar{p}_{k_i}^T Z_{k_i}^T H_{k_i} q_{k_i} + g_{k_i}^T q_{k_i} && \text{(26b)}\\
&\overset{(G2),(G3)}{\le} \quad -c_1 \|\bar{p}_{k_i}\|_2^2 + c_2 \|\bar{p}_{k_i}\|_2 \|c_{k_i}\| + c_3 \|c_{k_i}\| && \text{(26c)}\\
&\le \quad \chi(x_{k_i}) \left(-\epsilon\, c_1 + c_2 \theta(x_{k_i}) + \frac{c_3}{\epsilon}\theta(x_{k_i})\right) && \text{(26d)}
\end{aligned}$$

for some constants $c_1, c_2, c_3 > 0$, where we used $\chi(x_{k_i}) \ge \epsilon$ in the last inequality. If we now define

$$\epsilon_1 := \min\left\{\theta_{\rm inc}, \frac{\epsilon^2\, c_1}{2(c_2\, \epsilon + c_3)}\right\},$$

it follows for all x_{k_i} with $\theta(x_{k_i}) \le \epsilon_1$ that

$$m_{k_i}(\alpha) \le -\alpha \frac{\epsilon\, c_1}{2} \chi(x_{k_i}) \le -\alpha \frac{\epsilon^2\, c_1}{2}.$$

The claim follows after defining $\epsilon_2 := \frac{\epsilon^2\, c_1}{2}$. □

Lemma 3 *Suppose Assumption (G1) holds. Then there exist constants* $C_\theta, C_f > 0$*, so that for all* $k \notin \mathcal{R}_{\rm inc}$ *and* $\alpha \leq 1$

$$|\theta(x_k + \alpha d_k) - (1-\alpha)\theta(x_k)| \quad \leq \quad C_\theta \alpha^2 \|d_k\|^2 \tag{27a}$$
$$|f(x_k + \alpha d_k) - f(x_k) - m_k(\alpha)| \quad \leq \quad C_f \alpha^2 \|d_k\|^2. \tag{27b}$$

These inequalities follow directly from second order Taylor expansions and (3).

Finally, we show that Step 8 (feasibility restoration phase) of Algorithm I is well-defined. Unless the feasibility restoration phase terminates at a stationary point of the constraint violation it is essential that reducing the infeasibility measure $\theta(x)$ eventually leads to a point that is acceptable to the filter. This is guaranteed by the following lemma which shows that no (θ, f)-pair corresponding to a feasible point is ever included in the filter.

Lemma 4 *Suppose Assumptions G hold. Then*

$$\theta(x_k) = 0 \quad \Longrightarrow \quad m_k(\alpha) < 0 \qquad \textit{and} \tag{28}$$

$$\Theta_k := \min\{\theta : (\theta, f) \in \mathcal{F}_k\} > 0 \tag{29}$$

for all k *and* $\alpha \in (0,1]$.

Proof. If $\theta(x_k) = 0$, we have from Assumption (G5) that $k \notin \mathcal{R}_{\rm inc}$. In addition, it then follows $\chi(x_k) > 0$ because Algorithm I would have terminated otherwise in Step 2, in contrast to Assumptions G. Considering the decomposition (22), it follows as in (26) that

$$m_k(\alpha)/\alpha = g_k^T d_k \leq -c_1 \chi(x_k)^2 < 0,$$

i.e. (28) holds.

The proof of (29) is by induction. It is clear from Step 1 of Algorithm I, that the claim is valid for $k = 0$ since $\theta_{\max} > 0$. Suppose the claim is true for k. Then, if $\theta(x_k) > 0$ and the filter is augmented in iteration k, it is clear from the update rule (13), that $\Theta_{k+1} > 0$, since $\gamma_\theta \in (0,1)$. If on the other hand $\theta(x_k) = 0$, we have from (28) that $m_k(\alpha) < 0$ for all $\alpha \in (0,1]$, so that the f-type switching condition (9) is true for all trial step sizes. Therefore, Step 4.4 considers always "Case I", and the reason for α_k having been accepted must have been that α_k satisfies (11). Consequently, the filter is not augmented in Step 6. Hence, $\Theta_{k+1} = \Theta_k > 0$. □

3.3 Feasibility

In this section we show that under Assumptions G the sequence $\theta(x_k)$ converges to zero, i.e. all limit points of $\{x_k\}$ are feasible.

Lemma 5 *Suppose that Assumptions G hold, and that the filter is augmented only a finite number of times, i.e.* $|\mathcal{A}| < \infty$*. Then*

$$\lim_{k\to\infty} \theta(x_k) = 0. \tag{30}$$

Proof. Choose K, so that for all iterations $k \geq K$ the filter is not augmented in iteration k; in particular, $k \notin \mathcal{R}_{\rm inc} \subseteq \mathcal{A}$ for $k \geq K$. From Step 6 in Algorithm I we then have, that for all $k \geq K$ both conditions (9) and (11) are satisfied for α_k. We now distinguish two cases, where $k \notin \mathcal{A}$.

Case I ($s_f > 1$): From (9) it follows with M_m from Lemma 1 that

$$\delta[\theta(x_k)]^{s_\theta} < [-m_k(\alpha_k)]^{s_f}[\alpha_k]^{1-s_f} \leq M_m^{s_f}\alpha_k$$

and hence (since $1 - 1/s_f > 0$)

$$c_4[\theta(x_k)]^{s_\theta - \frac{s_\theta}{s_f}} < [\alpha_k]^{1-\frac{1}{s_f}} \qquad \text{with} \qquad c_4 := \left(\frac{\delta}{M_m^{s_f}}\right)^{1-\frac{1}{s_f}}.$$

This implies

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\overset{(11)}{\leq} \eta_f m_k(\alpha_k) \\ &\overset{(9)}{<} -\eta_f \delta^{\frac{1}{s_f}}[\alpha_k]^{1-\frac{1}{s_f}}[\theta(x_k)]^{\frac{s_\theta}{s_f}} \\ &< -\eta_f \delta^{\frac{1}{s_f}} c_4[\theta(x_k)]^{s_\theta}. \end{aligned}$$

Case II ($s_f = 1$): From (9) we have $\delta[\theta(x_k)]^{s_\theta} < -m_k(\alpha_k)$, so that from (11) we immediately obtain $f(x_{k+1}) - f(x_k) < -\eta_f \delta[\theta(x_k)]^{s_\theta}$.

In either case, we have for all $k \notin \mathcal{A}$ that

$$f(x_{k+1}) - f(x_k) < -\tilde{c}_4[\theta(x_k)]^{s_\theta} \tag{31}$$

for some $\tilde{c}_4 > 0$. Hence, for all $i = 1, 2, \ldots$,

$$\begin{aligned} f(x_{K+i}) &= f(x_K) + \sum_{k=K}^{K+i-1}(f(x_{k+1}) - f(x_k)) \\ &< f(x_K) - \tilde{c}_4 \sum_{k=K}^{K+i-1}[\theta(x_k)]^{s_\theta}. \end{aligned}$$

Since $f(x_{K+i})$ is bounded below as $i \to \infty$, the series on the right hand side in the last line is bounded, which in turn implies (30). □

Note that this result could be obtained with a simpler proof if the model $m_k(\alpha)$ has the particular form (10), but the above version also holds for the model (56) in Section 4.1.

The following lemma considers a subsequence $\{x_{k_i}\}$ with $k_i \in \mathcal{A}$ for all i.

Lemma 6 *Let $\{x_{k_i}\}$ be a subsequence of iterates generated by Algorithm I, so that the filter is augmented in iteration k_i, i.e. $k_i \in \mathcal{A}$ for all i. Furthermore assume that there exist constants $c_f \in \mathbb{R}$ and $C_\theta > 0$, so that*

$$f(x_{k_i}) \geq c_f \qquad \textit{and} \qquad \theta(x_{k_i}) \leq C_\theta$$

for all i (for example, if Assumptions (G1) holds). It then follows that

$$\lim_{i\to\infty} \theta(x_{k_i}) = 0.$$

Its proof can be found in [7, Lemma 3.3]. There, it is stated for slightly different circumstances, but it is easy to verify that it is still valid in our context.

The previous two lemmas prepare the proof of the following theorem.

Theorem 1 *Suppose Assumptions G hold. Then*

$$\lim_{k\to\infty} \theta(x_k) = 0.$$

Proof. In the case, that the filter is augmented only a finite number of times, Lemma 5 implies the claim. If in the other extreme there exists some $K \in \mathbb{N}$, so that the filter is updated by (13) in *all* iterations $k \geq K$, then the claim follows from Lemma 6. It remains to consider the case, where for all $K \in \mathbb{N}$ there exist $k_1, k_2 \geq K$ with $k_1 \in \mathcal{A}$ and $k_2 \notin \mathcal{A}$.

The proof is by contradiction. Suppose, $\limsup_k \theta(x_k) = M > 0$. Now construct two subsequences $\{x_{k_i}\}$ and $\{x_{l_i}\}$ of $\{x_k\}$ in the following way.

1. Set $i \leftarrow 0$ and $k_{-1} = -1$.
2. Pick $k_i > k_{i-1}$ with
$$\theta(x_{k_i}) \geq M/2 \tag{32}$$
and $k_i \notin \mathcal{A}$. (Note that Lemma 6 ensures the existence of $k_i \notin \mathcal{A}$ since otherwise $\theta(x_{k_i}) \to 0$.)
3. Choose $l_i := \min\{l \in \mathcal{A} : l > k_i\}$, i.e. l_i is the first iteration after k_i in which the filter is augmented.
4. Set $i \leftarrow i + 1$ and go back to Step 2.

Thus, every x_{k_i} satisfies (32), and for each x_{k_i} the iterate x_{l_i} is the first iterate after x_{k_i} for which $(\theta(x_{l_i}), f(x_{l_i}))$ is included in the filter.

Since (31) holds for all $k = k_i, \ldots, l_i - 1 \notin \mathcal{A}$, we obtain for all i

$$f(x_{l_i}) \leq f(x_{(k_i+1)}) < f(x_{k_i}) - \tilde{c}_4[M/2]^{s_\theta}. \tag{33}$$

This ensures that for all $K \in \mathbb{N}$ there exists some $i \geq K$ with $f(x_{k_{(i+1)}}) \geq f(x_{l_i})$ because otherwise (33) would imply

$$f(x_{k_{(i+1)}}) < f(x_{l_i}) < f(x_{k_i}) - \tilde{c}_4[M/2]^{s_\theta}$$

for all i and consequently $\lim_i f(x_{k_i}) = -\infty$ in contradiction to the fact that $\{f(x_k)\}$ is bounded below. Thus, there exists a subsequence $\{i_j\}$ of $\{i\}$ so that

$$f(x_{k_{(i_j+1)}}) \geq f(x_{l_{i_j}}). \tag{34}$$

Since $x_{k_{(i_j+1)}} \notin \mathcal{F}_{k_{(i_j+1)}} \supseteq \mathcal{F}_{l_{i_j}}$ and $l_{i_j} \in \mathcal{A}$, it follows from (34) and the filter update rule (13), that

$$\theta(x_{k_{(i_j+1)}}) \leq (1 - \gamma_\theta)\theta(x_{l_{i_j}}). \tag{35}$$

Since $l_{i_j} \in \mathcal{A}$ for all j, Lemma 6 yields $\lim_j \theta(x_{l_{i_j}}) = 0$, so that from (35) we obtain $\lim_j \theta(x_{k_{i_j}}) = 0$ in contradiction to (32). □

Remark 6 *As one can easily verify, if $s_f = 1$ is chosen in the f-type switching rule (9), then the proof of the previous theorem does not actually require the assumption that $\{f(x_k)\}$ and $\{\|\nabla f(x_k)\|\}$ are bounded above (see Assumption (G1)). This is important for the discussion of the interior point method in Section 4.3.*

3.4 Optimality

In this section we show that Assumptions G guarantee that the optimality measure $\chi(x_k)$ is not bounded away from zero, i.e. if $\{x_k\}$ is bounded, at least one limit point is a first order optimal point for the NLP (1).

The first lemma shows conditions under which it can be guaranteed that there exists a step length bounded away from zero so that the Armijo condition (11) for the objective function is satisfied.

Lemma 7 *Suppose Assumptions G hold. Let $\{x_{k_i}\}$ be a subsequence with $k_i \notin \mathcal{R}_{\rm inc}$ and $m_{k_i}(\alpha) \leq -\alpha\epsilon_2$ for a constant $\epsilon_2 > 0$ independent of k_i and for all $\alpha \in (0, 1]$. Then there exists some constant $\bar{\alpha} > 0$, so that for all k_i and $\alpha \leq \bar{\alpha}$*

$$f(x_{k_i} + \alpha d_{k_i}) - f(x_{k_i}) \leq \eta_f m_{k_i}(\alpha). \tag{36}$$

Proof. Let M_d and C_f be the constants from Lemma 1 and Lemma 3. It then follows for all $\alpha \leq \bar{\alpha}$ with $\bar{\alpha} := \frac{(1-\eta_f)\epsilon_2}{C_f M_d^2}$ that

$$\begin{aligned} & f(x_{k_i} + \alpha d_{k_i}) - f(x_{k_i}) - m_{k_i}(\alpha) \\ \stackrel{(27b)}{\leq} \;& C_f \alpha^2 \|d_{k_i}\|^2 \leq \alpha(1-\eta_f)\epsilon_2 \\ \leq \;& -(1-\eta_f) m_{k_i}(\alpha), \end{aligned}$$

which implies (36). □

Let us again first consider the "easy" case, in which the filter is augmented only a finite number of times.

Lemma 8 *Suppose that Assumptions G hold and that the filter is augmented only a finite number of times, i.e. $|\mathcal{A}| < \infty$. Then*

$$\lim_{k\to\infty} \chi(x_k) = 0.$$

Proof. Since $|\mathcal{A}| < \infty$, there exists $K \in \mathbb{N}$ so that $k \notin \mathcal{A}$ for all $k \geq K$. Suppose, the claim is not true, i.e. there exists a subsequence $\{x_{k_i}\}$ and a constant $\epsilon > 0$, so that $\chi(x_{k_i}) \geq \epsilon$ for all i. From (30) and Lemma 2 there exist $\epsilon_1, \epsilon_2 > 0$ and $\tilde{K} \geq K$, so that for all $k_i \geq \tilde{K}$ we have $\theta(x_{k_i}) \leq \epsilon_1$ and

$$m_{k_i}(\alpha) \leq -\alpha\epsilon_2 \quad \text{for all} \quad \alpha \in (0, 1]. \tag{37}$$

It then follows from (11) that for $k_i \geq \tilde{K}$

$$f(x_{k_i+1}) - f(x_{k_i}) \leq \eta_f m_{k_i}(\alpha_{k_i}) \leq -\alpha_{k_i}\eta_f\epsilon_2.$$

Reasoning similarly as in proof of Lemma 5, one can conclude that $\lim_i \alpha_{k_i} = 0$, since $f(x_{k_i})$ is bounded below and since $f(x_k)$ is monotonically decreasing (from (31)) for all $k \geq \tilde{K}$. We can now assume without loss of generality that $\tilde{K}$ is sufficiently large, so that $\alpha_{k_i} < 1$. This means that for $k_i \geq \tilde{K}$ the first trial step $\alpha_{k,0} = 1$ has not been accepted. The last rejected trial step size

$$\alpha_{k_i,l_i} \in [\alpha_{k_i}/\tau_2, \alpha_{k_i}/\tau_1] \tag{38}$$

during the backtracking line search procedure then satisfies (9) since $k_i \notin \mathcal{A}$ and $\alpha_{k_i,l_i} > \alpha_{k_i}$. Thus, it must have been rejected because it violates (11), i.e. it satisfies

$$f(x_{k_i} + \alpha_{k_i,l_i} d_{k_i}) - f(x_{k_i}) > \eta_f m_{k_i}(\alpha_{k_i,l_i}), \tag{39}$$

or it has been rejected because it is not acceptable to the current filter, i.e.

$$(\theta(x_{k_i} + \alpha_{k_i,l_i} d_{k_i}), f(x_{k_i} + \alpha_{k_i,l_i} d_{k_i})) \in \mathcal{F}_{k_i} = \mathcal{F}_K. \tag{40}$$

We conclude the proof by showing that neither (39) nor (40) can be true for sufficiently large k_i.

To (39): Since $\lim_i \alpha_{k_i} = 0$, we also have $\lim_i \alpha_{k_i,l_i} = 0$ (see (38)). In particular, for sufficiently large k_i we have $\alpha_{k_i,l_i} \le \bar{\alpha}$ with $\bar{\alpha}$ from Lemma 7, i.e. (39) cannot be satisfied for those k_i.

To (40): Let $\Theta_K := \min\{\theta : (\theta, f) \in \mathcal{F}_K\}$. From Lemma 4 we have $\Theta_K > 0$. Using Lemma 1 and Lemma 3, we then see that

$$\theta(x_{k_i} + \alpha_{k_i,l_i} d_{k_i}) \le (1 - \alpha_{k_i,l_i})\theta(x_{k_i}) + C_\theta M_d^2 [\alpha_{k_i,l_i}]^2.$$

Since $\lim_i \alpha_{k_i,l_i} = 0$ and from Theorem 1 also $\lim_i \theta(x_{k_i}) = 0$, it follows that for k_i sufficiently large we have $\theta(x_{k_i} + \alpha_{k_i,l_i} d_{k_i}) < \Theta_K$ which contradicts (40). $\square$

The next lemma establishes conditions under which a step size can be found that is acceptable to the current filter (see (12)).

Lemma 9 *Suppose Assumptions G hold. Let $\{x_{k_i}\}$ be a subsequence with $k_i \notin \mathcal{R}_{\text{inc}}$ and $m_{k_i}(\alpha) \le -\alpha\epsilon_2$ for a constant $\epsilon_2 > 0$ independent of k_i and for all $\alpha \in (0, 1]$. Then there exist constants $c_5, c_6 > 0$ so that*

$$(\theta(x_{k_i} + \alpha d_{k_i}), f(x_{k_i} + \alpha d_{k_i})) \notin \mathcal{F}_{k_i}$$

for all k_i and $\alpha \le \min\{c_5, c_6\theta(x_{k_i})\}$.

Proof. Let M_d, C_θ, and C_f be the constants from Lemma 1 and Lemma 3. Define $c_5 := \min\{1, \epsilon_2/(M_d^2\, C_f)\}$ and $c_6 := 1/(M_d^2\, C_\theta)$.

Now choose an iterate x_{k_i}. The mechanisms of Algorithm I ensure (see comment in Step 6), that

$$(\theta(x_{k_i}), f(x_{k_i})) \notin \mathcal{F}_{k_i}. \tag{41}$$

For $\alpha \le c_5$ we have $\alpha^2 \le \frac{\alpha\epsilon_2}{M_d^2\, C_f} \le \frac{-m_{k_i}(\alpha)}{C_f \|d_{k_i}\|^2}$, or equivalently

$$m_{k_i}(\alpha) + C_f \alpha^2 \|d_{k_i}\|^2 \le 0,$$

and it follows with (27b) that

$$f(x_{k_i} + \alpha d_{k_i}) \le f(x_{k_i}). \tag{42}$$

Similarly, for $\alpha \le c_6\theta(x_{k_i}) \le \frac{\theta(x_{k_i})}{\|d_{k_i}\|^2\, C_\theta}$, we have $-\alpha\theta(x_{k_i}) + C_\theta \alpha^2 \|d_{k_i}\|^2 \le 0$ and thus from (27a)

$$\theta(x_{k_i} + \alpha d_{k_i}) \le \theta(x_{k_i}). \tag{43}$$

The claim then follows from (41), (42) and (43) using (19). $\square$

The last lemma in this section shows that in iterations corresponding to a subsequence with only non-optimal limit points the filter is eventually not augmented. This result is used in the proof of the main global convergence theorem to yield a contradiction.

Lemma 10 *Suppose Assumptions G hold. Let $\{x_{k_i}\}$ be a subsequence with $\chi(x_{k_i}) \ge \epsilon$ for a constant $\epsilon > 0$ independent of k_i. Then there exists $K \in \mathbb{N}$, so that for all $k_i \ge K$ the filter is not augmented in iteration k_i, i.e. $k_i \notin \mathcal{A}$.*

Proof. Since by Theorem 1 we have $\lim_i \theta(x_{k_i}) = 0$, it follows from Lemma 2 that there exist constants $\epsilon_1, \epsilon_2 > 0$, so that

$$\theta(x_{k_i}) \leq \epsilon_1 \quad \text{and} \quad m_{k_i}(\alpha) \leq -\alpha\epsilon_2 \tag{44}$$

for k_i sufficiently large and $\alpha \in (0, 1]$; without loss of generality we can assume that (44) is valid for all k_i. We can now apply Lemma 7 and Lemma 9 to obtain the constants $\bar{\alpha}, c_5, c_6 > 0$. Choose $K \in \mathbb{N}$, so that for all $k_i \geq K$

$$\theta(x_{k_i}) < \min\left\{\theta_{\text{inc}}, \frac{\bar{\alpha}}{c_6}, \frac{c_5}{c_6}, \left[\frac{\tau_1 c_6 \epsilon_2^{s_f}}{\delta}\right]^{\frac{1}{s_\theta - 1}}\right\} \tag{45}$$

with τ_1 from Step 4.5. For all $k_i \geq K$ with $\theta(x_{k_i}) = 0$ we can argue as in the proof of Lemma 4 that both (9) and (11) hold in iteration k_i, so that $k_i \notin \mathcal{A}$.

For the remaining iterations $k_i \geq K$ with $\theta(x_{k_i}) > 0$ we note that (45) implies that $k_i \notin \mathcal{R}_{\text{inc}}$,

$$\frac{\delta\,[\theta(x_{k_i})]^{s_\theta}}{\epsilon_2^{s_f}} < \tau_1 c_6 \theta(x_{k_i}) \tag{46}$$

(since $s_\theta > 1$), as well as

$$c_6 \theta(x_{k_i}) < \min\{\bar{\alpha}, c_5\}. \tag{47}$$

Now choose an arbitrary $k_i \geq K$ with $\theta(x_{k_i}) > 0$ and define

$$\beta_{k_i} := c_6 \theta(x_{k_i}) \overset{(47)}{=} \min\{\bar{\alpha}, c_5, c_6\theta(x_{k_i})\}. \tag{48}$$

Lemma 7 and Lemma 9 then imply, that a trial step size $\alpha_{k_i,l} \leq \beta_{k_i}$ satisfies both

$$f(x_{k_i}(\alpha_{k_i,l})) \leq f(x_{k_i}) + \eta_f m_{k_i}(\alpha_{k_i,l}) \tag{49}$$

and

$$\Big(\theta(x_{k_i}(\alpha_{k_i,l})), f(x_{k_i}(\alpha_{k_i,l}))\Big) \notin \mathcal{F}_{k_i}. \tag{50}$$

If we now denote with $\alpha_{k_i,L}$ the first trial step size satisfying both (49) and (50), the backtracking line search procedure in Step 4.5 then implies that for $\alpha \geq \alpha_{k_i,L}$

$$\alpha \geq \tau_1 \beta_{k_i} \overset{(48)}{=} \tau_1 c_6 \theta(x_{k_i}) \overset{(46)}{>} \frac{\delta[\theta(x_{k_i})]^{s_\theta}}{\epsilon_2^{s_f}}$$

and therefore for $\alpha \geq \alpha_{k_i,L}$

$$\delta[\theta(x_{k_i})]^{s_\theta} < \alpha\epsilon_2^{s_f} = [\alpha]^{1-s_f}\,(\alpha\epsilon_2)^{s_f} \overset{(44)}{\leq} [\alpha]^{1-s_f}\,[-m_{k_i}(\alpha)]^{s_f}\,.$$

This means, $\alpha_{k_i,L}$ and all previous trial step sizes are f-step-sizes. Consequently, for all trial step sizes $\alpha_{k_i,l} \geq \alpha_{k_i,L}$, Case I is considered in Step 4.4, and by definition we have $\alpha_{k_i,L} \geq \alpha_{k_i}^{\min}$ (see discussion around Eqn. (16)). Hence, the method does not switch to the feasibility restoration phase in Step 4.2 for those trial step sizes. Therefore, $\alpha_{k_i,L}$ is indeed the accepted step size α_{k_i}. Since it satisfies both (9) and (49), the filter is not augmented in iteration k_i. □

We are now ready to prove the main global convergence result.

Theorem 2 *Suppose Assumptions G hold. Then*

$$\lim_{k\to\infty} \theta(x_k) = 0 \tag{51a}$$

$$\textit{and} \qquad \liminf_{k\to\infty} \chi(x_k) = 0. \tag{51b}$$

In other words, all limit points are feasible, and if $\{x_k\}$ is bounded, than there exists a limit point x_* of $\{x_k\}$ which is a first order optimal point for the equality constrained NLP (1).

Proof. Eqn. (51a) follows from Theorem 1. In order to show (51b), we consider two cases:

i) The filter is augmented only a finite number of times. Then Lemma 8 proves the claim.

ii) There exists a subsequence $\{x_{k_i}\}$, so that $k_i \in \mathcal{A}$ for all i. Now suppose, that $\limsup_i \chi(x_{k_i}) > 0$. Then there exists a subsequence $\{x_{k_{i_j}}\}$ of $\{x_{k_i}\}$ and a constant $\epsilon > 0$, so that $\lim_j \theta(x_{k_{i_j}}) = 0$ and $\chi(x_{k_{i_j}}) > \epsilon$ for all k_{i_j}. Applying Lemma 10 to $\{x_{k_{i_j}}\}$, we see that there is an iteration k_{i_j}, in which the filter is not augmented, i.e. $k_{i_j} \notin \mathcal{A}$. This contradicts the choice of $\{x_{k_i}\}$, so that $\lim_i \chi(x_{k_i}) = 0$, which proves (51b). □

Remark 7 *We do not think that it is possible to obtain a stronger result in Theorem 2 under Assumptions G, such as "$\lim_k \chi(x_k) = 0$." The reason for this is that arbitrarily close to a strict local solution the restoration phase might be invoked even though the search direction is very good. This can happen if the current filter contains information corresponding to previous iterates that lie in a different region of $\mathbb{R}^n$ but had values for θ and f similar to those for the current iterate. For example, if for the current iterate the pair $(\theta(x_k), f(x_k))$ is very close to the current filter (e.g. there exist filter pairs $(\bar\theta, \bar f) \in \mathcal{F}_k$ with $\bar\theta < \theta(x_k)$ and $\bar f \approx f(x_k)$) and the objective function f has to be increased in order to approach the optimal solution, then the trial step sizes can be repeatedly rejected in Step 4.3. In this case, $\alpha_{k,l}$ finally becomes smaller than $\alpha_k^{\min}$ and the restoration phase is triggered. Without making additional assumptions on the restoration phase we only know that the next iterate x_{k+1} returned from the restoration phase is acceptable to the augmented filter, but possibly far away from any KKT point. We believe that it is not possible under the current assumptions to exclude the chance that this situation occures repeatedly, in which case "$\lim_k \chi(x_k) = 0$" would not be valid.*

Remark 8 *It is possible to strengthen the convergence result under stronger assumptions. In addition to Assumptions G, suppose that x_* is a local solution of the NLP (1) satisfying the second order sufficient optimality conditions [17] with optimal multipliers λ_*. Let us further assume that the line search filter method generates multiplier iterates λ_k based on the linearization (3) by choosing $\lambda_{k+1} = \lambda_k + \alpha_k d_k^\lambda$ with*

$$d_k^\lambda := \lambda_k^+ - \lambda_k \tag{52}$$

in each iteration $k \notin \mathcal{R}_{\text{inc}}$. Also, assume that close to x_* the algorithm uses exact second derivatives, i.e. $H_k = \nabla^2 f(x_k) + \sum_i \lambda_k^{(i)} \nabla^2 c(x_k)$. Finally, suppose that, in the neighborhood of (x_*, λ_*), also the algorithm used for the restoration phase is taking steps (d_k, d_k^λ) generated from (3) and (52), where H_k is bounded and satisfies (20). Then, once the iterates (x_k, λ_k) are sufficiently close to (x_*, λ_*), the overall algorithm always takes fractions of steps (d_k, d_k^λ). The assumptions ensure that the KKT error (the norm of the left hand side of (2)) is monotonically decreased and that the iterates are attracted to (x_*, λ_*). As a second order sufficient optimal solution, (x_*, λ_*) is the only root of (2) *in a sufficiently small neighborhood. Therefore we obtain together with Theorem 2 that the entire sequence converges to the solution, once the iterates are sufficiently close.*

One way to construct a restoration phase that satisfies the condition necessary for this result is as follows. Suppose that we have a "rigorous" Algorithm R for the restoration phase, which either converges to a stationary point of the constraint violation or produces an acceptable new iterate for the filter method. If the restoration phase is now invoked at a point where the KKT error is small, then, instead of directly using Algorithm R, we first compute a search direction (d_k, d_k^λ) *from (3) and (52). If the new (intermediate) iterate obtained by taking the full step does not reduce the KKT error by a fixed fraction* $\kappa_R \in (0,1)$*, we switch to Algorithm R. Otherwise we continue taking steps from (3) and (52) (still formally within the restoration phase in Step 8), until finally either a new acceptable iterate* x_{k+1} *is obtained, or the method reverts to Algorithm R.*

4 Alternative Algorithms

4.1 Measures based on the augmented Lagrangian Function

The two measures $f(x)$ and $\theta(x)$ can be considered as the two components of the exact penalty function (7). Another popular choice for a merit function is the *augmented Lagrangian function* (see e.g. [2, 5, 18])

$$\ell_\rho(x,\lambda) := f(x) + \lambda^T c(x) + \frac{\rho}{2} c(x)^T c(x), \tag{53}$$

where λ are multiplier estimates corresponding to the equality constraints (1b). If λ_* is the vector of multipliers corresponding to a strict local solution x_* of the NLP (1), then there exists a penalty parameter $\rho > 0$, so that x_* is a strict local minimizer of $\ell_\rho(x,\lambda_*)$.

In the line search filter method described in Section 2 we can alternatively follow an approach based on the two components $\mathcal{L}(x,\lambda)$ (defined in (4)) and $\theta(x)$ (or equivalently $\theta(x)^2$) of the augmented Lagrangian function rather than based on the components of the exact penalty function. (Recently, Ulbrich [22] proposed a related approach using the Lagrangian function in a trust region filter method, including both global and local convergence results.) In Algorithm I we then replace all occurrences of the measure "$f(x)$" by "$\mathcal{L}(x,\lambda)$." In addition to the iterates x_k we now also keep iterates λ_k as estimates of the equality constraint multipliers, and compute in each iteration k a search direction d_k^λ for those variables. This search direction can be obtained with no additional computational cost from (52) with λ_k^+ from (3). Defining

$$\lambda_k(\alpha_{k,l}) := \lambda_k + \alpha_{k,l} d_k^\lambda, \tag{54}$$

the sufficient reduction criteria (8b) and (11) are then replaced by

$$\mathcal{L}(x_k(\alpha_{k,l}),\lambda_k(\alpha_{k,l})) \leq \mathcal{L}(x_k,\lambda_k) - \gamma_f \theta(x_k) \qquad \text{and} \tag{55a}$$

$$\mathcal{L}(x_k(\alpha_{k,l}),\lambda_k(\alpha_{k,l})) \leq \mathcal{L}(x_k,\lambda_k) + \eta_f m_k(\alpha_{k,l}), \tag{55b}$$

respectively, where the model m_k for $\mathcal{L}$ is now defined as

$$m_k(\alpha) := \alpha g_k^T d_k - \alpha \lambda_k^T c_k + \alpha(1-\alpha) c_k^T d_k^\lambda \tag{56}$$

which is obtained by Taylor expansions of $f(x)$ and $c(x)$ around (x_k, λ_k) into direction (d_k, d_k^λ) and the use of (3).

The f-type switching condition (9) remains unchanged, but the definition of the minimum step size (18) has to be modified to accommodate (55) and (56). The only requirements for this change are again that it is guaranteed that the method does not switch to the feasibility restoration phase in Step 4.2 as long as the f-type switching condition (9) is satisfied for a trial step size $\alpha \le \alpha_{k,l}$, and that the backtracking line search in Step 4 is finite. We also require that the multipliers λ_{k+1} that are used after the restoration phase has been called, are uniformly bounded (e.g. by choosing $\lambda_{k+1} = \lambda_k$ for $k \in \mathcal{R}$).

In order to see that the global convergence analysis in Section 3 still holds, let us briefly revisit the individual results. The first two bounds in Lemma 1 remain valid, so that with

$$\lambda_{k+1} \overset{(54)}{=} \lambda_k + \alpha_k d_k^\lambda \overset{(52)}{=} (1-\alpha_k)\lambda_k + \alpha_k \lambda_k^+$$

we obtain by induction that λ_k, and therefore also d_k^λ are uniformly bounded for all $k \notin \mathcal{R}_{\text{inc}}$. With this, also the last bound in (25) holds, as can be seen from (56). Since λ_k is bounded for all k, we further see that the sequence $\{\mathcal{L}(x_k, \lambda_k)\}$ is bounded below, a property used at several points in the analysis. It is then easy to verify, that Lemma 2 and Lemma 4 are still valid for the model definition (56), since the first equality in (26a) then becomes

$$m_{k_i}(\alpha)/\alpha = g_{k_i}^T d_{k_i} + O(\|c_{k_i}\|),$$

and thus only the constant c_3 in the proof may change. Furthermore, Lemma 3 still holds for the model definition (56) and with the measure "f" replaced by "$\mathcal{L}$", because

$$\begin{aligned}
& \mathcal{L}(x_k + \alpha d_k, \lambda_k + \alpha d_k^\lambda) - \mathcal{L}(x_k, \lambda_k) \\
= \;& f(x_k + \alpha d_k) - f(x_k) + (\lambda_k + \alpha d_k^\lambda)^T c(x_k + \alpha d_k) - \lambda_k^T c(x_k) \\
= \;& \alpha g_k^T d_k + O(\alpha^2 \|d_k\|^2) + (\lambda_k + \alpha d_k^\lambda)^T (c(x_k) + \alpha A_k^T d_k + O(\alpha^2\|d_k\|^2)) - \lambda_k^T c(x_k) \\
\overset{(3)}{=} \;& \alpha g_k^T d_k + (\lambda_k + \alpha d_k^\lambda)^T (1-\alpha) c(x_k) - \lambda_k^T c(x_k) + O(\alpha^2 \|d_k\|^2) \\
\overset{(56)}{=} \;& m_k(\alpha) + O(\alpha^2\|d_k\|).
\end{aligned}$$

Finally, the analysis in Sections 3.3 and 3.4 then holds with replacing "f" by "$\mathcal{L}$" where appropriate. The only point that deserves special attention is the proof of Lemma 8. Here, it is essential that the last rejected trial step size (38) satisfies the f-type switching condition (9), at least for k_i sufficiently large. To see that this is also true for the model definition (56), which is no longer linear in α, let us define the function

$$h_{k_i}(\alpha) := [-m_{k_i}(\alpha)]^{s_f}\ \alpha^{1-s_f} - \delta[\theta(x_{k_i})]^{s_\theta}.$$

This function is well defined for the considered k_i due to (37), and we have $h_{k_i}(\alpha_{k,l}) > 0$ if and only if (9) holds. Since we assume $\lim_i \theta(x_{k_i}) = 0$ and $\chi(x_{k_i}) \ge \epsilon$ in the proof, it can then be shown (using arguments similar to those in the proof of Lemma 2) that $h'_{k_i}(0) \ge \epsilon_3$ for some $\epsilon_3 > 0$ and k_i sufficiently large, and that $h''_{k_i}(0)$ is uniformly bounded. Since $\alpha_{k_i} \to 0$ and $h_{k_i}(\alpha_{k_i}) > 0$, it then follows that the f-type switching condition (9) holds for $\alpha_{k_i,l_i} \in [\alpha_{k_i}/\tau_2, \alpha_{k_i}/\tau_1]$ when k_i is sufficiently large.

4.2 Line Search Filter SQP Methods

In this section we show how Algorithm I can be applied to line search SQP methods for the solution of nonlinear optimization problems of the form

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{57a}$$
$$\text{subject to} \quad c(x) = 0 \tag{57b}$$
$$x \ge 0. \tag{57c}$$

We choose to consider only bounds of the form (57c) to simplify the presentation, but our discussion can easily be adapted to general bound constraints (such as "$x_L \leq x \leq x_U$").

At an iterate x_k, a line search SQP method obtains the search direction d_k as a solution of the quadratic program (QP)

$$\min_{d \in \mathbb{R}^n} \quad g_k^T d + \frac{1}{2} d^T H_k d \tag{58a}$$

$$\text{subject to} \quad A_k^T d + c(x_k) = 0 \tag{58b}$$

$$x_k + d \geq 0. \tag{58c}$$

As before, $g_k := \nabla f(x_k)$, $A_k := \nabla c(x_k)$. Furthermore, now H_k denotes a symmetric matrix approximating the Hessian of the Lagrangian

$$\mathcal{L}(x, \lambda, z) := f(x) + \lambda^T c(x) - v^T x \tag{59}$$

of the NLP (57). The vector $v \geq 0$ stands for the Lagrangian multipliers corresponding to the bound constraints (57c). We denote by λ_k^+ and $v_k^+ \geq 0$ some (not necessarily unique) multipliers corresponding to the QP solution d_k.

In the following analysis, we assume that the particular method for solving (58) is able to ensure that the QP Hessian H_k is positive definite in a certain space (see Assumption (G3*) below), possibly by modifying the matrix H_k. Then a (finite) solution of the QP exists, and the generated search direction d_k is a direction of sufficient decrease for the objective function if the constraint violation is small.

Starting from an initial point $x_0 \geq 0$, Algorithm I can then be used with the following modification:

- The computation of the search direction in Step 3 is replaced by the solution of the QP (58).
- The restoration phase is invoked from Step 3, if the QP (58) is unbounded, infeasible or "not sufficiently consistent" (see Assumption (G4*) below). As before, $\mathcal{R}_{\text{inc}}$ denotes the set of all iteration counters in which the restoration phase is invoked from Step 3.
- The feasibility restoration phase in Step 8 has to return an iterate x_{k+1} that satisfies the bound constraints (57c).

In order to state the assumptions necessary for a global convergence analysis let us define for each $k \notin \mathcal{R}_{\text{inc}}$ the set of coordinates that are active at the current point x_k *and* at $x_k + d_k$,

$$\mathcal{S}_k := \left\{ i \in \{1, \ldots, n\} : x_k^{(i)} = 0 \quad \text{and} \quad d_k^{(i)} = 0 \right\}.$$

For the purpose of the analysis we again consider a decomposition of the search direction

$$d_k = q_k + p_k \tag{60}$$

where q_k is now defined as the solution of the QP

$$\min_{q \in \mathbb{R}^n} \quad q^T q \tag{61a}$$

$$\text{subject to} \quad A_k^T q + c_k = 0 \tag{61b}$$

$$q^{(i)} = 0 \qquad \text{for } i \in \mathcal{S}_k \tag{61c}$$

$$x_k^{(i)} + q^{(i)} \geq 0 \qquad \text{for } i \notin \mathcal{S}_k. \tag{61d}$$

Therefore, q_k is the shortest vector that satisfies the constraints in the QP (58) and stays at those bounds that are active for all trial points (6). We further choose Z_k as an orthonormal null space matrix for the matrix

$$\left[\begin{array}{cccc} A_k & e_{j_1} & \cdots & e_{j_{l_k}} \end{array}\right]^T, \qquad \text{where} \qquad \mathcal{S}_k = \{j_1, \ldots, j_{l_k}\},$$

i.e. Z_k is a basis of the null space for the gradients of all equality constraints and bounds that are active at x_k and $x_k + d_k$. With this, we can compute $p_k = Z_k \bar{p}_k$ with $\bar{p}_k$ as the solution of the reduced QP (see e.g. [17])

$$\min_{\bar{p} \in \mathbb{R}^{n-m-l_k}} \qquad \left(Z_k^T g_k + Z_k^T H_k q_k\right)^T \bar{p} + \frac{1}{2} \bar{p}^T Z_k^T H_k Z_k \bar{p} \tag{62a}$$

$$\text{subject to} \qquad x_k + q_k + Z_k \bar{p} \geq 0. \tag{62b}$$

Note that the set $\mathcal{S}_k$ is not known before the QP (58) has been solved. The QPs (61) and (62) are defined only to state the assumptions below and are not a possible procedure to obtain the search direction d_k.

For the global convergence analysis of the filter line search SQP method we replace Assumptions (G3) and (G4) by

(G3*) *There exists a constant $M_H > 0$, so that for all $k \notin \mathcal{R}_{\text{inc}}$ we have*

$$\lambda_{\min}\left(Z_k^T H_k Z_k\right) \geq M_H, \tag{63}$$

where Z_k has been defined above.

(G4*) *There exists constants $M_q, M_\lambda, M_v > 0$, so that for all $k \notin \mathcal{R}_{\text{inc}}$ we have*

$$\|q_k\| \leq M_q \theta(x_k), \qquad \|\lambda_k^+\| \leq M_\lambda, \qquad \|v_k^+\| \leq M_v.$$

As Assumption (G3) for the original analysis, Assumption (G3*) is necessary to ensure descent in the objective function at points with small infeasibility. In order to ensure this condition, the algorithm could monitor the eigenvalues of the projection of H_k onto the null space of the gradients for all constraints active at x_k and $x_k + d_k$, and perform modifications of H_k if necessary[1]. Note that Assumption (G3*) is natural in the sense that if the method converges to a local solution x_* of the NLP (57) satisfying the strong second order optimality conditions, then the active set $\mathcal{S}_k$ finally becomes unchanged and Z_k is a null space matrix for the constraints active at x_*. Hence, no correction of the reduced Hessian is necessary close to x_*, if exact second derivatives are used and if λ_k converges to λ_*.

Assumption (G4*) is similar in spirit to the assumption expressed by Eqn. (2.10) for the trust region filter SQP method in [7]. Essentially, it implies that eventually the restoration phase is triggered from Step 3 if the constraints of the QP (58) are becoming increasingly degenerate close to feasible points.

It is straight-forward to verify that the proofs in Section 3 still hold under the modified Assumptions G. Only the proof of Lemma 2 requires special attention. Let us first state the KKT

[1] The solution d_k is not known before the QP (58) is solved. One possible way to find a suitable modification of H_k is to solve (58) repeatedly with $H_k = \nabla^2_{xx}\mathcal{L}(x_k, \lambda_k) + \xi I$ for an increasing sequence of modifications $\xi \geq 0$, until the QP is not unbounded and H_k has the required convexity properties expressed in (63).

conditions of the reduced QP (62), which have to be satisfied by the solution d_k and the corresponding multipliers,

$$\begin{aligned} Z_k^T H_k Z_k \bar{p}_k + (Z_k^T g_k + Z_k^T H_k q_k) - Z_k^T v_k^+ &= 0 && \text{(64a)}\\ x_k + q_k + Z_k \bar{p}_k &\geq 0 && \text{(64b)}\\ (x_k + q_k + Z_k \bar{p}_k)^T v_k^+ &= 0 && \text{(64c)}\\ v_k^+ &\geq 0. && \text{(64d)} \end{aligned}$$

For $k \notin \mathcal{R}_{\text{inc}}$ we then have

$$\begin{aligned} Z_k^T g_k &\overset{(64a)}{=} Z_k^T v_k^+ - Z_k^T H_k Z_k \bar{p}_k - Z_k^T H_k q_k \\ (x_k + q_k)^T v_k^+ &\overset{(64c)}{=} -(v_k^+)^T Z_k \bar{p}_k \end{aligned}$$

and therefore

$$\begin{aligned} g_k^T Z_k \bar{p}_k &= -(x_k + q_k)^T v_k^+ - \bar{p}_k^T Z_k^T H_k Z_k \bar{p}_k - \bar{p}_k^T Z_k^T H_k q_k \\ &\overset{(61c),(61d),(64d)}{\leq} -\bar{p}_k^T Z_k^T H_k Z_k \bar{p}_k - \bar{p}_k^T Z_k^T H_k q_k. \end{aligned}$$

This gives together with the modified Assumptions G

$$\begin{aligned} m_k(\alpha)/\alpha &\overset{(10)}{=} g_k^T d_k \overset{(60)}{=} g_k^T Z_k \bar{p}_k + g_k^T q_k \\ &\leq -\bar{p}_k^T Z_k^T H_k Z_k \bar{p}_k - \bar{p}_k^T Z_k^T H_k q_k + g_k^T q_k \\ &\leq -M_H [\chi(x_k)]^2 + O\left(\chi(x_k)\theta(x_k)\right) + O(\theta(x_k)), \end{aligned}$$

which corresponds to (26c). We can conclude the proof of Lemma 2 as before.

4.3 Line Search Filter Interior Point Methods

An alternative to active set methods for handling inequality constraints is offered by *interior point* or *barrier methods*. In this section we demonstrate how the line search filter method presented in Section 2 can be used within an interior point framework. The presented algorithm can be changed in an obvious way if (57c) is generalized to lower and upper bound constraints on all or only some variables.

The barrier method presented here can be of the primal or primal-dual type, and differs from the interior point filter algorithm proposed by Ulbrich et al. [21] in that the barrier parameter is kept constant for several iterations. This enables us to base the acceptance of trial steps directly on the (barrier) objective function instead of only on the norm of the optimality conditions. Therefore the presented method can be expected to be less likely to converge to saddle points or maxima than the algorithm proposed in [21].

A barrier method solves a sequence of *barrier problems*

$$\min_{x \in \mathbb{R}^n} \quad \varphi_\mu(x) := f(x) - \mu \sum_{i=1}^{n} \ln(x^{(i)}) \tag{65a}$$

$$\text{subject to} \quad c(x) = 0 \tag{65b}$$

for a decreasing sequence μ_l of *barrier parameters* with $\lim_l \mu_l = 0$. Local convergence of barrier methods as $\mu \to 0$ has been discussed in detail by other authors, in particular by Nash and Sofer

[16] for primal methods, and by Byrd et al. [4] and Gould et al. [12, 13] for primal-dual methods. In those approaches, the barrier problem (65) is solved to a certain tolerance $\epsilon > 0$ for a fixed value of the barrier parameter μ. The parameter μ is then decreased and the tolerance ϵ is tightened for the next barrier problem. For example, in [12] it is shown that if the parameters μ and ϵ are updated in a particular fashion, the new starting point (enhanced by an extrapolation step with the cost of one regular iteration that tries to follow the path defined by the optimality conditions for (65) as μ changes) eventually solves the next barrier problem well enough in order to satisfy the new tolerance. Then the barrier parameter μ is decreased again immediately (without taking an additional step), leading to a superlinear convergence rate of the overall interior point algorithm for solving the original NLP (57).

Consequently, the step acceptance criterion in the solution procedure for a fixed barrier parameter μ becomes irrelevant as soon as the (extrapolated) starting points are immediately accepted. Until then, we can consider the (approximate) solution of the individual barrier problems as independent procedures, similar to the approach taken by Byrd et al. in [3]. The focus of this paper are the properties of the line search filter approach, and we address therefore only the convergence properties of an algorithm for solving the barrier problem (65) for a *fixed* value of the barrier parameter μ. Some additional comments on the overall interior point method are given in Remark 9 at the end of this section.

The main idea is to apply the technique developed and analyzed in Sections 2 and 3 to solve the barrier problem (65), i.e. we replace all occurrences of "f" by "φ_μ". However, there are two issues that we have to consider:

1. The barrier objective function (65a) is only defined as long as all components of x are strictly within bounds, i.e. $x > 0$;

2. The barrier objective function and its derivatives become unbounded if any of the components of x approaches its bound.

In order to handle the first item, we enforce that all iterates x_k are strictly positive. For this purpose, we assume that the starting point satisfies $x_0 > 0$, and that an iterate returned from the restoration phase satisfies $x_{k+1} > 0$. We further define a maximal step size $\alpha_k^{\max} \in (0, 1]$ using the *fraction-to-the-boundary rule*

$$\alpha_k^{\max} := \max \{\alpha \in (0, 1] : x_k + \alpha d_k \geq (1 - \tau)x_k\} \tag{66}$$

for a fixed parameter $\tau \in (0, 1)$, usually chosen close to 1. With this, we start the backtracking line search in Step 4.1 of Algorithm I from $\alpha_{k,0} = \alpha_k^{\max}$. Then all trial points (6) lie strictly within bounds.

Addressing the second item, we show below that under additional assumptions the iterates generated by the modified Algorithm I (using (66)) are bounded away from the bounds (see Theorem 3), so that in turn the appropriate quantities in analysis of Section 3 are bounded. Here it is necessary to assume that the parameter s_f in the f-type switching condition (9) is chosen to be 1 (see Remark 6).

For later reference, let us restate the linear system (3) in the notation of this section. Recalling that "f" is replaced by "φ_μ", this system can be written as

$$\begin{bmatrix} W_k + \mu X_k^{-2} & A_k \\ A_k^T & 0 \end{bmatrix} \begin{pmatrix} d_k \\ \lambda_k^+ \end{pmatrix} = -\begin{pmatrix} \nabla f(x_k) - \mu X_k^{-1} e \\ c(x_k) \end{pmatrix}, \tag{67}$$

where $X_k := \mathrm{diag}(x_k)$, e is the vector of ones of appropriate dimension, and W_k is (an approximation of) the Hessian of the Lagrangian (59) for the *original* NLP (57). Note that the Hessian H_k in (3) is equal to $W_k + \mu X_k^{-2}$. It is easy to verify that the following arguments also hold if the primal Hessian "μX_k^{-2}" of the log-barrier terms is replaced by the primal-dual Hessian "$\Sigma_k = X_k^{-1} V_k$" (with variables $v_k > 0$), as long as there exists $m_\Sigma > 1$ such that

$$\frac{1}{m_\Sigma}\mu \leq v_k^{(i)} x_k^{(i)} \leq m_\Sigma \mu$$

for all i and k.

Next we state the assumptions necessary to show global convergence for the barrier line search filter algorithm.

Assumptions B. *Given a starting point $x_0 > 0$, let $\{x_k\}$ be the sequence generated by Algorithm I (adapted to the solution of the barrier problem and with $s_f = 1$ in (9)), where we assume that the feasibility restoration phase in Step 8 always terminates successfully with $x_{k+1} > 0$ and that the algorithm does not stop in Step 2 at a KKT point.*

(B1) *There exists an open set $\mathcal{C} \subseteq \mathbb{R}^n$ with $[x_k, x_k + \alpha_k^{\max} d_k] \subseteq \mathcal{C}$ for all $k \notin \mathcal{R}_{\text{inc}}$, so that f and c are differentiable on $\mathcal{C}$, and their function values, as well as their first derivatives, are bounded and Lipschitz-continuous over $\mathcal{C}$.*

(B2) *The matrices W_k approximating the Hessian of the Lagrangian of the* original *NLP (57) used in (67) are uniformly bounded for all $k \notin \mathcal{R}_{\text{inc}}$.*

(B3) *The matrices $H_k = W_k + \mu X_k^{-2}$ are uniformly positive definite on the null space of the Jacobian A_k^T. In other words, there exists a constant $M_H > 0$, so that for all $k \notin \mathcal{R}_{\text{inc}}$*

$$\lambda_{\min}\left(Z_k^T (W_k + \mu X_k^{-2}) Z_k\right) \geq M_H, \tag{68}$$

where the columns of $Z_k \in \mathbb{R}^{n \times (n-m)}$ form an orthonormal basis matrix of the null space of A_k^T.

(B4) *There exists a constant $M_A > 0$, so that for all $k \notin \mathcal{R}_{\text{inc}}$ we have*

$$\sigma_{\min}(A_k) \geq M_A. \tag{69}$$

(B5) *The iterates, for which the restoration phase is invoked from Step 3 (for example, because (68) or (69) are violated), are not arbitrarily close to the feasible region. In other words, there exists a constant $\theta_{\text{inc}} > 0$, so that $k \notin \mathcal{R}_{\text{inc}}$ whenever $\theta(x_k) \leq \theta_{\text{inc}}$*

(B6) *The iterates $\{x_k\}$ are bounded.*

(B7) *At all* feasible *limit points $\bar{x}$ of $\{x_k\}$, the gradients of the active constraints,*

$$\nabla c_1(\bar{x}), \ldots, \nabla c_m(\bar{x}), \qquad \text{and} \qquad e_i \text{ for } i \in \{j : \bar{x}^{(j)} = 0\}, \tag{70}$$

are linearly independent.

(B8) *There exist constants $\tilde{\delta}_\theta, \tilde{\delta}_x > 0$, so that whenever the restoration phase is called in Step 8 in an iteration $k \in \mathcal{R}$ with $\theta(x_k) \leq \tilde{\delta}_\theta$, it returns a new iterate with $x_{k+1}^{(i)} \geq x_k^{(i)}$ for all components satisfying $x_k^{(i)} \leq \tilde{\delta}_x$.*

Assumptions (B1)–(B5) are essentially identical to the original Assumptions (G1)–(G5). Note, however, that the boundedness assumptions in (B1) and (B2) pertain to the original functions and not those from the barrier problem (65), to which the line search filter algorithm is applied. Boundedness of the barrier function φ_μ cannot be assumed, as pointed out above. On the other hand, the lower bound (68) refers to the Hessian used for the step computation (67).

Assumption (B6) is necessary in order to guarantee that the barrier objective function $\varphi_\mu(x)$ is bounded below. Assumption (B7) implies Assumptions (B4) and (B5), if the strategy described at the end of Section 3.1 is used to define when $k \in \mathcal{R}_{\rm inc}$.

Assumption (B7) is considerably less restrictive than the regularity assumptions made for the global convergence analysis of the interior point methods proposed by El-Bakry et al. [6], Yamashita [28] and Yamashita et al. [29]. For those algorithms, it is essentially required that the gradients of all equality constraints and active inequality constraints (70) are linearly independent at *all* limit points, and not only at all *feasible* limit points. If those methods are applied to the example presented by Wächter and Biegler in [25] (which satisfies Assumption (B7)), they converge to a spurious solution that is neither feasible nor a stationary point for any norm of the constraint violation; for details see [25]. In contrast to this, the proposed algorithm is at least guaranteed to converge to a stationary point for the infeasibility (assuming that a reasonable restoration phase algorithm is used), and in practice converges to the solution [24]. We note here that the method presented by Tits et al. in [20] has a similar convergence guarantee as the proposed method, in the sense that the regularity assumption for the constraints in [20] only excludes infeasible limit points, at which there is no feasible descent direction for the constraint violation measured in the ℓ_1 norm.

To see that Assumption (B8) is reasonable, suppose that the gradients of the active constraints (70) are uniformly linearly independent at all feasible points $\bar{x}$ (this is similar to Assumption (B7)). By proof of contradiction one can then show that there exist constants $\tilde{\delta}_\theta, \tilde{\delta}_x > 0$ so that whenever $\theta(x_k) \leq \tilde{\delta}_\theta$ for $k \in \mathcal{R}$, then there exists a *feasible* point $\bar{x}_k$ with $\theta(\bar{x}_k) = 0$, $\bar{x}_k > 0$, and $\bar{x}_k^{(i)} \geq x_k^{(i)}$ for all i with $x_k^{(i)} \leq \tilde{\delta}_x$. The point $\bar{x}_k$ is a candidate for the point x_{k+1} returned from the restoration phase algorithm satisfying the condition in Assumption (B8).

To find an approximation to such a point $\bar{x}_k$, we may apply some algorithm for bound constraint optimization to the problem

$$\min \quad \|c(x)\|_2^2 + \rho\|x - x_k\|_2^2 \tag{71a}$$

$$\text{subject to} \quad x^{(i)} \geq \min\{\epsilon, x_k^{(i)}\} \quad \text{for } i = 1, \ldots, n. \tag{71b}$$

Here, the regularization term weighted by $\rho > 0$ aims to keep the solution of this problem in the neighborhood of x_k, and $\epsilon > 0$ is some small number that we introduce to make sure that the (approximate) solution for this problem is not arbitrarily close to the boundary. In order to find suitable values of ρ and ϵ one might start with some initial choices, and whenever the optimal solution of (71) does not reduce $\theta(x)$ sufficiently in order to be accepted in Step 8 as x_{k+1}, problem (71) could be resolved with smaller values of ρ or ϵ; Ulbrich et al. [21] outline a related restoration phase procedure. From the discussion in the previous paragraph, it is clear that a careful implementation of such a procedure should eventually produce (approximate) solutions $\bar{x}_k$ of (71) for $k \in \mathcal{R}$ that satisfy Assumption (B8)[2].

While a detailed discussion of the restoration phase algorithm is beyond the scope of this paper, we propose in Wächter and Biegler [27] a procedure for the restoration phase which applies the

[2]Recall that we assume here that the restoration phase always terminates successfully. Otherwise, this procedure should produce a limit point that is a local minimizer, or at least a stationary point, for the constraint violation within the bound constraints $x \geq 0$.

interior point filter algorithm recursively to a problem formulation similar to (71) and seems to perform well in practice.

Finally we remark that Assumption (B3) is weaker than the one made in an earlier version of our analysis [24].

The remainder of this section deals the with proof of the following theorem.

Theorem 3 *Suppose Assumptions B hold. Then there exists a constant* ϵ_x*, so that* $x_k \geq \epsilon_x e$ *for all* k*.*

This means that the iterates generated by Algorithm I (for the barrier algorithm) are bounded away from the boundary of the region defined by the bound constraints (57c). Once this is established, one can verify that then Assumptions B imply Assumptions G, and therefore the global convergence results from Section 3 hold. We only point out that Theorem 3 and Lemma 1 together with (66) establishes that the starting step size in the backtracking line search $\alpha_k^{\max}$ is uniformly bounded away from zero, a property necessary in the proofs of Lemmas 7, 8, and 9 (for details see also [24]).

In order to prove Theorem 3 we make use of the following lemma.

Lemma 11 *Suppose Assumptions B hold. Then, for a given subset of indices* $\mathcal{S} \subseteq \{1, \ldots, n\}$ *and a constant* $\delta_l > 0$*, there exist* $\delta_s, \delta_\theta > 0$ *so that* $d_k^{(i)} > 0$ *for* $i \in \mathcal{S}$ *whenever* $k \notin \mathcal{R}$ *and*

$$x_k \in L := \left\{ x \geq 0 : x^{(i)} \leq \delta_s \text{ for } i \in \mathcal{S}, x^{(i)} \geq \delta_l \text{ for } i \notin \mathcal{S}, \theta(x) \leq \delta_\theta \right\},$$

i.e. at sufficiently feasible points, the search direction points away from almost active bounds.

Proof. Let us denote with x_k^s the components of x_k in $\mathcal{S}$, and x_k^l the remaining ones. Without loss of generality we assume $x_k = [(x_k^s)\ (x_k^l)]$; similarly we define A_k^s, A_k^l etc. First, we rewrite the linear system (67) by scaling the first rows and columns by X_k^s,

$$\begin{bmatrix} X_k^s W_k^{ss} X_k^s + \mu I & X_k^s W_k^{sl} & X_k^s A_k^s \\ W_k^{ls} X_k^s & W_k^{ll} + \mu (X_k^l)^{-2} & A_k^l \\ (A_k^s)^T X_k^s & (A_k^l)^T & 0 \end{bmatrix} \begin{pmatrix} \tilde{d}_k^s \\ d_k^l \\ \lambda_k^+ \end{pmatrix} = - \begin{pmatrix} X_k^s g_k^s - \mu e \\ g_k^l - \mu (X_k^l)^{-1} e \\ c(x_k) \end{pmatrix}, \tag{72}$$

where we defined $\tilde{d}_k^s := (X_k^s)^{-1} d_k^s$.

For some initial choice of $\delta_s, \delta_\theta > 0$, let $\bar{x} \in L$ be a feasible point with $\bar{x}^s = 0$. We then have from Assumption (B7), that the columns of the matrix

$$\begin{bmatrix} [\nabla c(\bar{x})]^s & I \\ [\nabla c(\bar{x})]^l & 0 \end{bmatrix},$$

and therefore also the columns of $[\nabla c(\bar{x})]^l$, are linearly independent. Using a compactness argument and Assumption (B6), we can find a constant $m_\sigma > 0$, so that $\sigma_{\min}([\nabla c(\bar{x})]^l) \geq m_\sigma$ for all feasible limit points $\bar{x} \in L$ of $\{x_k\}$ with $\bar{x}^s = 0$. Therefore, we have from Assumption (B1) that

$$\sigma_{\min}(A_k^l) \geq \frac{m_\sigma}{2}$$

for all $x_k \in L$, if δ_θ and δ_s are chosen sufficiently small.

In addition, possibly after further decreasing δ_θ, it follows from Assumptions (B3) and (B5) that for all $x_k \in L$ the projection of $W_k^{ll} + \mu (X_k^l)^{-2}$ into the null space of $(A_k^l)^T$ is uniformly positive definite.

Together with the boundedness assumptions (B1) and (B2) we then see that (72) satisfies

$$\left(\begin{bmatrix} \mu I & 0 & 0 \\ 0 & W_k^{ll} + \mu(X_k^l)^{-2} & A_k^l \\ 0 & (A_k^l)^T & 0 \end{bmatrix} + O(\delta_s)\right)\begin{pmatrix} \tilde{d}_k^s \\ d_k^l \\ \lambda_k^+ \end{pmatrix} = -\begin{pmatrix} -\mu e \\ g_k^l - \mu(X_k^l)^{-1}e \\ c(x_k) \end{pmatrix} + O(\delta_s),$$

for $x_k \in L$, where the inverse of the matrix in the square brackets, as well as the right hand side, are uniformly bounded for δ_s sufficiently small. Therefore, for $x_k \in L$, we have that $\tilde{d}_k^s = e + O(\delta_s)$, and consequently $\tilde{d}_k^s > 0$, after possibly reducing δ_s even more. The claim then follows from $d_k^s = X_k^s \tilde{d}_k^s$. □

We finish with the proof of Theorem 3.

Proof. (of Theorem 3) We first show by contradiction, that there exist constants $\delta_x, \delta_\theta > 0$, so that $d_k^{(i)} > 0$ for all indices i with $x_k^{(i)} \leq \delta_x$ whenever $\theta(x_k) \leq \delta_\theta$ and $k \notin \mathcal{R}$.

Suppose this claim is not true. Then, there exists a subsequence $\{x_{k_j}\}$ of iterates with $k_j \notin \mathcal{R}$, $\lim_j \theta(x_{k_j}) = 0$ and $\lim_j x_{k_j}^{(s)} = 0$ for some index s, as well as $d_{k_j}^{(s)} \leq 0$ for all j. Let $\bar{x}$ be a limit point of $\{x_{k_j}\}$, and define $\mathcal{S} := \{i : \bar{x}^{(i)} = 0\}$ and $\delta_l := \min\{\bar{x}^{(i)}/2 : i \notin \mathcal{S}\} > 0$. Applying Lemma 11 we can conclude that $d_{k_j}^{(s)} > 0$ (since $s \in \mathcal{S}$) for j sufficiently large, in contradiction to the definition of the subsequence.

Since the filter mechanisms ensure $\lim_k \theta(x_k) = 0$ (even if the barrier objective function is unbounded above; see Remark 6), we can find K so that $\theta(x_k) \leq \min\{\delta_\theta, \tilde{\delta}_\theta\}$ for $k \geq K$ (recall the definition of $\tilde{\delta}_\theta$ and $\tilde{\delta}_x$ in Assumption (B8)). Define

$$\epsilon_x := \min\left\{(1-\tau)\min\{\delta_x, \tilde{\delta}_x\}, \min_i\{x_k^{(i)} : k \leq K\}\right\} > 0.$$

By definition it is clear that $x_k \geq \epsilon_x e$ for $k \leq K$, which can be used as the anchor for a proof by induction. Now suppose that $x_k \geq \epsilon_x e$ for some $k \geq K$. Since $d_k^{(i)} > 0$ for $x_k^{(i)} \leq \delta_x$ for $k \notin \mathcal{R}$, as well as from Assumption (B8), we see that we can only have $x_{k+1}^{(i)} < x_k^{(i)}$ for an index i if $x_k^{(i)} \geq \min\{\delta_x, \tilde{\delta}_x\}$. From (66) we then obtain $x_{k+1}^{(i)} \geq (1-\tau)x_k^{(i)} \geq (1-\tau)\min\{\delta_x, \tilde{\delta}_x\}$, so that overall $x_{k+1} \geq \epsilon_x e$. □

Remark 9 *For the overall barrier method as the barrier parameter μ is driven to zero, we may simply re-start Algorithm I by deleting the current filter whenever the barrier parameter changes. Alternatively, we may choose to store the values of the two terms $f(x_l)$ and $\sum_i \ln(x_l^{(i)})$ in the barrier function $\varphi_\mu(x_l)$ separately for each corner entry (14) in the filter, which would allow one to initialize the filter for the new barrier problem under consideration of already known information. Details on such a procedure are beyond the scope of this paper.*

5 Conclusions

A framework for line search filter methods that can be applied to barrier methods and active set SQP methods has been presented. Global convergence has been shown under mild assumptions, which are, in particular, less restrictive than those made previously for some line search interior point methods. The method also possesses favorable local convergence behavior, as we discuss in the companion paper [26]. We further proposed an alternative measure for the filter, using the Lagrangian function instead of the objective function, for which the global convergence properties still hold.

In a recent report [27] we present practical experience with the line search filter barrier method proposed in this paper. The numerical results on a large set of test problems show that the algorithm exhibits very good practical performance in terms of efficiency and robustness, and that it is competitive with other current NLP codes.

Acknowledgements.

Funding from the National Science Foundation (DMS-9706950) and by an unrestricted grant from the ExxonMobil Corporation is gratefully acknowledged. We are very indebted to Jorge Nocedal for intriguing discussions, helpful comments and encouragement regarding this paper. We thank Caroline Sainvitu and Philippe Toint for their careful study of the manuscript and for catching several typos. Last but not least, we are also very grateful to several referees for their insightful comments and suggestions, which greatly helped to improve the presentation of this paper.

References

[1] C. Audet and J. E. Dennis Jr. A pattern search filter method for nonlinear programming without derivatives. *SIAM Journal on Optimization*, 14(4):980–1010, 2004.

[2] Dimitri P. Bertsekas. *Constrainted Optimization and Lagrange Multiplier Methods*. Academic Press, New York, 1982.

[3] R. H. Byrd, J. Ch. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89:149–185, 2000.

[4] R. H. Byrd, G. Liu, and J. Nocedal. On the local behavior of an interior point method for nonlinear programming. In D. F. Griffiths and D. J. Higham, editors, *Numerical Analysis 1997*, pages 37–56. Addison–Wesley Longman, Reading, MA, USA, 1997.

[5] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. A globally convergent Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, 28(2):545–572, 1991.

[6] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang. On the formulation and theory of the Newton interior-point method for nonlinear programming. *Journal of Optimization Theory and Application*, 89(3):507–541, 1996.

[7] R. Fletcher, N. I. M. Gould, S. Leyffer, Ph. L. Toint, and A. Wächter. Global convergence of a trust-region SQP-filter algorithms for general nonlinear programming. *SIAM Journal on Optimization*, 13(3):635–659, 2002.

[8] R. Fletcher and S. Leyffer. A bundle filter method for nonsmooth nonlinear optimization. Technical Report NA/195, Department of Mathematics, University of Dundee, Scotland, December 1999.

[9] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming*, 91(2):239–269, 2002.

[10] R. Fletcher, S. Leyffer, and Ph. L. Toint. On the global convergence of a filter-SQP algorithm. *SIAM Journal on Optimization*, 13(1):44–59, 2002.

[11] D. M. Gay, M. L. Overton, and M. H. Wright. A primal-dual interior method for nonconvex nonlinear programming. In Y. Yuan, editor, *Advances in Nonlinear Programming*, pages 31–56. Kluwer Academic Publishers, Dordretch, 1998.

[12] N. I. M. Gould, D. Orban, A. Sartenaer, and Ph. L. Toint. Superlinear convergence of primal-dual interior point algorithms for nonlinear programming. *SIAM Journal on Optimization*, 11(4):974–1002, 2001.

[13] N. I. M. Gould, D. Orban, A. Sartenaer, and Ph. L. Toint. Componentwise fast convergence in the solution of full-rank systems of nonlinear equations. *Mathematical Programming*, 92(3):481 – 508, 2002.

[14] S.-P. Han. A globally convergent method for nonlinear programming. *Journal of Optimization Theory and Application*, 22:297–309, 1977.

[15] M. Marazzi and J. Nocedal. Feasibility control in nonlinear optimization. In A. DeVore, A. Iserles, and E. Suli, editors, *Foundations of Computational Mathematics*, volume 284 of *London Mathematical Society Lecture Note Series*, pages 125–154. Cambridge University Press, 2001.

[16] S. G. Nash and A. Sofer. Why extrapolation helps barrier methods. Technical report, Operations Research and Engineering Department, George Mason University, Fairfax, A 22030, USA, September 1998.

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

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

[19] M. J. D. Powell. A hybrid method for nonlinear equations. In P. Rabinowitz, editor, *Numerical Methods for Nonlinear Algebraic Equations*, pages 87–114. Gordon and Breach Science Publishers, London, New York, Paris, 1970.

[20] A. L. Tits, A. Wächter, S. Bakhtiari, T. J. Urban, and C. T. Lawrence. A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties. *SIAM Journal on Optimization*, 14(1):173–199, 2003.

[21] M. Ulbrich, S. Ulbrich, and L. N. Vicente. A globally convergent primal-dual interior-point filter method for nonlinear programming. *Mathematical Programming*, 100(2):379 – 410, 2004.

[22] S. Ulbrich. On the superlinear local convergence of a filter-SQP method. *Mathematical Programming*, 100(1):217–245, 2004.

[23] R. J. Vanderbei and D. F. Shanno. An interior-point algorithm for nonconvex nonlinear programming. *Computational Optimization and Applications*, 13:231–252, 1999.

[24] A. Wächter. *An Interior Point Algorithm for Large-Scale Nonlinear Optimization with Applications in Process Engineering.* PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA, January 2002.

[25] A. Wächter and L. T. Biegler. Failure of global convergence for a class of interior point methods for nonlinear programming. *Mathematical Programming*, 88(2):565–574, 2000.

[26] A. Wächter and L. T. Biegler. Line search filter methods for nonlinear programming: Local convergence. Technical report, Department of Chemical Engineering, Carnegie Mellon University, 2003.

[27] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Technical Report RC 23149, IBM T.J. Watson Research Center, Yorktown Heights, USA, March 2004. To appear in Mathematical Programming.

[28] H. Yamashita. A globally convergent primal-dual interior-point method for constrained optimization. *Optimization Methods and Software*, 10:443–469, 1998.

[29] H. Yamashita, H. Yabe, and T. Tanabe. A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization. Technical report, Mathematical System Institute, Inc., Tokyo, Japan, July 1997. Revised July 1998.