RC23036 (W0304-181) April 21, 2003
Computer Science

IBM Research Report

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

Andreas Wächter, Lorenz T. Biegler
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

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

Andreas Wächter* and Lorenz T. Biegler†

April 17, 2003

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 [8] have 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 [8] are encouraging, and subsequently global convergence results for related algorithms were established [6, 9]. Other researchers have also proposed global convergence results for different trust region based filter methods, such as for an interior point (IP) approach [20], a bundle method for non-smooth optimization [7], and a pattern search algorithm for derivative-free optimization [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 IP methods. The motivation given by Fletcher and Leyffer for the development of the filter method [8] is to avoid the necessity to determine a suitable value of the penalty parameter in the merit function. In addition, assuming that Newton directions are usually "good" directions (in particular if exact second derivative information is used) filter methods have the potential to be more efficient than algorithms based on merit functions, as they generally accept larger steps. However, 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

*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

are not even critical points for a measure of infeasibility, if the gradients of the constraints become linearly dependent at non-feasible points. In [17], Powell gives an example for this behavior. More recently, the authors demonstrated another global convergence problem for many line search IP methods on a simple well-posed example [24]. 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 measure of infeasibility (see also [14] 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 towards a solution of the problem, the proposed filter method reverts to a feasibility restoration phase, whose goal is to deliver a new iterate that is at least sufficiently less infeasible. As a consequence, the global convergence problems described above cannot occur.

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

In Section 2 we will 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. [6]. An important difference, however, lies in the condition that determines when to switch between certain sufficient decrease criteria; this modification allows us to show fast local convergence of the proposed line search filter method in the companion paper [26]. We will 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 will 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.

Notation. We will denote the i-th component of a vector $v \in \mathbb{R}^n$ by $v^{(i)}$. Norms $\|\cdot\|$ will denote a fixed vector norm and its compatible matrix norm unless otherwise noted. For brevity, we will use the convention $(x,\lambda) = (x^T,\lambda^T)^T$ for vectors x,λ. For a matrix A, we will 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 will 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 will 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

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

$$\text{s.t.} \qquad c(x) = 0 \tag{1b}$$

where the objective function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and the equality constraints $c : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ with $m < n$ are sufficiently smooth. We will 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 first order optimality conditions, or Karush-Kuhn-Tucker (KKT) conditions, (see e.g. [16]) 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). 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 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, $A_k := A(x_k)$, $g_k := g(x_k)$, $c_k := c(x_k)$, and 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 it, where λ_k is some estimate of the optimal multipliers corresponding to the equality constraints (1b). 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 will assume that the projection of the Hessian approximation H_k onto the null space of the constraint Jacobian is sufficiently 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 [13]

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

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

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

Under certain regularity assumptions it can be shown that a 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 [13].

In order to avoid the determination of an appropriate value of the penalty parameter ρ, Fletcher and Leyffer [8] proposed the concept of a *filter method* in the context of a trust region SQP algorithm. In the remainder of this section we will 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 will 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 will outline in the following; these are closely related to those proposed in [6]. (The overall line search filter algorithm is formally stated on page 7.)

1. *Sufficient Reduction.* Line search methods that use a merit function ensure *sufficient* progress towards the solution. For example, they may do so by enforcing an Armijo condition for the exact penalty function (7) (see e.g. [16]). Here, we borrow the idea from [6, 9] 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

$$\theta(x_k(\alpha_{k,l})) \leq (1 - \gamma_\theta)\theta(x_k) \tag{8a}$$

$$\text{or} \quad f(x_k(\alpha_{k,l})) \leq f(x_k) - \gamma_f \theta(x_k). \tag{8b}$$

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 converge 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 *switching condition*

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

holds with fixed constants $\delta > 0, s_\theta > 1, s_\varphi > 2s_\theta$, where

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

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

If the switching condition (9) holds, instead of insisting on (8), we require that an Armijo-type condition for the objective function,

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

is satisfied (see [6]). 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 switching condition (9) may no longer be valid, so that the method reverts to the acceptance criterion (8).

The switching condition (9) deserves some discussion. On the one hand, for global convergence we need to ensure that close to a feasible but non-optimal point $\bar{x}$ a new iterate indeed leads to progress in the objective function (and not only the infeasibility measure). Lemma 2 below will show that $m_k(\alpha) \leq -\alpha\epsilon$ for some $\epsilon > 0$ and all $\alpha \in (0, 1]$ for iterates x_k in a neighborhood of $\bar{x}$. Therefore, the switching condition is satisfied, if $\alpha_{k,l} > (\delta/\epsilon^{s_f})[\theta(x_k)]^{s_\theta}$. The fact that the right hand side is $o(\theta(x_k))$ allows us to show in Lemma 10 that sufficient decrease in the objective function (11) is indeed obtained by the new iterate close to $\bar{x}$. On the other hand, in order to show that full steps are taken in the neighborhood of a strict local solution x_* we need to ensure that then the Armijo condition (11) is only enforced (i.e. the switching condition is only true) if the progress predicted by the linear model m_k is large enough so that the full step, possibly improved by a second order correction step, is accepted. This is shown in Lemma 4 in the companion paper [26], and it is crucial for its proof that the switching condition with $\alpha_{k,0} = 1$ implies $\theta(x_k) = O(\|d_k\|^{\frac{s_f}{s_\theta}}) = o(\|d_k\|^2)$. Note that the switching conditions used in [6, 9] do not imply this latter relationship.

2. *Filter as taboo-region.* It is also necessary to avoid cycling. For example, this 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 [8] propose to 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 [6, 8], for the sake of a simplified notation we will 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 will 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 will 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 [6]). 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"

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

(see [6]).

It remains to decide which iterations should augment the filter. Since one motivation of the filter method is to make it generally less conservative than a penalty function approach, we do not want to augment the filter in every iteration. In addition, as we will 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 [6] is motivated by these considerations.

We will always augment the filter if for the accepted trial step size $\alpha_{k,l}$ the switching condition (9) or the Armijo condition (11) do not hold. Otherwise, if the filter is not augmented, the value of the objective function is strictly decreased (see Eq. (29) below). To see that this indeed prevents cycling let us assume for a moment that the algorithms 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.

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} merely by decreasing the constraint violation θ, so that x_{k+1} satisfies (8) and is also acceptable to the current filter. In this paper, we do not specify the particular procedure for this feasibility restoration phase. It could be any iterative algorithm for decreasing θ, possibly ignoring the objective function, and different methods could even be used at different stages of the optimization procedure.

Since we will make sure that a feasible iterate is never included in the filter, the algorithm for the feasibility restoration phase usually should be able to find a new acceptable iterate unless it converges to a stationary point of θ. The latter case may be important information for the user, as it indicates that the problem seems (at least locally) infeasible. 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.

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 switching condition (9) holds for some $\alpha \leq \alpha_{k,l}$. In this case, we will 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 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

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

However, if the 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

$$\theta(x_k + \alpha d_k) = \theta(x_k) - \alpha\theta(x_k) + O(\alpha^2) \quad (\text{since } A_k^T d_k + c(x_k) = 0) \tag{16a}$$

$$f(x_k + \alpha d_k) = f(x_k) + \alpha g_k^T d_k + O(\alpha^2) \tag{16b}$$

of the two measures. This predicts that the sufficient decrease condition for the infeasibility measure (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

$$\alpha_k^{\min} := \gamma_\alpha \cdot \begin{cases} \min\left\{\gamma_\theta, \frac{\gamma_f \theta(x_k)}{-g_k^T d_k}, \frac{\delta[\theta(x_k)]^{s_\theta}}{[-g_k^T d_k]^{s_f}}\right\} & \\ & \text{if } g_k^T d_k < 0 \\ \gamma_\theta & \text{otherwise} \end{cases} \tag{17}$$

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 (16) 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 $\theta(x_k)$ is not too 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.

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 $\theta_{\max} \in (\theta(x_0), \infty]$; $\gamma_\theta, \gamma_f \in (0,1)$; $\delta > 0$; $\gamma_\alpha \in (0,1]$; $s_\theta > 1$; $s_f > 2s_\theta$; $0 < \tau_1 \leq \tau_2 < 1$.

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

2. *Check convergence.* Stop, if x_k is a local solution (or at least 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 (almost) singular, go to 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 (17), 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. The switching condition (9) holds:* If the Armijo condition for the objective function (11) holds, accept the trial step and go to Step 5.
Otherwise, go to Step 4.5.

4.4.2. *Case II. The switching condition (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 one of the conditions (9) or (11) does not hold, augment the filter according to (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 according to (13) (for x_k) and continue with the regular iteration in Step 7.

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 (17) that* $\alpha_k^{\min} > 0$*. Therefore, the algorithm will either accept a new iterate in Step 4.4, or switch 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 *implies that* $g_k^T d_k < 0$ *(see e.g. Lemma 4). Therefore,* $\alpha_k^{\min} = 0$*, and the Armijo condition (11) is satisfied for a sufficiently small step size* $\alpha_{k,l}$*, i.e. a new iterate will be accepted in Step 4.4.1. Overall, we see that the inner loop in Step 4 will always terminate after a finite number of trial steps, and the algorithm is well-defined.*

Remark 2 *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.*

$$(\bar{\theta}, \bar{f}) \notin \mathcal{F}_k \quad \Longrightarrow \quad (\theta, f) \notin \mathcal{F}_k \text{ if } \theta \leq \bar{\theta} \text{ and } f \leq \bar{f}. \tag{18}$$

Remark 3 *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 switching rule (i.e. Step 4.4) so that (11) only has to be satisfied whenever* $\theta(x_k) \leq \theta_{\text{sml}}$ *for some* $\theta_{\text{sml}} > 0$ *without affecting the convergence properties of the method [23].*

Remark 4 *The proposed method has many similarities with the trust region filter SQP method proposed and analyzed in [6]. As pointed out above, we chose a modified 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 [6]* *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 had to be defined by means of a minimal step size (17). Due to these differences, the global convergence analysis presented in [6] does not apply to the proposed line search filter method.*

Notation. In the remainder of this paper we will denote the set of indices of those iterations, in which the filter has been augmented according to (13), 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 $\mathcal{R} \subseteq \mathbb{N}$ will be 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 will denote with $\mathcal{R}_{\text{inc}} \subseteq \mathcal{R}$ the set of those iteration counters, in which the linear system (3) is too ill-conditioned or singular, so that the restoration phase is invoked from Step 3. For the active set SQP approach discussed later in Section 4.2 these iterates will correspond to those where the QP is not (sufficiently) consistent.

3 Global Convergence

3.1 Assumptions and Preliminary Results

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

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 first-order optimal 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) *There exists a constant θ_{inc}, so that $k \notin \mathcal{R}_{\text{inc}}$ whenever $\theta(x_k) \leq \theta_{\text{inc}}$, i.e. the linear system (3) is "sufficiently consistent" and the restoration phase is not invoked from Step 3 close to feasible points.*

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

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

(G5) *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 $k \notin \mathcal{R}_{\text{inc}}$*

$$\lambda_{\min}\left(Z_k^T H_k Z_k\right) \geq M_H, \tag{19}$$

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.

Assumptions (G1) and (G2) merely establish smoothness and boundedness of the problem data. As we will see later in Lemma 2, Assumption (G5) ensures a certain descent property and it is similar to common assumptions on the reduced Hessian in SQP line search methods (see e.g. [16]). To guarantee this requirement in a practical implementation, one could compute a QR-factorization of A_k to obtain matrices $Z_k \in \mathbb{R}^{n\times(n-m)}$ and $Y_k \in \mathbb{R}^{n\times m}$ so that the columns of $[Z_k \; Y_k]$ 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. [10]). Then, the overall scaled search direction can be decomposed into two orthogonal components,

$$d_k = q_k + p_k, \qquad \text{where} \tag{20a}$$

$$q_k := Y_k\bar{q}_k \quad \text{and} \quad p_k := Z_k\bar{p}_k, \tag{20b}$$

with

$$\bar{q}_k := -\left[A_k^T Y_k\right]^{-1} c_k \tag{21a}$$

$$\bar{p}_k := -\left[Z_k^T H_k Z_k\right]^{-1} Z_k^T \left(g_k + H_k q_k\right) \tag{21b}$$

(see e.g. [16]). The eigenvalues for the reduced scaled Hessian in (21b) (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. [22]). Note, on the other hand, that (19) holds in the neighborhood of a local solution x_* satisfying the sufficient second order optimality conditions (see e.g. [16]), if H_k approaches the exact Hessian of the Lagrangian of the NLP (1). Then, close to x_*, no eigenvalue correction will be necessary and fast local convergence can be expected, assuming that full steps will be taken close to x_*.

The regularity requirement (G4) ensures that, whenever the gradients of the constraints become (nearly) linearly dependent, the method has to switch to the feasibility restoration phase in Step 3. In practice one could monitor the singular values of $Y_k^T A_k$ in (21a), which are identical to the singular values of A_k, as a criterion when to switch to the restoration phase in Step 3.

We can replace Assumptions (G3) and (G4) by the following assumption.

(G4*) *At all* feasible *points x the gradients of the constraints A_k are linearly independent.*

If (G4*) holds, there exists constants $b_1, b_2 > 0$, so that

$$\theta(x_k) \leq b_1 \qquad \Longrightarrow \qquad \sigma_{\min}(A_k) \geq b_2$$

due to the continuity of $\sigma_{\min}(A(x))$ as a function of x and the boundedness of the iterates. If we now decide to invoke the feasibility restoration phase in Step 3 whenever $\sigma_{\min}(A_k) \leq b_3\theta(x_k)$ for some fixed constant $b_3 > 0$, then Assumptions (G3) and (G4) hold.

Similar to the analysis in [6], we will 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 first order optimal 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 will define the criticality measure for iterations $k \notin \mathcal{R}_{\text{inc}}$ as

$$\chi(x_k) := \|\bar{p}_k\|_2\,, \tag{22}$$

with $\bar{p}_k$ from (21b). Note that this definition is unique, since p_k in (20a) 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 may 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_*. From Assumption (G3) we then have $k_i \notin \mathcal{R}_{\rm inc}$ for i sufficiently large. Furthermore, from Assumption (G4) and (21a) we have $\lim_i \bar{q}_{k_i} = 0$, and then from $\lim_i \chi(x_{k_i}) = 0$, (22), (21b), and Assumption (G5) 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. [16]).

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*

$$\|d_k\| \le M_d, \qquad \|\lambda_k^+\| \le M_\lambda, \qquad |m_k(\alpha)| \le M_m \alpha \tag{23}$$

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), (G4), and (G5) 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. Then the following statement is true:*

If $\{x_{k_i}\}$ is a subset 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 (G3), 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 (21a) and Assumption (G4)) it follows that for $k_i \notin \mathcal{R}_{\rm inc}$

$$\begin{aligned}
m_{k_i}(\alpha)/\alpha &= g_{k_i}^T d_{k_i} \overset{(20)}{=} g_{k_i}^T Z_{k_i} \bar{p}_{k_i} + g_{k_i}^T q_{k_i} \\
&\overset{(21b)}{=} -\bar{p}_{k_i}^T \left[Z_{k_i}^T H_{k_i} Z_{k_i} \right] \bar{p}_{k_i} - \bar{p}_{k_i}^T Z_{k_i}^T H_{k_i} q_{k_i} + g_{k_i}^T q_{k_i} \\
&\overset{(G2),(G5)}{\le} -c_1 \|\bar{p}_{k_i}\|_2^2 + c_2 \|\bar{p}_{k_i}\|_2 \|c_{k_i}\| + c_3 \|c_{k_i}\| \\
&\le \chi(x_{k_i}) \left(-\epsilon\, c_1 + c_2 \theta(x_{k_i}) + \frac{c_3}{\epsilon} \theta(x_{k_i}) \right)
\end{aligned} \tag{24}$$

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

$$m_{k_i}(\alpha) \le -\alpha \frac{\epsilon\, c_1}{2} \chi(x_{k_i}) \le -\alpha \frac{\epsilon^2\, c_1}{2} =: -\alpha \epsilon_2.$$

□

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* $\alpha \leq 1$

$$|\theta(x_k + \alpha d_k) - (1-\alpha)\theta(x_k)| \leq C_\theta \alpha^2 \|d_k\|^2 \tag{25a}$$
$$|f(x_k + \alpha d_k) - f(x_k) - m_k(\alpha)| \leq C_f \alpha^2 \|d_k\|^2. \tag{25b}$$

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*

$$\theta(x_k) = 0 \implies m_k(\alpha) < 0 \qquad \textit{and} \tag{26}$$

$$\Theta_k := \min\{\theta : (\theta, f) \in \mathcal{F}_k\} > 0 \tag{27}$$

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

Proof. If $\theta(x_k) = 0$, we have from Assumption (G3) 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 (20), it follows as in (24) that

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

i.e. (26) holds.

The proof of (27) 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 $\theta(x_k) = 0$, Lemma 2 applied to the singleton $\{x_k\}$ implies that $m_k(\alpha) < 0$ for all $\alpha \in (0, 1]$, so that the 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, $\Theta_{k+1} = \Theta_k > 0$. □

3.2 Feasibility

In this section we will 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*

$$\lim_{k\to\infty} \theta(x_k) = 0. \tag{28}$$

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. 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 $1 - 1/s_f > 0$)

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

which implies

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\overset{(11)}{\leq} \eta_f m_k(\alpha_k) \\ &\overset{(9)}{<} -\eta_f \delta^{\frac{1}{s_f}} [\alpha_k]^{1 - \frac{1}{s_f}} [\theta(x_k)]^{\frac{s_\theta}{s_f}} \\ &< -\eta_f \delta^{\frac{1}{s_f}} c_4 [\theta(x_k)]^{s_\theta}. \end{aligned} \tag{29}$$

Hence, for all $i = 1, 2, \ldots,$

$$\begin{aligned} f(x_{K+i}) &= f(x_K) + \sum_{k=K}^{K+i-1} (f(x_{k+1}) - f(x_k)) \\ &< f(x_K) - \eta_f \delta^{\frac{1}{s_f}} c_4 \sum_{k=K}^{K+i-1} [\theta(x_k)]^{s_\theta}. \end{aligned}$$

Since $f(x_{K+i})$ is bounded below, the series on the right hand side in the last line is bounded, which in turn implies (28). □

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 (50) in Section 4.1.

The following lemma considers a subsequence $\{x_{k_i}\}$ with $k_i \in \mathcal{A}$ for all i. Its proof can be found in [6, Lemma 3.3].

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

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

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
$$\theta(x_{k_i}) \geq M/2 \tag{30}$$
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 (30), 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 (29) holds for all $k = k_i, \ldots, l_i - 1 \notin \mathcal{A}$, we obtain for all i

$$f(x_{l_i}) \leq f(x_{(k_i+1)}) < f(x_{k_i}) - \eta_f \delta^{\frac{1}{s_f}} c_4 [M/2]^{s_\theta}. \tag{31}$$

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 (31) would imply
$$f(x_{k_{(i+1)}}) < f(x_{l_i}) < f(x_{k_i}) - \eta_f \delta^{\frac{1}{s_f}} c_4 [M/2]^{s_\theta}$$
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

$$f(x_{k_{(i_j+1)}}) \geq f(x_{l_{i_j}}). \tag{32}$$

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 (32) and the filter update rule (13), that

$$\theta(x_{k_{(i_j+1)}}) \leq (1-\gamma_\theta)\theta(x_{l_{i_j}}). \tag{33}$$

Since $l_{i_j} \in \mathcal{A}$ for all j, Lemma 6 yields $\lim_j \theta(x_{l_{i_j}}) = 0$, so that from (33) we obtain $\lim_j \theta(x_{k_{i_j}}) = 0$ in contradiction to (30). □

Remark 5 *It is easy to verify that the previous theorem is also valid if we do not assume that $\{f(x_k)\}$ is bounded above. This will be important for the discussion of the interior point method in Section 4.3.*

3.3 Optimality

In this section we will 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 $\alpha \leq \bar{\alpha}$*

$$f(x_{k_i} + \alpha d_{k_i}) - f(x_{k_i}) \leq \eta_f m_{k_i}(\alpha). \tag{34}$$

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

$$\begin{aligned} & f(x_{k_i} + \alpha d_{k_i}) - f(x_{k_i}) - m_{k_i}(\alpha) \\ \overset{(25b)}{\leq} \; & C_f \alpha^2 \|d_{k_i}\|^2 \leq \alpha(1-\eta_f)\epsilon_2 \\ \leq \; & -(1-\eta_f) m_{k_i}(\alpha), \end{aligned}$$

which implies (34). □

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 (28) 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

$$m_{k_i}(\alpha) \leq -\alpha\epsilon_2 \quad \text{for all} \quad \alpha \in (0,1]. \tag{35}$$

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 (29)) 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 $\alpha_{k_i,l_i} \in [\alpha_{k_i}/\tau_2, \alpha_{k_i}/\tau_1]$ 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

$$f(x_{k_i} + \alpha_{k_i,l_i} d_{k_i}) - f(x_{k_i}) > \eta_f m_{k_i}(\alpha_{k_i,l_i}), \tag{36}$$

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

$$(\theta(x_{k_i} + \alpha_{k_i,l_i} d_{k_i}), f(x_{k_i} + \alpha_{k_i,l_i} d_{k_i})) \in \mathcal{F}_{k_i} = \mathcal{F}_K. \tag{37}$$

We will conclude the proof by showing that neither (36) nor (37) can be true for sufficiently large k_i.

To (36): Since $\lim_i \alpha_{k_i} = 0$, we also have $\lim_i \alpha_{k_i,l_i} = 0$. In particular, for sufficiently large k_i we have $\alpha_{k_i,l_i} \leq \bar{\alpha}$ with $\bar{\alpha}$ from Lemma 7, i.e. (36) cannot be satisfied for those k_i.

To (37): 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

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

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 (37). □

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* $k_i \notin \mathcal{R}_{\rm inc}$ *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 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* $\alpha \leq \min\{c_5, c_6\theta(x_{k_i})\}$*.*

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

$$(\theta(x_{k_i}), f(x_{k_i})) \notin \mathcal{F}_{k_i}. \tag{38}$$

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

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

and it follows with (25b) that

$$f(x_{k_i} + \alpha d_{k_i}) \leq f(x_{k_i}). \tag{39}$$

Similarly, for $\alpha \leq c_6\theta(x_{k_i}) \leq \frac{\theta(x_{k_i})}{\|d_{k_i}\|^2\, C_\theta}$, we have $-\alpha\theta(x_{k_i}) + C_\theta\alpha^2\|d_{k_i}\|^2 \leq 0$ and thus from (25a)

$$\theta(x_{k_i} + \alpha d_{k_i}) \leq \theta(x_{k_i}). \tag{40}$$

The claim then follows from (38), (39) and (40) using (18). □

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 will be 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* $\chi(x_{k_i}) \geq \epsilon$ *for a constant* $\epsilon > 0$ *independent of* k_i*. Then there exists* $K \in \mathbb{N}$*, so that for all* $k_i \geq K$ *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

$$\theta(x_{k_i}) \leq \epsilon_1 \quad \text{and} \quad m_{k_i}(\alpha) \leq -\alpha\epsilon_2 \tag{41}$$

for k_i sufficiently large and $\alpha \in (0,1]$; without loss of generality we can assume that (41) 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 $k_i \geq K$

$$\theta(x_{k_i}) < \min\left\{\theta_{\rm inc}, \frac{\bar{\alpha}}{c_6}, \frac{c_5}{c_6}, \left[\frac{\tau_1 c_6 \epsilon_2^{s_f}}{\delta}\right]^{\frac{1}{s_\theta - 1}}\right\} \tag{42}$$

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 this implies that $k_i \notin \mathcal{R}_{\rm inc}$,

$$\frac{\delta\,[\theta(x_{k_i})]^{s_\theta}}{\epsilon_2^{s_f}} < \tau_1 c_6 \theta(x_{k_i}) \tag{43}$$

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

$$c_6\theta(x_{k_i}) < \min\{\bar{\alpha}, c_5\}. \tag{44}$$

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

$$\beta_{k_i} := c_6\theta(x_{k_i}) \stackrel{(44)}{=} \min\{\bar{\alpha}, c_5, c_6\theta(x_{k_i})\}. \tag{45}$$

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

$$f(x_{k_i}(\alpha_{k_i,l})) \leq f(x_{k_i}) + \eta_f m_{k_i}(\alpha_{k_i,l}) \tag{46}$$

and

$$\Big(\theta(x_{k_i}(\alpha_{k_i,l})), f(x_{k_i}(\alpha_{k_i,l}))\Big) \notin \mathcal{F}_{k_i}. \tag{47}$$

If we now denote with $\alpha_{k_i,L}$ the first trial step size satisfying both (46) and (47), the backtracking line search procedure in Step 4.5 then implies that for $\alpha \geq \alpha_{k_i,L}$

$$\alpha \geq \tau_1\beta_{k_i} \stackrel{(45)}{=} \tau_1 c_6\theta(x_{k_i}) \stackrel{(43)}{>} \frac{\delta[\theta(x_{k_i})]^{s_\theta}}{\epsilon_2^{s_f}}$$

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

$$\delta[\theta(x_{k_i})]^{s_\theta} < \alpha\epsilon_2^{s_f} = [\alpha]^{1-s_f}\,(\alpha\epsilon_2)^{s_f} \stackrel{(41)}{\leq} [\alpha]^{1-s_f}\,[-m_{k_i}(\alpha)]^{s_f}\,.$$

This means, the switching condition (9) is satisfied for $\alpha_{k_i,L}$ and all previous trial 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. We also have $\alpha_{k_i,l} \geq \alpha_{k_i}^{\min}$, i.e. the method does not switch to the feasibility restoration phase in Step 4.2 for those trial step sizes. Consequently, $\alpha_{k_i,L}$ is indeed the accepted step size α_{k_i}. Since it satisfies both (9) and (46), 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*

$$\lim_{k\to\infty} \theta(x_k) = 0 \tag{48a}$$

$$\text{and} \qquad \liminf_{k\to\infty} \chi(x_k) = 0. \tag{48b}$$

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. (48a) follows from Theorem 1. In order to show (48b), we have to 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 (48b). □

Remark 6 *It is not possible to obtain a stronger result in Theorem 2, such as "*$\lim_k \chi(x_k) = 0$*". The reason for this is that even 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 "old historic 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* $(\theta(x_k), f(x_k))$ *is very close to the current filter (e.g. there exists 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, the trial step sizes can be repeatedly rejected in Step 4.3 so that finally* $\alpha_{k,l}$ *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 less infeasible, but possibly far away from any KKT point.*

In order to avoid that x_{k+1} *diverts from a strict local solution* x_* *(satisfying the usual second order sufficient optimality conditions, see e.g. [16]), we propose the following procedure. If the restoration phase is invoked at points where the KKT error (the norm of the left hand side of (2)) is small, continue to take steps into the usual search directions* d_k *from (3) (now within the restoration phase), as long as the KKT error is decreased by a fixed fraction. If this is not possible, we have to revert to a different algorithm for the feasibility restoration phase. If* x_k *is sufficiently close to* x_**, the second order sufficient optimality conditions ensure that* x_* *is a point of attraction for Newton's method, so that this procedure will be able to eventually deliver a new iterate* x_{k+1} *which is sufficiently close to feasibility in order to be accepted by the current filter and at the same time approaches* x_**, so that overall* $\lim_k x_k = x_*$ *is guaranteed.*

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. [16])

$$\ell_\rho(x, \lambda) := f(x) + \lambda^T c(x) + \frac{\rho}{2} c(x)^T c(x), \tag{49}$$

where λ are multiplier estimates corresponding to the equality constraints (1b). If λ_* are the 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 augmented Lagrangian function rather than on the exact penalty function, by splitting the augmented Lagrangian function (49) into its two components $\mathcal{L}(x, \lambda)$ (defined in (4)) and $\theta(x)$ (or equivalently $\theta(x)^2$). 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, for example, with no additional computational cost as $d_k^\lambda := \lambda_k^+ - \lambda_k$ with λ_k^+ from (3). Defining

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

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

$$\begin{aligned} \mathcal{L}(x_k(\alpha_{k,l}), \lambda_k(\alpha_{k,l})) &\leq \mathcal{L}(x_k, \lambda_k) - \gamma_f \theta(x_k) \qquad \text{and} \\ \mathcal{L}(x_k(\alpha_{k,l}), \lambda_k(\alpha_{k,l})) &\leq \mathcal{L}(x_k, \lambda_k) + \eta_f m_k(\alpha_{k,l}), \end{aligned}$$

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

$$\begin{aligned} m_k(\alpha) &:= \alpha g_k^T d_k - \alpha \lambda_k^T c_k + \alpha(1-\alpha) c_k^T d_k^\lambda \\ &= \mathcal{L}(x_k + \alpha d_k, \lambda_k + \alpha d_k^\lambda) - \mathcal{L}(x_k, \lambda_k) + O(\alpha^2) \end{aligned} \quad (50)$$

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

The switching condition (9) remains unchanged, but the definition of the minimum step size (17) has to be changed accordingly. 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 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.

One can verify that the global convergence analysis in Section 3 still holds with minor modifications [23]. Recently, Ulbrich [21] has discussed a similar approach using the Lagrangian function in a trust region filter setting, including both global and local convergence results. Unlike the extension we propose in the companion paper, this algorithm does not use second order corrections for the local analysis.

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 with inequality constraints of the form

$$\min_{x \in \mathbb{R}^n} \quad f(x) \quad (51a)$$

$$\text{s.t.} \quad c^{\mathcal{E}}(x) = 0 \quad (51b)$$

$$c^{\mathcal{I}}(x) \ge 0, \quad (51c)$$

where the functions f and $c := (c^{\mathcal{E}}, c^{\mathcal{I}})$ have the smoothness properties of f and c in Assumptions (G1). A line search SQP method obtains search directions d_k as the solution of the quadratic program (QP)

$$\min_{d \in \mathbb{R}^n} \quad g_k^T d + \frac{1}{2} d^T H_k d \quad (52a)$$

$$\text{s.t.} \quad (A_k^{\mathcal{E}})^T d + c^{\mathcal{E}}(x_k) = 0 \quad (52b)$$

$$(A_k^{\mathcal{I}})^T d + c^{\mathcal{I}}(x_k) \ge 0, \quad (52c)$$

where $g_k := \nabla f(x_k)$, $A_k^{\mathcal{E}} := \nabla c^{\mathcal{E}}(x_k)$, $A_k^{\mathcal{I}} := \nabla c^{\mathcal{I}}(x_k)$, and H_k is (an approximation of the) Hessian of the Lagrangian

$$\mathcal{L}(x, \lambda, v) = f(x) + (c^{\mathcal{E}}(x))^T \lambda - (c^{\mathcal{I}}(x))^T v$$

of the NLP (51) with the Lagrange multipliers $v \ge 0$ corresponding to the inequality constraints (51c). We will denote the optimal QP multipliers corresponding to (52b) and (52c) with λ_k^+ and $v_k^+ \ge 0$, respectively.

We further define the infeasibility measure by

$$\theta(x) := \left\| \begin{pmatrix} c^{\mathcal{E}}(x) \\ c^{\mathcal{I}}(x)^{(-)} \end{pmatrix} \right\|,$$

where for a vector w the expression $w^{(-)}$ defines the vector with the components $\max\{0, -w^{(i)}\}$. 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 (52). The restoration phase is invoked in this step, if the QP (52) is infeasible or not "sufficiently consistent" (see Assumption (G4**) below).

In order to state the assumptions necessary for a global convergence analysis let us again consider a decomposition of the search direction

$$d_k = q_k + p_k \tag{53}$$

where q_k is now defined as the solution of the QP

$$\begin{aligned} \min_{q \in \mathbb{R}^n} \quad & q^T q \\ \text{s.t.} \quad & (A_k^{\mathcal{E}})^T q + c^{\mathcal{E}}(x_k) = 0 \\ & (A_k^{\mathcal{I}})^T q + c^{\mathcal{I}}(x_k) \geq 0, \end{aligned}$$

i.e. q_k is the shortest vector satisfying the constraints in the QP (52), and p_k is simply defined as $d_k - q_k$. With these definitions we can now replace Assumptions (G4) and (G5) by

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

$$\|d_k\| \leq M_d, \quad \|\lambda_k^+\| \leq M_\lambda, \quad \|v_k^+\| \leq M_v, \quad \|q_k\| \leq M_q \theta(x_k)$$

(G5**) *There exists a constant $M_H > 0$, so that for all $k \notin \mathcal{R}_{\text{inc}}$ we have*

$$d_k^T H_k d_k \geq M_H d_k^T d_k. \tag{54}$$

The last inequality in Assumption (G4**) is similar to the assumption expressed by Eq. (2.10) in [6]. Essentially, we assume that if the constraints of the QPs (52) become increasingly linearly dependent, eventually the restoration phase will be triggered in Step 3. Together with Assumption (G3) this assumption also means that we suppose that the QP (52) is sufficiently consistent when feasible points are approached.

Assumption (G5**) again ensures descent in the objective function at sufficiently feasible points. This assumption has been made previously for global convergence proofs of SQP line search methods (see e.g. [18]). However, this assumption can be rather strong since even close to a strict local solution the exact Hessian might have to be modified in order to satisfy (54). In [23] an alternative and more natural assumption is considered for an NLP formulation (56) which only allows bound constraints as inequality constraints.

In order to see that the global convergence analysis in Section 3 still holds under the modified Assumptions G, let us first define the criticality measure again as $\chi(x_k) := \|p_k\|_2$ for $k \notin \mathcal{R}_{\text{inc}}$, and it is straight-forward to verify that the proofs are still valid. Only the proof of Lemma 2 deserves special attention. From the optimality conditions for the QP (52) it follows in particular that

$$g_k + H_k d_k + A_k^{\mathcal{E}} \lambda_k^+ - A_k^{\mathcal{I}} v_k^+ = 0 \tag{55a}$$
$$\left((A_k^{\mathcal{I}})^T d_k + c^{\mathcal{I}}(x_k)\right)^T v_k^+ = 0 \tag{55b}$$
$$v_k^+ \geq 0, \tag{55c}$$

so that for all $k \notin \mathcal{R}_{\text{inc}}$

$$\begin{aligned} g_k^T d_k &\overset{(55a)}{=} -d_k^T H_k d_k - d_k^T A_k^{\mathcal{E}} \lambda_k^+ + d_k^T A_k^{\mathcal{I}} v_k^+ \\ &\overset{(52b),(55b)}{=} -d_k^T H_k d_k + c^{\mathcal{E}}(x_k)^T \lambda_k^+ - c^{\mathcal{I}}(x_k)^T v_k^+ \\ &\overset{(55c)}{\leq} -d_k^T H_k d_k + c^{\mathcal{E}}(x_k)^T \lambda_k^+ + \left(c^{\mathcal{I}}(x_k)^{(-)}\right)^T v_k^+ \\ &\overset{(53)}{\leq} -M_H[\chi(x_k)]^2 + O\left(\chi(x_k)\theta(x_k)\right) + O(\theta(x_k)) \end{aligned}$$

where we used Assumptions (G4**) and (G5**) in the last inequality. This corresponds to the second last line in (24), and 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 (IP)* or *barrier methods*. In this section we will assume that the general NLP is stated as

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{56a}$$

$$\text{s.t.} \quad c(x) = 0 \tag{56b}$$

$$x \geq 0, \tag{56c}$$

possibly after reformulating (51) using slack variables. The algorithm can be changed in an obvious way if (56c) is generalized to lower and upper bound constraints on only some or all variables.

The barrier method presented here can be of the primal or primal-dual type, and differs from the IP filter algorithm proposed by M. Ulbrich, S. Ulbrich, and Vicente [20] 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 [20]. Recently, Benson, Shanno, and Vanderbei [2] proposed several heuristics based on the idea of filter methods, for which improved efficiency compared to their previous merit function approach are reported. Their approach is different from the one proposed here in many aspects, and no global convergence analysis is given. Our assumptions made for the analysis of the interior point method are less restrictive than those made for previously proposed line search IP methods for NLP (e.g. [5, 27, 19]).

A barrier method solves a sequence of *barrier problems*

$$\min_{x \in \mathbb{R}^n} \quad \varphi_\mu(x) := f(x) - \mu \sum_{i=1}^{n} \ln(x^{(i)}) \tag{57a}$$

$$\text{s.t.} \quad c(x) = 0 \tag{57b}$$

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 [15] for primal methods, and by Gould, Orban, Sartenaer, and Toint [11, 12] for primal-dual methods. In those approaches, the barrier problem (57) 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. 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, will eventually solve the next barrier problem well enough in order to satisfy the new tolerance. Then the barrier parameter μ will be 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 (1).

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 in [3] and [4]). The focus of this paper are the properties of the line search filter approach, and we will therefore only address the convergence properties of an algorithm for solving the barrier problem (57) for a *fixed* value of the barrier parameter μ, and only give some additional comments on the overall IP method in Remark 7 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 (57), which only has equality constraints, i.e. we replace all occurrences of f in Algorithm I by φ_μ. However, there are two issues that we have to address:

1. The barrier objective function (57a) is only defined as long as all components of x are strictly positive, i.e. $x > 0$;

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

In order to handle the first point, we will enforce that all iterates x_k are strictly positive. For this purpose, we will assume, that the starting point satisfies $x_0 > 0$, and we further define the largest step size $\alpha_k^{\max} \in (0,1]$ that satisfies the *fraction-to-the-boundary rule*, that is

$$\alpha_k^{\max} := \max\left\{\alpha \in (0,1] : x_k + \alpha d_k \geq (1-\tau)x_k\right\} \tag{58}$$

for a fixed parameter $\tau \in (0,1)$, usually chosen close to 1. With this, we will start the backtracking line search in Step 4.1 of Algorithm I from $\alpha_{k,0} = \alpha_k^{\max}$. For later reference, let us state that the search directions in Step 3 are computed from

$$\begin{bmatrix} W_k + \mu X_k^{-2} & A_k \\ A_k^T & 0 \end{bmatrix} \begin{pmatrix} d_k \\ \lambda_k^+ \end{pmatrix} = -\begin{pmatrix} \nabla f(x_k) - \mu X_k^{-1} e \\ c_k \end{pmatrix}, \tag{59}$$

where $X_k := \operatorname{diag}(x_k)$, e is the vector of ones of appropriate dimension, and W_k is (an approximation of) the Hessian of the Lagrangian for the *original* NLP (56). It will be easy to verify that the analysis below also holds if the primal Hessian of the barrier "μX_k^{-2}" 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.

Let us first state the assumptions necessary for the analysis of the barrier filter line search method.

Assumptions B. *Let $\{x_k\}$ be the sequence generated by Algorithm I (adapted to the solution of the barrier problem), 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 first-order optimal point.*

(B1) *There exists an open set* $\mathcal{C} \subseteq \mathbb{R}^n$ *with* $[x_k, x_k + \alpha_k^{\max} d_k] \subseteq \mathcal{C}$*, 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 iterates* $\{x_k\}$ *are bounded.*

(B3) *The matrices* W_k *approximating the Hessian of the Lagrangian of the* original *NLP (56) in (59) are uniformly bounded.*

(B4) *At all* feasible *points* $\bar{x} \in \mathcal{C}$*, the gradients of the active constraints*

$$\nabla c_1(\bar{x}), \ldots, \nabla c_m(\bar{x}), \qquad \textit{and} \qquad e_i \textit{ for } i \in \{j : \bar{x}^{(j)} = 0\} \tag{60}$$

are linearly independent; e_i *being the* i*-th unit coordinate vector.*

(B5) *The matrices* $W_k + \mu X_k^{-2}$ *are uniformly positive definite on the null space of the Jacobian* A_k^T*.*

(B6) *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* $x_k^{(i)} \leq \tilde{\delta}_x$*.*

Assumption (B2) might seem somewhat strong since it explicitly excludes divergence of the iterates. However, this assumption is necessary in order to guarantee that the barrier objective function $\varphi_\mu(x)$ is bounded below.

Note that Assumption (B4) is considerably less restrictive than those made in the analysis of [5, 20, 27, 28], where it is essentially required that the gradients of all equality constraints and active inequality constraints (60) are linearly independent at *all* limit points, and not only at all *feasible* limit points. The assumptions made in [19] are weaker than this, but still require at all points linear independence of the gradients of all *active* equality and inequality constraints, also at infeasible points. Also note that Assumption (B4) is satisfied in the problematic example presented by the authors in [24], and that Assumption (B6) is reasonable in light of Assumption (B4).

Finally we remark that Assumption (B5) is weaker than the one we made in an earlier version of this paper [25].

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) will be bounded away from the boundary of the region defined by the bound constraints (56c). 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 remark that Lemma 1 together with (58) 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 [23]).

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

Lemma 11 *Suppose Assumptions (B1)-(B5) 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*

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

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.

From Assumptions (B1), (B2), and (B4) we know that there exists $m_\sigma > 0$

$$\sigma_{\min}(A_k^l) \geq m_\sigma \tag{61}$$

for all $x_k \in L$ if $\delta_\theta > 0$ is chosen sufficiently small. Furthermore, from Assumption (B5) it also follows 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.

Let us now rewrite the linear system (59) by scaling the first rows and columns by X_k^s:

$$\begin{bmatrix} X_k^s W_k^{ss} X_k^s + \mu I & X_k^s W_k^{sl} & X_k^s A_k^s \\ W_k^{ls} X_k^s & W_k^{ll} + \mu(X_k^l)^{-2} & A_k^l \\ (A_k^s)^T X_k^s & (A_k^l)^T & 0 \end{bmatrix} \begin{pmatrix} \tilde{d}_k^s \\ d_k^l \\ \lambda_k^+ \end{pmatrix} = - \begin{pmatrix} X_k^s g_k^s - \mu e \\ g_k^l - \mu(X_k^l)^{-1} e \\ c(x_k) \end{pmatrix} \tag{62}$$

where we defined $\tilde{d}_k^s := (X_k^s)^{-1} d_k^s$. Using the boundedness assumptions and the comments in the previous paragraph we see that this system satisfies (for $x_k \in L$)

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

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. Recalling that $d_k^s = X_k^s \tilde{d}_k^s$ proves the claim. □

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 exist an index s and a subsequence $\{x_{k_j}\}$ of iterates with $k_j \notin \mathcal{R}$, $\lim_j \theta(x_{k_j}) = 0$ and $\lim_j x_{k_j}^{(s)} = 0$, 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 objective function is unbounded above; see Remark 5), 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 (B6)). 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 $x_k^{(i)} \leq \delta_x$ and (6) for $k \notin \mathcal{R}$, as well as Assumption (B6), 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 the (58) 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 7 *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 IP methods. The method also possesses favorable local convergence behavior, as we discuss in the companion paper [26]. We also 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 future paper we will present practical experience with the line search filter barrier method proposed in this paper. So far, our numerical results are very promising [23].

References

[1] C. Audet and J. E. Dennis. A pattern search filter method for nonlinear programming without derivatives. Technical Report TR00-09, Department of Computational and Applied Mathematics, Rice University, Houston, TX, USA, March 2000.

[2] H. Y. Benson, D. F. Shanno, and R. J. Vanderbei. Interior-point methods for nonconvex nonlinear programming: Filter methods and merit functions. Technical report, Operations Research and Financial Engineering, Princeton University, Princeton, NJ, USA, December 2000.

[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*. Addison–Wesley Longman, Reading, MA, USA, 1997.

[5] 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.

[6] 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. Technical Report 99/03, Department of Mathematics, University of Namur, Belgium, May 1999. Revised October 2001. To appear in SIAM Journal on Optimization.

[7] 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.

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

[9] 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.

[10] 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.

[11] N. I. M. Gould, D. Orban, A. Sartenaer, and P. L. Toint. Componentwise fast convergence in the solution of full-rank systems of nonlinear equations. Technical Report TR/PA/00/56, CERFACS, Toulouse, France, 1999. Revised May 2001.

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

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

[14] 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.

[15] 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.

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

[17] 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.

[18] K. Schittkowski. The nonlinear programming method of Wilson, Han and Powell with an augmented Lagrangian type line search function. *Numerische Mathematik*, 38, 1981.

[19] A.L. Tits, T.J. Urban, S. Bakhtiari, and C.T. Lawrence. A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties. Technical Report TR 2001-3 R2, Department of Electrical and Computer Engineering and Institute for Systems Research, Univerity of Maryland, College Park, MD, July 2001.

[20] M. Ulbrich, S. Ulbrich, and L. N. Vicente. A globally convergent primal-dual interior-point filter method for nonconvex nonlinear programming. Technical Report TR00-12, Department of Computational and Applied Mathematics, Rice University, Houston, TX, USA, April 2000. Revised February 2002.

[21] S. Ulbrich. On the superlinear local convergence of a filter-SQP method. *Mathematical Programming.* to appear.

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

[23] 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.

[24] 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.

[25] A. Wächter and L. T. Biegler. Global and local convergence of line search filter methods for nonlinear programming. Technical Report CAPD B-01-09, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA, August 2001.

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

[28] 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.