Michael Ulbrich · Stefan Ulbrich · Luís N. Vicente

A globally convergent primal-dual interior-point filter method for nonlinear programming

Abstract. In this paper, the filter technique of Fletcher and Leyffer (1997) is used to globalize the primal-dual interior-point algorithm for nonlinear programming, avoiding the use of merit functions and the updating of penalty parameters.

The new algorithm decomposes the primal-dual step obtained from the perturbed first-order necessary conditions into a normal and a tangential step, whose sizes are controlled by a trust-region type parameter. Each entry in the filter is a pair of coordinates: one resulting from feasibility and centrality, and associated with the normal step; the other resulting from optimality (complementarity and duality), and related with the tangential step.

Global convergence to first-order critical points is proved for the new primal-dual interior-point filter algorithm.

Key words. interior-point methods – primal-dual – filter – global convergence

1. Introduction

In this paper we use the filter technique of Fletcher and Leyffer [14] to globalize the primal-dual interior-point method for nonlinear optimization. This technique incorporates the concept of nondominance (borrowed from multi-criteria optimization) to build a *filter* that is able to reject poor trial iterates and enforce global convergence from arbitrary starting points. The filter replaces the use of merit functions, avoiding therefore the update of penalty parameters associated with the penalization of the constraints in merit functions.

Since its first appearance in a 1997 paper by Fletcher and Leyffer [14], the filter technique has been mostly applied, so far, to SLP (sequential linear programming) and SQP (sequential quadratic programming) type methods [12, 14, 15]. Global conver-

gence to first-order critical points has been proved for SLP by Fletcher, Leyffer, and Toint [15] in 1998 and for SQP by Fletcher, Gould, Leyffer, Toint, and Wächter [12] in 1999. In the context of composite SQP for equality constrained optimization, Ulbrich and Ulbrich [22] have also proposed, based on filter ideas, a nonmonotone trust-region algorithm. Recently, Audet and Dennis [2] presented a pattern search filter method for derivative-free nonlinear programming. The filter idea has proven to be very successful numerically in the SLP/SQP framework [13], motivating its applications to interior-point methods.

Interior-point methods, although quite well studied for linear and convex programming, are still a very open topic of research in nonlinear programming. One of the issues is guaranteeing global convergence because there seems to be no ideal merit function. Several approaches for globalizing interior-point methods using different merit functions have been proposed. See the references [5, 9, 11, 16, 17, 23]. On the other hand, the local convergence properties of interior-point methods for nonlinear programming are quite well studied in the literature [6, 8, 11, 19, 24, 31], although difficulties arise when the limit point does not satisfy strict complementarity or linear independence of the gradients of the active or binding constraints [18, 21, 25, 30]. Since the appearance of this paper, two others have also appeared applying the filter technique to interior-point methods (see [3] and [26, 28]).

The primal-dual interior-point method is based on the application of Newton's method to a perturbed version of the first-order necessary conditions. The perturbation incorporates the (numerically efficient) notion of centrality, forcing the iterates to stay as much as possible away from the boundary of the feasible set. Our primal-dual interior-point filter method is partially motivated by the SQP-filter algorithm of Fletcher, Gould, Leyffer, Toint, and Wächter [12]. We also split the primal-dual step into *normal* and *tangential* steps and use a trust-region parameter to control the size of both steps. The normal step points towards the quasi-central path, trying to achieve an improvement in feasibility and centrality. The tangential step is designed to reduce the size of the gradient of the Lagrangian function (and complementarity). The algorithm incorporates also a restoration phase (proposed in [12, 14, 15] for SLP/SQP-filter) aimed to improve, if necessary, feasibility and centrality.

This paper is organized as follows. A brief outline of the basic concept of filter methods is given in Section 2. The primal-dual interior-point framework is presented in Section 3, where a number of estimates are presented for the composite primal-dual step (the proofs are postponed to an appendix). The filter mechanism and the primal-dual interior-point filter method are described in Section 4. Section 5 contains the proof of global convergence to first-order critical points. A restoration algorithm is proposed in Section 6 and some final remarks and open questions are stated in Section 7.

We use $\| \cdot \|$ to denote the Euclidean norm of a vector as well as the induced matrix norm. Given two vectors $u \in \mathbb{R}^{p_1}$ and $v \in \mathbb{R}^{p_2}$, we use (u, v) to represent the vector $w = (u^T, v^T)^T \in \mathbb{R}^{p_1+p_2}$. Finally, we pose the nonlinear programming problem in the general form

$$\min f(x) \quad \text{s.t.} \quad h(x) = 0, \quad x \geq 0, \tag{1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are twice continuously differentiable functions on an open set $\Omega \subset \mathbb{R}^n$.

2. Basic concept of a filter method

We begin by outlining the main concepts of the filter method by Fletcher and Leyffer [14], using the form analyzed by Fletcher, Gould, Leyffer, Toint, and Wächter [12]. This method is applicable to the general nonlinear programming problem. However, since in our interior-point approach the nonnegativity constraints $x \geq 0$ will be handled by the interior-point step calculation, it is sufficient to sketch the basic algorithmic framework of the original filter method for the simpler (equality constrained) problem

$$\min f(x) \quad \text{s.t.} \quad h(x) = 0, \tag{2}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $h : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are twice continuously differentiable functions on an open set $\Omega \subset \mathbb{R}^n$.

The concept of NLP filter methods originates from the observation that the solution of a nonlinear programming problem like (2) consists of the two competing aims of minimizing a measure of feasibility $\theta(x)$, e.g. $\theta(x) = \|h(x)\|$, and of minimizing the objective function $f(x)$. Hence, (2) can be seen as a bi-criteria optimization problem with the additional requirement that θ has some priority, since convergence to a feasible point must be ensured. Instead of combining the two objectives by using a penalty function, Fletcher and Leyffer [14] proposed the use of a *filter* to build the efficient border of the bi-criteria optimization problem of minimizing infeasibility and objective function value. The definition of a filter takes into account the fact that we would like to reduce both, $\theta(x)$ and $f(x)$. A filter $\mathcal{F}$ is a finite set of tuples $(\theta(x_j), f(x_j))$ – pairs in this case – that correspond to a collection of points x_j, with the additional requirement that no filter entry is dominated by any of the others. Hereby, following [14], a point x, or the corresponding pair $(\theta(x), f(x))$, is said to dominate a point x', or the corresponding pair $(\theta(x'), f(x'))$, if

$$\theta(x) \leq \theta(x') \quad \text{and} \quad f(x) \leq f(x').$$

If x dominates x', the latter is most probably of no real interest. A natural requirement for a new iterate is, therefore, that it should not be dominated by previous iterates. The filter serves the purpose of collecting information on selected previous iterates, and thus provides, in terms of dominance, a selection criteria for new iterates. However, it is obvious that the acceptance of iterates whenever they are not dominated by the filter (i.e., by any of the filter entries) does not exclude, for example, a clustering of iterates at an infeasible point. To avoid the acceptance of pairs that are arbitrarily close to the efficient border (i.e., to the boundary of the set of all pairs that are dominated by the filter), acceptability of x to the filter is defined in [12] in a more stringent way by requiring that for all filter entries $(\theta(x_j), f(x_j)) \in \mathcal{F}$ it holds that

$$\max\{\theta(x_j) - \theta(x), f(x_j) - f(x)\} > \gamma_{\mathcal{F}} \theta(x_j),$$

where $\gamma_{\mathcal{F}} \in (0, 1/2)$ is fixed. The original concept of nondominance is still used to add a point, or the corresponding pair $(\theta(x), f(x))$, to the filter: If x is added to the filter, then all entries that are dominated by x are removed from the filter.

In order to produce new iterates that are acceptable to the filter, Fletcher, Gould, Leyffer, Toint, and Wächter [12] proposed the use of a trust-region framework and the

decomposition of the step $s_k = s_k^n + s_k^t$ into a normal step s_k^n and a tangential step s_k^t. The normal step s_k^n is computed to yield linearized feasibility, i.e.,

$$h(x_k) + \nabla h(x_k)^T s_k^n = 0, \quad \|s_k^n\| \le \kappa_\Delta \Delta_k, \tag{3}$$

where $\kappa_\Delta \in (0, 1)$ is a constant. It is assumed that there exists a constant $\kappa_n > 0$ such that $\|s_k^n\| \le \kappa_n \theta(x_k)$ for all k. The tangential step is computed to satisfy

$$\nabla h(x_k)^T s_k^t = 0, \quad \|s_k^n + s_k^t\| \le \Delta_k,$$

and provide a fraction of the Cauchy decrease for a quadratic model m_k of f.

Problem (3) is infeasible if Δ_k is too small in comparison to $\theta(x_k)$. Therefore, it is required in [12] that

$$\|s_k^n\| \le \kappa_\Delta \Delta_k \min\{1, \kappa_\mu \Delta_k^\mu\} \tag{4}$$

with constants $\kappa_\mu > 0$ and $\mu \in (0, 1)$. In particular, for small Δ_k, the normal step s_k^n must be tiny compared to Δ_k. If it is not possible to compute a normal step satisfying (4), a restoration phase is entered with the goal of reducing the infeasibility, measured by θ, as much as needed. Then the full normal step can always be taken, and thus $\theta(x_k + s_k) = \mathcal{O}(\Delta_k^2)$. Possibly after reducing Δ_k (and reentering restoration if necessary), the new trial iterate will be acceptable to the filter if all filter entries (θ_j, f_j) satisfy $\theta_j > 0$, which is ensured by the mechanism of selecting new filter entries. If the filter test is passed, the decrease properties of the full step s_k on the model m_k are checked. If the predicted decrease for f is not promising, more precisely if $m_k(x_k) - m_k(x_k + s_k) < \kappa_\theta \theta(x_k)^2$ with a constant $\kappa_\theta \in (0, 1)$, then the infeasibility is considered to dominate the possible decrease in f. The new iterate is accepted and x_k is added to the filter. Otherwise, the f-decrease is required to satisfy the standard trust-region acceptance criterion

$$\frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)} \ge \eta$$

with preset $\eta \in (0, 1)$. If the test fails, Δ_k is reduced. Otherwise, the step is accepted and Δ_k is updated as in a standard trust-region algorithm. If the f-decrease is met, x_k is not added to the filter, since $\theta(x_k)$ can be very small in this case and adding x_k can enforce small trust-region radii to get acceptable points in later iterations.

The interior-point filter method introduced in this paper is inspired by the SQP filter method described in [12]. Essentially, our application of the filter concept to the globalization of interior-point methods has the following features:

- A primal-dual interior-point Newton step forms the basis for the trial step computation. The nonnegativity constraints are handled by a centering mechanism.
- We identify two objectives θ and θ_g, corresponding to feasibility and objective function value, such that an appropriate splitting of the step in "normal" and "tangential" components guarantees decrease for linearized models of θ and θ_g, respectively.
- An adaptation of the filter framework of [12] for the pair (θ, θ_g) is proposed.
- The proposed method does not require the recomputation of the normal and tangential steps as the step bound Δ is reduced.

3. Interior-point framework

We return to the nonlinear programming problem posed in the general form (1).

3.1. Step computation

Primal-dual interior-point methods are based on the idea of applying Newton's method to an appropriate perturbation of the first-order necessary optimality conditions (Karush-Kuhn-Tucker or KKT conditions). For the problem under consideration, the KKT conditions can be written in the form

$$\nabla_x \ell(x, y, z) = 0, \tag{5}$$
$$h(x) = 0, \tag{6}$$
$$Xz = 0, \tag{7}$$
$$x \geq 0, \quad z \geq 0, \tag{8}$$

where $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ are the Lagrange multipliers, ℓ denotes the Lagrange function

$$\ell(x, y, z) = f(x) + h(x)^T y - x^T z,$$

and X is the diagonal matrix of order n in which the i-th diagonal element is x_i. Under a constraint qualification the conditions (5)–(8) are necessary for x to be a local solution of (1).

We now perturb block (7) of the KKT system (5)–(7) and write

$$F_{\hat{\mu}}(x, y, z) \stackrel{\text{def}}{=} \begin{pmatrix} \nabla_x \ell(x, y, z) \\ h(x) \\ Xz - \hat{\mu}e \end{pmatrix} = 0,$$

where $\hat{\mu} > 0$. Throughout, we will work with $\hat{\mu} = \sigma\mu$, where $\sigma \in (0, 1)$ is a centering parameter, and

$$\mu = \frac{x^T z}{n}. \tag{9}$$

To abbreviate notation we set

$$w = (x, y, z) \quad \text{and} \quad \Delta w = (\Delta x, \Delta y, \Delta z).$$

The primal-dual Newton step Δw is determined by the Newton equation for the perturbed KKT system, i.e.,

$$F'_{\sigma\mu}(x, y, z)\Delta w = -F_{\sigma\mu}(x, y, z),$$

or, in detail, by

$$\begin{pmatrix} \nabla^2_{xx}\ell(x, y, z) & \nabla h(x) & -I \\ \nabla h(x)^T & 0 & 0 \\ Z & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = - \begin{pmatrix} \nabla_x \ell(x, y, z) \\ h(x) \\ Xz - \sigma\mu e \end{pmatrix}.$$

The choice $\hat{\mu} = \sigma\mu$ with centering parameter $\sigma \in (0, 1)$ and complementarity measure μ according to (9) ensures that the primal-dual Newton direction Δw is a descent direction for $x^T z/n$ and allows thus a dynamic reduction of μ. This choice is frequently used in the context of linear programming and was also used in the nonlinear programming algorithm of El-Bakry et al. [11].

To adapt the methodology of a filter as outlined in Section 2 to our interior-point context, we have to specify the two quantities for the filter entries, the first component corresponding to feasibility and the second corresponding to optimality. With this choice of filter components, we have to find a corresponding decomposition of the trial step into a normal step and a tangential step that yields a decrease of the feasibility- and optimality-component, respectively.

To motivate our choice of the components in the filter and the step decomposition, we rewrite the perturbed KKT-conditions in the form

$$F_{\sigma\mu}(x, y, z) = \begin{pmatrix} 0 \\ h(x) \\ Xz - \mu e \end{pmatrix} + \begin{pmatrix} \nabla_x \ell(x, y, z) \\ 0 \\ (1 - \sigma)\mu e \end{pmatrix} = 0. \tag{10}$$

The first term in the middle expression measures the proximity to the quasi-central path. We recall that the quasi-central path, parameterized by μ (see [1]), is defined by

$$P_\mu^q = \{(x, z) : h(x) = 0, \; Xz = \mu e\}.$$

Therefore, in terms of our filter approach, it seems natural to let the quasi-central path play the role of the feasible set $\{x : h(x) = 0\}$ in Section 2 and to choose the measure of *quasi-centrality*

$$\theta(w) = \|h(x)\| + \|Xz - (x^T z)/ne\|$$

as the first component in the filter. In this context it is important to mention that the central path for nonlinear programming

$$P_\mu^c = \{w = (x, y, z) : \nabla_x \ell(w) = 0, \; h(x) = 0, \; Xz = \mu e\},$$

parameterized by μ, is only guaranteed to exist (for sufficiently small μ) in the neighborhood of a point (x, y, z) that satisfies the second-order sufficient conditions, strict complementarity ($\max\{x_i, z_i\} > 0, i = 1, \dots, n$), and linear independence of the gradients of the active or binding constraints. (The linear independence assumption for the existence of the central path can be weakened to the Mangasarian-Fromowitz constraint qualification [30].)

The second term in the middle expression of (10) measures complementarity and criticality. For the second filter component we choose therefore the optimality measure

$$\theta_g(w) = x^T z/n + \|\nabla_x \ell(w)\|^2.$$

We are aware that this choice certainly allows some room for improvement, since the fact that we are dealing with a minimization problem is not very well reflected by θ_g. However, given that the investigation of filter methods is still in its beginnings, we think that our choice of the optimality measure is appropriate for the purpose of this paper. We

believe that our approach is also viable for other choices of θ_g, and return to this issue in Section 7.

With this choice of filter components it remains to define corresponding tangential and normal components of the trial step. We use the decomposition associated with the splitting (10). For the *normal* step $s^n = (\Delta x^n, \Delta y^n, \Delta z^n)$ we thus choose

$$F'_{\sigma\mu}(w)s^n = -\begin{pmatrix} 0 \\ h(x) \\ Xz - \mu e \end{pmatrix}, \tag{11}$$

whereas our *tangential* step $s^t = (\Delta x^t, \Delta y^t, \Delta z^t)$ is given by

$$F'_{\sigma\mu}(w)s^t = -\begin{pmatrix} \nabla_x \ell(w) \\ 0 \\ (1-\sigma)\mu e \end{pmatrix}. \tag{12}$$

Note that $\Delta w = s^n + s^t$. However, it will be crucial that we exploit the flexibility of the step splitting to introduce different step sizes for s^n and s^t in our trial step computation.

The adjectives normal and tangential are borrowed from the SQP context [4, 20] but have a slightly different flavor here. The normal step can be seen as a step towards the quasi-central path P_μ^q. The tangential step is the sum of a tangential component s_1^t

$$F'_{\sigma\mu}(w)s_1^t = -\begin{pmatrix} \nabla_x \ell(w) \\ 0 \\ 0 \end{pmatrix},$$

which attempts to reduce $\|\nabla_x \ell(w)\|$, and a predictor component s_2^t

$$F'_{\sigma\mu}(w)s_2^t = -\begin{pmatrix} 0 \\ 0 \\ (1-\sigma)\mu e \end{pmatrix},$$

which seeks the minimization of $\mu = x^T z/n$ (see, for instance, [29]). Therefore, the tangential step aims to reduce the optimality measure $\theta_g(w) = x^T z/n + \|\nabla_x \ell(w)\|^2$.

We introduce Δ as the positive scalar that primarily controls the length of the step taken along Δw, forcing the damped components $\alpha^n(\Delta)s^n$ and $\alpha^t(\Delta)s^t$, to satisfy

$$\|\alpha^n(\Delta)s^n\| \leq \Delta, \qquad \|\alpha^t(\Delta)s^t\| \leq \Delta.$$

Having these bounds in mind, and requiring explicitly $\alpha^t(\Delta) \leq \alpha^n(\Delta)$, we define the step sizes taken along the normal and tangential directions respectively as

$$\alpha^n(\Delta) = \min\left\{1, \frac{\Delta}{\|s^n\|}\right\}, \tag{13}$$

$$\alpha^t(\Delta) = \min\left\{\alpha^n(\Delta), \frac{\Delta}{\|s^t\|}\right\} = \min\left\{1, \frac{\Delta}{\|s^n\|}, \frac{\Delta}{\|s^t\|}\right\}. \tag{14}$$

Hereby, we use for $\Delta > 0$ the natural definition $\alpha^n(\Delta) = 1$ for $\|s^n\| = 0$ by using the convention $\min\{1, \infty\} = 1$. We also say that $\alpha^t(\Delta) = \alpha^n(\Delta)$ if $\|s^t\| = 0$, although

our algorithm cannot generate tangential steps for which $\|s^t\| = 0$ since the right-hand-side in (12) will never be zero if the iterates x and z are kept positive throughout. The requirement $\alpha^t(\Delta) \leq \alpha^n(\Delta)$ is mainly necessary to enforce the iterates to stay in the neighborhood $\mathcal{N}(\gamma, M)$ defined in (17), see Lemma 2.

Let also

$$w(\Delta) = (x(\Delta), y(\Delta), z(\Delta)) = w + \alpha^n(\Delta)s^n + \alpha^t(\Delta)s^t, \tag{15}$$

$$s(\Delta) = \big(s_x(\Delta), s_y(\Delta), s_z(\Delta)\big) = w(\Delta) - w = \alpha^n(\Delta)s^n + \alpha^t(\Delta)s^t. \tag{16}$$

Thus,

$$\|s(\Delta)\| \leq 2\Delta,$$

and Δ plays here a role comparable to the trust-region radius.

The scalars $\alpha^n(\Delta)$ and $\alpha^t(\Delta)$ will be such that positivity and some measure of centrality of the new iterate $w(\Delta)$ are maintained. However, both $\alpha^n(\Delta)$ and $\alpha^t(\Delta)$ depend on Δ, that in turn will be adjusted, not only to meet the purpose of positivity and centrality, but also to enforce global convergence.

We introduce the notation

$$\theta_h(w) = \|h(x)\|, \qquad \theta_c(w) = \left\| Xz - \frac{x^T z}{n}e \right\|, \qquad \theta_\ell(w) = \|\nabla_x \ell(w)\|,$$

which allows us to write the filter components as

$$\theta(w) = \theta_c(w) + \theta_h(w), \qquad \theta_g(w) = \frac{x^T z}{n} + \|\nabla_x \ell(w)\|^2.$$

Since Xz might not be zero, a point w that satisfies $\theta(w) = \theta_\ell(w) = 0$ and $(x, z) \geq 0$, might not be a KKT point. The definition of $\theta_g(w)$, however, guarantees that a point w satisfying $\theta(w) = \theta_g(w) = 0$ and $(x, z) \geq 0$, is indeed a KKT point.

With the purpose of achieving a reduction on the function θ_g, we introduce, at a given point w, the quadratic model

$$m(w(\Delta))$$
$$= \frac{x^T z}{n} + \frac{(x(\Delta) - x)^T z + (z(\Delta) - z)^T x}{n} + \|\nabla_x \ell(w) + \nabla_{wx}^2 \ell(w)(w(\Delta) - w)\|^2$$
$$= \frac{x(\Delta)^T z(\Delta) - (x(\Delta) - x)^T (z(\Delta) - z)}{n} + \|\nabla_x \ell(w) + \nabla_{wx}^2 \ell(w)(w(\Delta) - w)\|^2,$$

by adding to the linearization of $x^T z/n$ the squared norm of the linearization of $\nabla_x \ell(w)$. To simplify the notation we also define

$$\mu(\Delta) = \frac{x(\Delta)^T z(\Delta)}{n}.$$

In order to prevent $(x(\Delta), z(\Delta))$ from approaching the boundary of the positive orthant too rapidly we will keep the iteration in the neighborhood

$$\mathcal{N}(\gamma, M) = \left\{ w : (x, z) > 0, \ \ Xz \geq \gamma \frac{x^T z}{n}, \ \ \theta_h(w) + \theta_\ell(w) \leq M \frac{x^T z}{n} \right\} \tag{17}$$

with fixed $\gamma \in (0, 1)$ and $M > 0$. This is a frequently used centrality condition in infeasible interior-point methods for linear and convex quadratic programming, cf. [29] and the references therein, and has also proven its efficiency in the context of nonlinear programming [11]. We will see in the next subsection that $w \in \mathcal{N}(\gamma, M)$ implies $w(\Delta) \in \mathcal{N}(\gamma, M)$ whenever $\Delta \in (0, \Delta_{\min}]$ for a given constant $\Delta_{\min} > 0$.

3.2. Step estimates

The following lemma provides upper bounds for θ_h, θ_c, θ_ℓ, and θ_g at the new point $w(\Delta)$ in terms of Δ and of the corresponding values at the previous point w. It also provides a lower bound for the decrease produced on the quadratic model m by the step $w(\Delta) - w$.

Lemma 1. *There exist positive constants M_h, M_c, and M_ℓ, depending on an upper bound for θ_ℓ and on the Lipschitz constants of ∇h and $\nabla^2_{xw}\ell$, such that, for all $\Delta > 0$,*

$$\theta_h(w(\Delta)) \le (1 - \alpha^n(\Delta))\theta_h(w) + M_h\Delta^2, \tag{18}$$

$$\theta_c(w(\Delta)) \le (1 - \alpha^n(\Delta))\theta_c(w) + M_c\Delta^2, \tag{19}$$

$$\theta_\ell(w(\Delta)) \le (1 - \alpha^t(\Delta))\theta_\ell(w) + M_\ell\Delta^2. \tag{20}$$

Given an upper bound $\Delta_{ub} > 0$ for Δ, it also holds

$$\theta_g(w(\Delta)) \le \left(1 - \alpha^t(\Delta)(1 - \sigma)\right)\theta_g(w) + M_g\Delta^2, \tag{21}$$

for $0 < \Delta \le \Delta_{ub}$ and for some positive constant M_g.
Finally, for any $\Delta > 0$, we also have

$$m(w) - m(w(\Delta)) \ge \alpha^t(\Delta)(1 - \sigma)\theta_g(w). \tag{22}$$

Now we state a result that says that if the current point $w = (x, y, z)$ satisfies the centrality requirement $Xz \ge \gamma\mu e$, so does the next point $w(\Delta) = (x(\Delta), y(\Delta), z(\Delta))$, provided Δ is sufficiently small. A similar property is also stated for the inequalities $\theta_h(w) + \theta_\ell(w) \le M\mu$ and $(x, z) > 0$.

Lemma 2. *Let $\|F'_{\sigma\mu}(w)^{-1}\| \le C$ and, for $\gamma \in (0, 1)$ and $M > 0$, assume that*

$$Xz \ge \gamma\mu e, \quad \theta_h(w) + \theta_\ell(w) \le M\mu.$$

There exists a constant $\Delta_{\min}$ such that, if $0 < \Delta \le \Delta_{\min}$, then

$$X(\Delta)z(\Delta) \ge \gamma\mu(\Delta)e, \tag{23}$$

$$\theta_h(w(\Delta)) + \theta_\ell(w(\Delta)) \le M\mu(\Delta). \tag{24}$$

Furthermore, if $(x, z) > 0$, then, for all Δ in $(0, \Delta_{\min}]$,

$$(x(\Delta), z(\Delta)) > 0. \tag{25}$$

Thus, $w \in \mathcal{N}(\gamma, M)$ implies $w(\Delta) \in \mathcal{N}(\gamma, M)$ for all $\Delta \in (0, \Delta_{\min}]$.

The proofs of these results are quite technical and are left for an appendix of this paper.

4. The interior-point filter method

Our definition of a filter takes into account the fact that we would like to reduce both $\theta(w) = \theta_c(w) + \theta_h(w)$ and $\theta_g(w)$. Hence, we choose θ and θ_g to form a filter entry, where θ measures feasibility and θ_g measures optimality. We thus replace the objective function value by the criticality measure θ_g. Since we introduced the filter concept already in Section 2, we give here only formal definitions to set the notations for the rest of the paper.

Definition 1 (Dominance). *A point w, or the corresponding pair $(\theta(w), \theta_g(w))$, is said to dominate a point w', or the corresponding pair $(\theta(w'), \theta_g(w'))$, if*

$$\theta(w) \leq \theta(w') \quad and \quad \theta_g(w) \leq \theta_g(w'),$$

or, equivalently, if the following inequality is violated*:*

$$\max\{\theta(w) - \theta(w'), \theta_g(w) - \theta_g(w')\} > 0.$$

Definition 2 (Filter). *A filter is a finite subset $\mathcal{F} \subset \mathbb{R}^2$ consisting of pairs (θ^f, θ_g^f), with $\theta^f \stackrel{\text{def}}{=} \theta_h^f + \theta_c^f$, such that no pair can dominate any of the others.*

As pointed out in Section 2, the mere requirement that a new iterate is not dominated by any of the filter entries is too weak as an acceptance criterion. Instead, we require:

Definition 3. *Let $\gamma_{\mathcal{F}} \in (0, 1/2)$ be fixed. The point w is* acceptable to the filter $\mathcal{F}$ *if, for all $(\theta^f, \theta_g^f) \in \mathcal{F}$, it holds*

$$\max\{\theta^f - \theta(w), \theta_g^f - \theta_g(w)\} > \gamma_{\mathcal{F}}\theta^f.$$

In the course of the algorithm, we will add selected new points to the filter. This procedure is done in the following way:

Definition 4. *By adding w to the filter $\mathcal{F}$ we mean the following operation:*

$$\mathcal{F} \mapsto \mathcal{F} = \{(\theta(w), \theta_g(w))\} \cup$$
$$\left\{(\theta^f, \theta_g^f) \in \mathcal{F} : \min\{\theta^f - \theta(w), \theta_g^f - \theta_g(w)\} < 0\right\}.$$

Remark 1. Therefore, if w is added to the filter, all old entries that are dominated by the new entry are removed. □

Our primal-dual interior-point filter method generates iterates $w_{k+1} = w_k(\Delta_k) \neq w_k$ that are acceptable to the filter, but not all new iterates w_{k+1} are added to the filter.

In general, the primal-dual interior-point filter method imposes a sufficient reduction criterion relating the actual reduction in θ_g with the reduction predicted by its model m_k:

$$\rho_k \geq \eta$$

where

$$\rho_k \stackrel{\text{def}}{=} \frac{\theta_g(w_k) - \theta_g(w_k(\Delta_k))}{m_k(w_k) - m_k(w_k(\Delta_k))}$$

and $\eta \in (0, 1)$ is a preset constant.

However, the test of this condition is skipped if

$$m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa\theta(w_k)^2,$$

where $\kappa \in (0, 1)$ is a preset constant. In other words, the sufficient reduction criterion $\rho_k \geq \eta$ is only imposed when the reduction in the model m_k is sufficiently good compared with $\theta(w_k)^2$. In the situation where $\rho_k < \eta$ and $m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa\theta(w_k)^2$, the new iterate $w_{k+1} = w_k(\Delta_k)$ is accepted and the previous point w_k is added to the filter (guaranteeing that this new filter entry satisfies $\theta(w_k) > 0$). This criterion for adding points w_k to the filter prevents us from building up a filter for which the computation of acceptable points would require too small trust-region radii.

If $\rho_k \geq \eta$ and $m_k(w_k) - m_k(w_k(\Delta_k)) \geq \kappa\theta(w_k)^2$, the iterate w_k is not added to the filter. This situation is the only one where a new iterate $w_{k+1} = w_k(\Delta_k)$ is computed and the previous one, w_k, is not added to the filter.

If $\theta(w_k)$ is too large compared to Δ_k (or an appropriate power of Δ_k), the algorithm enters a restoration phase with the purpose of reducing θ. More precisely, a restoration algorithm is called if

$$\theta(w_k) > \Delta_k \min\{\gamma_1, \gamma_2\Delta_k^\beta\},$$

where γ_1, γ_2, and β are preset positive constants. The restoration algorithm must produce a new iterate w_{k+1} that is not only acceptable to the filter but also satisfies $\theta(w_{k+1}) \leq \Delta_k \min\{\gamma_1, \gamma_2\Delta_k^\beta\}$. In this situation, the previous iterate w_k is added to the filter (guaranteeing also that this new filter entry satisfies $\theta(w_k) > 0$). In Section 6, we propose a restoration algorithm, based on the primal-dual interior-point framework of this paper, that verifies the requirements of the restoration phase.

The primal-dual interior-point filter method satisfying the above features is now presented. Note that step 5 guarantees that the potentially new iterate $w_k(\Delta_k)$ is always acceptable to the filter.

Notation. In the following algorithm, the current iterate in iteration k is denoted by w_k and the normal and tangential trial steps are denoted by s_k^t and s_k^n, respectively. Further, the step sizes $\alpha_k^n(\Delta)$ and $\alpha_k^t(\Delta)$ are defined according to (13) and (14), respectively, with $s^n = s_k^n$ and $s^t = s_k^t$. Similarly, $w_k(\Delta)$ and $s_k(\Delta)$ are defined by (15) and (16), respectively, with $w = w_k$, $s^{n/t} = s_k^{n/t}$, and $\alpha^{n/t}(\Delta) = \alpha_k^{n/t}(\Delta)$.

Algorithm 1 (Primal-dual interior-point filter method).

0. Choose $\sigma \in (0, 1)$, $\nu \in (0, 1)$ $\gamma_1, \gamma_2 > 0$, $0 < \beta, \eta, \kappa < 1$, and $\gamma_{\mathcal{F}} \in (0, 1/2)$. Set $\mathcal{F} := \emptyset$. Choose $(x_0, z_0) > 0$ and y_0, and determine $\gamma \in (0, 1)$ such that $X_0 z_0 \geq \gamma\mu_0$ with $\mu_0 = x_0^T z_0/n$. Further, choose $M > 0$ such that $\theta_h(w_0) + \theta_\ell(w_0) \leq M\mu_0$. Choose $\Delta_0^{in} > 0$ and set $k := 0$.

1. Set $\mu_k := x_k^T z_k/n$ and compute s_k^n and s_k^t by solving the linear systems (11) and (12), respectively, with $(w, \mu) = (w_k, \mu_k)$.

2. Compute $\Delta_k' \in [0, \Delta_k^{in}]$ such that

$$x_k(\Delta) > 0, \quad z_k(\Delta) > 0, \quad X_k(\Delta)z_k(\Delta) \geq \gamma\mu_k(\Delta)e \quad \text{for all } \Delta \in [0, \Delta_k']$$

and such that Δ_k' is not smaller than the largest $\nu^r\Delta_k^{in}$, $r = 0, 1, \ldots$, having this property.

3. Compute the largest $\Delta_k'' = \nu^j \Delta_k'$, $j = 0, 1, \ldots$, such that

$$\theta_h(w_k(\Delta_k'')) + \theta_\ell(w_k(\Delta_k'')) \leq M\mu_k(\Delta_k'').$$

Set $\Delta_k := \Delta_k''$.

4. If $\theta(w_k) \leq \Delta_k \min\{\gamma_1, \gamma_2\Delta_k^\beta\}$ then continue in step 5. Otherwise **add** w_k to the filter and use a restoration algorithm to produce a point w_{k+1} such that:

$w_{k+1} \in \mathcal{N}(\gamma, M)$ with $\mu_{k+1} = x_{k+1}^T z_{k+1}/n$;
w_{k+1} is acceptable to the filter;
$\theta(w_{k+1}) \leq \Delta_{k+1} \min\{\gamma_1, \gamma_2\Delta_{k+1}^\beta\}$ with $\Delta_{k+1} = \Delta_k$.
Set $\Delta_{k+1}^{in} := \Delta_k$, $k := k + 1$, and go to step 1.

5. If $w_k(\Delta_k)$ is not acceptable to the filter (with w_k considered in the filter if $m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa\theta(w_k)^2$), then go to step 11.

6. If $m_k(w_k) - m_k(w_k(\Delta_k)) = 0$, then set $\rho_k := 0$. Otherwise, compute

$$\rho_k = \frac{\theta_g(w_k) - \theta_g(w_k(\Delta_k))}{m_k(w_k) - m_k(w_k(\Delta_k))}.$$

7. If $\rho_k < \eta$ and $m_k(w_k) - m_k(w_k(\Delta_k)) \geq \kappa\theta(w_k)^2$ then go to step 11.

8. If $m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa\theta(w_k)^2$ then **add** w_k to the filter.

9. Choose $\Delta_{k+1}^{in} \geq \Delta_k$.

10. Set $w_{k+1} := w_k(\Delta_k)$, $k := k + 1$, and go to step 1.

11. Set $w_{k+1} := w_k$, $s_{k+1}^n := s_k^n$, $s_{k+1}^t := s_k^t$, $\Delta_{k+1}' := \Delta_k/2$. Set $k := k + 1$ and go to step 3.

In practice, step 2 would be implemented as $\Delta_k' = \tau_k\hat{\Delta}_k'$, where $\hat{\Delta}_k'$ is the largest value of Δ such that $(x_k(\Delta), z_k(\Delta)) \geq 0$ and $X_k(\Delta)z_k(\Delta) \geq \gamma\mu_k(\Delta)e$ and τ_k is a parameter in $(\nu, 1)$ to enforce $(x_k(\Delta), z_k(\Delta)) > 0$. The adjustment of τ_k would be important to achieve a rapid rate of local convergence. We point out that the calculation of Δ_k is split in steps 2 and 3 for good reasons. In fact, in step 2 it is possible to determine explicitly Δ_k' (more precisely $\hat{\Delta}_k'$). However, because of the nonlinearity of θ_h and θ_ℓ, that is not the case in step 3, where we know from Lemma 2 that although there exists a sufficiently small Δ_k'' satisfying $\theta_h(w_k(\Delta_k'')) + \theta_\ell(w_k(\Delta_k'')) \leq M\mu_k(\Delta_k'')$, it cannot be determined explicitly.

In practice step 1 of the algorithm would start by checking the satisfaction of a stopping criterion of the form $\theta(w_k) + \theta_g(w_k) \leq \epsilon$, for small $\epsilon > 0$. To be able to analyze the asymptotic global convergence properties of the algorithm we did not include any stopping criterion.

5. Global convergence to first-order critical points

For the rest of this paper we assume that $\{w_k\}$ is a sequence of iterates generated by the primal-dual interior-point filter method (Algorithm 1). We will also impose the following assumptions.

Assumption 1.

(A1) The sequence $\{(x_k, y_k, z_k)\}$ is bounded.
(A2) The derivatives ∇h and $\nabla^2_{xw}\ell$ exist and are Lipschitz continuous in an open set containing all the iterates (x_k, y_k, z_k) and the line segments $[w_k, w_k + s_k(\Delta_k)]$.
(A3) There exists $C > 0$ such that for all k it holds $\|F'_{\sigma\mu_k}(w_k)^{-1}\| \le C$.

Remark 2. (A3) holds in a neighborhood of a regular point w^* satisfying the second-order sufficient conditions and strict complementarity, see for instance [11]. Moreover, conditions are given in [11] under which (A1) and (A3) are ensured if the iterates w_k are kept in $\mathcal{N}(\gamma, M)$ and only the boundedness of $\{x_k\}$ is assumed.

As we will see, these assumptions allow us to prove global convergence to KKT points. Since KKT points are feasible, we are restricting the analysis to problems that are not infeasible. It is the uniform boundedness of the inverse of the Jacobian given in Assumption (A3) that rules out infeasibility. □

The following simple result is a direct consequence of these assumptions and of Lemmas 1 and 2.

Lemma 3. *The following hold:*

i) *The sequences $\{\theta_h(w_k)\}$, $\{\theta_c(w_k)\}$, $\{\mu_k\}$, and $\{\theta_g(w_k)\}$ are bounded.*
ii) *The constants M_h, M_c, M_ℓ, M_g in Lemma 1 are bounded for all k.*
iii) *There exists $\Delta_{\min} > 0$ such that the conditions in steps 2 and 3 are satisfied for all Δ'_k, $\Delta''_k \in [0, \Delta_{\min}]$. Thus, steps 2 and 3 leave Δ^{in}_k unchanged for $0 \le \Delta^{in}_k \le \Delta_{\min}$ and we have $\Delta_k = \Delta^{in}_k$.*
iv) *It holds that $\max\{\|s^n_k\|, \|s^t_k\|\} \le C(M + (n^2 - n + 1)^{1/2})\mu_k$ for all k.*

For the result iv) we use that $\|Xz - \mu e\| \le (n^2 - n)^{1/2}\mu$ and $\|(1 - \sigma)\mu e\| \le n^{1/2}\mu$.

Given the fact that $\{(x_k, y_k, z_k)\}$ is bounded, the boundedness of the sequence $\{s(\Delta_k)\}$ follows from Lemma 3. (Note that $\|s^n_k\|$ and $\|s^t_k\|$ are bounded by iv) and α^n_k and α^t_k do not exceed one.)

We point out that all filter entries (θ^f, θ^f_g) obey $\theta^f > 0$. This fact is stated in the following lemma without proof and follows directly from the structure of the algorithm.

Lemma 4. *If w_k is added to the filter, then $\theta(w_k) > 0$.*

Proof. An iterate w_k is added to the filter either in step 4 or in step 8. In the first case (step 4), we see that $\theta(w_k) > \Delta_k \min\{\gamma_1, \gamma_2\Delta^\beta_k\} > 0$. In the second case (step 8), we conclude from Lemma 1, (22) that

$$\theta(w_k)^2 > \frac{1}{\kappa}(m_k(w_k) - m_k(w_k(\Delta_k)) \ge 0.$$

Thus, in both cases, $\theta(w_k) > 0$. □

We show in the next lemma that new iterates are always acceptable to the filter. The proof of this result requires no analysis and follows directly from the structure of the algorithm.

Lemma 5. *In all iterations $k \ge 0$, the current iterate w_k is acceptable to the filter.*

Proof. The proof is by induction. Since $\mathcal{F}$ is empty in iteration $k = 0$, the initial iterate w_0 is certainly acceptable to the filter. Now let $k \geq 0$ and assume that w_k is acceptable to the filter. The iterate w_{k+1} is either generated in step 4, or in step 10, or in step 11. If w_{k+1} is obtained in step 4, then w_k is added to the filter and a restoration algorithm is called that returns a point w_{k+1} that is acceptable to the filter. If w_{k+1} is obtained in step 10, then the new iterate $w_{k+1} = w_k(\Delta_k)$ passed the filter acceptance test in step 5 (with w_k considered in the filter in the case where w_k is added in step 8). Finally, if w_{k+1} results from step 11, then $w_{k+1} = w_k$ and w_k is not added to the filter. By the induction hypothesis, w_k is acceptable to the filter. Since the filter is not changed, the same holds true for $w_{k+1} = w_k$. □

The next three lemmas provide technical results needed to establish global convergence to first-order critical points. The first of these lemmas provides a crucial inequality showing that feasibility and centrality at $w_k(\Delta_k)$ are of the order of Δ_k^2.

Lemma 6. *There exists a $\Delta_r > 0$ such that, if $\Delta_k \leq \Delta_r$ in step 5, it holds that*

$$\theta(w_k(\Delta_k)) \leq (M_h + M_c)\Delta_k^2.$$

Proof. If step 5 is reached, then

$$\theta(w_k) \leq \gamma_2 \Delta_k^{1+\beta}.$$

Thus, by (A3) and (11), we have $\|s_k^n\| \leq C\gamma_2 \Delta_k^{1+\beta}$. In the case $\|s_k^n\| = 0$ we get for all $\Delta_k > 0$ that $\alpha_k^n = \min\{1, \Delta_k/\|s_k^n\|\} = 1$ by our natural convention $\min\{1, \infty\} = 1$. Otherwise, we have

$$\frac{\Delta_k}{\|s_k^n\|} \geq \frac{1}{C\gamma_2 \Delta_k^\beta}.$$

We see then that $\alpha_k^n = 1$ whenever

$$\Delta_k \leq \Delta_r \stackrel{\text{def}}{=} \left(\frac{1}{C\gamma_2}\right)^{\frac{1}{\beta}}.$$

Thus, we have in both cases $\alpha_k^n = 1$ for $\Delta_k \leq \Delta_r$. But then, by Lemma 1,

$$\theta(w_k(\Delta_k)) \leq (M_h + M_c)\Delta_k^2. \qquad \square$$

Remark 3. We stress that Lemma 6 as well as the next two Lemmas 7 and 8 make assertions on the situation in *step 5 of the algorithm*. Step 5 is preceded by step 4, and thus, in step 5 it always holds that

$$\theta(w_k) \leq \Delta_k \min\{\gamma_1, \gamma_2 \Delta_k^\beta\}, \tag{26}$$

since otherwise step 4 calls restoration instead of step 5. As we have already seen in Lemma 6, (26) makes it possible to show $\theta(w_k(\Delta_k)) = \mathcal{O}(\Delta_k^2)$ for sufficiently small Δ_k, which would be impossible without (26) holding. □

The next two lemmas deal also with step 5 of the primal-dual interior-point filter method. They provide sufficient conditions on the value of Δ_k for $w(\Delta_k)$ to be acceptable to the filter in step 5. In both lemmas we analyze the acceptability of $w(\Delta_k)$ to the filter by considering that the filter contains w_k if $m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa\theta(w_k)^2$, despite the fact that, in this situation, w_k will possibly be added to the filter only in step 8. Firstly we consider the case where w_k is bounded away from a KKT point and the filter has a finite number of entries.

Lemma 7. *Suppose that* $\theta(w_k) + \theta_g(w_k) \geq \epsilon > 0$ *for all k. There exists* $\Delta_a > 0$ *depending only on ϵ and on the values of the filter entries, such that, if*

$$0 < \Delta_k \leq \Delta_a,$$

then $w(\Delta_k)$ is in step 5 acceptable to the filter (with w_k considered in the filter when $m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa\theta(w_k)^2$).

Proof. Since $0 < \gamma_{\mathcal{F}} < 1/2 < 1$, we have from Lemma 4 that

$$\theta_{\mathcal{F}} = \min_{\mathcal{F}}(1 - \gamma_{\mathcal{F}})\theta^f > 0.$$

Consider first the case where $\theta(w_k) \geq \epsilon/2$. Then $w_k(\Delta_k)$ is acceptable to the filter (with w_k considered in the filter when $m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa\theta(w_k)^2$) if

$$\theta(w_k(\Delta_k)) \leq \frac{1}{2}\min\{\theta_{\mathcal{F}}, (1-\gamma_{\mathcal{F}})\epsilon/2\} < \min\{\theta_{\mathcal{F}}, (1-\gamma_{\mathcal{F}})\epsilon/2\}. \qquad (27)$$

We also know from Lemma 6 that

$$\theta(w_k(\Delta_k)) \leq (M_h + M_c)\Delta_k^2$$

holds for $\Delta_k \leq \Delta_r$. Thus, (27) is satisfied for $\Delta_k \leq \Delta_a^{(1)}$ with $\Delta_a^{(1)} > 0$ depending only on $\theta_{\mathcal{F}}$, ϵ, M_h, M_c, $\gamma_{\mathcal{F}}$, and Δ_r.

Otherwise we have $\theta_g(w_k) \geq \epsilon/2$. If w_k is not considered in the filter in step 5, then a similar argument, with $\theta(w_k(\Delta_k)) \leq \frac{1}{2}\theta_{\mathcal{F}}$ instead of (27), shows that if $\Delta_k \leq \Delta_a^{(1)}$ then $w_k(\Delta_k)$ is acceptable to the filter. Moreover $w_k(\Delta_k)$ is also acceptable, with w_k considered in the filter when $m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa\theta(w_k)^2$, if, in addition,

$$\theta_g(w_k(\Delta_k)) - \theta_g(w_k) < -\gamma_{\mathcal{F}}\theta(w_k). \qquad (28)$$

In the rest of the proof we show how this bound can be achieved for sufficiently small Δ_k. Since step 5 is reached, we know that

$$\theta(w_k) \leq \gamma_2\Delta_k^{1+\beta}.$$

On the other hand, we obtain from $\theta_g(w_k) \geq \epsilon/2$ and Lemma 1 with $0 < \Delta_k \leq \Delta_{ub}$ that

$$\theta_g(w_k(\Delta_k)) - \theta_g(w_k) \leq -(1-\sigma)\alpha_k^t\epsilon/2 + M_g\Delta_k^2.$$

Hence we need to show that

$$-(1-\sigma)\alpha_k^t\epsilon/2 + M_g\Delta_k^2 < -\gamma_{\mathcal{F}}\gamma_2\Delta_k^{1+\beta}.$$

Since $\|s_k^n\|$ and $\|s_k^t\|$ are bounded by a constant M_s and $\alpha_k^t = \min\{1, \frac{\Delta_k}{\|s_k^n\|}, \frac{\Delta_k}{\|s_k^t\|}\}$, we have for all $\Delta_k \leq M_s$, that $\alpha_k^t \geq \Delta_k/M_s$. Thus (28) holds if

$$M_g \Delta_k + \gamma_{\mathcal{F}} \gamma_2 \Delta_k^\beta \leq \frac{(1-\sigma)\epsilon}{4M_s} < \frac{(1-\sigma)\epsilon}{2M_s},$$

which in turn holds for all $\Delta_k \leq \Delta_a^{(2)}$ with $\Delta_a^{(2)} > 0$ depending only on ϵ, M_g, $\gamma_{\mathcal{F}}$, γ_2, β, σ, M_s, and Δ_{ub}. Taking $\Delta_a = \min\{\Delta_a^{(1)}, \Delta_a^{(2)}\}$ concludes the proof. $\square$

Secondly we look at the case where only the measure of optimality is bounded away from zero, but where we impose a condition relating $\theta(w_k)$ and Δ_k.

Lemma 8. *Suppose that for given $\epsilon > 0$*

$$\theta_g(w_k) \geq \epsilon \quad and \quad \theta(w_k) > \frac{\Delta_k}{2} \min\{\gamma_1, \gamma_2(\Delta_k/2)^\beta\}. \tag{29}$$

Then there exists $\Delta_f > 0$ such that, if

$$0 < \Delta_k \leq \Delta_f,$$

then $w(\Delta_k)$ is in step 5 acceptable to the filter (with w_k considered in the filter when $m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa \theta(w_k)^2$).

Proof. Since, by Lemma 5, w_k is acceptable to the filter, $w_k(\Delta_k)$ is acceptable to the filter (with w_k considered in the filter when $m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa \theta(w_k)^2$) if

$$\theta(w_k(\Delta_k)) \leq \theta(w_k)$$

and

$$\theta_g(w_k(\Delta_k)) < \theta_g(w_k) - \gamma_{\mathcal{F}} \theta(w_k). \tag{30}$$

We know from Lemma 6 that, if $\Delta_k \leq \Delta_r$,

$$\theta(w_k(\Delta_k)) \leq (M_h + M_c)\Delta_k^2.$$

Hence, $\theta(w_k(\Delta_k)) \leq \theta(w_k)$ is ensured by the second inequality in (29) if in addition

$$(M_h + M_c)\Delta_k \leq \frac{1}{2} \min\{\gamma_1, \gamma_2(\Delta_k/2)^\beta\}. \tag{31}$$

Moreover, when $0 < \Delta_k \leq \Delta_{ub}$, we have by Lemma 1 and the first inequality in (29) that

$$\theta_g(w_k(\Delta_k)) - \theta_g(w_k) \leq -\alpha_k^t(1-\sigma)\epsilon + M_g\Delta_k^2.$$

We have pointed out before that $\alpha_k^t \geq \Delta_k/M_s$ for all $\Delta_k \leq M_s$, see the end of the proof of Lemma 7. So,

$$\theta_g(w_k(\Delta_k)) - \theta_g(w_k) \leq \Delta_k \left(-\frac{(1-\sigma)\epsilon}{M_s} + M_g\Delta_k \right).$$

Since we are concerned with step 5 of the algorithm, we know that $\theta(w_k) \leq \gamma_2 \Delta_k^{1+\beta}$, see Remark 3, (26). Hence, we obtain (30) whenever

$$M_g \Delta_k + \gamma_{\mathcal{F}} \gamma_2 \Delta_k^{\beta} \leq \frac{(1-\sigma)\epsilon}{2M_s} < \frac{(1-\sigma)\epsilon}{M_s}. \tag{32}$$

The requirements $0 < \Delta_k \leq \min\{\Delta_r, \Delta_{ub}\}$, (31) and (32) on Δ_k are obviously satisfied if $0 < \Delta_k \leq \Delta_f$ with some constant $\Delta_f > 0$. $\square$

Now we are ready to derive asymptotic results. We appeal first to a commonly used argument in filter convergence proofs to show that $\liminf_{k \to \infty} \theta(w_k) = 0$ when infinitely many iterates are added to the filter.

Lemma 9. *From the moment that w_k is added to the filter, the filter always contains an entry that dominates w_k.*

Proof. Since w_k dominates w_k, the assertion is trivial as long as w_k remains in the filter. If w_k is removed from the filter then it is replaced by an iterate $w_{k'}$, $k' > k$, that dominates w_k. Thus, the assertion remains true as long as $w_{k'}$ stays in the filter. If $w_{k'}$ is removed from the filter then it is replaced by an iterate $w_{k''}$, $k'' > k'$, that dominates $w_{k'}$ and thus also dominates w_k by the transitivity of dominance. Thus, the result follows inductively. $\square$

Lemma 10. *Suppose there are infinitely many points added to the filter. Then there exists a subsequence $\{k_i\}$ such that w_{k_i} is added to the filter and*

$$\lim_{i \to \infty} \theta(w_{k_i}) = 0. \tag{33}$$

Proof. Let infinitely many points be added to the filter and set

$$\mathcal{A} = \{k : w_k \text{ is added to the filter}\}.$$

Assume (for deriving a contradiction) that the assertion is wrong. Since by Lemma 4 holds $\theta(w_k) > 0$ for all $k \in \mathcal{A}$, we then find $\epsilon > 0$ with $\theta(w_k) \geq \epsilon$ for all $k \in \mathcal{A}$.

For $k \in \mathcal{A}$, define the square

$$S_k = [\theta(w_k) - \gamma_{\mathcal{F}}\epsilon, \theta(w_k)] \times [\theta_g(w_k) - \gamma_{\mathcal{F}}\epsilon, \theta_g(w_k)].$$

We prove that for all $k, l \in \mathcal{A}$ with $k > l$ it holds

$$S_k \cap S_l = \emptyset. \tag{34}$$

In fact, at the time where w_k is added to the filter, the filter contains an entry $w_{l'}$ that dominates w_l according to Lemma 9. In addition, w_k is acceptable to the filter by Lemma 5, so that at least one of the following inequalities (35) or (36) holds:

$$\theta(w_k) < \theta(w_{l'}) - \gamma_{\mathcal{F}}\theta(w_{l'}) \leq \theta(w_l) - \gamma_{\mathcal{F}}\epsilon \tag{35}$$

$$\text{or} \quad \theta_g(w_k) < \theta_g(w_{l'}) - \gamma_{\mathcal{F}}\theta(w_{l'}) \leq \theta_g(w_l) - \gamma_{\mathcal{F}}\epsilon. \tag{36}$$

In either case, this implies (34). Therefore, all these infinitely many squares S_k, $k \in \mathcal{A}$, with area $(\gamma_{\mathcal{F}}\epsilon)^2$ are disjoint. Since the sequence $\{\theta(w_k), \theta_g(w_k)\}$ is bounded, we obtain the desired contradiction. $\square$

Remark 4. Note that Lemma 10 asserts (33) only for *some particular* subsequence $\{w_{k_i}\}$ of iterates added to the filter and not for *any* such subsequence. The reason is that acceptability of a pair does not imply acceptability of a dominated pair. In fact, let $P_1 = (1, 1)$ and $P_2 = (1/2, 1)$. Then P_2 is acceptable to P_1 and dominates P_1; nevertheless, all points in the set $[1 - \gamma_\mathcal{F}, \infty) \times [1 - \gamma_\mathcal{F}, 1 - \gamma_\mathcal{F}/2)$ *are* acceptable to P_2, but *are not* acceptable to P_1. Therefore, if P_1 is in the filter and P_2 enters the filter, then, by dominance, P_1 is removed, and this can result in points that are acceptable to the updated filter, but were not acceptable to the filter before the update.

If required, this effect can be circumvented in several ways. The easiest approach is to never remove dominated entries from the filter. Then the above proof can be easily modified to establish that (33) holds for any infinite subsequence of iterates that are added to the filter. An alternative to derive this stronger result, if one wishes to remove dominated filter entries, can also be obtained by slightly modifying the filter acceptance test, see [7] and [10, §15.5]. In fact, if we require

$$\text{either} \quad \theta^f - \theta(w) > \gamma_\mathcal{F}\theta^f \quad \text{or} \quad \theta_g^f - \theta_g(w) > \gamma_\mathcal{F}\theta(w),$$

then acceptability to a pair implies acceptability to all dominated pairs and it is straightforward to prove that (33) holds for any infinite subsequence of iterates added to the filter, see [10, Lem. 15.5.2]. □

Our next step is to show that in the case where infinitely many iterates are added to the filter there exists a subsequence of iterates that converges to a KKT-point. In fact, our previous result $\liminf_{k \to \infty} \theta(w_k) = 0$ can be extended to $\liminf_{k \to \infty} \theta(w_k) + \theta_g(w_k) = 0$. Since iterates are added to the filter only if restoration is invoked or in step 8, the sequence $\{k_i\}$ of Lemma 10 must contain either a subsequence where restoration is invoked, or a subsequence where the iterates are added to the filter in step 8. We start with the first case.

Lemma 11. *Suppose that there exists an infinite sequence $\{k_i\}$ of iterations at which restoration is invoked and for which holds that*

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

Then $\{k_i\}$ contains a subsequence $\{k_j'\}$ with

$$\lim_{j \to \infty} \theta(w_{k_j'}) = 0, \qquad \lim_{j \to \infty} \theta_g(w_{k_j'}) = 0.$$

Proof. Let k_i be a subsequence where restoration is invoked for every k_i (and thus w_{k_i} is added to the filter) such that $\lim_{i \to \infty} \theta(w_{k_i}) = 0$. Since the restoration is invoked it must hold that

$$\theta_{k_i} > \Delta_{k_i} \min\{\gamma_1, \gamma_2 \Delta_{k_i}^\beta\}. \tag{37}$$

Therefore, we have

$$0 = \lim_{i \to \infty} \theta_{k_i} = \lim_{i \to \infty} \Delta_{k_i}$$

and thus can find $K_0 > 0$ such that $\Delta_{k_i} < \nu\Delta_{\min}$ for all $k_i \geq K_0$ with $\Delta_{\min}$ from Lemma 3.iii, with $\nu \in (0, 1)$. We show next that $\Delta_{k_i-1} \leq 2\Delta_{k_i}$ for all $k_i \geq K_0$ which then yields

$$0 = \lim_{i\to\infty} \theta_{k_i} = \lim_{i\to\infty} \Delta_{k_i} = \lim_{i\to\infty} \Delta_{k_i-1}. \tag{38}$$

In fact, $\Delta_{k_i} < \nu\Delta_{\min}$ for $k_i \geq K_0$ shows that $\Delta_{k_i} = \Delta_{k_i}^{in}$ for $k_i \geq K_0$, since, by Lemma 3.iii, step 2 and step 3 yield only $\Delta_{k_i} \neq \Delta_{k_i}^{in}$ if $\Delta_{k_i}^{in} > \Delta_{\min}$. But then the result of step 2 and step 3 would be a radius $\Delta_{k_i} > \nu\Delta_{\min}$, which is not the case for $k_i \geq K_0$. Thus, we have $\Delta_{k_i} = \Delta_{k_i}^{in}$ for $k_i \geq K_0$ and conclude that $\Delta_{k_i} \geq \Delta_{k_i-1}/2$ for all $k_i \geq K_0$. Thus, (38) holds.

By (38) and Lemma 3.iii we know that for large enough i step 2 and step 3 do not change $\Delta_{k_i-1}^{in}$ and $\Delta_{k_i}^{in}$. Thus, we find $K_1 \geq K_0$ with

$$\Delta_{k_i-1} = \Delta_{k_i-1}^{in}, \quad \Delta_{k_i} = \Delta_{k_i}^{in} \quad \text{for all } k_i - 1 \geq K_1. \tag{39}$$

We show next that step 5 is reached in all iterations $k_i - 1 \geq K_1$. In fact, otherwise the restoration procedure is called in iteration $k_i - 1$. Thus, we have $\Delta_{k_i}^{in} = \Delta_{k_i-1}$ and consequently $\Delta_{k_i} = \Delta_{k_i-1}$ by (39). Since by our assumption the restoration is invoked in iteration $k_i - 1$, by using $\Delta_{k_i} = \Delta_{k_i-1}$ the outcome of the restoration is an iterate w_{k_i} with

$$\theta_{k_i} \leq \Delta_{k_i-1} \min\{\gamma_1, \gamma_2\Delta_{k_i-1}^{\beta}\} = \Delta_{k_i} \min\{\gamma_1, \gamma_2\Delta_{k_i}^{\beta}\},$$

which contradicts (37). Hence, step 5 is reached for all $k_i - 1 \geq K_1$ and thus in particular

$$\theta_{k_i-1} \leq \Delta_{k_i-1} \min\{\gamma_1, \gamma_2\Delta_{k_i-1}^{\beta}\}.$$

For the purpose of deriving a contradiction, suppose that $\theta_g(w_{k_i}) \geq \epsilon > 0$ for $k_i \geq K_2$ with some sufficiently large $K_2 \geq K_1$. We show next that then there exists $K_3 \geq K_2$ such that

$$\theta_g(w_{k_i-1}) \geq \epsilon/2 \quad \text{for all } k_i \geq K_3. \tag{40}$$

In fact, we have either $w_{k_i} = w_{k_i-1}$ or $w_{k_i} = w_{k_i-1}(\Delta_{k_i-1})$. In the first case (40) is obvious since then $\theta_g(w_{k_i-1}) = \theta_g(w_{k_i}) \geq \epsilon$ for all $k_i \geq K_2$. In the case $w_{k_i} = w_{k_i-1}(\Delta_{k_i-1})$ it follows from Lemma 1 that for $0 < \Delta_{k_i-1} \leq \Delta_{ub}$

$$\theta_g(w_{k_i}) = \theta_g(w_{k_i-1}(\Delta_{k_i-1})) \leq (1 - \alpha_{k_i-1}^t(1 - \sigma))\theta_g(w_{k_i-1}) + M_g\Delta_{k_i-1}^2,$$

and thus for $k_i \geq K_2$

$$\epsilon \leq \theta_g(w_{k_i}) \leq \theta_g(w_{k_i-1}) + M_g\Delta_{k_i-1}^2.$$

We can therefore conclude from (38) that (40) holds for $K_3 \geq K_2$ large enough.

Next, we show that step 7 must be reached for all iterations $k_i - 1$ with $k_i - 1 \geq K_4$ and $K_4 \geq K_3$ large enough. In fact, let Δ_f be the bound of Lemma 8 corresponding to $\epsilon/2$ instead of ϵ. By (38) we can find $K_4 \geq K_3$ such that $\Delta_{k_i-1} \leq \Delta_f$ holds for all $k_i - 1 \geq K_4$. Now assume that step 7 is not reached in iteration $k_i - 1 \geq K_4$. Then step

5 is followed by step 11 and thus $\theta_{k_i} = \theta_{k_i-1}$, and, using (39), $\Delta_{k_i} = \Delta_{k_i}^{in} = \Delta_{k_i-1}/2$. Therefore, by (37),

$$\theta_{k_i-1} > \frac{\Delta_{k_i-1}}{2} \min\{\gamma_1, \gamma_2(\Delta_{k_i-1}/2)^\beta\}.$$

Hence, we obtain from Lemma 8 that $w_{k_i-1}(\Delta_{k_i-1})$ was acceptable to the filter in step 5, since $k_i - 1 \geq K_4$ ensures $\Delta_{k_i-1} \leq \Delta_f$. Therefore, step 5 would not have branched to step 11 as assumed. Hence, step 7 is always reached in all iterations $k_i - 1 \geq K_4$.

We conclude the proof by showing the existence of $K_5 \geq K_4$ such that step 9 is reached for all iterations $k_i - 1$ with $k_i - 1 \geq K_5$. This assertion leads to a contradiction: by (39) and steps 9, 10 we have $\Delta_{k_i} = \Delta_{k_i}^{in} \geq \Delta_{k_i-1}$, $w_{k_i} = w_{k_i-1}(\Delta_{k_i-1})$. Thus, we obtain by Lemma 6 for all $\Delta_{k_i-1} \leq \Delta_r$ (which holds by (38) for all i large enough)

$$\theta_{k_i} = \theta(w_{k_i-1}(\Delta_{k_i-1})) \leq (M_h + M_c)\Delta_{k_i-1}^2 \leq (M_h + M_c)\Delta_{k_i}^2.$$

This contradicts (37) and (38).

It remains to show that step 9 is eventually reached in all iterations $k_i - 1$ with $k_i - 1 \geq K_5$, $K_5 \geq K_4$ large enough. We note that by (40) and Lemma 1 for all $k_i \geq K_4$

$$m_{k_i-1}(w_{k_i-1}) - m_{k_i-1}(w_{k_i-1}(\Delta_{k_i-1})) \geq \alpha_{k_i-1}^t(1-\sigma)\frac{\epsilon}{2} \geq \Delta_{k_i-1}(1-\sigma)\frac{\epsilon}{2M_s}.$$

Hereby, we use again the fact that $\alpha_{k_i-1}^t \geq \Delta_{k_i-1}/M_s$ if $\Delta_{k_i-1} \leq M_s$, which holds by (38) possibly after increasing K_4. On the other hand, we have

$$|m_{k_i-1}(w_{k_i-1}) - m_{k_i-1}(w_{k_i-1}(\Delta_{k_i-1})) - \theta_g(w_{k_i-1}) + \theta_g(w_{k_i-1}(\Delta_{k_i-1}))|$$
$$= O(\Delta_{k_i-1}^2).$$

The last two inequalities show that $\rho_{k_i-1} \to 1$ and hence there exists $K_5 \geq K_4$ such that step 9 is reached in all iterations $k_i - 1$ with $k_i - 1 \geq K_5$.

As we have already seen, this leads to a contradiction with $\theta_g(w_{k_i}) \geq \epsilon$ for all $k_i \geq K_2$. The proof is therefore completed since there exists a subsequence $\{k_j'\} \subset \{k_i\}$ for which $\lim_{j\to\infty} \theta(w_{k_j'}) = \lim_{j\to\infty} \theta_g(w_{k_j'}) = 0$. $\qquad\square$

The other situation is where the sequence $\{k_i\}$ of Lemma 10 contains a subsequence, where the iterates are added to the filter in step 8. As in the previous Lemma we have the following result.

Lemma 12. *Suppose that there exists an infinite sequence $\{k_i\}$ of iterations for which w_{k_i} is added to the filter in step 8 and, in addition, $\lim_{i\to\infty} \theta(w_{k_i}) = 0$. Then $\{k_i\}$ contains a subsequence $\{k_j'\}$ such that*

$$\lim_{j\to\infty} \theta(w_{k_j'}) = 0, \qquad \lim_{j\to\infty} \theta_g(w_{k_j'}) = 0.$$

Proof. Let $\{k_i\}$ be a sequence of iterations such that w_{k_i} is added to the filter in step 8 and $\lim_{i\to\infty} \theta(w_{k_i}) = 0$. Suppose now that $\theta_g(w_{k_i}) \geq \epsilon > 0$ for all $k_i \geq K_0$ for some $K_0 \geq 0$. By Lemma 1 and since w_{k_i} is added to the filter in step 8 we have

$$\alpha_{k_i}^t(1-\sigma)\epsilon \leq m_{k_i}(w_{k_i}) - m_{k_i}(w_{k_i}(\Delta_{k_i})) < \kappa\theta(w_{k_i})^2.$$

Thus, we obtain $\alpha_{k_i}^t \to 0$ and consequently $\Delta_{k_i} \to 0$. In particular, $\alpha_{k_i}^t \geq \Delta_{k_i}/M_s$ for large enough i, and since the restoration procedure is not called, we have $\theta(w_{k_i}) \leq \gamma_2 \Delta_{k_i}^{1+\beta}$ and conclude that

$$\Delta_{k_i}(1-\sigma)\epsilon/M_s \leq m_{k_i}(w_{k_i}) - m_{k_i}(w_{k_i}(\Delta_{k_i})) < \kappa(\gamma_2 \Delta_{k_i}^{1+\beta})^2$$

which is a contradiction to $\Delta_{k_i} \to 0$. □

We summarize both situations in the next theorem.

Theorem 1. *Suppose that infinitely many iterates are added to the filter. Then there exists a subsequence $\{k_j\}$ such that*

$$\lim_{j\to\infty} \theta(w_{k_j}) = 0, \qquad \lim_{j\to\infty} \theta_g(w_{k_j}) = 0.$$

Proof. By Lemma 10 there exists a sequence $\{k_i\}$ of iterates such that w_{k_i} is added to the filter and $\lim_{i\to\infty} \theta(w_{k_i}) = 0$. As we have already observed there exists either a subsequence $\{k_j'\}$ of $\{k_i\}$ such that $w_{k_j'}$ are added to the filter in the restoration or a subsequence $\{k_j'\}$ of $\{k_i\}$ such that $w_{k_j'}$ are added to the filter in step 8. In the first case the assertion follows from Lemma 11; in the second case from Lemma 12. □

It remains to consider the case where the algorithm runs infinitely but the filter is left with a finite number of entries.

Theorem 2. *Suppose that the algorithm runs infinitely and only finitely many iterates are added to the filter. Then*

$$\lim_{k\to\infty} \theta(w_k) = 0, \qquad \liminf_{k\to\infty} \theta_g(w_k) = 0.$$

Proof. The assumption says that for $k \geq K$, with K large enough, no further filter entry is added. Hence, the filter contains for all $k \geq K$ the same finitely many entries, and the restoration is never invoked. Thus, all new iterates $w_{k+1} \neq w_k$ are computed in step 10. We now show that step 10 is reached infinitely many times.

In fact, step 5 is reached in each iteration, and, by Lemma 7, step 7 is reached after finitely many reductions of Δ_k in step 11. Again, step 8 is reached after finitely many reductions of Δ_k. In fact, if $\theta(w_k) > 0$ then clearly

$$m_k(w_k) - m_k(w_k(\Delta_k)) < \kappa\theta(w_k)^2$$

for Δ_k sufficiently small and step 8 is reached. Otherwise, $\theta(w_k) = 0$ and $\theta_g(w_k) > 0$ and therefore $\rho_k \geq \eta$ for all Δ_k small enough (we may apply exactly the same arguments as at the end of the proof of Lemma 12). So, step 10 is always reached after finitely many reductions of Δ_k, producing always new iterates.

Since no further entry is added to the filter we know, cf. step 8, that in step 10 it always holds that

$$\theta_g(w_k) - \theta_g(w_{k+1}) \geq \eta(m_k(w_k) - m_k(w_k(\Delta_k))) \geq \eta\kappa\theta(w_k)^2.$$

Since this holds for all successful steps and $\{\theta_g(w_k)\}$ is bounded, we conclude that

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

Now assume that $\theta_g(w_k) \geq \epsilon > 0$ for all $k \geq K$ and some $\epsilon > 0$. Since the filter entries do not change for $k \geq K$, the test in step 5 is passed whenever $\Delta_k \leq \Delta_a$ (cf. Lemma 7). Also, since $\theta_g(w_k) \geq \epsilon > 0$, we obtain as before that $\rho_k \geq \eta$ whenever $\Delta_k \leq \Delta'$ for some $\Delta' > 0$. Finally, we know by Lemma 3.iii that for $\Delta_k^{in} \leq \Delta_{\min}$ steps 2 and 3 yield $\Delta_k = \Delta_k^{in}$. Hence, we see that $\Delta_k \geq \delta \stackrel{\text{def}}{=} \min\{\Delta_a/2, \Delta'/2, \nu\Delta_{\min}\} > 0$ for $k \geq K$. Thus, step 10 is reached for all successful steps with $\Delta_k \geq \delta > 0$ and we have, as above,

$$\theta_g(w_k) - \theta_g(w_{k+1}) \geq \eta(m_k(w_k) - m_k(w_k(\Delta_k))) \geq \eta(1-\sigma)\epsilon\alpha_k^t$$

$$\geq \eta(1-\sigma)\epsilon \min\left\{\frac{\delta}{M_s}, 1\right\}$$

where M_s is as before a uniform upper bound on $\max\{\|s_k^t\|, \|s_k^n\|\}$. This is again a contradiction to the boundedness of $\theta_g(w_k)$ and the proof is complete.   □

The main result is obtained by combining Theorems 1 and 2:

Corollary 1. *Under Assumption 1, the sequence of iterates $\{w_k\}$ generated by the primal-dual interior-point method (Algorithm 1) satisfies*

$$\liminf_{k\to\infty} \ \theta(w_k) + \theta_g(w_k) = 0.$$

6. A restoration algorithm

In this section we present restoration algorithms that can be used in step 4 of the primal-dual interior-point filter method (Algorithm 1). The purpose of a restoration algorithm is to find a point $w_{k+1} \in \mathcal{N}(\gamma, M)$ acceptable to the filter and such that $\theta(w_{k+1}) \leq \Delta_{k+1} \min\{\gamma_1, \gamma_2\Delta_{k+1}^\beta\}$ with $\Delta_{k+1} = \Delta_k$. Thus, the purpose of a restoration algorithm is to decrease the value of $\theta(w) = \theta_h(w) + \theta_c(w)$. To achieve this goal we introduce the function

$$\theta_2(w) \stackrel{\text{def}}{=} \frac{1}{2}\left(\theta_h(w)^2 + \theta_c(w)^2\right) = \frac{1}{2}\left(\|h(x)\|^2 + \|Xz - \mu e\|^2\right).$$

6.1. A Restoration Algorithm based on the KKT-Newton-System

The normal step s^n computed from (11) is a descent direction for $\theta_2(w)$. In fact,

$$\nabla\theta_2(w)^T s^n = (Xz - \mu e)^T (Z\Delta x^n + X\Delta z^n) + h(x)^T \nabla h(x)^T \Delta x^n$$

$$= -(Xz - \mu e)^T (Xz - \mu e) - h(x)^T h(x).$$

Thus, $\nabla\theta_2(w)^T s^n = -2\theta_2(w)$, and s^n is, in fact, a descent direction for $\theta_2(w)$. One can also show using

$$(Xz - \mu e)^T (1-\sigma)\mu e = (1-\sigma)\mu(n\mu - n\mu) = 0$$

that the tangential step (12) yields $\nabla\theta_2(w)^T s^t = 0$. We summarize these two properties for future reference:

$$\nabla\theta_2(w)^T s^n = -2\theta_2(w), \qquad \nabla\theta_2(w)^T s^t = 0. \tag{41}$$

The restoration algorithm presented here works with the step framework $w(\Delta) = w + \alpha^n(\Delta)s^n + \alpha^t(\Delta)s^t$, where $\alpha^n(\Delta)$, s^n, $\alpha^t(\Delta)$, and s^t are given by (13), (11), (14), and (12), respectively. Several other restoration algorithms are plausible but we chose the following one because it is consistent with the step calculation of our primal-dual interior-point filter method.

Algorithm 2 (Restoration algorithm).

0. Choose $\xi, \nu \in (0, 1)$. Set $w_k^0 := w_k$, $\Delta_k^0 := \Delta_k$, $j := 0$ and start with step 4.
1. If $\theta(w_k^j) \leq \Delta_k \min\{\gamma_1, \gamma_2\Delta_k^\beta\}$ and w_k^j is acceptable to the filter then set $w_{k+1} := w_k^j$ and stop restoration.
2. Set $\mu_k^j := (x_k^j)^T z_k^j/n$ and compute the steps $s_k^{n,j}$ and $s_k^{t,j}$ by solving the linear systems (11) and (12), respectively, with $(w, \mu) = (w_k^j, \mu_k^j)$.
3. Compute $\Delta_k^j \in (0, \Delta_k^{in,j}]$ such that

$$x_k^j(\Delta) > 0, \;\; z_k^j(\Delta) > 0, \;\; X_k^j(\Delta)z_k^j(\Delta) \geq \gamma\mu_k^j(\Delta)e \quad \text{for all } \Delta \in [0, \Delta_k^j] \tag{42}$$

and such that Δ_k^j is not smaller than the largest value $\nu^r \Delta_k^{in,j}$, $r = 0, 1, \ldots$, having this property.
4. If

$$\theta_2(w_k^j) - \theta_2(w_k^j(\Delta_k^j)) \geq -\xi\nabla\theta_2(w_k^j)^T s_k^j(\Delta_k^j), \tag{43}$$

$$\theta_h(w_k^j(\Delta_k^j)) + \theta_\ell(w_k^j(\Delta_k^j)) \leq M\mu_k^j(\Delta_k^j), \tag{44}$$

then choose $\Delta_k^{in,j+1} \geq \Delta_k^j$, set $w_k^{j+1} := w_k^j(\Delta_k^j)$, $j := j + 1$, and return to step 1. Otherwise set $\Delta_k^{j+1} = \Delta_k^j/2$, $j := j + 1$, and repeat step 4.

This restoration algorithm terminates successfully in a finite number of iterations as we prove in our final theorem. In analogy to Assumption 1 we require

Assumption 2. Assumption 1 holds for w_k, $s_k(\Delta_k)$, μ_k replaced by w_k^j, $s_k^j(\Delta_k^j)$, μ_k^j, respectively.

Theorem 3. *Under Assumption 2, the restoration algorithm 2 terminates in a finite number of iterations.*

Proof. Let us first consider the well-definedness of the algorithm. Hereby, the only critical issues are the computation of the trial steps $s_k^{n,j}$ and $s_k^{t,j}$, which is possible by Assumption 2 (A3), and the computation of Δ_k^j in step 3, which is possible by Lemma 2. Thus, the algorithm is well-defined.

For the rest of the proof, assume that the restoration algorithm does not terminate finitely. Let

$$\theta_\mathcal{F} = \min_\mathcal{F}(1 - \gamma_\mathcal{F})\theta^f.$$

Since $\gamma_{\mathcal{F}} \in (0, 1/2)$, we have from Lemma 4 that $\theta_{\mathcal{F}} > 0$, and w_k^j is acceptable to the filter if $\theta(w_k^j) \leq 2\sqrt{\theta_2(w_k^j)} \leq \theta_{\mathcal{F}}/2 < \theta_{\mathcal{F}}$. This condition and $\theta(w_k^j) \leq \Delta_k \min\{\gamma_1, \gamma_2 \Delta_k^{\beta}\}$ are eventually satisfied if

$$\liminf_{j \to \infty} \theta_2(w_k^j) = 0. \tag{45}$$

Hence, if the restoration does not terminate finitely, then there exists an $\epsilon > 0$ with $\theta_2(w_k^j) \geq \epsilon$ for all j. We show that this uniform bound will lead to a contradiction. In fact, from (41), we have

$$\nabla \theta_2(w_k^j)^T s_k^j(\Delta) = \alpha_k^{n,j}(\Delta) \nabla \theta_2(w_k^j)^T s_k^{n,j} + \alpha_k^{t,j}(\Delta) \nabla \theta_2(w_k^j)^T s_k^{t,j}$$
$$= -2\alpha_k^{n,j}(\Delta) \theta_2(w_k^j).$$

Moreover, there exists a constant $M_2 > 0$ such that

$$\theta_2(w_k^j) - \theta_2(w_k^j(\Delta)) = -\nabla \theta_2(w_k^j)^T s_k^j(\Delta) - M_2 \|s_k^j(\Delta)\|^2,$$

which in turn, appealing to $\alpha_k^{t,j}(\Delta) \leq \alpha_k^{n,j}(\Delta)$, implies

$$\theta_2(w_k^j) - \theta_2(w_k^j(\Delta)) \geq -\nabla \theta_2(w_k^j)^T s_k^j(\Delta) - 2M_2 \alpha_k^{n,j}(\Delta)^2 (\|s_k^{n,j}\|^2 + \|s_k^{t,j}\|^2).$$

Hence, (43) holds for all $\alpha_k^{n,j}(\Delta)$ such that

$$2(1 - \xi)\alpha_k^{n,j}(\Delta)\theta_2(w_k^j) \geq 2M_2 \alpha_k^{n,j}(\Delta)^2 (\|s_k^{n,j}\|^2 + \|s_k^{t,j}\|^2),$$

i.e., for all $\alpha_k^{n,j}(\Delta)$ such that

$$\alpha_k^{n,j}(\Delta) \leq \bar{\alpha}_k^{n,j} \stackrel{\text{def}}{=} \min\left\{1, \frac{(1 - \xi)\theta_2(w_k^j)}{M_2(\|s_k^{n,j}\|^2 + \|s_k^{t,j}\|^2)}\right\}.$$

From Lemmas 2 and 3.iii, we see finally that (42), (43), and (44) are satisfied for all Δ_k^j such that

$$0 < \Delta_k^j \leq \min\left\{\Delta_{\min}, \bar{\alpha}_k^{n,j}\|s_k^{n,j}\|\right\},$$

showing that all these conditions are satisfied after finitely many reductions of Δ_k^j in step 4.

Now, if $\theta_2(w_k^j) \geq \epsilon > 0$ holds for all j, then, since $\max\{\|s_k^{n,j}\|, \|s_k^{t,j}\|\} \leq M_s$,

$$\alpha_k^{n,j}(\Delta_k^j) \geq \frac{1}{2}\min\left\{\frac{\Delta_{\min}}{\|s_k^{n,j}\|}, \bar{\alpha}_k^{n,j}\right\} \geq \bar{\alpha} > 0$$

for some $\bar{\alpha} > 0$, and we conclude that

$$\theta_2(w_k^j) - \theta_2(w_k^j(\Delta_k^j)) \geq 2\xi\bar{\alpha}\epsilon,$$

which yields a contradiction. Hence, we have (45) and the finite termination is proved. □

6.2. Other restoration procedures and the Wächter-Biegler example

The restoration algorithm 2 uses the same step computation as the interior-point filter method. Global convergence is attained as long as Assumptions 1 and 2 hold. In practice it may happen that Assumption (A3) is violated, affecting not only Algorithm 1, but also the restoration algorithm 2. We will briefly discuss here the main issues related with this situation and leave a more detailed study for a forthcoming paper.

For the purpose of this discussion, let us consider the problem proposed by Wächter and Biegler in [27], which takes the following form:

$$\min_{x \in \mathbb{R}^3} f(x) \quad \text{s.t.} \quad x_1^2 - x_2 + a = 0, \quad x_1 - x_3 - b = 0, \quad x_2, x_3 \geq 0, \quad (46)$$

with $a \in \mathbb{R}$ and $b \geq 0$. The particular form of the (sufficiently smooth) objective is not relevant since the interesting properties of the problem are generated by the constraints. For brevity, we set $h_1(x) = x_1^2 - x_2 + a$ and $h_2(x) = x_1 - x_3 - b$. The problem is non-degenerate, because $\nabla h(x)$ has full rank and its condition number is uniformly bounded on sets where x_1 is bounded. Nevertheless, it was shown in [27] that for every initial point $x^0 \in (-\infty, 0) \times (0, \infty)^2$ with $r := \frac{h_1(x^0)}{|h_2(x^0)|} \geq 0$ and $a - rb \leq \min\{0, -a/2\}$ any method fails to converge if it computes the new primal iterate x^{k+1} from the old point $x^k \in \mathbb{R} \times (0, \infty)^2$ by a step of the following form:

$$x^{k+1} = x^k + \sigma_k s^k \in \mathbb{R} \times (0, \infty)^2, \quad \sigma_k \in (0, 1],$$
$$\text{where } s^k \text{ satisfies} \quad h(x^k) + \nabla h(x^k)^T s^k = 0. \quad (47)$$

Both Algorithms 1 and 2 can be easily adapted to allow for x_1 as a free variable.

It is straightforward to see that Algorithm 1, when no restoration is called or when the restoration algorithm is of the form described in Section 6.1, is contained in the class of algorithms described in the previous paragraph. In fact we have run this problem using our Matlab implementation of Algorithm 1, combined with Algorithm 2 for the restoration, and we have observed bad numerical behavior. We have seen that the iterates approached a point where Assumption (A3) is violated and that this was the cause for such bad numerical performance.

This situation may be avoided by using a restoration algorithm that terminates successfully even if assumption (A3) is violated. To illustrate this point, we have considered a restoration procedure that does not belong to the class (47). This alternative is based on the optimization problem in the variables x and z

$$\min \frac{1}{2}\left(\|h(x)\|^2 + \|Xz - \mu^c e\|^2\right) \quad \text{s.t.} \quad x \geq 0, \quad z \geq 0, \quad (48)$$

where $\mu^c \stackrel{\text{def}}{=} \mu_k$ denotes the value of μ when restoration is entered. Analogously, we denote by $w^c = (x^c, y^c, z^c) \stackrel{\text{def}}{=} (x_k, y_k, z_k)$ the iterate for which restoration is entered. Note that if there exists a strictly feasible point of (46), i.e., a point x such that $h(x) = 0$ and $x > 0$, then we can choose $z = \mu^c X^{-1} e$ and obtain a global minimizer of (48). Conversely, if there exists a strictly feasible point then any global minimizer (x, z) satisfies $h(x) = 0$, $x, z > 0$, and $Xz - \mu^c e = 0$. Therefore, any solver that is able to find a global minimizer $(\bar{x}, \bar{z})$ of (48) for which, in addition,

$$\theta_h(\bar{x}, y^c, \bar{z}) + \theta_\ell(\bar{x}, y^c, \bar{z}) \leq M\mu^c \quad (49)$$

holds, provides a valid restoration procedure (with $w_{k+1} = (\bar{x}, y^c, \bar{z})$). In order to include a device to enforce (49) we tried (among many other possible options) to augment the objective function in (48) by the penalization term $\frac{\rho_1}{2} \|\nabla_x \ell(x, y^c, z)\|^2$. In addition, it can be advantagous to add the regularization term $\frac{\rho_2}{2} \|(x, z) - (x^c, z^c)\|^2$ to make the Hessian of the restoration problem positive definite. Both parameters ρ_1, ρ_2 are chosen small, i.e., $0 \leq \rho_i \ll 1$. Since these modifications may lead to non-interior solutions, i.e., $x_i = 0$ or $z_i = 0$ for some i, we move the nonnegativity constraints slightly. The resulting problem is

$$\min \frac{1}{2} \left(\|h(x)\|^2 + \|Xz - \mu^c e\|^2 + \rho_1 \|\nabla_x \ell(x, y^c, z)\|^2 + \rho_2 \|(x, z) - (x^c, z^c)\|^2 \right)$$

$$\text{s.t.} \quad x, z \geq \delta,$$

where $\delta > 0$ is very small and should be chosen, e.g., depending on the value of μ^c. If required, the value of ρ_2 can be adjusted during the minimization process. The value of ρ_1 is kept small, but large enough to ensure (49). We applied Bertsekas' projected Newton method to solve this problem in the context of (46). We have numerically observed that Algorithm 1 using this alternative restoration procedure (adapted to the case where x_1 is unrestricted) was able to successfully solve problem (46).

The use of new alternatives for the restoration procedure, including the one presented above, is subject of ongoing research. Another alternative is for instance the use of the restoration algorithm 2 but with a suitable regularization for the matrix $F'_{\sigma\mu}(w)$. One should also consider a modification of the Algorithm 1 so that restoration is also called if $\|F'_{\sigma\mu_k}(w_k)^{-1}\|$ exceeds a prescribed, very large bound C' – larger than we expect or wish the constant C to be in assumption (A3). If the restoration procedure is able to find a new point w_{k+1} at which, besides the requirements of step 4 in Algorithm 1, $F'_{\sigma\mu_{k+1}}(w_{k+1})$ satisfies (A3) with a reasonable value of $C < C'$, then the chances of satisfying assumption (A3) will be improved so that the interior-point filter method can continue successfully.

7. Concluding remarks

The filter mechanism has been used for the first time to globalize primal-dual interior-point methods. Global convergence to first-order critical points has been proved, and the main result has been reported in Corollary 1.

The combination of interior-point and filter ideas led to a new class of algorithms. We have already tested our primal-dual interior-point filter method in Matlab for QP problems and small-scale NLP problems. The results are encouraging but there is still some work ahead. We are currently working on a fortran 90 implementation of our primal-dual interior-point filter method and we plan to report numerical results in a future paper.

This paper is hopefully a first step in this challenging topic. Several issues need to be addressed and better understood, and among them we highlight the following.

The new primal-dual interior-point algorithm uses a 2D filter: one dimension for feasibility and centrality combined and the other for the size of gradient of the Lagrangian (with complementarity added). An open question is the use of 3D filters. In a 3D filter,

one uses the first dimension for feasibility, the second for centrality, and the third for the size of the gradient of the Lagrangian.

As we have mentioned in the introduction, the rate of local convergence is well studied in the literature for primal-dual interior-point methods under the standard assumptions, and even in cases where some of these standard assumptions do not hold. The open issue in interior-point filter methods is whether the globalization scheme (where the filter plays an important role) becomes locally inactive to allow fast convergence rates.

Another interesting topic for future research is the choice of alternatives for the components used in the filter. As already mentioned, it would be desirable to replace the optimality measure θ_g by a function that reflects better the goal of minimizing f. Essentially, an appropriate candidate should be a function for which the tangential step s^t yields a fraction of Cauchy decrease close to the quasi-central path.

Appendix

The following lemma measures the decrease on complementarity obtained by the new iterate $w(\Delta)$ and is needed to prove Lemmas 1 and 2.

Lemma 13. *For all $\Delta > 0$ and all $i = 1, \ldots, n$ it holds*

$$x_i(\Delta)z_i(\Delta) \leq (1 - \alpha^n(\Delta))x_i z_i + \left(\alpha^n(\Delta) - \alpha^t(\Delta)(1 - \sigma)\right)\mu + 4\Delta^2, \qquad (50)$$

$$x_i(\Delta)z_i(\Delta) \geq (1 - \alpha^n(\Delta))x_i z_i + \left(\alpha^n(\Delta) - \alpha^t(\Delta)(1 - \sigma)\right)\mu - 4\Delta^2, \qquad (51)$$

$$\mu(\Delta) \leq \left(1 - \alpha^t(\Delta)(1 - \sigma)\right)\mu + 4\Delta^2,$$

$$\mu(\Delta) \geq \left(1 - \alpha^t(\Delta)(1 - \sigma)\right)\mu - 4\Delta^2. \qquad (52)$$

Proof. By the definition of s^n and s^t, we have

$$
\begin{aligned}
x_i(\Delta)z_i(\Delta) &= (x_i + \alpha^n(\Delta)\Delta x_i^n + \alpha^t(\Delta)\Delta x_i^t)(z_i + \alpha^n(\Delta)\Delta z_i^n + \alpha^t(\Delta)\Delta z_i^t) \\
&= x_i z_i + \alpha^n(\Delta)(z_i \Delta x_i^n + x_i \Delta z_i^n) + \alpha^t(\Delta)(z_i \Delta x_i^t + x_i \Delta z_i^t) \\
&\quad + (\alpha^n(\Delta)\Delta x_i^n + \alpha^t(\Delta)\Delta x_i^t)(\alpha^n(\Delta)\Delta z_i^n + \alpha^t(\Delta)\Delta z_i^t) \\
&= x_i z_i - \alpha^n(\Delta)(x_i z_i - \mu) - \alpha^t(\Delta)(1 - \sigma)\mu \\
&\quad + (x_i(\Delta) - x_i)(z_i(\Delta) - z_i).
\end{aligned}
$$

So, inequalities (50) and (51) follow from this derivation and

$$|x_i(\Delta) - x_i||z_i(\Delta) - z_i| \leq (2\Delta)^2.$$

Summing (50) and (51) over all i, dividing the result by n, and using $\mu = x^T z/n$, $\mu(\Delta) = x(\Delta)^T z(\Delta)/n$, yield (52). $\qquad\square$

We can now prove Lemmas 1 and 2.

Proof of Lemma 1

Proof. Denote by B_ℓ an upper bound for θ_ℓ, and by C_h and $C_{\ell'} > 1$ Lipschitz constants for ∇h and $\nabla^2_{xw}\ell$, respectively. We will prove Lemma 1 with

$$M_h = 2C_h, \qquad M_c = 8\sqrt{n},$$
$$M_\ell = 2C_{\ell'}, \qquad M_g = 4(1 + B_\ell C_{\ell'}) + 4C^2_{\ell'}\Delta^2_{ub}.$$

We note that $\nabla h(x)^T s_x(\Delta) = -\alpha^n(\Delta)h(x)$, and thus

$$\theta_h(w(\Delta)) = \|h(x(\Delta))\| = \left\| h(x) + \int_0^1 \nabla h(x + ts_x(\Delta))^T s_x(\Delta)dt \right\|$$

$$= \left\| (1 - \alpha^n(\Delta))h(x) + \int_0^1 \left(\nabla h(x + ts_x(\Delta)) - \nabla h(x) \right)^T s_x(\Delta)dt \right\|$$

$$\leq (1 - \alpha^n(\Delta))\theta_h(w) + C_h\|s_x(\Delta)\|^2 \int_0^1 tdt,$$

which proves (18).

Similarly, we have $\nabla^2_{xw}\ell(w)^T s(\Delta) = -\alpha^t(\Delta)\nabla_x\ell(w)$ and, as above, we get

$$\theta_\ell(w(\Delta)) \leq (1 - \alpha^t(\Delta))\theta_\ell(w) + \int_0^1 \|\nabla^2_{xw}\ell(w + ts(\Delta)) - \nabla^2_{xw}\ell(w)\|\|s(\Delta)\|dt,$$

which yields (20).

The estimate (19) follows from Lemma 13:

$$\pm\big(x_i(\Delta)z_i(\Delta) - \mu(\Delta)\big) \leq \pm\Big((1 - \alpha^n(\Delta))x_iz_i + \big(\alpha^n(\Delta) - \alpha^t(\Delta)(1 - \sigma)\big)\mu\Big) + 4\Delta^2$$

$$\mp \big(1 - \alpha^t(\Delta)(1 - \sigma)\big)\mu + 4\Delta^2$$

$$= \pm(1 - \alpha^n(\Delta))(x_iz_i - \mu) + 8\Delta^2.$$

Inequality (21) is derived by appealing to Lemma 13 and to the previously established inequality (20):

$$\theta_g(w(\Delta)) = \mu(\Delta) + \theta_\ell(w(\Delta))^2$$

$$\leq \big(1 - \alpha^t(\Delta)(1 - \sigma)\big)\mu + 4\Delta^2 + \big((1 - \alpha^t(\Delta))\theta_\ell(w) + 2C_{\ell'}\Delta^2\big)^2$$

$$\leq \big(1 - \alpha^t(\Delta)(1 - \sigma)\big)\theta_g(w) + \big(4 + 4(1 - \alpha^t(\Delta))\theta_\ell(w)\big)C_{\ell'}\Delta^2 + 4C^2_{\ell'}\Delta^4.$$

Finally, we have

$$m(w) - m(w(\Delta)) = \mu - \mu(\Delta) + (x(\Delta) - x)^T(z(\Delta) - z)/n$$

$$+ (1 - (1 - \alpha^t(\Delta))^2)\|\nabla_x\ell(w)\|^2$$

$$= \alpha^t(\Delta)(1 - \sigma)\mu + (1 - (1 - \alpha^t(\Delta))^2)\|\nabla_x\ell(w)\|^2$$

$$\geq \alpha^t(\Delta)(1 - \sigma)\theta_g(w),$$

and the proof is therefore completed. $\qquad\square$

Proof of Lemma 2

Proof. We will prove Lemma 2 with

$$\Delta_{\min} = \left\{ \frac{\sigma(1-\gamma)}{4(1+\gamma)C(M+n)}, \frac{\sigma M}{(M_h + M_\ell + 4M)C(M+n)} \right\}. \tag{53}$$

1. We first show that (23) holds for all $\Delta \in (0, \Delta_{\min}]$ with $\Delta_{\min}$ satisfying (53). From Lemma 13 we obtain

$$X(\Delta)z(\Delta) \geq \big(\gamma + (1-\gamma)\alpha^n(\Delta) - \alpha^t(\Delta)(1-\sigma)\big)\mu e - 4\Delta^2 e. \tag{54}$$

On the other hand, Lemma 13 also yields

$$\gamma\mu(\Delta) \leq \gamma\mu - \gamma\alpha^t(\Delta)(1-\sigma)\mu + 4\gamma\Delta^2.$$

Hence, $X(\Delta)z(\Delta) \geq \gamma\mu(\Delta)e$ holds whenever

$$4\Delta^2 \leq \frac{(1-\gamma)(\alpha^n(\Delta) - \alpha^t(\Delta)(1-\sigma))\mu}{1+\gamma}. \tag{55}$$

Since $\alpha^t(\Delta) \leq \alpha^n(\Delta)$, a sufficient condition for this inequality to hold is

$$\Delta^2 \leq \frac{\alpha^n(\Delta)\sigma(1-\gamma)\mu}{4(1+\gamma)},$$

which, by (13), is implied by

$$\Delta \leq \min\left\{ \sqrt{\frac{\sigma(1-\gamma)\mu}{4(1+\gamma)}}, \frac{\sigma(1-\gamma)\mu}{4(1+\gamma)\|s^n\|} \right\},$$

which in turn is true if

$$\Delta \leq \min\left\{ \sqrt{\frac{\sigma(1-\gamma)\mu}{4(1+\gamma)}}, \frac{\sigma(1-\gamma)}{4(1+\gamma)C(M+n)} \right\} \overset{\text{def}}{=} \delta_1(\mu),$$

since $\|s^n\| \leq C(M + (n^2 - n)^{1/2})\mu \leq C(M+n)\mu$.

On the other hand, we can also deduce that $\|s^t\| \leq C(M + n^{1/2})\mu \leq C(M+n)\mu$, and therefore

$$\delta \overset{\text{def}}{=} \max\{\|s^n\|, \|s^t\|\} \leq C(M+n)\mu. \tag{56}$$

We consider now two possible cases in order to show that (23) holds whenever

$$\min(\Delta, \delta) \leq \delta_1(\mu). \tag{57}$$

In the first case $\Delta \leq \delta$ we use that, as we have just shown, (23) holds provided $\Delta \leq \delta_1(\mu)$. Since $\Delta \leq \delta$ in this case, the inequality $\Delta \leq \delta_1(\mu)$ is equivalent to (57). In the case $\delta < \Delta$, we know that $\alpha^n(\Delta) = \alpha^t(\Delta) = 1$, $w(\Delta) = w(\delta)$, and $s(\Delta) = s(\delta)$. Thus, (23) is the same as $X(\delta)z(\delta) \geq \gamma\mu(\delta)$. One can apply the same algebraic arguments used

in the first paragraph of the proof to show that $X(\delta)z(\delta) \geq \gamma\mu(\delta)$ holds if $\delta \leq \delta_1(\mu)$. Now, since $\delta < \Delta$, we have that $\delta \leq \delta_1(\mu)$ is also equivalent to (57).

Hence, it remains to show (57) for $\Delta \in [0, \Delta_{\min}]$ with $\Delta_{\min}$ according to (53). Let

$$\mu_{\text{crit}}^1 \overset{\text{def}}{=} \frac{\sigma(1-\gamma)}{4(1+\gamma)C^2(M+n)^2}.$$

If $\mu \leq \mu_{\text{crit}}^1$, then $\delta_1(\mu)$ is given by its first expression, and (57) follows directly from (56), since by the definition of μ_{crit}^1 for $\mu \leq \mu_{\text{crit}}^1$ with (56) holds

$$\delta \leq C(M+n)\mu \leq \sqrt{C^2(M+n)^2\mu_{\text{crit}}^1\mu} = \sqrt{\frac{\sigma(1-\gamma)\mu}{4(1+\gamma)}} = \delta_1(\mu).$$

If $\mu > \mu_{\text{crit}}^1$, then $\delta_1(\mu)$ is given by its second expression, and (57) holds if

$$\min(\Delta, \delta) \leq \delta_1(\mu) = \delta_1(\mu_{\text{crit}}^1)$$

which is true if

$$\Delta \leq \delta_1(\mu_{\text{crit}}^1) = \frac{\sigma(1-\gamma)}{4(1+\gamma)C(M+n)}.$$

2. We prove now that (24) holds for all $\Delta \in (0, \Delta_{\min}]$ with $\Delta_{\min}$ satisfying (53). From Lemma 1 and $\alpha^t(\Delta) \leq \alpha^n(\Delta)$ we derive

$$\theta_\ell(w(\Delta)) \leq (1 - \alpha^t(\Delta))\theta_\ell(w) + M_\ell\Delta^2,$$
$$\theta_h(w(\Delta)) \leq (1 - \alpha^t(\Delta))\theta_h(w) + M_h\Delta^2.$$

Using $\theta_h(w) + \theta_\ell(w) \leq M\mu$ we get

$$\theta_h(w(\Delta)) + \theta_\ell(w(\Delta)) \leq (1 - \alpha^t(\Delta))M\mu + (M_h + M_\ell)\Delta^2.$$

On the other hand, by Lemma 13

$$M\mu(\Delta) \geq (1 - \alpha^t(\Delta))M\mu + \sigma\alpha^t(\Delta)M\mu - 4M\Delta^2.$$

Therefore, (24) holds whenever

$$(M_h + M_\ell + 4M)\Delta^2 \leq \sigma\alpha^t(\Delta)M\mu,$$

which, by (14), is implied by

$$\Delta \leq \min\left\{\sqrt{\frac{\sigma M\mu}{M_h + M_\ell + 4M}}, \frac{\sigma M\mu}{(M_h + M_\ell + 4M)\max\{\|s^n\|, \|s^t\|\}}\right\},$$

which in turn is true, by (56), if

$$\Delta \leq \min\left\{\sqrt{\frac{\sigma M\mu}{M_h + M_\ell + 4M}}, \frac{\sigma M}{(M_h + M_\ell + 4M)C(M+n)}\right\} \overset{\text{def}}{=} \delta_2(\mu).$$

From now on, this part of the proof follows exactly the same steps of part 1, with

$$\mu_{\mathrm{crit}}^2 \overset{\text{def}}{\equiv} \frac{\sigma M}{(M_h + M_\ell + 4M)C^2(M + n)^2},$$

replacing the role of μ_{crit}^1, and

$$0 < \Delta \le \delta_2(\mu_{\mathrm{crit}}^2) = \frac{\sigma M}{(M_h + M_\ell + 4M)C(M + n)}.$$

3. Finally we prove that (25) holds for all Δ such that (53) is satisfied. We know from part 1 that (54) and (55) are verified if $\Delta \in (0, \Delta_{\min}]$ with $\Delta_{\min}$ satisfying (53). It follows from (55) that

$$4\Delta^2 < \big(\alpha^n(\Delta) - \alpha^t(\Delta)(1 - \sigma)\big)\mu$$

So, from (54), we get

$$X(\Delta)z(\Delta) > \gamma\big(1 - \alpha^n(\Delta)\big)\mu e \ge 0,$$

for all Δ for which (53) is satisfied, and assertion (25) follows trivially. □

References

1. Argaez, M., Tapia, R.A.: On the global convergence of a modified augmented Lagrangian line search interior point Newton method for nonlinear programming. Tech. Report TR95–38, Department of Computational and Applied Mathematics, Rice University, 1995. Revised September 1999
2. Audet, C., Dennis, J.E.: A pattern search filter method for nonlinear programming without derivatives. Tech. Report TR00–09, Department of Computational and Applied Mathematics, Rice University, 2000
3. Benson, H.Y., Shanno, D.F., Vanderbei, R.J.: Interior-point methods for nonconvex nonlinear programming: Filter methods and merit functions. Tech. Report ORFE-00-06, Operations Research and Financial Engineering, Princeton University, 2000
4. Boggs, P.T., Tolle, J.W.: Sequential quadratic programming. In: Acta Numerica 1995, A. Iserles, (ed.), Cambridge University Press, Cambridge, London, New York, 1995, pp. 1–51
5. Byrd, R.H., Hribar, M.E., Nocedal, J.: An interior point algorithm for large-scale nonlinear programming. SIAM J. Optim. **9**, 877–900 (1999)
6. Byrd, R.H., Liu, G., Nocedal, J.: On the local behavior of an interior point method for nonlinear programming. In: Numerical analysis 1997 (Dundee) Pitman Res. Notes Math. Ser., 380, Longman, Harlow, 1998, pp. 37–56
7. Chin, C.M., Fletcher, R.: On the global convergence of an SLP-filter algorithm that takes EQP steps. Tech. Report NA/199, Department of Mathematics, University of Dundee, 2001
8. Coleman, T.F., Li, Y.: On the convergence of interior–reflective Newton methods for nonlinear minimization subject to bounds. Math. Program. **67**, 189–224 (1994)

9. Conn, A.R., Gould, N.I.M., Orban, D., Toint P.L.: A primal-dual trust-region algorithm for minimizing a non-convex function subject to general inequality and linear equality constraints. In: Nonlinear optimization and related topics (Erice, 1998), Kluwer Acad. Publ., Dordrecht, 2000, pp. 15–49
10. Conn, A.R., Gould, N.I.M., Toint P.L.: Trust-Region Methods. MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2000
11. El-Bakry, A.S., Tapia, R.A., Tsuchiya, T., Zhang, Y.: On the formulation and theory of the Newton interior–point method for nonlinear programming. J. Optim. Theor. Appl. **89**, 507–541 (1996)
12. Fletcher, R., Gould, N.I.M., Leyffer, S., Toint, P.L., Wächter, A.: Global convergence of trust-region SQP-filter algorithms for general nonlinear programming. SIAM J. Optim. **13**, 635–659 (2002)
13. Fletcher, R., Leyffer, S.: User manual for filterSQP. Tech. Report NA/181, Department of Mathematics, University of Dundee, 1998
14. Fletcher, R., Leyffer, S.: Nonlinear programming without a penalty function. Math. Programm. **91**, 239–269 (2002)
15. Fletcher, R., Leyffer, S., Toint, P.L.: On the global convergence of an SLP-filter algorithm. Tech. Report 98/13, Département de Mathématique, FUNDP, Namur, 1998
16. Forsgren, A., Gill, P.E.: Primal-dual interior methods for nonconvex nonlinear programming. SIAM J. Optim. **8**, 1132–1152 (1998)
17. Gay, D.M., Overton, M.L., Wright, M.H.: A primal-dual interior method for nonconvex nonlinear programming. In: Proceedings of the 1996 International Conference on Nonlinear Programming, Beijing, China, Y. Yuan, (ed.), Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 31–56
18. Heinkenschloss, M., Ulbrich, M., Ulbrich, S.: Superlinear and quadratic convergence of affine-scaling interior-point Newton methods for problems with simple bounds without strict complementarity assumption. Math. Program. **86**, 615–635 (1999)
19. Martinez, H.J., Parada, Z., Tapia, R.A.: On the characterization of q-superlinear convergence of quasi-Newton interior-point methods for nonlinear programming. Boletin de la Sociedad Matematica Mexicana **1**, 1–12 (1995)
20. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer–Verlag, Berlin, 1999
21. Tseng, P.: Error bounds and superlinear convergence analysis of some Newton-type methods in optimization. In: Nonlinear Optimization and Applications, Vol. **2**, Kluwer Academic Publishers B.V., 1998
22. Ulbrich, M., Ulbrich, S.: Nonmonotone trust region methods for nonlinear equality constrained optimization without a penalty function. Math. Program. **95**, 103–135 (2003)
23. Vanderbei, R.J., Shanno, D.F.: An interior-point algorithm for nonconvex nonlinear programming. Comput. Optim. Appl. **13**, 231–252 (1999)
24. Vicente, L.N.: Local convergence of the affine-scaling interior-point algorithm for nonlinear programming. Comput. Optim. Appl. **17**, 23–35 (2000)
25. Vicente, L.N., Wright, S.J.: Local convergence of a primal-dual method for degenerate nonlinear programming. Comput. Optim. Appl. **22**, 311–328 (2002)
26. Wächter, A.: An Interior Point Algorithm for Large-Scale Nonlinear Optimization with Applications in Process Engineering. PhD thesis, Department of Chemical Engineering, Carnegie Mellon University, 2002
27. Wächter, W., Biegler, L.T.: Failure of global convergence for a class of interior point methods for nonlinear programming. Math. Program. **88**, (2000)
28. Wächter, A., Biegler, L.T.: Global and local convergence of line search filter methods for nonlinear programming. Tech. Report B-01-09, CAPD, Department of Chemical Engineering, Carnegie Mellon University, 2001
29. Wright, S.J.: Primal-Dual Interior-Point Methods. SIAM, Philadelphia, 1997
30. Wright, S.J., Orban, D.: Properties of the log-barrier function on degenerate nonlinear programs. Math. Oper. Res. **27**, 585–613 (2003)
31. Yamashita, H., Yabe, H.: Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization. Math. Program. **75**, 377–397 (1996)