A Globally Convergent Primal-Dual Interior-Point Filter Method for Nonconvex Nonlinear Programming

Michael Ulbrich
Stefan Ulbrich
Luís N. Vicente

April 2000
TR00-12

Department of Computational and Applied Mathematics - MS134
Rice University
6100 Main Street
Houston, Texas 77005-1892
USA

Michael Ulbrich · Stefan Ulbrich · Luís N. Vicente

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

April 2000

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 [10] 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 discovery in 1997 by Fletcher and Leyffer [10], the filter technique has been mostly applied, so far, to SLP (sequential linear programming) and SQP (sequential quadratic programming) type methods [9, 10, 12]. Global convergence to first-order critical points has been proved for SLP by Fletcher, Leyffer, and Toint [12] in 1998 and for SQP by Fletcher, Gould, Leyffer, and Toint [9] in 1999. In the context of composite SQP for equality constrained optimization, Ulbrich and Ulbrich [19], have also

Michael Ulbrich: Zentrum Mathematik, Technische Universität München, 80290 München, Germany, E-mail: mulbrich@ma.tum.de. Support for this author was provided by CRPC grant CCR–9120008.

Stefan Ulbrich: Zentrum Mathematik, Technische Universität München, 80290 München, Germany, E-mail: sulbrich@ma.tum.de. Support for this author was provided by CRPC grant CCR–9120008.

Luís N. Vicente: Departamento de Matemática, Universidade de Coimbra, 3001-454 Coimbra, Portugal, E-mail: lvicente@mat.uc.pt. Support for this author was provided by Centro de Matemática da Universidade de Coimbra.

Mathematics Subject Classification (1991): 65K05, 90C06, 90C29, 90C30

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 [11], 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 [4, 7, 8, 13, 14, 20]. On the other hand, the local convergence properties of interior-point methods for nonlinear programming are quite well studied in the literature [5, 6, 8, 16, 21, 24], although difficulties arise when the limit point does not satisfy strict complementarity or linear independence of the gradients of the active or binding constraints [15, 18, 22].

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, and Toint [9]. 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 [9, 10, 12] 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 matrix or a vector. 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 [10], using the form analyzed by Fletcher, Gould, Leyffer, and Toint [9]. 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 proposed in [10] to use a *filter* to build the efficient frontier 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 [10], 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 acceptability of pairs that are arbitrarily close to the efficient border, acceptability of x to the filter is defined in [9] in a more stringent way by requiring that for all filter entries $(\theta(x_j), f(x_j)) \in \mathcal{F}$ holds

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

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 $(\theta(x), \theta_g(x))$, 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, and Toint [9] 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 yielding linearized feasibility, i.e.

$$h(x_k) + \nabla h(x_k)^T s_k^n = 0, \quad \|s_k^n\| \leq \Delta_k, \quad \|s_k^n\| \leq \kappa_n \, \theta_k \tag{3}$$

with some constant $\kappa_n > 0$. The tangential step is, in turn, computed verifying

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

and providing a fraction of Cauchy decrease for a quadratic model m_k of f. Problem (3) is infeasible if Δ_k is too small in comparison to θ_k. Therefore, it is required in [9] that the normal step s_k^n should be tiny compared to Δ_k, otherwise 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 $m_k(x_k) - m_k(x_k + s_k) < \kappa\theta_k^2$ with a constant $\kappa > 0$, 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

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

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 θ_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 sketched SQP filter method of [9]. Essentially, our application of the filter concept to the globalization of interior-point methods was led by the following considerations:

– A primal-dual interior-point Newton step forms the basis for the trial step computation. The nonnegativity constraints are handled by a centering mechanism.
– An identification of two objectives θ and θ_g, corresponding to feasibility and objective function value, such that an appropriate splitting of the step in a “normal” and “tangential” component guarantees decrease for linearized models of θ and θ_g, respectively.
– An adaptation of the filter framework of [9] for the pair (θ, θ_g).
– The need to consider trial steps that for different trust-regions radii do not require the recomputation of normal and tangential components.

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

$$\nabla_x \ell(x,y,z) = 0, \tag{4}$$
$$h(x) = 0, \tag{5}$$
$$Xz = 0, \tag{6}$$
$$x \geq 0, \quad z \geq 0. \tag{7}$$

Hereby, $y \in \mathbb{R}^m$ 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 (4)–(7) are necessary for x to be a local solution of (1).

We now perturb block (6) of the KKT system (4)-(6) and write

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

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

$$\mu = \frac{x^T z}{n}. \tag{8}$$

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 (8) 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. [8].

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. Along with the choice of the 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

$$F_{\sigma\mu}(x,y,z) = \begin{pmatrix} 0 \\ h(x) \\ Xz - \mu e \end{pmatrix} + \begin{pmatrix} \nabla_x \ell(x,y,z) \\ 0 \\ (1-\sigma)\mu e \end{pmatrix} = 0. \tag{9}$$

The first term in the middle expression measures the proximity to the quasi-central path. We recall that the quasi-central path, parametrized 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

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

parametrized 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 ($\max\{x_i, z_i\} > 0, i = 1, \ldots, n$), and linear independence of the gradients of the active or binding constraints.

The second term in the middle expression of (9) 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 provides some room for improvements, 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 the filter components it remains to define corresponding tangential and normal components of the trial step. We use the decomposition associated with the splitting (9). For the *normal* step $s^n = (\Delta x^n, \Delta y^n, \Delta z^n)$ we thus choose

$$F'_{\sigma\mu}(w)s^n = -\begin{pmatrix} 0 \\ h(x) \\ Xz - \mu e \end{pmatrix}, \tag{10}$$

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

$$F'_{\sigma\mu}(w)s^t = -\begin{pmatrix} \nabla_x \ell(w) \\ 0 \\ (1-\sigma)\mu e \end{pmatrix}. \tag{11}$$

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 stepsizes for s^n and s^t in our trial step computation.

The adjectives normal and tangential are borrowed from the SQP context [3,17] but have here a slightly different flavor. The normal step can be seen as a step towards the quasi-central path P^q_μ. The tangential step is the sum of a tangential component s^t_1

$$F'_{\sigma\mu}(w)s^t_1 = -\begin{pmatrix} \nabla_x \ell(w) \\ 0 \\ 0 \end{pmatrix}$$

that attempts to reduce $\|\nabla_x \ell(w)\|$, with a predictor component s^t_2

$$F'_{\sigma\mu}(w)s^t_2 = -\begin{pmatrix} 0 \\ 0 \\ (1-\sigma)\mu e \end{pmatrix}$$

that seeks the minimization of $\mu = x^T z/n$ (see, for instance, [23]). 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

$$\|\alpha_n(\Delta)s^n\| \le \Delta, \qquad \|\alpha_t(\Delta)s^t\| \le \Delta.$$

Having these bounds in mind, and requiring explicitly $\alpha_t(\Delta) \le \alpha_n(\Delta)$, we set

$$\alpha_n(\Delta) = \min\left\{1, \frac{\Delta}{\|s^n\|}\right\}, \tag{12}$$

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

Let also

$$w(\Delta) = (x(\Delta), y(\Delta), z(\Delta)) = w + \alpha_n(\Delta)s^n + \alpha_t(\Delta)s^t,$$
$$s(\Delta) = \big(s_x(\Delta), s_y(\Delta), s_z(\Delta)\big) = w(\Delta) - w = \alpha_n(\Delta)s^n + \alpha_t(\Delta)s^t.$$

Thus,

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

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

The scalars $\alpha_n(\Delta)$ and $\alpha_t(\Delta)$ define the steps taken along the normal and tangential directions, respectively, and 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 to rewrite 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 verifying $\theta(w) = \theta_g(w) = 0$ 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

$$\begin{aligned} & m(w(\Delta)) \\ &= \frac{x^T z}{n} + \frac{(x(\Delta) - x)^T z + (z(\Delta) - z)^T x}{n} + \|\nabla_x \ell(w) + \nabla^2_{wx} \ell(w)(w(\Delta) - w)\|^2 \\ &= \frac{x(\Delta)^T z(\Delta) - (x(\Delta) - x)^T (z(\Delta) - z)}{n} + \|\nabla_x \ell(w) + \nabla^2_{wx} \ell(w)(w(\Delta) - w)\|^2, \end{aligned}$$

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

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

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

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

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. [23] and the references therein, and has also proven its efficiency in the context of nonlinear programming [8]. We will show 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, at $w(\Delta)$, upper bounds on the values of θ_h, θ_c, θ_ℓ, and θ_g in terms of Δ and of the corresponding values at 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 $\Delta > 0$,*

$$\theta_h(w(\Delta)) \leq (1 - \alpha_n(\Delta))\theta_h(w) + M_h\Delta^2, \tag{14}$$

$$\theta_c(w(\Delta)) \leq (1 - \alpha_n(\Delta))\theta_c(w) + M_c\Delta^2, \tag{15}$$

$$\theta_\ell(w(\Delta)) \leq (1 - \alpha_t(\Delta))\theta_\ell(w) + M_\ell\Delta^2. \tag{16}$$

For Δ satisfying $0 < \Delta \leq \Delta_{ub}$, it also holds

$$\theta_g(w(\Delta)) \leq \big(1 - \alpha_t(\Delta)(1-\sigma)\big)\theta_g(w) + M_g\Delta^2, \tag{17}$$

for some positive constants M_g.

Finally, for any $\Delta > 0$, we also have

$$m(w) - m(w(\Delta)) \geq \alpha_t(\Delta)(1-\sigma)\theta_g(w).$$

Now we state a result that says that if the current point $w = (x, y, z)$ satisfies the centrality requirement $Xz \geq \gamma\mu e$, 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 $\theta_h(w) + \theta_\ell(w) \leq M\mu$ and $(x, z) > 0$.

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

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

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

$$X(\Delta)z(\Delta) \geq \gamma\mu(\Delta), \tag{18}$$

$$\theta_h(w(\Delta)) + \theta_\ell(w(\Delta)) \leq M\mu(\Delta). \tag{19}$$

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

$$(x(\Delta), z(\Delta)) > 0. \tag{20}$$

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

The proves 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

$$\theta(w) \le \theta(w') \quad \text{and} \quad \theta_g(w) \le \theta_g(w'),$$

or, alternatively, 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 a too weak acceptance criteria. 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:

$$\begin{aligned} \mathcal{F} \mapsto \mathcal{F} = &\{(\theta(w), \theta_g(w))\} \cup \\ &\big\{(\theta^f, \theta_g^f) \in \mathcal{F} \;:\; \min\{\theta^f - \theta(w), \theta_g^f - \theta_g(w)\} < 0\big\}. \end{aligned}$$

Remark. 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:

$$\rho_k \ge \eta$$

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 $\rho_k \ge \eta$ 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 selection criterion of adding points w_k to the filter prevents 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 to 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 are preset positive constants. The restoration algorithm must produce a new iterate w_{k+1} that is acceptable to the filter and for which holds $\theta(w_{k+1}) \leq \Delta_k \min\{\gamma_1, \gamma_2 \Delta_k^\beta\}$. 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, in any circumstances, acceptable to the filter.

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

0. Choose $\sigma \in (0,1)$, $\rho \in (0,1)$, $\gamma_1, \gamma_2 > 0$, $0 < \beta, \eta, \kappa < 1$, and $\gamma_{\mathcal{F}} \in (0, 1/2)$. 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 $\theta_h(w_0) + \theta_\ell(w_0) \leq M\mu_0$. Set $k := 0$ and $\Delta_0 := 1$.
1. Compute s_k^n and s_k^t by solving the linear systems (10) and (11), respectively, with $(w, \mu) = (w_k, \mu_k)$.
2. Compute Δ_k such that

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

3. Reduce Δ_k, if necessary, so that $\theta_h(w_k(\Delta_k)) + \theta_\ell(w_k(\Delta_k)) \leq M\mu_k(\Delta_k)$.
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 call a restoration algorithm that produces a point w_{k+1} such that:

   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 $k := k+1$ and return to step 1 for a new iteration with $\Delta_{k+1} := \Delta_k$.
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. 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. Set $\Delta_{k+1} \geq \Delta_k$.
10. Set $w_{k+1} := w_k(\Delta_k)$, $k := k+1$, and go to step 1.
11. Set $w_{k+1} := w_k$, $s^n_{k+1} := s^n_k$, $s^t_{k+1} := s^t_k$, $\mu_{k+1} := \mu_k$, $\Delta_{k+1} := \Delta_k/2$. Set $k := k+1$ and go to step 3.

In practice, step 2 would be implemented as $\Delta_k = \tau_k \Delta'_k$, where Δ'_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 $(0,1)$ 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 splitted in steps 2 and 3 for good reasons. In fact, in step 2 it is possible to determine explicitly Δ_k (more precisely Δ'_k). However, due to the nonlinearity of θ_h and θ_l, that is not the case in step 3, where we know from Lemma 2 that we can indeed find a Δ_k verifying $\theta_h(w_k(\Delta_k)) + \theta_\ell(w_k(\Delta_k)) \leq M\mu_k(\Delta_k)$ but we cannot explicitly determine it.

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 2

(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 $[w_k, w_k + s(\Delta_k)]$.
(A3) There exists $C > 0$ such that for all k it holds $\|F'_{\sigma\mu_k}(w_k)^{-1}\| \leq C$.

Remark. (A3) holds in a neighborhood of a regular point w^* satisfying the second-order sufficient conditions and strict complementarity, see for instance [8]. Moreover, in [8] conditions are given 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.
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* $0 \leq \Delta_k \leq \Delta_{\min}$.
iv) *It holds* $\max\{\|s^n_k\|, \|s^t_k\|\} \leq C(M + (n^2 - n + 1)^{1/2})\mu_k$ *for all* k.

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

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

We start the convergence theory by showing that new iterates are always acceptable to the filter and that all filter entries (θ^f, θ_g^f) obey $\theta^f > 0$, two facts that require no analysis and follow 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 both cases, we can see that $\theta(w_k) > 0$.

Lemma 5. *All new iterates $w_{k+1} \neq w_k$ are acceptable to the filter.*

Proof. Clearly w_0 is acceptable to the filter. Moreover, iterates with $\theta(w_k) = 0$ are never added to the filter. New iterates $w_{k+1} \neq w_k$ are produced either in the restoration or in step 11. The restoration terminates finitely with a point that is acceptable to the filter. Step 11 is only reached if the check in step 5 was successful, which ensures that w_{k+1} is acceptable even if w_k was possibly added to the filter in step 8.

The next three lemmas describe a few technical results that are 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*

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

Proof. If step 5 is reached, then

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

Thus, we have $\|s_k^n\| \leq C\gamma_2\Delta_k^{1+\beta}$, or equivalently,

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

We see then that $\alpha_n^k = 1$ whenever

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

But then, by Lemma 1,

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

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 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

$$\theta(w_k(\Delta_k)) \leq \frac{1}{2}\min\{\theta_{\mathcal{F}}, (1-\gamma_{\mathcal{F}})\epsilon/2\} < \min\{\theta_{\mathcal{F}}, (1-\gamma_{\mathcal{F}})\epsilon/2\}. \tag{21}$$

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, (21) is satisfied for $\Delta_k \leq \Delta_a^1$ with $\Delta_a^1 > 0$ depending only on $\theta_{\mathcal{F}}$, ϵ and, of course, on 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 (21), 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,

$$\theta_g(w_k(\Delta_k)) - \theta_g(w_k) < -\gamma_{\mathcal{F}}\theta(w_k). \tag{22}$$

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

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

Hence we need to show that

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

Since $\|s_k^n\|$ and $\|s_k^t\|$ are bounded by a constant M_s and $\alpha_t^k = \min\{1, \frac{\Delta_k}{\|s_k^n\|}, \frac{\Delta_k}{\|s_k^t\|}\}$, we have for all $\Delta_k \leq M_s$, that $\alpha_t^k \geq \Delta_k/M_s$. Thus (22) 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 and of course on the constants 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.

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* $\epsilon > 0$

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

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 was 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

$$\theta_g(w_k(\Delta_k)) < \theta_g(w_k) - \gamma_{\mathcal{F}}\theta(w_k). \tag{23}$$

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, if, in addition

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

Moreover, when $0 < \Delta_k \leq \Delta_{ub}$, we have by Lemma 1 that

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

We have pointed out before that $\alpha_t^k \geq \Delta_k/M_s$ for all $\Delta_k \leq M_s$. 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 $\theta(w_k) \leq \gamma_2\Delta_k^{1+\beta}$, we obtain (23) whenever

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

All these requirements on Δ_k are satisfied if $0 < \Delta_k \leq \Delta_f$ with some constant $\Delta_f > 0$.

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

Lemma 9. *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*

$$\lim_{i\to\infty} \theta(w_{k_i}) = 0. \tag{24}$$

Proof. Assume that for a given sequence $\{k_l\}$ with w_{k_l} added to the filter, we have $\theta(w_{k_l}) \geq \epsilon > 0$ for all l . Since w_{k_l} is added to the filter and by Lemma 5 all new produced points are acceptable to the filter, no further entry can be made in the square

$$[\theta(w_{k_l}), \theta(w_{k_l}) - \gamma_{\mathcal{F}}\epsilon] \times [\theta_g(w_{k_l}), \theta_g(w_{k_l}) - \gamma_{\mathcal{F}}\epsilon].$$

All these squares with area $(\gamma_{\mathcal{F}}\epsilon)^2$ are disjoint which contradicts the boundedness of the sequence $\{\theta(w_k), \theta_g(w_k)\}$. Thus, there exists a subsequence $\{k_i\} \subset \{k_l\}$ for which (24) holds.

One can prove more when infinitely many iterates are added to the filter. In fact, in this case $\liminf_{k\to\infty} \theta(w_k) = 0$ can be extended to $\liminf_{k\to\infty} \theta(w_k) + \theta_g(w_k) = 0$. We start with the case where restoration is responsible for the infinite number of entries in the filter.

Lemma 10. *Suppose that restoration is entered infinitely many times. 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. We know from Lemma 9 that there exists a subsequence $\{k_i\}$ (where restoration is entered for every k_i and w_{k_i} is added to the filter) such that $\lim_{i\to\infty} \theta(w_{k_i}) = 0$. Since the restoration is entered it must hold

$$\theta_{k_i} > \Delta_{k_i} \min\{\gamma_1, \gamma_2 \Delta_{k_i}^{\beta}\}. \tag{25}$$

But $\Delta_{k_i-1} \leq 2\Delta_{k_i}$ and therefore

$$0 = \lim_{i\to\infty} \theta_{k_i} = \lim_{i\to\infty} \Delta_{k_i} = \lim_{i\to\infty} \Delta_{k_i-1}. \tag{26}$$

By Lemma 3.iii we know that for large enough i step 2 and step 3 do not change Δ_{k_i-1} and Δ_{k_i}. Hence, in all iterations $k_i - 1 \geq K_1$, with K_1 large enough, step 5 is reached. In fact, otherwise the restoration is entered in iteration $k_i - 1$ and since step 2 and step 3 do not change Δ_{k_i} we would reach step 4 in iteration k_i with $\Delta_{k_i} = \Delta_{k_i-1}$ and

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

which contradicts (25). 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 all sufficiently large i. It follows from Lemma 1, with $0 < \Delta_{k_i-1} \leq \Delta_{ub}$, that

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

and thus

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

We can therefore conclude, from (26), and for large enough i, that

$$\theta_g(w_{k_i-1}) \geq \epsilon/2.$$

We show now that step 7 must be reached for $k_i - 1 \geq K_2$, with $K_2 \geq K_1$. Otherwise, step 5 is followed by step 11 and thus $\theta_{k_i} = \theta_{k_i-1}$, $\Delta_{k_i} = \Delta_{k_i-1}/2$. Thus, by (25),

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

Hence, we obtain by Lemma 8 that $w_{k_i-1}(\Delta_{k_i-1})$ is acceptable to the filter in step 5 whenever $\Delta_{k_i-1} \leq \Delta_f$. But, by (26), this last inequality holds for all i sufficiently large. Hence, there exists K_2 for which step 7 is always reached if $k_i - 1 \geq K_2$.

We conclude the proof by showing the existence of $K_3 \geq K_2$ for which step 9 is reached for all $k_i - 1 \geq K_3$. This assertion leads to a contradiction since then, by Lemma 6, for all $\Delta_{k_i-1} \leq \Delta_r$

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

which contradicts (25) and (26), since $\Delta_{k_i} \geq \Delta_{k_i-1}$ is not changed in step 2 and step 3.

It remains to show that step 9 is eventually reached. We note that by Lemma 1

$$m_{k_i-1}(w_{k_i-1}) - m_{k_i-1}(w_{k_i-1}(\Delta_{k_i-1})) \geq \alpha_t^{k_i-1}(1-\sigma)\epsilon/2.$$

We use again the fact that $\alpha_t^{k_i-1} \geq \Delta_{k_i-1}/M_s$ if $\Delta_{k_i-1} \leq M_s$. On the other hand, we have

$$|m_{k_i-1}(w_{k_i-1}) - m_{k_i-1}(w_{k_i-1}(\Delta_{k_i-1})) - \theta_g(w_{k_i-1}) + \theta_g(w_{k_i-1}(\Delta_{k_i-1}))| \\ \leq M_g \Delta^2_{k_i-1}.$$

This shows that $\rho_{k_i-1} \to 1$ and hence there exists K_3 such that step 9 is reached for all $k_i - 1 \geq K_3$. The proof is therefore completed since there exists a subsequence $\{k_j\} \subset \{k_i\}$ for which $\lim_{j\to\infty} \theta(w_{k_j}) = \lim_{j\to\infty} \theta_g(w_{k_j}) = 0$.

The other situation is when it is step 8 that causes the filter to contain an infinite number of entries. We summarize both situations in the next theorem.

Theorem 3. *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. The assertion is true, by Lemma 10, if restoration is entered infinitely many times. So, it remains to consider the case when, eventually, all the iterates that are added to the filter are added in step 8. In this case we can also appeal to Lemma 9 to conclude the existence of a subsequence $\{k_i\}$ such that $\lim_{i\to\infty}\theta(w_{k_i}) = 0$. Suppose now that $\theta_g(w_{k_i}) \geq \epsilon > 0$. Then

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

Thus, we obtain $\alpha_t^{k_i} \to 0$ and consequently $\Delta_{k_i} \to 0$. In particular, $\alpha_t^{k_i} \geq \Delta_{k_i}/M_s$ for large enough i, and since the restoration is not entered, we 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 $\Delta_{k_i} \to 0$.

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

Theorem 4. *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 entered. 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. If $\theta(w_k) = 0$ then $\theta_g(w_k) > 0$ and therefore $\rho_k \geq \eta$ for all Δ_k small enough. 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 that in step 10 it always holds

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

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 $\Delta' > 0$. Hence, Δ_k is eventually not decreased any more and therefore $\Delta_k \geq \delta > 0$ for all $k \geq K'$. Thus, step 10 is reached for all successful steps with $\Delta_k \geq \delta > 0$ and we have, as above,

$$\theta_g(w_k) - \theta_g(w_{k+1}) \geq \eta(m_k(w_k) - m_k(w_k(\Delta_k))) \geq \eta(1-\sigma)\epsilon\alpha_t^k \geq \eta(1-\sigma)\epsilon\frac{\delta}{M_s}$$

with M_s 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 3 and 4:

Corollary 1. *Under Assumption 2, 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 a restoration algorithm 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} acceptable to the filter and such that $\theta(w_{k+1}) \leq \Delta_k \min\{\gamma_1, \gamma_2\Delta_{k+1}^\beta\}$ with $\Delta_{k+1} = \Delta_k$. In a restoration algorithm, it is therefore desired to decrease the value of $\theta(w) = \theta_h(w) + \theta_c(w)$. To achieve this goal we introduce the function

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

The normal step s^n computed from (10) is a descent direction for $\theta_2(w)$. In fact,

$$\begin{aligned}\nabla\theta_2(w)^T s^n &= (Xz - \mu e)^T(Z\Delta x^n + X\Delta z^n) + h(x)^T\nabla h(x)^T\Delta x^n \\ &= -(Xz - \mu e)^T(Xz - \mu e) - h(x)^T h(x).\end{aligned}$$

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

$$\nabla\theta_2(w)^T s^n = -2\theta_2(w), \qquad \nabla\theta_2(w)^T s^t = 0. \tag{27}$$

The restoration algorithm that we present 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 (12), (10), (13), and (11), 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 5 (Restoration algorithm)

0. Choose $\nu \in (0,1)$. Set $w_k^0 := w_k$, $\Delta_k^0 := \Delta_k$, $j := 0$ and start with step 4.
1. If $\theta(w_k^j) \le \Delta_k \min\{\gamma_1, \gamma_2 \Delta_k^\beta\}$ and w_k^j is acceptable to the filter then set $w_{k+1} := w_k^j$ and stop restoration.
2. Compute s_k^n and s_k^t by solving the linear systems (10) and (11), respectively, with $(w,\mu) = (w_k^j, \mu_k^j)$.
3. Compute Δ_k^j such that

$$x_k^j(\Delta) > 0, \quad z_k^j(\Delta) > 0, \quad X_k^j(\Delta) z_k^j(\Delta) \ge \gamma \mu_k^j(\Delta) e \qquad \text{for all } \Delta \in [0, \Delta_k^j].$$

4. If

$$\frac{\theta_2(w_k^j) - \theta_2(w_k^j(\Delta_k^j))}{-\nabla\theta_2(w_k^j)^T s_k^j(\Delta_k^j)} \ge \nu, \tag{28}$$

$$\theta_h(w_k^j(\Delta_k^j)) + \theta_\ell(w_k^j(\Delta_k^j)) \le M \mu_k^j(\Delta_k^j), \tag{29}$$

then 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 last theorem.

Theorem 6. *The restoration algorithm 5 terminates in a finite number of iterations.*

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 $\theta(w_k^j) \le 2\sqrt{\theta_2(w_k^j)} \le 1/2\theta_{\mathcal{F}}$. This condition and $\theta(w_k^j) \le \Delta_k \min\{\gamma_1, \gamma_2 \Delta_k^\beta\}$ are eventually satisfied if

$$\liminf_{j\to\infty} \theta_2(w_k^j) = 0. \tag{30}$$

Hence, if the restoration does not terminate finitely, then there exists an $\epsilon > 0$ with $\theta_2(w_k^j) \ge \epsilon$ for all j. We show that this uniform bound will lead to a contradiction. In fact, from (27), we have

$$\begin{aligned}\nabla\theta_2(w_k^j)^T s_k^j(\Delta) &= \alpha_n^{k,j}(\Delta) \nabla\theta_2(w_k^j)^T s_{k,j}^n + \alpha_t^{k,j}(\Delta)\nabla\theta_2(w_k^j)^T s_{k,j}^t \\ &= -2\alpha_n^{k,j}(\Delta)\theta_2(w_k^j).\end{aligned}$$

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 $\alpha_t^{k,j}(\Delta) \le \alpha_n^{k,j}(\Delta)$, implies

$$\theta_2(w_k^j) - \theta_2(w_k^j(\Delta)) \ge -\nabla\theta_2(w_k^j)^T s_k^j(\Delta) - 2M_2 \alpha_n^{k,j}(\Delta)^2 (\|s_{k,j}^n\|^2 + \|s_{k,j}^t\|^2).$$

Hence, (28) holds for all $\alpha_n^{k,j}(\Delta)$ such that

$$2(1-\nu)\alpha_n^{k,j}(\Delta)\theta_2(w_k^j) \geq 2M_2\alpha_n^{k,j}(\Delta)^2(\|s_{k,j}^n\|^2 + \|s_{k,j}^t\|^2),$$

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

$$\alpha_n^{k,j}(\Delta) \leq \bar{\alpha}_n^{k,j} \stackrel{\text{def}}{=} \min\left\{1, \frac{(1-\nu)\theta_2(w_k^j)}{M_2(\|s_{k,j}^n\|^2 + \|s_{k,j}^t\|^2)}\right\}.$$

From Lemmas 2 and 3.iii, we see finally that (28) and (29) are satisfied for all Δ_k^j such

$$0 < \Delta_k^j \leq \min\left\{\Delta_{\min}, \bar{\alpha}_n^{k,j}\|s_{k,j}^n\|\right\},$$

showing that these two 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 $\max\{\|s_{k,j}^n\|, \|s_{k,j}^t\|\} \leq M_s$,

$$\alpha_n^{k,j}(\Delta_k^j) \geq \frac{1}{2}\min\left\{\frac{\Delta_{\min}}{\|s_{k,j}^n\|}, \bar{\alpha}_n^{k,j}\right\} \geq \bar{\alpha} > 0$$

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

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

which yields a contradiction. Hence, we have (30) and the finite termination is proved.

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 algorithm. We are currently working on an implementation of the algorithm. A preliminary version of the algorithm that was obtained by modifying a primal-dual interior-point QP solver performed very promising on QP test problems.

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 two.

The new primal-dual interior-point algorithm used 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 could use the first dimension for feasibility, the second for centrality, and the third for the size of the gradient of the Lagrangian.

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

$$x_i(\Delta)z_i(\Delta) \le (1-\alpha_n(\Delta))x_iz_i + \big(\alpha_n(\Delta) - \alpha_t(\Delta)(1-\sigma)\big)\mu + 4\Delta^2, \tag{31}$$

$$x_i(\Delta)z_i(\Delta) \ge (1-\alpha_n(\Delta))x_iz_i + \big(\alpha_n(\Delta) - \alpha_t(\Delta)(1-\sigma)\big)\mu - 4\Delta^2, \tag{32}$$

$$\mu(\Delta) \le \big(1-\alpha_t(\Delta)(1-\sigma)\big)\,\mu + 4\Delta^2, \quad \mu(\Delta) \ge \big(1-\alpha_t(\Delta)(1-\sigma)\big)\,\mu - 4\Delta^2. \tag{33}$$

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

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

So, inequalities (31) and (32) follow from this derivation and

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

Summing (31) and (32) over all i, dividing the result by n, and using $\mu = x^Ty/n$, $\mu(\Delta) = x(\Delta)^Ty(\Delta)/n$, yield (33).

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

$$\begin{aligned} M_h &= 2C_h, & M_c &= 8\sqrt{n}, \\ M_\ell &= 2C_{\ell'}, & M_g &= 4(1 + B_\ell C_{\ell'}) + 4C_{\ell'}^2\Delta_{ub}^2. \end{aligned}$$

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

$$\begin{aligned} \theta_h(w(\Delta)) = \|h(x(\Delta))\| &= \left\| h(x) + \int_0^1 \nabla h(x + ts_x(\Delta))^T s_x(\Delta)dt \right\| \\ &= \left\| (1-\alpha_n(\Delta))h(x) + \int_0^1 \big(\nabla h(x + ts_x(\Delta)) - \nabla h(x)\big)^T s_x(\Delta)dt \right\| \\ &\le (1-\alpha_n(\Delta))\theta_h(w) + C_h\|s_x(\Delta)\|^2 \int_0^1 t\,dt, \end{aligned}$$

which proves (14).

Similarly, we have $\nabla^2_{xw}\ell(w)^T w(\Delta) = -\alpha_t(\Delta)\nabla_x \ell(w)$ 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 (16).

The estimate (15) follows from Lemma 11:

$$\begin{aligned}\pm\big(x_i(\Delta)z_i(\Delta)-\mu(\Delta)\big) &\leq \pm\Big((1-\alpha_n(\Delta))x_i z_i + \big(\alpha_n(\Delta)-\alpha_t(\Delta)(1-\sigma)\big)\mu\Big)\\ &\quad + 4\Delta^2 \mp \big(1-\alpha_t(\Delta)(1-\sigma)\big)\mu + 4\Delta^2\\ &= \pm(1-\alpha_n(\Delta))(x_i z_i - \mu) + 8\Delta^2.\end{aligned}$$

Inequality (17) is derived by appealing to Lemma 11 and to the previously established inequality (16):

$$\begin{aligned}\theta_g(w(\Delta)) &= \mu(\Delta) + \theta_\ell(w(\Delta))^2\\ &\leq \big(1-\alpha_t(\Delta)(1-\sigma)\big)\mu + 4\Delta^2 + \big((1-\alpha_t(\Delta))\theta_\ell(w) + 2C_{\ell'}\Delta^2\big)^2\\ &\leq \big(1-\alpha_t(\Delta)(1-\sigma)\big)\theta_g(w) + \big(4+4(1-\alpha_t(\Delta))\theta_\ell(w)C_{\ell'}\big)\Delta^2 + 4C_{\ell'}^2\Delta^4.\end{aligned}$$

Finally, we have

$$\begin{aligned}m(w)-m(w(\Delta)) &= \mu - \mu(\Delta) + (x(\Delta)-x)^T(z(\Delta)-z)/n\\ &\quad + (1-(1-\alpha_t(\Delta))^2)\|\nabla_x\ell(w)\|^2\\ &= \alpha_t(\Delta)(1-\sigma)\mu + (1-(1-\alpha_t(\Delta))^2)\|\nabla_x\ell(w)\|^2\\ &\geq \alpha_t(\Delta)(1-\sigma)\theta_g(w),\end{aligned}$$

and the proof is therefore completed.

Proof of Lemma 2

Proof. We will prove Lemma 2 with

$$\Delta_{\min} = \left\{\frac{\sigma(1-\gamma)}{4(1+\gamma)C(M+n)}, \frac{\sigma M}{(M_h+M_\ell+4M)C(M+n)}\right\}. \tag{34}$$

1. We first show that (18) holds for all $\Delta > 0$ satisfying (34).

From Lemma 11 we obtain

$$X(\Delta)z(\Delta) \geq \big(\gamma + (1-\gamma)\alpha_n(\Delta) - \alpha_t(\Delta)(1-\sigma)\big)\mu e - 4\Delta^2 e. \tag{35}$$

On the other hand, Lemma 11 also yields

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

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

$$4\varDelta^2 \leq \frac{(1-\gamma)(\alpha_n(\varDelta) - \alpha_t(\varDelta)(1-\sigma))\mu}{1+\gamma}. \tag{36}$$

Since $\alpha_t(\varDelta) \leq \alpha_n(\varDelta)$, a sufficient condition for this inequality to hold is

$$\varDelta^2 \leq \frac{\alpha_n(\varDelta)\sigma(1-\gamma)\mu}{4(1+\gamma)},$$

which, by (12), is implied by

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

which in turn is true if

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

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

$$\delta \stackrel{\text{def}}{=} \max\{\|s^n\|, \|s^t\|\} \leq C(M+n)\mu. \tag{37}$$

We consider now two possible cases in order to show that (18) holds whenever

$$\min(\varDelta, \delta) \leq \delta_1(\mu). \tag{38}$$

In the first case $\varDelta \leq \delta$ we use that, as we have just shown, (18) holds provided $\varDelta \leq \delta_1(\mu)$. Since $\varDelta \leq \delta$ in this case, the inequality $\varDelta \leq \delta_1(\mu)$ is equivalent to (38). In the case $\delta < \varDelta$, we know that $\alpha_n(\varDelta) = \alpha_t(\varDelta) = 1$, $w(\varDelta) = w(\delta)$, and $s(\varDelta) = s(\delta)$. Thus, (18) 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 $\delta < \varDelta$, we have that $\delta \leq \delta_1(\mu)$ is also equivalent to (38).

Hence, it remains to show (38) for $\varDelta \in [0, \varDelta_{\min}]$ with $\varDelta_{\min}$ according to (34). Let

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

If $\mu \leq \mu^1_{\text{crit}}$, then $\delta_1(\mu)$ is given by its first expression, and (38) follows directly from (37), since by the definition of μ^1_{crit} for $\mu \leq \mu^1_{\text{crit}}$ with (37) holds

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

If $\mu > \mu^1_{\text{crit}}$, then $\delta_1(\mu)$ is given by its second expression, and (38) holds if

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

which is true if

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

2. We prove now that (19) holds for all $\Delta > 0$ satisfying (34).
From Lemma 1 and $\alpha_t(\Delta) \le \alpha_n(\Delta)$ we derive

$$\theta_\ell(w(\Delta)) \le (1-\alpha_t(\Delta))\theta_\ell(w) + M_\ell \Delta^2,$$
$$\theta_h(w(\Delta)) \le (1-\alpha_t(\Delta))\theta_h(w) + M_h \Delta^2.$$

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

$$\theta_h(w(\Delta)) + \theta_\ell(w(\Delta)) \le (1-\alpha_t(\Delta))M\mu + (M_h + M_\ell)\Delta^2.$$

On the other hand, by Lemma 11

$$M\mu(\Delta) \ge (1-\alpha_t(\Delta))M\mu + \sigma\alpha_t(\Delta)M\mu - 4M\Delta^2.$$

Therefore, (19) holds whenever

$$(M_h + M_\ell + 4M)\Delta^2 \le \sigma\alpha_t(\Delta)M\mu,$$

which, by (13), is implied by

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

which in turn is true, by (37), if

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

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

$$\mu_{\rm crit}^2 \stackrel{\rm def}{=} \frac{\sigma M}{(M_h + M_\ell + 4M)C^2(M+n)^2},$$

replacing the role of $\mu_{\rm crit}^1$, and

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

3. Finally we prove that (20) holds for all Δ such that (34) is satisfied. We know from part 1 that (35) and (36) are verified if Δ is such that (34) holds. It follows from (36) that

$$4\Delta^2 < \big(\alpha_n(\Delta) - \alpha_t(\Delta)(1-\sigma)\big)\mu$$

So, from (35), we get

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

for all Δ for which (34) is satisfied, and assertion (20) follows trivially.

Acknowledgments

This research collaboration on the globalization of interior-point methods using the filter technique resulted from conversations held during the First Workshop on Nonlinear Optimization: "Interior-Point and Filter Methods", organized in Coimbra in October 1999.

The authors would also like to acknowledge Matthias Heinkenschloss (Rice University) for his support and insightful comments.

Large parts of this work were done while the authors were visiting the Center for Research on Parallel Computation and the Department of Computational and Applied Mathematics at Rice University, which provided a very stimulating research environment. The authors would like to thank John Dennis and Matthias Heinkenschloss for their hospitality and support.

References

1. M. Argaez and R. A. Tapia, *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. C. Audet and J. E. Dennis, *A pattern search filter method for nonlinear programming without derivatives*, Tech. Report TR00–09, Department of Computational and Applied Mathematics, Rice University, 2000.
3. P. T. Boggs, *Sequential quadratic programming*, in Acta Numerica 1995, A. Iserles, ed., Cambridge University Press, Cambridge, London, New York, 1995, pp. 1–51.
4. R. H. Byrd, M. E. Hribar, and J. Nocedal, *An interior point algorithm for large-scale nonlinear programming*, SIAM J. Optim., 9 (1999), pp. 877–900.
5. R. H. Byrd, G. Liu, and J. Nocedal, *On the local behavior of an interior point method for nonlinear programming*, Tech. Report OTC 98/02, Optimization Technology Center, Northwestern University, 1998.
6. T. F. Coleman and Y. Li, *On the convergence of interior–reflective Newton methods for nonlinear minimization subject to bounds*, Math. Programming, 67 (1994), pp. 189–224.
7. A. R. Conn, N. Gould, D. Orban, and P. L. Toint, *A primal-dual trust-region algorithm for minimizing a non-convex function subject to general inequality and linear equality constraints*, Tech. Report 99/04, Dept. of Mathematics, FUNDP, Namur, 1999.
8. 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*, J. Optim. Theory Appl., 89 (1996), pp. 507–541.
9. R. Fletcher, N. Gould, S. Leyffer, and P. L. Toint, *Global convergence of trust-region SQP-filter algorithms for general nonlinear programming*, Tech. Report 99/03, Département de Mathématique, FUNDP, Namur, 1999.
10. R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function*, Tech. Report NA/171, Department of Mathematics, University of Dundee, 1997.
11. ———, *User manual for filterSQP*, Tech. Report NA/181, Department of Mathematics, University of Dundee, 1998.
12. R. Fletcher, S. Leyffer, and P. L. Toint, *On the global convergence of an SLP-filter algorithm*, Tech. Report 98/13, Département de Mathématique, FUNDP, Namur, 1998.
13. A. Forsgren and P. E. Gill, *Primal-dual interior methods for nonconvex nonlinear programming*, SIAM J. Optim., 8 (1998), pp. 1132–1152.
14. D. M. Gay, M. L. Overton, and M. H. Wright, *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.
15. M. Heinkenschloss, M. Ulbrich, and S. Ulbrich, *Superlinear and quadratic convergence of affine-scaling interior-point Newton methods for problems with simple bounds without strict complementarity assumption*, Math. Programming, 86 (1999), pp. 615–635.

16. H. J. MARTINEZ, Z. PARADA, AND R. A. TAPIA, *On the characterization of q-superlinear convergence of quasi-Newton interior-point methods for nonlinear programming*, Boletin de la Sociedad Matematica Mexicana, 1 (1995), pp. 1–12.
17. J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, Springer–Verlag, Berlin, Heidelberg, New York, Tokyo, 1999.
18. P. TSENG, *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.
19. S. ULBRICH AND M. ULBRICH, *Nonmonotone trust region methods for nonlinear equality constrained optimization without a penalty function*, Tech. Report, Department of Computational and Applied Mathematics, Rice University, 2000.
20. R. J. VANDERBEI AND D. F. SHANNO, *An interior-point algorithm for nonconvex nonlinear programming*, Comput. Optim. and Appl., 13 (1999), pp. 231–252.
21. L. N. VICENTE, *Local convergence of the affine-scaling interior-point algorithm for nonlinear programming*, Comput. Optim. and Appl., (2000, to appear).
22. L. N. VICENTE AND S. J. WRIGHT, *Local convergence of a primal-dual method for degenerate nonlinear programming*, Tech. Report 00-05, Departamento de Matemática, Universidade de Coimbra, 2000.
23. S. J. WRIGHT, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, 1997.
24. H. YAMASHITA AND H. YABE, *Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization*, Math. Programming, 75 (1996), pp. 377–397.