Rapid infeasibility detection in a mixed logarithmic barrier-augmented Lagrangian method for nonlinear optimization

Paul Armand, Ngoc Nguyen Tran

▶ To cite this version:

HAL Id: hal-01785374

https://hal.science/hal-01785374v1

Submitted on 4 May 2018

Rapid infeasibility detection in a mixed logarithmic barrier-augmented Lagrangian method for nonlinear optimization

Paul Armand[a] and Ngoc Nguyen Tran[b]

[a,b] Laboratoire XLIM, Université de Limoges, 123 av. A. Thomas, 87060, Limoges, France

ABSTRACT
We present a modification of a primal-dual algorithm [7] based on a mixed augmented Lagrangian and a log-barrier penalty function. The goal of this new feature is to quickly detect infeasibility. An additional parameter is introduced to balance the minimization of the objective function and the realization of the constraints. The rules of updating the parameters are based on [8]. The global convergence of the modified algorithm is analysed under mild assumptions. We also show that under a suitable choice of the parameters along the iterations, the rate of convergence of the algorithm to an infeasible stationary point is superlinear. This is the first local convergence result for the class of interior point methods in the infeasible case. We finally report some numerical experiments to show that this new algorithm is quite efficient to detect infeasibility and does not deteriorate the overall behavior in the general case.

KEYWORDS
infeasibility detection, nonlinear optimization, primal-dual methods, interior-point method, augmented Lagrangian method

1. Introduction

Nonlinear optimization algorithms focus on two tasks: minimizing the objective function and satisfying the constraints. When the set of the optimal solutions is nonempty, many efficient algorithms in the literature are designed to find a local optimal solution of the optimization problem. On the other hand, infeasible instances also appear quite a lot. They can arise, for example, in mathematical modeling, by varying the parameters of a model to study the system response. Infeasibility may also appear when solving a sequence of subproblems in an algorithm like a branch-and-bound method. Even if the problem is feasible, a numerical optimization algorithm may encounter difficulties in finding a feasible point. In that case it would be nice to quickly return an infeasible stationary point, to avoid long sequence of iterations or convergence to a spurious solution. In this context, a rapid infeasibility detection becomes an important issue in nonlinear optimization.

In the literature, a variety of algorithms using different approaches has been proposed to handle infeasibility. Martínez and Prudente [21], Birgin et al. [9, 10] and Gonçalves et al. [19] investigated infeasibility detection for augmented Lagrangian algorithms. Byrd et al. [12] and Burke et al. [11] have also proposed infeasibility detection

---

strategies for sequential quadratic programming methods. Especially, local convergence analyses in the infeasible case have been done in these papers. In the framework of interior point method, the solver IPOPT [25], which is a filter line search algorithm, uses a feasibility restoration phase. Besides finding a new acceptable iterate to the filter, a feature of this phase is to detect (local) infeasibility. The local convergence of IPOPT near an infeasible stationary point was not made. The algorithm of Curtis [14] is a combination of a penalty method and an interior point method. The numerical results showed that this approach is efficient to detect infeasibility. Nocedal et al. [22] presented a trust region interior point method with two phases: a main phase (solution of the barrier problem with a decrease of the barrier parameter) and the feasibility phase (minimization the feasibility violation measure). Numerical experiments showed that the new method is better than the older one KNITRO/CG [13] in detecting infeasible problems and there is no loss of robustness in solving feasible problems. Despite good numerical results obtained, there is no complete global or local convergence analysis for two latter algorithms ([14, 22]).

Recently, Armand and Omheni [7] proposed a nonlinear optimization algorithm, called SPDOPT, which is a mix of an interior point method and of an augmented Lagrangian method. The capability of this algorithm to detect infeasibility is closely related to the behaviors of the penalty parameter and of the dual variables. In particular, an infeasible stationary point can be detected when the sequence of dual variables tends to infinity and the penalty parameter converges to zero. Nevertheless, a fast infeasibility detection has not been observed in practice or theoretically proved for this algorithm. This fact motivated us to modify SPDOPT in order to accelerate the infeasibility detection without losing its good performances in the feasible case. More specifically, a new parameter, called the feasibility parameter, is introduced to balance the minimization of the barrier function and the realization of the equality constraints. If a nearly feasible point is detected, the feasibility parameter remains constant and our algorithm acts as the original algorithm. When the algorithm tends to an infeasible stationary point, the feasibility parameter acts as a barrier parameter. In this case, the exact solution of the perturbed system parametrized by the feasibility parameter defines a smooth trajectory. With a suitable rule of updating the feasibility parameter, the iterates tangentially follow this trajectory. Consequently, when the sequence of iterates converges to an infeasible stationary point, the algorithm can achieve a superlinear rate of convergence. To the best of our knowledge, this is the first local convergence analysis in the infeasible case related to interior point methods. The idea to introduce a new parameter is inspired from [8]. This one is a modification of SPDOPT-AL [6]. Besides analyzing the local convergence near an infeasible stationary point, numerical experiments demonstrated a better performance of the modified algorithm compared with the predecessor in detecting infeasibility.

The paper is organized as follows. The current section is an introduction with some notations which will be used throughout the paper. The algorithm is described in the next section. Section 3 and Section 4 are devoted to the global and the local convergence analysis. The implementation of our algorithm and some numerical experiments are reported in the last section.

Notation

Vector inequalities are understood componentwise. For a vector x, the capital letter stands for the diagonal matrix $X = \text{diag}(x)$ and the ith component of this vector is

denoted by $[x]_i$ or simply x_i when there is no ambiguity. The letter e is used for the vector of all ones. For two real vectors x and y of same lengths in $\mathbb{R}^n$, $x^\top y$ is their Euclidean scalar product and $\|x\| = (x^\top x)^{1/2}$ is the associated norm. The Hadamard product of two vectors x and y is defined by $[x \circ y]_i = [x]_i[y]_i$ for all indices i. The notation $x \perp y$ means that the vectors x and y are orthogonal. For a real matrix M, the induced matrix norm is $\|M\| = \max\{\|Md\| : \|d\| \le 1\}$. The inertia of a real symmetric matrix M, denoted $\text{In}(M) := (\iota_+, \iota_-, \iota_0)$, is the numbers of positive, negative and null eigenvalues. For a function f and an iterate x_k, to simplify the notation we denote $f_k = f(x_k)$. Likewise, f^* stands for $f(x^*)$, and so on. The positive part of a real number r is defined by $r^+ = \max\{r, 0\}$. Let $\{a_k\}$ and $\{b_k\}$ be nonnegative scalar sequences. We write $a_k = \text{O}(b_k)$, or equivalently $b_k = \Omega(a_k)$, if there exists a constant $c > 0$ such that $a_k \le c\, b_k$ for all $k \in \mathbb{N}$. We also write $a_k = \Theta(b_k)$ if $a_k = \text{O}(b_k)$ and $a_k = \Omega(b_k)$. The notation $a_k = \text{o}(b_k)$ means that there exists a positive sequence $\{\epsilon_k\}$ converging to zero such that $a_k = \epsilon_k b_k$ for all $k \in \mathbb{N}$.

2. Algorithm

Consider the following nonlinear optimization problem:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ f(x) \ \text{ subject to } c(x) = 0 \text{ and } x \ge 0, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable functions. Any optimization problem with equality and inequality constraints can be formulated under this form by possibly adding slack variables and splitting free variables. Let us recall some basic definitions about stationarity. A point $x \in \mathbb{R}^n$ is called a Fritz-John (FJ) point of problem (1) if there exists a nonzero vector $(u, y, z) \in \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^n$ such that

$$ug(x) + A(x)y - z = 0, \quad u \ge 0, \quad c(x) = 0 \quad \text{and} \quad 0 \le x \perp z \ge 0,$$

where we denote $g(x) = \nabla f(x)$ the gradient of f at x and $A(x) = \nabla c(x)$ the transpose of the Jacobian matrix of c at x. When $u > 0$, a FJ point x is called a Karush-Kuhn-Tucker (KKT) point. By dividing the first equation by u, we see that y/u and z/u are the vectors of Lagrange multipliers associated with the equality and the bound constraints. Whenever $u = 0$, a FJ point x is called a singular stationary point. Equivalently, a singular stationary point is a feasible solution for (1) at which the Mangasarian-Fromovitz constraint qualification (MFCQ) does not hold. The point x is called an infeasible stationary point, if there exists $z \in \mathbb{R}^n$ such that

$$c(x) \ne 0, \quad A(x)c(x) - z = 0 \quad \text{and} \quad 0 \le x \perp z \ge 0.$$

In other words, an infeasible stationary point is not feasible for (1), but is a KKT point for the feasibility problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|c(x)\|^2 \ \text{ subject to } x \ge 0. \tag{2}$$

3

To the problem (1), we associate the mixed penalty function

$$\varphi_{\rho,\lambda,\sigma,\mu}(x) = \rho f(x) + \lambda^\top c(x) + \tfrac{1}{2\sigma}\|c(x)\|^2 - \rho\mu \sum_i \log x_i, \tag{3}$$

where $\rho > 0$ is the *feasibility parameter*, $\lambda \in \mathbb{R}^m$ is an estimate of the vector of Lagrange multipliers associated with the equality constraints, $\sigma > 0$ is the *quadratic penalty parameter* and $\mu > 0$ is the *barrier parameter*. This penalty function is a mixed of the augmented Lagrangian and of the logarithmic barrier function. It can be interpreted as the augmented Lagrangian associated with the log-barrier problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \rho(f(x) - \mu \sum_i \log x_i), \quad \text{subject to } c(x) = 0. \tag{4}$$

The first order optimality conditions for (3) are

$$\rho g(x) + A(x)(\lambda + \tfrac{1}{\sigma}c(x)) - \rho\mu X^{-1}e = 0. \tag{5}$$

By introducing the dual variables $y = \lambda + \tfrac{1}{\sigma}c(x)$ and $z = \rho\mu X^{-1}e$, the equation (5) is equivalently formulated as

$$\Phi(w,\lambda,\rho,\sigma,\mu) := \begin{pmatrix} \rho g(x) + A(x)y - z \\ c(x) + \sigma(\lambda - y) \\ XZe - \rho\mu e \end{pmatrix} = 0, \tag{6}$$

where $w := (x,y,z) \in \mathbb{R}^N$, with $N = n + m + n$. This primal-dual system of equation can be seen as a perturbation of the FJ optimality conditions of problem (1). The equality constraints is perturbed thanks to the term $\sigma(\lambda - y)$. Note that when $\lambda = y$, this system is the primal-dual optimality system associated with the problem (4). Each complementarity products is perturbed thanks to the term $\rho\mu$. An important feature of our approach is that each time the parameters are updated, either ρ is reduced, or μ is reduced, but not both. We will see that, under some usual regularity assumptions, this allows to guarantee the superlinear convergence in case of convergence to a local minimum or to an infeasible stationary point.

The algorithm involves applying a Newton-type method for the solution of the system $\Phi = 0$, while updating the parameters along the iterations to control the convergence to a FJ point or to an infeasible stationary point. There are two kinds of iteration: outer and inner. At an outer iteration, the parameters are updated, while at an inner iteration the parameters are kept constant. At each iteration, outer or inner, a candidate iterate w^+ is computed as a solution of the following linear system:

$$J_{\rho,\theta,\delta}(w)(w^+ - w) = -\Phi(w,\lambda,\rho,\sigma,\mu)$$

where the matrix is defined as

$$J_{\rho,\theta,\delta}(w) := \begin{pmatrix} H_{\rho,\theta}(x,y) & A(x) & -I \\ A(x)^\top & -\delta I & 0 \\ Z & 0 & X \end{pmatrix},$$

with

$$H_{\rho,\theta}(x,y) := \rho \nabla^2 f(x) + \sum_{i=1}^{m} y_i \nabla^2 c_i(x) + \theta I.$$

The regularization parameters $\theta \geq 0$ and $\delta > 0$ are chosen such that the matrix

$$K_{\rho,\theta,\delta}(w) := H_{\rho,\theta}(x,y) + X^{-1}Z + \frac{1}{\delta} A(x)A(x)^\top$$

is positive definite. This property implies that not only the matrix $J_{\rho,\theta,\delta}(w)$ is nonsingular, but also that the solution of the linear system solved during the inner iterations is a descend direction for some merit function.

We now describe in detail the outer iteration algorithm. At the beginning, a starting point $w_0 = (x_0, y_0, z_0) \in \mathbb{R}^{2n+m}$ satisfying $v_0 = (x_0, z_0) > 0$ is chosen, then we set $\lambda_0 = y_0$. Besides, we choose $\sigma_0 > 0$, $\mu_0 > 0$, $\rho_0 = 1$ and four constants κ, d, $\bar{\tau} \in (0,1)$ and $l \in \mathbb{N}$. The outer iteration counter is set to $k = 0$. A feasibility tolerance $\epsilon > 0$ must be chosen. A flag is used to indicate if the algorithm is in the feasibility detection phase ($\mathtt{F} = 1$) or not ($\mathtt{F} = 0$). Initially $\mathtt{F} = 1$, then this value is kept until a feasible, or nearly feasible, point has been detected. The index i_k stores the last iteration number prior to k at which inequality (7) is satisfied. It is initialized to $i_0 = 0$.

The algorithm can be seen as an extension of [7, Algorithm 1]. The first four steps are related to the updating of the parameters and are drawn from the feasibility detection algorithm for the equality constrained case proposed in [8].

The first step aims to detect a feasible or nearly feasible point. Whenever this is the case, the flag is set to $\mathtt{F} = 0$ and this value is kept for all further iterations. In this case, the algorithm is exactly the same as [7, Algorithm 1]. In particular, the sequence $\{\rho_k\}$ is eventually constant. It is worth noting that this switching mechanism is necessary to avoid the undesirable situation in which the condition (7) is alternatively satisfied and not satisfied an infinite number of times, for example when the feasibility measure decreases very slowly. In this case, it would be difficult to distinguish between a convergence to a KKT point or to a singular stationary point. Moreover, in practice, a point is deemed to be feasible if the norm of the constraints is smaller than some predefined tolerance. Suppose, for example, that we want to minimize x in $\mathbb{R}$, subject to the constraint $e^x \leq 0$. Most of well established softwares for nonlinear numerical optimization return a message like "optimal solution found" when solving this problem. Hence, it seems natural to state that the problem is feasible, whenever the feasibility tolerance is small enough.

Depending on the reduction of the feasibility measure, the trial values ρ_k^+ and μ_k^+ for the feasibility and barrier parameters, as well as the new values σ_{k+1} and λ_{k+1} for the augmented Lagrangian parameters, are chosen. If the inequality (7) is satisfied, the feasibility parameter is kept constant, new values the barrier and quadratic penalty parameters are chosen, the Lagrange multiplier estimate is set to the current value of the dual variable. As shown in [6], from a theoretical point of view and in practice, it is better to force the convergence of $\{\sigma_k\}$ to zero to get a rapid rate of convergence. On the other hand, if the condition (7) is not satisfied, then there are two situations. The first situation is when the algorithm is still in the feasibility detection phase ($\mathtt{F} = 1$). In that case, the feasibility parameter is sufficiently reduced, while the barrier and quadratic penalty parameters are kept constant, and then the Lagrange multiplier estimate is rescaled. When these updates are always done from some iteration, this scaling of the

Algorithm 1 (kth outer iteration)

(1) If $\|c_k\| \le \epsilon$, then set $\mathtt{F} = 0$.

(2) Choose $\zeta_k > 0$ such that $\{\zeta_k\} \to 0$. If $k = 0$ or

$$\|c_k\| \le \kappa \max\{\|c_{i_j}\| : (k-l)^+ \le j \le k\} + \zeta_k \tag{7}$$

then set $i_{k+1} = k$ and go to Step 4, otherwise set $i_{k+1} = i_k$.

(3) If $\mathtt{F} = 1$, then choose $0 < \rho_k^+ \le d\rho_k$ and set $\sigma_{k+1} = \sigma_k, \mu_k^+ = \mu_k, \lambda_{k+1} = \rho_k^+ \lambda_k$, else set $\rho_k^+ = \rho_k$, choose $0 < \sigma_{k+1} \le d\sigma_k, 0 < \mu_k^+ \le d\mu_k$ and set $\lambda_{k+1} = \lambda_k$. Go to Step 5.

(4) Set $\rho_k^+ = \rho_k$, choose $0 < \sigma_{k+1} \le \sigma_k$ and $0 < \mu_k^+ \le d\mu_k$. Set $\lambda_{k+1} = y_k$.

(5) If $\rho_k^+ = \rho_k$, then choose $\delta_k = \Omega(\mu_k)$, else set $\delta_k = \sigma_k$. Choose regularization parameter $\theta_k \ge 0$ such that $K_{\rho_k^+, \theta_k, \delta_k}(w_k) \succ 0$. Set $J_k = J_{\rho_k^+, \theta_k, \delta_k}(w_k)$. Compute w_k^+ by solving the linear system

$$J_k(w_k^+ - w_k) = -\Phi(w_k, \lambda_{k+1}, \rho_k^+, \sigma_{k+1}, \mu_k^+). \tag{8}$$

(6) Choose $\tau_k \in [\bar{\tau}, 1[$. Compute α_k as the largest $\alpha \in (0,1]$ such that

$$v_k + \alpha(v_k^+ - v_k) \ge (1 - \tau_k)v_k, \tag{9}$$

where $v_k = (x_k, z_k)$. Choose a vector $a_k = (a_k^x, a_k^y, z_k^z) \in [\alpha_k, 1]^N$ such that $v_k + a_k^v \circ (v_k^+ - v_k) > 0$, where $a_k^v = (a_k^x, a_k^z)$.

(7) Set $\widehat{w}_k = w_k + a_k \circ (w_k^+ - w_k)$, $(\rho_{k+1}, \mu_{k+1}) = (\rho_k, \mu_k) + \alpha_k(\rho_k^+ - \rho_k, \mu_k^+ - \mu_k)$.

(8) Choose $\varepsilon_k > 0$ such that $\{\varepsilon_k\} \to 0$. If $\|\Phi(\widehat{w}_k, \lambda_{k+1}, \rho_{k+1}, \sigma_{k+1}, \mu_{k+1})\| \le \varepsilon_k$, then set $w_{k+1} = \widehat{w}_k$. Otherwise, apply the inner iteration algorithm to find w_{k+1} such that

$$\|\Phi(w_{k+1}, \lambda_{k+1}, \rho_{k+1}, \sigma_{k+1}, \mu_{k+1})\| \le \varepsilon_k \tag{10}$$

Lagrange multiplier estimate is useful for the convergence of the iterates to an infeasible stationary point (see Theorem 3.1-ii). The second situation is when the algorithm has left the feasibility detection phase ($\mathtt{F} = 0$). In that case, the feasibility parameter is kept constant, the barrier parameter is reduced as in a standard interior point method, the quadratic penalty parameter is also reduced to penalize the constraints violation and the Lagrange multiplier estimate is kept constant as in a classical augmented Lagrangian algorithm.

At Step 5, the choice of the regularization parameter δ_k of the regularized Jacobian matrix is done as follows. When the feasibility parameter is unchanged, because (7) is satisfied or $\mathtt{F} = 0$, we set $\delta_k = \Omega(\mu_k)$. This choice is imposed by the global convergence theory of the algorithm in the feasible case, see [5, Theorem 3.3] and [7, Theorem 4.2]. In case the feasibility parameter is reduced, we set $\delta_k = \sigma_k$. This choice is motivated to get a rapid convergence when the sequence of iterates converges to an infeasible stationary point, see Lemma 4.3 below.

Once the Newton iterate w_k^+ is computed at Step 5, the fraction to the boundary rule is applied to ensure the positivity of the primal and dual variables. As in [5], the step length can be selected componentwise to calculate the trial iterate $\widehat{w}_k$. The

values of barrier and feasibility parameters are then updated according to formulae at Step 7. These formulae avoid too large discrepancies between the magnitude of these parameters and the one of $\|\Phi\|$ and increase robustness [4].

Finally, at Step 8, if the trial iterate $\widehat{w}_k$ satisfies a sufficient reduction of the residual norm of the perturbed optimality conditions, then $w_{k+1} = \widehat{w}_k$. If this is not the case, a sequence of inner iterations with all the parameters fixed to their current values will be carried out to find the new iterate w_{k+1}.

The inner iteration algorithm is a backtracking line search applied to the primal-dual merit function

$$\mathcal{M}_{\rho,\lambda,\sigma,\mu}(w) = \varphi_{\rho,\lambda,\sigma,\mu}(x) + \nu_1\psi_{\sigma,\lambda}(x,y) + \nu_2\pi_{\rho,\mu}(x,z)$$

where $\varphi_{\rho,\lambda,\sigma,\mu}(x)$ is defined by (3),

$$\psi_{\sigma,\lambda}(x,y) = \frac{1}{2\sigma}\|c(x) + \sigma(\lambda - y)\|^2 \quad \text{and} \quad \pi_{\rho,\mu}(x,z) = x^\top z - \rho\mu\sum_{i=1}^{p}\log(x_iz_i),$$

where $\nu_1 > 0, \nu_2 > 0$ are scaling parameters. This is motivated by the fact that the first order optimality conditions for minimizing this merit function correspond to (6). We refer the reader to [7, Algorithm 2] for a complete description of this algorithm. To simplify the presentation, we consider that the quadratic parameter σ_{k+1} remains constant all along the inner iterations, while in [7] it can be increased. This choice has no impact from the theoretical point of view and in our numerical experiments, the value of this parameter is allowed to increase during the inner iterations. It has been shown that if the sequence of primal inner iterates remains in a compact set, then the inner iteration algorithm succeeds in finding an iterate that satisfies (10) after a finite number of iterations [7, Theorem 4.1].

We end this section by showing some properties related to the behavior of the parameters with respect to the feasibility detection by means of the criterion (7).

Lemma 2.1. *Assume that Algorithm 1 generates an infinite sequence of iterates $\{w_k\}$. Let $\mathcal{K} \subset \mathbb{N}$ be the set of iteration indices at which the inequality (7) is satisfied.*

(i) *If $\mathcal{K}$ is infinite, then $\{c_k\}_{k\in\mathcal{K}}$ converges to zero and $\{\rho_k\}$ is eventually constant.*
(ii) *If $\mathcal{K}$ is finite, then $\liminf\|c_k\| > 0$.*

In addition, suppose that the sequence $\{H_{\rho_k^+,\theta_k}(x_k,y_k), g_k, A_k\}$ is bounded and that the matrices $K_{\rho_k^+,\theta_k,\delta_k}(w_k)$ are uniformly positive definite for $k \in \mathbb{N}$.

(iii) *If $\mathcal{K}$ is infinite, then the sequence $\{\mu_k\}$ converges to zero.*
(iv) *If $\mathcal{K}$ is finite, then the sequences $\{\sigma_k\rho_k\}$ and $\{\sigma_k\lambda_k\}$ converge to zero.*

Proof. The outcomes (i) and (ii) are proved in [8, Lemma 1], while (iii) is a direct consequence of [7, Theorem 4.2 (iv)].

To prove (iv), suppose that $\mathcal{K}$ is finite and let us define $k_0 := \max\mathcal{K}$. Two cases are considered. The first case is when $\|c_k\| \leq \epsilon$ at some iteration k. We then have $\mathtt{F} = 0$ for all further iterations. The update of the parameters at Step 3 implies that both sequences $\{\rho_k\}$ and $\{\lambda_k\}$ are eventually constant and $\{\sigma_k\}$ tends to zero. It follows that the two sequences $\{\sigma_k\rho_k\}$ and $\{\sigma_k\lambda_k\}$ converge to zero. The second case is when $\|c_k\| > \epsilon$ for all $k \in \mathbb{N}$, which implies that $\mathtt{F} = 1$ at each iteration. In that case, for all $k \geq k_0 + 1$, $\rho_k^+ \leq d\rho_k$, $\sigma_{k+1} = \sigma_{k_0}$, $\mu_{k+1} = \mu_{k_0}$ and $\|\lambda_{k+1}\| \leq d\rho_k\|\lambda_k\| \leq d\rho_k\|y_{k_0}\|$.

By using similar arguments as in the proof of [7, Theorem 4.2 (iv)] and [5, Theorem 3.3], we will show that $\{\rho_k\}$ converges to zero, which will imply that the two sequences $\{\sigma_k\rho_k\}$ and $\{\sigma_k\lambda_k\}$ converge to zero. The proof is by contradiction by supposing that $\{\rho_k\}$ is bounded away from zero by some constant $\bar{\rho} > 0$. For all $k \geq k_0$, we have $\rho_k^+ \leq d\rho_k$ with $d \in (0, 1)$. From Step 7 of Algorithm 1, for all $k \geq k_0$ we have

$$\begin{aligned} \rho_{k+1} &= \rho_k + \alpha_k(\rho_k^+ - \rho_k) \\ &\leq (1 - (1-d)\alpha_k)\rho_k. \end{aligned}$$

Because $\{\rho_k\}$ is supposed to be bounded away from zero, this inequality implies that $\{\alpha_k\}$ converges to zero. We will get a contradiction by showing that $\{\alpha_k\}$ is bounded away from zero. The inequality (33) in [7, Theorem 4.2] shows that for all k large enough, we have

$$\frac{\bar{\tau}}{1 + \|w_k^+ - w_k\|/\sqrt{\bar{\rho}}} \leq \alpha_k. \tag{11}$$

Recall that the inequality (10) is satisfied at each iteration with a sequence $\{\varepsilon_k\}$ going to zero. Therefore, the sequence $\{\Phi(w_k, \lambda_k, \rho_k, \sigma_k, \mu_k)\}$ converges to zero. In particular, $\{X_k Z_k e - \rho_k \mu_k e\}$ tends to zero. Consequently, for k large enough

$$[x_k]_i[z_k]_i \geq \frac{\bar{\rho}\mu_{k_0}}{2}, \quad \text{for all } i = 1, ..., n.$$

Keeping in mind all the assumptions and the previous lower bound on $\{x_k \circ z_k\}$, [2, Theorem 1] shows that there exists a constant $K_1 > 0$, such that for all $k \geq 0$

$$\|J_{\rho_k^+, \theta_k, \delta_k}(w_k)^{-1}\| \leq K_1. \tag{12}$$

From the definition of $\Phi(\cdot)$, for all $k \geq 0$ we have

$$\begin{aligned} \Phi(w_k, \lambda_{k+1}, \rho_k^+, \sigma_{k+1}, \mu_k^+) &= \Phi(w_k, \lambda_k, \rho_k, \sigma_{k+1}, \mu_k^+) \\ &\quad + ((\rho_k^+ - \rho_k)g_k^\top, \sigma_{k+1}(\lambda_{k+1} - \lambda_k)^\top, (\rho_k - \rho_k^+)\mu_k^+ e^\top)^\top. \end{aligned}$$

By using (8), (10), (12) and noting that $\sigma_{k+1} = \sigma_k = \sigma_{k_0}$, $\mu_k^+ = \mu_k = \mu_{k_0}$, $\lambda_{k+1} = \rho_k^+ \lambda_k$ and $\rho_k^+ \leq \rho_k \leq 1$, for k large enough we then get

$$\begin{aligned} \|w_k^+ - w_k\| &= \|J_{\rho_k^+, \theta_k, \delta_k}(w_k)^{-1}\Phi(w_k, \lambda_{k+1}, \rho_k^+, \sigma_{k+1}, \mu_k^+)\| \\ &\leq K_1(\|\Phi(w_k, \lambda_k, \rho_k, \sigma_k, \mu_k)\| + |\rho_k - \rho_k^+|(\|g_k\| + \mu_{k_0}\|e\|) + \sigma_{k_0}|\rho_k^+ - 1|\|\lambda_k\|) \\ &\leq K_1(\varepsilon_k + \|g_k\| + \mu_{k_0}\|e\| + \sigma_{k_0}\|\lambda_k\|). \end{aligned}$$

Because $\{\varepsilon_k\}$ tends to zero, $\{g_k\}$ is bounded and $\|\lambda_k\| \leq d\|y_{k_0}\|$, we deduce that $\|w_k^+ - w_k\|$ is bounded from above, which contradicts inequality (11) and the fact that $\{\alpha_k\}$ is supposed to converge to zero. We then deduce that $\rho_k \to 0$, which completes the proof of (iv). $\qquad\square$

3. Global convergence analysis

For the global convergence of the inner iterations see [7, Theorem 4.1]. Roughly speaking, this theorem shows that under some standard assumptions, the inner iteration algorithm is able to find a new iterate w_{k+1} satisfying the stopping test (10) after a finite number of iterations. Hence, we can assume that the inner iteration algorithm successfully terminates each time it is called at Step 8.

We then have the following result about the global convergence of Algorithm 1.

Theorem 3.1. *Assume that all the assumptions of Lemma 2.1 are satisfied and that Algorithm 1 generates an infinite sequence $\{w_k\}$. Let $\mathcal{K} \subset \mathbb{N}$ be the set of iteration indices for which the condition (7) is satisfied.*

(i) If $\mathcal{K}$ is infinite, then $\rho_k = \bar\rho > 0$ for k large enough and the iterates approach stationary of the problem (1), i.e., the sequences $\{\bar\rho g_k + A_k y_k - z_k\}$, $\{c_k\}_{\mathcal{K}}$ and $\{X_k z_k\}$ converge to zero.

(ii) If $\mathcal{K}$ is finite, then $\{\|c_k\|\}$ is bounded away from zero and the sequence of primal iterates approaches stationarity of the feasibility problem, i.e., there exists a sequence $\{u_k\} \subset \mathbb{R}_+^n$ such that $\lim \|(A_k c_k - u_k, X_k u_k)\| = 0$.

Proof. Let us denote $\Phi_k := \Phi(w_k, \lambda_k, \rho_k, \sigma_k, \mu_k)$ for $k \in \mathbb{N}$. Step 8 of Algorithm 1 implies that $\{\Phi_k\}$ converges to zero.

Suppose that $\mathcal{K}$ is infinite. Lemma 2.1-(i) shows that $\{c_k\}_{\mathcal{K}}$ tends to zero and that $\rho_k = \bar\rho$ for sufficiently large k. The first block of Φ_k is $\rho_k g_k + A_k y_k - z_k$, therefore $\lim \bar\rho g_k + A_k y_k - z_k = 0$. The third block of Φ_k is $X_k Z_k e - \rho_k \mu_k e$. Lemma 2.1-(iii) implies that $\{\mu_k\}$ tends to zero. Therefore $\{X_k z_k\}$ tends to zero, which concludes the proof of outcome (i).

Suppose now that $\mathcal{K}$ is finite. Lemma 2.1-(ii) implies that the sequence $\{\|c_k\|\}$ is bounded away from zero. Let us define $u_k := \sigma_k z_k > 0$ for $k \in \mathbb{N}$. For all $k \in \mathbb{N}$, we have

$$A_k c_k - u_k = \sigma_k(\rho_k g_k + A_k y_k - z_k) - \sigma_k \rho_k g_k + A_k(c_k + \sigma_k(\lambda_k - y_k)) - \sigma_k A_k \lambda_k$$

and

$$X_k u_k = \sigma_k(X_k Z_k e - \rho_k \mu_k e) + \sigma_k \rho_k \mu_k e.$$

By taking the norm on both sides, for all k we then get

$$\|A_k c_k - u_k\| + \|X_k u_k\| \leq (2\sigma_k + \|g_k\| + 2\|A_k\| + \mu_k \|e\|) \max\{\|\Phi_k\|, \sigma_k \rho_k, \sigma_k \|\lambda_k\|\}.$$

By assumptions, the sequences $\{g_k\}$ and $\{A_k\}$ are bounded. The first assertion of this proof and Lemma 2.1-(iv) imply that the second term of this inequality tends to zero. We then deduce that $\{(A_k c_k - u_k, X_k u_k)\}$ converges to zero, which ends the proof. $\square$

The next result summarizes the behavior of the Algorithm 1 under a mild and usual assumption about the boundedness of the sequence of primal iterates.

Theorem 3.2. *Assume that Algorithm 1 generates an infinite sequence of iterates $\{w_k\}$ such that $\{x_k\}$ lies in a compact set. Assume also that the regularization parameters chosen at Step 5 are such that the matrices $K_{\rho_k^+, \theta_k, \delta_k}(w_k)$ are uniformly positive definite for $k \in \mathbb{N}$.*

9

(i) *Any feasible limit point of $\{x_k\}$ is a Fritz-John point of problem* (1).

(ii) *If $\{x_k\}$ has no feasible limit point, then any limit point is an infeasible stationary point of problem* (1).

Proof. The compactness assumption implies that the sequence $\{(g_k, A_k)\}$ is bounded.

Assume that $\{x_k\}$ has a feasible limit point $\bar{x}$, i.e., $\bar{c} = 0$. Lemma 2.1-(ii) implies that the updating condition (7) is satisfied an infinite number of times. Let $\mathcal{K} \subset \mathbb{N}$ such that $\lim_{\mathcal{K}} x_k = \bar{x}$. From Lemma 2.1-(i), we have $\rho_k = \bar{\rho}$ for k large enough. We consider two situations regarding the boundedness of the sequence $\{y_k\}_{\mathcal{K}}$. Suppose that $\{y_k\}_{\mathcal{K}}$ is bounded. With the compactness of $\{x_k\}$ we have that the sequence $\{H_{\rho_k^+, \theta_k}(x_k, y_k)\}$ is bounded and so Theorem 3.1 applies. The sequence $\{z_k\}$ is also bounded. Indeed, for all $k \in \mathbb{N}$, we have

$$\|z_k\| \le \|\rho_k g_k + A_k y_k - z_k\| + \|\rho_k g_k + A_k y_k\|.$$

The first term on the right hand side tends to zero and the second one is bounded. Consequently, by virtue of Theorem (3.1)-(i), any limit point of the sequence $\{x_k, y_k, z_k\}_{\mathcal{K}}$ satisfies the FJ conditions. Because $\bar{\rho} > 0$, $\bar{x}$ is a KKT point of problem (1). The second situation is when $\{y_k\}_{\mathcal{K}}$ is unbounded. For all $k \in \mathcal{K}$, let $(a_k, b_k) := \frac{(y_k, z_k)}{\|(y_k, z_k)\|}$. Note that $b_k > 0$, because of Step 6 of Algorithm 1. By taking a subsequence if necessary, we can assume that $\lim_{\mathcal{K}}(a_k, b_k) = (\bar{a}, \bar{b}) \ne 0$, with $\bar{b} \ge 0$. For all $k \in \mathbb{N}$ we have

$$\|A_k a_k - b_k\| \le \frac{1}{\|(y_k, z_k)\|} \left(\|\rho_k g_k + A_k y_k - z_k\| + \|\rho_k g_k\| \right)$$

Because $\{\rho_k g_k + A_k y_k - z_k\}$ converges to zero, $\{g_k\}$ is bounded, $\{\rho_k\}$ is eventually constant and $\{y_k\}_{\mathcal{K}}$ is unbounded, we get $\bar{A}\bar{a} - \bar{b} = 0$. For all $k \in \mathbb{N}$ we have

$$\|X_k b_k\| = \frac{1}{\|(y_k, z_k)\|}(\|X_k Z_k e - \rho_k \mu_k e\| + \rho_k \mu_k \|e\|)$$

Because $\{X_k Z_k e - \rho_k \mu_k e\}$ tends to zero and $\{(y_k, z_k)\}_{\mathcal{K}}$ is unbounded, we also get $\bar{X}\bar{b} = 0$. We can conclude that $\bar{x}$ is a singular stationary point of problem (1).

Let us now consider the second outcome for which any limit point of $\{x_k\}$ is infeasible. It follows from Lemma 2.1-(i) that the Step 3 of Algorithm 1 is executed at all iteration $k \ge k_0$ for some $k_0 \in \mathbb{N}$. There are two cases depending on the boundedness of $\{y_k\}$. If $\{y_k\}$ is bounded, then the sequence $\{H_{\rho_k^+, \theta_k}(x_k, y_k)\}$ is bounded. Therefore, Theorem 3.1-(ii) shows that any limit point $\bar{x}$ of $\{x_k\}$ is an infeasible stationary point of (1). In the second case, $\{y_k\}$ is unbounded. From the convergence to zero of the sequence $\{c_k + \sigma_k(\lambda_k - y_k)\}$ and the boundedness of $\{\lambda_k\}$ (since $\|\lambda_k\| \le \|\lambda_{k_0}\|$ for all $k \ge k_0$), we deduce that the sequence $\{\sigma_k\}$ tends to zero. As a consequence, the sequences $\{\sigma_k \rho_k\}$ and $\{\sigma_k \lambda_k\}$ converge to zero. We then obtain the same conclusion as Lemma 2.1-(iv). It suffices to repeat the proof of Theorem (3.1)-(ii) to show that any limit point $\{x_k\}$ is an infeasible stationary point of (1). $\qquad\square$

4. Asymptotic behavior

There are two cases to analyse. The first one is when $\{w_k\}$ converges to a primal-dual solution of problem (1). Because the flag F is switched to zero at some iteration,

the feasibility parameter becomes constant after a finite number of iterations. Consequently, the analysis in [23, Section 1.4] can be directly applied to demonstrate that under some classical assumptions and a suitable choice of the parameters, the rate of convergence of the sequence $\{w_k\}$ is superlinear. The second case is when Algorithm 1 generates a convergent sequence $\{x_k\}$ to an infeasible stationary point $x^* \in \mathbb{R}^n$ of the problem (1). This section concentrates to this case.

4.1. Assumptions and basic results

The first assumption is about the regularity of the problem data. This assumption is standard in our framework.

Assumption 1. *The function f and c are twice continuously differentiable and their second derivatives are Lipschitz continuous over an open neighborhood of x^*.*

To analyse the rate of convergence of the sequence of iterates to an infeasible stationary point, a natural assumption is that the whole sequence of iterates converges to such a point.

Assumption 2. *Algorithm 1 generates an infinite sequence $\{w_k\}$ which converges to $w^* = (x^*, y^*, z^*) \in \mathbb{R}^{2n+m}$, where x^* is an infeasible stationary point of problem (1).*

This assumption implies that the algorithm always stays in feasibility detection phase, i.e., the feasibility flag keeps the value $\mathtt{F} = 1$ all along the iterations. More precisely, some direct consequences are the following.

Lemma 4.1. *Under Assumption 2, $\mathtt{F} = 1$ at each iteration, there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$, $\sigma_k = \sigma := \sigma_{k_0} > 0$ and $\mu_k = \mu := \mu_{k_0} > 0$, $\{\rho_k\}$ converges to zero and $\|\lambda_{k+1}\| = \mathrm{o}(\rho_k^+)$.*

Proof. Assumption 2 implies that $\{c_k\}$ is bounded away from zero. By virtue of Lemma 2.1-(i), the inequality (7) is satisfied only a finite number of times. It follows that Step 3 of Algorithm 1 is always executed for k sufficiently large. For all $k \in \mathbb{N}$, we have

$$\|c_k\| \leq \|c_k + \sigma_k(\lambda_k - y_k)\| + \sigma_k\|\lambda_k\| + \sigma_k\|y_k\|.$$

Inequality (10) implies that the first term of the right-hand side tends to zero. The update of λ_k at Step 3 implies that $\{\lambda_k\}$ remains bounded. Since $\{y_k\}$ is supposed to be convergent, the sequence $\{\sigma_k\}$ cannot go to zero, which implies that $\mathtt{F} = 1$ at each iteration. Hence, there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$, $\sigma_k = \sigma_{k_0}, \mu_k = \mu_{k_0}$ and $\lambda_{k+1}/\rho_k^+ = \rho_{k-1}^+ ... \rho_{k_0}^+ \lambda_{k_0}$. By using the same arguments as in the proof of Lemma 2.1-(iv), we deduce that the sequence $\{\rho_k\}$ converge to zero. Thank to the fact that $\rho_{k-1}^+ \leq \rho_k$, we get $\|\lambda_{k+1}\| = \mathrm{o}(\rho_k^+)$. $\square$

For $w = (x, y, z) \in \mathbb{R}^{2n+m}$ and $\rho > 0$, let us define

$$F(w, \rho) = \begin{pmatrix} \rho g(x) + A(x)y - z \\ c(x) - \sigma y \\ XZe - \rho\mu e \end{pmatrix},$$

where σ and μ are the values defined by Lemma 4.1. From the fact that

$\lim \Phi(w_k, \lambda_k, \rho_k, \sigma_k, \mu_k) = \Phi(w^*, 0, 0, \sigma, \mu) = F(w^*, 0)$, one has $y^* = \frac{1}{\sigma}c^*$, $z^* = \frac{1}{\sigma}A^*c^*$ and $0 \leq x^* \perp z^* \geq 0$.

Let us denote the Hessian matrix of the function $\frac{1}{2}\|c\|^2$ at $x \in \mathbb{R}^n$ by

$$S(x) := \sum_{i=1}^{m} c_i(x)\nabla^2 c_i(x) + A(x)A(x)^\top.$$

The set of active bounds is denoted by $\mathcal{A} := \{i : [x^*]_i = 0\}$.

Assumption 3. *The second order sufficient optimality conditions (SOSC) for the feasibility problem (2) hold at x^*, i.e., $u^\top S^* u > 0$ for all $u \neq 0$ satisfying $[u]_i = 0$ for all $i \in \mathcal{A}$.*

Assumption 4. *Strict complementarity holds at w^*, that is*

$$a := \min\{[x^*]_i + [z^*]_i : i = 1, ..., n\} > 0.$$

The next lemma summarizes some basic results which are direct consequences of these assumptions.

Lemma 4.2. *Under Assumptions 1-4, there exist positive constants r^*, ρ^*, K, C, L, L_1, L_2 and a continuously differentiable function $\mathbf{w} : (-\rho^*, \rho^*) \to \mathbb{R}^N$, such that for all $w, w' \in B(w^*, r^*)$ and for all $\rho, \rho' \in (-\rho^*, \rho^*)$, we have*

(i) $\|F_w'(w, \rho)^{-1}\| \leq K$,
(ii) $F(w, \rho) = 0$ *if and only if* $\mathbf{w}(\rho) = w$,
(iii) $\|\mathbf{w}(\rho) - \mathbf{w}(\rho')\| \leq C|\rho - \rho'|$,
(iv) $\|F_w'(w, \rho) - F_w'(w', \rho)\| \leq L(\|w - w'\|)$,
(v) $L_1\|w - w'\| \leq \|F(w, \rho) - F(w', \rho)\| \leq L_2\|w - w'\|$.

Proof. In order to prove (i), we only need to show that the matrix $F_w'(w^*, 0)$ is nonsingular. Let $u \in \mathbb{R}^N$ such that $F_w'(w^*, 0)u = 0$. By writing $u := (u_1, u_2, u_3)$ and by using the fact that $y^* = \frac{1}{\sigma}c^*$, we have

$$\begin{pmatrix} \frac{1}{\sigma}\sum_i c_i^* \nabla^2 c_i^* & A^* & -I \\ A^{*\top} & -\sigma I & 0 \\ Z^* & 0 & X^* \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The third equation of this linear system and Assumption 4 imply that $[u_1]_i = 0$ for all $i \in \mathcal{A}$ and $[u_3]_i = 0$ for all $i \notin \mathcal{A}$. Consequently, one has $u_1^\top u_3 = 0$. The second equation of the linear system gives us $u_2 = \frac{1}{\sigma}A^{*\top}u_1$. Substituting this equality into the first equation and premultiplying by $u_1^\top$, we get $\frac{1}{\sigma}u_1^\top S^* u_1 = 0$. By Assumption 3 we deduce that $u_1 = 0$, from which we deduce that $u_2 = 0$ and $u_3 = 0$.

The properties (ii) and (iii) are direct consequences of the implicit function theorem.
The Lipschitz continuity of F_w', property (iv), follows from Assumption 1.
The last assertion (v) is a consequence of [16, Lemma 15]. $\qquad \square$

Our asymptotic analysis also requires some specifications on the choice of the feasibility parameter. For all $k \in \mathbb{N}$, the trial value of feasibility parameter ρ_k^+ is chosen

such that

$$\theta_1 \rho_k^{1+t} \le \rho_k^+ \le \theta_2 \rho_k \tag{13}$$

for some constants $\theta_1 > 0$, $\theta_2 \in (0,1)$ and $t \in (0,1)$. From the fact that $\{\rho_k\}$ goes to zero, the left inequality of (13) implies that

$$\rho_k^2 = o(\rho_k^+). \tag{14}$$

At last, the parameter τ_k used in (9) must satisfy the following condition:

$$\lim_{k \to \infty} (1 - \tau_k) \frac{\rho_k}{\rho_k^+} = 0. \tag{15}$$

The conditions (13) and (15) are standard in the asymptotic analysis of this family of primal-dual methods, see e.g., [1, 5]. The left inequality of (13) means that the parameter of the path $\mathbf{w}$ must converge to zero with a subquadratic rate of convergence in order that the Newton iterates, which converge naturally with a quadratic rate, catch the path prior the parameter is nearly zero.

For k large enough, we have $\rho_k^+ < \rho_k$ and $\sigma_k = \sigma$. Therefore, at Step 5 of Algorithm 1, the regularization parameter of the matrix J_k is set to $\delta_k = \sigma$. The next lemma shows that the matrix J_k coincides with the Jacobian of $F(\cdot, \rho_k^+)$ at w_k when the feasibility parameter goes to zero.

Lemma 4.3. *Under Assumptions 1-4, for all $k \in \mathbb{N}$ large enough, one has*

$$J_k = F'_w(w_k, \rho_k^+) \quad and \quad \|J_k^{-1}\| \le K,$$

where K is defined by Lemma 4.2.

Proof. For the first claim, it suffices to show that $\theta_k = 0$ for k large enough. This is true whenever $K_{\rho_k^+, 0, \sigma}(w_k)$ is positive definite. For all $k \in \mathbb{N}$ we have

$$
\begin{aligned}
K_{\rho_k^+, 0, \sigma}(w_k) &= H_{\rho_k^+, 0}(x_k, y_k) + \frac{1}{\sigma} A_k A_k^\top + X_k^{-1} Z_k \\
&= \frac{1}{\sigma}(S_k - S^*) + H_{\rho_k^+, 0}(x_k, y_k - \frac{1}{\sigma} c_k) + \frac{1}{\sigma} S^* + X_k^{-1} Z_k.
\end{aligned}
$$

The first two matrices tend to zero when k tends to infinity because $\{(x_k, y_k)\}$ and $\{\rho_k^+\}$ respectively converge to $(x^*, \frac{1}{\sigma} c^*)$ and zero. It remains to show that the matrices $\frac{1}{\sigma} S^* + X_k^{-1} Z_k$ are uniformly positive definite for k large enough. Let us define the $n \times n$ diagonal matrix E, whose ith diagonal element is equal to one if $i \in \mathcal{A}$ and zero otherwise. Assumption 3 means that S^* is positive definite on the null space of E. Therefore, from Debreu's Lemma [15], there exists $\gamma > 0$ such that the matrix $\frac{1}{\sigma} S^* + \gamma E$ is positive definite. For $k \in \mathbb{N}$, let us define the diagonal matrix Ξ_k whose ith diagonal element is $[z_k]_i / [x_k]_i$ if $i \in \mathcal{A}$ and zero otherwise. Because of the strict complementarity assumption, the matrices $X_k^{-1} Z_k - \Xi_k$ tend to zero and each nonzero component of Ξ_k goes to infinity. It follows that $\Xi_k - \gamma E$ is positive semidefinite for all k large enough. By writing

$$\frac{1}{\sigma} S^* + X_k^{-1} Z_k = \frac{1}{\sigma} S^* + \gamma E + \Xi_k - \gamma E + X_k^{-1} Z_k - \Xi_k,$$

we deduce that the matrices $\frac{1}{\sigma}S^* + X_k^{-1}Z_k$ are uniformly positive definite for all k large enough.

The second assertion follows directly from the first claim and Lemma 4.2-(i). $\qquad\square$

Throughout this section, we assume that Assumption 1-4 are satisfied. The following result gives an estimate of the distance of the Newton iterate to the trajectory $\mathbf{w}$.

Lemma 4.4. *There exists $M > 0$, such that for all sufficiently large k*

$$\|w_k^+ - \mathbf{w}(\rho_k^+)\| \leq M(\|w_k - \mathbf{w}(\rho_k)\|^2 + \rho_k^2 + \|\lambda_{k+1}\|).$$

Proof. Let $k \in \mathbb{N}$ be large enough such that $\rho_k \leq \rho^*$ and $w_k \in B(w^*, r^*)$. Define $e_k := \mathbf{w}(\rho_k^+) - w_k$. From the linear system (8), then using $\Phi(w, \lambda, \rho, \sigma, \mu) = F(w, \rho) + (0, \sigma\lambda^\top, 0)^\top$ and $F(\mathbf{w}(\rho_k^+), \rho_k^+) = 0$, we have

$$
\begin{aligned}
w_k^+ - \mathbf{w}(\rho_k^+) &= -J_k^{-1}\Phi(w_k, \lambda_{k+1}, \rho_k^+, \sigma, \mu) - e_k \\
&= J_k^{-1}\left(F(\mathbf{w}(\rho_k^+), \rho_k^+) - F(w_k, \rho_k^+) - (0, \sigma\lambda_{k+1}^\top, 0)^\top - J_k e_k\right) \\
&= J_k^{-1}\int_0^1 (F_w'(w_k + te_k, \rho_k^+) - F_w'(w_k, \rho_k^+))e_k\,dt \\
&\quad + J_k^{-1}(F_w'(w_k, \rho_k^+) - J_k)e_k - J_k^{-1}(0, \sigma\lambda_{k+1}^\top, 0)^\top.
\end{aligned}
$$

By taking norm on both sides and next applying Lemma 4.2-(iv) and Lemma 4.3, we then have

$$\|w_k^+ - \mathbf{w}(\rho_k^+)\| \leq \frac{1}{2}KL\|e_k\|^2 + K\sigma\|\lambda_{k+1}\|.$$

By applying Lemma 4.2-(iii) and by using the inequality $\rho_k^+ \leq \rho_k$, we also get

$$
\begin{aligned}
\|e_k\| &\leq \|w_k - \mathbf{w}(\rho_k)\| + \|\mathbf{w}(\rho_k) - \mathbf{w}(\rho_k^+)\| \\
&\leq \|w_k - \mathbf{w}(\rho_k)\| + C\rho_k.
\end{aligned}
$$

Finally, by using the inequality $(a + b)^2 \leq 2(a^2 + b^2)$ for two real numbers a and b, we deduce that

$$\|w_k^+ - \mathbf{w}(\rho_k^+)\| \leq KL\|w_k - \mathbf{w}(\rho_k)\|^2 + KLC^2\rho_k^2 + K\sigma\|\lambda_{k+1}\|.$$

Set $M = K\max\{L, LC^2, \sigma\}$ to complete the proof. $\qquad\square$

4.2. Asymptotic result

According to Assumptions 1-4 and Lemma 4.4, the convergence to w^* of w_k implies that there exists $\bar{r} \in (0, \min\{r^*, a\})$, where a is defined in Assumption 4, such that for all k large enough, w_k, w_{k+1}, $\mathbf{w}(\rho_k)$, $\mathbf{w}(\rho_k^+)$, and $\mathbf{w}(\rho_{k+1})$ belong to $B(w^*, \bar{r})$ and $\rho_k \in (0, \rho^*)$. Without loss of generality, we can assume that these properties are true for all $k \in \mathbb{N}$.

The following lemma gives an evaluation of the distance between the next iterate w_{k+1} and the path $\mathbf{w}$.

Lemma 4.5. *There exists a constant $C_1 > 0$ such that for all $k \in \mathbb{N}$,*

$$\|w_{k+1} - \mathbf{w}(\rho_{k+1})\| \le C_1(\|\widehat{w}_k - \mathbf{w}(\rho_{k+1})\| + \|\lambda_{k+1}\|). \tag{16}$$

Proof. If there is no inner iteration, we then have $w_{k+1} = \widehat{w}_k$, therefore the inequality (16) holds trivially with $C_1 = 1$. Suppose now that w_{k+1} is obtained by applying a sequence of inner iterations. By virtue of Lemma 4.2-(v), Lemma 4.2-(ii) and Step 8 of Algorithm 1, there exists an index $k_0 \in \mathbb{N}$ such that for all $k \ge k_0$,

$$
\begin{aligned}
L_1\|w_{k+1} - \mathbf{w}(\rho_{k+1})\| &\le \|F(w_{k+1}, \rho_{k+1}) - F(\mathbf{w}(\rho_{k+1}), \rho_{k+1})\| \\
&= \|\Phi(w_{k+1}, \lambda_{k+1}, \rho_{k+1}, \sigma, \mu) - (0, \sigma\lambda_{k+1}^\top, 0)^\top\| \\
&\le \varepsilon_k + \sigma\|\lambda_{k+1}\| \\
&< \|\Phi(\widehat{w}_k, \lambda_{k+1}, \rho_{k+1}, \sigma, \mu)\| + \sigma\|\lambda_{k+1}\| \\
&\le \|F(\widehat{w}_k, \rho_{k+1}) - F(\mathbf{w}(\rho_{k+1}), \rho_{k+1})\| + 2\sigma\|\lambda_{k+1}\| \\
&\le L_2\|\widehat{w}_k - \mathbf{w}(\rho_{k+1})\| + 2\sigma\|\lambda_{k+1}\|.
\end{aligned}
$$

By defining $C_1 := \frac{1}{L_1}\max\{L_2, 2\sigma\}$, the inequality (16) is true for all $k \ge k_0$. This inequality also holds for all $k < k_0$ by increasing the constant C_1 if necessary. $\qquad\square$

The next lemma gives a lower bound on the step length computed by applying the fraction to the boundary rule (9). The proof is given in [1, Corollary 1].

Lemma 4.6. *For all $k \in \mathbb{N}$, the step length α_k computed by (9) satisfies*

$$1 - \alpha_k \le 1 - \tau_k + b\|w_k - w_k^+\|, \tag{17}$$

where $b = \dfrac{1}{a - \bar{r}} > 0$.

The first step of the asymptotic analysis is to show that the distance of the iterates to the trajectory $\mathbf{w}$ is upper bounded by a constant times the feasibility parameter.

Lemma 4.7. *The iterates w_k generated by Algorithm 1 satisfy*

$$\|w_k - \mathbf{w}(\rho_k)\| = \mathrm{O}(\rho_k)$$

Proof. First of all, let us prove that there exist constants $D_1 \in (0, 1)$, $D_2 > 0$ and $D_3 > 0$ such that for all $k \in \mathbb{N}$,

$$d_{k+1} \le d_k^2 + D_1 \frac{\rho_{k+1}}{\rho_k} d_k + D_2 \rho_{k+1}, \tag{18}$$

where $d_k := D_3\|w_k - \mathbf{w}(\rho_k)\|$. Indeed, if a sequence $\{d_k\}$ satisfies this inequality, it has been proved in [1, Lemma 7] that $d_k = \mathrm{O}(\rho_k)$ and thus the conclusion of the lemma will follow.

Let us choose $k \ge 0$. Invoking inequality (16) and Lemma 4.2-(iii), we get

$$
\begin{aligned}
C_1^{-1}\|w_{k+1} - \mathbf{w}(\rho_{k+1})\| &\le \|\widehat{w}_k - \mathbf{w}(\rho_k^+)\| + \|\mathbf{w}(\rho_k^+) - \mathbf{w}(\rho_{k+1})\| + \|\lambda_{k+1}\| \\
&\le \|\widehat{w}_k - \mathbf{w}(\rho_k^+)\| + C|\rho_k^+ - \rho_{k+1}| + \|\lambda_{k+1}\|. \tag{19}
\end{aligned}
$$

By using the definition of ρ_{k+1} at Step 7 of Algorithm 1, one has

$$
\begin{aligned}
|\rho_k^+ - \rho_{k+1}| &= (1 - \alpha_k)(\rho_k - \rho_k^+) \\
&\leq (1 - \alpha_k)\rho_k \\
&\leq (1 - \tau_k + b\|w_k - w_k^+\|)\rho_k,
\end{aligned}
\tag{20}
$$

where the last inequality is due to (17). In the same manner, we also have

$$
\widehat{w}_k - \mathbf{w}(\rho_k^+) = a_k \circ (w_k^+ - \mathbf{w}(\rho_k^+)) + (e - a_k) \circ (w_k - \mathbf{w}(\rho_k^+)).
$$

Taking the norm on both sides and noting that $\|a_k\|_\infty \leq 1$ and $\|e - a_k\|_\infty \leq 1 - \alpha_k$, we obtain

$$
\begin{aligned}
\|\widehat{w}_k - \mathbf{w}(\rho_k^+)\| &\leq \|a_k\|_\infty \|w_k^+ - \mathbf{w}(\rho_k^+)\| + \|e - a_k\|_\infty \|w_k - \mathbf{w}(\rho_k^+)\| \\
&\leq \|w_k^+ - \mathbf{w}(\rho_k^+)\| + (1 - \alpha_k)\|w_k - \mathbf{w}(\rho_k^+)\|.
\end{aligned}
$$

Applying Lemma 4.2-(iii), Lemma 4.4 and inequality (17) to the previous inequality, we deduce

$$
\begin{aligned}
\|\widehat{w}_k - \mathbf{w}(\rho_k^+)\| &\leq \|w_k^+ - \mathbf{w}(\rho_k^+)\| + (1 - \alpha_k)(\|w_k - \mathbf{w}(\rho_k)\| + C\rho_k) \\
&\leq M(\|w_k - \mathbf{w}(\rho_k)\|^2 + \rho_k^2 + \|\lambda_{k+1}\|) \\
&\quad + (1 - \tau_k + b\|w_k - w_k^+\|)(\|w_k - \mathbf{w}(\rho_k)\| + C\rho_k).
\end{aligned}
\tag{21}
$$

By using Lemma 4.2-(iii), Lemma 4.4, $\|w_k - \mathbf{w}(\rho_k)\| \leq 2\bar{r}$ and $\rho_k \leq \rho^*$, we obtain

$$
\begin{aligned}
\|w_k - w_k^+\| &\leq \|w_k - \mathbf{w}(\rho_k)\| + \|\mathbf{w}(\rho_k) - \mathbf{w}(\rho_k^+)\| + \|\mathbf{w}(\rho_k^+) - w_k^+\| \\
&\leq \|w_k - \mathbf{w}(\rho_k)\| + C\rho_k + M(\|w_k - \mathbf{w}(\rho_k)\|^2 + \rho_k^2 + \|\lambda_{k+1}\|) \\
&\leq K_1 \|w_k - \mathbf{w}(\rho_k)\| + K_2 \rho_k + M\|\lambda_{k+1}\|,
\end{aligned}
\tag{22}
$$

where $K_1 = 1 + 2M\bar{r}$ and $K_2 = C + M\rho^*$. Combining (20)-(22) into (19), using again $\|w_k - \mathbf{w}(\rho_k)\| \leq 2\bar{r}$ and also the inequality $ab \leq \frac{1}{2}(a^2 + b^2)$ for two real numbers a and b, we obtain

$$
\begin{aligned}
C_1^{-1}\|w_{k+1} - \mathbf{w}(\rho_{k+1})\| &\leq C_2 \|w_k - \mathbf{w}(\rho_k)\|^2 + (1 - \tau_k)\|w_k - \mathbf{w}(\rho_k)\| \\
&\quad + C_3 \rho_k^2 + 2C(1 - \tau_k)\rho_k + C_4\|\lambda_{k+1}\|,
\end{aligned}
$$

where $C_2 = M + bK_1 + bCK_1 + \frac{1}{2}bK_2$, $C_3 = M + 2bCK_2 + bCK_1 + \frac{1}{2}bK_2$ and $C_4 = M + 2bCM + 2bM\bar{r} + 1$. By using the properties (14) and (15), we get

$$
C_1^{-1}\|w_{k+1} - \mathbf{w}(\rho_{k+1})\| \leq C_2 \|w_k - \mathbf{w}(\rho_k)\|^2 + \mathrm{o}\left(\frac{\rho_k^+}{\rho_k}\right)\|w_k - \mathbf{w}(\rho_k)\| + \mathrm{o}(\rho_k^+) + C_4\|\lambda_{k+1}\|.
$$

Now, we use $\|\lambda_{k+1}\| = \mathrm{o}(\rho_k^+)$ given in Lemma 4.1 and $\rho_k^+ \leq \rho_{k+1}$ to get

$$
\|w_{k+1} - \mathbf{w}(\rho_{k+1})\| \leq C_1 C_2 \|w_k - \mathbf{w}(\rho_k)\|^2 + \mathrm{o}\left(\frac{\rho_{k+1}}{\rho_k}\right)\|w_k - \mathbf{w}(\rho_k)\| + \mathrm{o}(\rho_{k+1}).
$$

16

Multiplying both sides by $D_3 := C_1 C_2$, for k large, we get

$$d_{k+1} \le d_k^2 + o\left(\frac{\rho_{k+1}}{\rho_k}\right) d_k + o(\rho_{k+1}).$$

By increasing the constant $D_2 > 0$ if necessary, inequality (18) is satisfied for all $k \in \mathbb{N}$. $\qquad\square$

Lemma 4.8. *The Newton-like iterate w_k^+ generated at Step 5 of Algorithm 1 satisfies*

$$\|w_k - w_k^+\| = O(\rho_k).$$

Proof. Let $k \in \mathbb{N}$. In view of Lemmas 4.4, 4.7, 4.1 and the relation (14), we get

$$\|w_k^+ - \mathbf{w}(\rho_k^+)\| \le M(\|w_k - \mathbf{w}(\rho_k)\|^2 + \rho_k^2 + \|\lambda_{k+1}\|) = o(\rho_k^+) \qquad (23)$$

By virtue of Lemmas 4.7, 4.2-(iii), (23) and (13), we then have

$$
\begin{aligned}
\|w_k - w_k^+\| &\le &\|w_k - \mathbf{w}(\rho_k)\| + \|\mathbf{w}(\rho_k) - \mathbf{w}(\rho_k^+)\| + \|\mathbf{w}(\rho_k^+) - w_k^+\| \\
&\le& O(\rho_k) + C|\rho_k - \rho_k^+| + o(\rho_k^+) \\
&=& O(\rho_k).
\end{aligned}
$$

$\qquad\square$

The following lemma shows that iterate $\widehat{w}_k$ is asymptotically tangent to the trajectory $\mathbf{w}$.

Lemma 4.9. *The candidate iterate $\widehat{w}_k$ computed at Step 7 of Algorithm 1 satisfies*

$$\widehat{w}_k = \mathbf{w}(\rho_{k+1}) + o(\rho_{k+1}).$$

Proof. Let $k \in \mathbb{N}$. A first order Taylor expansion of $\mathbf{w}$ at $\rho = 0$ and the definition of ρ_{k+1} at Step 7 of Algorithm 1 yield

$$
\begin{aligned}
\mathbf{w}(\rho_{k+1}) &=& w^* + \mathbf{w}'(0)\rho_{k+1} + o(\rho_{k+1}) \\
&=& \alpha_k(w^* + \mathbf{w}'(0)\rho_k^+) + (1 - \alpha_k)(w^* + \mathbf{w}'(0)\rho_k) + o(\rho_{k+1}) \\
&=& \alpha_k(\mathbf{w}(\rho_k^+) + o(\rho_k^+)) + (1 - \alpha_k)(\mathbf{w}(\rho_k) + o(\rho_k)) + o(\rho_{k+1}) \\
&=& \alpha_k \mathbf{w}(\rho_k^+) + (1 - \alpha_k)\mathbf{w}(\rho_k) + o(\rho_{k+1})
\end{aligned}
$$

By using the definition of $\widehat{w}_k$ at Step 7 of Algorithm 1, we get

$$
\begin{aligned}
\widehat{w}_k - \mathbf{w}(\rho_{k+1}) &=& a_k \circ w_k^+ + (e - a_k) \circ w_k - \mathbf{w}(\rho_{k+1}) \\
&=& a_k \circ (w_k^+ - \mathbf{w}(\rho_k^+)) + (e - a_k) \circ (w_k - \mathbf{w}(\rho_k)) \\
&& + (a_k - \alpha_k e) \circ (\mathbf{w}(\rho_k^+) - \mathbf{w}(\rho_k)) + o(\rho_{k+1}).
\end{aligned}
$$

Taking the norm on both sides, using $\|a_k\|_\infty \le 1$, $\|e - a_k\|_\infty \le 1 - \alpha_k$ and $\|a_k - \alpha_k e\|_\infty \le 1 - \alpha_k$ and then invoking Lemma 4.2-(iii), Lemma 4.7 and inequality (23), we deduce

17

that

$$
\begin{aligned}
\|\widehat{w}_k - \mathbf{w}(\rho_{k+1})\| &\leq \|a_k\|_\infty \|w_k^+ - \mathbf{w}(\rho_k^+)\| + \|e - a_k\|_\infty \|w_k - \mathbf{w}(\rho_k)\| \\
&\quad + \|a_k - \alpha_k e\|_\infty \|\mathbf{w}(\rho_k^+) - \mathbf{w}(\rho_k)\| + \mathrm{o}(\rho_{k+1}) \\
&\leq \|w_k^+ - \mathbf{w}(\rho_k^+)\| + (1 - \alpha_k)\|w_k - \mathbf{w}(\rho_k)\| \\
&\quad + C(1 - \alpha_k)\rho_k + \mathrm{o}(\rho_{k+1}) \\
&= \mathrm{o}(\rho_k^+) + \mathrm{O}((1 - \alpha_k)\rho_k) + \mathrm{o}(\rho_{k+1}) \qquad (24)
\end{aligned}
$$

The bound (17) on the step length gives

$$
(1 - \alpha_k)\rho_k \leq (1 - \tau_k + b\|w_k - w_k^+\|)\rho_k.
$$

Using Lemma 4.8, properties (15) and (14), we deduce that

$$
(1 - \alpha_k)\rho_k = \mathrm{o}(\rho_k^+) + \mathrm{O}(\rho_k^2) = \mathrm{o}(\rho_k^+) \qquad (25)
$$

We conclude by substituting (25) into (24) and by using the fact that $\rho_k^+ \leq \rho_{k+1}$. $\quad\square$

We now state the main result of this section, which states that the sequence of iterates becomes asymptotically tangent to the trajectory $\mathbf{w}$.

Theorem 4.10. *Under Assumptions 1-4, if the feasibility parameter satisfies (13), if the fraction to the boundary parameter satisfies (15), if the tolerance is such that $\varepsilon_k = \Omega(\rho_{k+1})$, then we have the following results*

(i) *Algorithm 1 asymptotically need no inner iterations, i.e., $w_{k+1} = \widehat{w}_k$ for k large enough.*

(ii) *The iterates generated by Algorithm 1 satisfy*

$$
w_k = \mathbf{w}(\rho_k) + \mathrm{o}(\rho_k).
$$

(iii) *The unit step is asymptotically accepted by the fraction to the boundary rule, i.e., $\alpha_k = 1$ for k large enough.*

In particular, $\{w_k\}$ and $\{\rho_k\}$ have the same rate of convergence, i.e.,

$$
\|w_k - w^*\| = \Theta(\rho_k). \qquad (26)
$$

Proof. To prove the result (i), it suffices to show that the stopping condition in Step 8 of Algorithm 1 is satisfied for k large enough. According to Lemmas 4.2-(ii), 4.2-(v), 4.9 and 4.1 with noting that $\rho_k^+ \leq \rho_{k+1}$, we have

$$
\begin{aligned}
\|\Phi(\widehat{w}_k, \lambda_{k+1}, \rho_{k+1}, \sigma, \mu)\| &\leq \|F(\widehat{w}_k, \rho_{k+1}) - F(\mathbf{w}(\rho_{k+1}), \rho_{k+1}) - (0, \sigma\lambda_{k+1}^\top, 0)^\top\| \\
&\leq L_2 \|\widehat{w}_k - \mathbf{w}(\rho_{k+1})\| + \sigma\|\lambda_{k+1}\| \\
&= \mathrm{o}(\rho_{k+1}) \\
&< \varepsilon_k.
\end{aligned}
$$

The second result (ii) follows directly from (i) and Lemma 4.9.

The proof of the acceptation of the unit step obtained by the fraction to the boundary rule (9) is similar to the one in [5, Lemma 4.20].

18

From (ii) we have $w_k - w^* = \mathbf{w}'(0)\rho_k + \mathrm{o}(\rho_k)$. The strict complementarity assumption implies that $\mathbf{w}'(0) \neq 0$, from which we deduce (26). $\qquad\square$

5. Numerical experiments

We refer to our algorithm as SPDOPT-ID (Strongly Primal-Dual Optimization with Infeasibility Detection) which has been implemented in C. We compared it with SPDOPT [7] on two sets of problems. The *standard* set consists of 186 problems in the Hock and Schittkowski collection [20] with at least one inequality. The *infeasible* set is created from the *standard* set by adding the constraint $c_1^2 + 1 = 0$ or $(x_1 - u_1)^2 + 1 = 0$, where c_1 is the first component of c and u_1 is a bound of the first variable ($x_1 \leq u_1$ or $x_1 \geq u_1$). It is clear that all problems of the latter set is infeasible.

With a default starting point x_0 and $z_0 = (1, \ldots, 1)^\top$, y_0 is defined as the least squares solution of $g_0 + A_0 y - z_0 = 0$. The barrier parameter is initialized by $\mu_0 = 0.1$. If this parameter is updated with a trial value $\mu_k^+ < \mu_k$, we adopt the rule described in [5, Algorithm 2].

The feasibility parameter is initially set to $\rho_0 = 1$. When $\mathtt{F} = 1$, a trial value of the feasibility parameter in Step 3 is updated as follows

$$\rho_k^+ = \min\{0.2\rho_k, \rho_k^{1.4}\}.$$

This choice of ρ_k^+ satisfies the requirement (13) with $t = 0.4, \theta_1 = 1, \theta_2 = 0.2$. A lower bound of 10^{-16} is also imposed on this parameter.

For the fraction to the boundary rule, the choice $\tau_k = \max\{0.99, 1 - \rho_k\mu_k\}$ verifies condition (15). The regularization parameter δ_k is updated by the following rule: if $\mathtt{F} = 1$ and the condition (7) is not satisfied, then $\delta_k = \sigma_k$; otherwise, $\delta_k = \max\{10^{-2}\mu_k, 10^{-8}\}$. It is easy to verify that all assumptions of δ_k in the global and the local analysis are fulfilled. The parameters σ_k and θ_k are updated as in [7, Algorithm 1].

If $\|(g_k + A_k y_k/\rho_k - z_k/\rho_k, c_k, x_k \circ z_k/\rho_k)\|_\infty \leq \varepsilon_{\mathrm{tol}}$ with $\varepsilon_{\mathrm{tol}} = 10^{-8}$, the algorithm is terminated and an optimal solution is declared to be found. Otherwise, if $\|c_k\| > \varepsilon_{\mathrm{tol}}, \|\Phi(w_k, 0, 0, \sigma_k, \mu_k)\|_\infty \leq \varepsilon_{\mathrm{tol}}$ and $\rho_k \leq \varepsilon_{\mathrm{tol}}$, the algorithm is stopped at an infeasible stationary point. For SPDOPT, the stopping conditions $\|c_k\| > \varepsilon_{\mathrm{tol}}, \|A_k c_k \circ x_k\|_\infty \leq \varepsilon_{\mathrm{tol}}$ and $\sigma_k \leq \varepsilon_{\mathrm{tol}}$ are added to terminate this algorithm at an infeasible stationary point.

For the aim of getting a fast local convergence when the algorithm converges to an infeasible stationary point, the feasibility tolerance at Step 1 is set to $\epsilon = \varepsilon_{\mathrm{tol}}$. At Step 2 of Algorithm 1, we choose $c = 0.9, \ell = 2$ and $\zeta_k = 10\sigma_k\rho_k$ for all iteration k. The sequence of tolerance $\{\varepsilon_k\}$ in Step 8 is defined by the following formula

$$\varepsilon_k = 0.9 \max\{\|\Phi(w_i, \lambda_i, \rho_i, \sigma_i, \mu_i)\| : (k-4)^+ \leq i \leq k\} + 10 \min\{\alpha_k^x, \alpha_k^z\}^{0.2}\mu_{k+1}\rho_{k+1}.$$

By applying [3, Proposition 1], it is easy to see that $\{\varepsilon_k\}$ converges to zero. This choice meets the requirements to get a fast convergence in both feasible case, i.e., $\varepsilon_k = \Omega(\mu_{k+1})$, and in the infeasible case, i.e., $\varepsilon_k = \Omega(\rho_{k+1})$.

The linear solver MA57 [18] is used for all the algorithms. The maximum number of iterations, counting both the inner and the outer iterations, is limited to 3000.

For the standard problems, only 182 problems solved by at least one of two algorithms are selected for the comparison purpose (problems $\mathtt{hs099}$, $\mathtt{hs102}$, $\mathtt{hs103}$ and $\mathtt{s332}$ have not been solved). Figure 1 gives us the performance profiles of Dolan and
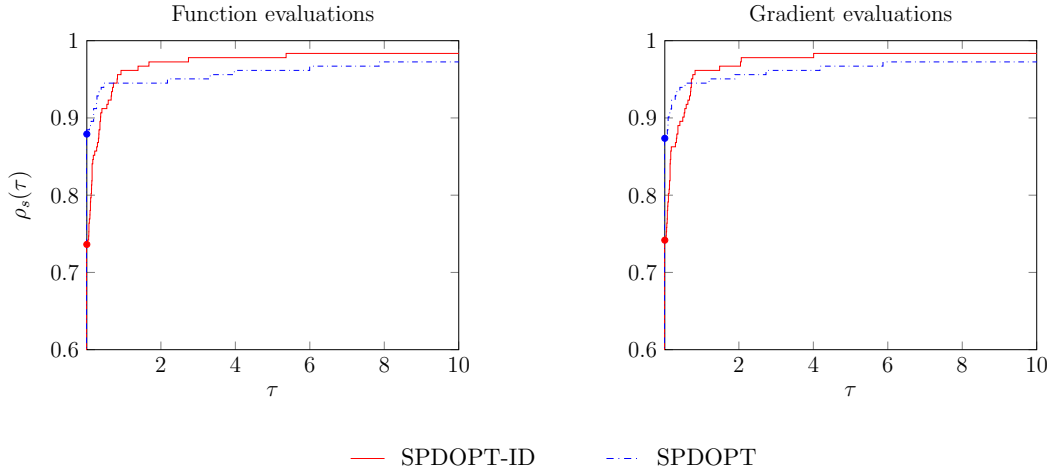
Figure 1. Performance profiles comparing the two algorithms on the set of standard problems

Moré [17] on the numbers of function and gradient evaluations. These profiles show that SPDOPT-ID and its predecessor SPDOPT have their own strengths in terms of efficiency and robustness. In particular, SPDOPT is slightly more efficient than SPDOPT-ID (14%). In term of robustness, SPDOPT-ID are succeeded in 180 problems which are more 2 problems than SPDOPT. We can conclude that the infeasibility detection does not change too much the performances of the original algorithm (SP-DOPT) for solving standard problems.

Figure 2 shows the performances of these algorithms in terms of numbers of function and gradient evaluations on a set of 174 infeasible problems (the problems `hs044, hs057, hs064, hs083, hs084, hs099, hs105, s220, s331, s332, s357, s376` have been eliminated since two algorithms cannot detect the infeasibility). We observe that SPDOPT-ID is the most efficient algorithm for detecting infeasible problems. In particular, the efficiency rate of SPDOPT-ID is over 95%. In term of robustness, SPDOPT-ID is also more robust than SPDOPT since they can detect more than 93% of problems (174 problems), while the rate of SPDOPT is only 40%.
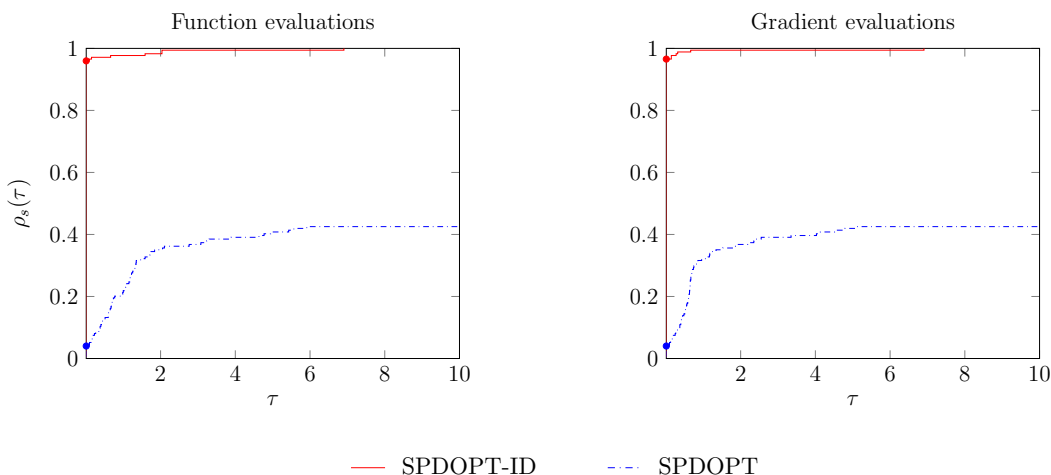
Figure 2. Performance profiles comparing the two algorithms on the set of infeasible problems

20

Figure 3 gives a general overview about the performances of these algorithms on the set of 356 problems including both standard and infeasible problems. We are able to see that SPDOPT-ID is more efficient and more robust than SPDOPT.
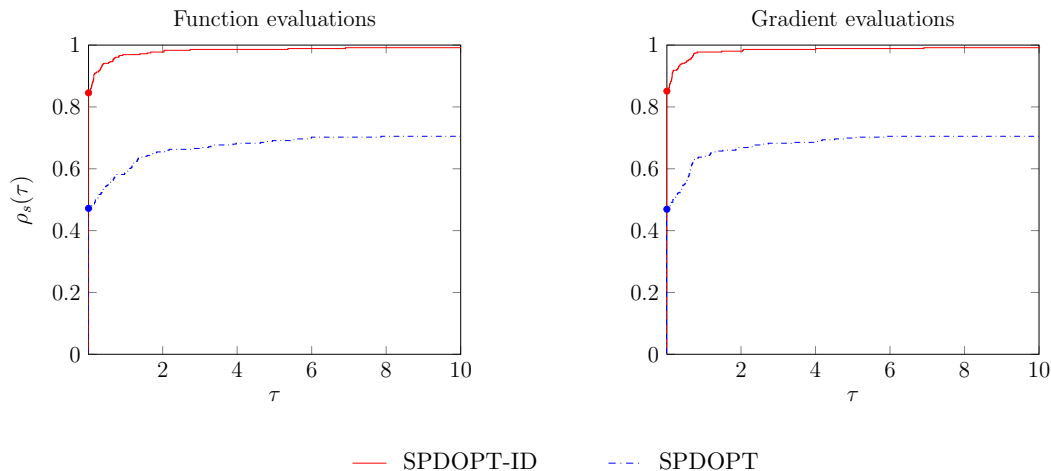
Figure 3. Comparison of the two algorithms on the set of standard and infeasible problems

6. Conclusion

Both in theory and practice, the capability of the mixed logarithmic barrier-augmented Lagrangian algorithm [7] to quickly detect infeasibility has been improved. But when the feasibility measure becomes smaller than the feasibility tolerance ($\|c^*\| \leq \epsilon$), the new algorithm exhibits the same behavior as the original one. More precisely, in that case, the quadratic penalty parameter goes to zero and the multipliers associated to the equality constraints become unbounded. An open question would be to find an interior point algorithm with a superlinear rate of convergence in any case, even if the sequence stays infeasible, but becomes nearly feasible (think about the realization of the constraint $e^x \leq 0$). In practice we can always choose a feasibility tolerance "small enough", but from a conceptual point of view this not entirely satisfactory.

This paper completes the local convergence analysis of an interior point method for nonlinear optimization in the difficult case of infeasible problems. Note especially that no assumption on the linear independence of the gradient of active constraints is used in our analysis, contrary to the analyses of Byrd et al. [12] and Burke et al. [11]. This comes from the fact that in the infeasible case, the quadratics penalty parameter provides a natural regularization of the matrix of the linear system to solve at each iteration. As a consequence, an open question is whether we can analyse the local behavior of the mixed penalty method [7] in the feasible case without any constraint qualification. We note that the local convergence analysis of interior point methods were considered under MFCQ (see, e.g., Wright and Orban [26], Vicente and Wright [24]) or the full rank of the Jacobian of equality constraints (see, e.g., Yamashita and Yabe [27]). On the other hand, when solving large scale problems, the algorithm [7] has good performances on degenerate problems, even if MFCQ is not valid.

References

[1] P. Armand and J. Benoist, *A local convergence property of primal-dual methods for non-linear programming*, Math. Program. 115 (2008), pp. 199–222.

[2] P. Armand and J. Benoist, *Uniform boundedness of the inverse of a Jacobian matrix arising in regularized interior-point methods*, Math. Program. 137 (2013), pp. 587–592.

[3] P. Armand, J. Benoist, R. Omheni, and V. Pateloup, *Study of a primal-dual algorithm for equality constrained minimization*, Comput. Optim. Appl. 59 (2014), pp. 405–433.

[4] P. Armand, J. Benoist, and D. Orban, *Dynamic updates of the barrier parameter in primal-dual methods for nonlinear programming*, Comput. Optim. Appl. 41 (2008), pp. 1–25.

[5] P. Armand, J. Benoist, and D. Orban, *From global to local convergence of interior methods for nonlinear optimization*, Optim. Methods Softw. 28 (2013), pp. 1051–1080.

[6] P. Armand and R. Omheni, *A globally and quadratically convergent primal-dual augmented Lagrangian algorithm for equality constrained optimization*, Optim. Methods Softw. 32 (2017), pp. 1–21.

[7] P. Armand and R. Omheni, *A mixed logarithmic barrier-augmented Lagrangian method for nonlinear optimization*, J. Optim. Theory Appl. 173 (2017), pp. 523–547.

[8] P. Armand and N.N. Tran, *An augmented lagrangian method for equality constrained optimization with rapid infeasibility detection capabilities*, Tech. Rep., XLIM Laboratory, 2017.

[9] E.G. Birgin, J.M. Martínez, and L.F. Prudente, *Augmented Lagrangians with possible infeasibility and finite termination for global nonlinear programming*, J. Global Optim. 58 (2014), pp. 207–242.

[10] E.G. Birgin, J.M. Martínez, and L.F. Prudente, *Optimality properties of an augmented Lagrangian method on infeasible problems*, Comput. Optim. Appl. 60 (2015), pp. 609–631.

[11] J.V. Burke, F.E. Curtis, and H. Wang, *A sequential quadratic optimization algorithm with rapid infeasibility detection*, SIAM J. Optim. 24 (2014), pp. 839–872.

[12] R.H. Byrd, F.E. Curtis, and J. Nocedal, *Infeasibility detection and SQP methods for nonlinear optimization*, SIAM J. Optim. 20 (2010), pp. 2281–2299.

[13] R.H. Byrd, J. Nocedal, and R.A. Waltz, *KNITRO: An integrated package for nonlinear optimization*, in *Large-scale nonlinear optimization*, Nonconvex Optim. Appl. Vol. 83, Springer, New York, 2006, pp. 35–59.

[14] F.E. Curtis, *A penalty-interior-point algorithm for nonlinear constrained optimization*, Math. Program. Comput. 4 (2012), pp. 181–209.

[15] G. Debreu, *Definite and semidefinite quadratic forms*, Econometrica 20 (1952), pp. 295–300.

[16] J.E. Dennis Jr. and R.B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Classics in Applied Mathematics Vol. 16, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

[17] E.D. Dolan and J.J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program. 91 (2002), pp. 201–213.

[18] I.S. Duff, *MA57—a code for the solution of sparse symmetric definite and indefinite systems*, ACM Trans. Math. Software 30 (2004), pp. 118–144.

[19] M.L.N. Gonçalves, J.G. Melo, and L.F. Prudente, *Augmented Lagrangian methods for nonlinear programming with possible infeasibility*, J. Global Optim. 63 (2015), pp. 297–318.

[20] W. Hock and K. Schittkowski, *Test examples for nonlinear programming codes*, Lecture Notes in Economics and Mathematical Systems Vol. 187, Springer-Verlag, Berlin-New York, 1981.

[21] J.M. Martínez and L.d.F. Prudente, *Handling infeasibility in a large-scale nonlinear optimization algorithm*, Numer. Algorithms 60 (2012), pp. 263–277.

[22] J. Nocedal, F. Öztoprak, and R.A. Waltz, *An interior point method for nonlinear programming with infeasibility detection capabilities*, Optim. Methods Softw. 29 (2014), pp.

837–854.

[23] R. Omheni, *Méthodes primales-duales régularisées pour l'optimisation non linéaire avec contraintes*, Ph.D. diss., Université de Limoges, École Doctorale S2IM, 2014.

[24] L.N. Vicente and S.J. Wright, *Local convergence of a primal-dual method for degenerate nonlinear programming*, Comput. Optim. Appl. 22 (2002), pp. 311–328.

[25] A. Wächter and L.T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Program. 106 (2006), pp. 25–57.

[26] S.J. Wright and D. Orban, *Properties of the log-barrier function on degenerate nonlinear programs*, Math. Oper. Res. 27 (2002), pp. 585–613.

[27] H. Yamashita and H. Yabe, *Quadratic convergence of a primal-dual interior point method for degenerate nonlinear optimization problems*, Comput. Optim. Appl. 31 (2005), pp. 123–143.