Switching stepsize strategies for PDIP

George Tzallas-Regas Berç Rustem

April 11, 2009

Abstract

In this chapter we present a primal-dual interior point algorithm for solving constrained nonlinear programming problems. *Switching rules* are implemented that aim at exploiting the merits and avoiding the drawbacks of three different merit functions. The penalty parameter is determined using an *adaptive penalty strategy* that ensures a descent property for the merit function. The descent property is assured without requiring positive definiteness of the Hessian used in the subproblem generating search direction. It is also shown that the penalty parameter does not increase indefinitely, dispensing thus with the various numerical problems occurring otherwise. Global convergence of the algorithm is achieved through the monotonic decrease of a properly chosen merit function. The algorithm is shown to possess convergent stepsizes, and therefore does not impede superlinear convergence under standard assumptions.

Keywords: Nonlinear programming, PDIP, global convergence, superlinear convergence, steplength convergence, merit functions, switching merit functions

1 Introduction

We consider the nonlinearly constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \begin{array}{rcl} \mathbf{h}(\mathbf{x}) & = & \mathbf{0} \\ \mathbf{x} & \geq & \mathbf{0} \end{array} \end{array} \tag{NLPIP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$. The *Lagrangian function* of the problem is

$$\mathcal{L}(\mathbf{w}) = f(\mathbf{x}) - \langle \mathbf{y}, \mathbf{h}(\mathbf{x}) \rangle - \langle \mathbf{z}, \mathbf{x} \rangle,$$

where $\mathbf{w}^T = (\mathbf{x}^T, \mathbf{y}^T, \mathbf{z}^T)$ and $\boldsymbol{y}, \boldsymbol{z}$ are the Lagrange multipliers for the equality and inequality constraints, respectively.

We propose a line search PDIP algorithm that is intended to converge to a minimum of problem (NLPIP) even when started far from the solution. We also intend to show that in a neighbourhood of the solution the fast convergence properties of the underlying Newton iteration will not be impeded by the line search procedure. Our method uses three line search procedures. It switches primarily between the l_∞ and the l_2 penalty function, in order to guide the iterates. If necessary, a switch is made to the squared euclidean norm of the perturbed KKT conditions merit function so as to exploit its fast local convergence. Switches are performed in a manner that complements the advantages of the merit functions and at the same time avoids their disadvantages. Unlike standard line search PDIP algorithms, we dispense with the assumption of positive definiteness of the approximations to the Hessian of the Lagrangian, for the purposes of achieving descent in the merit function.

Remark 1.1 (Notation) *The following notation will be used throughout the chapter. Algorithmic parameters will be typeset using lowercase greek letters. Vectors will be typeset using boldface lowercase characters, for example primal variables will be written as* $\mathbf{x}$. *Matrices will be typeset using uppercase boldface characters, for example* $\mathbf{A}$. *For a vector* $\mathbf{x}$, $\mathbf{X}$ *will be the diagonal matrix, whose diagonal elements are the elements of vector* $\mathbf{x}$. *Vector* $\mathbf{e}$ *denotes the vector of all ones, the length of which will be clear from the context.* $\mathbf{e}_i$ *denotes the* i*-th column of the unity matrix, and as previously, its length will depend on the context. Matrix* $\mathbf{I}_q$ *denotes the unity matrix of order* q *(*q *will vary depending). Superscript* T *will be used to denote the transpose of a matrix or vector. Superscript* (k) *denotes sequence elements and superscript* $*$ *accumulation points or the sought optimal value.* $[x]_i$ *denotes the* i*-th component of vector* $\mathbf{x}$. *The gradient of a function with respect to the primal variables will be denoted by* $\boldsymbol{\nabla}$, *and* $\boldsymbol{\nabla}^2$ *will denote the Hessian of a function with respect to the primal variables. Explicit subscripts will be used otherwise. When applied to a function of the form* $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{\nabla}\mathbf{h}(\mathbf{x})$ *denotes the* $n \times m$ *matrix, each column of which is* $\boldsymbol{\nabla}[h]_i(\mathbf{x})$. $\|\cdot\|$ *will be taken to mean the Euclidean norm of a vector or the induced operator norm and an explicit subscript will be used for different norms. We also define the following set*

$$I_a^b = \left\{ i \,\middle|\, \lim \inf_{k\to\infty} [a^{(k)}]_i = b \right\}$$

which will be useful to as. Subscript a *will be a vector (either the primal or the dual variables), the allowed values of* i *will depend on the length of* a, *and superscript* b *will be a real value (usually* 0 *or* ∞*).*

In order to simplify the presentation we shall suppress iterate superscripts in the proofs. The current estimates will appear unadorned (for example $\mathbf{x}^{(k)}$ *will appear as* $\mathbf{x}$*) and a bar will adorn next estimates (for example* $\mathbf{x}^{(k+1)}$ *will appear as* $\bar{\mathbf{x}}$*).*

We note that for demonstration of the general line search framework we chose to use $(\alpha, \mathbf{w})$ *instead of a more specific* $(\alpha_x, \mathbf{x})$ *or* $(\mathbf{A}, \mathbf{w})$*. Also any dependence of the barrier-penalty function* Φ *and of* ϕ *on either the penalty parameter* σ *or the barrier parameter* ρ *is dropped. The algorithm uses both primal and primal-dual merit functions, and these details depend on the type of the merit function. We shall present the Armijo condition in detail when each merit function is presented in Section 4.*

This chapter is organized as follows. In Section 2 we present the basic primal-dual framework for problem (NLPIP). The formulation of the algorithm is presented in Section 3. Section 4 contains a description of the merit functions employed and the switches between the merit functions. We also describe the update of the penalty parameter in the same section. The line search procedure details are presented in Section 5. The update of the barrier parameter is given in Section 6. Global convergence results are displayed in Section 8. Section 9 contains step size convergence results.

2 Primal-Dual framework

The KKT conditions of problem (NLPIP) can be written in matrix form as

$$\mathbf{F}(\mathbf{w}) = \begin{pmatrix} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{w}) \\ \mathbf{h}(\mathbf{x}) \\ \mathbf{XZe} \end{pmatrix} = \mathbf{0}, \qquad \text{(KKTIP)}$$

where, as is traditional in primal-dual interior point methods, we use $(\boldsymbol{y}, \boldsymbol{z})$ to denote the Lagrange multiplier vectors associated with equalities and inequalities, respectively and $\mathbf{w}$ to denote the triple $(\mathbf{x}, \mathbf{y}, \mathbf{z})$. $\mathbf{X}$ denotes a diagonal matrix with diagonal $\mathbf{x}$ (analogous notation is employed for other quantities) and $\mathbf{e}$ is a vector of all ones whose dimension varies with the context. In this notation the *Lagrangian function* of problem (NLPIP) can be written as

$$\mathcal{L}(\mathbf{w}) = f(\mathbf{x}) + \langle \mathbf{y}, \mathbf{h}(\mathbf{x}) \rangle - \langle \mathbf{z}, \mathbf{x} \rangle.$$

If we use the *logarithmic barrier function*

$$B(\mathbf{x}) = -\sum_{i=1}^{p} \log\left([g]_i(\mathbf{x})\right), \qquad (1)$$

problem (NLPIP) can be written equivalently as

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x};\rho) \triangleq f(\mathbf{x}) - \rho B(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{aligned} \tag{BNLP}$$

where $\mathbf{x} > 0$. The KKT conditions of this problem are

$$\nabla f(\mathbf{x}) - \nabla \mathbf{h}(\mathbf{x})\mathbf{y} - \rho \mathbf{X}^{-1}\mathbf{e} = \mathbf{0} \tag{2a}$$

$$\mathbf{h}(\mathbf{x}) = \mathbf{0}. \tag{2b}$$

Introducing $\mathbf{z} = \rho \mathbf{X}^{-1}\mathbf{e}$, (2) are written as

$$\nabla f(\mathbf{x}) - \nabla \mathbf{h}(\mathbf{x})\mathbf{y} - \mathbf{z} = \mathbf{0} \tag{3a}$$

$$\mathbf{h}(\mathbf{x}) = \mathbf{0} \tag{3b}$$

$$\mathbf{z} = \rho \mathbf{X}^{-1}\mathbf{e} \tag{3c}$$

or using the Lagrangian function, we can write in matrix form

$$\mathbf{F}(\mathbf{w};\rho) = \begin{pmatrix} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{w}) \\ \mathbf{h}(\mathbf{x}) \\ \mathbf{XZe} - \rho \mathbf{e} \end{pmatrix} = \mathbf{0}, \tag{PRTKKT}$$

which are called the *perturbed KKT conditions*. It is obvious from Eqs. (PRTKKT), (KKTIP) that the perturbed KKT conditions differ from the KKT conditions of the original problem only in the complementarity conditions.

As $\rho \to 0$, the perturbed KKT conditions approximate the original KKT conditions more and more accurately and $\mathbf{w}(\rho)$ converges to the solution of the KKT conditions along the central path.

The primal-dual interior point framework involves inner and outer iterations. Inner iterations are dominated by the generation of a search direction $\delta \mathbf{w}^{(k)}$. This is done solving the perturbed KKT conditions (PRTLNS)

$$\begin{pmatrix} \nabla^2_{\mathbf{x}} \mathcal{L}(\mathbf{w}^{(k)}) & -\nabla \mathbf{h}(\mathbf{x}^{(k)}) & -\mathbf{I} \\ \nabla \mathbf{h}(\mathbf{x}^{(k)})^T & \mathbf{0} & \mathbf{0} \\ \mathbf{Z}^{(k)} & \mathbf{0} & \mathbf{X}^{(k)} \end{pmatrix} \delta \mathbf{w}^{(k)} = - \begin{pmatrix} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{w}) \\ \mathbf{h}(\mathbf{x}) \\ \mathbf{XZe} - \rho \mathbf{e} \end{pmatrix}. \tag{PRTLNS}$$

The next iterate is generated using

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \mathbf{A}^{(k)} \delta \mathbf{w}^{(k)},$$

where $\mathbf{A}^{(k)} = \text{diag}(\alpha_x^{(k)}\mathbf{I}_n, \alpha_y^{(k)}\mathbf{I}_m, \alpha_z^{(k)}\mathbf{I}_p)$ is the line search parameter. The step-lengths $\alpha_x, \alpha_y, \alpha_z \in (0,1]$ can be equal or different to each other. El-Bakry et al. [6] discuss this choice. The value of α_x is decided by a line search that uses an *Armijo condition* of the form

$$\Phi(\mathbf{x}^{(k)} + \alpha\delta\mathbf{x}^{(k)}) - \Phi(\mathbf{x}^{(k)}) \leq -\alpha c\phi(\delta\mathbf{x}^{(k)}),$$

where Φ is an appropriately chosen merit function, ϕ is a term that quantifies the decrease in the merit function and $c \in (0, \tilde{c})$ is a predetermined constant. The value of $\tilde{c}$ depends on the choice of ϕ.

In addition, the values of α_x, α_z are decided so as to ensure positiveness of $(\mathbf{x}^{(k+1)}, \mathbf{z}^{(k+1)})$ for all k. Line search strategies for primal-dual interior point methods are also discussed in [4, 9, 30, 2].

When an approximation to a point that satisfies the perturbed KKT conditions is found, the value of the barrier parameter is reduced to a strictly smaller value, and the calculation is repeated until ρ becomes zero.

The first order change of the perturbed KKT conditions (PRTLNS) can be written analytically as

$$\begin{aligned}
\mathbf{B}\delta\mathbf{x} - \nabla\mathbf{h}(\mathbf{x})\delta\mathbf{y} - \delta\mathbf{z} &= -\nabla f(\mathbf{x}) + \nabla\mathbf{h}(\mathbf{x})\mathbf{y} + \mathbf{z} \\
\nabla\mathbf{h}(\mathbf{x})^T\delta\mathbf{x} &= -\mathbf{h}(\mathbf{x}) \\
\mathbf{Z}\delta\mathbf{x} + \mathbf{X}\delta\mathbf{z} &= -\mathbf{X}\mathbf{Z}\mathbf{e} - \rho\mathbf{e}.
\end{aligned}$$

Solving the last of these for $\delta\mathbf{z}$ and substituting in the first one we obtain

$$\begin{aligned}
(\mathbf{B} + \mathbf{X}^{-1}\mathbf{Z})\delta\mathbf{x} - \nabla\mathbf{h}(\mathbf{x})\delta\mathbf{y} &= -\nabla f(\mathbf{x}) + \nabla\mathbf{h}(\mathbf{x})\mathbf{y} + \rho\mathbf{X}^{-1}\mathbf{e} && \text{(4a)} \\
\nabla\mathbf{h}(\mathbf{x})^T\delta\mathbf{x} &= -\mathbf{h}(\mathbf{x}) && \text{(4b)} \\
\delta\mathbf{z} &= -\mathbf{X}^{-1}\mathbf{Z}\delta\mathbf{x} - \mathbf{z} + \rho\mathbf{X}^{-1}\mathbf{e}. && \text{(4c)}
\end{aligned}$$

If we set $\mathbf{N} = \mathbf{B} + \mathbf{X}^{-1}\mathbf{Z}$ and $\nabla f(\mathbf{x};\rho) = \nabla f(\mathbf{x}) - \rho\mathbf{X}^{-1}\mathbf{e}$, then the following *reduced KKT conditions* can be obtained

$$\begin{aligned}
\mathbf{N}\delta\mathbf{x} - \nabla\mathbf{h}(\mathbf{x})\hat{\mathbf{y}} &= -\nabla f(\mathbf{x};\rho) && \text{(5a)} \\
\nabla\mathbf{h}(\mathbf{x})^T\delta\mathbf{x} &= -\mathbf{h}(\mathbf{x}), && \text{(5b)}
\end{aligned}$$

where $\hat{\mathbf{y}} = \mathbf{y} + \delta\mathbf{y}$. Using an *orthogonal decomposition* we can write the primal search direction as

$$\delta\mathbf{x}^{(k)} = \mathbf{u}^{(k)} + \mathbf{v}^{(k)}, \tag{6}$$

where $\mathbf{v}^{(k)}$ is in the column space of the constraint gradients $\boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^{(k)})$ and $\mathbf{u}^{(k)}$ in the null space of $\boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^{(k)})^T$.

The following assumptions are invoked throughout the paper

(A1) Functions f and $\mathbf{h}$ are twice continuously differentiable.

(A2) There exists constant $\beta_1 > 0$ such that

$$\|\mathbf{B}^{(k)}\| \leq \beta_1, \quad \forall k.$$

(A3) There exist constant $\beta_2 > 0$ such that at each iteration

$$\langle \mathbf{u}, \mathbf{B}^{(k)} \rangle \mathbf{u} \geq \beta_2 \|\mathbf{u}\|,$$

for all $\mathbf{u} \neq \mathbf{0}$ that satisfy $\boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^{(k)})^T \mathbf{u} = \mathbf{0}$.

(A4) The constraint normals of the equality constraints are Lipschitz continuous,

$$\boldsymbol{\nabla}[h]_i(\mathbf{x}) \in Lip_\gamma(D), \quad i = 1, \ldots, m.$$

(A5) The columns of matrix $[\boldsymbol{\nabla}\mathbf{h}(\mathbf{x}), \mathbf{e}_i, i \in I_x^0]$ are linearly independent.

(A6) If $[\mathbf{z}^*]_i > 0$ then $[\mathbf{x}^*]_i = 0$ for all $i = 1, \ldots, n$ and vice versa.

(A7) For all $\mathbf{u} \neq \mathbf{0}$ that satisfy $\boldsymbol{\nabla}[h]_i(\mathbf{x})^T \mathbf{u} = 0$, $i = 1, \ldots, m$ and $\mathbf{e}_i{}^T \mathbf{u} = 0$, $i \in I_x^0$ there holds

$$\langle \mathbf{u}, \boldsymbol{\nabla}^2_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \rangle \mathbf{u} > 0.$$

Assumptions that pertain to $\mathbf{B}^{(k)}$, will also be used for $\mathbf{N}^{(k)}$, since their difference is $\mathbf{X}^{-1^{(k)}}\mathbf{Z}^{(k)}$ which is positive definite for $\mathbf{x}^{(k)}, \mathbf{z}^{(k)} > 0$.

3 Algorithm formulation

The algorithm involves outer and inner iterations. Outer iterations check if the perturbed KKT conditions are approximately satisfied, and if not the barrier parameter is decreased, with the intention to make it approach zero.

Inner iterations solve the Newton system (PRTLNS), in order to generate a search direction. Our algorithm contains switches that activate three different line search procedures. The choice of the line search procedure is indicated by $t_1 \in \{1, 2, 3\}$. The penalty parameter is updated accordingly (increased), so as to achieve sufficient decrease in the merit function used in the line search procedure. The stepsize parameter is subsequently calculated, and the new iterates are generated. The inner iterations are repeated, until a convergence criterion is satisfied. In that case, an outer iteration takes place, and the procedure is repeated.

Algorithm 1 gives the outline of the algorithm. We note that due to space limitations the details of the update of the penalty parameter, the barrier parameter and the line search procedure are not presented here, but are listed separately in the mentioned sections.

Algorithm 1 Merit function switching PDIP algorithm

Choose $\epsilon_h, \epsilon_d, \epsilon_0 \in (0, 1)$
Choose $\delta, q \in [0, +\infty)$, $\zeta, \theta, \eta, \in (0, 1)$, $c_1 \in (0, 1), c_2, c_3 \in (0, \frac{1}{2})$, $m, M, M_0 > 0$
Choose $\mathbf{x}^{(0)} \in \mathbb{R}^n$, $\mathbf{z}^{(0)} \in \mathbb{R}^p$, $\mathbf{y}^{(0)} \in \mathbb{R}^m$, $\mathbf{B}^{(0)} \in \mathbb{R}^{n \times n}$
Choose $\sigma^{(0)} \in (0, +\infty)$, $\rho^{(0)} \in (0, +\infty)$
Set $l := 0$
repeat
    Set $k := 0$
    repeat
        Solve the Newton system (PRTLNS) to obtain $\delta\mathbf{w}^{(k)} = (\delta\mathbf{x}^{(k)}, \delta\mathbf{y}^{(k)}, \delta\mathbf{z}^{(k)})$
        Choose $\sigma^{(k+1)}$ and t_1 as described in Algorithm 2 (Section 4)
        Choose $\boldsymbol{A}^{(k)}$ as described in Algorithm 3 (Section 5)
        Set $\mathbf{w}^{(k+1)} := \mathbf{w}^{(k)} + \boldsymbol{A}^{(k)}\delta\mathbf{w}^{(k)}$
        Update $\mathbf{B}^{(k)}$ to $\mathbf{B}^{(k+1)}$
        Set $k := k + 1$
    until Inner-convergence
    Choose $\rho^{(l+1)}$ as described in Algorithm 4 (Section 6)
    Set $l := l + 1$
until Outer-convergence

During the inner iterations the barrier parameter is constant. Therefore we drop the dependence of ρ on l in Sections 4, 5 and write $\rho^{(l)}$ as ρ.

4 Merit functions and penalty parameter

In Sections 4.1–4.3 we introduce the merit functions we are going to use in the algorithm. In Section 4.4 we give the details for activating a merit function and the update of the penalty parameter.

4.1 The l_∞ merit function

For this line search procedure we employ the l_∞ *merit function*

$$\Phi_1(\mathbf{x};\sigma,\rho) = f(\mathbf{x};\rho) + \sigma P(\mathbf{x}),$$

where

$$P_1(\mathbf{x}) = \max\left\{0, \left|[h]_1(\mathbf{x})\right|, \ldots, \left|[h]_m(\mathbf{x})\right|\right\}$$

and $\sigma \geq 0$ is a *penalty parameter.* The right hand side of the Armijo condition is

$$\phi_1(\mathbf{x},\hat{\mathbf{y}};\sigma) = \sigma P(\mathbf{x}) - \hat{\mathbf{y}}^T\mathbf{h}(\mathbf{x}).$$

The *Armijo condition* is

$$\Phi(\mathbf{x}^{(k+1)};\sigma^{(k+1)},\rho) - \Phi(\mathbf{x}^{(k)};\sigma^{(k+1)},\rho) \leq -\alpha_x c_1 \phi(\mathbf{x}^{(k)},\hat{\mathbf{y}}^{(k)};\sigma^{(k+1)}), \tag{7}$$

where $c_1 \in (0,1)$. This line search procedure corresponds to $t_1 = 1$ in Algorithm 2.

This merit function was introduced for a line search SQP framework by Pshenichnyi [25], Pshenichnyi and Danilin [26] and was also used by Polak and Mayne [21], Mayne and Polak [16]. We note that the right hand-side of the Armijo condition is different to the one proposed in the aforementioned papers.

4.2 The l_2 merit function

This merit function is the l_2 *merit function* used by Akrotirianakis and Rustem [2]. It corresponds to $t_1 = 2$ in Algorithm 2. In detail

$$\Phi_2(\mathbf{x};\sigma,\rho) = f(\mathbf{x};\rho) + \sigma P_2(\mathbf{x}),$$

where the penalty term is given by

$$P_2(\mathbf{x}) = \|\mathbf{h}(\mathbf{x})\|^2.$$

The right hand side of the Armijo line search procedure is the directional derivative of the merit function (with respect to $\mathbf{x}$), *i.e.*

$$\phi_2(\mathbf{x},\delta\mathbf{x};\sigma,\rho) = -\langle \delta\mathbf{x}, \boldsymbol{\nabla}\Phi(\mathbf{x};\sigma,\rho)\rangle.$$

The *Armijo condition* for this merit function is

$$\Phi_2(\mathbf{x}^{(k+1)};\sigma^{(k+1)},\rho) - \Phi_2(\mathbf{x}^{(k)};\sigma^{(k+1)},\rho) \leq -\alpha_x c_2 \phi_2(\mathbf{x}^{(k)},\delta\mathbf{x}^{(k)};\sigma^{(k+1)},\rho), \tag{8}$$

where $c_2 \in (0, \frac{1}{2})$. The gradient of the merit function is

$$\boldsymbol{\nabla}\Phi_2(\mathbf{x};\sigma,\rho) = \boldsymbol{\nabla} f(\mathbf{x};\rho) + \sigma \boldsymbol{\nabla}\mathbf{h}(\mathbf{x})\mathbf{h}(\mathbf{x}). \tag{9}$$

By considering Eqs. (4b), (9) we can write

$$\phi_2(\mathbf{x}^{(k)}, \delta\mathbf{x}^{(k)}; \sigma^{(k+1)}, \rho) = -\langle \delta\mathbf{x}^{(k)}, \boldsymbol{\nabla} f(\mathbf{x}^{(k)};\rho)\rangle + \sigma^{(k+1)} \|\mathbf{h}(\mathbf{x}^{(k)})\|^2. \tag{10}$$

4.3 The perturbed KKT residual

This merit function is the squared Euclidean of the perturbed KKT conditions (PRTKKT), *i.e.* for $t_1 = 3$ in Algorithm 2 we use

$$\Phi_3(\mathbf{w};\rho) = \frac{1}{2}\|\mathbf{F}(\mathbf{w};\rho)\|^2,$$

where $\mathbf{F}(\mathbf{w};\rho)$ is given by (3). The right hand side of the Armijo condition is the negative of the directional derivative of this merit function along the primal search direction. It can be shown using (PRTLNS) that

$$\phi_3(\mathbf{w}^{(k)};\rho) = -\langle \delta\mathbf{w}^{(k)}, \boldsymbol{\nabla}\Phi_3(\mathbf{w}^{(k)};\rho)\rangle = \|\mathbf{F}(\mathbf{w}^{(k)};\rho)\|^2.$$

The *Armijo condition* is

$$\Phi_3(\mathbf{w}^{(k)} + \alpha\delta\mathbf{w}^{(k)};\rho) - \Phi_3(\mathbf{w}^{(k)};\rho) \le -\alpha c_3 \phi_3(\mathbf{w}^{(k)};\rho), \tag{11}$$

where $c_3 \in (0, \frac{1}{2})$. The local and global convergence properties of this line search procedure have been analyzed thoroughly by El-Bakry et al. [6], and we shall not analyze it further in our convergence proofs.

4.4 Merit function switches

The choice of the merit functions is motivated by the following remarks:

- The squared norm of the KKT conditions (Φ_3) is forceful in dispatching the sequence $\{\mathbf{w}^{(k)}\}$ to its optimal solution when used in a close neighbourhood of the latter. However it may, on occasion, allow the algorithm to converge to a saddle point or a maximum. It can be shown to converge to unity stepsizes and consequently to allow superlinear convergence.
- Φ_2 converges to a minimum and stepsizes converge to unity, but the penalty parameter may go to infinity.

t_1	$\tau^{(k)} + \langle \delta\mathbf{x}^{(k)}, \boldsymbol{\nabla} f(\mathbf{x}^{(k)};\rho)\rangle > 0$	$\|\delta\mathbf{x}^{(k)}\| > \epsilon_d$	$\|\mathbf{h}(\mathbf{x}^{(k)})\|^2 \leq \epsilon_h$
1	T	T	T
1	T	T	F
3	T	F	T
2	T	F	F
2	F	T	T
2	F	T	F
2	F	F	T
2	F	F	F

Table 1: PDIP Switch motivation

- Φ_1 also converges to a minimum and is less problematic unless the multipliers tend to infinity, but it cannot be shown to attain unity stepsizes.

We use several switches to exploit the advantages and avoid the disadvantages of these merit functions. Table (1) shows the truth-values of the three conditions on which our switches depend on. Our motivation is to activate a line search procedure that will

1. guarantee descent of the activated merit function
2. allow admission of unity stepsizes in a neighbourhood of the solution
3. not result in large values of the penalty parameter.

In Table (1) the first column, labelled t_1, shows which line search procedure, if activated, guarantees these three targets. The other three columns are conditions based on which these targets are satisfied. These conditions have been identified as necessary for achieving our goals. In Table (1)

$$\tau^{(k)} = r^{(k)}\|\delta\mathbf{x}^{(k)}\|^2, \quad r^{(k)} > 0. \tag{12}$$

Based on these remarks, we formulate Algorithm 2. The algorithm implements switches that combine three line search procedures so as to complement their merits and avoid their drawbacks. Depending on the generated direction, and the feasibility of the current point, the algorithm will activate a line search procedure by choosing an appropriate merit function. The choice of the merit function is signified using switch $t_1 \in \{1, 2, 3\}$. The update of the penalty parameter has two targets. We seek on the one hand to guarantee that the search direction is a descent one for the activated merit function and on the other hand to avoid updating the penalty parameter indefinitely.

For the computation of $\tau^{(k)}$ from (12) we choose $r^{(k)}$ chosen to lie in

$$r^{(k)} \in (0, \tilde{r}^{(k)}] \tag{13}$$

where

$$\tilde{r}^{(k)} = \frac{1}{2}\frac{\langle \mathbf{u}^{(k)}, \mathbf{N}^{(k)} \rangle \mathbf{u}^{(k)}}{\|\mathbf{u}^{(k)}\|^2} \cdot \frac{\|\delta\mathbf{x}^{(k)}\|^2}{\|\mathbf{u}^{(k)}\|^2}. \tag{14}$$

Remark 4.1 *We note that $\tilde{r}^{(k)} > 0$ under Assumption (A3), as long as $\delta\mathbf{x}^{(k)} \neq \mathbf{0}$.*

Remark 4.2 *The introduction of $r^{(k)}$ is used in order to achieve descent in the activated merit function without assuming positive definiteness of the Hessian of the Lagrangian.*

For the computation of ϑ at iteration k in Step 1 we use

$$\vartheta = \begin{cases} 0, & \text{if } \|\mathbf{h}(\mathbf{x})\| \leq \sqrt{\epsilon_h} \\ 0, & \text{if } \|\mathbf{h}(\mathbf{x})\| > \sqrt{\epsilon_h} \text{ and } \langle \delta\mathbf{x}, \mathbf{N} \rangle \delta\mathbf{x} \geq 0 \\ \frac{\tau - \langle \delta\mathbf{x}, \mathbf{N} \rangle \delta\mathbf{x}}{\sqrt{\epsilon_h} P_1(\mathbf{x})}, & \text{if } \|\mathbf{h}(\mathbf{x})\| > \sqrt{\epsilon_h} \text{ and } \langle \delta\mathbf{x}, \mathbf{N} \rangle \delta\mathbf{x} < 0. \end{cases} \tag{15}$$

Remark 4.3 *From Eq. (15) we can see that $\vartheta^{(k)} \geq 0$. The addition of $\vartheta^{(k)}$ in the update of the penalty parameter guarantees that*

$$\langle \delta\mathbf{x}^{(k)}, \boldsymbol{\nabla}\Phi_1(\mathbf{x}^{(k)}; \sigma^{(k+1)}, \rho) \rangle \leq -\tau^{(k)}$$

instead of the traditional

$$\langle \delta\mathbf{x}^{(k)}, \boldsymbol{\nabla}\Phi_1(\mathbf{x}^{(k)}; \sigma^{(k+1)}, \rho) \rangle \leq -\langle \delta\mathbf{x}^{(k)}, \mathbf{N}^{(k)} \rangle \delta\mathbf{x}^{(k)}$$

which is associated with positive definite Hessian approximations $\mathbf{N}^{(k)}$. The latter case corresponds to $\vartheta^{(k)} = 0$, which occurs if $\langle \delta\mathbf{x}^{(k)}, \mathbf{N}^{(k)} \rangle \delta\mathbf{x}^{(k)} \geq 0$, or if we are feasible, which by Assumption (A3) means that $\langle \delta\mathbf{x}^{(k)}, \mathbf{N}^{(k)} \rangle \delta\mathbf{x}^{(k)} \geq 0$.

Remark 4.4 *The term $\sqrt{\epsilon_h}$ in the denominator of $\vartheta^{(k)}$ will be useful in the proof of Lemma (9).*

Finally for the computation of β we set

$$\beta = \max\left\{1, \frac{\sqrt{p}}{\sqrt{\epsilon_h}}\right\}. \tag{16}$$

Algorithm 2 Merit function switch and penalty adaptive update (PDIP)

Compute $r^{(k)}$ from (14), $\tau^{(k)}$ from (12), $\vartheta^{(k)}$ from (15), $\beta^{(k)}$ from (16)
if $(\tau^{(k)} + \langle \delta\mathbf{x}^{(k)}, \boldsymbol{\nabla} f(\mathbf{x}^{(k)}; \rho)\rangle > 0)$ **then**
  if $(\|\delta\mathbf{x}^{(k)}\| > \epsilon_d)$ **then**
    $t_1 := 1$
    $\sigma^{(k+1)} := \max\left\{\beta\|\hat{\mathbf{y}}^{(k)}\|_1 + \vartheta^{(k)} + \delta, \sigma^{(k)}\right\}$
  else
    if $(\|\mathbf{h}(\mathbf{x}^{(k)})\|^2 \leq \epsilon_h)$ **then**
      $t_1 := 3$
      $\sigma^{(k+1)} := \sigma^{(k)}$
    else
      $t_1 := 2$
      if $(\tau^{(k)} + \langle \delta\mathbf{x}^{(k)}, \boldsymbol{\nabla} f(\mathbf{x}^{(k)}; \rho)\rangle \leq \sigma^{(k)}\|\mathbf{h}(\mathbf{x}^{(k)})\|^2)$ **then**
        $\sigma^{(k+1)} := \sigma^{(k)}$
      else
        $\sigma^{(k+1)} := \max\left\{\dfrac{\tau^{(k)} + \langle \delta\mathbf{x}^{(k)}, \boldsymbol{\nabla} f(\mathbf{x}^{(k)}; \rho)\rangle}{\|\mathbf{h}(\mathbf{x}^{(k)})\|^2}, \sigma^{(k)} + \delta\right\}$
      end if
    end if
  end if
else
  $t_1 := 2$
  $\sigma^{(k+1)} := \sigma^{(k)}$
end if

5 Line search procedure

In order to obtain a globally convergent algorithm, the Newton iteration generating the search direction has to be stabilized. The new iterates are updated using the formula

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \mathbf{A}^{(k)}\delta\mathbf{w}^{(k)} \tag{17}$$

where $\mathbf{A}^{(k)} = (\alpha_x^{(k)}\mathbf{I}_n, \alpha_y^{(k)}\mathbf{I}_m, \alpha_z^{(k)}\mathbf{I}_p)$. The main goal is to decrease the value of the merit function, and in our algorithm we use variants of Armijo's rule to achieve this.

Our algorithm uses three different merit functions.

- Φ_3, introduced in Section 4.3, is a primal-dual merit function. For this merit function we use a common step length for both the primal and the dual variables ($\alpha_x = \alpha_y = \alpha_z$). This merit function was used in a line search procedure by El-Bakry et al. [6]. Steps 4–11 of Algorithm 3 give details of the procedure that corresponds to this merit function. We will not analyze this strategy. The interested reader may consult [6] for more details.
- Φ_1, Φ_2 are *barrier-penalty functions* of the primal category. For these merit functions Armijo's rule is used to determine the primal step size $\alpha_x^{(k)}$, and box constraints for the dual step size $\alpha_z^{(k)}$ that corresponds to the $\mathbf{z}$ variables. The step size $\alpha_y^{(k)}$ for the $\mathbf{y}$ variables can be set either to unity, or to be equal to $\alpha_z^{(k)}$. Steps 13–25 give details of the procedure that corresponds to these merit functions.

The step acceptance criterion $SA(\alpha, \mathbf{w}^{(k)}, \delta\mathbf{w}^{(k)}, t_1)$ is discussed in Section 5.1. In what follows we present the line search procedure that corresponds to the primal merit functions (Φ_1 and Φ_2).

The line search procedure starts by computing the maximum allowable step size towards the feasible region of the primal variables using the formula

$$\bar{\alpha}_x^{(k)} = \min_{i=1,\ldots,n} \left\{ \frac{-[\mathbf{x}^{(k)}]_i}{[\delta\mathbf{x}^{(k)}]_i} : [\delta\mathbf{x}^{(k)}]_i < 0 \right\}.$$

The barrier term of the logarithmic barrier function is well behaved if all primal iterates are strictly positive. In order to obtain $\mathbf{x}^{(k+1)} > 0$ from (17), we confine the primal step size within $[0, \bar{\alpha}_x^{(k)})$. This is achieved using

$$\hat{\alpha}_x^{(k)} = \min\left\{1, \zeta\bar{\alpha}_x^{(k)}\right\}$$

Algorithm 3 Line search procedure (PDIP)

$\bar{\alpha}_x^{(k)} := \min_{i=1,\ldots,n} \left\{ \frac{-[\mathbf{x}^{(k)}]_i}{[\delta \mathbf{x}^{(k)}]_i} : [\delta \mathbf{x}^{(k)}]_i < 0 \right\}$

$j^{(k)} := 0$

if ($t_1 == 3$) **then**

  $\bar{\alpha}_z^{(k)} := \min_{i=1,\ldots,p} \left\{ \frac{-[\mathbf{z}^{(k)}]_i}{[\delta \mathbf{z}^{(k)}]_i} : [\delta \mathbf{z}^{(k)}]_i < 0 \right\}$

  $\hat{\alpha}^{(k)} := \min \left\{ 1, \zeta \bar{\alpha}_x^{(k)}, \zeta \bar{\alpha}_z^{(k)} \right\}$

  $\alpha^{(k)} := \theta^{j^{(k)}} \hat{\alpha}^{(k)}$

  while $SA(\alpha^{(k)}, \mathbf{w}^{(k)}, \delta \mathbf{w}^{(k)}, t_1)$ **do**

    $j^{(k)} := j^{(k)} + 1$

    $\alpha^{(k)} := \theta^{j^{(k)}} \hat{\alpha}^{(k)}$

  end while

  $\boldsymbol{\alpha}^{(k)} := (\alpha^{(k)} \mathbf{e}, \alpha^{(k)} \mathbf{e}, \alpha^{(k)} \mathbf{e})$

else

  $\hat{\alpha}_x^{(k)} := \min \left\{ 1, \zeta \bar{\alpha}_x^{(k)} \right\}$

  $\alpha_x^{(k)} := \theta^{j^{(k)}} \hat{\alpha}_x^{(k)}$

  while $SA(\alpha_x^{(k)}, \mathbf{x}^{(k)}, \delta \mathbf{x}^{(k)}, t_1)$ **do**

    $j^{(k)} := j^{(k)} + 1$

    $\alpha_x^{(k)} := \theta^{j^{(k)}} \hat{\alpha}_x^{(k)}$

  end while

  $[LB^{(k)}]_i := \min \left\{ \frac{1}{2} m\rho, [\mathbf{x}^{(k+1)}]_i [\mathbf{z}^{(k)}]_i \right\}$

  $[UB^{(k)}]_i := \min \left\{ 2M\rho, [\mathbf{x}^{(k+1)}]_i [\mathbf{z}^{(k)}]_i \right\}$

  for $(i = 1 \ldots n)$ **do**

    $[\alpha_z^{(k)}]_i := \max_{[\alpha]_i} \left\{ [LB^{(k)}]_i \leq [\mathbf{x}^{(k+1)}]_i [\mathbf{z}^{(k)} + \alpha \delta \mathbf{z}^{(k)}]_i \leq [UB^{(k)}]_i \right\}$

  end for

  $\alpha_z^{(k)} := \min \left\{ 1, \min_{i=1,\ldots,p} \left\{ [\alpha_z^{(k)}]_i \right\} \right\}$

  $\boldsymbol{\alpha}^{(k)} := (\alpha_x^{(k)} \mathbf{e}, \alpha_y^{(k)} \mathbf{e}, \alpha_z^{(k)} \mathbf{e})$

end if

$\mathbf{A}^{(k)} := \mathrm{diag}(\boldsymbol{\alpha}^{(k)})$

as the initial step size, where $\zeta \in (0,1)$ is a predetermined constant. This rule also guarantees that the unity step size is never exceeded. The step size is given by

$$\alpha_x^{(k)} = \theta^{j^{(k)}} \hat{\alpha}_x^{(k)},$$

where $\theta \in (0,1)$ and $j^{(k)}$ is the smallest nonnegative integer such that

$$SA(\alpha_x^{(k)}, \mathbf{x}^{(k)}, \delta\mathbf{x}^{(k)}, t_1)$$

is satisfied.

The rule for the dual variables uses information provided by the primal variables $\mathbf{x}^{(k+1)}$. This rule was used by Akrotirianakis and Rustem [2]. A step $[\alpha_z^{(k)}]_i$ is calculated along the dual direction $[\delta\mathbf{z}^{(k)}]_i$ so that the following box constraints are satisfied

$$[\alpha_z^{(k)}]_i = \max_{\alpha>0} \left\{ [LB^{(k)}]_i \leq [\mathbf{x}^{(k+1)}]_i [\mathbf{z}^{(k)} + \alpha\delta\mathbf{z}^{(k)}]_i \leq [UB^{(k)}]_i \right\}. \tag{18}$$

The box lower and upper bounds are defined as

$$[LB^{(k)}]_i = \min \left\{ \frac{1}{2} m\rho, [\mathbf{x}^{(k+1)}]_i [\mathbf{z}^{(k)}]_i \right\} \tag{19a}$$

$$[UB^{(k)}]_i = \max \left\{ 2M\rho, [\mathbf{x}^{(k+1)}]_i [\mathbf{z}^{(k)}]_i \right\} \tag{19b}$$

for $i = 1, \ldots, n$. The parameters m, M are chosen such that

$$0 < m \leq \min \left\{ 1, \frac{(1-\zeta)(1-\frac{\zeta}{(M_0)^\rho}) \min_i \left\{ [\mathbf{x}^{(k)}]_i [\mathbf{z}^{(k)}]_i \right\}}{\rho} \right\}, \tag{20}$$

and

$$M \geq \max \left\{ 1, \frac{\max_i \left\{ [\mathbf{x}^{(k)}]_i [\mathbf{z}^{(k)}]_i \right\}}{\rho} \right\} > 0, \tag{21}$$

where M_0 is a large positive number. These parameters are fixed to constants which satisfy (20), (21).

Remark 5.1 *Note that we could replace the terms* $[\mathbf{x}^{(k)}]_i[\mathbf{z}^{(k)}]_i$ *in* (20), (21) *by* $[\mathbf{x}^{(l)}]_i[\mathbf{z}^{(l)}]_i$ *and fix the two parameters to constants for the duration of inner iterations, i.e. have them changed when the barrier parameter* ρ *changes.*

The dual step length $[\alpha_z^{(k)}]_i$ is the minimum of all step lengths $[\alpha_z^{(k)}]_i$, that are not greater than one, *i.e.*

$$\alpha_z^{(k)} := \min \left\{ 1, \min_{i=1,\ldots,p} \left\{ [\alpha_z^{(k)}]_i \right\} \right\}.$$

Finally, we chose to set the step $\alpha_y^{(k)}$ for the $\mathbf{y}$ variables equal to $\alpha_z^{(k)}$.

5.1 Step acceptance

In Algorithm 3 the algorithm decides a value for $t_1 \in \{1, 2, 3\}$, which signifies the line search procedure that will be activated. As presented in Sections 4.1–4.3 the step acceptance rule requires that the merit function that corresponds to t_1 decreases sufficiently at each iteration (inequalities (7), (8) and (11) respectively). In order to prove convergence, we need to keep the iterates in a compact region. We achieve that through the *monotonic decrease* of one of the merit functions used by the algorithm.

There are two cases to consider in terms of the value the algorithm assigns to t_1 in Algorithm 3, at each iteration:

1. If the algorithm, after a certain iteration, chooses the same value for t_1, then Φ_{t_1} is monotonically decreasing. The iterates lie an a compact region and under appropriate assumptions the sequence converges.
2. If the algorithm chooses different values for t_1 from iteration to iteration, then we cannot deduce the monotonic decrease of one of the merit functions, and the iterates may not lie in a compact region. Convergence is not easy to establish in this case. In the next paragraph we discuss how to overcome this difficulty.

The general form of the Armijo condition mandates that $\alpha \in (0, 1]$ satisfies

$$\Phi_{t_1}(\mathbf{w}^{(k)} + \alpha\delta\mathbf{w}^{(k)}; \sigma^{(k+1)}, \rho) - \Phi_{t_1}(\mathbf{w}^{(k)}; \sigma^{(k+1)}, \rho) \leq -\alpha c_{t_1} \phi_{t_1}(\delta\mathbf{w}^{(k)}; \rho), \quad \text{(22a)}$$

which requires sufficient decrease of the activated merit function. As stated, (22a) is a generalization and takes the form of (7), (8), (11) for $t_1 \in \{1, 2, 3\}$ respectively. Since condition (22a) may not, on its own, guarantee monotonic decrease of Φ_1, Φ_2 or Φ_3 as discussed, we also require that Φ_2 decreases from iteration to iteration. That is we require that, along with the previous condition, $\alpha \in (0, 1]$ satisfies

$$\Phi_2(\mathbf{x}^{(k)} + \alpha\delta\mathbf{x}^{(k)}; \sigma^{(k+1)}, \rho) \leq \Phi_2(\mathbf{x}^{(k)}; \sigma^{(k+1)}, \rho). \quad \text{(22b)}$$

If $t_1 = 2$, then (22a) implies (22b), as long as $\delta\mathbf{x}^{(k)} \neq \mathbf{0}$. In Section 8.2 we show that the update of the penalty parameter in Algorithm 3 is sufficient to guarantee that the search direction is a direction of descent for Φ_2. In this sense, condition (22b) is milder than a condition of the form (22a), because it requires decrease, not sufficient decrease. Hence condition (22b) is not too hard to achieve. The step acceptance criterion for our algorithm is

$$SA(\alpha, \mathbf{w}^{(k)}, \delta\mathbf{w}^{(k)}, t_1) = \begin{cases} \text{true}, & \text{if } \alpha \in (0, 1] \text{ satisfies both (22a), (22b)} \\ \text{false}, & \text{otherwise} \end{cases}$$

6 Barrier parameter

The barrier parameter should be reduced to zero gradually, so that the perturbed KKT conditions (PRTKKT) resemble more and more the original KKT conditions (KKTIP). Then the points generated by the inner iteration satisfy, in the limit, the original KKT conditions. Interior point methods are very sensitive on the convergence properties of the barrier parameter. If convergence to zero is fast, the algorithm may fail as observed by Wright [27].

Bearing this in mind we adopt the barrier reduction strategy that was used by Akrotirianakis and Rustem [2]. It is a hybrid of the strategies of Lasdon et al. [14] and Gay et al. [9]. The new value of the barrier parameter ρ is determined by taking into consideration the distance of the current point $\mathbf{w}^{(k)}$ from the central path and the optimal solution of the initial problem. The details can be found in Algorithm 4.

Algorithm 4 Barrier parameter choice (PDIP)

1: Set $\rho^{(l+1)} := \min\left\{0.95\rho^{(l)}, 0.01(0.95)^k \|\mathbf{F}(\mathbf{w}^{(k)})\|\right\}$.
2: **if** ($\|\mathbf{F}(\mathbf{w}^{(k)}; \rho^{(l)})\| \leq 0.1\eta\rho^{(l)}$) **then**
3:    **if** ($\rho^{(l)} < 10^{-4}$) **then**
4:       Set $\rho^{(l+1)} := \min\left\{0.85\rho^{(l+1)}, 0.01(0.85)^{k+2q} \|\mathbf{F}(\mathbf{w}^{(k)})\|\right\}$
5:    **else**
6:       Set $\rho^{(l+1)} := \min\left\{0.85\rho^{(l+1)}, 0.01(0.85)^{k+q} \|\mathbf{F}(\mathbf{w}^{(k)})\|\right\}$
7:    **end if**
8: **end if**

In Algorithm 4, $\mathbf{F}(\mathbf{w}^{(k)}, \rho^{(l)})$ and $\mathbf{F}(\mathbf{w}^{(k)})$ represent the perturbed and unperturbed KKT conditions of problem (NLPIP), as given by (3) and (KKTIP), respectively. The condition of Step 2 checks closeness of the current iterate to the central path. The condition of Step 3 checks closeness of the current iterate to the solution. If the current iterate is close to the central path and to the optimal solution, then the barrier parameter is reduced at a fast rate, signified by factor $(0.85)^{2q}$, where $q > 0$. If on the other hand, it is close to the central path, the barrier parameter is decreased at a lower rate than before, as factor 0.85^q shows.

7 Convergence criteria

In this section we present convergence criteria for the outer iterations as well as for the inner iterations. The convergence criterion for the outer iterations is

$$\frac{\|\mathbf{F}(\mathbf{x}^{(l)}, \mathbf{y}^{(l)}, \mathbf{z}^{(l)})\|}{1 + \|\mathbf{x}^{(l)}, \mathbf{y}^{(l)}, \mathbf{z}^{(l)}\|} \leq \epsilon_0, \tag{23}$$

where ϵ_0 is a predetermined constant. If this criterion is satisfied, then the algorithm will terminate. Such a rule has also been used by El-Bakry et al. [6] and Akrotirianakis and Rustem [2] for example.

The convergence criterion for the inner iterations is

$$\|\mathbf{F}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}, \mathbf{z}^{(k)}; \rho^{(l)})\| \leq \eta\rho^{(l)}. \tag{24}$$

This is a termination criterion that has been used in primal-dual frameworks by Akrotirianakis and Rustem [2] and Yamashita [28] for example, in order to find points that are close to the central path defined by $\rho^{(l)}$.

8 Global convergence

In this section we prove convergence of the proposed algorithm from remote points. We use a monotonic decrease argument in order to achieve our main result which is Theorem (19).

In Section 8.1 we are concerned with properties of the l_∞ merit function, that will help us establish global convergence. We also display the behaviour of inner iterations, if the algorithm always chooses $t_1 = 1$ after a certain iteration. In Section 8.2 we present properties associated with the l_2 merit function, and also display the behaviour of the inner iterations, for problems for which the algorithm always chooses $t_1 = 2$ after a certain iteration. In Section 8.3 we are concerned with properties of the algorithm on general problems, *i.e.* problems for which the algorithm may not activate a single merit function.

For convenience we shall use the following notation

$$G_F = \int_0^1 (1-t)\langle \delta\mathbf{x}^{(k)}, \nabla^2_{\mathbf{x}} F(\mathbf{x}^{(k)} + t\alpha\delta\mathbf{x}^{(k)})\rangle \delta\mathbf{x}^{(k)} dt,$$

where F is either the objective function $f = f(\mathbf{x})$, the barrier term $B = B(\mathbf{x})$ or the logarithmic barrier function $f_\rho = f(\mathbf{x}; \rho) = f(\mathbf{x}) - \rho B(\mathbf{x})$.

8.1 Properties of the l_∞ merit function

The next lemma is a result about the sign of ϕ_1. It will be used subsequently to establish further results and global convergence of the algorithm.

Lemma 1 (The sign of the descent function)
Assume that A1 holds. If $\sigma^{(k+1)}$ *is sufficiently large, i.e.*

$$\sigma^{(k+1)} \geq \|\hat{\mathbf{y}}^{(k)}\|_1, \tag{25}$$

then

$$\phi_1(\mathbf{x}^{(k)}, \hat{\mathbf{y}}^{(k)}; \sigma^{(k+1)}) \geq 0.$$

Proof : We have that

$$\begin{aligned} -\langle \hat{\mathbf{y}}^{(k)}, \mathbf{h}(\mathbf{x}^{(k)}) \rangle &\geq -P_1(\mathbf{x}^{(k)}) \sum_{i=1}^{m} \left| [\hat{\mathbf{y}}^{(k)}]_i \right| \\ &= -P_1(\mathbf{x}^{(k)}) \|\hat{\mathbf{y}}^{(k)}\|_1 \\ &\geq -\sigma^{(k+1)} P_1(\mathbf{x}^{(k)}), \end{aligned}$$

where the first inequality follows from the definition of the penalty term. This completes the proof of the lemma. □

Whenever in this section we say that $\sigma^{(k+1)}$ is sufficiently large we mean that it satisfies inequality (25). The following bound is necessary for establishing global and local convergence of the algorithm, when the current line search strategy is activated by the algorithm.

Lemma 2
Assume that Assumptions (A1), (A3) hold and that $\sigma^{(k+1)}$ *is sufficiently large. For the iterations* k *for which there holds*

$$\tau^{(k)} + \langle \delta\mathbf{x}^{(k)}, \boldsymbol{\nabla} f(\mathbf{x}^{(k)}; \rho) \rangle > 0 \tag{26}$$

there exists $\beta_5 > 0$ *bounded from above such that*

$$\|\delta\mathbf{x}^{(k)}\| \leq \beta_5 P_1(\mathbf{x}^{(k)}).$$

Proof : As in Powell and Yuan [24] let $\beta_8 > 0$ satisfy the condition

$$2\beta_8 \|\mathbf{N}^{(k)}\| + \beta_8^2 \|\mathbf{N}^{(k)}\| \leq \frac{1}{2}\beta_2 \tag{27}$$

for all k. From the orthogonal decomposition of the primal search direction into two orthogonal components $\mathbf{v}, \mathbf{u}$ and since $\mathbf{v}$ is the shortest vector $\delta\mathbf{x}$ that can satisfy (4b) we have the relation

$$\|\mathbf{v}^{(k)}\| \leq \beta_4 \|\mathbf{h}(\mathbf{x}^{(k)})\|, \tag{28}$$

for a positive constant β_4. When $\|\mathbf{v}^{(k)}\| \geq \beta_8 \|\mathbf{u}^{(k)}\|$, the bound

$$\begin{aligned}
\|\mathbf{h}(\mathbf{x}^{(k)})\|^2 &\geq \beta_4^{-2}\|\mathbf{u}^{(k)}\| \geq \frac{1}{\beta_4^2(1+\beta_8^{-1})^2}(\|\mathbf{v}^{(k)}\| + \|\mathbf{u}^{(k)}\|)^2 \\
&= \frac{1}{\beta_4^2(1+\beta_8^{-1})^2}\|\delta\mathbf{x}^{(k)}\|^2 \qquad (29)
\end{aligned}$$

is satisfied. Therefore from norm equivalence we obtain

$$\|\delta\mathbf{x}^{(k)}\| \leq \beta_5 P_1(\mathbf{x}^{(k)}).$$

for

$$\beta_5 \geq \beta_4(1+\beta_8^{-1}). \qquad (30)$$

Since β_4 is a positive constant and $\beta_8 > 0$, we can deduce that the right hand side of (30) is not infinite and therefore β_5 is bounded from above.

Alternatively, when $\|\mathbf{v}^{(k)}\| < \beta_8\|\mathbf{u}^{(k)}\|$, the Cauchy-Schwarz inequality, (28) and Assumption (A3) imply

$$\begin{aligned}
\langle \delta\mathbf{x}^{(k)}, \mathbf{N}^{(k)} \rangle \delta\mathbf{x}^{(k)} &= \langle (\mathbf{v}^{(k)} + \mathbf{u}^{(k)}), \mathbf{N}^{(k)} \rangle (\mathbf{v}^{(k)} + \mathbf{u}^{(k)}) \geq \|\mathbf{u}^{(k)}\|^2(\beta_2 - 2\beta_8\|\mathbf{N}\| - \beta_8^2\|\mathbf{N}\|) \\
&\geq \frac{1}{2}\beta_2\|\mathbf{u}^{(k)}\|^2 > \frac{\beta_2}{2(1+\beta_8)^2}(\|\mathbf{v}^{(k)}\| + \|\mathbf{u}^{(k)}\|)^2 \\
&= \frac{\beta_2}{2(1+\beta_8)^2}\|\delta\mathbf{x}^{(k)}\|^2. \qquad (31)
\end{aligned}$$

Using (5a), (5b) we can write

$$\begin{aligned}
\sigma^{(k+1)} P_1(\mathbf{x}) &\geq \sum_{i=1}^{m} [\hat{\mathbf{y}}^{(k)}]_i P_1(\mathbf{x}^{(k)}) \\
&\geq -\langle \hat{\mathbf{y}}^{(k)}, \mathbf{h}(\mathbf{x}^{(k)}) \rangle \\
&= \langle \delta\mathbf{x}^{(k)}, \boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^{(k)})\hat{\mathbf{y}}^{(k)} \rangle \\
&= (\langle \delta\mathbf{x}^{(k)}, \mathbf{N}^{(k)} \rangle \delta\mathbf{x}^{(k)} + \langle \delta\mathbf{x}^{(k)}, \boldsymbol{\nabla} f(\mathbf{x}^{(k)}; \rho) \rangle) \\
&\geq \frac{\beta_2}{2(1+\beta_8)^2}\|\delta\mathbf{x}^{(k)}\|^2 + \langle \delta\mathbf{x}^{(k)}, \boldsymbol{\nabla} f(\mathbf{x}^{(k)}; \rho) \rangle,
\end{aligned}$$

which gives

$$\|\delta\mathbf{x}^{(k)}\| \leq \beta_5 P_1(\mathbf{x}^{(k)}) \qquad (32)$$

for $r^{(k)}$ computed by (14) and

$$\beta_5 \geq \sigma^{(k+1)} \frac{\|\delta \mathbf{x}^{(k)}\|}{\frac{\beta_2}{2(1+\beta_8)^2}\|\delta \mathbf{x}^{(k)}\|^2 + \langle \delta \mathbf{x}^{(k)}, \boldsymbol{\nabla} f(\mathbf{x}^{(k)};\rho)\rangle}. \tag{33}$$

If (26) is satisfied, then $\delta \mathbf{x}^{(k)} \neq \mathbf{0}$ for all such k. In such cases the fractional part on the right hand side of (33) is bounded from above. If $\sigma^{(k+1)}$ does not tend to infinity, then β_5 satisfying (33) is not infinite. □

Lemma 3
Assume that A1 holds. If $\sigma^{(k+1)}$ is sufficiently large, then there exists $\beta_6 > 0$, such that

$$P_1(\mathbf{x}^{(k)}) \leq \beta_6 \phi_1(\mathbf{x}^{(k)}, \hat{\mathbf{y}}^{(k)}; \sigma^{(k+1)}).$$

Proof : From the definition of ϕ_1 we have that

$$\begin{aligned}
\phi_1(\mathbf{x}^{(k)}, \hat{\mathbf{y}}^{(k)}; \sigma^{(k+1)}) &= \sigma^{(k+1)} P_1(\mathbf{x}^{(k)}) - \langle \hat{\mathbf{y}}^{(k)}, \mathbf{h}(\mathbf{x}^{(k)})\rangle \\
&\geq \sigma^{(k+1)} P_1(\mathbf{x}^{(k)}) - \sum_{i=1}^{m} \left| [\hat{\mathbf{y}}^{(k)}]_i \right| P_1(\mathbf{x}^{(k)}) \\
&\geq \left(\sigma^{(k+1)} - \|\hat{\mathbf{y}}^{(k)}\|_1\right) P_1(\mathbf{x}^{(k)}).
\end{aligned}$$

Therefore if

$$\sigma^{(k+1)} > \|\hat{\mathbf{y}}^{(k)}\|_1,$$

then if we set

$$\beta_6 = \frac{1}{\sigma^{(k+1)} - \|\hat{\mathbf{y}}^{(k)}\|_1}$$

the lemma holds. □

Lemma 4
If the assumptions of Lemma (3) hold, then there exists $\beta_7 > 0$ bounded from above such that

$$\|\delta \mathbf{x}^{(k)}\| \leq \beta_7 \phi_1(\mathbf{x}^{(k)}, \hat{\mathbf{y}}^{(k)}; \sigma^{(k+1)}).$$

Proof : A direct consequence of Lemmas (2), (3) for $\beta_7 = \beta_5\beta_6$. □

Remark 8.1 *Quantity ψ has been introduced to cope with the fact that we do not assume that $\mathbf{N}^{(k)}$ is positived definite.*

In order to simplify notation, the primal step size $\alpha_x^{(k)}$ will be denoted as $\alpha^{(k)}$.

Theorem 1 (Decrease of $\Phi(\mathbf{x}^{(k)}; \sigma^{(k+1)}, \rho)$)
Let Assumptions (A1), (A2), (A4) hold and also let condition (26) *be satisfied. If $\sigma^{(k+1)}$ is sufficiently large, then for ψ sufficiently large, there exists $\alpha_x^{(k)} \in (\frac{1}{\psi+0.5}, 1]$ such that the Armijo condition* (7) *is satisfied for $c_1 \in (0,1)$.*

Proof : We first find suitable bounds on the components of Φ_1, and then use them in conjunction with the previous lemmas to show that the Armijo condition (7) is well defined.

A *first order Taylor series expansion* on $[h]_i(\mathbf{x})$ around $\bar{\mathbf{x}} = \mathbf{x} + \alpha\delta\mathbf{x}$ yields

$$\begin{aligned}
[h]_i(\bar{\mathbf{x}}) &= [h]_i(\mathbf{x}) + \alpha\langle \delta\mathbf{x}, \boldsymbol{\nabla}[h]_i(\mathbf{x} + \xi_i\alpha\delta\mathbf{x})\rangle, \qquad 0 \le \xi_i \le 1 \\
&= [h]_i(\mathbf{x}) + \alpha\langle \delta\mathbf{x}, \boldsymbol{\nabla}[h]_i(\mathbf{x})\rangle + \alpha\langle \delta\mathbf{x}, (\boldsymbol{\nabla}[h]_i(\mathbf{x} + \xi_i\alpha\delta\mathbf{x}) - \boldsymbol{\nabla}[h]_i(\mathbf{x}))\rangle \\
&\le [h]_i(\mathbf{x}) + \alpha\langle \delta\mathbf{x}, \boldsymbol{\nabla}[h]_i(\mathbf{x})\rangle + \alpha\|\delta\mathbf{x}\|\|\boldsymbol{\nabla}[h]_i(\mathbf{x} + \xi_i\alpha\delta\mathbf{x}) - \boldsymbol{\nabla}[h]_i(\mathbf{x})\| \\
&\le [h]_i(\mathbf{x}) + \alpha\langle \delta\mathbf{x}, \boldsymbol{\nabla}[h]_i(\mathbf{x})\rangle + \alpha^2\gamma\|\delta\mathbf{x}\|^2 \\
&= [h]_i(\mathbf{x}) - \alpha[h]_i(\mathbf{x}) + \alpha^2\gamma\|\delta\mathbf{x}\|^2,
\end{aligned}$$

where in the last inequality we have used Assumption (A4) and the last equality (4b). Maximizing both sides for all $i = 1, \ldots, m$ we obtain

$$P_1(\bar{\mathbf{x}}) \le P_1(\mathbf{x}) - \alpha P_1(\mathbf{x}) + \alpha^2\gamma\|\delta\mathbf{x}\|^2. \tag{34}$$

From a *second order Taylor series expansion* for terms of the logarithmic barrier function

$$\begin{aligned}
\log[\bar{\mathbf{x}}]_i &= \log[\mathbf{x}]_i + \alpha[\delta\mathbf{x}]_i([\mathbf{x}]_i)^{-1} + \\
&\quad - \alpha^2 \int_0^1 (1-t)[\delta\mathbf{x}]_i[\mathbf{x} + t\alpha\delta\mathbf{x}]_i^{-2}[\delta\mathbf{x}]_i dt.
\end{aligned}$$

Summing the above over i we obtain

$$\begin{aligned}
\sum_{i=1}^{n} \log[\bar{\mathbf{x}}]_i &= \sum_{i=1}^{n} \log[\mathbf{x}]_i + \alpha\langle \delta\mathbf{x}, \mathbf{X}^{-1}\mathbf{e}\rangle + \\
&\quad - \alpha^2 \int_0^1 (1-t)\langle \delta\mathbf{x}, (\mathbf{X} + t\alpha\delta\mathbf{X})^{-2}\rangle\delta\mathbf{x} dt,
\end{aligned}$$

or

$$B(\bar{\mathbf{x}}) = B(\mathbf{x}) + \alpha\langle \delta\mathbf{x}, \mathbf{X}^{-1}\mathbf{e}\rangle + \alpha^2 G_B. \tag{35}$$

In a similar manner for the objective function

$$f(\bar{\mathbf{x}}) = f(\mathbf{x}) + \alpha\langle\delta\mathbf{x}, \boldsymbol{\nabla} f(\mathbf{x})\rangle + \alpha^2 G_f, \tag{36}$$

If we combine (35), (36) using $f(\mathbf{x};\rho) = f(\mathbf{x}) - \rho B(\mathbf{x})$ we can write

$$\begin{aligned}
f(\bar{\mathbf{x}};\rho) &= f(\bar{\mathbf{x}}) - \rho B(\bar{\mathbf{x}}) \\
&= f(\mathbf{x}) + \alpha\langle\delta\mathbf{x}, \boldsymbol{\nabla} f(\mathbf{x})\rangle + \alpha^2 G_f \\
&\quad - \rho B(\mathbf{x}) - \rho\alpha\langle\delta\mathbf{x}, \mathbf{X}^{-1}\mathbf{e}\rangle - \rho\alpha^2 G_B \\
&= f(\mathbf{x};\rho) + \alpha\langle\delta\mathbf{x}, \boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle + \alpha^2 G_{f_\rho} \\
&\leq f(\mathbf{x};\rho) + \alpha\langle\delta\mathbf{x}, \boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle + \frac{\alpha^2}{2}\langle\delta\mathbf{x}, \mathbf{N}\rangle\delta\mathbf{x} + \eta\alpha^2\|\delta\mathbf{x}\|^2 \qquad (37)
\end{aligned}$$

where

$$\eta = \int_0^1 (1-t)\|\boldsymbol{\nabla}^2 f(\mathbf{x} + t\alpha\delta\mathbf{x};\rho) - \mathbf{N}\| dt.$$

Using (5a) we can write

$$\alpha\langle\delta\mathbf{x}, \boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle = -\alpha\langle\delta\mathbf{x}, \boldsymbol{\nabla}\mathbf{h}(\mathbf{x})\hat{\mathbf{y}}\rangle - \alpha\langle\delta\mathbf{x}, \mathbf{N}\rangle\delta\mathbf{x}$$

and substituting in (37) we obtain

$$\begin{aligned}
f(\bar{\mathbf{x}};\rho) &\leq f(\mathbf{x};\rho) - \alpha\langle\delta\mathbf{x}, \boldsymbol{\nabla}\mathbf{h}(\mathbf{x})\hat{\mathbf{y}}\rangle + \left(\frac{\alpha^2}{2} - \alpha\right)\langle\delta\mathbf{x}, \mathbf{N}\rangle\delta\mathbf{x} + \eta\alpha^2\|\delta\mathbf{x}\|^2 \\
&\leq f(\mathbf{x};\rho) - \alpha\langle\delta\mathbf{x}, \boldsymbol{\nabla}\mathbf{h}(\mathbf{x})\hat{\mathbf{y}}\rangle + \left(\alpha - \frac{\alpha^2}{2}\right)\|\mathbf{N}\|\|\delta\mathbf{x}\|^2 + \eta\alpha^2\|\delta\mathbf{x}\|^2 \\
&\leq f(\mathbf{x};\rho) - \alpha\langle\delta\mathbf{x}, \boldsymbol{\nabla}\mathbf{h}(\mathbf{x})\hat{\mathbf{y}}\rangle + \left(\alpha - \frac{\alpha^2}{2}\right)\beta_1\|\delta\mathbf{x}\|^2 + \\
&\quad + \eta\alpha^2\|\delta\mathbf{x}\|^2. \qquad (38)
\end{aligned}$$

The inequality before last follows from the Cauchy-Schwarz inequality since $\frac{\alpha^2}{2} - \alpha < 0$. The last inequality follows from Assumption (A2). If we choose $\psi \geq \frac{1}{2}$, then for

$$\alpha \geq \frac{1}{\psi + \frac{1}{2}} \tag{39}$$

we can write $\alpha - \frac{\alpha^2}{2} \leq \psi\alpha^2$ and (38) becomes

$$\begin{aligned}
f(\bar{\mathbf{x}};\rho) &\leq f(\mathbf{x};\rho) - \alpha\langle\delta\mathbf{x}, \boldsymbol{\nabla}\mathbf{h}(\mathbf{x})\hat{\mathbf{y}}\rangle + \psi\beta_1\alpha^2\|\delta\mathbf{x}\|^2 + \eta\alpha^2\|\delta\mathbf{x}\|^2 \\
&= f(\mathbf{x};\rho) - \alpha\langle\delta\mathbf{x}, \boldsymbol{\nabla}\mathbf{h}(\mathbf{x})\hat{\mathbf{y}}\rangle + (\psi\beta_1 + \eta)\alpha^2\|\delta\mathbf{x}\|^2. \qquad (40)
\end{aligned}$$

Using the definition of Φ_1, ϕ_1 and combining (34), (40) we obtain

$$\begin{aligned}
\Phi_1(\bar{\mathbf{x}};\bar{\sigma},\rho) &= f(\bar{\mathbf{x}};\rho)+\bar{\sigma}P_1(\bar{\mathbf{x}}) \\
&\leq f(\mathbf{x};\rho)-\alpha\langle\delta\mathbf{x},\boldsymbol{\nabla}\mathbf{h}(\mathbf{x})\hat{\mathbf{y}}\rangle+(\psi\beta_1+\eta)\alpha^2\|\delta\mathbf{x}\|^2 \\
&\quad +\bar{\sigma}P_1(\mathbf{x})-\bar{\sigma}\alpha P_1(\mathbf{x})+\bar{\sigma}\alpha^2\gamma\|\delta\mathbf{x}\|^2 \\
&= f(\mathbf{x};\rho)+\bar{\sigma}P_1(\mathbf{x})-\alpha\left(\bar{\sigma}P_1(\mathbf{x})+\langle\delta\mathbf{x},\boldsymbol{\nabla}\mathbf{h}(\mathbf{x})\hat{\mathbf{y}}\rangle\right) \\
&\quad +(\psi\beta_1+\eta+\bar{\sigma}\gamma)\alpha^2\|\delta\mathbf{x}\|^2 \\
&= \Phi_1(\mathbf{x};\bar{\sigma},\rho)-\alpha\phi_1(\mathbf{x},\hat{\mathbf{y}};\bar{\sigma})+(\psi\beta_1+\eta+\bar{\sigma}\gamma)\alpha^2\|\delta\mathbf{x}\|^2.
\end{aligned}$$

For convenience we introduce $\Delta\Phi_1 = \Phi_1(\bar{\mathbf{x}};\bar{\sigma},\rho)-\Phi_1(\mathbf{x};\bar{\sigma},\rho)$. The conditions of Lemmas 2, 4 are satisfied, therefore

$$\begin{aligned}
\Delta\Phi_1 &\leq -\alpha\phi_1(\mathbf{x},\hat{\mathbf{y}};\bar{\sigma})+(\psi\beta_1+\eta+\bar{\sigma}\gamma)\alpha^2\|\delta\mathbf{x}\|^2 \\
&\leq -\alpha\phi_1(\mathbf{x},\hat{\mathbf{y}};\bar{\sigma})+(\psi\beta_1+\eta+\bar{\sigma}\gamma)\alpha^2\beta_5P_1(\mathbf{x})\|\delta\mathbf{x}\| \\
&\leq -\alpha\phi_1(\mathbf{x},\hat{\mathbf{y}};\bar{\sigma})+(\psi\beta_1+\eta+\bar{\sigma}\gamma)\alpha^2\beta_5P_1(\mathbf{x})\beta_7\phi_1(\mathbf{x},\hat{\mathbf{y}};\bar{\sigma}) \\
&= -\alpha\phi_1(\mathbf{x},\hat{\mathbf{y}};\bar{\sigma})\left(1-\alpha\beta_5\beta_7P_1(\mathbf{x})(\psi\beta_1+\eta+\bar{\sigma}\gamma)\right). \qquad (41)
\end{aligned}$$

From inequality (41), and since $0<c_1<1$, there is $\alpha\in(0,1]$ satisfying (39) such that

$$c \leq 1-\alpha\beta_5\beta_7P_1(\mathbf{x})(\psi\beta_1+\eta+\bar{\sigma}\gamma)\leq 1. \qquad (42)$$

By Lemma (1), we have that $\phi_1(\mathbf{x},\hat{\mathbf{y}};\bar{\sigma})\geq 0$, therefore inequality (7) must be satisfied for this α. Assume that $\tilde{\alpha}$ is the largest step in the interval $(\frac{1}{\psi+0.5},1]$ satisfying inequality (7). Then for every $\alpha\leq\tilde{\alpha}$ Armijo's condition is satisfied and the selected $\alpha\in[\theta\tilde{\alpha},\tilde{\alpha}]$. □

The rest of the properties of this section concern problems for which there exists $k_1\geq 0$ such that the algorithm sets $t_1=1$ for all $k\geq k_1$. In such situations Φ_1 is monotonically decreasing, using the previous theorem. We can then prove the following lemma[1].

[1]We do not suppress iterate scripts in this proof

Lemma 5
Let Assumptions (A1), (A2), (A4) hold, and let for $k_0 \geq 0$ and $k \geq k_0$, the set

$$\mathcal{F}_1 = \left\{ \mathbf{x} > 0 \,\middle|\, \Phi_1(\mathbf{x};\sigma,\rho) \leq \Phi_1(\mathbf{x}^{(k_0)};\sigma,\rho) \right\} \tag{43}$$

be compact. In addition, if there exists an integer k_1, such that for $k \geq k_1$ the algorithm chooses $t_1 = 1$, then for all $k \geq \max\{k_0, k_1\}$ we have that

$$\lim_{k\to\infty} \phi_1(\mathbf{x}^{(k)}, \hat{\mathbf{y}}^{(k)}; \sigma^{(k+1)}) = 0. \tag{44}$$

Proof : The scalar $c_1 \in (0,1)$ in the Armijo condition (7) corresponding to $t_1 = 1$ (Step 15), determines a stepsize $\alpha^{(k)}$ such that

$$c_1 \leq 1 - \alpha^{(k)}\beta_5\beta_7 P_1(\mathbf{x}^{(k)})(\psi\beta_1 + \eta^{(k)} + \sigma^{(k+1)}\gamma) \leq 1.$$

Solving for $\alpha^{(k)}$ we obtain

$$\alpha^{(k)} \leq \frac{1 - c_1}{\beta_5\beta_7 P_1(\mathbf{x}^{(k)})\left(\psi\beta_1 + \eta^{(k)} + \gamma\sigma^{(k+1)}\right)}. \tag{45}$$

Therefore, in order to satisfy the Armijo condition, the largest value the stepsize parameter $\alpha^{(k)}$ can take is

$$\tilde{\alpha}^{(k)} = \min\left\{1, \frac{1 - c_1}{\beta_5\beta_7 P_1(\mathbf{x}^{(k)})\left(\psi\beta_1 + \eta^{(k)} + \gamma\sigma^{(k+1)}\right)}\right\}.$$

As f is twice continuously differentiable and the level set $\mathcal{F}_1$ is bounded, it follows that there exists a scalar $\tilde{\eta}$ such that

$$\eta \leq \tilde{\eta} < \infty.$$

Also from continuity of P_1 we always have that

$$\alpha^{(k)} \geq \tilde{\alpha}^{(k)} > 0.$$

As shown in Lemmas 3, 4, quantities β_5, β_7 are bounded from above. We also show in Lemma (16) that $\sigma^{(k)}$ does not go to infinity. Taking also into account the continuity of P_1, we always have that

$$\alpha^{(k)} \geq \tilde{\alpha}^{(k)} > 0.$$

Furthermore, from the Armijo rule (7) and the sign of ϕ_1 we have that

$$\Phi_1(\mathbf{x}^{(k+1)};\sigma,\rho) - \Phi_1(\mathbf{x}^{(k)};\sigma,\rho) \leq c_1\alpha^{(k)}\phi_1(\mathbf{x}^{(k)}, \hat{\mathbf{y}}^{(k)}; \sigma^{(k+1)}) \leq 0. \tag{46}$$

By the boundedness assumption on $\mathcal{F}_1$, we deduce that

$$\lim_{k\to\infty} \left| \{\Phi_1(\mathbf{x}^{(k+1)};\sigma,\rho) - \Phi_1(\mathbf{x}^{(k)};\sigma,\rho)\} \right| = 0$$

and since $c_1, \alpha^{(k)} > 0$, the lemma follows from (46). □

Lemma 6
Let the assumptions of the previous lemma hold. Then

$$\lim_{k\to\infty} \|\delta \mathbf{x}^{(k)}\| = 0. \tag{47}$$

for all $k \geq \max\{k_0, k_1\}$.

Proof : The result follows directly from Lemmas (4), (5). □

8.2 Properties of the l_2 merit function

Lemmas (7)–(11) show that if the penalty parameter is chosen according to Algorithm 2, then $\delta \mathbf{x}^{(k)}$ is a direction of descent for Φ_2. This result means that imposing (22b) in (22a) in order to establish monotonic decrease of Φ_2 is not so hard to achieve.

Lemma 7
Let Assumption (A1) hold. If $\delta \mathbf{w}^{(k)}$ *is calculated by* (PRTLNS) *and* $\sigma^{(k+1)}$ *is chosen as in Step 13 of Algorithm 2, then* Φ_2 *decreases along* $\delta \mathbf{x}^{(k)}$.

Proof : Using the fact that in Step 13 of Algorithm 2

$$r\|\delta\mathbf{x}\|^2 + \langle \delta\mathbf{x}, \boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle \leq \sigma\|\mathbf{h}(\mathbf{x})\|^2,$$

we have that

$$-\phi_2(\mathbf{x}, \hat{\mathbf{y}}; \bar{\sigma}, \rho) = \langle \delta\mathbf{x}, \boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle - \bar{\sigma}\|\mathbf{h}(\mathbf{x})\|^2 \leq -r\|\delta\mathbf{x}\|^2,$$

which is the required result. □

Lemma 8
Let Assumption (A1) hold. If $\delta \mathbf{w}^{(k)}$ *is calculated by* (PRTLNS) *and* $\sigma^{(k+1)}$ *is chosen as in Step 15 of Algorithm 2, then* Φ_2 *decreases along* $\delta \mathbf{x}^{(k)}$.

Proof : In Step 15 of Algorithm 2 there holds

$$r\|\delta\mathbf{x}\|^2 + \langle \delta\mathbf{x}, \boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle \leq \sigma\|\mathbf{h}(\mathbf{x})\|^2.$$

If the penalty parameter is updated as in Step 15, then from (10)

$$-\phi_2(\mathbf{x}, \hat{\mathbf{y}}; \bar{\sigma}, \rho) = \langle \delta\mathbf{x}, \boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle - \bar{\sigma}\|\mathbf{h}(\mathbf{x})\|^2 \leq -r\|\delta\mathbf{x}\|^2,$$

which is the required result. □

Lemma 9
Let Assumption (A1) hold. If $\delta\mathbf{w}^{(k)}$ is calculated by (PRTLNS) *and $\sigma^{(k+1)}$ is chosen as in Step 5 of Algorithm 2, then Φ_2 decreases along $\delta\mathbf{x}^{(k)}$.*

Proof : If $\mathbf{x}$ is feasible then any descent property associated with Φ_1 will also hold for Φ_2. Therefore we assume that $\mathbf{x}$ is not feasible and in fact that $\|\mathbf{h}(\mathbf{x})\| > \sqrt{\epsilon_h}$.

From the update of the penalty parameter in Step 5 of Algorithm 2 and the non-feasibility assumption we obtain

$$\begin{aligned}
\bar{\sigma} &\geq \frac{\sqrt{m}}{\sqrt{\epsilon_h}}\|\hat{\mathbf{y}}\|_1 + \vartheta \\
&> \frac{\sqrt{m}\|\hat{\mathbf{y}}\|_1}{\|\mathbf{h}(\mathbf{x})\|} + \frac{\vartheta\sqrt{\epsilon_h}}{\|\mathbf{h}(\mathbf{x})\|} \\
&= \frac{\|\hat{\mathbf{y}}\|_1\sqrt{m}\|\mathbf{h}(\mathbf{x})\|}{\|\mathbf{h}(\mathbf{x})\|^2} + \frac{\vartheta\sqrt{\epsilon_h}P_1(\mathbf{x})}{\|\mathbf{h}(\mathbf{x})\|_\infty\|\mathbf{h}(\mathbf{x})\|} \\
&\geq \frac{\|\hat{\mathbf{y}}\|_1\|\mathbf{h}(\mathbf{x})\|_1}{\|\mathbf{h}(\mathbf{x})\|^2} + \frac{\vartheta\sqrt{\epsilon_h}P_1(\mathbf{x})}{\|\mathbf{h}(\mathbf{x})\|^2} && \text{(48a)} \\
&= \frac{\sum_{i=1}^m |[\hat{\mathbf{y}}]_i| \sum_{i=1}^m |[h]_i(\mathbf{x})|}{\|\mathbf{h}(\mathbf{x})\|^2} + \frac{\vartheta\sqrt{\epsilon_h}P_1(\mathbf{x})}{\|\mathbf{h}(\mathbf{x})\|^2}, && \text{(48b)}
\end{aligned}$$

where (48a) holds from norm equivalence theorems [see 20, for example], and (48b) by the definition of the l_1 norm.

If we continue further we obtain

$$\begin{aligned}
\bar{\sigma} &\geq \frac{\sum_{i=1}^m |[\hat{\mathbf{y}}]_i[h]_i(\mathbf{x})|}{\|\mathbf{h}(\mathbf{x})\|^2} + \frac{\vartheta\sqrt{\epsilon_h}P_1(\mathbf{x})}{\|\mathbf{h}(\mathbf{x})\|^2} \\
&= \frac{\sum_{i=1}^m [\hat{\mathbf{y}}]_i[h]_i(\mathbf{x})}{\|\mathbf{h}(\mathbf{x})\|^2} + \frac{\vartheta\sqrt{\epsilon_h}P_1(\mathbf{x})}{\|\mathbf{h}(\mathbf{x})\|^2} && \text{(49a)} \\
&= \frac{\langle\delta\mathbf{x}, \boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle + \langle\delta\mathbf{x}, \mathbf{N}\rangle\delta\mathbf{x} + \vartheta\sqrt{\epsilon_h}P_1(\mathbf{x})}{\|\mathbf{h}(\mathbf{x})\|^2}, && \text{(49b)}
\end{aligned}$$

where equality (49a) follows from (5a) and (26) and equality (49b) from (5a).

If $\langle\delta\mathbf{x}, \mathbf{N}\rangle\delta\mathbf{x} < 0$, then using the definition (15) of ϑ we can write (49b) as

$$\bar{\sigma} \geq \frac{\langle\delta\mathbf{x}, \boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle + r\|\delta\mathbf{x}\|^2}{\|\mathbf{h}(\mathbf{x})\|^2}.$$

Therefore $\bar{\sigma}$ is large enough to guarantee descent of Φ_2 along $\delta\mathbf{x}$, and in fact

$$-\phi_2(\mathbf{x},\hat{\mathbf{y}};\bar{\sigma},\rho) \leq -r\|\delta\mathbf{x}\|^2 \leq 0.$$

If $\langle\delta\mathbf{x},\mathbf{N}\rangle\delta\mathbf{x} \geq 0$, using again the definition of ϑ we can write (49b) as

$$\bar{\sigma} \quad \geq \quad \frac{\langle\delta\mathbf{x},\boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle + \langle\delta\mathbf{x},\mathbf{N}\rangle\delta\mathbf{x}}{\|\mathbf{h}(\mathbf{x})\|^2}.$$

Therefore $\bar{\sigma}$ is large enough to guarantee descent of Φ_2 along $\delta\mathbf{x}$, and in fact if, for the iterations for which there holds $\langle\delta\mathbf{x},\mathbf{N}\rangle\delta\mathbf{x} \geq 0$, there exists $\beta_3 > 0$ such that

$$\langle\delta\mathbf{x},\mathbf{N}\rangle\delta\mathbf{x} \geq \beta_3\|\delta\mathbf{x}\|^2 \tag{50}$$

then

$$-\phi_2(\mathbf{x},\hat{\mathbf{y}};\bar{\sigma},\rho) \leq -\langle\delta\mathbf{x},\mathbf{N}\rangle\delta\mathbf{x} \leq -\beta_3\|\delta\mathbf{x}\|^2 \leq 0.$$

□

Lemma 10
Let Assumption (A1) hold. If $\delta\mathbf{w}^{(k)}$ is calculated by (PRTLNS) *and $\sigma^{(k+1)}$ is chosen as in Step 21 of Algorithm 2, then Φ_2 decreases along $\delta\mathbf{x}^{(k)}$.*

Proof : In Step 21 of Algorithm 2 there holds

$$r\|\delta\mathbf{x}\|^2 + \langle\delta\mathbf{x},\boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle \leq 0,$$

then

$$\begin{aligned} -\phi_2(\mathbf{x},\hat{\mathbf{y}};\bar{\sigma},\rho) &= \langle\delta\mathbf{x},\boldsymbol{\nabla} f(\mathbf{x};\rho)\rangle - \bar{\sigma}\|\mathbf{h}(\mathbf{x})\|^2 \\ &\leq -r\|\delta\mathbf{x}\|^2 - \bar{\sigma}\|\mathbf{h}(\mathbf{x})\|^2 \\ &\leq -r\|\delta\mathbf{x}\|^2, \end{aligned}$$

which is the required result. □

In the next lemma we show that Φ_2 decreases even when the constraints are satisfied and the penalty parameter is not updated (Step 9).

Lemma 11
Let Assumptions (A1), (A3) be satisfied. If $\mathbf{h}(\mathbf{x}^{(k)}) = \mathbf{0}$, $\delta\mathbf{w}^{(k)}$ is calculated by (PRTLNS) *and $\sigma^{(k+1)}$ is chosen as in Step 9 of Algorithm 2, then Φ_2 decreases along $\delta\mathbf{x}^{(k)}$.*

Proof : Since we are feasible the range space component of the search direction will vanish, and $\delta\mathbf{x} = \mathbf{u}$. If we premultiply (5a) by $\mathbf{u}$ and use (10) and Assumption (A3) we obtain

$$\begin{aligned} -\phi_2(\mathbf{x}, \hat{\mathbf{y}}; \bar{\sigma}, \rho) &= \langle \mathbf{u}, \boldsymbol{\nabla} f(\mathbf{x}; \rho)\rangle = -\langle \mathbf{u}, \mathbf{N}\rangle \mathbf{u} \\ &\leq -\beta_2 \|\mathbf{u}\|^2 = -\beta_2 \|\delta\mathbf{x}\|^2 \end{aligned}$$

which proves the lemma. □

The next corollary summarizes the above results.

Corollary 8.1
Let Assumptions (A1), (A3) be satisfied. If $\delta\mathbf{w}^{(k)}$ is calculated by (PRTLNS) *and $\sigma^{(k+1)}$ is chosen as in Algorithm 2, then Φ_2 decreases along $\delta\mathbf{x}^{(k)}$ and moreover*

$$-\phi_2(\mathbf{x}^{(k)}; \sigma^{(k+1)}, \rho) \leq -\beta_9 \|\delta\mathbf{x}^{(k)}\|^2,$$

where $\beta_9 = \min\left\{r^{(k)}, \beta_2, \beta_3\right\}$, β_3 is defined in (50)*.*

The monotonic decrease of the l_2 merit function has been established in [2], and we shall not replicate it here. We only mention the next result from this paper, without proof, because it will be used in Theorem (2).

Lemma 12 *Let Assumptions (A1)–(A3) hold. Let $\sigma^{(k+1)}$ be sufficiently large and assume that there exists $k_0 \geq 0$, such that for $k \geq k_0$ the set*

$$\mathcal{F}_2 = \left\{\mathbf{x} > 0 \,\middle|\, \Phi_2(\mathbf{x}; \sigma^*, \rho) \leq \Phi_2(\mathbf{x}^{(k_0)}; \sigma^*, \rho)\right\} \tag{51}$$

be compact. In addition, if there exists an integer $k_2 \geq 0$, such that for $k \geq k_2$ the algorithm chooses $t_1 = 2$, then for all $k \geq \max\{k_0, k_2\}$ we have that

$$\lim_{k\to\infty} \|\delta\mathbf{x}^{(k)}\| = 0.$$

8.3 Properties of the switches

The algorithm can guarantee the monotonic decrease of a merit function, for problems for which after a certain iteration, the value of t_1 does not change. For problems for which this behaviour does not occur, the addition of (22b) to (22a) guarantees that Φ_2 is monotonically decreasing.

The following result states that the primal variables are bounded away from zero[2].

Corollary 8.2 *The sequence* $\{\mathbf{x}^{(k)}\}$ *of primal variables generated by Algorithm 1, with* ρ *fixed, is bounded away from zero.*

Proof : Assume to the contrary that $\{\mathbf{x}^{(k)}\} \to \infty$. Then $\{B(\mathbf{x}^{(k)})\} \to \infty$ and sequence $\{\Phi_2(\mathbf{x}^{(k)}; \sigma^*, \rho)\} \to \infty$, which contradicts the monotonic decrease of Φ_2. □

In the next lemma we show that the lower bounds of the box constraints (Algorithm 3) are bounded above and away from zero.

Lemma 13 *Assume that* ρ *is fixed and that the primal variables are bounded above and away from zero. Then the lower bounds* $[LB^{(k)}]_i$ *and upper bounds* $[UB^{(k)}]_i$, $i = 1, \ldots, n$ *of the box constraints are bounded away from zero and bounded from above, respectively.*

Proof : The proof can be found in [28]. □

The following lemma is necessary to establish boundedness of the sequence $\{\delta\mathbf{x}^{(k)}, \mathbf{y}^{(k)} + \delta\mathbf{y}^{(k)}, \delta\mathbf{z}^{(k)}\}$.

Lemma 14
Let $\{\mathbf{w}^{(k)}\}$ *be a sequence of vectors generated by Algorithm 1 for a fixed value of* ρ. *Then the matrix sequence* $\{\mathbf{R}^{(k)^{-1}}\}$ *is bounded, where* $\mathbf{R}$ *is the matrix of the reduced system* (5)

$$\mathbf{R} = \begin{pmatrix} \mathbf{0} & \nabla\mathbf{h}(\mathbf{x})^T \\ -\nabla\mathbf{h}(\mathbf{x}) & \mathbf{N} \end{pmatrix}.$$

Proof : The proof is taken from [2]. The inverse of the partitioned matrix, suppressing iterates, is

$$\mathbf{R}^{-1} = \begin{pmatrix} \mathbf{Q} & -\mathbf{Q}\nabla\mathbf{h}(\mathbf{x})^T\mathbf{N}^{-1} \\ \mathbf{N}^{-1}\nabla\mathbf{h}(\mathbf{x})\mathbf{Q} & \mathbf{N}^{-1} - \mathbf{N}^{-1}\nabla\mathbf{h}(\mathbf{x})\mathbf{Q}\nabla\mathbf{h}(\mathbf{x})^T\mathbf{N}^{-1} \end{pmatrix}$$

where $\mathbf{Q}^{(k)} = \left(\nabla\mathbf{h}(\mathbf{x}^{(k)})^T\mathbf{N}^{(k)^{-1}}\nabla\mathbf{h}(\mathbf{x}^{(k)})\right)^{-1}$. From Assumption (A7), Corollary (8.2) and Lemma (13) we have that matrices $\mathbf{N}^{(k)^{-1}}$ and $\mathbf{Q}^{(k)}$ exist and are bounded. Hence matrix $\mathbf{R}^{(k)^{-1}}$ is bounded, since all the involved matrices are bounded. □

[2]We do not suppress iterate subscripts in most of the proofs of this section

Lemma 15 *Let* $\{\mathbf{w}^{(k)}\}$ *be a sequence of vectors generated by Algorithm 1 for* ρ *fixed. Then the sequence of vectors* $\{\delta\mathbf{x}^{(k)}, \mathbf{y}^{(k)} + \delta\mathbf{y}^{(k)}, \delta\mathbf{z}^{(k)}\}$ *is bounded.*

Proof : The reduced system (5) can be written in matrix form as

$$\begin{pmatrix} \mathbf{0} & \boldsymbol{\nabla}\mathbf{h}(\mathbf{x})^T \\ -\boldsymbol{\nabla}\mathbf{h}(\mathbf{x}) & \mathbf{N} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{y}} \\ \delta\mathbf{x} \end{pmatrix} = -\begin{pmatrix} \mathbf{h}(\mathbf{x}) \\ \boldsymbol{\nabla} f(\mathbf{x};\rho) \end{pmatrix}. \tag{52}$$

From Lemma (14) we have that the inverse of the reduced KKT conditions exists and is bounded. Therefore the sequences $\{\delta\mathbf{x}^{(k)}\}$ and $\{\hat{\mathbf{y}}^{(k)}\}$ are also bounded. By considering Eq. (4c) we may deduce that sequence $\{\delta\mathbf{z}^{(k)}\}$ is also bounded, as a sum of bounded sequences. □

We are now in position to show finiteness of the penalty parameter.

Lemma 16 (Finiteness of the penalty parameter) *Let Assumption (A1) hold. Also assume that* $\{\mathbf{x}^{(k)}\}$ *is bounded. Then the penalty parameter is increased finitely often, that is there exists an integer* $k^* \geq 0$ *such that for all* $k \geq k^*$ *we have that* $\sigma^{(k)} \in [0, +\infty)$.

Proof : We shall proceed by contradiction. Assume that $\sigma^{(k)} \to \infty$ as $k \to \infty$. By the boundedness of sequence $\{\mathbf{y}^{(k)}+\delta\mathbf{y}^{(k)}\}$, this can only happen in Step 15. The penalty parameter in that step becomes unbounded, only if $\|\mathbf{h}(\mathbf{x}^{(k)})\| \to 0$, that is if there exists an integer k_h such that, for all $k \geq k_h$ there holds

$$0 < \|\mathbf{h}(\mathbf{x}^{(k)})\| \leq \epsilon_h.$$

But if this is the case, then the penalty parameter is not updated (Step 9). Therefore the maximum value $\sigma^{(k)}$ can take is

$$\sigma^* = \max\left\{\frac{\tilde{\mathcal{D}}}{\epsilon_h}, \tilde{y}\right\}$$

where

$$\tilde{y} \geq \|\mathbf{y}^{(k)} + \delta\mathbf{y}^{(k)}\|$$

and $\tilde{\mathcal{D}}$ is a finite value[3]. Thence $\sigma^* < \infty$, which contradicts our assumption that $\sigma^{(k)} \to \infty$ as $k \to \infty$. Therefore the conclusion of the lemma holds. □

In order to prove the next lemma, we need to assume that the level sets $\mathcal{F}_1, \mathcal{F}_2$ introduced by (43), (51) are compact for sufficiently large k (for all

[3]the bound of $\vartheta^{(k)}$ has been factored in $\tilde{y}$

$k \geq k_0$, as introduced in Lemmas (6), (12)). For the case $t_1 = 3$ the reader is referred to El-Bakry et al. [6].

Lemma 17
Let Assumptions (A1)–(A3) hold and let the barrier parameter be fixed. Also assume that for sufficiently large k, the level sets of the activated merit functions $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ are compact and that $\{\mathbf{x}^{(k)}\}$ is bounded. Then

$$\lim_{k\to\infty} \|\delta\mathbf{x}^{(k)}\| = 0.$$

Proof : At each iteration, the algorithm generates $(\delta\mathbf{x}^{(k)}, \delta\mathbf{y}^{(k)}, \delta\mathbf{z}^{(k)})$ which is a KKT point of (PRTLNS). If

$$\delta\mathbf{x}^{(k)} = \mathbf{0},$$

then the lemma is proven.

Assume that there exists $k_{t_1} \geq 0$, such that for all $k \geq k_{t_1}$ the algorithm chooses the same value for $t_1 \in \{1, 2, 3\}$. In such a case, Lemmas (6), (12) prove the required result for $t_1 \in \{1, 2\}$. For $t_1 = 3$, a similar a similar result can be deduced.

Assume now that no such k_t exists. That is the algorithm activates different merit functions from iteration to iteration. Also assume, for contradiction, that $\delta\mathbf{x}^{(k)} \neq \mathbf{0}$. That there exists an accumulation point, say $\mathbf{x}^*$, of sequence $\{\mathbf{x}^{(k)}\}$ is guaranteed by the fact that Φ_2 is monotonically decreasing which ensures that $\mathbf{x}^{(k)} \in \mathcal{F}_2$, $\mathcal{F}_2$ compact. Also by Assumption (A2), Corollary (8.2) and Lemmas 13, 15 we have that $\mathbf{N}^*$ is a point of accumulation of $\{\mathbf{N}^{(k)}\}$. Without loss of generality, we may assume that

$$\mathbf{x}^{(k)} \to \mathbf{x}^* \quad \text{and} \quad \mathbf{N}^{(k)} \to \mathbf{N}^*.$$

Assume that $(\delta\mathbf{x}^*, \delta\mathbf{y}^*, \delta\mathbf{z}^*)$ is a KKT point of (PRTLNS). If $\delta\mathbf{x}^* = \mathbf{0}$ then the lemma is proven. Suppose to the contrary, that $\delta\mathbf{x}^* \neq \mathbf{0}$. Let α_x^* be chosen so that $SA(\alpha_x^*, \mathbf{x}^*, \delta\mathbf{x}^*, t_1)$ is satisfied. From (22b) we have that α_x^* satisfies

$$\Phi_2(\mathbf{x}^* + \alpha_x^*\delta\mathbf{x}^*; \sigma^*, \rho) < \Phi_2(\mathbf{x}^*; \sigma^*, \rho).$$

Since

$$\mathbf{x}^{(k)} + \alpha^*\delta\mathbf{x}^{(k)} \to \mathbf{x}^* + \alpha^*\delta\mathbf{x}^*,$$

it follows that for sufficiently large k, we have

$$\Phi_2(\mathbf{x}^{(k)} + \alpha^*\delta\mathbf{x}^{(k)}; \sigma^*, \rho) < \Phi_2(\mathbf{x}^*; \sigma^*, \rho). \tag{53}$$

But then, from the monotonic decrease of Φ_2, we have that

$$\begin{aligned}\Phi_2(\mathbf{x}^*;\sigma^*,\rho) &< \Phi_2(\mathbf{x}^{(k+1)};\sigma^*,\rho)\\ &\leq \Phi_2(\mathbf{x}^{(k)}+\alpha^*\delta\mathbf{x}^{(k)};\sigma^*,\rho)\end{aligned}$$

which contradicts (53). Hence $\delta\mathbf{x}^{(k)} \to \mathbf{0}$ for sufficiently large k. □

Theorem 2
Let the assumptions of the previous lemma hold. Then for ρ fixed, the algorithm converges asymptotically to a point satisfying the KKT conditions (3) *of problem (BNLP).*

Proof : The proof is divided into three parts. In the first part we show that the dual step sizes converge to unity. In the second part we show that the complementarity condition (4c) is satisfied and finally in the third and last part we show that (4a) is satisfied.

Let $\mathbf{x}^*(\rho), \mathbf{y}^*(\rho), \mathbf{z}^*(\rho)$ be such that

$$\lim_{k\to\infty}\mathbf{x}^{(k)} = \mathbf{x}^*(\rho), \quad \lim_{k\to\infty}\mathbf{y}^{(k)} = \mathbf{y}^*(\rho), \quad \lim_{k\to\infty}\mathbf{z}^{(k)} = \mathbf{z}^*(\rho)$$

for all $k_* \leq k \in \mathcal{K} \subseteq \{1,2,\ldots\}$. The existence of such points is ensured since by Lemmas 13, 15 the sequence $\{\mathbf{x}^{(k)}(\rho), \mathbf{y}^{(k)}(\rho), \mathbf{z}^{(k)}(\rho)\}$ is bounded for ρ fixed and Φ_2 decreases at each iteration, thereby ensuring that $\mathbf{x}^{(k)} \in \mathcal{F}_2$, with $\mathcal{F}_2$ compact.

In this first part of this proof we show that the dual step size $\alpha_z^{(k)} \to 1$, by showing that

$$\lim_{k\to\infty}\|\mathbf{z}^{(k)}+\delta\mathbf{z}^{(k)}-\rho{\mathbf{X}^{(k+1)}}^{-1}\mathbf{e}\| = 0. \tag{54}$$

We begin by adding $-\rho{\mathbf{X}^{(k+1)}}^{-1}\mathbf{e}$ to both sides of (4c), therefore obtaining

$$\begin{aligned}\|\mathbf{z}^{(k)}+\delta\mathbf{z}^{(k)}-\rho{\mathbf{X}^{(k+1)}}^{-1}\mathbf{e}\| &\leq \|-{\mathbf{X}^{(k)}}^{-1}\mathbf{Z}^{(k)}\|\|\delta\mathbf{x}^{(k)}\| + \\ &\quad + \rho\|{\mathbf{X}^{(k)}}^{-1}-{\mathbf{X}^{(k+1)}}^{-1}\|\|\mathbf{e}\|. \end{aligned}\tag{55}$$

We also have that

$$\begin{aligned}\|{\mathbf{X}^{(k)}}^{-1}-{\mathbf{X}^{(k+1)}}^{-1}\|^2 &\leq n\max_{1\leq i\leq n}\left\{\left(\frac{1}{[x^{(k)}]_i}-\frac{1}{[x^{(k+1)}]_i}\right)^2\right\}\\ &= n\max_{1\leq i\leq n}\left\{\frac{(\alpha_x^{(k)})^2([\delta x^{(k)}]_i)^2}{([x^{(k)}]_i)^2([x^{(k+1)}]_i)^2}\right\}.\end{aligned}$$

Since the sequence $\{\mathbf{x}^{(k)}\}$ is bounded away from zero, $\alpha_x^{(k)} \in (0,1]$ and $([\delta x^{(k)}]_i)^2 \leq \|\delta \mathbf{x}^{(k)}\|^2$ the last inequality can be written as

$$\lim_{k\to\infty} \|{\mathbf{X}^{(k)}}^{-1} - {\mathbf{X}^{(k+1)}}^{-1}\|^2 \leq n \lim_{k\to\infty} \max_{1\leq i\leq n} \left\{ \frac{\|\delta \mathbf{x}^{(k)}\|^2}{([x^{(k)}]_i)^2([x^{(k+1)}]_i)^2} \right\}. \quad (56)$$

Therefore for k sufficiently large, from Lemma (17) and the above, we have from (55) that (54) holds, which proves that $\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \delta \mathbf{z}^{(k)}$ for k sufficiently large.

In the second part of the proof we show that the complementarity condition (3c) is satisfied asymptotically. For sufficiently large k, the complementarity condition becomes using (4c)

$$\mathbf{X}^{(k+1)}\mathbf{z}^{(k+1)} = \mathbf{X}^{(k+1)}(\mathbf{z}^{(k)}+\delta\mathbf{z}^{(k)}) = \mathbf{X}^{(k+1)}{\mathbf{X}^{(k)}}^{-1}(-\mathbf{Z}^{(k)}\delta\mathbf{x}^{(k)}+\rho\mathbf{e}). \quad (57)$$

If we write the elements of the diagonal matrix $\mathbf{X}^{(k+1)}{\mathbf{X}^{(k)}}^{-1}$ as

$$\frac{[x^{(k+1)}]_i}{[x^{(k)}]_i} = 1 + \alpha_x^{(k)} \frac{[\delta x^{(k)}]_i}{[x^{(k)}]_i}, \qquad i = 1, \dots, n$$

and use Lemma (17) we deduce that

$$\lim_{k\to\infty} \mathbf{X}^{(k+1)}{\mathbf{X}^{(k)}}^{-1} = \mathbf{I}_n. \quad (58)$$

Therefore for sufficiently large k, Eq. (57) yields using Lemma (17), Eq. (58)

$$\mathbf{X}^{(k+1)}\mathbf{z}^{(k+1)} = \mathbf{X}^*(\rho)\mathbf{z}^*(\rho) = \rho\mathbf{e}, \quad (59)$$

which shows satisfaction of the complementarity condition (3c).

Again, for k sufficiently large, using Lemma (17) we have from Eq. (5b) that

$$\lim_{k\to\infty} \boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^{(k)})^T \delta\mathbf{x}^{(k)} = \mathbf{h}(\mathbf{x}^*(\rho)) = \mathbf{0}, \quad (60)$$

which proves satisfaction of the equality constraints Eq. (3b).

In the last part of this proof, we show satisfaction of Eq. (3a), to finish the proof. From the first equation (5a) of the reduced system we have that

$$\boldsymbol{\nabla} f(\mathbf{x}^{(k)};\rho) - \boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^{(k)})\hat{\mathbf{y}}^{(k)} = -\mathbf{N}^{(k)}\delta\mathbf{x}^{(k)}.$$

For sufficiently large k the above becomes using Lemma (17)

$$\lim_{k\to\infty} \|\boldsymbol{\nabla} f(\mathbf{x}^{(k)};\rho) - \boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^{(k)})\mathbf{y}^{(k+1)}\| = 0. \quad (61)$$

From Assumptions (A1), (A5) the last equation yields, using again Lemma (17)

$$\lim_{k\to\infty} \|\nabla f(\mathbf{x}^{(k+1)};\rho) - \nabla \mathbf{h}(\mathbf{x}^{(k+1)})\mathbf{y}^{(k+1)}\| = 0, \tag{62}$$

or equivalently that

$$\nabla f(\mathbf{x}^*(\rho);\rho) - \nabla \mathbf{h}(\mathbf{x}^*(\rho))\mathbf{y}^*(\rho) = \mathbf{0}.$$

Therefore $(\mathbf{x}^*(\rho), \mathbf{y}^*(\rho), \mathbf{z}^*(\rho))$ is a solution of the perturbed KKT conditions (3). □

The proofs so far have showed convergence of the iterates to approximate central points, *i.e.* to points that satisfy the KKT conditions of problem (BNLP). In the remaining part we show that the sequence of approximate central points converges indeed to a KKT point $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)$ of the original problem (NLPIP).

For a given $\epsilon > 0$ sufficiently small, consider the set of all approximate central points generated by the algorithm

$$\mathcal{F}_\epsilon = \left\{\mathbf{w} \,\middle|\, \epsilon \leq \|\mathbf{F}(\mathbf{w};\rho)\| \leq \|\mathbf{F}(\mathbf{w}^{(0)};\rho^{(0)})\|, \quad \rho < \rho^{(0)}\right\},$$

where $\mathbf{F}(\mathbf{w};\rho)$ are the KKT conditions of the logarithmic barrier problem (BNLP) given by (PRTKKT). The line search rules described in Section 5 guarantee that $\mathbf{x}^{(k)}, \mathbf{z}^{(k)} \in \mathcal{F}_\epsilon$ are bounded away from zero for $k \geq 0$, $\epsilon > 0$. This in turn means that $\langle \mathbf{x}^{(k)}, \mathbf{z}^{(k)} \rangle$ is bounded away from zero in $\mathcal{F}_\epsilon$. In the following lemma we prove boundedness of $\mathbf{y}^{(k)}$.

Lemma 18
If Assumptions (A1), (A5) hold and the primal iterates $\mathbf{x}^{(k)}$ are in a compact set for $k \geq 0$, then there exists a constant $c_4 > 0$ such that

$$\|\mathbf{y}^{(k)}\| \leq c_4(1 + \|\mathbf{z}^{(k)}\|).$$

Proof : The proof is patterned after [2]. By defining $\mathbf{b}^{(k)} = \nabla f(\mathbf{x}^{(k)}) - \mathbf{z}^{(k)} - \nabla \mathbf{h}(\mathbf{x}^{(k)})\mathbf{y}^{(k)}$ and solving for $\nabla \mathbf{h}(\mathbf{x}^{(k)})\mathbf{y}^{(k)}$ we obtain

$$\nabla \mathbf{h}(\mathbf{x}^{(k)})\mathbf{y}^{(k)} = \nabla f(\mathbf{x}^{(k)}) - \mathbf{z}^{(k)} - \mathbf{b}^{(k)}.$$

From our assumption the above equations can be written as

$$\begin{aligned}\mathbf{y}^{(k)} &= \left(\nabla \mathbf{h}(\mathbf{x}^{(k)})^T \nabla \mathbf{h}(\mathbf{x}^{(k)})\right)^{-1} \nabla \mathbf{h}(\mathbf{x}^{(k)})^T \left(\nabla f(\mathbf{x}^{(k)}) - \mathbf{z}^{(k)} - \mathbf{b}^{(k)}\right) \\ &\quad - \left(\nabla \mathbf{h}(\mathbf{x}^{(k)})^T \nabla \mathbf{h}(\mathbf{x}^{(k)})\right)^{-1} \nabla \mathbf{h}(\mathbf{x}^{(k)})^T \mathbf{z}^{(k)}.\end{aligned}$$

If we set

$$
\begin{aligned}
b_1 &= \|\left(\nabla \mathbf{h}(\mathbf{x}^{(k)})^T \nabla \mathbf{h}(\mathbf{x}^{(k)})\right)^{-1} \nabla \mathbf{h}(\mathbf{x}^{(k)})\| && \text{(63a)}\\
b_2 &= b_1 \|\nabla f(\mathbf{x}^{(k)}) - \mathbf{z}^{(k)} - \mathbf{b}^{(k)}\| && \text{(63b)}
\end{aligned}
$$

and take norms in both sides of the last equation we obtain

$$
\begin{aligned}
\|\mathbf{y}^{(k)}\| &\leq b_2 \|\nabla f(\mathbf{x}^{(k)}) - \mathbf{z}^{(k)} - \mathbf{b}^{(k)}\| + b_1 \|\mathbf{z}^{(k)}\| \\
&\leq c_4 (1 + \|\mathbf{z}^{(k)}\|), && \text{(64)}
\end{aligned}
$$

where the constant c_4 is defined as

$$c_4 \geq \max \{b_2, b_1\},$$

which is finite, according to our assumptions. □

Lemma 19
Let Assumptions (A1), (A5) hold. Also assume that $\{\mathbf{x}^{(k)}\}$ is bounded. If $(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}, \mathbf{z}^{(k)}) \in \mathcal{F}_\epsilon$ for all $k \geq 0$, then the sequence $\{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}, \mathbf{z}^{(k)})\}$ is bounded above.

Proof : The proof is taken from [2]. From Lemma (18), it suffices to prove that the sequence $\{\mathbf{z}^{(k)}\}$ is bounded from above. Assume that there exists a non-empty set I_z^∞. From the boundedness of the sequences $\{[x^{(k)}]_i [z^{(k)}]_i\}$, $i = 1, \ldots, n$, we obtain

$$\liminf_{k \to \infty} [x^{(k)}]_i = 0, \qquad i \in I_z^\infty.$$

Furthermore from the definition of I_x^0, it is evident that $I_z^\infty \subseteq I_x^0$.

From (24) and the fact that the barrier parameter goes to zero, we have that the sequence

$$\{\|\nabla f(\mathbf{x}^{(k)}) - \mathbf{z}^{(k)} - \nabla \mathbf{h}(\mathbf{x}^{(k)}) \mathbf{y}^{(k)}\|\}$$

is bounded. Using this and the fact that $\{\|\nabla f(\mathbf{x}^{(k)})\|\}$ is bounded, we conclude that $\{\| - \mathbf{z}^{(k)} - \nabla \mathbf{h}(\mathbf{x}^{(k)}) \mathbf{y}^{(k)}\|\}$ is also bounded. Hence we have that

$$\frac{\|\mathbf{z}^{(k)} + \nabla \mathbf{h}(\mathbf{x}^{(k)}) \mathbf{y}^{(k)}\|}{\|(\mathbf{y}^{(k)}, \mathbf{z}^{(k)})\|} \to 0. \tag{65}$$

If we set $\mathbf{u}^{(k)} = \frac{(\mathbf{y}^{(k)}, \mathbf{z}^{(k)})}{\|(\mathbf{y}^{(k)}, \mathbf{z}^{(k)})\|}$, we have that $\{\mathbf{u}^{(k)}\}$ is bounded and that $\{\mathbf{u}^{(k)}\} \to \mathbf{u}^*$. It is clear that $\|\mathbf{u}^*\| = 1$ and that $[u^*]_i$, $i \notin I_z^\infty$, i.e. $\{[z^{(k)}]_i\}$, are zero. If $\hat{\mathbf{u}}^*$ is the vector consisting of the components of $[u^*]_i$, $i \in I_z^\infty$, then $\|\hat{\mathbf{u}}^*\| = \|\mathbf{u}^*\| = 1$. Furthermore, from (65) we have

$$\begin{aligned} \frac{\mathbf{z}^{(k)} + \boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^{(k)})\mathbf{y}^{(k)}}{\|(\mathbf{y}^{(k)}, \mathbf{z}^{(k)})\|} &= \frac{[\boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^{(k)}), \mathbf{I}_m](\mathbf{y}^{(k)}, \mathbf{z}^{(k)})}{\|(\mathbf{y}^{(k)}, \mathbf{z}^{(k)})\|} \\ &= [\boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^{(k)}), \mathbf{e}_i : i \in I_x^0]\hat{\mathbf{u}}^* \to 0. \end{aligned}$$

However, this result contradicts Assumption (A5). Hence I_z^∞ is empty, that is to say, for all indices $i = 1, \ldots, p$, elements of $\{[z^{(k)}]_i\}$ are bounded. Consequently, $\{\mathbf{z}^{(k)}\}$ is also bounded. □

The following theorem shows that the sequence of approximate central path points converge to a KKT point of problem (NLPIP).

Theorem 3 *Let $\{\rho^{(l)}\}$ be a monotonically decreasing sequence converging to zero. Also let $(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}, \mathbf{z}^{(k)})$ be a sequence of approximate central points satisfying*

$$\|\mathbf{F}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}, \mathbf{z}^{(k)}; \rho^{(l)})\| \leq \eta\rho^{(l)}. \tag{66}$$

If $\{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}, \mathbf{z}^{(k)})\} \to (\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)$, then the limit point satisfies

$$\mathbf{F}(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*; 0) = \begin{pmatrix} \boldsymbol{\nabla} f(\mathbf{x}^*) - \mathbf{z}^* - \boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^*)\mathbf{y}^* \\ \mathbf{h}(\mathbf{x}^*) \\ \mathbf{X}^*\mathbf{Z}^*\mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}.$$

Proof : From Lemma (18) the sequence $\{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}, \mathbf{z}^{(k)})\}$ is bounded and remains in the compact set $\mathcal{F}_\epsilon$. Therefore the existence of its limit point $\{(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)\}$ is guaranteed. Since $\rho^{(l)} \to 0$ and from (66) we have that

$$\lim_{k\to\infty} \|\mathbf{F}(\mathbf{x}^{(k)}, \mathbf{y}^{(k)}, \mathbf{z}^{(k)})\| = 0.$$

Therefore

$$\begin{aligned} \boldsymbol{\nabla} f(\mathbf{x}^*) - \mathbf{z}^* - \boldsymbol{\nabla}\mathbf{h}(\mathbf{x}^*)\mathbf{y}^* &= \mathbf{0} \\ \mathbf{h}(\mathbf{x}^*) &= \mathbf{0} \\ \mathbf{X}^*\mathbf{Z}^*\mathbf{e} &= \mathbf{0} \end{aligned}$$

which is the required result. □

9 Local convergence

Local convergence results for primal-dual interior point methods have been given by Zhang and Tapia [31], Zhang et al. [32] and Zhang et al. [33] for linear and quadratic programs. Mizuno et al. [19] have given similar results for a predictor-corrector algorithm. Local convergence results for general nonlinear programming problems have been given by McCormick [17, 18], El-Bakry et al. [6] and Yamashita and Yabe [29].

In this section we present local convergence properties of the proposed algorithm. According to Powell [23], Han [12] and Bertsekas [3] it is necessary to show that the Newton direction is accepted untruncated by the line search procedure, in order to ensure superlinear convergence. Establishing such a property is in general difficult for non-differentiable merit functions.

We have isolated the use of the line search procedure activated when $t_1 = 1$ in order to demonstrate the overall convergence of the primal stepsizes to unity. Local convergence properties of the l_2 merit function have been given by Akrotirianakis and Rustem [1]; similarly for the euclidean norm of the perturbed KKT constraints, local convergence properties are displayed by El-Bakry et al. [6]. In the following result we show the attainment of unity stepsizes by the l_∞ merit function.

Theorem 4 (Steplength convergence)
Suppose that the sequence $\{\mathbf{x}^{(k)}\}$ converges to a KKT point. Then for the line search procedure employing the l_∞ merit function ($t_1 = 1$), we have that for sufficiently large $k \geq k_0$

$$\{\alpha^{(k)}\} \rightarrow 1.$$

Proof : Following the argument of Lemma (5) we have that the largest value the line search parameter can take is

$$\tilde{\alpha}^{(k)} = \min\left\{1, \frac{1 - c_1}{\beta_5 \beta_7 P_1(\mathbf{x}^{(k)})(\psi\beta_1 + \eta^{(k)} + \sigma^{(k+1)}\gamma)}\right\}.$$

The terms in parentheses multiplied by $P_1(\mathbf{x}^{(k)})$ are bounded. Also for sufficiently large k we have that $P_1(\mathbf{x}^{(k)}) \rightarrow 0$ and therefore for sufficiently large k we obtain $\tilde{\alpha}^{(k)} = 1$.

Since $\alpha^{(k)}$ is obtained by reducing the maximum allowable step length $\tilde{\alpha}^{(k)}$ until Armijo's condition is satisfied, it follows that Armijo's condition will be satisfied immediately without truncations. □

10 Numerical results for PDIP

For the numerical testing of the PDIP method we implemented the algorithm using `ANSI C`. The problems solved are the ones found in the collection of Hock and Schittkowski [13]. The majority of the problems were coded using `ANSI C` and interfaced with the solver. The algorithm was also interfaced with the powerful mathematical programming language `AMPL` [8, 7]. Problems of the Hock and Schittkowski collection that were hard to code were solved using `AMPL`.

There are two options to calculate the Hessian of the Lagrangian. The first option involves the use of the *damped BFGS* formula introduced by Powell [22]. In the second option the exact Hessian is calculated, with second derivatives calculated either programmatically or from `AMPL`. In the first case the Hessian matrix $\mathbf{N}^{(k)}$ is always positive definite. In the second case though such property is not ensured, and the primal search direction $\delta\mathbf{x}^{(k)}$ is not guaranteed to decrease the merit function. In order to prevent this behaviour we modify the Hessian of the Lagrangian by replacing it with a positive definite matrix $\mathbf{G}^{(k)}$. This matrix is generated using the *indefinite Cholesky factorization* ($\mathbf{LDL}^T$) described in Gill et al. [10] and also discussed in Dennis, Jr. and Schnabel [5] for the unconstrained minimization case.

The algorithmic parameters are as follows. The initial value for the barrier parameter is $\rho^{(0)} = 1.0$. For the penalty parameter we set $\sigma^{(0)} = 0.0$. In the switches we set $\epsilon_h = 10^{-6}$, $\epsilon_d = 10^{-4}$ and $\delta = 10.0$. In the line search procedure $\theta = 0.5$, $c = 10^{-4}$ and $\zeta = 0.9$. In the barrier parameter choice $\eta = 10^4$ and $q = 3.0$. Finally, in the stopping criterion of the outer iterations we used $\epsilon_0 = 10^{-8}$.

Table (2) summarizes the numerical results for the test problems found in the Hock and Schittkowski collection. Table (3) summarizes results obtained without imposing the monotonic decrease condition (22b) for Φ_2. We only impose the sufficient decrease criterion of the activated merit function (22a) (Section 5.1).

In these tables the following abbreviations are used:

- No is the name of the problem.
- k is the number of iterations to find the optimum solution.
- σ^* is the final value of the penalty parameter.
- k^* is the iteration after which the search direction decreases the merit function without truncations.

- Φ_1 shows the number of times the l_∞ merit function was used in the problem.
- Φ_2 shows the number of times the l_2 merit function was used in the problem.
- Φ_3 shows the number of times the Euclidean norm of the perturbed KKT conditions was used in the problem.

Table 2: Hock & Schittkowski results for monotonic PDIP

No	k	σ^*	k^*	Φ_1	Φ_2	Φ_3
1	19	0	4	0	19	0
2	14	0	0	0	14	0
3	10	0	0	0	10	0
4	10	0	0	0	10	0
5	11	0	0	0	5	6
6	4	10	0	3	1	0
7	90	10.2887	82	89	1	0
8	3	10	0	3	0	0
9	4	0	0	0	3	1
10	16	10.5	15	15	1	0
11	13	42.9593	0	3	10	0
12	15	10.5	14	14	1	0
14	11	13.4412	10	8	3	0
15	20	86,026.8	9	10	10	0
16	16	11.8813	0	2	14	0
17	15	16.9823	0	1	14	0
18	10	10.1972	0	2	8	0
19	18	111,632	0	8	8	2
20	15	268.736	0	4	11	0
21	10	0	0	0	10	0
22	11	11.3389	0	5	6	0
23	13	0	0	0	13	0

Continued on next page

Table 2 – continued from previous page

No	k	σ^*	k^*	Φ_1	Φ_2	Φ_3
24	15	0	0	0	15	0
25	35	1	27	0	10	25
26	26	10.8882	5	9	17	0
27	23	18.571	20	20	1	2
28	2	0	0	0	1	1
29	38	10.8485	0	9	29	0
30	12	0	0	0	12	0
31	12	0	0	0	12	0
32	50	0	0	0	50	0
33	15	0	0	0	15	0
34	33	10.4788	0	18	15	0
35	11	0	0	0	11	0
36	14	0	0	0	14	0
37	53	0	0	0	53	0
38	18	0	0	0	5	13
39	15	18.1385	12	14	0	1
41	65	14.2846	0	1	64	0
42	8	14.5356	7	7	1	0
43	15	13	14	14	1	0
44	19	0	10	0	19	0
45	18	0	0	0	17	1
46	20	0	0	0	19	1
47	19	13.5829	0	5	14	0
48	2	0	0	0	1	1
49	16	0	0	0	14	2
50	10	0	0	0	8	2
51	2	0	0	0	1	1
52	2	23.9312	0	1	1	0
53	9	20.4432	8	8	1	0
54	8	11.1705	6	3	2	3
55	19	0	0	0	19	0
56	68	11.7848	0	5	59	4

Continued on next page

Table 2 – continued from previous page

No	k	σ^*	k^*	Φ_1	Φ_2	Φ_3
57	17	10.8513	0	1	16	0
59	90	10.0055	85	3	87	0
60	11	0	0	0	11	0
61	10	13.943	9	9	1	0
62	10	0	0	0	1	9
63	13	11.5264	0	10	1	2
64	36	373,382	31	34	1	1
65	12	10.0818	0	4	8	0
66	33	10.8686	4	18	15	0
70	12	10.0503	5	3	6	3
71	58	10.7122	0	10	43	5
73	13	0	0	0	13	0
74	9	51.5832	8	8	1	0
75	12	2,809.09	0	9	2	1
76	11	0	0	0	11	0
77	7	0	0	0	7	0
78	73	12.4074	0	3	5	65
79	6	0	0	0	6	0
80	7	10.0833	0	4	2	1
83	15	1,634.36	0	6	9	0
84	32	10.0002	22	4	28	0
86	15	35.0904	0	4	11	0
87	25	69.9099	14	6	19	0
93	372	39,887	139	7	7	358
95	20	10.0322	3	4	16	0
96	20	10.0305	3	3	17	0
100	13	11.5091	12	11	2	0
101	55	8,344.63	46	47	8	0
102	313	3,921.79	303	310	3	0
103	18	2,275.56	3	10	8	0
104	46	22.0406	8	9	37	0
105	16	0	4	0	13	3

Continued on next page

Table 2 – continued from previous page

No	k	σ^*	k^*	Φ_1	Φ_2	Φ_3
107	14	44,272.8	0	9	2	3
108	27	51.5038	0	11	16	0
109	37	63.2423	33	26	11	0
110	10	0	0	0	10	0
111	90	49.6809	0	24	1	65
112	27	0	25	0	25	2
113	14	14.2271	0	1	13	0
114	372	15,969.2	0	98	64	210
117	31	1	23	0	31	0
118	15	0	0	0	15	0
119	16	5,370.34	0	1	15	0

Table 3: Hock & Schittkowski results for non-monotonic PDIP

No	k	σ^*	k^*	Φ_1	Φ_2	Φ_3
1	27	0	15	0	27	0
2	14	0	1	0	14	0
3	10	0	0	0	10	0
4	10	0	0	0	10	0
5	11	0	0	0	5	6
6	4	10	0	1	3	0
7	9	10.2887	2	8	0	1
8	3	10	0	1	2	0
9	1	0	0	0	0	1
10	16	10.5	0	12	1	3
11	13	13.0494	0	8	1	4
12	12	10.5	0	10	1	1
14	12	13.4412	1	6	3	3
15	20	86,026.8	9	10	10	0
16	16	11.8813	0	2	14	0
17	15	16.9823	0	1	14	0
18	10	10.1972	0	2	8	0

Continued on next page

Table 3 – continued from previous page

No	k	σ^*	k^*	Φ_1	Φ_2	Φ_3
19	18	111,632	0	8	8	2
20	15	268.736	0	4	11	0
21	10	0	0	0	10	0
22	11	11.3389	0	5	6	0
23	13	0	0	0	13	0
24	15	0	0	0	15	0
25	35	1	27	0	10	25
26	23	10.1082	0	8	15	0
27	39	10.7088	31	34	2	3
28	2	0	0	0	1	1
29	39	10.7278	0	12	27	0
30	12	0	0	0	12	0
31	12	0	0	0	12	0
32	50	0	0	0	50	0
33	15	0	0	0	15	0
34	33	10.4788	0	18	15	0
35	11	0	0	0	11	0
36	14	0	0	0	14	0
37	53	0	0	0	53	0
38	18	0	0	0	5	13
39	20	17.5716	3	13	0	7
40	19	13.8682	0	5	0	14
41	65	14.2846	0	1	64	0
42	10	14.5356	0	5	1	4
43	14	13	1	9	1	4
44	19	0	10	0	19	0
45	18	0	0	0	17	1
46	21	10.0681	1	2	19	0
47	18	16.7967	1	4	14	0
48	2	0	0	0	0	2
49	18	0	0	0	17	1
50	9	0	0	0	5	4

Continued on next page

Table 3 – continued from previous page

No	k	σ^*	k^*	Φ_1	Φ_2	Φ_3
51	2	0	0	0	1	1
52	2	80.0769	0	2	0	0
53	10	20.4432	0	5	1	4
55	19	1	0	0	19	0
56	1	0	0	0	0	1
57	15	0	13	0	15	0
59	8	0	0	0	0	8
60	11	0	0	0	11	0
61	6	13.2042	0	4	1	1
62	10	0	0	0	1	9
63	30	11.5264	0	10	1	19
64	37	373,382	31	34	1	2
65	12	10.0818	0	4	8	0
66	33	10.8686	4	18	15	0
71	58	10.7122	0	10	43	5
72	16	49,167.5	0	15	1	0
73	13	0	0	0	13	0
74	10	51.5832	0	5	1	4
75	12	2,809.09	0	9	2	1
76	11	0	0	0	11	0
77	8	0	0	0	6	2
78	73	12.4074	0	3	5	65
79	6	0	0	0	6	0
80	11	10.0833	0	4	2	5
81	17	62.312	13	8	8	1
83	15	1,634.36	0	6	9	0
84	32	10.0002	22	4	28	0
86	15	35.0904	0	4	11	0
87	25	69.9099	14	6	19	0
93	371	48,721.5	13	24	30	317
95	20	10.0322	3	4	16	0
96	20	10.0305	3	3	17	0

Continued on next page

Table 3 – continued from previous page

No	k	σ^*	k^*	Φ_1	Φ_2	Φ_3
97	31	11.3978	29	14	14	3
98	23	11.1843	0	5	17	1
100	11	11.5099	0	7	2	2
103	18	2,275.56	3	10	8	0
104	45	23.2979	0	8	37	0
105	16	1	4	0	13	3
107	9	13,644.9	0	7	1	1
108	88	13.3575	0	5	80	3
110	10	0	0	0	10	0
111	12	51.4597	0	7	1	4
112	23	1	0	0	21	2
113	14	1	0	0	14	0
114	372	15,969.4	0	98	64	210
116	48	12,780.7	22	14	33	1
117	31	1	23	0	31	0
118	15	1	0	0	15	0
119	16	5,370.34	0	1	15	0

All the numerical results were obtained by using the exact Hessian. The problems that do not appear in this table were not solved to the desired accuracy. We observe from Table (3) that in all cases the value of the penalty parameter is well behaved in all cases, with a maximum value of 373,382 (problem 064). We also observe that during the latter stages of the calculations unity stepsizes are employed for all problems.

For problem 093 in Table (2) (Table (3)), convergence is attained after 371 (372) iterations, which is quite large. What is particularly interesting with this problem, is that the generated search direction is last truncated at iteration 139 (13). The reason for such slow convergence was mainly the length of the Newton direction. The value of the maximum element of the diagonal perturbation added to the Hessian of the Lagrangian using the indefinite Cholesky factorization [10] is on average 360. A similar comment goes for problem 114 (Tables (2), (3)). For this problem our algorithm converges after 372 iterations. The search direction is never truncated and as previously the average value of the maximum element of the perturbation matrix is 200. In a relevant discussion, Gould and Leyffer [11] (also [15]) state that modifications of the Hessian of (PRTLNS) might, in the worst case, have the effect of over emphasizing one large negative eigenvalue at the expense of the remaining small, positive ones, and in producing a direction which is essentially steepest descent. We believe that the behaviour of the algorithm for these problems can be interpreted under this remark. The average value of the perturbation matrix, as decided by the modified Cholesky factorization, is relatively large which might force (PRTLNS) to generate the steepest descent direction. If the steepest descent direction is short in length then, although it is accepted from the stepsize strategy without truncation, the iterates do not escape the region in which the Hessian is indefinite, and this effect is repeated.

For problems 78, 81, 103 and 116 the initial value of the barrier parameter was set to $\rho^{(0)} = 10^2$ in order to avoid exceeding the maximum number of iterations.

We also introduced a trap in order to cater for cases in which Assumption (A5) cannot be met. Namely, if the value of the penalty parameter as obtained in Step 5 of Algorithm 2 was above ϵ_z, then the penalty parameter is updated as in Step 15 of the same algorithm. We set $\epsilon_z = 10^6$ in our numerical testing.

From Tables (2) and (3) it is observed that the two line search procedures, the monotonic and the non-monotonic, performed equally well, in terms of the number of problems solved. The non-monotonic version though performed slightly better in terms of the number of iterations needed for convergence.

11 Conclusions

A primal-dual interior point algorithm has been presented for general nonlinear programming problems. The algorithm is shown to achieve convergence from arbitrary starting points. It is also shown that in the neighbourhood of the solution unity stepsizes will be accepted, giving thus rise to fast local convergence. The adaptive penalty strategy is also such that the penalty parameter does not grow indefinitely. The innovation of the algorithm lies in the fact that it uses three different line search procedures. Each line search procedure, when used individually, has its own advantages and disadvantages. Switching rules are implemented so as to complement the merits of the line search procedures, avoiding at the same time their drawbacks. Numerical experimentation shows that the sought targets are fulfilled. In addition the value of the penalty parameter is kept very low at all times dispensing thus with the associated numerical problems. The algorithm can therefore be part of the arsenal of optimizers for solving general nonlinear programming problems. We hope that further experimentation will allow to formulate new algorithms which in the same sense achieve the same goals.

References

[1] Ioannis Akrotirianakis and Berç Rustem. A globally convergent interior point algorithm for non-linear problems. Technical Report 97/14, Department of Computing, Imperial College, London, 1997.

[2] Ioannis Akrotirianakis and Berç Rustem. A globally convergent interior point algorithm for non-linear problems. *Journal of Optimization Theory and Applications*, 125(3):497–521, 2005.

[3] Dimitri P. Bertsekas. Penalty and multiplier methods. In *Nonlinear optimization (Proc. Internat. Summer School, Univ. Bergamo, Bergamo, 1979)*, pages 253–278. Birkhäuser Boston, Mass., 1980.

[4] M. G. Breitfeld and David F. Shanno. Preliminary computational experience with modified log–barrier functions for large–scale nonlinear programming. In W. W. Hager, D. W. Hearn, and P. M. Pardalos, editors, *Large–Scale Optimization : The State–of–the–Art*, pages 45–67. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[5] John E. Dennis, Jr. and Robert B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. ISBN 0-89871-364-1. Corrected reprint of the 1983 original.

[6] A. S. El-Bakry, Richard A. Tapia, T. Tsuchiya, and Yin Zhang. On the formulation and theory of the Newton interior-point method for nonlinear programming. *Journal of Optimization Theory and Applications*, 89(3):507–541, 1996. ISSN 0022-3239.

[7] R. Fourer, David M. Gay, and B. W. Kernighan. *AMPL – A Modeling Language for Mathematical Programming.* The Scientific Press, 1993.

[8] David M. Gay. Hooking your solver to AMPL. Technical Report 97-4-06, Bell Laboratories, Murray Hill, NJ, 1997. URL `http://www.ampl.com/REFS/hooking2.ps.gz`.

[9] David M. Gay, Michael L. Overton, and Margaret H. Wright. A primal-dual interior method for nonconvex nonlinear programming. Technical Report 97-4-08, Bell Laboratories, Murray Hill, NJ, 1997. URL `http://cm.bell-labs.com/cm/cs/doc/97/4-08.ps.gz`.

[10] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Practical Optimization.* Academic Press, London, 1981. ISBN 0-12-283952-8.

[11] Nicholas I. M. Gould and Sven Leyffer. An introduction to algorithms for nonlinear optimization. Technical Report RAL-TR-2002-031, Rutherford Appleton Laboratory, Chilton, England, 2002.

[12] Shih Ping Han. Superlinearly convergent variable metric algorithms for general nonlinear programming problems. *Mathematical Programming*, 11:263–282, 1976.

[13] W. Hock and Klaus Schittkowski. Test examples for nonlinear programming codes. In *Lecture Notes in Economics and Mathematical Systems #187*. Springer-Verlag, Berlin, Heidelberg and New York, 1981.

[14] Leon S. Lasdon, John Plummer, and Gang Yu. Primal-dual and primal interior point algorithms for general nonlinear programs. *ORSA J. Comput.*, 7(3):321–332, 1995. ISSN 0899-1499.

[15] Donald W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.*, 11:431–441, 1963.

[16] D. Q. Mayne and Elijah Polak. A superlinearly convergent algorithm for constrained optimization problems. *Mathematical Programming Studies*, 16:45–61, 1982.

[17] Garth P. McCormick. The superlinear convergence of a nonlinear primal-dual algorithm. Technical Report T–550/91, School of Engineering and Applied Science, George Washington University, Washington, D.C., 1991.

[18] Garth P. McCormick. Resolving the Shell Dual with a nonlinear primal–dual algorithm. In P. M. Pardalos, editor, *Advances in Optimization and Parallel Computing*, pages 233–246. North Holland, Amsterdam, The Netherlands, 1992.

[19] S. Mizuno, M. J. Todd, and Y. Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, 18(4):964–981, 1993.

[20] James M. Ortega and Werner C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Acedemic Press, New York, 1970. Computer Science and Applied Mathematics.

[21] Elijah Polak and D. Q. Mayne. A robust secant method for optimization problems with inequality constraints. *Journal of Optimization Theory and Applications*, 33(4):463–477, 1981.

[22] Michael J. D. Powell. A fast algorithm for nonlinearly constrained optimization calculations. In G.A. Watson, editor, *Lecture notes in Mathematics, Numerical Analysis, Dundee 1977*, Dundee, Scotland, 1977. Springer-Verlag.

[23] Michael J. D. Powell. The convergence of variable metric methods for nonlinearly constrained optimization calculations. In Olvi L. Mangasarian, R. R. Meyer, and Stephen M. Robinson, editors, *Nonlinear Programming 3*, pages 27–63. Academic Press, New York, NY, 1978.

[24] Michael J. D. Powell and Y. Yuan. A recursive quadratic programming algorithm that uses differentiable penalty functions. *Mathematical Programming*, 7:265–278, 1986.

[25] Boris N. Pshenichnyi. Algorithms for general mathematical programming. *Kibernetica*, 5:120–125, 1970.

[26] Boris N. Pshenichnyi and Yu M. Danilin. *Numerical Methods in Extremal problems.* Hayka, 1975. English translation, 1978.

[27] Stephen J. Wright. *Primal-Dual Interior-Point Methods.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1997. ISBN 0-89871-382-X.

[28] Hiroshi Yamashita. A globally convergent primal-dual interior point method for constrained optimization. Technical report, Mathematical Systems Inc, 2-5-3 Shinjuku, Shinjuku-ku, Tokyo, Japan, 1995.

[29] Hiroshi Yamashita and Hiroshi Yabe. Superlinear and quadratic convergence of some primal–dual interior point methods for constrained optimization. *Mathematical Programming*, 75:377–397, 1996.

[30] Hiroshi Yamashita and Hiroshi Yabe. An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimization. *SIAM Journal on Optimization*, 14(2):479–499 (electronic), 2003. ISSN 1095-7189.

[31] Yin Zhang and Richard A. Tapia. On the superlinear and quadratic convergence of primal–dual interior point linear programming algorithms revisited. *Journal of Optimization Theory and Applications*, 73(2):229–242, 1992.

[32] Yin Zhang, Richard A. Tapia, and John E. Dennis, Jr. On the superlinear and quadratic convergence of primal–dual interior point linear programming algorithms. *SIAM Journal on Optimization*, 2(2):304–324, 1992.

[33] Yin Zhang, Richard A. Tapia, and F. Potra. On the superlinear convergence of interior point algorithms for a general class of problems. *SIAM Journal on Optimization*, 3(2):413–422, 1993.