ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
"Antonio Ruberti"
CONSIGLIO NAZIONALE DELLE RICERCHE

G. Di Pillo, G. Liuzzi, S. Lucidi

AN EXACT PENALTY-LAGRANGIAN APPROACH FOR LARGE-SCALE NONLINEAR PROGRAMMING

R. 10-03 2010

G. Di Pillo – Dipartimento di Informatica e Sistemistica "A. Ruberti", "Sapienza" Università di Roma, Via Ariosto 25 - 00185 Rome, Italy, dipillo@dis.uniroma1.it.

G. Liuzzi – Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti", Consiglio Nazionale delle Ricerche, Viale Manzoni 30 - 00185 Rome, Italy, liuzzi@iasi.cnr.it.

S. Lucidi – Dipartimento di Informatica e Sistemistica "A. Ruberti", "Sapienza" Università di Roma, Via Ariosto 25 - 00185 Rome, Italy, lucidi@dis.uniroma1.it.

This work was partially supported by MIUR PRIN National Research Program 20079PLLN7 "Nonlinear Optimization, Variational Inequalities and Equilibrium Problems".

ISSN: 1128–3378

Collana dei Rapporti
Istituto di Analisi dei Sistemi ed Informatica “Antonio Ruberti”
Consiglio Nazionale delle Ricerche

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: iasi@iasi.cnr.it
URL: http://www.iasi.cnr.it

Abstract

Nonlinear programming problems with equality constraints and bound constraints on the variables are considered. The presence of bound constraints in the definition of the problem is exploited as much as possible. To this aim an efficient search direction is defined which is able to produce a locally and superlinearly convergent algorithm and that can be computed in an efficient way by using a truncated scheme suitable for large scale problems. Then, an exact merit function is considered whose analytical expression again exploits the particular structure of the problem, by using an exact augmented Lagrangian approach for equality constraints and an exact penalty approach for the bound constraints. It is proved that the search direction and the merit function have some strong connections which can be the basis to define a globally convergent algorithm with superlinear convergence rate for the solution of the constrained problem.

1. Introduction

We are interested in the solution of smooth constrained optimization problems. In particular, we consider problems with equality constraints and upper and lower bounds on the variables, that is

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h(x) = 0 \\ & l \le x \le u \end{array} \tag{1}$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$ with $m \le n$, and $l, u \in \mathbb{R}^n$. We denote by

$$\mathcal{F} = \{x \in \mathbb{R}^n : h(x) = 0, l \le x \le u\}$$

the feasible set of Problem (1). In the paper we assume that

$$-\infty < l < u < \infty$$

and that f and h are twice continuously differentiable functions.

Remark 1. *By virtue of the assumption on the bounds l and u, the set $\{x \in \Re^n : l \le x \le u\}$ is compact. Hence, the feasible set $\mathcal{F}$ is compact as well and, if $\mathcal{F} \neq \emptyset$, the problem admits a global solution.*

Any nonlinear constrained optimization problem can be stated as an equality constrained problem with lower or upper bounds on the variables. Hence, apparently, requiring every variable to be bounded both from below and above could be too restrictive an assumption and, as a result, the problem could lose some generality. However, it is always the case in practical applications that all the variables have either explicit or hidden bounds. Problem (1) allows to model these applications, at the very most, by making explicit hidden bounds or introducing fictitious (large) ones. In addition, we remark that this model is the one adopted by most interior point methods.

We denote by $L(x, \lambda)$ the Lagrangian function of Problem (1) restricted to the equality constraints,

$$L(x, \lambda) = f(x) + \lambda^T h(x),$$

where $\lambda \in \mathbb{R}^m$ is the vector of Karush-Kuhn-Tucker (KKT) multipliers associated to the constraints $h(x) = 0$. Then, the complete Lagrangian function for Problem (1) can be written as:

$$\tilde{L}(x, \lambda, \rho, \sigma) = L(x, \lambda) + \rho^T (l - x) + \sigma^T (x - u),$$

where $\rho, \sigma \in \mathbb{R}^n$ are KKT multipliers associated, respectively, with the lower and the upper bound constraints. We denote by $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma}) \in \mathbb{R}^{3n+m}$ a KKT-tuple for Problem (1), namely a point such that $\bar{x}$ is feasible and:

$$\begin{array}{l} \nabla_x L(\bar{x}, \bar{\lambda}) - \bar{\rho} + \bar{\sigma} = 0, \\ \bar{\rho}^T (l - \bar{x}) = 0, \quad \bar{\rho} \ge 0, \\ \bar{\sigma}^T (\bar{x} - u) = 0, \quad \bar{\sigma} \ge 0. \end{array}$$

At a given point $x \in \mathbb{R}^n$ we associate the sets of indices:

$$\begin{array}{l} \mathcal{L}_0(x) = \{i : \ x_i = l_i\}, \\ \mathcal{U}_0(x) = \{i : \ x_i = u_i\}, \\ \mathcal{M}_0(x) = \{i : \ l_i < x_i < u_i\}. \end{array}$$

It is clear that if $x \in \mathcal{F}$, then $\{\mathcal{L}_0(x), \mathcal{U}_0(x), \mathcal{M}_0(x)\}$ constitutes a partition of the index set $\{1, \dots, n\}$; in particular $\mathcal{L}_0(x), \mathcal{U}_0(x)$ are the index sets of those variables, respectively, at their lower and upper bounds, while $\mathcal{M}_0(x)$ is the set of free variables, that is, those for which the bound constraints are not active.
With reference to a KKT-tuple $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$, let us also consider the following index sets:

$$\begin{array}{l} \mathcal{L}_+(\bar{x}, \bar{\rho}) = \{i \in \mathcal{L}_0(\bar{x}) : \ \bar{\rho}_i > 0\}, \\ \mathcal{U}_+(\bar{x}, \bar{\sigma}) = \{i \in \mathcal{U}_0(\bar{x}) : \ \bar{\sigma}_i > 0\}. \end{array}$$

Then, the *strict complementarity* condition holds at $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$ if $\mathcal{L}_0(\bar{x}) = \mathcal{L}_+(\bar{x}, \bar{\rho})$ and $\mathcal{U}_0(\bar{x}) = \mathcal{U}_+(\bar{x}, \bar{\sigma})$.

In recent years many algorithms have been proposed for the solution of nonlinearly constrained optimization problems, see e.g. the references [5, 6, 3, 2, 12, 10, 8, 1, 7, 14, 17, 13] for augmented Lagrangian and penalty methods, and [21, 30, 4, 11, 20, 22, 23, 29, 31] for a selection of recent papers on filter methods. Most of these methods are based on the idea of computing, by means of some local and possibly efficient algorithm, a search direction which is then used to update the current iterate. A globalization strategy is needed to promote convergence of the method to stationary points of the constrained problem and relies on the connections between the search direction itself and a suitable merit function. Furthermore, the mentioned connection should guarantee that the globalization strategy does not deteriorate the possible efficiency of the search direction. In this paper we are particularly interested in studying these theoretical connections with reference to a particular choice of a local search direction and of an exact merit function. This idea has been recently investigated in the paper [13] where an exact augmented Lagrangian function is used as a globalization strategy for the solution of large scale problems.

In this paper we exploit as much as possible the presence of bound constraints in the definition of Problem (1). To this aim we define an efficient search direction which is able to produce a locally and superlinearly convergent algorithm [5, 6, 18, 13, 26] and can be computed in an efficient way by using a truncated scheme suitable for large scale problems [13]. Then, we consider an exact merit function whose analytical expression again exploits the particular structure of Problem (1) [14, 16, 12] by taking advantage of the presence of lower and upper bounds on the variables. We prove that the search direction and the merit function have some strong connections which can be the basis to define a globally convergent algorithm to solve Problem (1).

In the paper we shall make use of the following notation. We denote by $\|\cdot\|$ the ℓ_2 (Euclidean) norm. We denote by E and I, respectively, the $n\times n$ and the $m\times m$ identity matrix.

Given a vector $v\in \mathbb{R}^p$, we indicate by $\mathrm{diag}\{v\}$ the $p\times p$ diagonal matrix with entries $v_i, 1\le i\le p$ in the main diagonal. Given a vector $y\in\mathbb{R}^p$ and a vector $v\in\mathbb{R}^p$ with components $v_i>0$, we denote $\|y\|_v^2 = y^T\mathrm{diag}\{v\}y = \sum_{i=1}^p v_i y_i^2$.

Given two vectors $v, w \in\mathbb{R}^p$

- the operation $\max\{v,w\}$ is intended component-wise, namely $\max\{v,w\}$ denotes the p-dimensional vector with components $\max\{v_i,w_i\}$, $i=1,\dots,p$;
- $v\circ w$ denotes their Hadamard product, that is the p-dimensional vector with components v_iw_i, $i=1,\dots,p$.

For short notation, the Hadamard product $w\circ w$ of a vector w by itself is denoted by w^2 and, given a positive scalar q, the vector with components $1/w_i^q$ is denoted by w^{-q}.

Given an $n\times m$ matrix Q we denote by q_i^T the i-th row of Q.

Let $\mathcal{K}\subseteq\{1,\dots,n\}$ be an index subset, $v\in\mathbb{R}^n$ be a vector and Q be a $n\times m$ matrix. We denote by $v_\mathcal{K}$ the subvector of v with components v_i such that $i\in\mathcal{K}$ and by $Q_\mathcal{K}$ the submatrix of Q made up of the rows q_i^T with $i\in\mathcal{K}$. For the transpose of $Q_\mathcal{K}$, $(Q_\mathcal{K})^T$, we adopt the short notation $Q_\mathcal{K}^T$. Moreover, let $\mathcal{H}\subseteq\{1,\dots,m\}$ be an index subset, we denote by $Q_{\mathcal{K},\mathcal{H}}$ the submatrix made up of the elements q_{ij} with $i\in\mathcal{K}$ and $j\in\mathcal{H}$. Note that we can write:

$$Q_{\mathcal{K},\mathcal{H}} = E_\mathcal{K} Q\ I_\mathcal{H}^T.$$

If $\{\mathcal{K}_1,\mathcal{K}_2\}$ constitute a partition of the index set $\{1,\dots,n\}$ we can write:

$$\begin{array}{l} v_{\mathcal{K}_i} = E_{\mathcal{K}_i} v, \quad i=1,2\\ v = E_{\mathcal{K}_1}^T v_{\mathcal{K}_1} + E_{\mathcal{K}_2}^T v_{\mathcal{K}_2}. \end{array}$$

Let $\mathcal{K}\subseteq\{1,\dots,n\}$ be an index subset, we will denote by $\bar{\mathcal{K}}$ the subset $\{1,\dots,n\}\backslash\mathcal{K}$.

In the paper we shall make use of the following definitions.

Definition 1 (Linear independence constraint qualification (LICQ)) *A point $x\in\mathcal{F}$ satisfies the LICQ if the gradients of the active constraints are linearly independent. Due to the bound constraints, this is equivalent to requiring that the matrix $\nabla h(x)_{\mathcal{M}_0(x)}$ is full rank.*

Definition 2 (Extended Mangasarian-Fromowitz constraint qualification (EMFCQ)) *Given a point $x\in\Re^n$, it satisfies the EMFCQ if there exist no $\tilde w_i$, $i\in I_u=\{i: x_i-u_i\ge 0\}$, $\hat w_i$, $i\in I_l=\{i: l_i-x_i\ge 0\}$, v_j, $j=1,\dots,m$, such that*

$$\sum_{j=1}^m v_j\nabla h_j(x) + \sum_{i\in I_u}\tilde w_i e_i - \sum_{i\in I_l}\hat w_i e_i = 0$$

with $\tilde w_i\ge 0$, $i\in I_u$, $\hat w_i\ge 0$, $i\in I_l$, and v_j, $j=1,\dots,m$, $\tilde w_i$, $i\in I_u$, $\hat w_i$, $i\in I_l$ not all zero.

Definition 3 (Strong second order sufficient condition (SSOSC)) *A KKT-tuple* $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$ *satisfies the SSOSC if the inequality*

$$y^T \nabla_x^2 L(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma}) y > 0$$

holds for all vectors $y \in \Re^n, y \neq 0$, *with*

$$y_i = 0, \quad i \in \mathcal{L}_+(\bar{x}, \bar{\rho}) \cup \mathcal{U}_+(\bar{x}, \bar{\sigma}),$$

and such that

$$\nabla h_j(\bar{x})^T y = 0, \quad j = 1, \dots, m.$$

The paper is organized as follows. In Section 2 we introduce the new exact penalty-augmented Lagrangian function for Problem (1) and state its main exactness properties. In Section 3 we propose a procedure for the computation of a truncated Newton-type direction for Problem (1) and show some interesting properties enjoyed by the direction. Section 4 is devoted to the analysis of the connections between the truncated direction and the exact merit function. Namely, for sufficiently small values of the penalty parameter and sufficiently close to a KKT-tuple, the truncated direction is a good descent direction for the merit function. In Section 5 we describe the overall exact penalty-augmented Lagrangian (`ExPAL`) solution algorithm for Problem (1) and give its main convergence properties. In Appendix A we report the proofs of a technical analytical result. In Appendix B we give an explicit expression for the directional derivative of the merit function. Other technical results are proved in Appendix C.

2. An exact penalty-Lagrangian function

In this section we give the expression of the merit function $P(x, \lambda; \varepsilon)$ that we employ for solving Problem (1) and we report its main properties. Following with minor modifications the approach described in [14] and [15], the merit function is obtained by combining an exact augmented Lagrangian approach for the equality constraints, and a continously differentiable exact penalty approach for the bound constraints, making use of suitable multiplier functions for the bound constraints. We point out that the evaluation of the multiplier functions for the bound constraints does not require any matrix inversion, as instead usually happens for the multiplier functions employed in continuously differentiable exact penalty functions for general constraints.

2.1. The merit function $P(x, \lambda; \varepsilon)$

The merit function is defined on a open relaxation of the feasible set, namely on the set

$$\mathcal{S} = \{x \in \mathbb{R}^n : \quad l - \alpha < x < u + \alpha\}$$

where α is a positive vector in $\mathbb{R}^n$.
We introduce the vector functions

$$\begin{aligned} r(x) &= \alpha - l + x, \\ s(x) &= \alpha + u - x. \end{aligned}$$

Note that in terms of the functions $r(x), s(x)$, the set $\mathcal{S}$ can be described also as $\mathcal{S} = \{x \in \mathbb{R}^n : \ r(x) > 0, \ s(x) > 0\}$.
We will make use of the following continuously differentiable multiplier functions $\rho(x, \lambda)$ and $\sigma(x, \lambda)$ associated respectively with the lower and upper bound constraints, already considered in [14, 15]:

$$\begin{aligned} \rho(x, \lambda) &= \left[(l-x)^2 + (x-u)^2\right]^{-1} \circ (x-u)^2 \circ \nabla_x L(x, \lambda) \\ \sigma(x, \lambda) &= -\left[(l-x)^2 + (x-u)^2\right]^{-1} \circ (l-x)^2 \circ \nabla_x L(x, \lambda); \end{aligned}$$

however, for convenience, we will adopt for the multiplier functions the following expressions:

$$\begin{aligned} \rho(x, \lambda) &= w(x) \circ (l-x)^{-2} \circ \nabla_x L(x, \lambda) \\ \sigma(x, \lambda) &= -w(x) \circ (x-u)^{-2} \circ \nabla_x L(x, \lambda), \end{aligned} \tag{2}$$

with

$$w(x) = (l-x)^2 \circ \left[(l-x)^2 + (x-u)^2\right]^{-1} \circ (x-u)^2.$$

It is worth noting that, if $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$ is any KKT-tuple for Problem (1), then the vector functions $\rho(x,\lambda)$ and $\sigma(x,\lambda)$ are such that:

$$\rho(\bar{x},\bar{\lambda}) = \bar{\rho}, \quad \sigma(\bar{x},\bar{\lambda}) = \bar{\sigma}.$$

Then, the continuously differentiable merit function P of concern here can be written as:

$$\begin{aligned} P(x,\lambda;\varepsilon) &= f(x) + \lambda^T h(x) + \rho(x,\lambda)^T \phi_l(x,\lambda;\varepsilon) + \sigma(x,\lambda)^T \phi_u(x,\lambda;\varepsilon) \\ &\quad + \frac{1}{2\varepsilon}\left[\|h(x)\|^2 + \|\phi_l(x,\lambda;\varepsilon)\|^2_{r(x)^{-1}} + \|\phi_u(x,\lambda;\varepsilon)\|^2_{s(x)^{-1}}\right] + \|\varphi(x,\lambda)\|^2, \end{aligned}$$

where

$$\begin{aligned} \phi_l(x,\lambda;\varepsilon) &= \max\{l-x, -\varepsilon r(x)\circ\rho(x,\lambda)\}, \\ \phi_u(x,\lambda;\varepsilon) &= \max\{x-u, -\varepsilon s(x)\circ\sigma(x,\lambda)\}, \end{aligned}$$

and

$$\varphi(x,\lambda) = \nabla h(x)^T \left[w(x)\circ\nabla_x L(x,\lambda)\right] + \|h(x)\|^2\lambda.$$

The structure of the function P derives from the one of the classical Augmented Lagrangian Function (see ([25]), ([28])), its distinguishing points are:

- the multiplier functions $\rho(x,\lambda)$ and $\sigma(x,\lambda)$ that estimate the multipliers ρ and σ associated with the bound constraints;
- the barrier terms $r(x)^{-1}$ and $s(x)^{-1}$ that guarantee that an unconstrained minimization of the penalty function produces points which belong to the relaxation of the feasible set $\mathcal{S}$;
- the term $\|\varphi(x,\lambda)\|^2$ which forces the variable λ to satisfy the KKT conditions and prevents unboundedness of sequences $\{\lambda^k\}$.

2.2. The gradient of $P(x,\lambda;\varepsilon)$

The function P is continuously differentiable. It can be verified that its gradient is given by:

$$\begin{aligned} \nabla_x P(x,\lambda;\varepsilon) &= \frac{1}{\varepsilon}\nabla h(x)h(x) \\ &\quad + \left[\nabla_x\rho(x,\lambda) - \frac{1}{\varepsilon}\mathrm{diag}\{r(x)^{-1}\}\right]\phi_l(x,\lambda;\varepsilon) \\ &\quad + \left[\nabla_x\sigma(x,\lambda) + \frac{1}{\varepsilon}\mathrm{diag}\{s(x)^{-1}\}\right]\phi_u(x,\lambda;\varepsilon) \\ &\quad - \frac{1}{2\varepsilon}r(x)^{-2}\circ\phi_l(x,\lambda;\varepsilon)^2 + \frac{1}{2\varepsilon}s(x)^{-2}\circ\phi_u(x,\lambda;\varepsilon)^2 \\ &\quad + 2\nabla_x\varphi(x,\lambda)\varphi(x,\lambda) \\ \nabla_\lambda P(x,\lambda;\varepsilon) &= h(x) + \nabla_\lambda\sigma(x,\lambda)\phi_u(x,\lambda;\varepsilon) + \nabla_\lambda\rho(x,\lambda)\phi_l(x,\lambda;\varepsilon) \\ &\quad + 2\nabla_\lambda\varphi(x,\lambda)\varphi(x,\lambda), \end{aligned}$$

where

$$
\begin{aligned}
\nabla_x \varphi &= \nabla_x^2 L(x,\lambda)\mathrm{diag}\{w(x)\}\nabla h(x) + 2\nabla h(x) h(x)\lambda^T \\
&\quad + \sum_{j=1}^{m} \nabla^2 h_j(x)\mathrm{diag}\{w(x)\}\nabla_x L(x,\lambda) i_j^T \\
&\quad -2\mathrm{diag}\{w(x)^2 \circ \left[(l-x)^{-3} - (x-u)^{-3}\right] \circ \nabla_x L(x,\lambda)\}\nabla h(x) \\
\nabla_\lambda \varphi &= \nabla h(x)^T \mathrm{diag}\{w(x)\}\nabla h(x) + \|h(x)\|^2 I \\
\nabla_x \rho(x,\lambda) &= \nabla_x^2 L(x,\lambda)\mathrm{diag}\{w(x)\circ (l-x)^{-2}\} + \Phi(x,\lambda) \\
\nabla_x \sigma(x,\lambda) &= -\nabla_x^2 L(x,\lambda)\mathrm{diag}\{w(x)\circ (x-u)^{-2}\} + \Phi(x,\lambda) \\
\nabla_\lambda \rho(x,\lambda) &= \nabla h(x)^T \mathrm{diag}\{w(x)\circ (l-x)^{-2}\} \\
\nabla_\lambda \sigma(x,\lambda) &= -\nabla h(x)^T \mathrm{diag}\{w(x)\circ (x-u)^{-2}\}
\end{aligned}
\tag{3}
$$

with

$$
\Phi(x,\lambda) = \quad -2\mathrm{diag}\{(u-l)\circ (l-x)^{-3} \circ w(x)^2 \circ (x-u)^{-3} \circ \nabla_x L(x,\lambda)\}.
$$

2.3. Exactness properties of the function $P(x, \lambda; \varepsilon)$

In this subsection we summarize the exactness properties of the merit function $P(x, \lambda; \varepsilon)$. These properties can be proved, with minor changes, as in [14] and [15]. In essence, function $P(x, \lambda; \varepsilon)$ is exact in the sense that for sufficiently small values of the penalty parameter ε, a correspondence can be established between KKT-tuples of Problem (1) and stationary points of $P(x, \lambda; \varepsilon)$, as well as between local/global solutions of Problem (1) and local/global minimizers of $P(x, \lambda; \varepsilon)$.

More precisely, let us introduce the following assumptions:

Assumption 1. *Every $x \in \mathcal{F}$ satisfies the LICQ.*

Assumption 2. *Every point $x \in \mathcal{S} \setminus \mathcal{F}$ satisfies the EMFCQ.*

The following proposition states the relevant exactness properties of the merit function $P(x, \lambda; \varepsilon)$.

Proposition 2.1. *Let $P(x, \lambda; \varepsilon)$ be the merit function defined in subsection 2.1. Then,*

(i) *For all $\varepsilon > 0$, if $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$ is a KKT-tuple of Problem (1) then $(\bar{x}, \bar{\lambda})$ is a stationary point of P;*

(ii) *for all $\varepsilon > 0$, if $(\bar{x}, \bar{\lambda}) \in \mathcal{S} \times \Re^m$ is a stationary point of P such that $h(\bar{x}) = 0$, $\phi_l(\bar{x}, \bar{\lambda}; \varepsilon) = 0$ and $\phi_u(\bar{x}, \bar{\lambda}; \varepsilon) = 0$, then $(\bar{x}, \bar{\lambda}, \rho(\bar{x}, \bar{\lambda}), \sigma(\bar{x}, \bar{\lambda}))$ is a KKT-tuple of Problem (1).*

Assume further that Assumptions 1 and 2 hold:

(iii) *a value $\bar{\varepsilon} > 0$ exists, such that for all $\varepsilon \in (0, \bar{\varepsilon}]$, if $(\bar{x}, \bar{\lambda})$ is a stationary point (local/global minimizer) of P then $(\bar{x}, \bar{\lambda}, \rho(\bar{x}, \bar{\lambda}), \sigma(\bar{x}, \bar{\lambda}))$ is a KKT-tuple (local/global minimizer) of Problem (1);*

(iv) *a value $\bar{\varepsilon} > 0$ exists, such that for all $\varepsilon \in (0, \bar{\varepsilon}]$, if $\bar{x}$ is a global minimum point of Problem (1) with KKT multipliers $\bar{\lambda}, \bar{\rho}, \bar{\sigma}$, then $(\bar{x}, \bar{\lambda})$ is a global minimum point of P in $\mathcal{S} \times \Re^m$.*

Given a point $(x^0, \lambda^0) \in \mathcal{S} \times \Re^m$ let us denote by $\Omega(x^0, \lambda^0; \varepsilon)$ the level set:

$$
\Omega(x^0, \lambda^0; \varepsilon) = \{(x, \lambda) \in \mathcal{S} \times \Re^m : P(x, \lambda; \varepsilon) \le P(x^0, \lambda^0; \varepsilon)\}.
$$

In addition to the above exactness properties, in the following proposition we report some properties that are needed for the definition of an algorithm exploiting the merit function.

Proposition 2.2. *Assume that Assumptions 1 and 2 hold.*

(i) *Let (x^0, λ^0) be a given point in $\mathcal{S} \times \Re^m$, then there exists a value $\bar{\varepsilon} > 0$ such that, for any $\varepsilon \in (0, \bar{\varepsilon}]$, the level set $\Omega(x^0, \lambda^0; \varepsilon)$ is compact. In particular, there exist positive constants $c_1(x^0, \lambda^0)$ and $c_2(x^0, \lambda^0)$ such that*

$$
\Omega(x^0, \lambda^0; \varepsilon) \subseteq \{(x, \lambda) : x \in \mathcal{S},\ \|\lambda\| \le \frac{1}{\varepsilon^{1/2}} c_1(x^0, \lambda^0) + c_2(x^0, \lambda^0)\};
$$

moreover, if $x^0 \in \mathcal{F}$ then $c_1(x^0, \lambda^0) = 0$.

(ii) *A value $\bar{\varepsilon} > 0$ exists, such that for all $\varepsilon \in (0, \bar{\varepsilon}]$, and all $(x, \lambda) \in \Omega(x^0, \lambda^0; \varepsilon)$ we have:*

$$
\|\nabla P(x, \lambda; \varepsilon)\| \ge \|h(x)\| + \|\phi_l(x, \lambda; \varepsilon)\| + \|\phi_u(x, \lambda; \varepsilon)\|.
$$

3. An efficient search direction

As already said in Section 1, we solve Problem (1) by making use of a line search procedure for minimizing the exact penalty-Lagrangian function $P(x, \lambda; \varepsilon)$ in the product space $(x, \lambda) \in \mathrm{IR}^n \times \mathrm{IR}^m$.

In this section we prove that the approach proposed in [5] can be adapted to propose a good descent direction $d = (d_x, d_\lambda) \in \mathrm{IR}^n \times \mathrm{IR}^m$ for the minimization of function $P(x, \lambda; \varepsilon)$ (see also [18, 13]). This direction can be evaluated by solving (possibly approximately) a linear system.
As we will see, the direction is able to provide, under weak assumptions, local convergence with superlinear convergence rate to the minimization procedure, provided that the penalty parameter ε is small enough.

3.1. A property of the multiplier functions $\rho(x, \lambda), \sigma(x, \lambda)$

In this subsection we are concerned with the multiplier functions $\rho(x, \lambda), \sigma(x, \lambda)$ that appear in the definition of the merit function P. They can be used to provide an estimation of the bound constrains active at the solution. More precisely, let us introduce the index sets $\mathcal{L}(x, \lambda)$, $\mathcal{U}(x, \lambda)$ and $\mathcal{M}(x, \lambda)$ defined as:

$$\begin{aligned} \mathcal{L}(x, \lambda) &= \{i : l_i - x_i \geq -\varepsilon r_i(x) \rho_i(x, \lambda)\}, \\ \mathcal{U}(x, \lambda) &= \{i : x_i - u_i \geq -\varepsilon s_i(x) \sigma_i(x, \lambda)\}, \\ \mathcal{M}(x, \lambda) &= \{1, \dots, n\} \setminus (\mathcal{L}(x, \lambda) \cup \mathcal{U}(x, \lambda)). \end{aligned} \tag{4}$$

By the same arguments developed in [18], it is possible to prove the following proposition which shows that, at least in a neighborhood of a KKT-tuple, the preceding sets of indices can be used to estimate the indices of the active and nonactive bounds.

Proposition 3.1. *Let $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$ be a KKT-tuple for Problem (1). Then there exists a neighborhood $\mathcal{B}(\bar{x}, \bar{\lambda})$ of $(\bar{x}, \bar{\lambda})$ such that for all $(x, \lambda) \in \mathcal{B}(\bar{x}, \bar{\lambda})$:*

$$\begin{aligned} \mathcal{L}_+(\bar{x}, \bar{\rho}) \quad &\subseteq \mathcal{L}(x, \lambda) \quad \subseteq \mathcal{L}_0(\bar{x}), \\ \mathcal{U}_+(\bar{x}, \bar{\sigma}) \quad &\subseteq \mathcal{U}(x, \lambda) \quad \subseteq \mathcal{U}_0(\bar{x}), \\ \mathcal{M}_0(\bar{x}) \quad &\subseteq \mathcal{M}(x, \lambda) \quad \subseteq \{i : \ \bar{\rho}_i = 0, \ \bar{\sigma}_i = 0\}. \end{aligned} \tag{5}$$

Moreover if the strict complementarity holds at $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$, then, for all $(x, \lambda) \in \mathcal{B}(\bar{x}, \bar{\lambda})$ it results

$$\begin{aligned} \mathcal{L}(x, \lambda) &= \mathcal{L}_0(\bar{x}), \\ \mathcal{U}(x, \lambda) &= \mathcal{U}_0(\bar{x}), \\ \mathcal{M}(x, \lambda) &= \mathcal{M}_0(\bar{x}). \end{aligned}$$

□

The proposition states that in a neighborhood of a KKT-tuple we can identify exactly the binding bounds with strictly positive multipliers. Hence if strict complementarity holds, our estimate is actually exact.
Note that, far from a KKT-tuple, it may happen that $\mathcal{L}(x, \lambda) \bigcap \mathcal{U}(x, \lambda) \neq \{\emptyset\}$, that is the sets $\mathcal{L}$ and $\mathcal{U}$ could be not disjoint. In the following we will assume that $\mathcal{L}(x, \lambda)$ and $\mathcal{U}(x, \lambda)$ are always disjoint, by putting in only one set the common indices, if any.

For simplicity, in the following we will denote by $\bar{\mathcal{M}}(x, \lambda)$ the index subset defined as

$$\bar{\mathcal{M}}(x, \lambda) = \mathcal{L}(x, \lambda) \cup \mathcal{U}(x, \lambda).$$

By these notations, $\mathcal{M}(x, \lambda)$ collects the indices of variables x_i estimated to be non-binding, and $\bar{\mathcal{M}}(x, \lambda)$ collects the indices of variables x_i estimated to satisfy a lower or on an upper bound at the solution of Problem (1).

3.2. Search direction definition

In order to define the search direction we introduce a suitable system of equations. This system is obtained by "transforming" the equality and inequality KKT conditions into equations by means of efficient estimates of the active constraints at a KKT-tuple. In this way the system has the property that its solutions are KKT-tuples of the nonlinear programming problem, and vice versa. By using the sets of indices defined in subsection 3.1, the system of equations is given by

$$F(x,\lambda) = \begin{bmatrix} (\nabla_x L(x,\lambda))_{\mathcal{M}(x,\lambda)} \\ h(x) \\ (l-x)_{\mathcal{L}(x,\lambda)} \\ (x-u)_{\mathcal{U}(x,\lambda)} \end{bmatrix} = 0. \tag{6}$$

The importance of the preceding system of equations relies on the fact that its solutions are strictly connected to the KKT-tuples of Problem (1). In fact we have the following result.

Proposition 3.2. (i) *Every KKT-tuple $(\bar{x},\bar{\lambda},\bar{\rho},\bar{\sigma})$ of Problem (1) is such that $(\bar{x},\bar{\lambda})$ is a solution of system (6).* (ii) *Every solution $(\bar{x},\bar{\lambda})$ of system (6) is such that $(\bar{x},\bar{\lambda},\rho(\bar{x},\bar{\lambda}),\sigma(\bar{x},\bar{\lambda}))$ is a KKT-tuple of Problem (1).*

Proof. Point (i) Recalling Proposition 3.1 it easy to verify that if $(\bar{x},\bar{\lambda},\bar{\rho},\bar{\sigma})$ is a KKT-tuple of Problem (1) then $(\bar{x},\bar{\lambda})$ is a solution of system (6).
Point (ii) Now we state the converse result. If $(\bar{x},\bar{\lambda})$ is a solution of system (6) then

$$\begin{aligned} (\nabla_x L(\bar{x},\bar{\lambda}))_i &= 0, \quad i \in \mathcal{M}(\bar{x},\bar{\lambda}), & (7)\\ h_j(\bar{x}) &= 0, \quad j = 1,\dots,m, \\ (l-\bar{x})_i &= 0, \quad i \in \mathcal{L}(\bar{x},\bar{\lambda}), & (8)\\ (\bar{x}-u)_i &= 0, \quad i \in \mathcal{U}(\bar{x},\bar{\lambda}). & (9) \end{aligned}$$

By using (7) and (2) we get that

$$\rho_i(\bar{x},\bar{\lambda}) = 0, \quad \sigma_i(\bar{x},\bar{\lambda}) = 0, \qquad i \in \mathcal{M}(\bar{x},\bar{\lambda}). \tag{10}$$

Then (8)-(9) and the definitions of $\mathcal{L}(\bar{x},\bar{\lambda})$, $\mathcal{U}(\bar{x},\bar{\lambda})$, given by (4), yield

$$\rho_i(\bar{x},\bar{\lambda}) \geq 0, \qquad i \in \mathcal{L}(\bar{x},\bar{\lambda}), \tag{11}$$

$$\sigma_i(\bar{x},\bar{\lambda}) \geq 0, \qquad i \in \mathcal{U}(\bar{x},\bar{\lambda}). \tag{12}$$

Now (9) and (2) imply:

$$\rho_i(\bar{x},\bar{\lambda}) = 0, \qquad i \in \mathcal{U}(\bar{x},\bar{\lambda}), \tag{13}$$

while (8) and (2) imply:

$$\sigma_i(\bar{x},\bar{\lambda}) = 0, \qquad i \in \mathcal{L}(\bar{x},\bar{\lambda}). \tag{14}$$

Then (10) and the definition of the set $\mathcal{M}(\bar{x},\bar{\lambda})$ imply that

$$(l-\bar{x})_i < 0, \qquad (\bar{x}-u)_i < 0, \qquad i \in \mathcal{M}(\bar{x},\bar{\lambda}). \tag{15}$$

By using (8), (9), (10), (11), (12), (13)-(15) we have that

$$\begin{aligned} l-\bar{x} \leq 0 \qquad & \bar{x}-u \leq 0 \\ \rho(\bar{x},\bar{\lambda}) \geq 0, \qquad & \sigma(\bar{x},\bar{\lambda}) \geq 0, \\ \rho(\bar{x},\bar{\lambda})^T(l-\bar{x}) = 0, \quad & \sigma(\bar{x},\bar{\lambda})^T(\bar{x}-u) = 0. \end{aligned} \tag{16}$$

Now, by using again (2) we have :

$$\begin{aligned} \rho_i(\bar{x},\bar{\lambda}) &= (\nabla_x L(\bar{x},\bar{\lambda}))_i, \quad i \in \mathcal{L}(\bar{x},\bar{\lambda}), \\ \sigma_i(\bar{x},\bar{\lambda}) &= -(\nabla_x L(\bar{x},\bar{\lambda}))_i, \quad i \in \mathcal{U}(\bar{x},\bar{\lambda}). \end{aligned}$$

and this, together with (16), implies that $(\bar{x},\bar{\lambda},\rho(\bar{x},\bar{\lambda}),\sigma(\bar{x},\bar{\lambda}))$ is a KKT-tuple for Problem (1). □

For simplicity, let us now denote by the apex k the quantities evaluated at (x^k,λ^k); in particular for the sets in (4) we have:

$$\mathcal{L}^k = \mathcal{L}(x^k,\lambda^k), \ \mathcal{U}^k = \mathcal{U}(x^k,\lambda^k), \ \mathcal{M}^k = \mathcal{M}(x^k,\lambda^k).$$

If the sets $\mathcal{L}^k$, $\mathcal{U}^k$ and $\mathcal{M}^k$ were constant in a neighborhood of the point (x^k,λ^k) the Newton direction $(d_x^k,\ d_\lambda^k)$ for the equation (6) at the point (x^k,λ^k) would be obtained by solving:

$$\begin{bmatrix} \nabla_x^2 L^k_{\mathcal{M}^k} & \nabla h^k_{\mathcal{M}^k} \\ (\nabla h^k)^T & 0 \\ -E_{\mathcal{L}^k} & 0 \\ E_{\mathcal{U}^k} & 0 \end{bmatrix} \begin{bmatrix} d_x \\ d_\lambda \end{bmatrix} = - \begin{bmatrix} (\nabla_x L^k)_{\mathcal{M}^k} \\ h^k \\ (l-x^k)_{\mathcal{L}^k} \\ (x^k-u)_{\mathcal{U}^k} \end{bmatrix}. \tag{17}$$

However, in the neighborhood of the point (x^k, λ^k) the sets $\mathcal{L}^k$, $\mathcal{U}^k$ and $\mathcal{M}^k$ are not constant unless $(x^k, \lambda^k, \rho(x^k, \lambda^k), \sigma(x^k, \lambda^k))$ stays in a neighborhood of a KKT-tuple which satisfies the strict complementarity condition. A distinguishing feature of our approach consists in proving that the direction (d_x, d_λ) obtained as solution of system (17) is efficient in terms of convergence and rate of convergence even if the strict complementarity assumption is not satisfied.

Before going into details, we observe that system (17) can be written in a more compact form. Indeed the components in $\mathcal{L}^k$,$\mathcal{U}^k$ of the direction d_x can be obtained directly as:

$$\begin{aligned} (d_x)_{\mathcal{L}^k} &= (l - x^k)_{\mathcal{L}^k} \\ (d_x)_{\mathcal{U}^k} &= (u - x^k)_{\mathcal{U}^k} \end{aligned} \tag{18}$$

whereas, writing $d_x = E^T_{\mathcal{M}^k}(d_x)_{\mathcal{M}^k} + E^T_{\bar{\mathcal{M}}}(d_x)_{\bar{\mathcal{M}}}$, the components $(d_x)_{\mathcal{M}^k}$ and (d_λ) can be obtained by solving the reduced system

$$\begin{bmatrix} \nabla^2_x L^k_{\mathcal{M}^k,\mathcal{M}^k} & \nabla h^k_{\mathcal{M}^k} \\ (\nabla h^k_{\mathcal{M}^k})^T & 0 \end{bmatrix} \begin{bmatrix} (d_x)_{\mathcal{M}^k} \\ d_\lambda \end{bmatrix} = -\tilde{F}(x^k, \lambda^k) \tag{19}$$

where

$$\tilde{F}(x^k, \lambda^k) = \begin{bmatrix} (\nabla^2_x L^k_{\mathcal{M}^k,\bar{\mathcal{M}}^k})(d_x)_{\bar{\mathcal{M}}^k} + (\nabla_x L^k)_{\mathcal{M}^k} \\ h^k + (\nabla h^k_{\bar{\mathcal{M}}^k})^T (d_x)_{\bar{\mathcal{M}}^k} \end{bmatrix}.$$

We remark that the dimension of the linear system (19) is $m + |\mathcal{M}^k|$, that is the number of equality constraints plus the number of variables estimated to be not binding. We observe that the non-singularity of the coefficient matrix of system (17) is equivalent to the non-singularity of the matrix in the l.h.s. of (19):

$$\begin{bmatrix} \nabla^2_x L^k_{\mathcal{M}^k,\mathcal{M}^k} & \nabla h^k_{\mathcal{M}^k} \\ (\nabla h^k_{\mathcal{M}^k})^T & 0 \end{bmatrix}. \tag{20}$$

We point out that non-singularity of matrix (20) can be ensured also in a neighborhood of a point that satisfies LICQ and SSOSC [19]. In particular, we have:

Remark 2. *If $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$ is a KKT-tuple that satisfies the LICQ and the SSOSC, by repeating the proof of Proposition 3.1 of [18], we can conclude that there exists a neighborhood $\mathcal{B}(\bar{x}, \bar{\lambda})$ of $(\bar{x}, \bar{\lambda})$, such that the matrix (20) is non singular for all $(x, \lambda) \in \mathcal{B}(\bar{x}, \bar{\lambda})$.*

3.3. An inexact computation procedure for the direction d

In this subsection we require the following assumption to hold true:

Assumption 3. The matrix $\nabla h_{\mathcal{M}^k}(x^k)$ is full rank.

When dealing with large dimensional problems, it is convenient to resort to some approximate, rather then exact, solution of system (19), that can be evaluated with less computational effort, provided that good convergence properties are retained. In this subsection we describe a suitable inexact computation procedure for evaluating the direction d^k at a point (x^k, λ^k).

For the solution of system (19), taking into account its structure, and in particular the presence of the null submatrix in the last m equations, we follow the approach proposed in [9]. This approach needs Assumption 3 which we suppose to hold true throughout this entire subsection.

The computation is based on the decomposition of the vector $(d_x)_{\mathcal{M}^k}$ into a horizontal component $d_o \in \Re^{|\mathcal{M}^k|}$ belonging to the null space $\mathcal{N}((\nabla h^k_{\mathcal{M}^k})^T)$ and a vertical component $d_v \in \Re^{|\mathcal{M}^k|}$ belonging to the range space $\mathcal{R}(\nabla h^k_{\mathcal{M}^k})$, i.e.

$$(d_x)_{\mathcal{M}^k} = d_o + d_v$$

with

$$(d_v \quad \nabla h^k_{\mathcal{M}^k})^T d_o = 0. \tag{21}$$

By this decomposition we obtain the vertical component d_v as an exact solution of the last m equations of system (19):

$$(\nabla h^k_{\mathcal{M}^k})^T d_v = -h^k - (\nabla h^k_{\bar{\mathcal{M}}^k})^T (d^k_x)_{\bar{\mathcal{M}}^k}; \tag{22}$$

in particular, we take d_v^k as the minimum norm solution of equation (22), that is:

$$d_v^k = -\nabla h^k_{\mathcal{M}^k} \left(\nabla {h^k_{\mathcal{M}^k}}^T \nabla h^k_{\mathcal{M}^k} \right)^{-1} \left(h^k + (\nabla h^k_{\bar{\mathcal{M}}^k})^T (d_x^k)_{\bar{\mathcal{M}}^k} \right).$$

Then we can introduce the projection operator $\mathcal{P}^k$ on the null space of $(\nabla h^k_{\mathcal{M}^k})^T$, given by

$$\mathcal{P}^k = E_{\mathcal{M}^k,\mathcal{M}^k} - \nabla h^k_{\mathcal{M}^k} \left((\nabla h^k_{\mathcal{M}^k})^T \nabla h^k_{\mathcal{M}^k} \right)^{-1} (\nabla h^k_{\mathcal{M}^k})^T. \tag{23}$$

By noting that $\mathcal{P}^k \nabla h^k_{\mathcal{M}^k}\, d_\lambda = 0$, we can break the first $|\mathcal{M}^k|$ equations of system (19) into a system of $2|\mathcal{M}^k|$ equations, which, omitting the arguments, can be written as:

$$\mathcal{P}^k \left(\nabla_x^2 L^k_{\mathcal{M}^k,\mathcal{M}^k} (d_x)_{\mathcal{M}^k} + \nabla_x^2 L^k_{\mathcal{M}^k,\bar{\mathcal{M}}^k} (d_x)_{\bar{\mathcal{M}}^k} + \nabla_x L^k_{\mathcal{M}^k} \right) = 0, \tag{24}$$

$$(E_{\mathcal{M}^k,\mathcal{M}^k} - \mathcal{P}^k) \left(\nabla_x^2 L^k_{\mathcal{M}^k,\mathcal{M}^k} (d_x)_{\mathcal{M}^k} + \nabla_x^2 L^k_{\mathcal{M}^k,\bar{\mathcal{M}}^k} (d_x)_{\bar{\mathcal{M}}^k} + \nabla h^k_{\mathcal{M}^k} d_\lambda + \nabla_x L^k_{\mathcal{M}^k} \right) = 0. \tag{25}$$

Recalling that $(d_x)_{\mathcal{M}^k} = d_o + d_v$, once obtained d_v, equation (24) constitutes a system of $|\mathcal{M}^k|$ equations in the horizontal component d_o. This system represents the core of the computation and we solve it with increasing accuracy. In particular, we let d_o solve equation (24) with a residual error vector $\tau^k \in \mathbb{R}^{|\mathcal{M}^k|}$, namely we let:

$$\mathcal{P}^k \nabla_x^2 L^k_{\mathcal{M}^k,\mathcal{M}^k} d_o^k = -\mathcal{P}^k \left(\nabla_x^2 L^k_{\mathcal{M}^k,\mathcal{M}^k} d_v^k + \nabla_x^2 L^k_{\mathcal{M}^k,\bar{\mathcal{M}}^k} (d_x^k)_{\bar{\mathcal{M}}^k} + \nabla_x L^k_{\mathcal{M}^k} \right) + \tau^k,$$

where τ^k will be driven to zero fast enough to guarantee a superlinear rate of local convergence.
Having calculated the component $(d_x^k)_{\mathcal{M}^k}$, we can compute d_λ^k by solving exactly system (25).

The Inexact Computation Procedure (ICP) of the search direction can be summarized in the following way.

Inexact Computation Procedure (ICP)

The direction $d^k = (d_x^k, d_\lambda^k)$ is given by:

$$(d_x^k)_{\mathcal{L}^k} = (l - x^k)_{\mathcal{L}^k},$$
$$(d_x^k)_{\mathcal{U}^k} = (u - x^k)_{\mathcal{U}^k},$$
$$(d_x^k)_{\mathcal{M}^k} = d_o^k + d_v^k, \qquad \text{where:}$$

$$d_v^k = -\nabla h^k_{\mathcal{M}^k} \left(\nabla {h^k_{\mathcal{M}^k}}^T \nabla h^k_{\mathcal{M}^k} \right)^{-1} \left(h^k + (\nabla h^k_{\bar{\mathcal{M}}^k})^T (d_x^k)_{\bar{\mathcal{M}}^k} \right); \tag{26}$$

d_o^k satisfies

$$\mathcal{P}^k \nabla_x^2 L^k_{\mathcal{M}^k,\mathcal{M}^k} d_o^k = -\mathcal{P}^k \left(\nabla_x^2 L^k_{\mathcal{M}^k,\mathcal{M}^k} d_v^k + \nabla_x^2 L^k_{\mathcal{M}^k,\bar{\mathcal{M}}_k} (d_x)_{\bar{\mathcal{M}}_k} + \nabla_x L^k_{\mathcal{M}^k} \right) + \tau^k; \tag{27}$$

$$d_\lambda^k = -\left(\nabla {h^k_{\mathcal{M}^k}}^T \nabla h^k_{\mathcal{M}^k} \right)^{-1} (\nabla h^k_{\mathcal{M}^k})^T \left(\nabla_x^2 L^k_{\mathcal{M}^k,\mathcal{M}^k} d_x^k + \nabla_x L^k_{\mathcal{M}^k} \right). \tag{28}$$

By letting

$$y^k = \mathcal{P}^k \left(\nabla_x^2 L^k_{\mathcal{M}^k,\mathcal{M}^k} d_v^k + \nabla_x^2 L^k_{\mathcal{M}^k,\bar{\mathcal{M}}^k} (d_x^k)_{\bar{\mathcal{M}}^k} + \nabla_x L^k_{\mathcal{M}^k} \right), \tag{29}$$

equation (27) can be written in a more compact form:

$$\mathcal{P}^k \nabla_x^2 L^k_{\mathcal{M}^k,\mathcal{M}^k} d_o^k = -y^k + \tau^k.$$

Remark 3. *In the following we suppose that whenever d^k is referenced, Assumption 3 holds.*

3.4. A truncated procedure to compute d_o

In the procedure ICP, it remains to explain how to obtain a solution d_o^k of system (27). Since, as already said, the main aim of the paper is the definition of an algorithm suitable to tackle large scale problems, we resort to a truncated conjugate gradient-type method.
We note that, by posing

$$H^k = \nabla_x^2 L_{\mathcal{M}^k,\mathcal{M}^k}(x^k, \lambda^k)$$

and

$$d_o^k = \mathcal{P}^k v$$

where $v \in \Re^{|\mathcal{M}^k|}$ and $\mathcal{P}^k$ is given by (23), system (27) can be rewritten as a symmetric system of the type

$$\mathcal{P}^k H^k \mathcal{P}^k v = -y^k + \tau^k, \tag{30}$$

with y^k defined in (29).
Essentially, the computation of d_o^k is based on the use of a CG-type algorithm for computing a stationary point of the quadratic function

$$\omega(v) = (y^k)^T v + \frac{1}{2} v^T \mathcal{P}^k H^k \mathcal{P}^k v.$$

Following [9], the projection matrix $\mathcal{P}^k$ can be seen as a preconditioning matrix. Preconditioned Conjugate Gradient (PCG) methods differ in the updating formulae of standard CG methods for the presence of a preconditioning matrix $\mathcal{P}$. However, all the computations for solving system (30) can be directly referred to the original variable d_o^k rather than to v. In this way it is possible to derive an iterative procedure that using the same computations of a conjugate gradient method, except for an additional projection per iteration, computes d_o^k with a controlled accuracy.
Standard PCG algorithms generate sequences $\{t^i\}, \{s^i\}, \{r^i\}$, where t^i is the approximated solution of system (27) at the i-th inner iteration of the algorithm, $s^0, \ldots, s^i$ are the conjugate directions and r^i denotes the residual ζ.
The general scheme is

$$\begin{aligned} t^{i+1} &= t^i + \nu^i s^i, \\ r^{i+1} &= r^i - \nu^i \mathcal{P}^k H^k s^i, \\ s^{i+1} &= r^{i+1} + \beta^i s^i, \end{aligned}$$

where the stepsize ν^i would minimize the quadratic function ω along the direction s^i if ω were strictly convex, and β^i is chosen so as to maintain conjugacy among the directions s^i .

Standard PCG algorithms stop

- *either* if the residual $\|r^i\|$ is below a given tolerance,
- *or* if a negative curvature direction s^i is found (namely $(s^i)^T H^k s^i < 0$).

The PCG procedure that we use is derived from the approach described in [24], which differs from the standard PCG algorithms outlined above. The main difference is in the fact of computing and using as search direction, a good truncated solution of system (27) even if negative curvature directions are met. In particular, if a negative curvature direction s^i is found, the algorithm does not stop, but continues to generate conjugate directions in order to get better truncated solutions of system (27).

Together with the sequence $\{t^i\}$, the modified PCG algorithm also generates a sequence $\{\theta^i\}$ of vectors $\theta^i \in \Re^n$. This sequence differs from $\{t^i\}$ only if a negative curvature direction s^i is found. In this case the updating rule is $\theta^{i+1} = \theta^i - \nu^i s^i$, that is, the opposite direction $-s^i$ is taken.

The new truncated algorithm stops

- *either* if the residual $\|r^i\|$ is below a given tolerance $\eta^k \|y^k\|$, with $\eta^k > 0$,
- *or* if $|(s^i)^T H s^i| \le c_1 \|s^i\|^2$, with $c_1 > 0$.

Then d_o^k is set either to t^i or to $-t^i$ or to θ^i, depending on the outcome of a suitable test able to guarantee global convergence. Similarly to the unconstrained case, this test ensures that the produced direction satisfies an angle condition with respect to the opposite of the initial residual $-r^0 = y^k$.

The proposed iterative scheme, referred to as Preconditioned Conjugate Gradient Procedure (PCGP), is reported in the following.

Preconditioned Conjugate Gradient Procedure (PCGP)

Data. $c_1 > 0$, $c_2 > 0$ and $\eta^k > 0$.

Initialization Set $t^\circ = 0$, $\theta^\circ = 0$, $r^\circ = -y^k$, $s^\circ = r^\circ$, and $i = 0$.

While $\Big(\; \|r^i\| > \eta^k \|y^k\| \Big)$

If $\Big(\; |{s^i}^T H^k s^i| \leq c_1 \|s^i\|^2 \; \Big)$ set $d_o^k = \begin{cases} r^\circ & \text{if } \; i = 0 \\ \theta^i & \text{if } \; i > 0 \end{cases}$ and **Return.**

Compute

$$
\begin{aligned}
t^{i+1} &= t^i + \nu^i s^i, && \text{with} \qquad \nu^i = \frac{(s^i)^T r^i}{(s^i)^T H^k s^i}, \\
r^{i+1} &= r^i - \nu^i \mathcal{P}^k H^k s^i, && \\
s^{i+1} &= r^{i+1} + \beta^i s^i && \text{with} \qquad \beta^i = \frac{(r^{i+1})^T r^{i+1}}{(r^i)^T r^i}, \\
\theta^{i+1} &= \begin{cases} \theta^i - \nu^i s^i & \text{if } {s^i}^T H^k s^i < -c_1 \|s^i\|^2, \\ \theta^i + \nu^i s^i & \text{if } {s^i}^T H^k s^i > c_1 \|s^i\|^2. \end{cases}
\end{aligned}
$$

Set $\; i = i + 1$

End while

If ($|(y^k)^T t^i| \geq c_2 \|y^k\|^2$) **then** set

$$
d_o^k = \begin{cases} t^i & \text{if } (y^k)^T t^i \leq 0; \\ -t^i & \text{if } (y^k)^T t^i > 0. \end{cases}
$$

else set $\qquad d_o^k = \theta^i.$

Return.

In the above truncated scheme, we assume that the parameter η^k is taken as:

$$
\eta^k = \eta(x^k, \lambda^k), \tag{31}
$$

where η is a nonnegative continuous function such that $\eta(\bar{x}, \bar{\lambda}) = 0$ if $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$ is a KKT-tuple. Moreover, whenever $d_o^k = t^i$ we have that $\tau^k = r^i$.

As a direct consequence of the results proved in [24], the following proposition holds true.

Proposition 3.3. *Let d_o^k be calculated by PCGP, then the following inequalities hold.*

$$
\begin{aligned}
{d_o^k}^T y^k &\leq -\zeta_1 \left\| y^k \right\|^2, && \text{(32a)} \\
\left\| d_o^k \right\| &\leq \;\; \zeta_2 \left\| y^k \right\|, && \text{(32b)}
\end{aligned}
$$

where ζ_1, ζ_2 are two positive constants not depending on the index k.

Proof. See the proof of Theorem 2.2(c) of [24]. □

Conditions (32a) and (32b) of Proposition 3.3 are the extensions to the constrained case of similar conditions used in the truncated Newton method proposed in [24] for unconstrained minimization problems. Roughly speaking, the direction d_o^k should contribute in locating stationary points of the function $L(\mathcal{P}^k x, \lambda^k)$ and the the vector y^k can be seen as the gradient of the function $L(\mathcal{P}^k x, \lambda^k)$. Then conditions (32a) and (32b) guarantee that the direction d_o^k is a suitable bounded descent direction for $L(\mathcal{P}x, \lambda)$.

The next proposition proves that in a neighbourhood of a KKT-tuple $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$ satisfying LICQ and SSOSC, and for sufficiently small values of c_1 appearing in PCGP, the horizontal step d_o^k satisfies the stopping criterion of PCGP. Furthermore it shows that the truncation error can be forced to zero rapidly enough to guarantee the superlinear convergence rate of the local algorithm.

Proposition 3.4. *Let $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$ be a KKT-tuple for Problem (1) which satisfies LICQ and SSOSC and let $\{(x^k, \lambda^k)\}$ be a sequence converging to $(\bar{x}, \bar{\lambda})$. Then an integer $\bar{k}$ and a positive constant $\bar{c}_1$ exist such that for all $k \geq \bar{k}$ and $c_1 \leq \bar{c}_1$, the horizontal step d_o^k, obtained by PCGP, is such that*

(i) $\|\tau^k\| \leq \eta^k \|y^k\|$,

(ii) *if η^k converges to zero, then*

$$\lim_{k\to\infty} \frac{\|\tau^k\|}{\|y^k\|} = 0.$$

Proof. The proof follows by analogous reasoning as in the proof of Proposition 2.6 of [13]. □

In next Section, it will be shown that the fact that the horizontal component d_o^k satisfies (32) plays a fundamental role in order to guarantee the global convergence of the sequence $\{x^k\}$ produced by the overall algorithm.

We point out that the linear algebra computations needed to evaluate d_v^k as in (26) and d_λ^k as in (28) can be performed in a very efficient way, as shown in [9, 13].

3.5. Properties of the search direction d

The main aim of this subsection is to show that, if $\{x^k, \lambda^k\}$ is a bounded sequence, and $\{d^k\}$ is the sequence obtained by the ICP and such that Proposition 3.3 holds, then there is a correspondence between the fact that the sequence $\{x^k, \lambda^k\}$ has limit points $(\bar{x}, \bar{\lambda})$ such that $(\bar{x}, \bar{\lambda}, \rho((\bar{x}, \bar{\lambda}), \sigma(\bar{x}, \bar{\lambda})))$ is a KKT-tuple of Problem 1 and the fact that $lim_{k\to\infty}\|d^k\| = 0$. This correspondence is essential in establishing the global convergence of the algorithm that employs d as search direction in the minimization of the function $P(x, \lambda; \varepsilon)$.

In order to show the correspondence we need to establish some technical inequalities that will be instrumental also for subsequent developments. To this aim we introduce the following notation:

$$z = \nabla_x^2 L d_x + \nabla h d_\lambda + \nabla_x L.$$

In order to state the main theoretical result of the subsection we need to introduce a technical proposition, whose proof is given in Appendix A for the interested reader.

Proposition 3.5. *Assume that the direction d^k is computed by the ICP, and that the sequence $\{x^k, \lambda^k\}$ is bounded. Then*

(i) *a positive constant $\tilde{c}$, not depending on k, exists such that*

$$\|(\rho^k - z^k)_{\mathcal{L}}\| + \|\rho^k_{\mathcal{U}\cup\mathcal{M}}\| + \|(\sigma^k + z^k)_{\mathcal{U}}\| + \|\sigma^k_{\mathcal{L}\cup\mathcal{M}}\| \leq \tilde{c}\|d^k\|;$$

(ii) *a positive constant $\hat{c}$, not depending on k, exists such that:*

$$\|\nabla_x L(x^k, \lambda^k) - \rho^k + \sigma^k\| \leq \hat{c}\|d^k\|.$$

Proof. See Appendix A. □

Now we are ready to introduce the main result concerning the truncated direction. Roughly speaking the proposition states that if we have a bounded sequence of pairs $\{(x^k, \lambda^k)\}$ and a sequence of directions computed by means of Procedure ICP $\{d^k\}$, then $\|d^k\|$ tends to zero if and only if every accumulation point of the sequence $\{(x^k, \lambda^k, \rho(x^k, \lambda^k), \sigma(x^k, \lambda^k))\}$ is a KKT-tuple for Problem (1).

Proposition 3.6. *Suppose that Assumption 1 holds. Assume the sequence $\{x^k, \lambda^k\}$ is bounded. Then,*

(i) *for every k, let d^k be computed by the ICP. If it results that*

$$\lim_{k\to\infty} \|d^k\| = 0,$$

then every accumulation point $(\tilde{x}, \tilde{\lambda}, \tilde{\rho}, \tilde{\sigma})$ of $\{(x^k, \lambda^k, \rho(x^k, \lambda^k), \sigma(x^k, \lambda^k))\}$ is a KKT-tuple for Problem (1).

(ii) *If every accumulation point $(\tilde{x}, \tilde{\lambda}, \tilde{\rho}, \tilde{\sigma})$ of $\{(x^k, \lambda^k, \rho(x^k, \lambda^k), \sigma(x^k, \lambda^k))\}$ is a KKT-tuple for Problem (1) then, eventually d^k can be computed by the ICP and*

$$\lim_{k\to\infty} \|d^k\| = 0.$$

Proof. Point (i). Let $(\tilde{x}, \tilde{\lambda})$ be an accumulation point of the sequence $\{(x^k, \lambda^k)\}$, we can extract a subsequence such that

$$\begin{array}{l}(x^k, \lambda^k) \to (\tilde{x}, \tilde{\lambda}), \\ \mathcal{L}^k \to \mathcal{L}(\tilde{x}, \tilde{\lambda}) = \tilde{\mathcal{L}},\ \mathcal{U}^k \to \mathcal{U}(\tilde{x}, \tilde{\lambda}) = \tilde{\mathcal{U}}, \mathcal{M}^k \to \mathcal{M}(\tilde{x}, \tilde{\lambda}) = \tilde{\mathcal{M}} \quad \text{for all } k.\end{array}$$

By assumption we have that $d_x^k \to 0$, $d_\lambda^k \to 0$. Then, by point (i) of Proposition 3.5, we have $z_{\mathcal{L}}^k \to \rho_{\tilde{\mathcal{L}}}(\tilde{x}, \tilde{\lambda}) = \tilde{\rho}_{\mathcal{L}}$, $z_{\mathcal{U}} \to -\sigma_{\tilde{\mathcal{U}}}(\tilde{x}, \tilde{\lambda}) = -\tilde{\sigma}_{\mathcal{U}}$ and

$$\tilde{\sigma}_{\tilde{\mathcal{L}} \cup \tilde{\mathcal{M}}} = 0, \qquad \tilde{\rho}_{\tilde{\mathcal{U}} \cup \tilde{\mathcal{M}}} = 0. \tag{33}$$

Taking the limit in (17), we have that

$$h(\tilde{x}) = 0, \qquad (l - \tilde{x})_{\tilde{\mathcal{L}}} = 0, \qquad (\tilde{x} - u)_{\tilde{\mathcal{U}}} = 0. \tag{34}$$

Moreover, by point (ii) of Proposition 3.5, we have that

$$\nabla_x L(\tilde{x}, \tilde{\lambda}) - \tilde{\rho} + \tilde{\sigma} = 0. \tag{35}$$

Recalling (4), we get, using (34) and (33)

$$\rho_{\tilde{\mathcal{L}}}(\tilde{x}, \tilde{\lambda}) \geq 0,\ \sigma_{\tilde{\mathcal{U}}}(\tilde{x}, \tilde{\lambda}) \geq 0,\ (l - \tilde{x})_{\tilde{\mathcal{U}} \cup \tilde{\mathcal{M}}} \leq 0,\ (\tilde{x} - u)_{\tilde{\mathcal{L}} \cup \tilde{\mathcal{M}}} \leq 0. \tag{36}$$

Hence, (33)–(36) imply that $(\tilde{x}, \tilde{\lambda}, \tilde{\rho}, \tilde{\sigma})$ is a KKT-tuple.

Point (ii). Assume now that every accumulation point $(\tilde{x}, \tilde{\lambda}, \tilde{\rho}, \tilde{\sigma})$ of $\{(x^k, \lambda^k, \rho(x^k, \lambda^k), \sigma(x^k, \lambda^k))\}$ is a KKT-tuple for Problem (1). By Proposition 3.1 and taking into account Assumption 1, we have by continuity that, for k sufficiently large Assumption 3 is satisfied. Now, we proceed by contradiction. Let us suppose that

$$\lim_{k \to \infty} \|d^k\| = 0,$$

does not hold. In this case we can extract subsequences, that we rename again $\{x^k, \lambda^k\}$, $\{d_x^k, d_\lambda^k\}$, such that

$$\begin{aligned} &\lim_{k \to \infty} (x^k, \lambda^k) = (\tilde{x}, \tilde{\lambda}), \\ &\lim_{k \to \infty} (d_x^k, d_\lambda^k) = (\tilde{d}_x, \tilde{d}_\lambda), \\ &\lim_{k \to \infty} \rho(x^k, \lambda^k) = \rho(\tilde{x}, \tilde{\lambda}) = \tilde{\rho}, \\ &\lim_{k \to \infty} \sigma(x^k, \lambda^k) = \sigma(\tilde{x}, \tilde{\lambda}) = \tilde{\sigma}, \\ &\lim_{k \to \infty} \|d^k\| > 0, \end{aligned} \tag{37}$$

where $(\tilde{x}, \tilde{\lambda}, \tilde{\rho}, \tilde{\sigma})$ is a KKT-tuple.
Since the number of possibly different estimates $\mathcal{L}^k$, $\mathcal{M}^k$, $\mathcal{U}^k$ is finite, also in this case we assume, without loss of generality, that $\mathcal{L}^k = \tilde{\mathcal{L}}$, $\mathcal{M}^k = \tilde{\mathcal{M}}$, $\mathcal{U}^k = \tilde{\mathcal{U}}$.
By Proposition 3.1 we have that

$$\begin{array}{l}\tilde{\mathcal{L}} \subseteq \mathcal{L}_0(\tilde{x}) \\ \tilde{\mathcal{U}} \subseteq \mathcal{U}_0(\tilde{x}) \\ \tilde{\mathcal{M}} \subseteq \{i : \tilde{\rho}_i = 0, \tilde{\sigma}_i = 0\},\end{array}$$

which implies

$$\begin{aligned} &\lim_{k \to \infty} (x^k - l)_{\tilde{\mathcal{L}}} = 0, \\ &\lim_{k \to \infty} (u - x^k)_{\tilde{\mathcal{U}}} = 0, \\ &\lim_{k \to \infty} \rho(x^k, \lambda^k)_{\tilde{\mathcal{M}} \bigcup \tilde{\mathcal{U}}} = 0 \\ &\lim_{k \to \infty} \sigma(x^k, \lambda^k)_{\tilde{\mathcal{L}} \bigcup \tilde{\mathcal{M}}} = 0. \end{aligned} \tag{38}$$

Now from (18) and (38) we obtain:

$$\begin{aligned} &\lim_{k \to \infty} (d_x^k)_{\tilde{\mathcal{L}}} = 0, \\ &\lim_{k \to \infty} (d_x^k)_{\tilde{\mathcal{U}}} = 0. \end{aligned} \tag{39}$$

Then by using the fact that $(\tilde{x}, \tilde{\lambda}, \tilde{\rho}, \tilde{\sigma})$ is a KKT-tuple, (39), (26), (32b) and (28) we have

$$\begin{aligned}\lim_{k\to\infty} d_v^k &= 0,\\ \lim_{k\to\infty} d_o^k &= 0,\\ \lim_{k\to\infty} d_\lambda^k &= 0.\end{aligned} \tag{40}$$

Finally (38), (39), (40) contradict (37). □

4. Connections between the truncated direction and the merit function

In this section we show that the direction d is a descent direction for the merit function $P(x, \lambda; \varepsilon)$ whenever the current point (x^k, λ^k) is sufficiently close to a pair $(\bar{x}, \bar{\lambda})$ of a KKT-tuple, and the penalty coefficient ε is sufficiently small. To this aim we have to analyze the structure of the directional derivative of the merit function. By recalling the gradient formulas (8), (9), the directional derivative $\nabla P(x, \lambda; \varepsilon)^T d$ is given by:

$$\begin{aligned}\nabla P(x,\lambda;\varepsilon)^T d = {} & d_x^T \nabla_x P(x,\lambda;\varepsilon) + d_\lambda^T \nabla_\lambda P(x,\lambda;\varepsilon)\\ = {} & \frac{1}{\varepsilon} d_x^T \nabla h(x) h(x) + d_\lambda^T h(x)\\ & +2\left[d_x^T \nabla_x \varphi(x,\lambda) + d_\lambda^T \nabla_\lambda \varphi(x,\lambda)\right] \varphi(x,\lambda)\\ & + \left[d_x^T \nabla_x \rho(x,\lambda) + d_\lambda^T \nabla_\lambda \rho(x,\lambda)\right] \ \phi_l(x,\lambda;\varepsilon)\\ & + \left[d_x^T \nabla_x \sigma(x,\lambda) + d_\lambda^T \nabla_\lambda \sigma(x,\lambda)\right] \ \phi_u(x,\lambda;\varepsilon)\\ & -\frac{1}{\varepsilon} d_x^T \operatorname{diag}\{r(x)^{-1}\} \ \phi_l(x,\lambda;\varepsilon) + \frac{1}{\varepsilon} d_x^T \operatorname{diag}\{s(x)^{-1}\} \ \phi_u(x,\lambda;\varepsilon)\\ & -\frac{1}{2\varepsilon} \operatorname{diag}\{r(x)^{-2}\} \ \phi_l(x,\lambda;\varepsilon)^2 + \frac{1}{2\varepsilon} \operatorname{diag}\{s(x)^{-2}\} \ \phi_u(x,\lambda;\varepsilon)^2\end{aligned} \tag{41}$$

The following proposition establishes a technical result needed in the proof of the main result of this subsection. The proof is reported in Appendix C.

Proposition 4.1. *Let $\{(x^k, \lambda^k)\}$ and $\{d^k\}$ be bounded sequences and let d^k be obtained by ICP choosing $\mathcal{L}^k$, $\mathcal{U}^k$ and $\mathcal{M}^k$ as in (4). Then we can write for the directional derivative $d^T \nabla P(x, \lambda; \varepsilon)$:*

$$d_x^{k\,T} \nabla_x P(x^k, \lambda^k; \varepsilon) + d_\lambda^{k\,T} \nabla_\lambda P(x^k, \lambda^k; \varepsilon) = \psi_1(x^k, \lambda^k, d^k; \varepsilon) + \psi_2(x^k, \lambda^k, d^k; \tau), \tag{42}$$

where ψ_1 and ψ_2 are given in Appendix B and

(i) *the term $\psi_1(x^k, \lambda^k, d^k; \varepsilon)$ is such that, $\bar{\varepsilon}$ and $\gamma > 0$ exist for which, the following inequality holds for every $\varepsilon \in (0, \bar{\varepsilon}]$*

$$\psi_1(x^k, \lambda^k, d^k; \varepsilon) \leq -\gamma \|d^k\|^2; \tag{43}$$

(ii) *the term $\psi_2(x, \lambda, d, \tau)$ satisfies the following inequality*

$$\psi_2(x^k, \lambda^k, d^k; \tau) \leq \xi_1(x, \lambda)\|d\| \, \|\tau\| + \xi_2(x, \lambda)\|d\|^2$$

where $\xi_1(x, \lambda)$ and $\xi_2(x, \lambda)$ are nonnegative continuous functions and $\xi_2(\bar{x}, \bar{\lambda}) = 0$ if $(\bar{x}, \bar{\lambda}, \sigma(\bar{x}, \bar{\lambda}), \rho(\bar{x}, \bar{\lambda}))$ satisfies the KKT conditions.

Proof. The proof of this proposition is quite technical and follows the same arguments of the proof of Proposition 3.5 of [13]. Hence, the proof is available in Appendix C for the interested reader. □

Now we can state the main result of the subsection, namely, that the truncated direction satisfies an angle condition with the gradient of the merit function $P(x, \lambda; \varepsilon)$.

Proposition 4.2. *Suppose that Assumption 1 holds. Let $(\bar{x}, \bar{\lambda}, \bar{\rho}, \bar{\sigma})$ be a KKT-tuple for Problem (1) which satisfies SSOSC and let $\{(x^k, \lambda^k)\}$ be a sequence converging to $(\bar{x}, \bar{\lambda})$. Assume that, in procedure ICP, we take $\mathcal{L}^k$, $\mathcal{U}^k$ and $\mathcal{M}^k$ as in (4) and that, in PCGP, we take $c_1 \leq \bar{c}_1$ (where $\bar{c}_1$ is introduced in Proposition 3.4) and η^k as in (31).*
Then $\bar{\varepsilon}$, and $\gamma > 0$ exist such that, for all $\varepsilon \in (0, \bar{\varepsilon}]$ and for sufficiently large k, it results

$$\nabla P(x^k, \lambda^k; \varepsilon)^T d^k \leq -\frac{\gamma}{2}\|d^k\|^2. \tag{44}$$

Proof. By assumption, the sequence $\{(x^k, \lambda^k)\}$ is convergent to $(\bar{x}, \bar{\lambda})$, and hence it is bounded; on the other hand, the instructions of procedure ICP, Assumption 3 and (32b) ensure that also the sequence $\{d^k\}$ is bounded. Therefore we can conclude that, for sufficiently large k, the assumptions of Proposition 4.1 hold.
By Proposition 3.1, the index sets $\mathcal{L}^k$, $\mathcal{U}^k$ and $\mathcal{M}^k$ satisfy properties (5), point (ii) of Proposition 3.6 ensures that, for sufficiently large k, the direction d^k can be computed by procedure ICP.

From (42) and points (i) and (ii) of Proposition 4.1, we have

$$\nabla P(x^k, \lambda^k; \varepsilon)^T d^k \leq \psi_1(x^k, \lambda^k, d^k; \varepsilon) + \xi_1(x^k, \lambda^k)\|d^k\|\|\tau^k\| + \xi_2(x^k, \lambda^k)\|d^k\|^2. \tag{45}$$

Then, since (x^k, λ^k) satisfies the SSOSC, taking $c_1 \leq \bar{c}_1$ and $\eta^k = \eta(x^k, \lambda^k)$ in PCGP, we have from (45), for sufficiently large k:

$$\nabla P(x^k, \lambda^k; \varepsilon)^T d^k \leq \psi_1(x^k, \lambda^k, d^k; \varepsilon) + \xi_1(x^k, \lambda^k)\eta(x^k, \lambda^k)\|d^k\|\|y^k\| + \xi_2(x^k, \lambda^k)\|d^k\|^2. \tag{46}$$

By (32a) we have:

$$\rho_1\|y^k\|^2 \leq |{d_o^k}^T y^k| \leq \|d_o^k\|\|y^k\|,$$

from which, recalling (21), we obtain

$$\|y^k\| \leq \frac{1}{\rho_1}\|d_o^k\| \leq \frac{1}{\rho_1}\|d_o^k + d_v^k\| = \frac{1}{\rho_1}\|d^k\|.$$

Using this inequality in (46) we obtain:

$$\nabla P(x^k, \lambda^k; \varepsilon)^T d^k \leq \psi_1(x^k, \lambda^k, d^k; \varepsilon) + \frac{1}{\rho_1}\xi_1(x^k, \lambda^k)\eta(x^k, \lambda^k)\|d^k\|^2 + \xi_2(x^k, \lambda^k)\|d^k\|^2. \tag{47}$$

Then we obtain from (47)

$$\nabla P(x^k, \lambda^k; \varepsilon)^T d^k \quad \leq \quad -\gamma\|d^k\|^2 + \frac{1}{\rho_1}\xi_1(x^k, \lambda^k)\eta(x^k, \lambda^k)\|d^k\|^2 + \xi_1(x^k, \lambda^k)\|d^k\|^2.$$

Finally, by the properties of the functions ψ_1, ψ_2 and η, we get the result. □

In the following proposition we show that if (44) holds, then direction d^k used within a suitable linesearch technique, is able to enforce convergence to KKT tuples.

Proposition 4.3. *Let $\{(x^k, \lambda^k)\}$ be a bounded sequence, and let d^k be obtained by ICP, choosing $\mathcal{U}^k$ and $\mathcal{L}^k$ as in (4). Assume that $\tilde{\gamma} > 0$ exists such that , for all k, it results*

$$\nabla P(x^k, \lambda^k; \varepsilon)^T d^k \leq -\tilde{\gamma}\|d^k\|^2, \tag{48}$$

and that the following limit holds

$$\lim_{k\to\infty} \nabla P(x^k, \lambda^k; \varepsilon)^T d^k = 0. \tag{49}$$

Then every accumulation point $(\bar{x}, \bar{\lambda})$ of the sequence $\{(x^k, \lambda^k)\}$ yields a KKT-tuple $(\bar{x}, \bar{\lambda}, \rho(\bar{x}, \bar{\lambda}), \sigma(\bar{x}, \bar{\lambda}))$ of Problem (1).

Proof. The proof follows from point (i) of Proposition 3.6, noting that (48) and (49) imply that $\lim_{k\to\infty} d^k = 0$.
□

We point out that property (49) can easily be ensured by employing standard line-search techniques along the direction d^k.

5. The overall algorithm

In this section we introduce an algorithm model for the solution of Problem (1). It is based on the computation of a search direction as described in Section 3.3 and on the exact penalty-augmented Lagrangian function introduced in Section 2 as a mean to enforce global convergence.
The main iteration of the algorithmic scheme is defined as a primal-dual line-search (LS) based method:

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix} + \mu^k \begin{pmatrix} p_x^k \\ p_\lambda^k \end{pmatrix} \tag{50}$$

where μ^k is the step taken along the search direction p^k, which is either the truncated direction d^k or a gradient related descent direction for P, that we denote by z^k.

The overall Exact Penaly-Augmented Lagrangian (ExPAL) algorithm can be stated as follows:

ExPAL Algorithm model

Data. Let $x^0 \in \mathcal{S}$, $\lambda^0 \in R^m$, $\alpha > 0$; $\varepsilon > 0$, $\zeta \in (0,1)$; $\delta > 0$; set $k = 0$.

While $\Big((x^k, \lambda^k, \rho(x^k, \lambda^k), \sigma(x^k, \lambda^k))$ is not a KKT-tuple $\Big)$ **Do**

Step 2. If Assumption 3 holds, compute the truncated direction d^k

If $\left(\nabla P(x^k, \lambda^k; \varepsilon)^T d^k \le -\frac{\varepsilon}{2} \|d^k\|^2 \right)$ **then** set $p^k = d^k$ and go to **Step 5**,

If $\left(\|d^k\| + \|(\nabla h^k{}_{\mathcal{M}})^T \nabla_x L^k_{\mathcal{M}} + \|h^k\| \lambda^k\| \le \delta \right)$ **and** $\left(\psi_1(x^k, \lambda^k, d^k; \varepsilon) \not\le -\varepsilon \|d^k\|^2 \right)$ **then** go to **Step 4**,

Step 3. Compute the gradient related direction z^k

If $\Big(\nabla P(x^k, \lambda^k; \varepsilon)^T z^k \le -\varepsilon \Big(\|\phi_u(x^k, \lambda^k; \varepsilon)\|^2 + \|\phi_l(x^k, \lambda^k; \varepsilon)\|^2 + \|h(x^k)\|^2 \Big) \Big)$

then set $p^k = z^k$ and go to **Step 5**,

Step 4. Update the penalty parameter

Set $\varepsilon = \zeta \varepsilon$.

If $\left(P(x^k, \lambda^k; \varepsilon) \le P(x^0, \lambda^0; \varepsilon) \right)$ **then** set $(x^{k+1}, \lambda^{k+1}) = (x^k, \lambda^k)$,

Else set $(x^{k+1}, \lambda^{k+1}) = (x^0, \lambda^0)$.

Set $k = k + 1$ and go to **Step 2**.

Step 5. Compute $(x^{k+1}, \lambda^{k+1}) = \mathrm{LS}[(x^k, \lambda^k; p^k)]$

Set $k = k + 1$.

End While

Roughly speaking, the algorithm can be described by the following steps.

1. In the first step, (provided that Assumption 3 is satisfied) we compute the truncated direction d^k and we assess its suitability as search direction. Namely, we check whether it is a "good" descent direction for the exact merit function. If this is the case, p^k is set equal to d^k and a line search is performed along it (go to Step 5). Otherwise, we check whether the penalty parameter is to be decreased. Indeed, as pointed out by the theoretical analysis carried out in Section 4, d^k is a good descent direction for $P(x, \lambda; \varepsilon)$ only if the current iterate is sufficiently close to a pair $(\bar{x}, \bar{\lambda})$ of a KKT-tuple of Problem (1) and the penalty parameter is sufficiently small. Hence, d^k could be rejected as search direction either because we are "too far" from a KKT-tuple (and we proceed to Step 3) or because the penalty parameter is too big (and we jump to Step 4).
2. At Step 3 of the Algorithm, we compute a gradient related direction z^k for the exact merit function. By definition, z^k is a good descent direction but, prior to using it in a line search (Step 5), we perform a test for penalty parameter updating. If the latter test is satisfied we proceed to Step 4 for penalty parameter updating. Otherwise we jump to Step 5.
3. Step 4 of the Algorithm is devoted to the updating of the penalty parameter. In this step we decrease the penalty parameter by setting $\varepsilon = \zeta\varepsilon$ (with $\zeta \in (0, 1)$). Then we set $(x^{k+1}, \lambda^{k+1}) = (x^k, \lambda^k)$ if $(x^k, \lambda^k) \in \Omega(x^0, \lambda^0; \varepsilon)$, otherwise we set $(x^{k+1}, \lambda^{k+1}) = (x^0, \lambda^0)$. Finally the iteration index k is incremented and we start a new iteration from Step 2.
4. At Step 5 the new iterate is determined by doing a line search along the search direction p^k.

The line-search procedure employed can be an Armijo-type procedure LS[$(x^k,\ \lambda^k; p^k)$] which returns (x^{k+1}, λ^{k+1}) as in (50), using a value μ^k which is the largest value among $1, 1/2, 1/4, \ldots$ such that:

$$\begin{aligned}&x^k + \mu^k p_x^k \in \mathcal{S},\\ &P(x^k + \mu^k p_x^k, \lambda^k + \mu^k p_\lambda^k; \varepsilon) \le P(x^k, \lambda^k; \varepsilon) + \xi \mu^k \nabla P(x^k, \lambda^k; \varepsilon)^T p^k.\end{aligned}$$

with $\xi \in (0, 1/2)$

Of course, the above monotone line search procedure can be substituted by any other monotone or nonmonotone line-search scheme [24]; indeed, what really matters is that the procedure is able to guarantee that $\{x^k\} \in \mathcal{S}$ and that

$$\lim_{k\to\infty} \nabla P(x^k, \lambda^k; \varepsilon)^T p^k = 0, \tag{51a}$$

$$\lim_{k\to\infty} \mu^k \nabla P(x^k, \lambda^k; \varepsilon)^T p^k = 0. \tag{51b}$$

We remark that by "gradient related" direction, at Step 3 of Algorithm `ExPAL`, we mean a direction z^k that satisfies the following relations:

$$\nabla P(x^k, \lambda^k; \varepsilon)^T z^k \le -c_3 \|\nabla P(x^k, \lambda^k; \varepsilon^k)\|^2, \tag{52a}$$

$$\|z^k\| \le c_4 \|\nabla P(x^k, \lambda^k; \varepsilon^k)\|. \tag{52b}$$

The convergence of Algorithm `ExPAL` is addressed by the following proposition.

Proposition 5.1. *Suppose that Assumptions 1 and 2 hold. Let $\{(x^k, \lambda^k)\}$ be the sequence of points produced by Algorithm* `ExPAL`*, then:*

(i) *after a finite number of steps the penalty parameter ε is no longer updated;*

(ii) *the sequence $\{(x^k, \lambda^k)\}$ is bounded;*

(iii) *every accumulation point of the sequence $\{(x^k, \lambda^k)\}$ yields a KKT-tuple;*

(iv) *if $(\bar{x}, \bar{\lambda})$ is an accumulation point of the sequence $\{(x^k, \lambda^k)\}$ such that $(\bar{x}, \bar{\lambda}, \rho(\bar{x}, \bar{\lambda}), \sigma(\bar{x}, \bar{\lambda}))$ is an isolated KKT-tuple, then the whole sequence $\{(x^k, \lambda^k)\}$ converges to $(\bar{x}, \bar{\lambda})$.*

Assume further that the sequence $\{(x^k, \lambda^k)\}$ converges to a point $(\bar{x}, \bar{\lambda})$ such that $(\bar{x}, \bar{\lambda}, \rho(\bar{x}, \bar{\lambda}), \sigma(\bar{x}, \bar{\lambda}))$ is a KKT-tuple that satisfies the SSOSC, then eventually

(v) the truncated direction is taken as search direction;

(vi) the rate of convergence is superlinear.

Proof. Point (i). The proof is by contradiction. Suppose that the assertion is false. Then the penalty parameter is updated an infinite number of times. Since the penalty parameter can be updated by the instructions either of Step 2 or of Step 3, we have that there exist sequences $\{\varepsilon^i\}$, $\{(x^{\ell^i}, \lambda^{\ell^i})\}$, $\{p^{\ell^i}\}$ such that $\{\varepsilon^i\} \downarrow 0$ and either

$$\|p^{\ell^i}\| + \| \left((\nabla h(x^{\ell^i})_{\mathcal{M}})^T (\nabla h(x^{\ell^i})_{\mathcal{M}}) + \|h(x^{\ell^i})\| I \right) \lambda^{\ell^i} + (\nabla h(x^{\ell^i})_{\mathcal{M}})^T \nabla f(x^{\ell^i})_{\mathcal{M}} \| \le \delta, \tag{53}$$

$$\psi_1(x^{\ell^i}, \lambda^{\ell^i}, p^{\ell^i}; \varepsilon^i) > -\varepsilon^i \|p^{\ell^i}\|^2, \tag{54}$$

or

$$\nabla P(x^{\ell^i}, \lambda^{\ell^i}; \varepsilon^i)^T p^{\ell^i} > -\varepsilon^i \Big(\|\phi_u(x^{\ell^i}, \lambda^{\ell^i}; \varepsilon)\|^2 + \|\phi_l(x^{\ell^i}, \lambda^{\ell^i}; \varepsilon)\|^2 + \|h(x^{\ell^i})\|^2 \Big). \tag{55}$$

Since, by the instructions at Step 4, we have that $\{(x^{\ell^i}, \lambda^{\ell^i})\} \in \Omega(x^0, \lambda^0; \varepsilon^i)$, point (i) of Proposition 2.1 implies that $\{x^{\ell^i}\}$ is bounded.

Assume first that (53) and (54) hold. Assumption 1 implies that the matrix $\big((\nabla h(x^{\ell^i})_{\mathcal{M}})^T\ (\nabla h(x^{\ell^i})_{\mathcal{M}}) + \|h(x^{\ell^i})\| I\big)$ in (53) is non-singular (see [27]), so that (53) yields that $\{\lambda^{\ell^i}\}$ and $\{p^{\ell^i}\}$ are bounded. Therefore the assumptions of point (i) of Proposition 4.1 are satisfied and hence we get a contradiction between the inequality (43) of Proposition 4.1 and (54).

In the second case, recalling (52) and point (iii) of Proposition 2.1, we have that, for sufficiently small values of ε^i,

$$\begin{aligned}\nabla P(x^{\ell^i}, \lambda^{\ell^i}; \varepsilon^i)^T p^{\ell^i} &\le -c_3 \|\nabla P(x^{\ell^i}, \lambda^{\ell^i}; \varepsilon^i)\|^2 \\ &\le -c_3 \Big(\|\phi_u(x^{\ell^i}, \lambda^{\ell^i}; \varepsilon)\|^2 + \|\phi_l(x^{\ell^i}, \lambda^{\ell^i}; \varepsilon)\|^2 + \|h(x^{\ell^i})\|^2 \Big),\end{aligned}$$

which, for i sufficiently large, contradicts (55). This completes the proof of point (i).

By the above point (i), from now on, we assume for the sake of simplicity, that ε stays fixed at the value $\hat{\varepsilon}$.

Point (ii). The proof of this point follows directly from point (i) above and the compactness of $\Omega(x^0, \lambda^0; \hat{\varepsilon})$, stated in point (i) of Proposition 2.2.

Point (iii). Let $\{(x^k, \lambda^k)\}_K$ be a subsequence of $\{(x^k, \lambda^k)\}$ converging to $(\bar{x}, \bar{\lambda})$. We show that $(\bar{x}, \bar{\lambda}, \rho(\bar{x}, \bar{\lambda}), \sigma(\bar{x}, \bar{\lambda}))$ is a KKT-tuple. In fact, by point (ii) above, the sequence $\{(x^k, \lambda^k)\}$ is bounded, and by property (51a) of the line-search procedure

$$\lim_{k\to\infty, k\in K} \nabla P(x^k, \lambda^k; \hat{\varepsilon})^T p^k = 0. \tag{56}$$

Moreover, since the index sets $\mathcal{L}^k$, $\mathcal{U}^k$ and $\mathcal{M}^k$ satisfy property (5), an integer $\bar{k}$ exists such that, for all $k \ge \bar{k}$, Assumption 1 implies that Assumption 3 is satisfied. Now, if for an infinite number of $k \in K$ with $k \ge \bar{k}$, p^k is the truncated direction d^k which satisfies

$$\nabla P(x^k, \lambda^k; \hat{\varepsilon})^T d^k \le -\hat{\varepsilon}/2 \|d^k\|^2,$$

then Proposition 4.3 ensures that $(\bar{x}, \bar{\lambda}, \rho(\bar{x}, \bar{\lambda}), \sigma(\bar{x}, \bar{\lambda}))$ is a KKT-tuple.
On the other hand, if for an infinite number of $k \in K$, p^k is the gradient related direction z^k which satisfies the inequalities

$$\begin{aligned}\nabla P(x^k, \lambda^k; \hat{\varepsilon})^T z^k &\le -\hat{\varepsilon} \Big(\|\phi_u(x^k, \lambda^k; \varepsilon)\|^2 + \|\phi_l(x^k, \lambda^k; \varepsilon)\|^2 + \|h(x^k)\|^2 \Big), \\ \nabla P(x^k, \lambda^k; \hat{\varepsilon})^T z^k &\le -c_3 \|\nabla P(x^k, \lambda^k; \hat{\varepsilon})\|^2,\end{aligned}$$

then (56) implies that

$$\nabla P(\bar{x}, \bar{\lambda}; \hat{\varepsilon}) = 0,$$
$$\phi_u(\bar{x}, \bar{\lambda}; \varepsilon) = 0, \quad \phi_l(\bar{x}, \bar{\lambda}; \varepsilon) = 0, \quad h(\bar{x}) = 0,$$

which, recalling (ii) of Proposition 2.1, show that $(\bar{x}, \bar{\lambda}, \rho(\bar{x}, \bar{\lambda}), \sigma(\bar{x}, \bar{\lambda}))$ is a KKT-tuple.

Point (iv). We recall that, by property (51b) of the line-search procedure, it results

$$\lim_{k\to\infty} \mu^k \nabla P(x^k, \lambda^k; \hat{\varepsilon})^T p^k = 0.$$

By the properties of directions d^k and z^k, it follows that

$$\lim_{k\to\infty} \mu^k p^k = 0,$$

which yields

$$\lim_{k\to\infty} \left\| \begin{pmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} \right\| = 0.$$

Now the proof of points (iv), (v) and (vi) follows reasoning as in the proof of Proposition 4.4 in [13]. □

6. Conclusions

In this paper we have shown that a nonlinear programming problem with equality constraints and bounds on the variables can be tackled by using an exact merit function based on an exact augmented Lagrangian approach for equality constraints and on an exact penalty approach for the bound constraints. Such a merit function can be used in connection with a search direction which can be computed in an efficient way by using a truncated scheme, and which is able to produce a locally and superlinearly convergent algorithm for the solution of the problem. At each iteration, the search direction requires only the solution of a linear system whose dimension is equal at most to the number of the free variables plus the number of the equality constraints. The connection between the merit function and the search direction is established by the fact that the search direction is proved to be a descent direction for the merit function, provided that the penalty parameter is small enough and that the current primal-dual pair (x^k, λ^k) belongs to a neighborhood of a pair $(\bar{x}, \bar{\lambda})$ of a KKT-tuple of the problem. This connection allows to endow the locally efficient algorithm with global convergence property. On these bases, it is possible to describe a globally and superlinearly convergent algorithm model that exploits the particular structure of the problem constraints so as to limit the computational burden as much as possible, thus making suitable the algorithm for large dimensional instances. A distinguishing feature of the algorithm model is that, due to the exactness of the merit function, the penalty parameter is updated a finite number of times, thus preventing ill-conditioning. A numerical implementation of the algorithm model ExPAL is out of the scope of this paper, and will be matter of future work. We point out that the implementation of algorithm ExPAL requires analogous effort as the implementation of an algorithm for unconstrained minimization of a nonlinear function. However, there are some points that deserve particular attention. They are: the tests at Steps 2 and 3 of the Algorithm, that are needed to assess the goodness of both the search direction and of the current value of the penalty parameter, and the particular choice of the gradient related direction z used when d is not of "sufficient" descent. The ultimate performance of the algorithm can also be sensibly influenced by the implementation of procedure ICP and in particular by the definition of the sequence $\{\eta^k\}$ needed in the implementation of PCGP and which determines the behavior in the limit of the conjugate gradient residual. All these aspects should be carefully taken into account in a practical implementation of ExPAL.

Appendix

A. Proof of Proposition 3.5

For simplicity we omit the iteration index k.

Preliminary from (32a) we get:

$$\zeta_1 \|y\|^2 \le -(d_o)^T y \le \|d_o\| \, \|y\| \,,$$

so that

$$\|y\| \le \frac{1}{\zeta_1} \|d_o\| \,.$$

Then, from the definition (29) of y, it follows that

$$\left\|\mathcal{P}\left(\nabla_x L\right)_{\mathcal{M}}\right\| \le \frac{1}{\zeta_1}\|d_o\| + \left\|\mathcal{P}(\nabla_x^2 L_{\mathcal{M},\mathcal{M}} d_v + \nabla_x^2 L_{\mathcal{M},\bar{\mathcal{M}}}(d_x)_{\bar{\mathcal{M}}})\right\|. \tag{57}$$

Point (i). We first consider the term $\|(\rho - z)_{\mathcal{L}}\|$. Recalling the definitions (2) of the multiplier functions we have:

$$\begin{aligned}((l-x)^2 + (x-u)^2)_{\mathcal{L}} \circ (\rho - z)_{\mathcal{L}} &= (x-u)^2_{\mathcal{L}} \circ \nabla_x L_{\mathcal{L}} - ((l-x)^2 + (x-u)^2)_{\mathcal{L}} \circ z_{\mathcal{L}} \\ &= (x-u)^2_{\mathcal{L}} \circ (\nabla_x L - z)_{\mathcal{L}} - (l-x)^2_{\mathcal{L}} \circ z_{\mathcal{L}} \\ &= -(x-u)^2_{\mathcal{L}} \circ (\nabla_x^2 L d_x + \nabla h d_\lambda)_{\mathcal{L}} - (l-x)^2_{\mathcal{L}} \circ z_{\mathcal{L}}\end{aligned}$$

so that we can write:

$$(\rho - z)_{\mathcal{L}} = -\big((l-x)^2 + (x-u)^2\big)^{-1}_{\mathcal{L}} \circ \big((x-u)^2_{\mathcal{L}} \circ (\nabla_x^2 L d_x + \nabla h d_\lambda)_{\mathcal{L}} + z_{\mathcal{L}} \circ (d_x^2)_{\mathcal{L}}\big).$$

From the last equation we get that there exists $\tilde{c}_1$ such that

$$\|(\rho - z)_{\mathcal{L}}\| \le \tilde{c}_1 \|d\|.$$

We now consider the term $\|\rho_{\mathcal{M}}\|$. We can write:

$$\begin{aligned}\rho_{\mathcal{M}} &= w_{\mathcal{M}} \circ (l-x)^{-2}_{\mathcal{M}} \circ (\nabla_x L)_{\mathcal{M}} \\ &= w_{\mathcal{M}} \circ (l-x)^{-2}_{\mathcal{M}} \circ (E_{\mathcal{M},\mathcal{M}} - \mathcal{P})(\nabla_x L)_{\mathcal{M}} \\ &\quad + w_{\mathcal{M}} \circ (l-x)^{-2}_{\mathcal{M}} \circ \mathcal{P}(\nabla_x L)_{\mathcal{M}}.\end{aligned} \tag{58}$$

From (23) and (28) we get:

$$(E_{\mathcal{M},\mathcal{M}} - \mathcal{P})(\nabla_x L)_{\mathcal{M}} = -\nabla h_{\mathcal{M}} \left((\nabla h_{\mathcal{M}})^T \nabla h_{\mathcal{M}}\right)^{-1} (\nabla h_{\mathcal{M}})^T \nabla_x^2 L_{\mathcal{M}} d_x - \nabla h d_\lambda.$$

Substituting in (58) and recalling (57), we can conclude that there exists $\tilde{c}_2$ such that:

$$\|\rho_{\mathcal{M}}\| \leq \tilde{c}_2 \|d\|.$$

Then we consider the term $\|\rho_{\mathcal{U}}\|$. We have:

$$\rho_{\mathcal{U}} = w_{\mathcal{U}} \circ (l-x)_{\mathcal{U}}^{-2} \circ (\nabla_x L)_{\mathcal{U}} = [(l-x)_{\mathcal{U}}^2 + (x-u)_{\mathcal{U}}^2]^{-1} \circ (\nabla_x L)_{\mathcal{U}} \circ (d_x)_{\mathcal{U}}^2,$$

and therefore we can conclude that there exists $\tilde{c}_3$ such that:

$$\|\rho_{\mathcal{U}}\| \leq \tilde{c}_3 \|d\|.$$

By treating in a similar way the terms $\|(\sigma+z)_{\mathcal{U}}\|$, $\|\sigma_{\mathcal{L}}\|$, $\|\sigma_{\mathcal{M}}\|$ we get the result.

Point (ii). As concerns $\|\nabla_x L - \rho + \sigma\|$ we have that

$$\begin{aligned}
\|\nabla_x L - \rho + \sigma\| &= \|(\nabla_x L - \rho + \sigma)_{\mathcal{L}}\| + \|(\nabla_x L - \rho + \sigma)_{\mathcal{M}}\| + \|(\nabla_x L - \rho + \sigma)_{\mathcal{U}}\| \\
&= \|\left(\nabla_x L + \sigma\right)_{\mathcal{L}} - z_{\mathcal{L}} + (z_{\mathcal{L}} - \rho_{\mathcal{L}})\| \\
&+ \|\left(\nabla_x L - \rho + \sigma\right)_{\mathcal{M}}\| \\
&+ \|\left(\nabla_x L - \rho\right)_{\mathcal{U}} - z_{\mathcal{U}} + (z_{\mathcal{U}} + \sigma_{\mathcal{U}})\|,
\end{aligned}$$

so that

$$\begin{aligned}
\|\nabla_x L - \rho + \sigma\| &\leq \|\left(\nabla_x L + \sigma\right)_{\mathcal{L}} - z_{\mathcal{L}} + (z_{\mathcal{L}} - \rho_{\mathcal{L}})\| \\
&+ \|\mathcal{P}\left(\nabla_x L - \rho + \sigma\right)_{\mathcal{M}}\| \\
&+ \|(E_{\mathcal{M},\mathcal{M}} - \mathcal{P})\left(\nabla_x L - \rho + \sigma\right)_{\mathcal{M}}\| \\
&+ \|\left(\nabla_x L - \rho\right)_{\mathcal{U}} - z_{\mathcal{U}} + (z_{\mathcal{U}} + \sigma_{\mathcal{U}})\|.
\end{aligned}$$

Let us consider the term $\|\left(\nabla_x L + \sigma\right)_{\mathcal{L}} - z_{\mathcal{L}} + (z_{\mathcal{L}} - \rho_{\mathcal{L}})\|$, for which it holds that

$$\begin{aligned}
&\|\left(\nabla_x L + \sigma\right)_{\mathcal{L}} - z_{\mathcal{L}} + (z_{\mathcal{L}} - \rho_{\mathcal{L}})\| \\
&\quad \leq \|\left(\nabla_x L\right)_{\mathcal{L}} - z_{\mathcal{L}}\| + \|z_{\mathcal{L}} - \rho_{\mathcal{L}}\| + \|\sigma_{\mathcal{L}}\|.
\end{aligned}$$

By the expression of z we have

$$\begin{aligned}
&\|\left(\nabla_x L\right)_{\mathcal{L}} - z_{\mathcal{L}}\| + \|z_{\mathcal{L}} - \rho_{\mathcal{L}}\| + \|\sigma_{\mathcal{L}}\| \\
&\quad = \|\left(\nabla_x^2 L d_x + \nabla h d_\lambda\right)_{\mathcal{L}}\| + \|z_{\mathcal{L}} - \rho_{\mathcal{L}}\| + \|\sigma_{\mathcal{L}}\| \\
&\quad \leq c_1\|d_x\| + c_2\|d_\lambda\| + \|z_{\mathcal{L}} - \rho_{\mathcal{L}}\| + \|\sigma_{\mathcal{L}}\|,
\end{aligned}$$

so that we get:

$$\begin{aligned}
&\|\left(\nabla_x L + \sigma\right)_{\mathcal{L}} - z_{\mathcal{L}} + (z_{\mathcal{L}} - \rho_{\mathcal{L}})\| \\
&\quad \leq c_1\|d_x\| + c_2\|d_\lambda\| + \|z_{\mathcal{L}} - \rho_{\mathcal{L}}\| + \|\sigma_{\mathcal{L}}\|.
\end{aligned}$$

In a similar way we get for the term $\|\left(\nabla_x L - \rho\right)_{\mathcal{U}} - z_{\mathcal{U}} + (z_{\mathcal{U}} + \sigma_{\mathcal{U}})\|$ that:

$$\begin{aligned}
&\|\left(\nabla_x L - \rho\right)_{\mathcal{U}} - z_{\mathcal{U}} + (z_{\mathcal{U}} + \sigma_{\mathcal{U}})\| \\
&\quad \leq c_3\|d_x\| + c_4\|d_\lambda\| + \|z_{\mathcal{U}} + \sigma_{\mathcal{U}}\| + \|\rho_{\mathcal{U}}\|.
\end{aligned}$$

As concerns the term $\|\mathcal{P}\left(\nabla_x L - \rho + \sigma\right)_{\mathcal{M}}\|$, by (57), we can write

$$\|\mathcal{P}\left(\nabla_x L - \rho + \sigma\right)_{\mathcal{M}}\| \leq c_5(\|d_x\| + \|\sigma_{\mathcal{M}}\| + \|\rho_{\mathcal{M}}\|).$$

Finally we consider the term $\|(E_{\mathcal{M},\mathcal{M}} - \mathcal{P})\left(\nabla_x L - \rho + \sigma\right)_{\mathcal{M}}\|$ for which, by (57) and the expression of d_λ, we can write

$$\|(E_{\mathcal{M},\mathcal{M}} - \mathcal{P})\left(\nabla_x L - \rho + \sigma\right)_{\mathcal{M}}\| \leq c_6(\|d_x\| + \|d_\lambda\| + \|\sigma_{\mathcal{M}}\| + \|\rho_{\mathcal{M}}\|).$$

Thus, a constant c_7 exists such that

$$\|\nabla_x L - \rho + \sigma\| \leq c_7\left(\|d_x\| + \|d_\lambda\| + \|(z-\rho)_{\mathcal{L}}\| + \|\sigma_{\mathcal{L}\cup\mathcal{M}}\| + \|(z+\sigma)_{\mathcal{U}}\| + \|\rho_{\mathcal{U}\cup\mathcal{M}}\|\right).$$

Now, the result follows by recalling point (i). □

B. Expressions of ψ_1 and ψ_2

The directional derivative of the exact-augmented Lagrangian function $P(x, \lambda; \varepsilon)$ along direction d can be written as in (42) where ψ_1 and ψ_2 denote the following quantities:

$$
\begin{aligned}
\psi_1(x, \lambda, d; \varepsilon) = &-\frac{1}{\varepsilon}\|h(x)\|^2 \quad - d_\lambda^T \nabla h(x)^T d_x \\
&-\frac{1}{\varepsilon} d_x^T \operatorname{diag}\{r(x)^{-1}\} E_{\mathcal{L}}^T (d_x)_{\mathcal{L}} \quad -\frac{1}{\varepsilon} d_x^T \operatorname{diag}\{s(x)^{-1}\} E_{\mathcal{U}}^T (d_x)_{\mathcal{U}} \\
&+\big(d_x^T \nabla_x \rho(x, \lambda) + d_\lambda^T \nabla_\lambda \rho(x, \lambda)\big) E_{\mathcal{L}}^T (d_x)_{\mathcal{L}} \\
&-\big(d_x^T \nabla_x \sigma(x, \lambda) + d_\lambda^T \nabla_\lambda \sigma(x, \lambda)\big) E_{\mathcal{U}}^T (d_x)_{\mathcal{U}} \\
&-\varepsilon\big(d_x^T \nabla_x \rho(x, \lambda) + d_\lambda^T \nabla_\lambda \rho(x, \lambda)\big) E_{\mathcal{U}\cup\mathcal{M}}^T (r(x) \circ \rho(x, \lambda))_{\mathcal{U}\cup\mathcal{M}} \\
&-\varepsilon\big(d_x^T \nabla_x \sigma(x, \lambda) + d_\lambda^T \nabla_\lambda \sigma(x, \lambda)\big) E_{\mathcal{L}\cup\mathcal{M}}^T (s(x) \circ \sigma(x, \lambda))_{\mathcal{L}\cup\mathcal{M}} \\
&+\frac{\varepsilon}{2}\Big(d_x^T E_{\mathcal{U}\cup\mathcal{M}}^T (w(x) \circ (l - x)^{-2} \circ \rho(x, \lambda))_{\mathcal{U}\cup\mathcal{M}} \\
&\qquad - d_x^T E_{\mathcal{L}\cup\mathcal{M}}^T (w(x) \circ (x - u)^{-2} \circ \sigma(x, \lambda))_{\mathcal{L}\cup\mathcal{M}}\Big) \\
&+d_x^T \Big(E_{\mathcal{U}\cup\mathcal{M}}^T (\rho(x, \lambda))_{\mathcal{U}\cup\mathcal{M}} - E_{\mathcal{L}\cup\mathcal{M}}^T (\sigma(x, \lambda))_{\mathcal{L}\cup\mathcal{M}}\Big) \\
&-\frac{1}{2\varepsilon} d_x^T \operatorname{diag}\{r(x)^{-2}\} E_{\mathcal{L}}^T (d_x \circ (l - x))_{\mathcal{L}} + \frac{1}{2\varepsilon} d_x^T \operatorname{diag}\{s(x)^{-2}\} E_{\mathcal{U}}^T (d_x \circ (u - x))_{\mathcal{U}} \\
&-2\left\|(\nabla h(x))_{\mathcal{M}}^T \big(w(x) \circ (\nabla_x^2 L(x, \lambda) d_x + \nabla h(x) d_\lambda)\big)_{\mathcal{M}} + \|h(x)\|^2 d_\lambda\right\|^2,
\end{aligned}
$$

$$\psi_2(x,\lambda,d;\tau) = 2R(x,\lambda,d,\tau) \tag{59}$$
$$+2\Bigg(\Big((d_\lambda^T\nabla h(x)^T + d_x^T\nabla_x^2 L(x,\lambda))\mathrm{diag}\{w(x)\}\nabla h(x) + d_\lambda^T\|h(x)\|^2\Big)$$
$$\cdot\Big(\nabla h(x)^T\Big[E_{\mathcal{U}}^T\Big((l-x)^2\circ\rho(x,\lambda)\Big)_{\mathcal{U}} - E_{\mathcal{L}}^T\Big((u-x)^2\circ\sigma(x,\lambda)\Big)_{\mathcal{L}}\Big]$$
$$-(h(x)^T\nabla h(x)^T d_x)(\lambda - d_\lambda)\Big)$$
$$+\Big(\nabla h(x)_{\mathcal{M}}^T[(w(x)\circ\big(\nabla_x^2 L(x,\lambda)d_x + \nabla h(x)d_\lambda)]_{\mathcal{M}}\Big)^T$$
$$\cdot\Big(\nabla h(x)_{\mathcal{M}}^T[(l-x)^2\circ\rho(x,\lambda)]_{\mathcal{M}} + \|h(x)\|^2 d_\lambda\Big)\Bigg)$$
$$+2d_x^T\Bigg(2\nabla h(x)h(x)\lambda^T + \sum_{j=1}^m \nabla^2 h_j(x)\left((l-x)^2\circ\rho(x,\lambda)\right)i_j^T$$
$$-2\mathrm{diag}\{w(x)\circ(l-x)^{-1}\circ\rho(x,\lambda)\}\nabla h(x)$$
$$-2\mathrm{diag}\{w(x)\circ(x-u)^{-1}\circ\sigma(x,\lambda)\}\nabla h(x)\Bigg)$$
$$\cdot\Bigg((\nabla h(x))_{\mathcal{U}}^T\left[(l-x)^2\circ\rho(x,\lambda)\right]_{\mathcal{U}} - (\nabla h(x))_{\mathcal{L}}^T\left[(u-x)^2\circ\sigma(x,\lambda)\right]_{\mathcal{L}}$$
$$+(\nabla h(x))_{\mathcal{M}}^T\left[(l-x)^2\circ\rho(x,\lambda)\right]_{\mathcal{M}} - (h(x)^T\nabla h(x)^T d_x)\lambda\Bigg),$$

where the scalar $R(x,\lambda,d,\tau)$ satisfies

$$R(x,\lambda,d,\tau) \le \xi_1(x,\lambda)\|d\|\|\tau\|, \tag{60}$$

with $\xi_1(x,\lambda)$ a positive continuous function.

C. Proof of Proposition 4.1

For simplicity we omit the iteration index k when it is not required.

We note that, from (17), we can write for $h(x)$

$$h(x) = -(\nabla h(x))^T d_x. \tag{61}$$

Moreover, recalling (4), (18), we can write for the terms $\phi_l(x,\lambda;\varepsilon)$ and $\phi_u(x,\lambda;\varepsilon)$:

$$\begin{aligned}\phi_l(x,\lambda;\varepsilon) \quad &= E_{\mathcal{L}}^T(\max\{l-x, -\varepsilon r(x)\circ\rho(x,\lambda)\})_{\mathcal{L}}\\ &+E_{\mathcal{U}\cup\mathcal{M}}^T(\max\{l-x,-\varepsilon r(x)\circ\rho(x,\lambda)\})_{\mathcal{U}\cup\mathcal{M}}\\ &= E_{\mathcal{L}}^T(d_x)_{\mathcal{L}} - \varepsilon E_{\mathcal{U}\cup\mathcal{M}}^T(r(x)\circ\rho(x,\lambda))_{\mathcal{U}\cup\mathcal{M}},\end{aligned} \tag{62}$$

$$\begin{aligned}\phi_u(x,\lambda;\varepsilon) \quad &= E_{\mathcal{U}}^T(\max\{x-u,-\varepsilon s(x)\circ\sigma(x,\lambda)\})_{\mathcal{U}}\\ &+E_{\mathcal{L}\cup\mathcal{M}}^T(\max\{x-u,-\varepsilon s(x)\circ\sigma(x,\lambda)\})_{\mathcal{L}\cup\mathcal{M}}\\ &= -E_{\mathcal{U}}^T(d_x)_{\mathcal{U}} - \varepsilon E_{\mathcal{L}\cup\mathcal{M}}^T(s(x)\circ\sigma(x,\lambda))_{\mathcal{L}\cup\mathcal{M}}.\end{aligned} \tag{63}$$

Thus, by using (61), (62), (63) in the expression (41) of the directional derivative, we have, after some reordering:

$$d_x^T \nabla_x P(x,\lambda;\varepsilon) + d_\lambda{}^T \nabla_\lambda P(x,\lambda;\varepsilon) = -\frac{1}{\varepsilon}\|h(x)\|^2 \quad - d_\lambda^T \nabla h(x)^T d_x \tag{64}$$
$$-\frac{1}{\varepsilon} d_x^T \operatorname{diag}\{r(x)^{-1}\} E_{\mathcal{L}}^T (d_x)_{\mathcal{L}} \quad -\frac{1}{\varepsilon} d_x^T \operatorname{diag}\{s(x)^{-1}\} E_{\mathcal{U}}^T (d_x)_{\mathcal{U}}$$
$$+\big(d_x^T \nabla_x \rho(x,\lambda) + d_\lambda^T \nabla_\lambda \rho(x,\lambda)\big) E_{\mathcal{L}}^T (d_x)_{\mathcal{L}}$$
$$-\big(d_x^T \nabla_x \sigma(x,\lambda) + d_\lambda^T \nabla_\lambda \sigma(x,\lambda)\big) E_{\mathcal{U}}^T (d_x)_{\mathcal{U}}$$
$$-\varepsilon\big(d_x^T \nabla_x \rho(x,\lambda) + d_\lambda^T \nabla_\lambda \rho(x,\lambda)\big) E_{\mathcal{U}\cup\mathcal{M}}^T (r(x)\circ\rho(x,\lambda))_{\mathcal{U}\cup\mathcal{M}}$$
$$-\varepsilon\big(d_x^T \nabla_x \sigma(x,\lambda) + d_\lambda^T \nabla_\lambda \sigma(x,\lambda)\big) E_{\mathcal{L}\cup\mathcal{M}}^T (s(x)\circ\sigma(x,\lambda))_{\mathcal{L}\cup\mathcal{M}}$$
$$+\frac{\varepsilon}{2}\Big(d_x^T E_{\mathcal{U}\cup\mathcal{M}}^T (w(x)\circ(l-x)^{-2}\circ\rho(x,\lambda))_{\mathcal{U}\cup\mathcal{M}}$$
$$-d_x^T E_{\mathcal{L}\cup\mathcal{M}}^T (w(x)\circ(x-u)^{-2}\circ\sigma(x,\lambda))_{\mathcal{L}\cup\mathcal{M}}\Big)$$
$$+d_x^T\Big(E_{\mathcal{U}\cup\mathcal{M}}^T(\rho(x,\lambda))_{\mathcal{U}\cup\mathcal{M}} - E_{\mathcal{L}\cup\mathcal{M}}^T(\sigma(x,\lambda))_{\mathcal{L}\cup\mathcal{M}}\Big)$$
$$-\frac{1}{2\varepsilon} d_x^T \operatorname{diag}\{r(x)^{-2}\} E_{\mathcal{L}}^T (d_x\circ(l-x))_{\mathcal{L}} + \frac{1}{2\varepsilon} d_x^T \operatorname{diag}\{s(x)^{-2}\} E_{\mathcal{U}}^T (d_x\circ(u-x))_{\mathcal{U}}$$
$$+2\Big(d_x^T \nabla_x \varphi(x,\lambda) + d_\lambda^T \nabla_\lambda \varphi(x,\lambda)\Big)\varphi(x,\lambda).$$

By recalling the expressions of $\nabla_x\varphi$ and $\nabla_\lambda\varphi$ in (3), the last term of the r.h.s. of equation (64), omitting the factor 2, can be rewritten as

$$(d_\lambda^T \nabla_\lambda \varphi(x,\lambda) + d_x^T \nabla_x \varphi(x,\lambda))\varphi(x,\lambda) =$$
$$\left(\Big(d_\lambda^T \nabla h(x)^T + d_x^T \nabla_x^2 L(x,\lambda)\Big)\operatorname{diag}\{w(x)\}\nabla h(x) + d_\lambda^T \|h(x)\|^2\right)\varphi(x,\lambda) \tag{65a}$$
$$+d_x^T\left(2\nabla h(x) h(x)\lambda^T + \sum_{j=1}^m \nabla^2 h_j(x) w(x)\circ\nabla_x L(x,\lambda) i_j^T\right. \tag{65b}$$
$$\left. -2\operatorname{diag}\{w(x)^2\circ\big[(l-x)^{-3}-(x-u)^{-3}\big]\circ\nabla_x L(x,\lambda)\}\nabla h(x)\right)\varphi(x,\lambda). \tag{65c}$$

Let us consider the first factor of term (65a) in the above equation, its transpose can be rewritten as

$$\nabla h(x)^T\Big(w(x)\circ\big(\nabla_x^2 L(x,\lambda)d_x + \nabla h(x)d_\lambda\big)\Big) + \|h(x)\|^2 d_\lambda =$$
$$\nabla h(x)^T\Big(E_{\mathcal{M}}^T[w(x)\circ\big(\nabla_x^2 L(x,\lambda)d_x + \nabla h(x)d_\lambda\big)]_{\mathcal{M}}$$
$$+E_{\bar{\mathcal{M}}}^T[w(x)\circ\big(\nabla_x^2 L(x,\lambda)d_x + \nabla h(x)d_\lambda\big)]_{\bar{\mathcal{M}}}\Big) + \|h(x)\|^2 d_\lambda,$$

while $\varphi(x,\lambda)$ can be rewritten as

$$\varphi(x,\lambda) = \nabla h(x)^T\Big(w(x)\circ\nabla_x L(x,\lambda)\Big) + \|h(x)\|^2\lambda =$$
$$\nabla h(x)^T\Big(E_{\mathcal{M}}^T[w(x)\circ\nabla_x L(x,\lambda)]_{\mathcal{M}}$$
$$+E_{\bar{\mathcal{M}}}^T[w(x)\circ\nabla_x L(x,\lambda)]_{\bar{\mathcal{M}}}\Big) + \|h(x)\|^2\lambda.$$

By using the two preceding relations, by adding and subtracting $\|h(x)\|^2 d_\lambda$ and by (27) and (28)

$$[\nabla_x L(x,\lambda)]_{\mathcal{M}} = -\big[\nabla_x^2 L(x,\lambda)d_x + \nabla h(x)d_\lambda\big]_{\mathcal{M}} + \tau$$

the term in (65a) can be rewritten as:

$$\begin{aligned}&\Big((d_\lambda^T\nabla h(x)^T + d_x^T\nabla_x^2 L(x,\lambda))\mathrm{diag}\{w(x)\}\nabla h(x) + d_\lambda^T\|h(x)\|^2\Big)\varphi(x,\lambda) = \\ &-\Big\|(\nabla h(x))_{\mathcal{M}}^T[w(x)\circ\big(\nabla_x^2 L(x,\lambda)d_x + \nabla h(x)d_\lambda)]_{\mathcal{M}} + \|h(x)\|^2 d_\lambda\Big\|^2 \\ &+\Big((d_\lambda^T\nabla h(x)^T + d_x^T\nabla_x^2 L(x,\lambda))\mathrm{diag}\{w(x)\}\nabla h(x) + d_\lambda^T\|h(x)\|^2\Big) \\ &\cdot\Big((\nabla h(x))_{\bar{\mathcal{M}}}^T[w(x)\circ\nabla_x L(x,\lambda)]_{\bar{\mathcal{M}}} + \|h(x)\|^2(\lambda - d_\lambda)\Big) \\ &+\Big((\nabla h(x))_{\bar{\mathcal{M}}}^T[w(x)\circ\big(\nabla_x^2 L(x,\lambda)d_x + \nabla h(x)d_\lambda)]_{\bar{\mathcal{M}}}\Big)^T \\ &\cdot\Big((\nabla h(x))_{\mathcal{M}}^T[w(x)\circ\nabla_x L(x,\lambda)]_{\mathcal{M}} + \|h(x)\|^2 d_\lambda\Big) + R(x,\lambda,d,\tau),\end{aligned} \tag{66}$$

where the scalar $R(x,\lambda,d,\tau)$ satisfies

$$R(x,\lambda,d,\tau) \le \xi_1(x,\lambda)\|d\|\|\tau\|, \tag{67}$$

with $\xi_1(x,\lambda)$ a positive continuous function.

Thus, by substituting (66) in (65a) we obtain

$$\begin{aligned}&(d_\lambda^T\nabla_\lambda\varphi(x,\lambda) + d_x^T\nabla_x\varphi(x,\lambda))\varphi(x,\lambda) = \\ &-\Big\|(\nabla h(x))_{\mathcal{M}}^T[w(x)\circ\big(\nabla_x^2 L(x,\lambda)d_x + \nabla h(x)d_\lambda)]_{\mathcal{M}} + \|h(x)\|^2 d_\lambda\Big\|^2 \\ &+\Big((d_\lambda^T\nabla h(x)^T + d_x^T\nabla_x^2 L(x,\lambda))\mathrm{diag}\{w(x)\}\nabla h(x) + d_\lambda^T\|h(x)\|^2\Big) \\ &\cdot\Big((\nabla h(x))_{\bar{\mathcal{M}}}^T[w(x)\circ\nabla_x L(x,\lambda)]_{\bar{\mathcal{M}}} + \|h(x)\|^2(\lambda - d_\lambda)\Big) \\ &+\Big((\nabla h(x))_{\bar{\mathcal{M}}}^T[w(x)\circ\big(\nabla_x^2 L(x,\lambda)d_x + \nabla h(x)d_\lambda)]_{\bar{\mathcal{M}}}\Big)^T \\ &\cdot\Big((\nabla h(x))_{\mathcal{M}}^T[w(x)\circ\nabla_x L(x,\lambda)]_{\mathcal{M}} + \|h(x)\|^2 d_\lambda\Big) + R(x,\lambda,d,\tau) \\ &+d_x^T\Bigg(2\nabla h(x)h(x)\lambda^T + \sum_{j=1}^m \nabla^2 h_j(x)w(x)\circ\nabla_x L(x,\lambda)i_j^T \\ &-2\mathrm{diag}\{(l-x)^{-3}\circ w(x)^2\circ(x-u)^{-3}\circ\big[(x-u)^3\ -(l-x)^3\big]\circ\nabla_x L(x,\lambda)\}\nabla h(x)\Bigg)\varphi(x,\lambda).\end{aligned}$$

Then, we can verify that the directional derivative of the merit function can be written as in (42)

$$d_x^T\nabla_x P(x\lambda;\varepsilon) + {d_\lambda}^T\nabla_\lambda P(x,\lambda;\varepsilon) = \psi_1(x,\lambda,d;\varepsilon) + \psi_2(x,\lambda,d;\tau),$$

where ψ_1 and ψ_2 are the quantities given in Appendix B.

Now we can prove points (i) and (ii) of the proposition.

Point (i) of Proposition 4.1.

First we consider the term $d_\lambda^T\nabla h(x)^T d_x$ that appears in ψ_1. By using (22) and the fact that $\nabla h^T d_x = -h(x)$, we obtain

$$d_\lambda^T\nabla h(x)^T d_x = -d_\lambda^T h(x) = d_\lambda^T\nabla h_{\mathcal{M}}^T d_v + d_\lambda^T\nabla h_{\bar{\mathcal{M}}}^T(d_x)_{\bar{\mathcal{M}}}. \tag{68}$$

Then we consider the term

$$d_x^T\Big(E_{\mathcal{U}\cup\mathcal{M}}^T(\rho(x,\lambda))_{\mathcal{U}\cup\mathcal{M}} - E_{\mathcal{L}\cup\mathcal{M}}^T(\sigma(x,\lambda))_{\mathcal{L}\cup\mathcal{M}}\Big).$$

It can be rewritten as

$$(d_x)_{\mathcal{U}}^T(\rho(x,\lambda)_{\mathcal{U}} - (d_x)_{\mathcal{L}}^T\sigma(x,\lambda)_{\mathcal{L}} + (d_x)_{\mathcal{M}}^T(\rho(x,\lambda) - \sigma(x,\lambda))_{\mathcal{M}}.$$

Noting that by the definition of $\rho(x,\lambda)$ and $\sigma(x,\lambda)$ it results,

$$\rho(x,\lambda) - \sigma(x,\lambda) = \nabla_x L(x,\lambda),$$

we obtain that

$$(d_x)^T_{\mathcal{M}}(\rho(x,\lambda)-\sigma(x,\lambda))_{\mathcal{M}} = (d_x)^T_{\mathcal{M}}(\nabla_x L)_{\mathcal{M}}. \tag{69}$$

We can write

$$(d_x)^T_{\mathcal{M}}(\nabla_x L)_{\mathcal{M}} = (d_x)^T_{\mathcal{M}}\big(P(\nabla_x L)_{\mathcal{M}} + (E_{\mathcal{M},\mathcal{M}} - P)(\nabla_x L)_{\mathcal{M}}\big). \tag{70}$$

Now we note that, by using properties (32a) and (32b), we have

$$d_o{}^T y \le -\frac{\zeta_1}{\zeta_2^2}\|d_o\|^2,$$

from which, by recalling the definition (29) of y and the fact that $\mathcal{P}d_o = \mathcal{P}(d_x)_{\mathcal{M}}$ we get:

$$(d_x)^T_{\mathcal{M}}\mathcal{P}(\nabla_x L)_{\mathcal{M}} \le -\frac{\zeta_1}{\zeta_2^2}\|d_o\|^2 - d_o^T\mathcal{P}\big(\nabla_x^2 L_{\mathcal{M},\mathcal{M}} d_v + \nabla_x^2 L_{\mathcal{M},\bar{\mathcal{M}}}(d_x)_{\bar{\mathcal{M}}}\big). \tag{71}$$

Moreover, by recalling (25) we have

$$\begin{aligned}(d_x)^T_{\mathcal{M}}(E_{\mathcal{M},\mathcal{M}} - \mathcal{P})(\nabla_x L)_{\mathcal{M}} = -(d_x)^T_{\mathcal{M}}(E_{\mathcal{M},\mathcal{M}} - \mathcal{P})\Big(\nabla_x^2 L_{\mathcal{M}} d_x + \nabla h_{\mathcal{M}} d_\lambda\Big) = \\ -(d_v)^T\Big(\nabla_x^2 L_{\mathcal{M}} d_x + \nabla h_{\mathcal{M}} d_\lambda\Big).\end{aligned} \tag{72}$$

Then we have from (70), (71), (72):

$$\begin{aligned}(d_x)^T_{\mathcal{M}}(\nabla_x L)_{\mathcal{M}} \le -\frac{\zeta_1}{\zeta_2^2}\|d_o\|^2 - 2d_o{}^T\nabla_x^2 L_{\mathcal{M},\mathcal{M}} d_v - d_v{}^T\nabla_x^2 L_{\mathcal{M},\mathcal{M}} d_v \\ -(d_x)^T_{\mathcal{M}}\nabla_x^2 L_{\mathcal{M},\bar{\mathcal{M}}}(d_x)_{\bar{\mathcal{M}}} - d_v^T\nabla h_{\mathcal{M}} d_\lambda.\end{aligned} \tag{73}$$

Taking into account (68), (69), (73), we obtain for ψ_1:

$$\begin{aligned}\psi_1 \le & -\frac{\zeta_1}{\zeta_2^2}\|d_o\|^2 - 2d_v{}^T(\nabla_x^2 L_{\mathcal{M},\mathcal{M}} d_o + \nabla h_{\mathcal{M}} d_\lambda) \\ & -d_v{}^T\nabla_x^2 L_{\mathcal{M},\mathcal{M}} d_v - d_\lambda^T\nabla h^T_{\bar{\mathcal{M}}}(d_x)_{\bar{\mathcal{M}}} \\ & -(d_x)^T_{\mathcal{M}}\nabla_x^2 L^T_{\mathcal{M},\bar{\mathcal{M}}}(d_x)_{\bar{\mathcal{M}}} \\ & +(d_x)^T_{\mathcal{U}}\rho(x,\lambda)_{\mathcal{U}} - (d_x)^T_{\mathcal{L}}\sigma(x,\lambda)_{\mathcal{L}} \\ & -\frac{1}{\varepsilon}\|h(x)\|^2 - \frac{1}{\varepsilon}d_x^T\mathrm{diag}\{r(x)^{-1}\}\left(E^T_{\mathcal{L}}(d_x)_{\mathcal{L}} + \frac{1}{2}\mathrm{diag}\{r(x)^{-1}\}E^T_{\mathcal{L}}(d_x\circ(l-x))_{\mathcal{L}}\right) \\ & -\frac{1}{\varepsilon}d_x^T\mathrm{diag}\{s(x)^{-1}\}\left(E^T_{\mathcal{U}}(d_x)_{\mathcal{U}} + \frac{1}{2}\mathrm{diag}\{s(x)^{-1}\}E^T_{\mathcal{U}}(d_x\circ(x-u))_{\mathcal{U}}\right) \\ & +\big(d_x^T\nabla_x\rho(x,\lambda) + d_\lambda^T\nabla_\lambda\rho(x,\lambda)\big)E^T_{\mathcal{L}}(d_x)_{\mathcal{L}} \\ & -\big(d_x^T\nabla_x\sigma(x,\lambda) + d_\lambda^T\nabla_\lambda\sigma(x,\lambda)\big)E^T_{\mathcal{U}}(d_x)_{\mathcal{U}} \\ & -\varepsilon\big(d_x^T\nabla_x\rho(x,\lambda) + d_\lambda^T\nabla_\lambda\rho(x,\lambda)\big)E^T_{\mathcal{U}\cup\mathcal{M}}(r(x)\circ\rho(x,\lambda))_{\mathcal{U}\cup\mathcal{M}} \\ & -\varepsilon\big(d_x^T\nabla_x\sigma(x,\lambda) + d_\lambda^T\nabla_\lambda\sigma(x,\lambda)\big)E^T_{\mathcal{L}\cup\mathcal{M}}(s(x)\circ\sigma(x,\lambda))_{\mathcal{L}\cup\mathcal{M}} \\ & +\frac{\varepsilon}{2}\Big((d_x)^T_{\mathcal{U}\cup\mathcal{M}}(w(x)\circ(l-x)^{-2}\circ\rho(x,\lambda))_{\mathcal{U}\cup\mathcal{M}} \\ & \qquad -(d_x)^T_{\mathcal{L}\cup\mathcal{M}}(w(x)\circ(x-u)^{-2}\circ\sigma(x,\lambda))_{\mathcal{L}\cup\mathcal{M}}\Big) \\ & -2\left\|(\nabla h(x))^T_{\mathcal{M}}\big(w(x)\circ\big(\nabla_x^2 L(x,\lambda)d_x + \nabla h(x)d_\lambda\big)\big)_{\mathcal{M}} + \|h(x)\|^2 d_\lambda\right\|^2.\end{aligned} \tag{74}$$

Now we proceed by contradiction. Assume that there exist sequences $\{\varepsilon^i\}$, $\{\gamma^i\}$, $\{(x^i,\lambda^i)\} \subseteq \{(x^k,\lambda^k)\}$ and $\{d^i\} \subseteq \{d^k\}$ such that

$$\begin{aligned} &\varepsilon^i \downarrow 0, \qquad \gamma^i \downarrow 0, \qquad (x^i,\lambda^i) \to (\tilde{x},\tilde{\lambda}), && (75)\\ &d^i \neq 0, \qquad \frac{d^i}{\|d^i\|} \to \tilde{d} \neq 0, \\ &\psi_1(x^i,\lambda^i,d^i;\varepsilon^i) > -\gamma^i \|d^i\|^2, && (76)\end{aligned}$$

and, without loss of generality,

$$\mathcal{L}^i = \mathcal{L}, \qquad \mathcal{U}^i = \mathcal{U}, \qquad \mathcal{M}^i = \mathcal{M}.$$

By (74) and (76), dividing by $\|d^i\|^2$ and denoting $\tilde{d^i} = d^i/\|d^i\|$, we get

$$\begin{aligned} 0 > &-\gamma^i + \frac{\zeta_1}{\zeta_2^2}\left\|\tilde{d_o^i}\right\|^2 + 2(\tilde{d_v^i})^T(\nabla_x^2 L(x^i,\lambda^i)_{\mathcal{M},\mathcal{M}}\tilde{d_o^i} + \nabla h(x^i)_{\mathcal{M}}\tilde{d_\lambda^i}) \qquad (77)\\ &+(\tilde{d_v^i})^T\nabla_x^2 L(x^i,\lambda^i)_{\mathcal{M},\mathcal{M}}\tilde{d_v^i} + (\tilde{d_\lambda^i})^T\nabla h(x^i)_{\bar{\mathcal{M}}}^T(\tilde{d_x^i})_{\bar{\mathcal{M}}} + (\tilde{d_x^i})_{\mathcal{M}}^T\nabla_x^2 L(x^i,\lambda^i)_{\mathcal{M},\bar{\mathcal{M}}}(\tilde{d_x^i})_{\bar{\mathcal{M}}}\\ &+(\tilde{d_x^i})_{\mathcal{U}}^T\rho_{\mathcal{U}}(x^i,\lambda^i)/\|d^i\| - (\tilde{d_x^i})_{\mathcal{L}}^T\sigma_{\mathcal{L}}(x^i,\lambda^i)/\|d^i\|\\ &+\frac{1}{\varepsilon^i}\frac{\|h(x^i)\|^2}{\|d^i\|^2} + \frac{1}{\varepsilon^i}(\tilde{d_x^i})^T\operatorname{diag}\{r(x^i)^{-1}\}\left(E + \frac{1}{2}\operatorname{diag}\left\{r(x^i)^{-1}\circ(l-x^i)\right\}\right)E_{\mathcal{L}}^T(\tilde{d_x^i})_{\mathcal{L}}\\ &+\frac{1}{\varepsilon^i}(\tilde{d_x^i})^T\operatorname{diag}\{s(x^i)^{-1}\}\left(E + \frac{1}{2}\operatorname{diag}\{s(x^i)^{-1}\circ(x^i-u)\}\right)E_{\mathcal{U}}^T(\tilde{d_x^i})_{\mathcal{U}}\\ &-\big((\tilde{d_x^i})^T\nabla_x\rho(x^i,\lambda^i) + (\tilde{d_\lambda^i})^T\nabla_\lambda\rho(x^i,\lambda^i)\big)E_{\mathcal{L}}^T(\tilde{d_x^i})_{\mathcal{L}}\\ &+\big((\tilde{d_x^i})^T\nabla_x\sigma(x^i,\lambda^i) + (\tilde{d_\lambda^i})^T\nabla_\lambda\sigma(x^i,\lambda^i)\big)E_{\mathcal{U}}^T(\tilde{d_x^i})_{\mathcal{U}}\\ &+\varepsilon^i\big((\tilde{d_x^i})^T\nabla_x\rho(x^i,\lambda^i) + (\tilde{d_\lambda^i})^T\nabla_\lambda\rho(x^i,\lambda^i)\big)E_{\mathcal{U}\cup\mathcal{M}}^T(r(x^i)\circ\rho(x^i,\lambda^i))_{\mathcal{U}\cup\mathcal{M}}/\|d^i\|\\ &+\varepsilon^i\big((\tilde{d_x^i})^T\nabla_x\sigma(x^i,\lambda^i) + (\tilde{d_\lambda^i})^T\nabla_\lambda\sigma(x^i,\lambda^i)\big)E_{\mathcal{L}\cup\mathcal{M}}^T(s(x^i)\circ\sigma(x^i,\lambda^i))_{\mathcal{L}\cup\mathcal{M}}/\|d^i\|\\ &-\frac{\varepsilon^i}{2}\Big((\tilde{d_x^i})_{\mathcal{U}\cup\mathcal{M}}^T(w(x^i)\circ(l-x^i)^{-2}\circ\rho(x^i,\lambda^i))_{\mathcal{U}\cup\mathcal{M}}\\ &\qquad -(\tilde{d_x^i})_{\mathcal{L}\cup\mathcal{M}}^T(w(x^i)\circ(x^i-u)^{-2}\circ\sigma(x^i,\lambda^i))_{\mathcal{L}\cup\mathcal{M}}\Big)/\|d^i\|\\ &+2\left\|\nabla h(x^i)_{\mathcal{M}}^T\big(w(x^i)\circ(\nabla_x^2 L(x^i,\lambda^i)\tilde{d_x^i} + \nabla h(x^i)\tilde{d_\lambda^i})\big)_{\mathcal{M}} + \|h(x^i)\|^2\tilde{d_\lambda^i}\right\|^2.\end{aligned}$$

Using (75), the continuity assumption and the fact that $\{(x^i,\lambda^i)\}$ is bounded and that

$$\operatorname{diag}\{r(x^i)^{-1}\}\left(E + \frac{1}{2}\operatorname{diag}\{r(x^i)^{-1}\circ(l-x^i)\}\right), \quad \operatorname{diag}\{s(x^i)^{-1}\}\left(E + \frac{1}{2}\operatorname{diag}\{s(x^i)^{-1}\circ(x^i-u)\}\right)$$

are positive definite for $x^i \in \mathcal{S}$, we obtain from (77):

$$\begin{aligned} \lim_{i\to\infty}\frac{\|h(x^i)\|^2}{\|d^i\|^2} &= 0, && (78)\\ \lim_{i\to\infty}(\tilde{d_x^i})_{\mathcal{L}} = \lim_{i\to\infty}\frac{(l-x^i)_{\mathcal{L}}}{\|d^i\|} &= 0, && (79)\\ \lim_{i\to\infty}(\tilde{d_x^i})_{\mathcal{U}} = \lim_{i\to\infty}\frac{(x^i-u)_{\mathcal{U}}}{\|d^i\|} &= 0. && (80)\end{aligned}$$

Hence, by recalling the boundedness of $\{d^i\}$, it follows that

$$\begin{aligned} &h(\tilde{x}) = 0, && (81)\\ &\tilde{x}^i = l^i, \qquad i \in \mathcal{L}, && (82)\\ &\tilde{x}^i = u^i, \qquad i \in \mathcal{U}. && (83)\end{aligned}$$

Moreover, by the definition of $\mathcal{U}$ and $\mathcal{L}$ and by (75), we obtain, for $i \in \mathcal{M}$,

$$l^i \le \tilde{x}^i \le u^i,$$

so that $\tilde{x} \in \mathcal{F}$.
From the definition (26) of d_v and by (78)–(80), we get

$$\lim_{i\to\infty} \|\tilde{d_v^i}\| = 0. \tag{84}$$

From (77), (81), (84), (82), (83), and the boundedness of d^i, we obtain that

$$\lim_{i\to\infty} \|\tilde{d_o^i}\|^2 = 0, \tag{85}$$

$$\left\|\nabla h(\tilde{x})^T_{\mathcal{M}}[w(\tilde{x}) \circ \left(\nabla^2_x L(\tilde{x})\tilde{d}_x + \nabla h(\tilde{x})\tilde{d}_\lambda\right)]_{\mathcal{M}}\right\|^2 = 0. \tag{86}$$

Now, (85), (84), (82) and (83) imply that

$$\tilde{d}_x = 0. \tag{87}$$

Hence, from (86) we obtain

$$\left\|\nabla h(\tilde{x})^T_{\mathcal{M}} \left[\mathrm{diag}\{w(\tilde{x})\}\nabla h(\tilde{x})\right]_{\mathcal{M}} \tilde{d}_\lambda\right\|^2 = 0,$$

which, by recalling Assumption 1 and the fact that $\tilde{x} \in \mathcal{F}$, implies

$$\tilde{d}_\lambda = 0. \tag{88}$$

Finally, (87) and (88), contradict the fact that $\|\tilde{d}\| = 1$. □

Point (ii) of Proposition 4.1.

Recalling the definitions of $\rho(x, \lambda)$, $\sigma(x, \lambda)$ and $w(x)$, we can rewrite (59) in the following way:

$$\psi_2(x, \lambda, d; \tau) = 2R(x, \lambda, d, \tau^k)$$

$$+2\Bigg(\left((d_\lambda^T \nabla h(x)^T + d_x^T \nabla^2_x L(x, \lambda))\mathrm{diag}\{w(x)\}\nabla h(x) + d_\lambda^T \|h(x)\|^2\right) \tag{89a}$$

$$\cdot\Big(\nabla h(x)^T \Big[E_{\mathcal{U}}^T \left(d_x \circ (u - x) \circ w(x) \circ (x - u)^{-2} \circ \nabla_x L(x, \lambda)\right)_{\mathcal{U}} \tag{89b}$$

$$+E_{\mathcal{L}}^T \left(d_x \circ (l - x) \circ w(x) \circ (l - x)^{-2} \nabla_x L(x, \lambda)\right)_{\mathcal{L}}\Big] \tag{89c}$$

$$-(h(x)^T \nabla h(x)^T d_x)(\lambda - d_\lambda)\Big) \tag{89d}$$

$$+\left(\nabla h(x)^T_{\mathcal{M}} \Big[w(x) \circ \left(\nabla^2_x L(x, \lambda)d_x + \nabla h(x)d_\lambda\right)\Big]_{\bar{\mathcal{M}}}\right)^T \tag{89e}$$

$$\cdot\left(\nabla h(x)^T_{\mathcal{M}} \left((l - x)^2 \circ \rho(x, \lambda)\right)_{\mathcal{M}} + \|h(x)\|^2 d_\lambda\right)\Bigg) \tag{89f}$$

$$+2d_x^T \Bigg(2\nabla h(x)h(x)\lambda^T + \sum_{j=1}^{m} \nabla^2 h_j(x)w(x) \circ (\nabla_x L(x, \lambda))\, i_j^T \tag{89g}$$

$$-2\mathrm{diag}\{(l - x)^{-3} \circ w(x)^2 \circ (x - u)^{-3} \circ \left[(x - u)^3 \; -(l - x)^3\right] \circ \nabla_x L(x, \lambda)\}\nabla h(x)\Bigg) \tag{89h}$$

$$\cdot\Bigg(\nabla h(x)^T_{\mathcal{U}} \left[d_x \circ (u - x)^{-1} \circ w(x) \circ \nabla_x L(x, \lambda)\right]_{\mathcal{U}} \tag{89i}$$

$$+\nabla h(x)^E_{\mathcal{L}} \left[d_x \circ (l - x)^{-1} \circ w(x) \circ \nabla_x L(x, \lambda)\right]_{\mathcal{L}} \tag{89j}$$

$$+\nabla h(x)^T_{\mathcal{M}} \left[(l - x)^2 \circ \rho(x, \lambda)\right]_{\mathcal{M}} - (h(x)^T \nabla h(x)^T d_x)\lambda\Bigg). \tag{89k}$$

As concerns terms (89a)–(89d), by the continuity assumptions and by noting that $(l-x)_{\mathcal{L}}$, $(u-x)_{\mathcal{U}}$ and $h(x)$ go to zero as x approaches a KKT point of the problem, they can be bounded from above by the quantity

$$\tilde{\xi}_2(x,\lambda)\|d\|^2, \tag{90}$$

where $\tilde{\xi}_2(x,\lambda)$ is a nonnegative continuous functions such that $\tilde{\xi}_2(\bar{x},\bar{\lambda}) = 0$ if $(\bar{x},\bar{\lambda},\sigma(\bar{x},\bar{\lambda}),\rho(\bar{x},\bar{\lambda}))$ satisfies the KKT conditions.
As regards terms (89e)–(89f), by point (i) of Proposition 3.5, they can be bounded from above by

$$\hat{\xi}_2(x,\lambda)\|d\|^2, \tag{91}$$

where, again, $\hat{\xi}_2(x,\lambda)$ is a nonnegative continuous functions such that $\hat{\xi}_2(\bar{x},\bar{\lambda}) = 0$ if $(\bar{x},\bar{\lambda},\sigma(\bar{x},\bar{\lambda}),\rho(\bar{x},\bar{\lambda}))$ satisfies the KKT conditions.
Finally, (89g)–(89k), by the continuity assumptions and by noting that $(l-x)_{\mathcal{L}}$, $(x-u)_{\mathcal{U}}$, $h(x)$, $(\nabla_x L(x,\lambda))_{\mathcal{M}}$ go to zero when (x,λ) approaches a KKT pair of the problem, can be bounded from above by

$$\check{\xi}_2(x,\lambda)\|d\|^2, \tag{92}$$

where, $\check{\xi}_2(x,\lambda)$ is a nonnegative continuous functions such that $\check{\xi}_2(\bar{x},\bar{\lambda}) = 0$ if $(\bar{x},\bar{\lambda},\sigma(\bar{x},\bar{\lambda}),\rho(\bar{x},\bar{\lambda}))$ satisfies the KKT conditions.
Finally, by (67), (90), (91) and (92) we get

$$\psi_2(x^k,\lambda^k,d^k;\tau) \le \xi_1(x,\lambda)\|d\|\,\|\tau\| + \xi_2(x,\lambda)\|d\|^2.$$

□

References

[1] R. Andreani, E. Birgin, J. Martinez, and M. Schuverdt, "On augmented lagrangian methods with general lower-level constraints," *SIAM Journal on Optimization*, vol. 18, pp. 1286–1309, 2007.

[2] M. Bartholomew-Biggs, "On the convergence of some constrained minimization algorithm based on the recursive quadratic programming," *Journal of the Institute of Mathematics and its Applications*, vol. 21, pp. 67–82, 1978.

[3] M. Bartholomew-Biggs, "Recursive quadratic programming based on the augmented lagrangian," *Mathematical programming Study*, vol. 31, pp. 21–41, 1987.

[4] H. Benson, D. Shanno, and R. Vanderbei, "Interior-point methods for nonconvex nonlinear programming: Filter-methods and merit functions," *Computational Optimization and Applications*, vol. 23, no. 2, pp. 257–272, 2002.

[5] D. Bertsekas, *Constrained optimization and Lagrange multiplier methods.* New York, (NY) USA: Academic Press, 1982.

[6] D. Bertsekas, *Nonlinear Programming.* Massachusetts, USA: Athena Scientific, 1999.

[7] E. Birgin and J. Martinez, "Improving ultimate convergence of an augmented lagrangian method," *Optimization Methods and Software*, vol. 23, pp. 177–195, 2008.

[8] R. Byrd, N. Gould, J. Nocedal, and R. Waltz, "An algorithm for nonlinear optimization using linear programming and equality constrained subproblems," *Mathematical Programmming Ser. B*, vol. 100, pp. 27–48, 2004.

[9] R. Byrd, M. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM J. on Optimization*, vol. 9, pp. 877–900, 1999.

[10] R. Byrd, J. Nocedal, and R. Waltz, "Steering exact penalty methods for nonlinear programming," *Optimization Methods and Software*, vol. 23, pp. 197–213, 2008.

[11] C. Chin and R. Fletcher, "On the global convergence of an slp-filter algorithm that takes eqp steps," *Mathematical Programming*, vol. 96, no. 1, pp. 161–177, 2003.

[12] A. Conn, N. Gould, and P. Toint, "A globally convergent lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds," *Mathematics of Computation*, vol. 66, no. 217, pp. 261–288, 1997.

[13] G. Di Pillo, G. Liuzzi, S. Lucidi, and L. Palagi, "A truncated Newton method in an augmented lagrangian framework for nonlinear programming," *Computational Optimization and Applications*, vol. 45, pp. 311–352, 2010.

[14] G. Di Pillo, S. Lucidi, and L. Palagi, "An exact penalty-lagrangian approach for a class of constrained optimization problems with bounded variables," *Optimization*, vol. 28, pp. 129–148, 1993.

[15] G. Di Pillo, S. Lucidi, and L. Palagi, "A shifted-barrier primal-dual algorithm model for linearly constrained optimization problems," *Computational Optimization and Applications*, vol. 12, pp. 157–188, 1999.

[16] G. Di Pillo, S. Lucidi, and L. Palagi, "A superlinearly convergent primal-dual algorithm for constrained optimization problems with bounded variables," *Optimization Methods and Software*, vol. 14, pp. 49–73, 2000.

[17] F. Facchinei, G. Liuzzi, and S. Lucidi, "A truncated Newton method for the solution of large-scale inequality constrained minimization problems," *Computational Optimization and Applications*, vol. 25, pp. 85–122, 2003.

[18] F. Facchinei and S. Lucidi, "Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problem," *Journal of Optimization Theory and Applications*, vol. 85, pp. 265–289, 1995.

[19] A. Fiacco and G. McCormick, *Nonlinear Programming: Sequential Unconstrained Optimization Techniques*. New York, (NY) USA: John Wiley and Sons, 1986.

[20] R. Fletcher, N. Gould, S. Leyffer, P. Toint, and A. Wächter, "Global convergence of trust-region sqp-filter algorithms for general nonlinear programming," *SIAM Journal on Optimization*, vol. 13, no. 3, pp. 635–659, 2002.

[21] R. Fletcher and S. Leyffer, "Nonlinear programming without a penalty function," *Mathematical Programming*, vol. 91, no. 2, pp. 239–269, 2002.

[22] R. Fletcher, S. Leyffer, and P. Toint, "On the global convergence of a filter-sqp algorithm," *SIAM Journal on Optimization*, vol. 13, no. 1, pp. 44–59, 2002.

[23] C. Gonzaga, E. Karas, and M. Vanti, "A globally convergent filter method for nonlinear programming," *SIAM Journal on Optimization*, vol. 14, no. 3, pp. 646–669, 2003.

[24] L. Grippo, F. Lampariello, and S. Lucidi, "A truncated Newton method with nonmonotone line search for unconstrained optimization," *Journal of Optimization Theory and Applications*, vol. 60, pp. 401–419, 1989.

[25] M. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications*, vol. 4, pp. 303–320, 1969.

[26] H. Kleinmichel, C. Richter, and K. Schönefeld, "On the class of hybrid methods for smooth constrained optimization," *Journal of Optimization Theory and Applications*, vol. 73, pp. 465–499, 1992.

[27] S. Lucidi, "New results on a continuously differentiable exact penalty function," *SIAM Journal on Optimization*, vol. 2, pp. 558–574, 1992.

[28] R. Rockafellar, "Augmented lagrangian multiplier functions and duality in nonconvex programming," *SIAM Journal on Control and Optimization*, vol. 12, pp. 268–285, 1974.

[29] S. Ulbrich, "On the superlinear local convergence of a filter-sqp method," *Mathematical Programming*, vol. 100, pp. 217–245, 2004.

[30] A. Wächter and L. Biegler, "Line search filter methods for nonlinear programming: Motivation and global convergence," *SIAM Journal on Optimization*, vol. 16, no. 1, pp. 1–31, 2005.

[31] A. Wächter and L. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 22–57, 2006.