A primal-dual algorithm for nonlinear programming exploiting negative curvature directions

Gianni Di Pillo*, Giampaolo Liuzzi†, Stefano Lucidi*

Abstract

In this paper we propose a primal-dual algorithm for the solution of inequality constrained optimization problems. The distinguishing feature of the proposed algorithm is that of exploiting as much as possible the local non-convexity of the problem to the aim of producing a sequence of points converging to second order stationary points. In the unconstrained case this task is accomplished by computing a suitable negative curvature direction of the objective function. In the constrained case it is possible to gain analogous information by exploiting the non-convexity of a particular exact merit function. The algorithm hinges on the idea of comparing, at every iteration, the relative effects of two directions and then selecting the more promising one. The first direction conveys first order information on the problem and can be used to define a sequence of points converging toward a KKT pair of the problem. Whereas, the second direction conveys information on the local non-convexity of the problem and can be used to drive the algorithm away from regions of non-convexity. We propose a proper selection rule for these two directions which, under suitable assumptions, is able to generate a sequence of points that is globally convergent to KKT pairs that satisfy the second order necessary optimality conditions, with superlinear convergence rate if the KKT pair satisfies also the strong second order sufficiency optimality conditions.

AMS subject classification. 90C30.

Keywords: nonlinear programming, augmented Lagrangian function, primal-dual algorithm, second order necessary optimality conditions, negative curvature direction.

* Università di Roma “La Sapienza”, Dipartimento di Informatica e Sistemistica “A. Ruberti”, Via Ariosto 25, 00185 Roma, Italy
e-mail (Di Pillo): `dipillo@dis.uniroma1.it`, (Lucidi): `lucidi@dis.uniroma1.it`

† Consiglio Nazionale delle Ricerche, Istituto di Analisi dei Sistemi ed Informatica “A. Ruberti”, Viale Manzoni 30, 00185 Roma, Italy
e-mail (Liuzzi): `liuzzi@iasi.cnr.it`

1 Introduction

We consider the smooth constrained optimization problem:

$$\begin{array}{ll} \min & f(x) \\ & g(x) \leq 0, \end{array} \tag{1}$$

where $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ are three times continuously differentiable functions. However, we point out that computation of third-order derivatives is never required by the proposed algorithm.

A Karush-Kuhn-Tucker (KKT) pair for Problem (1) is a pair $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^{n+m}$ such that

$$\nabla_x L(\bar{x}, \bar{\lambda}) = 0, \quad \bar{\lambda}' g(\bar{x}) = 0, \quad g(\bar{x}) \leq 0, \quad \bar{\lambda} \geq 0, \tag{2}$$

where $L(x, \lambda) = f(x) + \lambda' g(x)$ is the Lagrangian function for Problem (1) and $\lambda \in \mathbb{R}^m$ is the KKT multiplier. The *strict complementarity* condition holds at $(\bar{x}, \bar{\lambda})$ if $\bar{\lambda}_j > 0$ for all j such that $g_j(\bar{x}) = 0$.

If the *linear independence constraints qualification* (LICQ) holds at $\bar{x}$, namely if the gradients $\nabla g_j(\bar{x})$ with $j : g_j(\bar{x}) = 0$ are linearly independent, then (2) are first order necessary optimality conditions for Problem (1).

The second order necessary optimality conditions (SONC) are satisfied in a KKT pair $(\bar{x}, \bar{\lambda})$ if

$$y' \nabla_x^2 L(\bar{x}, \bar{\lambda}) y \geq 0, \quad \forall\, y : \nabla g_j(\bar{x})' y = 0 \quad \text{with } j : g_j(\bar{x}) = 0. \tag{3}$$

The strong second order sufficiency optimality conditions (SSOSC) are satisfied in a KKT pair $(\bar{x}, \bar{\lambda})$ if

$$y' \nabla_x^2 L(\bar{x}, \bar{\lambda}) y > 0, \quad \forall\, y \neq 0 : \nabla g_j(\bar{x})' y = 0 \quad \text{with } j : g_j(\bar{x}) = 0. \tag{4}$$

Standard algorithms for constrained minimization usually generate sequences converging to KKT pairs. In this paper we define a primal-dual algorithm model, having the potential for large-scale problems, that generates a sequence $\{(x^k, \lambda^k)\}$ converging to KKT pairs satisfying also the second order necessary conditions for optimality. Of course, convergence to second order stationary points allows us to better select among the points candidate to be solutions of Problem (1).

Trust-region algorithms convergent to second order stationary points have been developed for equality constrained and box constrained problems [2, 4, 7, 13, 26, 29, 31]. In [5] an interior point primal-dual trust-region method for problems with general inequality constraints and linear equality constraints has been proposed. In [30] an infeasible interior point method based on a trust region strategy has been proposed that uses a log-barrier function for the slack variables. Line search algorithms have been proposed at the beginning for the linear inequality constrained case [19, 23], and then also for the more complex nonlinear inequality constrained Problem (1). In particular, in [27] the inequality constrained problem is reduced, by the introduction of slack variables, to an equality constrained one which is then dealt with by means of an exact penalty function. In [1], a negative curvature Armijo type linesearch approach is used in connection with a sequential penalty approach. In [17], a curvilinear search approach has been proposed in connection with an

exact penalty function. More recently, in [24] an interior point method with a curvilinear search has been proposed which uses negative curvature directions in connection with an augmented Lagrangian function with the additional restriction on the infeasibility of the current iterate.
Of course, the definition of algorithms converging to second order points needs the use of second order information of the constrained problem that requires additional computational burden with respect to first order convergent algorithms.
This paper aims to combine the use of an exact augmented Lagrangian function, as described in [12], with the idea presented in [18] where an adaptive strategy for the selection of the search direction is used. Indeed our algorithm scheme belongs to the class of linesearch methods and it is based on the unconstrained reformulation of the constrained problem by means of an exact augmented Lagrangian function.

The paper is organized as follows. In section 2 we describe the exact augmented Lagrangian function L_a employed in the paper. In particular we use the exact augmented Lagrangian function L_a studied in [11], where it is shown that the original constrained Problem (1) is equivalent to the unconstrained minimization of L_a for sufficiently small values of a penalty parameter ϵ. We report the main exactness results that we need in the paper. Furthermore, we perform some second order analysis that plays a key role in the definition of the algorithm. Namely we show that points satisfying the SONC for Problem (1) correspond to points satisfying some kind of second order optimality condition for the unconstrained problem (which is not twice continuously differentiable).
Section 3 is devoted to the computation of the search directions. In particular, we define two directions d_P, which is a positive curvature direction for the augmented Lagrangian L_a, and $\hat{d}_S$, which is a negative curvature direction and is able to enforce convergence toward KKT pairs that satisfy SONC.
Section 4 is dedicated to the definition of an adaptive linesearch technique (ALS) for the minimization of L_a for a fixed value of the penalty parameter ϵ.
In Section 5 we introduce the overall Second Order Lagrangian Algorithm (SOLA) converging to points satisfying the SONC. Updating rules for ϵ that guarantee that it eventually stays fixed and that exactness properties are met, are defined here.
In Section 6 we prove that the convergence rate is superlinear if the sequence produced by SOLA converges towards a KKT pair which satisfies also the SSOSC for Problem (1).

We conclude this section by introducing some notation.
We denote by $\mathcal{F} = \{x \in \mathbb{R}^n : g(x) \leq 0\}$ the feasible set of Problem (1). At a given point $x \in \mathbb{R}^n$, not necessarily feasible, we associate the index sets:

$$A_0(x) = \{j : \ g_j(x) = 0\}, \quad N_0(x) = \{j : \ g_j(x) < 0\}.$$

Given a vector $v \in \mathbb{R}^p$, we indicate by the uppercase V the diagonal matrix $V = \mathrm{diag}_{1 \leq i \leq p}\{v_i\}$.
Let $K \subseteq \{1, \dots, p\}$ be an index subset, we denote by v_K the subvector of v with components v_i such that $i \in K$.
Given two vectors $v, w \in \mathbb{R}^p$, the operation $\max\{v, w\}$ is intended component-wise, namely $\max\{v, w\}$ denotes the vector with components $\max\{v_i, w_i\}$.
Given a symmetric matrix Q, we denote by $\lambda_m(Q)$ and $\lambda_M(Q)$ the smallest and largest eigenvalues of Q, respectively.
We denote by $\|\cdot\|_p$ the ℓ_p norm, and when p is not specified we intend $p = 2$.

2 The augmented Lagrangian function

In the definition of the algorithm we make use of the following augmented Lagrangian function, introduced in [11]:

$$\begin{aligned} L_a(x,\lambda;\epsilon) &= f(x) + \lambda' \max\{g(x), -\epsilon p(x,\lambda)\lambda\} + \frac{\|\max\{g(x), -\epsilon p(x,\lambda)\lambda\}\|^2}{2\epsilon p(x,\lambda)} \\ &\quad + \|\nabla g(x)' \nabla_x L(x,\lambda) + G(x)^2 \lambda\|^2, \end{aligned}$$

where $\epsilon > 0$ is a penalty parameter and

$$p(x,\lambda) = \frac{a(x)}{1+\|\lambda\|^2}\ , \tag{5}$$

with

$$a(x) = \alpha - \|\max\{g(x),0\}\|_s^s\ ,$$

and the scalars α, s are such that $\alpha > 0$ and $s \geq 3$.

The function $L_a(x,\lambda;\epsilon)$ is defined on the set

$$\mathcal{P} = \{x \in \mathbb{R}^n :\ \alpha - \|\max\{g(x),0\}\|_s^s > 0\},$$

which is an open perturbation of the feasible set $\mathcal{F}$, so that $\mathcal{F} \subset \mathcal{P}$.

We refer to [11] and [10] for a detailed discussion of the rationale behind the structure of the augmented Lagrangian $L_a(x,\lambda;\epsilon)$.

We point out that, given any point $x^0 \in \mathbb{R}^n$, it is easy to select values α and s that appear in the definition of $\mathcal{P}$, such that $x^0 \in \mathcal{P}$. Hence, given a point $(x^0,\lambda^0) \in \mathcal{P} \times \mathbb{R}^m$, we can introduce the level set of L_a defined by:

$$\Omega^0(\epsilon) = \{(x,\lambda) \in \mathcal{P} \times \mathbb{R}^m : L_a(x,\lambda;\epsilon) \leq L_a(x^0,\lambda^0;\epsilon)\}.$$

As we said before, our aim is to solve Problem (1) by an unconstrained minimization of L_a on $\mathcal{P} \times \mathbb{R}^m$. Therefore we are interested in the correspondence between stationary points of L_a belonging to $\Omega^0(\epsilon)$ and KKT pairs of Problem (1), as well as in the correspondence between local (global) minimizers of L_a belonging to $\Omega^0(\epsilon)$ and local (global) solutions of Problem (1).

The exactness properties of the function L_a employed in this paper can be stated under the following assumptions, which are discussed in details in [11]:

Assumption 1 *One of the two following conditions is satisfied:*

(a) $x^0 \in \mathcal{F}$ *and* $f(x)$ *is coercive on the closure* $\bar{\mathcal{P}}$ *of* $\mathcal{P}$ *, i.e. for any* $\{x^k\} \subseteq \mathcal{P}$ *with* $\|x^k\| \to \infty$ *we have* $f(x^k) \to \infty$*;*

(b) *the set $\bar{\mathcal{P}}$ is bounded and at every point $x \in \mathcal{P}/\mathcal{F}$ it results:*

$$\sum_{i:g_i(x)>0} c_i(x)\nabla g_i(x) \neq 0,$$

where

$$c_i(x) = \left[1 + \frac{s}{2}\frac{\|\max\{g(x),0\}\|^2 g_i(x)^{(s-2)}}{\alpha - \|\max\{g(x),0\}\|_s^s}\right] g_i(x).$$

Assumption 2 *For every $x \in \mathcal{F}$ the gradients $\nabla g_i(x)$ with $i \in A_0(x)$ are linearly independent.*

Assumptions 1 and 2 are reasonable enough. In fact, Assumption 1(a) is equivalent to the compactness of the level sets of the objective function on the set $\mathcal{P}$ and it is similar to the one usually used in the unconstrained case. Assumption 1(b) is a weakening of the Mangasarian-Fromovitz constraint qualification condition and it ensures the existence of a feasible solution of the constrained problem. Assumptions 2 requires that the LICQ holds at every feasible point and it guarantees the existence and uniqueness of the KKT multipliers.

In the sequel, we assume that Assumptions 1 and 2 hold.

The next proposition ensures that the essentially unconstrained problem

$$\min_{(x,\lambda)\in\mathcal{P}\times\mathbb{R}^m} L_a(x,\lambda;\epsilon)$$

is well defined, in the sense that a solution exists, and establishes the main exactness properties of the augmented Lagrangian function L_a.

Proposition 2.1 (See [11]) *Under Assumptions 1 and 2 it results:*

- *(a) for every $\epsilon > 0$ the level set $\Omega^0(\epsilon)$ is compact;*
- *(b) a value $\bar{\epsilon} > 0$ exists such that for all $\epsilon \in (0, \bar{\epsilon}]$,*
    - (i) *if $(\bar{x}, \bar{\lambda}) \in \Omega^0(\epsilon)$ is a stationary point of $L_a(x,\lambda;\epsilon)$, the pair $(\bar{x}, \bar{\lambda})$ is a KKT pair for Problem (1).*
    - (ii) *if $(\bar{x}, \bar{\lambda}) \in \Omega^0(\epsilon)$ is a global minimum point of $L_a(x,\lambda;\epsilon)$, then $\bar{x}$ is a global minimum point for Problem (1) and $\bar{\lambda}$ is the corresponding multiplier, and conversely.*

From the definition and the differentiability assumptions on f and g, it follows that the function $L_a(x,\lambda;\epsilon)$ is an SC^1 function for all $(x,\lambda) \in \mathcal{P} \times \mathbb{R}^m$, that is a continuously differentiable function with a semismooth gradient (see [28]). The gradient of L_a is given by:

$$\begin{aligned}\nabla_x L_a(x,\lambda;\epsilon) = \nabla_x L(x,\lambda) + \frac{1}{\epsilon p(x,\lambda)}\nabla g(x)\max\{g(x), -\epsilon p(x,\lambda)\lambda\} \\ + \frac{s}{2\epsilon a(x) p(x,\lambda)}\|\max\{g(x), -\epsilon p(x,\lambda)\lambda\}\|^2 \sum_{i=1}^m \nabla g_i(x)\max\{g_i(x),0\}^{s-1}\end{aligned}$$

$$+2\left[\nabla_x^2 L(x,\lambda)\nabla g(x) + \sum_{i=1}^{m} \nabla^2 g_i(x)\nabla_x L(x,\lambda)e_i' + 2\nabla g(x)G(x)\Lambda\right]\left[M(x)\lambda + \nabla g(x)'\nabla f(x)\right],$$

$$\nabla_\lambda L_a(x,\lambda;\epsilon) = \max\{g(x), -\epsilon p(x,\lambda)\lambda\} + \frac{1}{\epsilon a(x)}\,\|\max\{g(x), -\epsilon p(x,\lambda)\lambda\}\|^2\lambda$$
$$+2M(x)\left[M(x)\lambda + \nabla g(x)'\nabla f(x)\right]$$

where $M(x)$ is given by

$$M(x) \quad = \quad \nabla g(x)'\nabla g(x) + G^2(x), \qquad (6)$$

and e_i denotes the ith column of the $m \times m$ identity matrix.

Since L_a is an SC^1 function in $\mathcal{P} \times \mathbb{R}^m$, its generalized Hessian $\partial^2 L_a(x,\lambda;\epsilon)$, in Clarke's sense, can be defined [3]. For SC^1 functions a second-order Taylor-like expansion is possible, as stated in the following proposition.

Proposition 2.2 (See [21]) *Let $h : \mathbb{R}^n \to \mathbb{R}$ be an SC^1 function on the open set $\mathcal{O}$ and let x and y be two points in $\mathcal{O}$ such that $[x,y]$ is contained in $\mathcal{O}$. Then*

$$h(y) = h(x) + \nabla h(x)'(y-x) + \frac{1}{2}(y-x)'\Phi(y-x),$$

for some $\Phi \in \partial^2 h(z)$ and for some $z \in (x,y)$, where $\partial^2 h$ denotes the generalized Hessian of h.

By exploiting the piecewise smooth structure of the gradient of L_a, it is possible to describe the structure of the generalized Hessian $\partial^2 L_a$ in a neighborhood of a KKT pair of Problem (1). To this aim we consider a partition of the index set $\{1,\dots,m\}$ into the subsets $A \subseteq \{1,\dots,m\}$, $N = \{1,\dots,m\} \setminus A$, and we partition the vectors g and λ according to these index sets: $g = (g_A' \ g_N')'$ and $\lambda = (\lambda_A' \ \lambda_N')'$. Then we introduce the $(n+m) \times (n+m)$ symmetric matrix $H(x,\lambda;\epsilon,A)$ given block-wise by:

$$H_{xx}(x,\lambda;\epsilon,A) = \nabla_x^2 L(x,\lambda) + \frac{1}{\epsilon p(x,\lambda)}\nabla g_A(x)\nabla g_A(x)' + 2\nabla_x^2 L(x,\lambda)\nabla g(x)\nabla g(x)'\nabla_x^2 L(x,\lambda),$$

$$\begin{aligned} H_{x\lambda}(x,\lambda;\epsilon,A) &= \begin{bmatrix}\nabla g_A(x) & 0\end{bmatrix} + 2\nabla_x^2 L(x,\lambda)\nabla g(x) M_N(x), \\ H_{\lambda\lambda}(x,\lambda;\epsilon,A) &= -\epsilon p(x,\lambda)\begin{bmatrix}0 & 0\\ 0 & I_N\end{bmatrix} + 2M_N(x)M_N(x), \end{aligned} \qquad (7)$$

where

$$M_N(x) = \nabla g(x)'\nabla g(x) + \begin{pmatrix}0 & 0\\ 0 & G_N(x)^2\end{pmatrix}, \qquad (8)$$

I_N is the identity matrix of dimension $|N|$ and 0 is a zero matrix of proper dimensions. In correspondence of a KKT pair $(\bar{x},\bar{\lambda})$ we define the index set of the strictly active constraints, namely $A_+(\bar{x},\bar{\lambda}) = \{j \in A_0(\bar{x}) : \ \bar{\lambda}_j > 0\}$.

Proposition 2.3 (See [11], Proposition 6.1) *For every KKT pair* $(\bar{x},\bar{\lambda})$ *of Problem (1) and every given* ϵ*, a neighborhood* $\mathcal{B}$ *of* $(\bar{x},\bar{\lambda})$ *exists such that, for all* (x,λ) *in* $\mathcal{B}$*, we have* $\partial^2 L_a(x,\lambda;\epsilon) = \mathrm{co}\{\partial_B^2 L_a(x,\lambda;\epsilon)\}$*, where:*

$$\partial_B^2 L_a(x,\lambda;\epsilon) = \{H(x,\lambda;\epsilon,A) + K(x,\lambda;\epsilon,A) : \; A \in \mathcal{A}\},$$

with

$$\mathcal{A} = \{A : \; A_+(\bar{x},\bar{\lambda}) \subseteq A \subseteq A_0(\bar{x})\},$$

$H(x,\lambda;\epsilon,A)$ *is given by (7) and* $K(x,\lambda;\epsilon,A)$ *is a matrix such that* $\|K(x,\lambda;\epsilon,A)\| \leq \zeta(x,\lambda)$*, with* $\zeta(x,\lambda)$ *a nonnegative continuous function such that* $\zeta(\bar{x},\bar{\lambda}) = 0$*.*

We note that at a KKT pair where the strict complementarity holds, $A_+(\bar{x},\bar{\lambda}) = A_0(\bar{x})$, and $\partial^2 L_a(\bar{x},\bar{\lambda};\epsilon)$ reduces to a singleton; therefore in this case the generalized Hessian can be further characterized in a neighborhood of the KKT pair.

Proposition 2.4 (See [11], Proposition 6.2) *For every KKT pair* $(\bar{x},\bar{\lambda})$ *of Problem (1) where the strict complementarity holds, and for every given* ϵ*, a neighborhood* $\mathcal{B}$ *of* $(\bar{x},\bar{\lambda})$ *exists such that, for all* (x,λ) *in* $\mathcal{B}$*,* L_a *is twice continuously differentiable, with Hessian matrix given by:*

$$\nabla^2 L_a(x,\lambda;\epsilon) = H(x,\lambda;\epsilon,A_0(\bar{x})) + K(x,\lambda;\epsilon,A_0(\bar{x})),$$

where H *and* K *are matrices like in Proposition 2.3.*

The next proposition establishes that, for sufficiently small values of ϵ, KKT pairs of Problem (1) satisfying the SSOSC are strict local minimizers of L_a which satisfy also the second order sufficient optimality condition for SC^1 functions.

Proposition 2.5 (See [11], Theorem 6.3) *Let* $(\bar{x},\bar{\lambda})$ *be a KKT pair of Problem (1) which satisfies the SSOSC and LICQ. Then there exists an* $\bar{\epsilon}$ *such that, for all* $\epsilon \in (0,\bar{\epsilon}]$*,* $(\bar{x},\bar{\lambda})$ *is an isolated local minimum point for* $L_a(x,\lambda;\epsilon)$*, and all matrices in* $\partial^2 L_a(\bar{x},\bar{\lambda};\epsilon)$ *are positive definite.*

The next proposition provides a basis for the construction of an algorithm converging to second order stationary points of Problem (1).

Proposition 2.6 (See [12], Proposition 2.7) *Let* $(\bar{x},\bar{\lambda})$ *be a KKT pair of Problem (1) and let* $\epsilon > 0$ *be given. If a positive semidefinite matrix* $W \in \partial_B^2 L_a(\bar{x},\bar{\lambda};\epsilon)$ *exists, then the pair* $(\bar{x},\bar{\lambda})$ *satisfies the second order necessary conditions (3).*

The exactness properties of the augmented Lagrangian function L_a, described in the previous section, give us the possibility of defining constrained minimization algorithms by drawing inspiration from the approaches developed in the field of unconstrained optimization. In particular, the merit function L_a allows us to extend the approach of defining curvilinear search algorithms in order to get global convergence towards second order stationary points.

However, as already pointed out in [17], this extension is not trivial since we must cope with the following difficulties:

- the correspondence between the stationary points of the augmented Lagrangian function L_a and the KKT pairs of the original constrained problem holds only for values of the penalty parameter ϵ smaller than the unknown threshold value $\bar{\epsilon}$; therefore any constrained optimization method, based on the minimization of the Lagrangian function L_a, must be able itself to locate suitable values of the penalty parameter;
- the augmented Lagrangian function L_a is not twice continuously differentiable everywhere and the evaluation of the second order derivatives, where they exist, requires the use of the third order derivatives of the functions f and g.

In this paper we describe a new primal-dual algorithm model by extending the approach proposed in [17], without requiring the third order derivatives of f and g. The algorithm is defined in the extended space of the primal-dual variables (x, λ). We denote the vector of the variable in the $n+m$ space as $z = (x', \ \lambda')'$ and all vectors are assumed partitioned accordingly.

3 Computation of the search direction

We begin this section by introducing the matrix Q^k defined as in [12], that is

$$Q^k = H(x^k, \lambda^k; \epsilon, A_{\oplus}^k), \tag{9}$$

where $H(x^k, \lambda^k; \epsilon, A_{\oplus}^k)$ is given by (7) and the set of indices $A_{\oplus}^k$ is such that a neighborhood $\mathcal{B}(\bar{x}, \bar{\lambda})$ of a KKT pair $(\bar{x}, \bar{\lambda})$ exists such that, for all $(x^k, \lambda^k) \in \mathcal{B}(\bar{x}, \bar{\lambda})$, $A_{\oplus}^k = A_0(\bar{x})$. We refer the reader to papers [15] and [16] for examples of sets $A_{\oplus}^k$ that satisfy the preceding requirement. We recall that, in the following, $d = (d_x', \ d_\lambda')'$.
The following proposition establishes the asymptotic connections between the matrix Q^k and the generalized Hessian at a first order stationary point of L_a.

Proposition 3.1 *Let $\{x^k, \lambda^k\}$ be a sequence converging to a first order stationary point $(\bar{x}, \bar{\lambda})$ of the function L_a and $\{Q^k\}$ a sequence of matrices defined by (9). Then,*

$$\lim_{k\to\infty} dist(Q^k | \partial_B^2 L_a(\bar{x}, \bar{\lambda}; \epsilon)) = 0.$$

Furthermore, for every sequence of matrices $\{W^k\}$, with $W^k \in \partial^2 L_a(x^k, \lambda^k; \epsilon)$, and every sequence of directions $\{d^k\}$, we have

$$(d^k)'(W^k - Q^k)d^k \leq \delta^k,$$

where $\{\delta^k\}$ is a sequence of numbers converging to 0.

Proof. The proof easily follows from Proposition 2.3 and Proposition 5.1 in [12]. $\triangleleft$

The main idea for computing the search direction for minimizing L_a consists in using an iterative scheme to solve the system

$$Q^k d = -\nabla L_a^k.$$

This can be done by means of a conjugate gradient method within a truncated scheme thus allowing the solution of large-scale problems. We consider the minimization of the quadratic form

$$q(d) = \frac{1}{2} d' Q^k d + (\nabla L_a^k)' d,$$

as described in references [18, 20]. To simplify notation, in the remainder of this section we omit the iteration index k. Then, we recall that the conjugate gradient method either stops at the first iteration, thus producing $p^0 = -\nabla L_a$, or it performs m iterations, with $1 \le m \le n$, returning $m+1$ conjugate vectors with respect to matrix Q

$$p^0, p^1, \dots, p^m.$$

Following the idea proposed in [18] for unconstrained optimization, given a number $\rho > 0$, we can define the following disjoint sets of indices

$$\begin{aligned} I_P &= \{ i \le m : (p^i)' Q p^i \ge \rho \|p^i\|^2 \}, \\ I_S &= \{ i \le m : (p^i)' Q p^i \le -\rho \|p^i\|^2 \} \end{aligned}$$

which, in case $m \ge 1$, cannot be both empty. Assuming $m \ge 1$, let us define the following directions

$$d_P = \begin{cases} -\displaystyle\sum_{i \in I_P} \frac{\nabla {L_a}' p^i}{{p^i}' Q p^i} p^i, & I_P \ne \emptyset \\ \quad 0 & I_P = \emptyset \end{cases} \tag{10}$$

$$d_S = \begin{cases} -\displaystyle\sum_{i \in I_S} \frac{\nabla {L_a}' p^i}{|{p^i}' Q p^i|} p^i, & I_P \ne \emptyset \\ \quad 0 & I_S = \emptyset \end{cases} \tag{11}$$

which, by definition, are such that:

$$\begin{aligned} {d_P}' Q d_P &\ge 0, & (12) \\ {d_S}' Q d_S &\le 0. & (13) \end{aligned}$$

Therefore, d_P is a positive curvature direction whereas d_S is a negative curvature direction. The following proposition holds.

Proposition 3.2 *A constant $c > 0$ exists, such that*

$$\max\{\|d_P\|, \|d_S\|\} \le c \|\nabla L_a\|. \tag{14}$$

Proof. The proof follows by considering point (b) of Theorem 2.2 in [20]. $\triangleleft$

In order to guarantee convergence to second order stationary points, a modification of direction d_S is necessary. In particular, we consider a direction $\hat{d}_S$ obtained by adding to d_S another direction d_N, which is able to guarantee desired second order properties. Hence, let

$$\hat{d}_S = d_S + d_N$$

where d_N is such that the following Condition holds

Condition 1 *Direction d_N satisfies*

(a) $d_N = 0$ *if* Q *is positive semidefinite;*

(b) $\nabla {L_a}' d_N \le 0$ *and* $d'_N Q d_N < 0$*;*

(c) $(d_S + d_N)' Q (d_S + d_N) < 0$*;*

(d) *let* $\{d_S^k\}$, $\{d_N^k\}$ *and* $\{Q^k\}$ *be sequences of directions and matrices. Then,* $\{d_N^k\}$ *is bounded and, if* $\lim\limits_{k\to\infty} (d_S^k + d_N^k)' Q^k (d_S^k + d_N^k) = 0$ *then* $\lim\limits_{k\to\infty} \lambda_m(Q^k) = 0$.

We point out that a direction d_N satisfying Condition 1 can be computed in different ways, see, for example, [22, 25].

Let us now define

$$q(w) = \nabla {L_a}' w + 1/2 w' Q w.$$

Then, we show that directions d_P and $\hat{d}_S$ have some interesting properties that will be exploited for algorithmic purposes.

Proposition 3.3 *Directions d_P and $\hat{d}_S$ are such that:*

- *if* $q(d_P) \le q(\hat{d}_S)$, d_P *satisfies* $\nabla {L_a}' d_P \le -\frac{1}{2} \frac{\|\nabla L_a\|^2}{\|Q\|}$*;*
- *if* $q(d_P) > q(\hat{d}_S)$, $\hat{d}_S$ *satisfies* $q(\hat{d}_S) \le -\frac{2}{3} \frac{\|\nabla L_a\|^2}{\|Q\|}$.

Proof. Suppose first that $q(d_P) \le q(\hat{d}_S)$. In this case, considering that ${d_P}' Q d_P \ge 0$ and ${\hat{d}_S}{}' Q \hat{d}_S < 0$, we have that

$$\nabla {L_a}' \hat{d}_S > q(\hat{d}_S) \ge q(d_P) \ge \nabla {L_a}' d_P. \tag{15}$$

Let us consider first the case when $I_S = \emptyset$. Then, considering that $p^0 = -\nabla L_a$,

$$\nabla {L_a}' d_P = -\sum_{i\in I_P} \frac{(\nabla {L_a}' p^i)^2}{{p^i}' Q p^i} \le -\frac{(\nabla {L_a}' p^0)^2}{{p^0}' Q p^0} \le -\frac{\|\nabla L_a\|^2}{\|Q\|}.$$

Otherwise, if $I_S \ne \emptyset$, from (10), (15) and point (b) of Theorem 2.2 of [20], it results

$$-\frac{\|\nabla L_a\|^2}{\|Q\|} \ge -\nabla {L_a}' \left(\sum_{i\in I_S} \frac{\nabla {L_a}' p^i}{|{p^i}' Q p^i|} p^i + \sum_{i\in I_P} \frac{\nabla {L_a}' p^i}{{p^i}' Q p^i} p^i \right) \ge 2 \nabla {L_a}' d_P.$$

Consider now the case when $q(d_P) > q(\hat{d}_S)$. By (10) we have that

$$\begin{aligned} q(\hat{d}_S) &< \nabla {L_a}' d_P + \frac{1}{2} {d_P}' Q d_P = \nabla {L_a}' d_P + \frac{1}{2} \sum_{i\in I_P} \frac{(\nabla {L_a}' p^i)^2}{{p^i}' Q p^i} \\ &= \nabla {L_a}' d_P - \frac{1}{2} \nabla {L_a}' \left(-\sum_{i\in I_P} \frac{(\nabla {L_a}' p^i)}{{p^i}' Q p^i} p^i \right) = \frac{1}{2} \nabla {L_a}' d_P. \end{aligned} \tag{16}$$

On the other hand, by points (b) and (c) of Condition 1, we know that

$$q(\hat{d}_S) \leq \nabla {L_a}' \hat{d}_S. \tag{17}$$

From relations (16) and (17) we get that

$$\frac{3}{2} q(\hat{d}_S) \leq \frac{1}{2} \nabla {L_a}' (d_P + \hat{d}_S)$$

which , by (10), (11) and point (b) of Theorem 2.2 in [20], implies that

$$\frac{3}{2} q(\hat{d}_S) \leq - \frac{\|\nabla L_a\|^2}{\|Q\|}.$$

which completes the proof. $\triangleleft$

4 The adaptive linesearch ALS

In this section we focus our attention on the definition of a suitable linesearch scheme. The procedure, given the two directions d_P and $\hat{d}_S$, selects one of them and performs a suitable linesearch along it. For this reason we call the procedure adaptive linesearch (ALS). We show that ALS is able to enforce convergence towards stationary points of the augmented Lagrangian function L_a assuming that ϵ stays fixed. The problem of adjusting the value of the penalty parameter and the convergence towards second order stationary points of the original constrained Problem (1) will be addressed in the next section.

To simplify notation, we omit the iteration index k whenever it is unnecessary.

For a fixed value of the penalty parameter ϵ, the adaptive linesearch ALS takes as input $(x, \lambda) = z \in \Omega^0(\epsilon)$, the values $L_a = L_a(x, \lambda; \epsilon)$, $\nabla L_a = \nabla L_a(x, \lambda; \epsilon)$, the matrix $Q \in \mathbb{R}^{(n+m)\times(n+m)}$, the directions $d_P, \hat{d}_S$ and it gives in output the point $(\hat{x}, \hat{\lambda}) = \hat{z} \in \Omega^0(\epsilon)$.

Adaptive linesearch: ALS$(x, \lambda, d_P, \hat{d}_S; \epsilon)$

Step 0. **Data**. constants $\beta \in (0,1)$, $\sigma > 0$ and $\mu \in (0, \frac{1}{2})$ are given.

Step 1. **Choice of the search direction**.
If $q(d_P) \leq q(\hat{d}_S)$ **Then** *go to Step 2.* **Else** *go to Step 3.*

Step 2. **Linesearch along a gradient-related direction**.
Set $d = d_P$ *and compute* $\eta = \beta^\ell$ *where* ℓ *is the smallest nonnegative integer such that*

$$L_a(z + \eta d; \epsilon) \leq L_a(z; \epsilon) + \mu\eta \nabla {L_a}' d \quad \text{and} \quad x + \eta d_x \in \mathcal{P}, \tag{18}$$

and go to Step 4.

Step 3. **Linesearch along a negative curvature direction**.
Set $d = \hat{d}_S$. *If*

$$L_a(z + \sigma d; \epsilon) \leq L_a(z; \epsilon) + \mu \left(\sigma \nabla {L_a}' d + \frac{1}{2}(\sigma)^2 d' Q d \right) \quad \textit{and} \quad x + \sigma d_x \in \mathcal{P}, \tag{19}$$

compute $\eta = \beta^\ell \sigma$, *where* ℓ *is the largest non-positive integer such that*

$$\left\{ \begin{array}{l} L_a(z + \eta d; \epsilon) \leq L_a(z; \epsilon) + \mu \left(\eta \nabla {L_a}' d + \frac{1}{2}(\eta)^2 d' Q d \right) \\ \text{and} \quad x + \eta d_x \in \mathcal{P}, \end{array} \right. \tag{20}$$

and

$$\left\langle \begin{array}{l} L_a(z + \frac{\eta}{\beta} d; \epsilon) > L_a(z; \epsilon) + \mu \left(\frac{\eta}{\beta} \nabla {L_a}' d + \frac{1}{2} \left(\frac{\eta}{\beta} \right)^2 d' Q d \right) \\ \text{or} \quad x + \frac{\eta}{\beta} d_x \notin \mathcal{P}. \end{array} \right. \tag{21}$$

Otherwise compute $\eta = \beta^\ell \sigma$, *where* ℓ *is the smallest positive integer such that (20) holds.*

Step 4. **Update**. *Set* $\hat{z} = z + \eta d$ *and return* $\hat{z}$.

We note that the adaptive linesearch ALS differs from unconstrained curvilinear searches for the fact that the trial points are accepted only if the x-component belongs to $\mathcal{P}$. The matrix Q plays a role similar to the Hessian matrix in unconstrained algorithms; it is given by (9), as it has been shown in Section 3, so as to provide some kind of second order information on the original constrained problem to the algorithm.

First we prove that ALS is well-defined.

Proposition 4.1 *For any fixed value* ϵ, *let the matrix* Q *and the directions* d_P, $\hat{d}_S$ *be defined as in Section 3. Then procedure ALS computes a value* $\eta > 0$.

Proof. We proceed by considering the two cases (i) $\nabla L_a(z; \epsilon) \neq 0$ and (ii) $\nabla L_a(z; \epsilon) = 0$.

(i) In case procedure ALS selects direction d_P, the result follows from standard arguments of Armijo-type linesearch along a descent direction.

Let us now assume that procedure ALS selects direction $\hat{d}_S$. If (19) is satisfied, the existence of a finite ℓ is implied by (20) and the compactness of the level set $\Omega_0(\epsilon)$. Assume now that (19) fails. We note that $\mathcal{P}$ is an open set, therefore the test $x + \eta d_x \in \mathcal{P}$ is satisfied for sufficiently small values of η. Moreover, by Proposition 2.2, we can write

$$L_a\left(z + \eta d; \epsilon\right) = L_a(z; \epsilon) + \eta \nabla L_a(z; \epsilon)' \hat{d}_S + \frac{1}{2}\eta^2 (\hat{d}_S)' W \hat{d}_S \tag{22}$$

for some symmetric matrix W belonging to $\partial^2 L_a(u; \epsilon)$ where $u = z + \omega\eta\hat{d}_S$ for some $\omega \in (0, 1)$.

Now assume, by contradiction, that a sequence $\{\eta^j\}$ exists such that $\eta^j \to 0$ for $j \to \infty$ and

$$L_a\left(z + \eta^j d; \epsilon\right) > L_a(z; \epsilon) + \mu\eta^j \nabla L_a(z; \epsilon)' \hat{d}_S + \mu\frac{1}{2}(\eta^j)^2 (\hat{d}_S)' Q \hat{d}_S. \tag{23}$$

By considering (22), we get

$$0 > (\mu - 1)\eta^j \nabla L_a(z; \epsilon)' \hat{d}_S - \frac{1}{2}(\eta^j)^2 (\hat{d}_S)' W \hat{d}_S + \mu\frac{1}{2}(\eta^j)^2 (\hat{d}_S)' Q \hat{d}_S. \tag{24}$$

Dividing (24) by η^j and taking the limit for $j \to \infty$, and considering that $\mu < 1$, we get

$$\nabla L_a(z; \epsilon)' \hat{d}_S \geq 0.$$

On the other hand, by (11) and point (b) of Condition 1, we know that

$$\nabla L_a(z; \epsilon)' \hat{d}_S < 0,$$

thus completing the proof.

(ii) Let us now suppose that $\nabla L_a(z; \epsilon) = 0$. In this case, by definition, we have that $d_P = 0$ and $\hat{d}_S = d_N$, so that ALS selects direction $\hat{d}_S$. Furthermore, by Condition 1 it results $(\hat{d}_S)' Q \hat{d}_S < 0$. Again, if relation (19) is satisfied, the existence of a finite ℓ is guaranteed by (20) and the compactness of the level set $\Omega_0(\epsilon)$.

Otherwise, assume that (19) is not satisfied. By the fact that $\mathcal{P}$ is an open set, it follows that $x + \eta d \in \mathcal{P}$ for η sufficiently small. Hence, by Proposition 2.2, we can write

$$L_a\left(z + \eta d; \epsilon\right) = L_a(z; \epsilon) + \frac{1}{2}\eta^2 (\hat{d}_S)' W^j \hat{d}_S \tag{25}$$

for some symmetric matrix W^j belonging to $\partial^2 L_a(u; \epsilon)$ where $u = z + \omega\eta\hat{d}_S$ for some $\omega \in (0, 1)$.

Assume, by contradiction, that a sequence $\{\eta^j\}$ exists such that $\eta^j \to 0$ for $j \to \infty$ and

$$L_a\left(z + \eta^j d; \epsilon\right) > L_a(z; \epsilon) + \mu\frac{1}{2}(\eta^j)^2 (\hat{d}_S)' Q \hat{d}_S. \tag{26}$$

By considering (25) and (26) we get

$$0 > -\frac{1}{2}(\eta^j)^2(\hat{d}_S)'W^j\hat{d}_S + \mu\frac{1}{2}(\eta^j)^2(\hat{d}_S)'Q\hat{d}_S,$$

which, dividing both sides by $(\eta^j)^2/2$ and by adding and subtracting $(\hat{d}_S)'Q\hat{d}_S$, gives

$$0 < (\hat{d}_S)'(W^j - Q)\hat{d}_S + (1-\mu)(\hat{d}_S)'Q\hat{d}_S.$$

Now, considering Proposition 3.1 we get

$$0 < (1-\mu)(\hat{d}_S)'Q\hat{d}_S + \delta^j \tag{27}$$

where $\delta^j \to 0$ since $z + \eta^j d \to z$ where $\nabla L_a(z;\epsilon) = 0$. Taking the limit for $j \to \infty$ in (27) we obtain

$$0 < (1-\mu)(\hat{d}_S)'Q\hat{d}_S$$

which contradicts the assumption $\nabla L_a(z;\epsilon) = 0$ then $(\hat{d}_S)'Q\hat{d}_S < 0$. ◁

As already pointed out in Section 2, we recall that the open perturbation $\mathcal{P}$ of the feasible set is chosen in such a way that the initial point $x^0 \in \mathcal{P}$. Hence, the linesearch procedure ALS, as stated in Proposition 4.1, is well defined, in the sense that a value $\eta > 0$ is always computed such that, in particular, $x + \eta d_x \in \mathcal{P}$.

In the following we denote by $(x^{k+1}, \lambda^{k+1}) = ALS(x^k, \lambda^k, d_P^k, \hat{d}_S^k; \epsilon)$ the new point produced by ALS for a given value of ϵ and we show that ALS is able to produce a sequence of points $\{(x^k, \lambda^k)\}$ globally convergent towards stationary points of the augmented Lagrangian function L_a. The overall algorithm SOLA converging to second order stationary points and including the adjustment rule for ϵ is presented in Section 5.

Now, we can prove the following result.

Proposition 4.2 *For any fixed value ϵ, let the matrix Q^k and the directions d_P^k, $\hat{d}_S^k$ be defined as in Section 3. Let $\{x^k, \lambda^k\}$ be an infinite sequence produced by procedure ALS, then*

$$\lim_{k\to\infty} \nabla L_a(x^k, \lambda^k; \epsilon) = 0. \tag{28}$$

Moreover for every infinite index set $\bar{K}$ such that $d^k = \hat{d}_S^k$ for all $k \in \bar{K}$

$$\lim_{k\to\infty, k\in\bar{K}} (\hat{d}_S^k)'Q^k\hat{d}_S^k = 0. \tag{29}$$

Proof. By the compactness of the level set $\Omega_0(\epsilon)$, we have that the sequence $\{x^k, \lambda^k\}$ admits a limit point $(\tilde{x}, \tilde{\lambda})$. Let K_P and K_S be index sets of two subsequences of iterates converging to $(\tilde{x}, \tilde{\lambda})$ such that

(i) $d^k = d_P^k$ for all $k \in K_P$ and (18) holds;

(ii) $d^k = \hat{d}_S^k$ for all $k \in K_S$ and (20) holds.

By (18) and (20) we know that the sequence $\{L_a(x^k,\lambda^k;\epsilon)\}$ is non-increasing which considering the compactness of $\Omega_0(\epsilon)$ yields that $\{L_a(x^k,\lambda^k;\epsilon)\}$ admits a limit point.
Suppose first that K_P is infinite. In order to prove that $\nabla L_a(\tilde{x},\tilde{\lambda};\epsilon)=0$ we proceed by contradiction. Suppose that $\|\nabla L_a(x^k,\lambda^k;\epsilon)\|>\tau>0$ for all $k\in K_P$.
Then we have

$$|L_a(x^{k+1},\lambda^{k+1};\epsilon)-L_a(x^k,\lambda^k;\epsilon)|\geq\mu\eta^k|\nabla L_a(x^k,\lambda^k;\epsilon)'d_P^k|.$$

It follows that $\eta^k|\nabla L_a(x^k,\lambda^k;\epsilon)'d_P^k|\to 0$ as $k\to\infty$, $k\in K_P$. Therefore, either $\eta^k\to 0$ or $|\nabla L_a(x^k,\lambda^k;\epsilon)'d_P^k|\to 0$, as $k\to\infty$, $k\in K_P$.
Suppose first that $\eta^k\to 0$, as $k\to\infty$, $k\in K_P$. Since

$$L_a\left(z^k+\frac{\eta^k}{\beta}d_P^k;\epsilon\right)-L_a(z_k;\epsilon)>\mu\frac{\eta^k}{\beta}(\nabla L_a^k)'d_P^k,$$

then, by the mean value theorem we have, for $k\in K_P$,

$$\frac{\eta^k}{\beta}\nabla L_a(z^k+\delta\frac{\eta^k}{\beta}d_P^k;\epsilon)'d^k>\mu\frac{\eta^k}{\beta}(\nabla L_a^k)'d_P^k,$$

for some $\delta\in(0,1)$. Dividing by η^k/β and by $\|d_P^k\|$, we obtain

$$\frac{\nabla L_a(z^k+\delta\frac{\eta^k}{\beta}d_P^k;\epsilon)'d_P^k}{\|d_P^k\|}>\mu\frac{(\nabla L_a^k)'d_P^k}{\|d_P^k\|}\tag{30}$$

for $k\in K_P$. Now, we can extract a subsequence whose indices lie in the set $K_P'\subseteq K_P$ such that

$$z^k\to\tilde{z}\quad\text{and}\quad\frac{d_P^k}{\|d_P^k\|}\to\tilde{d}$$

for $k\in K_P'$. From (30), taking the limit as $k\to\infty, k\in K_P'$ we obtain that

$$(1-\mu)\nabla L_a(\tilde{z};\epsilon)'\tilde{d}\geq 0.$$

Since $1-\mu>0$ and $(\nabla L_a^k)'d_P^k<0$ for all $k\in K_P'$ we have that $\nabla L_a(\tilde{z};\epsilon)'\tilde{d}=0$ which implies, by using Proposition 3.3, $\nabla L_a(\tilde{z};\epsilon)=0$ and this contradicts the fact that $\|\nabla L_a^k\|>\tau>0$. Hence η^k cannot tend to zero for $k\in K_P$. This implies that there exists a subsequence $K_P''\subseteq K_P$ such that $\left|(\nabla L_a^k)'d_P^k\right|\to 0$ as $k\to\infty, k\in K_P''$. Proposition 3.3 and the continuity of the gradient imply that $\nabla L_a(\tilde{z};\epsilon)=0$, which again contradicts the assumption that $\|\nabla L_a^k\|>\tau>0$. Hence this latter assumption is itself impossible and we conclude that $\nabla L_a(\tilde{z};\epsilon)=0$ whenever K_P is infinite.
Now, suppose that K_S is infinite. In this case, it follows from (20) that

$$\left|L_a(z^{k+1};\epsilon)-L_a(z^k;\epsilon)\right|\geq\mu\left|\eta^k(\nabla L_a^k)'\hat{d}_S^k+\frac{1}{2}(\eta^k)^2(\hat{d}_S^k)'Q^k\hat{d}_S^k\right|.$$

for $k\in K_S$, and hence that

$$\left|\eta^k(\nabla L_a^k)'\hat{d}_S^k+\frac{1}{2}(\eta^k)^2(\hat{d}_S^k)'Q^k\hat{d}_S^k\right|\to 0\quad\text{as}\quad k\to\infty.$$

Therefore, either

$$(\nabla L_a^k)'\hat{d}_S^k \to 0 \quad \text{and} \quad (\hat{d}_S^k)'Q^k\hat{d}_S^k \to 0 \quad (k \to \infty, k \in K_S) \tag{31}$$

or $\eta^k \to 0$ when $k \to \infty, k \in K_S$. Let us first suppose that (31) holds and recall that, by definition of K_S, for every $k \in K_S$ we have $q(d_p^k) > q(\hat{d}_s^k)$. Hence, by Proposition 3.3,

$$q(\hat{d}_s^k) \leq -\frac{2}{3\lambda_M(Q^k)}\|\nabla L_a^k\|^2. \tag{32}$$

Hence, (28) follows from (31) and (32), whereas (29) follows from the second part of (31). Let us now assume, on the other hand, that $\eta^k \to 0$, $k \to \infty, k \in K_S$, we have

$$L_a\left(z^k + \frac{\eta^k}{\beta}\hat{d}_S^k;\epsilon\right) - L_a(z^k;\epsilon) > \mu\left(\frac{\eta^k}{\beta}(\nabla L_a^k)'\hat{d}_S^k + \frac{1}{2}\left(\frac{\eta^k}{\beta}\right)^2(\hat{d}_S^k)^TQ^k\hat{d}_S^k\right),$$

which, by Proposition 2.2, can be rewritten as

$$\frac{\eta^k}{\beta}(\nabla L_a^k)'\hat{d}_S^k + \frac{1}{2}\frac{(\eta^k)^2}{\beta^2}(\hat{d}_S^k)'W^k\hat{d}_S^k > \mu\left(\frac{\eta^k}{\beta}(\nabla L_a^k)'\hat{d}_S^k + \frac{1}{2}\left(\frac{\eta^k}{\beta}\right)^2(\hat{d}_S^k)'Q^k\hat{d}_S^k\right) \tag{33}$$

for some $\delta \in (0,1)$, $k \in K_S$, and with W^k belonging to $\partial^2 L_a(u;\epsilon)$ where $u = (z^k + \delta\frac{\eta^k}{\beta}\hat{d}_S^k)$. Dividing (33) by $\|\hat{d}_S^k\|$ and by Condition 1, we obtain

$$0 \leq (\mu - 1)\left[(\nabla L_a^k)'\hat{d}_S^k + \frac{1}{2}\frac{\eta^k}{\beta}(\hat{d}_S^k)'Q^k\hat{d}_S^k\right] < \frac{1}{2}\frac{\eta^k}{\beta}(\hat{d}_S^k)'\left[W^k - Q^k\right]\hat{d}_S^k. \tag{34}$$

Taking the limit for $k \to \infty$, $k \in K_S$, in (34), recalling that $\eta^k \to 0$, we obtain

$$(\nabla L_a^k)'\hat{d}_S^k \to 0. \tag{35}$$

Furthermore, by considering that by construction of $\hat{d}_S^k$ it results $(\nabla L_a^k)'\hat{d}_S^k \leq 0$ and recalling (34), we have

$$0 \leq (\mu - 1)\frac{1}{2}\frac{\eta^k}{\beta}(\hat{d}_S^k)'Q^k\hat{d}_S^k < \frac{1}{2}\frac{\eta^k}{\beta}(\hat{d}_S^k)'\left[W^k - Q^k\right]\hat{d}_S^k.$$

Dividing the above relation by η^k/β and taking the limit for $k \to \infty, k \in K_S$, we obtain

$$(\hat{d}_S^k)'Q^k\hat{d}_S^k \to 0. \tag{36}$$

Again by Proposition 3.3, we have that (32) holds so that, (28) follows from (35) and (32), whereas (29) follows from (36). ◁

Now we report a technical result which will be used to prove convergence to second order stationary points of the overall algorithm. It basically states that in a neighborhood of a KKT pair that does not satisfy the second order necessary optimality condition, Procedure ALS selects $\hat{d}_S^k$ as search direction.

Proposition 4.3 *Let* $\{(x^k,\lambda^k)\}_{\mathcal{K}}$ *be a subsequence converging to* $(\hat{x},\hat{\lambda})$ *which is a KKT pair of Problem (1) that does not satisfy the second order necessary optimality condition. Let* $\{Q^k\}$, $\{d_P^k\}$ *and* $\{\hat{d}_S^k\}$ *be sequences of matrices and directions defined as in Section 3. Then, for* $k \in \mathcal{K}$ *and sufficiently large, Step 1 of Procedure ALS selects*

$$d^k = \hat{d}_S^k.$$

Proof. Let us assume that the sequence $\{(x^k,\lambda^k)\}_{\mathcal{K}}$ converges to $(\hat{x},\hat{\lambda})$ which is a KKT pair of Problem (1) that does not satisfy the second order necessary optimality condition. Then it is our aim to prove that, for $k \in \mathcal{K}$ and k sufficiently large,

$$q(\hat{d}_S^k) < q(d_P^k),$$

so that $d^k = \hat{d}_S^k$.
First of all, we note that, by construction of d_P^k,

$$q(d_P^k) = (\nabla L_a^k)' d_P^k + \frac{1}{2}(d_P^k)' Q^k d_P^k \geq (\nabla L_a^k)' d_P^k.$$

Therefore, taking the limit,

$$\lim_{k\to\infty, k\in K} q(d_P^k) \geq 0. \tag{37}$$

Analogously, we have

$$q(\hat{d}_S^k) = (\nabla L_a^k)' \hat{d}_S^k + \frac{1}{2}(\hat{d}_S^k)' Q^k \hat{d}_S^k.$$

Since $\{(x^k,\lambda^k)\}_{\mathcal{K}}$ is converging to a KKT pair $(\hat{x},\hat{\lambda})$ which does not satisfy the second order necessary optimality conditions, by Proposition 2.6, we have that every matrix $W \in \partial_B^2 L_a(\bar{x},\bar{\lambda};\epsilon)$ is such that $\lambda_m(W) < 0$. Now, by Proposition 3.1 and for $k \in K$ sufficiently large, $\lambda_m(Q^k) < 0$ as well.
By Proposition 3.2 and Condition 1, it follows that $\hat{d}_S^k$ is bounded and $(\nabla L_a^k)'\hat{d}_S^k \to 0$ as $k \in K$ tends to ∞. Hence, by point (d) of Condition 1,

$$\lim_{k\in K, k\to\infty} q(\hat{d}_S^k) = \lim_{k\in K, k\to\infty} \frac{1}{2}(\hat{d}_S^k)' Q^k \hat{d}_S^k < 0,$$

which, along with relation (37), shows that $q(\hat{d}_S^k) < q(d_P^k)$, thus completing the proof. $\triangleleft$

5 The overall algorithm SOLA

In this section we present Algorithm SOLA which combines the adaptive linesearch techniques (ALS) for the minimization of the augmented Lagrangian function L_a with a suitable updating rule for the penalty parameter ϵ. We show that, under some additional conditions on Q^k, d_P^k, d_S^k, the limit points generated by Algorithm SOLA are also second order stationary points of Problem (1). This motivates the name SOLA, which stands for Second Order Lagrangian Algorithm. We recall that ALS accepts as inputs (x^k,λ^k), $L_a(x^k,\lambda^k;\epsilon)$, $\nabla L_a(x^k,\lambda^k;\epsilon)$, $Q^k, d_P^k, \hat{d}_S^k$, and returns the new point $(x^{k+1},\lambda^{k+1}) = ALS(x^k,\lambda^k,d_P^k,\hat{d}_S^k;\epsilon)$.

```
Algorithm SOLA (Second Order Augmented Lagrangian Algorithm)

Data: $(y^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^m$, and $\epsilon^0 > 0$. Choose $\alpha > 0$ and $s \geq 3$ such that $y^0 \in \mathcal{P}$.

Step 0: Set $j = 0$ and $(x^0, \lambda^0) = (y^0, \mu^0)$ (outer iteration).

Step 1: Set $k = 0$ (inner iteration).

While (Stopping condition not satisfied) do

    If $\|\nabla L_a(x^k, \lambda^k; \epsilon^j)\| \geq \|\max\{g(x^k), -\epsilon^j p(x^k, \lambda^k)\lambda^k\}\|$ then
        compute two directions $d_P^k$ and $\hat{d}_S^k$
        compute $(x^{k+1}, \lambda^{k+1}) = ALS(x^k, \lambda^k, d_P^k, \hat{d}_S^k; \epsilon^j)$
        set $k = k + 1$;
    else (update $\epsilon$ and restart the inner iteration)
        Set $\epsilon^{j+1} \in (0, \epsilon^j)$, $(y^{j+1}, \mu^{j+1}) = (x^k, \lambda^k)$, $j = j + 1$.
        If $L_a(y^j, \mu^j; \epsilon^j) \leq L_a(y^0, \mu^0; \epsilon^j)$ set $(x^0, \lambda^0) = (y^j, \mu^j)$
        else set $(x^0, \lambda^0) = (y^0, \mu^0)$.
        Go to Step 1.
    End If

End while
```

Algorithm SOLA can be seen as an enhancement of algorithm ALFA, proposed in [11]. In the definition of ALFA, an iteration map $T[z^k]$ which returns the value z^{k+1} is used; T must be such that, for every fixed value of ϵ and every starting point $z^0 \in \mathcal{P} \times \mathbb{R}^m$, the sequence $\{z^k\}$ belongs to the level set $\Omega(z^0, \epsilon)$ and all limits points are stationary points of L_a (Assumption A4 of [11]). This requirement is satisfied by ALS by Proposition 4.1. Hence, in the convergence analysis of SOLA, we can use part of the analysis developed for ALFA.

Proposition 5.1 *Let $\{x^k, \lambda^k\}$ be a sequence converging to a KKT pair $(\bar{x}, \bar{\lambda})$ of Problem (1). Let the matrices Q^k and the directions d_S^k be defined as in Section 3. If*

$$\lim_{k \to \infty} (\hat{d}_S^k)' Q^k \hat{d}_S^k = 0,$$

then $(\bar{x}, \bar{\lambda})$ satisfies the second order necessary conditions for Problem (1).

Proof. The proof follows from point (d) of Condition 1, Proposition 2.6 and Proposition 3.1. $\triangleleft$

The following proposition establishes the main result of this paper.

Proposition 5.2 *Let the matrices* Q^k *and the directions* d_P^k, $\hat{d}_S^k$ *used in procedure ALS, be defined as in Section 3. Then, after having updated the penalty parameter* ϵ *at most a finite number of times, Algorithm SOLA produces an infinite sequence* $\{(x^k, \lambda^k)\}$ *such that every limit point* (x^*, λ^*) *of* $\{(x^k, \lambda^k)\}$ *is a KKT pair of Problem (1) which satisfies SONC.*

Proof. Convergence of the sequence $\{(x^k, \lambda^k)\}$ generated by Algorithm SOLA to KKT pairs of Problem 1 follows using the same arguments of the convergence analysis of algorithm ALFA of [11]. Actually SOLA differs from ALFA in the use of procedure ALS. By Proposition 4.2 ALS satisfies Assumption A4 of [11] required by Algorithm ALFA on the iteration map T. Hence, by using Theorem 7.2 of [11], we have that the penalty parameter is updated a finite number of times and that the sequence produced by Algorithm SOLA converges to a KKT pair (x^*, λ^*) of Problem (1).
In order to prove convergence to a point that satisfies SONC we proceed by contradiction and suppose that an infinite index set $\mathcal{K}$ exists such that

$$\{(x^k, \lambda^k)\}_{\mathcal{K}} \to (\bar{x}, \bar{\lambda})$$

with $(\bar{x}, \bar{\lambda})$ a KKT pair of Problem (1) which does not satisfy SONC. By Proposition 4.3, we know that for k sufficiently large, $k \in \mathcal{K}$, $d^k = \hat{d}_S^k$ in procedure ALS. Hence, by Proposition 4.2, it results

$$\lim_{k\to\infty, k\in\mathcal{K}} (\hat{d}_S^k)' Q^k \hat{d}_S^k = 0,$$

which, by Proposition 5.1, turns out to be a contradiction thus proving the result. $\triangleleft$

Finally, we point out that far from a KKT pair, namely whenever

$$\|\nabla L_a(x^k, \lambda^k; \epsilon)\| + \|\max\{g(x^k), -\epsilon p(x^k, \lambda^k)\lambda^k\}\|$$

is large, we can also set $\hat{d}_S^k = d_S^k$, namely we can set $d_N^k = 0$, in Algorithm SOLA, since we can approach KKT pairs by using only the directions d_P^k and d_S^k.

6 Superlinear convergence rate

In order to prove that the sequence of iterates produced by Algorithm SOLA has superlinear convergence rate under the SSOSC, we need to preliminary state some technical results. The first proposition shows that, if $\{(x^k, \lambda^k)\}$ is converging towards a KKT pair $(\bar{x}, \bar{\lambda})$ which satisfies SSOSC and LICQ, then, eventually, the matrices Q^k are all positive definite.

Proposition 6.1 *Let* $(\bar{x}, \bar{\lambda})$ *be a KKT pair of Problem (1) which satisfies SSOSC and LICQ, let* $\{(x^k, \lambda^k)\}$ *be a sequence converging to* $(\bar{x}, \bar{\lambda})$*. Then an integer* $\bar{k}$ *and* $\bar{\epsilon} > 0$ *exist such that for all* $k \geq \bar{k}$ *and* $\epsilon \in (0, \bar{\epsilon}]$*, the matrices* Q^k *are positive definite.*

Proof. By Proposition 2.5, we know that an $\bar{\epsilon} > 0$ exists such that, for all $\epsilon \in (0, \bar{\epsilon}]$, $(\bar{x}, \bar{\lambda})$ is an isolated local minimum point for $L_a(x, \lambda; \epsilon)$, and all matrices in $\partial^2 L_a(\bar{x}, \bar{\lambda}; \epsilon)$ are positive definite. Furthermore, by Proposition 2.3 and by the Carathéodory theorem, every matrix $\bar{H}(\bar{x}, \bar{\lambda}; \epsilon)$ in $\partial^2 L_a(\bar{x}, \bar{\lambda}; \epsilon)$ can be written in the form

$$\bar{H}(\bar{x}, \bar{\lambda}; \epsilon) = \sum_{i=1}^{t} \beta_i H(\bar{x}, \bar{\lambda}; \epsilon, A_i),$$

where $t \leq (n+m)^2+1$, $\beta_i \geq 0$, $\sum_{i=1}^{t} \beta_i = 1$, and $A_i \in \mathcal{A}$. Then, each matrix $H(\bar{x}, \bar{\lambda}; \epsilon, A_i)$ is positive definite as well. Hence, by continuity, we have that a neighborhood $\mathcal{B}(\bar{x}, \bar{\lambda})$ of $(\bar{x}, \bar{\lambda})$ exists such that $H(x, \lambda; \epsilon, A_i)$ are positive definite for every $(x, \lambda) \in \mathcal{B}(\bar{x}, \bar{\lambda})$ so that also the matrix $H(x, \lambda; \epsilon, A_\oplus)$ is positive definite. Since $\{(x^k, \lambda^k)\}$ is converging to $(\bar{x}, \bar{\lambda})$, we get that an index $\bar{k}$ exists such that, for all $k \geq \bar{k}$, all the matrices Q^k are positive definite. ◁

The next proposition shows that, if $\{(x^k, \lambda^k)\}$ is converging towards a KKT pair $(\bar{x}, \bar{\lambda})$ which satisfies SSOSC and LICQ, then, eventually, $d_S^k = 0$ so that d_P^k is guaranteed to satisfy the stopping condition of the truncated scheme, namely

$$\|r^k\| = \|Q^k d_P^k - \nabla L_a^k\| \leq \xi^k \|\nabla L_a^k\|.$$

Proposition 6.2 *Let $(\bar{x}, \bar{\lambda})$ be a KKT pair of Problem (1) which satisfies SSOSC and LICQ, let $\{(x^k, \lambda^k)\}$ be a sequence converging to $(\bar{x}, \bar{\lambda})$. Then an integer $\bar{k}$, positive constant $\bar{c}_1$ and $\bar{\epsilon} > 0$ exist such that for all $k \geq \bar{k}$, $c_1 \leq \bar{c}_1$ and $\epsilon \in (0, \bar{\epsilon}]$ it results:*

(i) $d_S^k = 0$;

(ii) $\|r^k\| \leq \xi^k \|\nabla L_a^k\|$;

(iii) if ξ^k converges to zero, then

$$\lim_{k \to \infty} \frac{\|r^k\|}{\|\nabla L_a^k\|} = 0. \tag{38}$$

Proof. Point (i). By Proposition 6.1, we know that an integer $\bar{k}$ and $\bar{\epsilon} > 0$ exist such that, for all $k \geq \bar{k}$ and $\epsilon \in (0, \bar{\epsilon}]$, all the matrices Q^k are positive definite. Hence, the conjugate gradient scheme will not generate any negative curvature direction and $d_S^k = 0$.
Points (ii) and (iii) follows from Proposition 2.6 in [9] and point (i) above. ◁
The next two propositions establish properties that the direction d^k computed as described in Sections 3 and 4 eventually satisfies.

Proposition 6.3 *Let $(\bar{x}, \bar{\lambda})$ be a KKT pair of Problem (1) which satisfies SSOSC and LICQ and let $\{(x^k, \lambda^k)\}$ be a sequence converging to $(\bar{x}, \bar{\lambda})$. Let d^k be calculated by the conjugate gradient scheme proposed in [18] applied to the solution of system $Q^k d = -\nabla L_a^k$. Then, eventually, the following inequalities hold.*

$$(d^k)' \nabla L_a^k \leq -\rho 1 \|\nabla L_a^k\|^2, \tag{39}$$

$$\|d^k\| \leq \rho_2 \|\nabla L_a^k\|, \tag{40}$$

$$(d^k)' \nabla L_a^k \leq -\frac{\rho_1}{\rho_2^2} \|d^k\|^2, \tag{41}$$

where ρ_1 and ρ_2 are two positive constants not depending on the index k.

Proof. We recall that, by Proposition 6.2, for k sufficiently large, $d_S^k = 0$, that is no negative curvature direction is generated by the conjugate gradient of [18]. Hence, the conjugate gradient of [18] turns out to be exactly the same as the one proposed in [20]. Then, satisfaction of (39), (40) and (41) is proved in Theorem 2.2(c) of [20]. ◁

Proposition 6.4 (Proposition 2.7 in [9]) *Let $(\bar{x}, \bar{\lambda})$ be a KKT pair of Problem (1) which satisfies SSOSC and LICQ and let $\{(x^k, \lambda^k)\}$ be a sequence converging to $(\bar{x}, \bar{\lambda})$ and $\{d^k\}$ be the sequence of direction computed as described in Sections 3 and 4. Moreover, assume that (38) holds. Then, we have*

$$\lim_{k\to\infty} \frac{\left\| \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix} + d^k - \begin{pmatrix} \bar{x} \\ \bar{\lambda} \end{pmatrix} \right\|}{\left\| \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \bar{x} \\ \bar{\lambda} \end{pmatrix} \right\|} = 0. \tag{42}$$

Proof. We recall that, by Propositions 6.1 and 6.2 and by Condition 1, for k sufficiently large, $\hat{d}_S^k = 0$. Thus, Step 1 of the ALS procedure will eventually accept d_P^k as search direction. Then, the result follows by considering (38) and reasoning as in [6]. $\triangleleft$

Now we are ready to state the main proposition concerning the superlinear rate of convergence of the sequence of iterates produced by the algorithm.

Proposition 6.5 *Let $(\bar{x}, \bar{\lambda})$ be a KKT pair of Problem (1) which satisfies SSOSC and LICQ and let $\{(x^k, \lambda^k)\}$ be a sequence converging to $(\bar{x}, \bar{\lambda})$ and $\{d^k\}$ be the sequence of directions computed as described in Sections 3 and 4. Then, eventually the unit stepsize is accepted by the ALS procedure, and the rate of convergence is superlinear.*

Proof. By Proposition 6.3, we know that eventually d^k satisfies (41). This fact, along with Proposition 6.4 and the consideration that L_a is an SC^1 function, allow us to apply Theorem 3.3 of [14] which guarantees that, eventually, the unit stepsize is always accepted by the ALS procedure. This fact, and again Proposition 6.4, conclude the proof. $\triangleleft$

7 Conclusions

In this paper we propose a method for the solution of inequality constrained nonlinear programming problem which is globally convergent to KKT pairs that satisfy second order necessary optimality conditions. To this aim we minimize an augmented Lagrangian function, which enjoys strong exactness properties, on the product space of primal and dual variables. The minimization algorithm hinges on the idea of comparing, at every iteration, the relative effects of two directions, one of them being a negative curvature direction, and then selecting the more promising one. The selection rule is such that convergence to second order stationary points can be guaranteed. Moreover, under the strong second order sufficiency optimality condition, the convergence rate of the algorithm is superlinear.

As a final remark, we note that the problem considered in the paper is the general inequality constrained optimization problem. The extension to the case when both, equality and inequality constraints are present is straightforward but cumbersome. It can be done as, e.g., in [8] and [10].

References

[1] Auslender, A. Penalty methods for computing points that satisfy second order necessary conditions. *Mathematical Programming* **17**, 229–238 (1979).

[2] Byrd, R.H., Schnabel, R.B., Shultz, G.A. A trust region algorithm for nonlinearly constrained optimization. *SIAM Journal on Numerical Analysis* **24**, 1152–1170 (1987).

[3] Clarke, F.H. *Optimization and Nonsmooth Analysis,* John Wiley & Sons, New York (1983).

[4] Coleman, T.F., Yuan, W. A new trust-region algorithm for equality constrained optimization. *Computational Optimization and Applications*, **21**, 177–199 (2002).

[5] Conn, A.R., Gould, N.I.M., Orban, D., Toint, Ph.L. A primal-dual trust region algorithm for non-convex nonlinear programming. *Math. programming, Ser. B* **87**, 215–249 (2000).

[6] Dembo, R.S., Eisenstat, S.C., Steihaug, T. Inexact Newton Methods, *SIAM Journal on Numerical Anlysis* **19**, 400-408 (1982).

[7] Dennis, J.E., Vicente, L.N. On the convergence theory of trust-region-based algorithms for equality-constrained optimization, *SIAM Journal on Optimization* **7**, 527-550 (1994).

[8] Di Pillo, G., Grippo, L. Exact penalty functions in constrained optimization. *SIAM J. on Control and Optimization* **27**, 1333–1360 (1989).

[9] Di Pillo, G., Liuzzi, G., Lucidi, S., Palagi, L. A truncated Newton method in an augmented Lagrangian framework for nonlinear programming. *Computational Optimization and Applications* **45**, pp. 311-352 (2010).

[10] Di Pillo, G., Lucidi, S. On exact Augmented Lagrangian functions in nonlinear programming. In *Nonlinear Optimization and Applications*, G. Di Pillo and F. Giannessi, eds., Plenum Press, New York, pp. 85–100 (1996).

[11] Di Pillo, G., Lucidi, S. An augmented Lagrangian function with improved exactness properties. *SIAM Journal on Optimization*, **12**, pp. 376–406 (2001).

[12] Di Pillo, G., Lucidi, S., Palagi, L. Convergence to second-order stationary points of a primal-dual algorithm model for nonlinear programming. *Mathematics of Operations Research*, **30**, pp. 897–915 (2005).

[13] El-Alem, M. Convergence to a second-order point of a trust-region algorithm with a nonmonotonic penalty parameter for constrained optimization, *Journal of Optimization Theory and Applications*, **91**, 61–79 (1996).

[14] Facchinei, F. Minimization of SC1 functions and the Maratos effect. *Operations Research Letters*, **17**, 131–137 (1995).

[15] Facchinei, F., Fischer, A., Kanzow, C. On the accurate identification of active constraints. *SIAM Journal on Optimization*, **9**, 14–32 (1998).

[16] Facchinei, F., Lucidi, S. Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problems. *Journal of Optimization Theory and Applications*, **85**, 265–289 (1995).

[17] Facchinei, F., Lucidi, S. Convergence to second order stationary points in inequality constrained optimization. *Mathematics of Operations Research*, **23**, 746–766 (1998).

[18] Fasano, G., Lucidi, S. A nonmonotone truncated Newton-Krylov method exploiting negative curvature directions, for large scale unconstrained optimization. *Optimization Letters*, **3**, 521–535 (2009).

[19] Forsgren, A., Murray, W. Newton methods for large-scale linear inequality constrained minimization. *SIAM Journal on Optimization* **7**, 162-176 (1997).

[20] Grippo, L. , Lampariello, F., Lucidi, S. A truncated Newton method with nomonotone linesearch for unconstrained optimization. *Journal of Optimization Theory and Applications*, **60**, pp. 401–419 (1989).

[21] Hiriart-Urruty, J.-B., Strodiot, J.J., Nguyen, V.H. Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data. *Applied Mathematics and Optimization*, **11**, 43–56 (1984).

[22] Lucidi, S., Rochetich, F., Roma, M. Curvilinear stabilization techniques for truncated Newton methods in large scale unconstrained optimization: The complete results, *SIAM Journal on Optimization*, **8**, 916–939 (1998).

[23] McCormick, G.P. *Nonlinear Programming: Theory, Algorithms and Applications,* John Wiley & Sons, New York (1983).

[24] Moguerza, J.M., Prieto, F.J. An augmented Lagrangian interior point method using directions of negative curvature. *Mathematical Programming*, **95**, 573–616 (2003).

[25] Moré, J.J., Sorensen, D.C. On the use of directions of negative curvature in a modified Newton method. *Mathematical Programming*, **16**, 1–20 (1979).

[26] Moré, J.J., Sorensen, D.C. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, **4**, 553–572 (1983).

[27] Mukai, H., Polak, E. A second-order method for the general nonlinear programming problem. *Journal of Optimization Theory and Applications*, **26**, 515–532 (1978).

[28] Qi, L., Sun J. A nonsmooth version of Newton's method. *Mathematical Programming*, **58**, 353–367 (1993).

[29] Sorensen, D.C. Newton's method with a model trust region modification. *SIAM Journal on Numerical Analysis*, **19**, 409–426 (1982).

[30] Tseng, P. A convergent infeasible interior-point trust-region methods for constrained optimization. *SIAM Journal on Optimization*, **13**, 432–469 (2002).

[31] Vicente, L.N. Dennis, J. E., Heinkenschloss, M., Vicente, L.N. Trust-region interior-point SQP algorithms for a class of nonlinear programming problems, *SIAM Journal on Control and Optimization*, **36** 1750–1794 (1998).