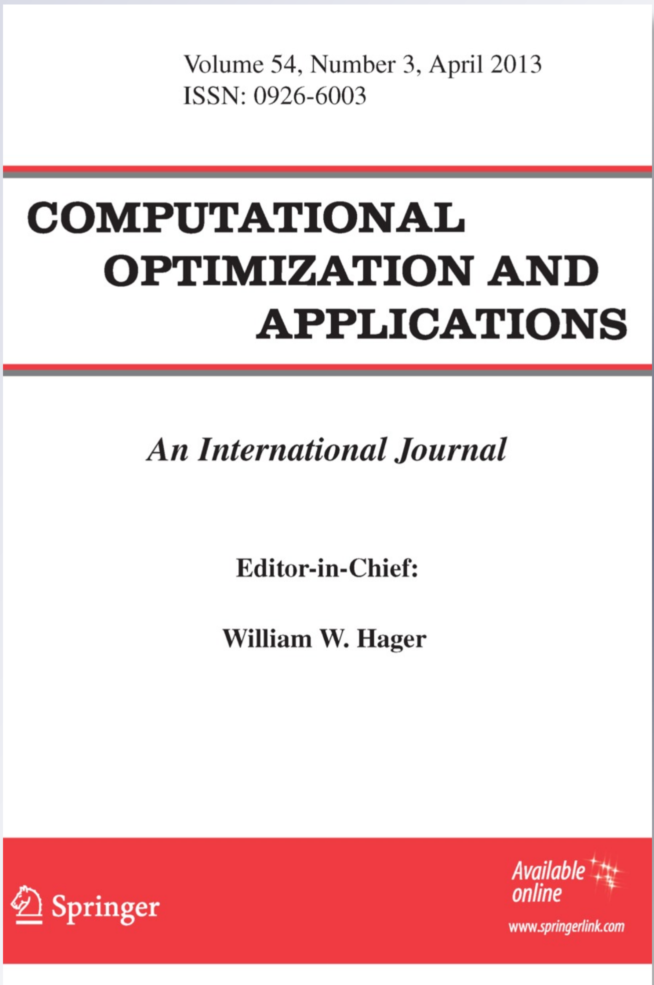
An inexact restoration strategy for the globalization of the sSQP method

D. Fernández, E. A. Pilotta & G. A. Torres

Computational Optimization and Applications
An International Journal

ISSN 0926-6003
Volume 54
Number 3

Comput Optim Appl (2013) 54:595-617
DOI 10.1007/s10589-012-9502-y



Springer

Comput Optim Appl (2013) 54:595–617
DOI 10.1007/s10589-012-9502-y

An inexact restoration strategy for the globalization of the sSQP method

D. Fernández · E.A. Pilotta · G.A. Torres

Received: 8 August 2011 / Published online: 22 August 2012

Abstract A globally convergent algorithm based on the stabilized sequential quadratic programming (sSQP) method is presented in order to solve optimization problems with equality constraints and bounds. This formulation has attractive features in the sense that constraint qualifications are not needed at all. In contrast with classic globalization strategies for Newton-like methods, we do not make use of merit functions. Our scheme is based on performing corrections on the solutions of the subproblems by using an inexact restoration procedure. The presented method is well defined and any accumulation point of the generated primal sequence is either a Karush-Kuhn-Tucker point or a stationary (maybe feasible) point of the problem of minimizing the infeasibility. Also, under suitable hypotheses, the sequence generated by the algorithm converges Q-linearly. Numerical experiments are given to confirm theoretical results.

Keywords Sequential quadratic programming · Nonlinear programming · Global convergence · Augmented Lagrangian

This work was carried out with the aid of grants from ANPCyT, CONICET and SECyT-UNC.

D. Fernández · E.A. Pilotta (✉) · G.A. Torres
Facultad de Matemática, Astronomía y Física, CIEM (CONICET), Universidad Nacional de Córdoba, Medina Allende s/n, Ciudad Universitaria, 5000 Córdoba, Argentina
e-mail: pilotta@famaf.unc.edu.ar

D. Fernández
e-mail: dfernandez@famaf.unc.edu.ar

G.A. Torres
e-mail: torres@famaf.unc.edu.ar

1 Introduction

Given $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ twice continuously differentiable, we want to solve the following constrained nonlinear program,

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \quad x \in \Omega, \end{array} \tag{1}$$

where $\Omega = \{x \in \mathbb{R}^n \mid a \leq x \leq b\}$ with $a, b \in \mathbb{R}^n$. The natural residual $\sigma : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ associated to problem (1) is given by

$$\sigma(x, \lambda) = \left\| \begin{bmatrix} \Pi_\Omega(x - \frac{\partial L}{\partial x}(x, \lambda)) - x \\ h(x) \end{bmatrix} \right\|, \tag{2}$$

where Π_Ω denotes the orthogonal projection onto Ω and $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ denotes the Lagrangian function of problem (1), i.e., $L(x, \lambda) = f(x) + \langle \lambda, h(x) \rangle$. Thus, x is a stationary point of problem (1) with associated Lagrange multipliers λ if and only if $\sigma(x, \lambda) = 0$.

In order to solve (1), we propose to use the stabilized sequential quadratic programming (sSQP) method with a suitable strategy to force its global convergence. Recall that at a given primal-dual iterate $(x^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^m$, a quasi-Newton sSQP subproblem has the form

$$\begin{array}{ll} \underset{(x,\lambda)}{\text{minimize}} & \left\langle \nabla f(x^k), x - x^k \right\rangle + \frac{1}{2}\left\langle Q_k(x - x^k), x - x^k \right\rangle + \frac{1}{2\rho_k}\|\lambda\|^2 \\ \text{subject to} & h(x^k) + \nabla h(x^k)^\top (x - x^k) - \frac{1}{\rho_k}(\lambda - \lambda^k) = 0, \quad x \in \Omega, \end{array} \tag{3}$$

where $\rho_k > 0$ is a penalty parameter and $Q_k \in \mathbb{R}^{n \times n}$ is an approximation to the Hessian of the Lagrangian of the problem (1).

The sSQP method was studied by Wright [33–35] to deal with optimization problems with degenerate constraints. This method is known to be locally convergent with quadratic/superlinear rate near any solution with associated Lagrange multipliers satisfying the second-order sufficient condition (SOSC), even in those cases where no constraint qualifications are satisfied at this solution (see [16, 17, 20]). For equality constrained problems, local convergence has recently been studied in [22] for solutions with noncritical Lagrange multipliers (weaker than SOSC) and without any constraint qualification assumptions. A quasi-Newton strategy was studied in [15], showing that the classical BFGS update can be used to generate a locally superlinear convergent primal-dual sequence.

Before introducing the globally convergent method, we shall explain the mathematical concepts that it involves. To this end, for $\lambda^k \in \mathbb{R}^m$ and $\rho_k > 0$ fixed, we define the following auxiliary functions

$$F_k(x, \lambda) = f(x) + \frac{1}{2\rho_k}\|\lambda\|^2, \tag{4}$$

$$H_k(x, \lambda) = h(x) - \frac{1}{\rho_k}\left(\lambda - \lambda^k\right), \tag{5}$$

and for each $x \in \mathbb{R}^n$ we consider the point

$$Y^k(x) = \big(x, \lambda^k + \rho_k h(x)\big). \tag{6}$$

Note that with this choice we have $H_k(Y^k(x)) = 0$ for all $x \in \mathbb{R}^n$. Also, after some algebraics it can be shown that for $Y^k = Y^k(x^k)$ and $Y = (x, \lambda)$ the problem

$$\begin{array}{ll} \underset{Y}{\text{minimize}} & \big\langle \nabla F_k\big(Y^k\big), Y - Y^k\big\rangle + \dfrac{1}{2}\left\langle \begin{bmatrix} Q_k & 0 \\ 0 & \frac{1}{\rho_k} I \end{bmatrix}\big(Y - Y^k\big), Y - Y^k \right\rangle \\ \text{subject to} & \nabla H_k\big(Y^k\big)^\top \big(Y - Y^k\big) = 0, \quad Y \in \Omega \times \mathbb{R}^m. \end{array} \tag{7}$$

is equivalent to the problem (3). Therefore, the k-th quasi-Newton sSQP subproblem (3) associated to the problem (1) is the same as a quasi-Newton SQP subproblem (7) associated to the problem

$$\begin{array}{ll} \underset{Y}{\text{minimize}} & F_k(Y) \\ \text{subject to} & H_k(Y) = 0, \quad Y \in \Omega \times \mathbb{R}^m. \end{array} \tag{8}$$

Thus, local convergence of the sSQP method is obtained by solving the approximate problem (7) instead of the subproblem (8). It is a well known fact [16, 20] that near to a primal-dual pair satisfying SOSC the local quadratic convergence is obtained solving just the quadratic problem (3) (or problem (7)). When the current iterate (x^k, λ^k) is far from a primal-dual pair satisfying SOSC, it is not clear why a single quadratic problem is enough. Therefore, we propose to solve problem (8) by using inexact restoration ideas.

Inexact restoration (IR) methods were introduced in [26] and modified in [6, 18, 25]. A survey on this subject can be found in [27]. The advantage of using these methods to solve problem (8) is the fact that a feasible point is always known since $H_k(Y^k(x)) = 0$ for all x (avoiding one of the phases of each iteration) and that subproblem (7) provides a suitable tangent direction that satisfies sufficient conditions for convergence, according to [18].

In this paper we develop a hybrid method that combines two well known strategies taking advantage of their individual features. On one hand, we have feasibility and good local behavior of the sSQP method. On the other hand, we obtain global convergence by updating the penalty parameter in the same way as the penalty parameter is updated in augmented Lagrangian methods. Moreover, the IR scheme is computationally attractive, in the sense that the restoration phase is straightforward, and therefore we need to solve only linearly constrained quadratic problems to obtain the inexact solution of the subproblem.

The paper is structured as follows. In Sect. 2 the last results on IR methods are summarized. The proposed algorithm and its well-definition is described in Sect. 3. The main result, global convergence of the sequence generated by the algorithm, is presented in Sect. 4. Local convergence and penalty boundedness results are treated in Sect. 5. Section 6 is devoted to numerical experiments and conclusions are given in Sect. 7.

In what follows we describe our notation. We use $\langle \cdot, \cdot \rangle$ to denote the Euclidean inner product and $\| \cdot \|$ its associated norm. When in matrix notation, vectors are considered columns. We denote by I the identity matrix and by e the vector of ones (the dimension is always clear from the context). For a function $g:\mathbb{R}^n\to\mathbb{R}$, ∇g is a column vector where the i-th component is $\frac{\partial g}{\partial x_i}$. For a function $G:\mathbb{R}^n\to\mathbb{R}^m$, ∇G is a $n\times m$ matrix where the i,j component is $\frac{\partial G_j}{\partial x_i}$. The normal cone to a set Ω at x is defined by $\mathcal{N}_\Omega(x)=\{v\in\mathbb{R}^m \mid \langle v, y-x\rangle \le 0\ \forall\, y\in\Omega\}$ if $x\in\Omega$, or $\mathcal{N}_\Omega(x)=\emptyset$ otherwise.

2 IR methods

Each iteration of IR methods is divided in two phases: restoration and minimization. In the restoration phase, given an iterate X^k, an intermediate point Y^k is computed (called restored point) in order to improve feasibility without deteriorating the objective function value. A merit function is defined combining feasibility and optimality, including a penalty parameter that changes between different iterations. In the minimization phase a line search is performed to the merit function along a direction D^k belonging to the first order feasible direction set at Y^k.

In order to solve the problem

$$\begin{array}{ll} \text{minimize} & F(x) \\ \text{subject to} & H(x)=0, \quad x\in\mathcal{W}, \end{array} \tag{9}$$

where $\mathcal{W}$ is a convex and compact set, we describe the Fischer-Friedlander IR model algorithm:

Algorithm 1 (Fischer-Friedlander model algorithm) Let $r\in(0,1)$, β, η, $\bar\eta$, τ be fixed.

Step 0: *Initialization*
Choose $X^0\in\mathcal{W}$ and $\theta_0\in(0,1)$. Set $j=0$.
Step 1: *Inexact restoration*
Compute $Y^j\in\mathcal{W}$ such that:

$$\|H(Y^j)\| \le r\|H(X^j)\|, \tag{10}$$

$$F(Y^j) \le F(X^j)+\beta\|H(X^j)\|. \tag{11}$$

Step 2: *Search direction*
Compute $D^j\in\mathbb{R}^n$ such that $Y^j+D^j\in\mathcal{W}$ and

$$F(Y^j+tD^j) \le F(Y^j)-\eta t\|D^j\|^2, \tag{12}$$

$$\|H(Y^j+tD^j)\| \le \|H(Y^j)\|+\bar\eta t^2\|D^j\|^2, \tag{13}$$

holds for all $t\in[0,\tau]$.
Step 3: *Penalty parameter*
Determine $\theta_{j+1}\in\{2^{-i}\theta_j : i\in\mathbb{N}\cup\{0\}\}$ as large as possible such that:

$$\phi(Y^j,\theta_{j+1})-\phi(X^j,\theta_{j+1}) \le \frac{(r-1)}{2}\left(\|H(X^j)\|-\|H(Y^j)\|\right), \tag{14}$$

where $\phi(X,\theta) = \theta F(X) + (1-\theta)\|H(X)\|$ is a merit function.
Step 4: *Line search*
Determine $t_j \in \{2^{-i} : i \in \mathbb{N} \cup \{0\}\}$ as large as possible such that:

$$\phi\big(Y^j + t_j D^j, \theta_{j+1}\big) - \phi\big(X^j, \theta_{j+1}\big) \le \frac{(r-1)}{2}\big(\big\|H\big(X^j\big)\big\| - \big\|H\big(Y^j\big)\big\|\big). \tag{15}$$

Step 5: *Update*
Set $X^{j+1} = Y^j + t_j D^j$ and $j = j+1$. Go to Step 1.

The main result in [18] is that any sequence of search directions generated by Algorithm 1 tends to zero.

Theorem 1 [18, Theorem 2] *Suppose that Step* 1 *of Algorithm* 1 *is well defined. Then,*

$$\lim_{j\to\infty} D^j = 0. \tag{16}$$

3 Description of the algorithm

We begin this section by introducing the proposed algorithm.

Algorithm 2 Let $\gamma \in (0,1)$, $r \in (0,1)$, $\varepsilon > 0$, $\{\epsilon_k\}$ with $\epsilon_k \searrow 0$ and $\alpha_L, \alpha_U > 0$. For a current parameter ρ_k we call

$$\Pi_k(x,\lambda) = \big(x, \max\big\{-\alpha_L\sqrt{\rho_k}e, \min\{\lambda, \alpha_U\sqrt{\rho_k}e\}\big\}\big). \tag{17}$$

Step 0: *Initialization*
Choose $X^0 = (x^0, \lambda^0) \in \Omega \times \mathbb{R}^m$ an arbitrary initial approximation, $\rho_0 > 0$ an initial parameter, $\psi_{-1} = \sigma(X^0)$ and $k = 0$.
Step 1: *Stopping criterion*
If the condition

$$\sigma\big(X^k\big) \le \varepsilon \tag{18}$$

is satisfied, terminate the execution of the algorithm, declaring that the residual is less than the tolerance ε.
Step 2: *Solve subproblem*

Step 2.0: Set $X^{k,0} = (x^{k,0}, \lambda^{k,0}) = (x^k, \lambda^k)$, $\theta_0 \in (0,1)$, $Q_{k,0} \in \mathbb{R}^{n\times n}$ a symmetric positive definite and $j = 0$.
Step 2.1: Set $Y^{k,j} = (x^{k,j}, \lambda^k + \rho_k h(x^{k,j}))$.
Step 2.2: Find $D^{k,j} \in \mathbb{R}^n \times \mathbb{R}^m$ solution of

$$\begin{aligned} \underset{D}{\text{minimize}} \quad & \big\langle \nabla F_k\big(Y^{k,j}\big), D\big\rangle + \frac{1}{2}\left\langle \begin{bmatrix} Q_{k,j} & 0 \\ 0 & \frac{1}{\rho_k} I \end{bmatrix} D, D \right\rangle \\ \text{subject to} \quad & \nabla H_k\big(Y^{k,j}\big)^\top D = 0, \quad Y^{k,j} + D \in \Omega \times \mathbb{R}^m. \end{aligned} \tag{19}$$

If $\|D^{k,j}\| < \epsilon_k$ then set $X^{k+1} = \Pi_k(Y^{k,j} + D^{k,j})$ and go to Step 3.
Step 2.3: Determine $\theta_{j+1} \in \{2^{-i}\theta_j : i \in \mathbb{N} \cup \{0\}\}$ as large as possible such that:

$$\phi_k\big(Y^{k,j}, \theta_{j+1}\big) - \phi_k\big(X^{k,j}, \theta_{j+1}\big) \le \frac{(r-1)}{2}\big\|H_k\big(X^{k,j}\big)\big\|, \tag{20}$$

where $\phi_k(X,\theta) = \theta F_k(X) + (1-\theta)\|H_k(X)\|$ is a merit function.
Step 2.4: Determine $t_j \in \{2^{-i} : i \in \mathbb{N} \cup \{0\}\}$ as large as possible such that:

$$\phi_k\big(Y^{k,j} + t_j D^{k,j}, \theta_{j+1}\big) - \phi_k\big(X^{k,j}, \theta_{j+1}\big) \le \frac{(r-1)}{2}\big\|H_k\big(X^{k,j}\big)\big\|. \tag{21}$$

Step 2.5: Set $X^{k,j+1} = Y^{k,j} + t_j D^{k,j}$, and choose $Q_{k,j+1} \in \mathbb{R}^{n\times n}$ symmetric positive definite, and $j = j+1$. Go to Step 2.1.

Step 3: *Update the penalty parameter*
Set $\psi_k = \min\{\psi_{k-1}, \sigma(X^k)\}$. If $\|h(x^{k+1})\| > \gamma\|h(x^k)\|$ and $\sigma(X^{k+1}) > \gamma\psi_k$, then set $\rho_{k+1} = 10\rho_k$. Otherwise, set $\rho_{k+1} = \rho_k$.

Besides that, set $k = k+1$ and go to Step 1.

Remark 1 Subproblem (19) is similar to (7) but centered at $Y^{k,j} = Y^k(x^{k,j})$ instead of $Y^k = Y^k(x^k)$ (see (6)). Also, it can be seen that if $D^{k,j}$ is a solution of (19), then $(x,\lambda) = Y^{k,j} + D^{k,j}$ is a solution of

$$\begin{array}{ll} \underset{(x,\lambda)}{\text{minimize}} & \big\langle\nabla f\big(x^{k,j}\big), x - x^{k,j}\big\rangle + \frac{1}{2}\big\langle Q_{k,j}\big(x - x^{k,j}\big), x - x^{k,j}\big\rangle + \frac{1}{2\rho_k}\|\lambda\|^2 \\ \text{subject to} & h\big(x^{k,j}\big) + \nabla h\big(x^{k,j}\big)^\top\big(x - x^{k,j}\big) - \frac{1}{\rho_k}\big(\lambda - \lambda^k\big) = 0, \quad x \in \Omega, \end{array} \tag{22}$$

which is similar to (3) but centered at $x^{k,j}$ instead of x^k. When $X^{k+1} = Y^{k,0} + D^{k,0}$ then the sequence generated by Algorithm 2 is the same as the sequence generated by the sSQP method. This update can be generated if (a) ρ_k is large enough, implying that Π_k (17) shall be avoided; and (b) the current iterate is near to a primal-dual pair satisfying the SOSC, which guarantees that $D^{k,0}$ shall be small.

The remaining part of this section is devoted to show that Algorithm 2 is well defined, which depends on the consistency of Step 2.

Let us define the function

$$\mu^k(x,y) = \lambda^k + \rho_k h(x) + \rho_k \nabla h(x)^\top(y-x), \quad x, y \in \Omega. \tag{23}$$

Since h and ∇h are continuous and Ω is a compact set, we deduce that there exists λ_L^k and λ_U^k depending on (λ^k, ρ_k) such that λ^k and $\mu^k(x,y)$ belong to the interior of $[\lambda_L^k, \lambda_U^k]$ for all $x, y \in \Omega$. Thus, there exists a compact set $\mathcal{W}_k = \Omega \times [\lambda_L^k, \lambda_U^k] \subset \mathbb{R}^n \times \mathbb{R}^m$ such that problem (8) is equivalent to

$$\begin{array}{ll} \underset{(x,\lambda)}{\text{minimize}} & F_k(x,\lambda) \\ \text{subject to} & H_k(x,\lambda) = 0, \quad (x,\lambda) \in \mathcal{W}_k. \end{array} \tag{24}$$

It can be seen that Step 2 of Algorithm 2 is a direct application of Algorithm 1 applied to the problem (24). Therefore, we have to show that hypotheses of Theorem 1 hold.

Notice that if f and h are twice continuously differentiable in Ω, then F_k and H_k are twice continuously differentiable in $\mathcal{W}_k$.

In the following two lemmas we prove that conditions (10), (11), (12) and (13) of Algorithm 1 are satisfied.

Lemma 1 *Let, for a fixed* k, $\{X^{k,j}\}$, $\{Y^{k,j}\}$ *and* $\{D^{k,j}\}$ *be the sequences generated by Algorithm 2. Then*

(a) $Y^{k,j} + D^{k,j}$ *and* $X^{k,j}$ *belong to* $\mathcal{W}_k$ *for all* j *if* $x^k \in \Omega$.
(b) *There exists* $\beta_k > 0$ *such that*

$$\left\| H_k\left(Y^{k,j}\right)\right\| \le r \left\| H_k\left(X^{k,j}\right)\right\|, \tag{25}$$

$$F_k\left(Y^{k,j}\right) \le F_k\left(X^{k,j}\right) + \beta_k \left\| H_k\left(X^{k,j}\right)\right\|, \tag{26}$$

for all $r > 0$.

Proof We will prove (a) by induction in j. From the Step 2.0 and the definition of λ_L^k and λ_U^k, and the fact that $x^k \in \Omega$, we have that $X^{k,0} = (x^{k,0}, \lambda^{k,0}) = (x^k, \lambda^k) \in \mathcal{W}_k$. Let $j \ge 0$ and suppose that $X^{k,j} \in \mathcal{W}_k$. Let us define $(x, \lambda) = Y^{k,j} + D^{k,j}$. From the definition of $Y^{k,j}$ we have $D^{k,j} = (x - x^{k,j}, \lambda - (\lambda^k + \rho_k h(x^{k,j})))$. From the equality constraint in (19) we obtain

$$\begin{aligned} 0 &= \nabla H_k\left(Y^{k,j}\right)^\top D^{k,j} \\ &= \nabla h\left(x^{k,j}\right)^\top \left(x - x^{k,j}\right) - \frac{1}{\rho_k}\left(\lambda - \lambda^k\right) + h\left(x^{k,j}\right) \end{aligned}$$

Solving for λ and using (23) we get $\lambda = \mu^k(x^{k,j}, x)$. Now, on one hand we have that $x \in \Omega$ since $Y^{k,j} + D^{k,j} \in \Omega \times \mathbb{R}^m$. On the other hand we have that $x^{k,j} \in \Omega$ because $X^{k,j} \in \mathcal{W}_k$. Combining these facts we obtain that $\lambda = \mu^k(x^{k,j}, x) \in [\lambda_L^k, \lambda_U^k]$ and therefore $Y^{k,j} + D^{k,j} \in \mathcal{W}_k$. Since $Y^{k,j} = (x^{k,j}, \mu^k(x^{k,j}, x^{k,j})) \in \mathcal{W}_k$, $t_j \in [0,1]$ and the convexity of $\mathcal{W}_k$, we have that $X^{k,j+1} = Y^{k,j} + t_j D^{k,j} = (1 - t_j)Y^{k,j} + t_j(Y^{k,j} + D^{k,j}) \in \mathcal{W}_k$.

Next, we will prove (b). By the definition of $Y^{k,j}$ in Step 2.1, we can see that condition (25) holds since $H_k(Y^{k,j}) = 0$.

Using that $X^{k,j} = (x^{k,j}, \lambda^{k,j})$ and $\lambda^k + \rho_k h(x^{k,j}) = \lambda^{k,j} + \rho_k H_k(X^{k,j})$, we can see that

$$\begin{aligned} F_k\left(Y^{k,j}\right) - F_k\left(X^{k,j}\right) &= \frac{1}{2\rho_k}\left(\left\|\lambda^{k,j} + \rho_k H_k\left(X^{k,j}\right)\right\|^2 - \left\|\lambda^{k,j}\right\|^2\right) \\ &= \frac{1}{2\rho_k}\left(\left\|\rho_k H_k\left(X^{k,j}\right)\right\|^2 + 2\rho_k\left\langle H_k\left(X^{k,j}\right), \lambda^{k,j}\right\rangle\right) \end{aligned}$$

$$\leq \|H_k(X^{k,j})\| \left(\frac{\rho_k}{2} \|H_k(X^{k,j})\| + \|\lambda^{k,j}\| \right)$$
$$\leq \beta_k \|H_k(X^{k,j})\|,$$

where $\beta_k > 0$ is a constant that exists because of the continuity of H_k, the compactness of $\mathcal{W}_k$ and the fact that $X^{k,j} \in \mathcal{W}_k$. Therefore, condition (26) holds. □

It remains to prove that the direction $D^{k,j}$ generated by the subproblem (19) satisfies conditions (12) and (13).

Lemma 2 *Suppose that, for a fixed k, matrices $\{Q_{k,j}\}$ are uniformly positive definite. Then there exist positive constants η_k, $\bar{\eta}_k$ and τ_k such that*

$$F_k(Y^{k,j} + tD^{k,j}) \leq F_k(Y^{k,j}) - \eta_k t \|D^{k,j}\|^2, \tag{27}$$

$$\|H_k(Y^{k,j} + tD^{k,j})\| \leq \|H_k(Y^{k,j})\| + \bar{\eta}_k t^2 \|D^{k,j}\|^2, \tag{28}$$

hold for all $t \in [0, \tau_k]$.

Proof Since $Y^{k,j} \in \Omega \times \mathbb{R}^m$ then $D = 0$ is feasible for the problem (19). Hence, the solution $D^{k,j}$ satisfies

$$\langle \nabla F_k(Y^{k,j}), D^{k,j} \rangle + \frac{1}{2} \left\langle \begin{bmatrix} Q_{k,j} & 0 \\ 0 & \frac{1}{\rho_k} I \end{bmatrix} D^{k,j}, D^{k,j} \right\rangle \leq 0.$$

Assuming that matrices $Q_{k,j}$ are uniformly positive definite, the there exists a constant $c_k > 0$ such that

$$\langle \nabla F_k(Y^{k,j}), D^{k,j} \rangle \leq -\frac{c_k}{2} \|D^{k,j}\|^2. \tag{29}$$

Let $L_k > 0$ be the Lipschitzian modulus of ∇F_k and ∇H_k (because of smoothness of f and h).

By the Taylor's formula we obtain

$$F_k(Y + tD) = F_k(Y) + t\langle \nabla F_k(Y), D \rangle + t \int_0^1 \langle \nabla F_k(Y + stD) - \nabla F_k(Y), D \rangle ds,$$

then by using (29) and Lipschitzianity of ∇F_k we get

$$\begin{aligned} F_k(Y^{k,j} + tD^{k,j}) &\leq F_k(Y^{k,j}) - \frac{c_k t}{2} \|D^{k,j}\|^2 + \frac{L_k t^2}{2} \|D^{k,j}\|^2 \\ &= F_k(Y^{k,j}) - \left(\frac{c_k}{2} - \frac{L_k t}{2} \right) t \|D^{k,j}\|^2, \end{aligned}$$

for all $t \in [0, 1]$. Therefore, (27) is valid for all $t \in [0, \tau_k]$ with $\tau_k = \min\{1, \frac{c_k}{2L_k}\}$ and $\eta_k = c_k/4$.

Similarly, using that $\nabla H_k(Y^{k,j})^\top D^{k,j} = 0$ and Lipschitzianity of ∇H_k we have

$$\|H_k(Y^{k,j} + tD^{k,j})\| \le \|H_k(Y^{k,j})\| + \frac{L_k t^2}{2}\|D^{k,j}\|^2,$$

for all $t \in [0, 1]$. Thus, (28) holds for all $t \in [0, 1]$ with $\bar{\eta}_k = L_k/2$. Therefore, (27) and (28) are valid for all $t \in [0, \tau_k]$ with η_k, $\bar{\eta}_k$ and τ_k as defined in this proof. □

The next lemma assures us that a sequence $\{X^k\}$ can be generated by Algorithm 2.

Lemma 3 *Algorithm 2 is well defined and generates sequences* $\{(x^k, \lambda^k)\}$, *where* $x^k \in \Omega$ *and* $\lambda^{k+1} \in [-\alpha_L\sqrt{\rho_k}e, \alpha_U\sqrt{\rho_k}e]$ *for all* k.

Proof Observe that $x^0 \in \Omega$ (from Step 0). Let us assume that $x^k \in \Omega$ for $k \ge 0$. Because of Lemmas 1 and 2, the compactness of $\mathcal{W}_k$ and Theorem 1, we have that the directions $D^{k,j}$ converge to zero when j tends to infinity. Thus, the condition $\|D^{k,j}\| \le \epsilon_k$ is satisfied for j sufficiently large, so Step 2 of Algorithm 2 is executed only a finite number of iterations. Therefore, the next iterate X^{k+1} can be generated. Since $Y^{k,j} + D^{k,j} \in \Omega \times \mathbb{R}^m$, $X^{k+1} = \Pi_k(Y^{k,j} + D^{k,j})$ and Π_k leaves invariant the primal part, we obtain that $X^{k+1} = (x^{k+1}, \lambda^{k+1}) \in \Omega \times [-\alpha_L\sqrt{\rho_k}e, \alpha_U\sqrt{\rho_k}e]$. □

We should stress that no constraint qualification assumptions were needed to guarantee neither the feasibility of the subproblem (19) nor the success of execution of Step 2.1 of Algorithm 2. In [18, Lemma 2] the Mangasarian–Fromovitz constraint qualification was required.

4 Convergence analysis

In this section we will prove that any accumulation point of the sequence generated by Algorithm 2 is either a stationary point of problem (1), or a stationary (maybe feasible) point of the squared norm of infeasibility. We will show that no constraint qualification is needed in order to prove global convergence results.

The proposed method is related with an inexact augmented Lagrangian method. The augmented Lagrangian method, also known as the method of multipliers, is based on the minimization of the augmented Lagrangian function [21, 31], $\bar{L}(x, \lambda, \rho) : \mathbb{R}^n \times \mathbb{R}^m \times (0, +\infty) \to \mathbb{R}$, defined by

$$\bar{L}(x, \lambda, \rho) = f(x) + \frac{1}{2\rho}\|\lambda + \rho h(x)\|^2.$$

Recall that, at a given multiplier estimate $\lambda^k \in \mathbb{R}^m$ and a penalty parameter $\rho_k > 0$, the (exact) augmented Lagrangian method generates the next iterate (x^{k+1}, λ^{k+1}) such that

$$x^{k+1} \text{ is a solution of } \underset{x\in\Omega}{\text{minimize}}\, \bar{L}(x, \lambda^k, \rho_k), \tag{30}$$

$$\lambda^{k+1} = \lambda^k + \rho_k h(x^{k+1}). \tag{31}$$

The augmented Lagrangian method had been studied by many authors [2–4, 7–10, 12, 13, 24, 29, 30], among other literature (see also [5, 28]).

Numerical implementations attempt to solve (30) inexactly, by using a suitable criterion. For example, some codes based on the augmented Lagrangian method, such as LANCELOT [11] and ALGENCAN [1], define $x^{k+1} = x$ if the residual of the minimization problem in (30) at x is less than some tolerance ϵ_k, i.e.,

$$\left\| \Pi_\Omega\left(x - \frac{\partial \bar{L}}{\partial x}\big(x, \lambda^k, \rho_k\big)\right) - x \right\| \le \epsilon_k.$$

We can see that problem (30)–(31) is equivalent to problem (8). Such equivalence comes from (8) by solving $H_k(x, \lambda) = 0$ for λ and replacing it in the objective function F_k. Due to this connection, the global convergence theory of Algorithm 2 is an adaptation of the standard augmented Lagrangian theory. The connection between the sequence generated by Algorithm 2 and the sequence generated by the sSQP method is given by the next statement.

Proposition 1 *Algorithm 2 generates sequences* $\{x^k\}, \{y^k\}, \{\lambda^k\}, \{\nu^k\}, \{\rho_k\}$ *and* $\{M_k\}$ *satisfying*

$$\big\langle \nabla f\big(y^k\big) + M_k\big(x^{k+1} - y^k\big) + \nabla h\big(y^k\big)\nu^{k+1}, y - x^{k+1}\big\rangle \ge 0, \quad \forall\, y \in \Omega, \tag{32}$$

$$h\big(y^k\big) + \nabla h\big(y^k\big)^\top\big(x^{k+1} - y^k\big) - \frac{1}{\rho_k}\big(\nu^{k+1} - \lambda^k\big) = 0, \tag{33}$$

$$\big\|x^{k+1} - y^k\big\|^2 + \big\|\nu^{k+1} - \big(\lambda^k + \rho_k h\big(y^k\big)\big)\big\|^2 < \epsilon_k^2. \tag{34}$$

Proof Note that the optimality conditions of problem (19) are

$$\left\langle \nabla F_k(Y^{k,j}) + \begin{bmatrix} Q_{k,j} & 0 \\ 0 & \frac{1}{\rho_k} I \end{bmatrix} D^{k,j} + \nabla H_k\big(Y^{k,j}\big)\xi^{k,j}, Y - Y^{k,j} - D^{k,j} \right\rangle \ge 0,$$

$$\nabla H_k\big(Y^{k,j}\big)^\top D^{k,j} = 0,$$

for all $Y \in \Omega \times \mathbb{R}^m$, where $Y^{k,j} + D^{k,j} \in \Omega \times \mathbb{R}^m$ and $\xi^{k,j} \in \mathbb{R}^m$ is an associated Lagrange multiplier.

Let $j(k)$ be the index where $\|D^{k,j(k)}\| < \epsilon_k$. Let us call $y^k = x^{k,j(k)}$, the primal component of $Y^{k,j(k)}$, ν^{k+1} the dual component of $Y^{k,j(k)} + D^{k,j(k)}$, $\xi^k = \xi^{k,j(k)}$ and $M_k = Q_{k,j(k)}$.

Since $X^{k+1} = (x^{k+1}, \lambda^{k+1}) = \Pi_k(Y^{k,j(k)} + D^{k,j(k)})$ and the projection Π_k (17) leaves invariant the primal part, we have that $Y^{k,j(k)} + D^{k,j(k)} = (x^{k+1}, \nu^{k+1})$. Hence, the optimality conditions can be rewritten in the following form

$$\big\langle \nabla f\big(y^k\big) + M_k\big(x^{k+1} - y^k\big) + \nabla h\big(y^k\big)\xi^k, y - x^{k+1}\big\rangle \ge 0, \quad \forall\, y \in \Omega,$$

$$\frac{1}{\rho_k}\big(\lambda^k + \rho_k h\big(y^k\big)\big) + \frac{1}{\rho_k}\big(\nu^{k+1} - \big(\lambda^k + \rho_k h\big(y^k\big)\big)\big) - \frac{1}{\rho_k}\xi^k = 0,$$

$$\nabla h\big(y^k\big)^\top\big(x^{k+1} - y^k\big) - \frac{1}{\rho_k}\big(\nu^{k+1} - \big(\lambda^k + \rho_k h\big(y^k\big)\big)\big) = 0.$$

Notice that from the second relation we obtain $\nu^{k+1} = \xi^k$. Therefore, $D^{k,j(k)}$ is a solution of (19) if and only if

$$\langle \nabla f(y^k) + M_k(x^{k+1} - y^k) + \nabla h(y^k)\nu^{k+1}, y - x^{k+1} \rangle \geq 0, \quad \forall y \in \Omega, \tag{35}$$

$$h(y^k) + \nabla h(y^k)^\top (x^{k+1} - y^k) - \frac{1}{\rho_k}(\nu^{k+1} - \lambda^k) = 0. \tag{36}$$

With this notation, $\|D^{k,j(k)}\| < \epsilon_k$ is equivalent to

$$\|x^{k+1} - y^k\|^2 + \|\nu^{k+1} - (\lambda^k + \rho_k h(y^k))\|^2 < \epsilon_k^2. \tag{37}$$

□

Remark 2 According to the augmented Lagrangian method, the Lagrange multipliers are updated as in (31). In practical implementations, a projection onto a fixed box B is usually performed, that is,

$$\lambda^{k+1} = \Pi_B(\lambda^k + \rho_k h(x^{k+1})).$$

However, in the proposed method the Lagrange multipliers are updated by projecting ν^{k+1} given in (33) onto the variable box $\Lambda_k = [-\alpha_L\sqrt{\rho_k}e, \alpha_U\sqrt{\rho_k}e]$, that is,

$$\lambda^{k+1} = \Pi_{\Lambda_k}(\lambda^k + \rho_k h(y^k) + \rho_k \nabla h(y^k)^\top (x^{k+1} - y^k)).$$

Therefore, even when these two methods are connected, the Lagrange multipliers are updated in a different way.

The next auxiliary proposition gives a relation between the Lagrange multiplier approximation and the penalty parameter.

Proposition 2 *The sequence $\{\lambda^k/\rho_k\}$ is convergent to zero if ρ_k tends to infinity.*

Proof From the definition of Π_k (17), we have that λ^{k+1} belongs to the closed set $[-\alpha_L\sqrt{\rho_k}e, \alpha_U\sqrt{\rho_k}e]$. If $\rho_{k+1} > \rho_k$, from the update of the penalty parameter, we get

$$\frac{\lambda^{k+1}}{\rho_{k+1}} \in \left[-\frac{\alpha_L}{\sqrt{10\rho_{k+1}}}, \frac{\alpha_U}{\sqrt{10\rho_{k+1}}}\right].$$

On the other hand, if $\rho_{k+1} = \rho_k$, then

$$\frac{\lambda^{k+1}}{\rho_{k+1}} \in \left[-\frac{\alpha_L}{\sqrt{\rho_{k+1}}}, \frac{\alpha_U}{\sqrt{\rho_{k+1}}}\right].$$

In both cases, if ρ_{k+1} tends to infinity, the proposition holds. □

Proposition 2 helps us to prove the following global convergence theorem.

Theorem 2 *Let $\bar{x}$ be a limit point of the sequence $\{x^k\}$ generated by Algorithm 2 and assume that matrices $\{M_k\}$ are uniformly bounded.*

1. *If $\{\rho_k\}$ remains bounded, then $\bar{x}$ is a stationary point of problem (1).*
2. *If $\{\rho_k\}$ tends to infinity, then $\bar{x}$ is a stationary point of the problem*

$$\underset{x\in\Omega}{\text{minimize}}\ \frac{1}{2}\left\|h(x)\right\|^2. \tag{38}$$

Proof Let $\bar{x}$ be a limit point of $\{x^{k+1}\}$, i.e., there exists an index subset $\mathcal{K}$ such that

$$\lim_{k\in\mathcal{K}} x^{k+1} = \bar{x}. \tag{39}$$

Since ϵ_k tends to zero, and using (39) and (34) we get

$$\lim_{k\in\mathcal{K}} y^k = \bar{x}. \tag{40}$$

Proof of 1. Let us consider first the case when $\{\rho_k\}$ remains bounded. By the updating formula, we have that there exists $k_0 \in \mathbb{N}$ such that $\rho_k = \bar{\rho}$ for all $k \geq k_0$. Then, λ^{k+1} belongs to the closed set $[-\alpha_L\sqrt{\bar{\rho}}e, \alpha_U\sqrt{\bar{\rho}}e]$ for all $k \geq k_0$, that is, $\{\lambda^{k+1}\}$ is bounded.

From Step 3 of Algorithm 2 we have that $\sigma(X^{k+1}) \leq \gamma\psi_k$ or $\|h(x^{k+1})\| \leq \gamma\|h(x^k)\|$ for all $k \geq k_0$. Let $\mathcal{K}_1$ be the index set defined by

$$\mathcal{K}_1 = \left\{k \in \mathcal{K} \mid \sigma\left(X^{k+1}\right) \leq \gamma\psi_k\right\}.$$

In what follows we will consider two subcases: when $\mathcal{K}_1$ is finite or is infinite.

(a) Suppose that $\mathcal{K}_1$ has infinite many elements. Since the sequence $\{\psi_k\}$ is nonincreasing and nonnegative, it converges to some $\bar{\psi} \geq 0$. From the definition of $\mathcal{K}_1$ and observing that $\psi_{k+1} \leq \sigma(X^{k+1})$, we deduce that $\bar{\psi} \leq \gamma\bar{\psi}$ by taking limits for $k \in \mathcal{K}_1$. Hence, $\bar{\psi} = 0$ because $\gamma \in (0, 1)$. Since $\{\lambda^{k+1}\}$ is bounded, taking subsequences if necessary, we can guarantee the existence of $\bar{\lambda}$ such that $\lim_{k\in\mathcal{K}_1} \lambda^{k+1} = \bar{\lambda}$. Thus, taking limits for $k \in \mathcal{K}_1$ we have that $\sigma(\bar{x}, \bar{\lambda}) \leq \gamma\bar{\psi} = 0$. Hence, we conclude that $\bar{x}$ is a stationary point of problem (1).

(b) Suppose that $\mathcal{K}_1$ has finite many elements. Then there exists $k_1 \geq k_0$ such that $\|h(x^{k+1})\| \leq \gamma\|h(x^k)\|$ for all $k \geq k_1$. Taking limits for $k \in \mathcal{K}$ and using (39) we have that $h(\bar{x}) = 0$. Passing onto a subsequence if necessary, assume that $\lim_{k\in\mathcal{K}} \lambda^k = \bar{\lambda}$ (because of the boundedness of $\{\lambda^k\}$). Taking limits in (34) for $k \in \mathcal{K}$, using (40) and the facts that $h(\bar{x}) = 0$ and $\rho_k = \bar{\rho}$ for k large enough, we deduce that

$$\lim_{k\in\mathcal{K}} \nu^{k+1} = \bar{\lambda}. \tag{41}$$

From (39), (40), (41) and the fact that $\{M_k\}$ are uniformly bounded, taking limits in (32) for $k \in \mathcal{K}$, we conclude that

$$\left\langle \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\lambda}, y - \bar{x}\right\rangle \geq 0, \quad \forall\, y \in \Omega.$$

This condition is equivalent to $\Pi_\Omega(\bar{x} - \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda})) - \bar{x} = 0$. Since $h(\bar{x}) = 0$ we deduce that $\sigma(\bar{x}, \bar{\lambda}) = 0$. Hence, $\bar{x}$ is a stationary point of problem (1).

Proof of 2. Let us consider the case when $\{\rho_k\}$ tends to infinity. Taking limits in (33) we obtain

$$\lim_{k \in \mathcal{K}} \frac{\nu^{k+1}}{\rho_k} = h(\bar{x}). \tag{42}$$

where we have used (39), (40) and the fact that $\{\lambda^k/\rho_k\}$ converges to zero by Proposition 2.

Dividing (32) by ρ_k, and using (39), (40), (42) and the fact that $\{M_k\}$ are uniformly bounded, taking limits for $k \in \mathcal{K}$, we obtain that

$$\left\langle \nabla h(\bar{x}) h(\bar{x}), y - \bar{x} \right\rangle \geq 0, \quad \forall\, y \in \Omega.$$

Hence, we conclude that $\bar{x}$ is a stationary point of problem (38). □

Remark 3 For item 2 of the previous Theorem, we can not assure that the limit point $\bar{x}$ is a stationary point of problem (1) without assuming any constraint qualifications at all. However, following the lines in [2, Theorem 4.2], it can be seen that $\bar{x}$ is a stationary point of problem (1) if we suppose that $\bar{x}$ is feasible and satisfies the CPLD constraint qualification.

5 Penalty boundedness results

From now on we will prove that, under suitable conditions, the sequence of penalty parameters $\{\rho_k\}$ generated by Algorithm 2 remains bounded. Let us consider the following assumptions:

Assumption A1 $\{x^k\}$ converges to a feasible point $\bar{x}$.

Assumption A2 There is no vector $\lambda \neq 0$ such that $-\nabla h(\bar{x})\lambda \in \mathcal{N}_\Omega(\bar{x})$ and there is only one vector $\bar{\lambda}$ of associated multipliers (this condition is equivalent to the Strict Mangasarian–Fromovitz constraint qualification).

Assumption A3 There exists $k_0 \in \mathbb{N}$ such that $\bar{\lambda} \in (-\alpha_L \sqrt{\rho_{k_0}} e, \alpha_U \sqrt{\rho_{k_0}} e)$.

Assumption A4 The second order sufficient optimality conditions are satisfied at $(\bar{x}, \bar{\lambda})$, where $\bar{\lambda}$ is a Lagrange multiplier associated to $\bar{x}$. That is,

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}) d, d \right\rangle > 0 \quad \forall d \in \mathcal{C} \setminus \{0\}, \tag{43}$$

where

$$\mathcal{C} = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \langle \nabla f(\bar{x}), d \rangle = 0,\ \nabla h(\bar{x})^\top d = 0, \\ d_i \leq 0 \text{ if } \bar{x}_i = b_i, d_i \geq 0 \text{ if } \bar{x}_i = a_i, i = 1, \ldots, n \end{array} \right\}. \tag{44}$$

Assumption A5 The sequence $\{\epsilon_k\}$ is chosen according to

$$\epsilon_k \le \chi\left(\sigma\left(x^k, \lambda^k\right)\right), \tag{45}$$

where $\chi : (0, +\infty) \to (0, +\infty)$ is such that $\lim_{t\to 0} \chi(t)/t = 0$.

We will prove a lemma that establishes the convergence of the dual sequence $\{\lambda^k\}$.

Lemma 4 *Let Assumptions A1, A2 and A3 hold. Then* $\lim_{k\to\infty} \lambda^k = \bar{\lambda}$.

Proof By A1, x^k converges to $\bar{x}$. By (34), y^k converges to $\bar{x}$.

Suppose that the sequence $\{\nu^{k+1}\}$ is unbounded. Taking subsequences if necessary, assume that $\nu^{k+1}/\|\nu^{k+1}\|$ converges to a unitary vector $\bar{\nu}$. Dividing (32) by $\|\nu^{k+1}\|$ and taking limits, we obtain that

$$\left\langle \nabla h(\bar{x})\bar{\nu}, y - \bar{x} \right\rangle \ge 0, \quad \forall\, y \in \Omega,$$

that is equivalent to $-\nabla h(\bar{x})\bar{\nu} \in \mathcal{N}_\Omega(\bar{x})$. From A2 we have that $\bar{\nu} = 0$ and this leads us to a contradiction.

Since $\{\nu^{k+1}\}$ is bounded, there exists at least a limit point $\bar{\nu}$. Passing onto subsequences if necessary and taking limits in (32) we get

$$\left\langle \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\nu}, y - \bar{x} \right\rangle \ge 0, \quad \forall\, y \in \Omega.$$

This means that $\bar{\nu}$ is a Lagrange multiplier associated to $\bar{x}$. From A2 we conclude that $\bar{\nu} = \bar{\lambda}$, and therefore $\{\nu^{k+1}\}$ converges to $\bar{\lambda}$.

By A3, $\nu^{k+1} \in [-\alpha_L\sqrt{\rho_k}e, \alpha_U\sqrt{\rho_k}e]$ for k large enough. Thus $\lambda^{k+1} = \nu^{k+1}$ since no projection is needed. □

The next lemma gives a relation between the natural residual (2) and the distance to the solution.

Lemma 5 *If A1–A4 hold, then there exist* $k_0 \in \mathbb{N}$, $\beta_1, \beta_2 > 0$ *such that for all* $k \ge k_0$,

$$\beta_1\left\|\left(x^k, \lambda^k\right) - (\bar{x}, \bar{\lambda})\right\| \le \sigma\left(x^k, \lambda^k\right) \le \beta_2\left\|\left(x^k, \lambda^k\right) - (\bar{x}, \bar{\lambda})\right\|. \tag{46}$$

Proof By Lipschitz continuity of σ and the fact that $\sigma(\bar{x}, \bar{\lambda}) = 0$ we guarantee the existence of β_2 satisfying the right–hand side inequality. By Assumption A4 and [17, Lemma 5, Theorem 2] there exists $\beta_1 > 0$ such that $\sigma(x, \lambda) \ge \beta_1\|(x, \lambda) - (\bar{x}, \bar{\lambda})\|$ for all (x, λ) close enough to $(\bar{x}, \bar{\lambda})$. From A1 and Lemma 4 we have that (x^k, λ^k) converges to $(\bar{x}, \bar{\lambda})$ and this concludes the proof. □

The following lemma is a technical result that will be used in the next theorem.

Lemma 6 *Let us assume that A1–A5 hold. Then there exist* $k_1 \in \mathbb{N}$, $c_1, c_2 > 0$ *such that*

$$\left(1 - \frac{c_2}{\rho_k}\right)\sigma\left(x^{k+1}, \lambda^{k+1}\right) \le \left(c_1\eta_k + \frac{c_2}{\rho_k}\right)\sigma\left(x^k, \lambda^k\right), \tag{47}$$

where

$$\eta_k = \frac{\chi(\sigma(x^k, \lambda^k))}{\sigma(x^k, \lambda^k)}.$$

Proof By Taylor expansion centered at y^k we get

$$\frac{\partial L}{\partial x}\big(x^{k+1}, \lambda^{k+1}\big) = \frac{\partial L}{\partial x}\big(y^k, \lambda^{k+1}\big) + \frac{\partial^2 L}{\partial x^2}\big(y^k, \lambda^{k+1}\big)\big(x^{k+1} - y^k\big) + o\big(\|x^{k+1} - y^k\|\big),$$

and therefore

$$\begin{aligned}
&\frac{\partial L}{\partial x}\big(x^{k+1}, \lambda^{k+1}\big) - \left[\frac{\partial L}{\partial x}\big(y^k, \nu^{k+1}\big) + M_k\big(x^{k+1} - y^k\big)\right] \\
&\quad = \left(\frac{\partial^2 L}{\partial x^2}\big(y^k, \lambda^{k+1}\big) - M_k\right)\big(x^{k+1} - y^k\big) + o\big(\|x^{k+1} - y^k\|\big) \\
&\quad = O\big(\|x^{k+1} - y^k\|\big),
\end{aligned} \tag{48}$$

where we are using that $\lambda^{k+1} = \nu^{k+1}$ for k large enough (see Lemma 4), the continuity of the second derivative of L with respect to x and the fact that $\{M_k\}$ is uniformly bounded.

By definition of projection and (32) we have that

$$x^{k+1} = \Pi_\Omega\left(x^{k+1} - \left[\frac{\partial L}{\partial x}\big(y^k, \nu^{k+1}\big) + M_k\big(x^{k+1} - y^k\big)\right]\right). \tag{49}$$

Since Π_Ω is nonexpansive, using (48) and (49) we obtain

$$\left\|\Pi_\Omega\left(x^{k+1} - \frac{\partial L}{\partial x}\big(x^{k+1}, \lambda^{k+1}\big)\right) - x^{k+1}\right\| \le O\big(\|x^{k+1} - y^k\|\big). \tag{50}$$

On the other hand, by using (33) and the fact that $\lambda^{k+1} = \nu^{k+1}$ for k large enough, we get

$$\begin{aligned}
h\big(x^{k+1}\big) &= h\big(y^k\big) + \nabla h\big(y^k\big)^\top\big(x^{k+1} - y^k\big) + o\big(\|x^{k+1} - y^k\|\big) \\
&= \frac{1}{\rho_k}\big(\lambda^{k+1} - \lambda^k\big) + o\big(\|x^{k+1} - y^k\|\big).
\end{aligned}$$

Then, by the previous two equations, there exist $k_1 \in \mathbb{N}$, $c_1, c_2 > 0$ such that for all $k \ge k_1$,

$$\begin{aligned}
\sigma\big(x^{k+1}, \lambda^{k+1}\big) &\le O\big(\|x^{k+1} - y^k\|\big) + \sqrt{2}\|h\big(x^{k+1}\big)\| \\
&\le O\big(\|x^{k+1} - y^k\|\big) + \frac{\sqrt{2}}{\rho_k}\|\lambda^{k+1} - \lambda^k\|
\end{aligned}$$

$$\leq c_1\epsilon_k + \frac{\sqrt{2}}{\rho_k}\|\lambda^{k+1} - \bar{\lambda}\| + \frac{\sqrt{2}}{\rho_k}\|\lambda^k - \bar{\lambda}\|$$
$$\leq c_1\epsilon_k + \frac{c_2}{\rho_k}\sigma\left(x^{k+1}, \lambda^{k+1}\right) + \frac{c_2}{\rho_k}\sigma\left(x^k, \lambda^k\right),$$

where in the third inequality we use (34) and for the last inequality we use (46).

Thus, by using A5, we conclude that

$$\left(1 - \frac{c_2}{\rho_k}\right)\sigma\left(x^{k+1}, \lambda^{k+1}\right) \leq \left(c_1\eta_k + \frac{c_2}{\rho_k}\right)\sigma\left(x^k, \lambda^k\right). \qquad \square$$

Now, under the set of assumptions of this section, we prove the following result about the boundedness of the penalty parameter.

Theorem 3 *Suppose that Assumptions A1–A5 hold. Then, the sequence of penalty parameters $\{\rho_k\}$ is bounded.*

Proof By contradiction, suppose that $\lim_{k\to\infty}\rho_k = \infty$. Since $\lim_{k\to\infty}\eta_k = 0$, then for k sufficiently large we have

$$1 - \frac{c_2}{\rho_k} > \frac{1}{2} \quad \text{and} \quad c_1\eta_k + \frac{c_2}{\rho_k} < \frac{\gamma}{2},$$

where γ is a parameter defined in Algorithm 2. Hence, by (47),

$$\sigma\left(x^{k+1}, \lambda^{k+1}\right) \leq \gamma\sigma\left(x^k, \lambda^k\right),$$

for k large enough.

Since $\gamma < 1$, $\{\sigma(X^k)\}$ is a strictly decreasing sequence, which implies that $\psi_k = \min\{\psi_{k-1}, \sigma(X^k)\} = \sigma(X^k)$ for k sufficiently large. Thus, $\sigma(X^{k+1}) \leq \gamma\psi_k$. Therefore, by Step 3 of the Algorithm 2 we conclude that $\rho_{k+1} = \rho_k$ for k large enough, in contradiction with the initial assumption. $\square$

Theorem 4 *Let us assume that A1–A5 hold. Then, given $q \in (0, 1)$ there exists $\bar{\rho}$ such that if $\rho_{\bar{k}} \geq \bar{\rho}$ for some $\bar{k}$, it holds that the sequence $\{(x^k, \lambda^k)\}$ converges Q-linearly to $(\bar{x}, \bar{\lambda})$ with rate equal to q.*

Proof Let us define $\bar{\rho} \geq (q\beta_1 + \beta_2)c_2/(q\beta_1)$, where β_1 and β_2 are the constants defined as in (46), and c_2 is given by Lemma 6. Due to the fact that $\{\rho_k\}$ is nondecreasing, for all $k \geq \bar{k}$ we have that

$$\frac{c_2}{\rho_k} \leq \frac{q\beta_1}{q\beta_1 + \beta_2} \quad \text{and} \quad \left(1 - \frac{c_2}{\rho_k}\right)^{-1} \leq \frac{q\beta_1 + \beta_2}{\beta_2}. \tag{51}$$

Hence,

$$\begin{aligned}
\left\|\left(x^{k+1},\lambda^{k+1}\right)-(\bar{x},\bar{\lambda})\right\| &\le \frac{1}{\beta_1}\sigma\left(x^{k+1},\lambda^{k+1}\right)\\
&\le \frac{1}{\beta_1}\left(1-\frac{c_2}{\rho_k}\right)^{-1}\left(c_1\eta_k+\frac{c_2}{\rho_k}\right)\sigma\left(x^k,\lambda^k\right)\\
&\le \left(\frac{q\beta_1+\beta_2}{\beta_1\beta_2}c_1\eta_k+\frac{q}{\beta_2}\right)\sigma\left(x^k,\lambda^k\right)\\
&\le \left(\frac{q\beta_1+\beta_2}{\beta_1}c_1\eta_k+q\right)\left\|\left(x^k,\lambda^k\right)-(\bar{x},\bar{\lambda})\right\|,
\end{aligned}$$

where for the first inequality we use the left-hand side of (46), the second inequality comes from (47), for the third inequality we use (51) and the last inequality follows from the right-hand side relation in (46).

Since $\{(x^k,\lambda^k)\}$ converges to $(\bar{x},\bar{\lambda})$ and $\lim_{k\to\infty}\eta_k=0$, we conclude that the primal-dual sequence converges with Q-linear rate equal to q. □

6 Numerical experiments

In this section we show preliminary numerical results obtained by using Algorithm 2. We have considered the whole set of nonlinear equality constrained problems from the CUTEr collection [14]. A total of 71 nonlinear equality constrained problems were identified. Algorithm 2 was written in Fortran 2003 and compiled with the Intel Compiler 12.0 on a PC running Linux.

The following choices were made and used on all test problems:

- Algorithmic parameters: $\gamma=0.99$, $\varepsilon=0.5\times10^{-6}$, $\epsilon_k=1/(k+1)^2$ for all $k\ge0$, $r=0.99$ and $\alpha_L=\alpha_U=100$.
- Initialization parameters: $\rho_0=0.01$, $\theta_0=0.9$, $Q_{k,0}$ is the identity matrix for all $k\ge0$.
- Starting points: λ^0 the origin, and x^0 is taken from the corresponding problem from the CUTEr collection.
- For solving the quadratic programming problem (19) we used an implementation of the Goldfarb-Idnani algorithm [19] written by B. Turlach [32].

We remark that we only want to show viability of the proposed approach. An optimal choice of the parameters and an appropriate quadratic programming solver is out of the scope of this paper. Since the Goldfarb-Idnani algorithm performs factorizations, it is not convenient to deal with large-scale problems, thus we exclude the following problems: LUKVLE1-LUKVLE18, EIGENA2, EIGENB2, EIGENC2, ELEC, GRIDNETE, GRIDNETH, ORTHRDS2, ORTHREGA, ORTHREGC, ORTHREGD, ORTHRGDM, ORTHRGDS. The remaining 41 problems were tested with the proposed method. In the following problems: BT7, BT8, COOLHANS, HS26, HS40, HS46, HS47, HS56, HS78, MSS1, MWRIGHT the quadratic solver failed.

Table 1 Numerical experiments from the CUTEr collection

Name	n	m	$\bar{\rho}$	$\sigma(\bar{x}, \bar{\lambda})$	qp calls
BT1	2	1	1.0E+02	0.3492216E–07	5, 2, 1
BT2	3	1	1.0E+09	0.1455329E–07	1, 1, 5
BT3	5	3	*	*	*
BT4	3	2	1.0E+07	0.4203023E–11	1, 1, 3
BT5	2	2	1.0E–01	0.4927916E–06	1, 1, 1
BT6	5	2	1.0E+06	0.1040323E–06	1, 1, 4
BT9	4	2	1.0E+00	0.4629197E–06	1, 1, 1
BT10	2	2	1.0E+00	0.4827694E–06	1, 1, 1
BT11	5	3	1.0E+03	0.2520146E–06	2, 1, 3
BT12	5	3	1.0E+00	0.4814203E–06	1, 1, 1
BYRDSPHR	3	2	1.0E–02	0.4981593E–06	1, 1, 1
DIXCHLNG	10	5	1.0E+04	0.1259454E–06	12, 2, 11
HS6	2	1	1.0E+01	0.4110476E–06	1, 1, 1
HS7	2	1	1.0E+02	0.1332312E–07	1, 3, 1
HS8	2	2	1.0E+00	0.4417480E–06	1, 1, 1
HS9	2	1	1.0E+00	0.4878523E–06	1, 1, 1
HS27	3	1	1.0E+06	0.1687748E–09	1, 1, 6
HS28	3	1	1.0E+04	0.3856428E–06	1, 2, 1
HS39	4	2	1.0E+02	0.4629197E–06	1, 1, 1
HS42	4	2	1.0E+03	0.7307294E–08	1, 1, 2
HS48	5	2	1.0E+12	0.2884765E–13	1, 2, 5
HS49	5	2	*	*	*
HS50	5	3	1.0E+06	0.7299110E–07	1, 1, 2
HS51	5	3	1.0E+13	0.1055938E–13	1, 3, 4
HS52	5	3	1.0E+02	0.2636263E–07	3, 2, 4
HS61	3	2	1.0E+02	0.3309869E–07	1, 2, 1
HS77	5	2	1.0E+06	0.1130220E–07	1, 1, 3
HS79	5	3	1.0E+05	0.2686601E–08	1, 1, 2
MARATOS	2	1	1.0E–01	0.3845145E–06	1, 1, 1
ORTHREGB	27	6	1.0E–01	0.3513701E–06	1, 1, 1
S316-322	2	1	1.0E+03	0.3563555E–07	2, 1, 4

In Table 1 we report, for each problem, the problem name in the CUTEr collection, the number of variables n, the number of constraints m, the last penalty parameter $\bar{\rho}$, the natural residual $\sigma(\bar{x}, \bar{\lambda})$ and number of calls to the quadratic solver in the last three iterations. We remark that in problems BT3 and HS49 the algorithm stopped because the inexact restoration penalty parameter θ_j was less than a prescribed tolerance.

In most of the problems we perform a few calls to the quadratic solver in the last iterations. Notice that if we perform one call to the quadratic solver, it means that $\|D^{k,0}\| < \epsilon_k$ (see Algorithm 2). According to Remark 1 the last iterations of Algorithm 2 are the same as the iterations of the sSQP method.

In Table 2 we report, for the same set of problems, a comparison by using Algorithm 2, ALGENCAN [1] and LANCELOT [23] with the default algorithmic parameters. First of all we indicate the problem name, the number of variables n, the number of constraints m, the number of outer iterations, the number of inner iterations and the functional value $f(\bar{x})$. For LANCELOT, we consider as outer iterations those reported as "iterations and function evaluations" (see [11, pp. 148]) without pointing out inner iterations. The first row of each problem corresponds to Algorithm 2, the second one to ALGENCAN and the last one to LANCELOT.

From Table 2 we can see that the functional values $f(\bar{x})$ given by the three solvers are the same in almost all problems. However, in problem DIXCHLNG, ALGENCAN found a different local solution with functional value greater than the obtained by our implementation and LANCELOT. As before, our implementation failed to solve problems BT3 and HS49 at the backtracking Step 2.3. In all cases the CPU time was negligible for the three solvers, therefore no comparison can be made.

Although the number of outer and inner iterations in our implementation are greater than those given by ALGENCAN and LANCELOT, we must stress that Algorithm 2 perform a standard backtracking procedure at Steps 2.3 and 2.4, while an ad hoc strategy is used by the other solvers. Also, concerning the update of the penalty parameter (Step 3), we generate a monotone increasing sequence ρ_k while the other two solvers allow a possible decrease of the penalty parameter.

7 Conclusions

In this paper we present a new hybrid method for solving optimization problems with equality constraints and bounds. The proposed method is based on the sSQP method and a combination between an IR method and an augmented Lagrangian-like penalty parameter update strategy. Global convergence is obtained by using a suitable penalty parameter update joint with an IR method to solve inexactly the subproblems without modifying the structure of the sSQP subproblems. So, the good local behavior of the sSQP method is inherited.

Since our method does not change the sSQP subproblems it preserves some known properties as solvability and well-conditioned subproblems without any constraint qualification assumptions. These features make this formulation very attractive. Besides that, this method presents an interesting connection between augmented Lagrangian methods and inexact restoration methods.

It has been proved that the algorithm is well defined and that any limit point of the sequence generated by the algorithm converges to a KKT point or to a stationary (maybe feasible) point of the problem that minimizes the infeasibility, depending on the boundedness of the sequence of the penalty parameters.

Moreover, if the sequence generated by the algorithm converges to a feasible point and the strict Mangasarian-Fromovitz constraint qualification and the second order sufficient optimality conditions hold, then the penalty parameter remains bounded and the primal-dual sequence converges Q-linearly.

Regarding numerical experiments, the algorithm was implemented in Fortran 2003 and tested on a set of problems from the CUTEr collection validating the theoretical results.

Table 2 Numerical experiments from the CUTEr collection

Name	n	m	Outer iter.	Inner iter.	$f(\bar{x})$
BT1	2	1	11	24	−1.0000E+00
			5	28	−1.0000E+00
			53		−1.0000E+00
BT2	3	1	35	79	3.2568E−02
			13	30	3.2568E−02
			27		3.2568E−02
BT3	5	3	*	*	*
			7	9	4.0930E+00
			9		4.0930E+00
BT4	3	2	33	1062	−3.7048E+00
			6	14	−4.5511E+01
			23		−4.5511E+01
BT5	2	2	24	32	9.6172E+02
			4	14	9.6172E+02
			18		9.6172E+02
BT6	5	2	26	42	2.7704E−01
			9	21	2.7704E−01
			28		2.7705E−01
BT9	4	2	353	1763	−1.0000E+00
			17	47	−1.0000E+00
			20		−1.0000E+00
BT10	2	2	99	153	−1.0000E+00
			17	41	−1.0000E+00
			17		−1.0000E+00
BT11	5	3	700	5309	8.2489E−01
			13	34	8.2489E−01
			19		8.2489E−01
BT12	5	3	42	75	6.1881E+00
			4	7	6.1881E+00
			18		6.1881E+00
BYRDSPHR	3	2	370	489	−4.6833E+00
			15	31	−4.6833E+00
			45		−4.6833E+00
DIXCHLNG	10	5	310	4802	2.9210E−12
			5	26	2.4719E+03
			36		7.1292E−10
HS6	2	1	26	32	5.1491E−14
			3	10	1.7280E−23
			56		2.7850E−14

Table 2 (*Continued*)

Name	n	m	Outer iter.	Inner iter.	$f(\bar{x})$
HS7	2	1	41	75	−1.7321E+00
			14	27	−1.7321E+00
			18		−1.7321E+00
HS8	2	2	16	20	−1.0000E+00
			4	10	−1.0000E+00
			11		−1.0000E+00
HS9	2	1	150	154	−5.0000E−01
			4	5	−5.0000E−01
			4		−5.0000E−01
HS27	3	1	27	116	4.0000E−02
			7	17	4.0000E−02
			16		4.0000E−02
HS28	3	1	24	42	3.5435E−14
			1	1	6.4812E−27
			3		4.8934E−30
HS39	4	2	353	1763	−1.0000E+00
			17	47	−1.0000E+00
			20		−1.0000E+00
HS42	4	2	115	520	1.3858E+01
			5	12	1.3858E+01
			4		1.3858E+01
HS48	5	2	50	80	2.9336E−30
			1	2	1.2586E−17
			3		6.9025E−31
HS49	5	2	*	*	*
			5	16	1.3806E−09
			15		2.2188E−08
HS50	5	3	43	78	6.8830E−16
			3	9	3.5338E−13
			12		5.3045E−13
HS51	5	3	47	63	5.0536E−31
			1	2	5.5467E−31
			2		2.5884E−31
HS52	5	3	750	4888	5.3266E+00
			6	8	5.3266E+00
			6		5.3266E+00
HS61	3	2	32	68	−1.4365E+02
			5	14	−1.4365E+02
			19		−1.4365E+02

Table 2 (*Continued*)

Name	n	m	Outer iter.	Inner iter.	$f(\bar{x})$
HS77	5	2	47	90	2.4151E−01
			8	22	2.4151E−01
			24		2.4151E−01
HS79	5	3	36	61	7.8777E−02
			13	25	7.8777E−02
			9		7.8777E−02
MARATOS	2	1	45	45	−1.0000E+00
			9	14	−1.0000E+00
			7		−1.0000E+00
ORTHREGB	27	6	11	12	3.0043E−14
			5	7	4.7394E−20
			98		1.8487E−12
S316-322	2	1	241	798	3.3431E+02
			6	11	3.3431E+02
			23		3.3431E+02

Acknowledgements The authors are thankful to the anonymous referees for the careful reading of the manuscript, helpful comments and suggestions.

References

1. ALGENCAN: http://www.ime.usp.br/~egbirgin/tango/
2. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. **18**(4), 1286–1309 (2007)
3. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. Math. Program., Ser. B **111**(1–2), 5–32 (2008)
4. Bertsekas, D.P.: Multiplier methods: a survey. Automatica J. IFAC **12**(2), 133–145 (1976)
5. Bertsekas, D.P.: Constrained optimization and lagrange multiplier methods. In: Computer Science and Applied Mathematics. Academic Press, New York (1982)
6. Birgin, E.G., Martínez, J.M.: Local convergence of an inexact-restoration method and numerical experiments. J. Optim. Theory Appl. **127**(2), 229–247 (2005)
7. Birgin, E.G., Castillo, R.A., Martínez, J.M.: Numerical comparison of augmented Lagrangian algorithms for nonconvex problems. Comput. Optim. Appl. **31**(1), 31–55 (2005)
8. Birgin, E.G., Floudas, C.A., Martínez, J.M.: Global minimization using an augmented Lagrangian method with variable lower-level constraints. Math. Program., Ser. A **125**(1), 139–162 (2010)
9. Birgin, E.G., Fernández, D., Martínez, J.M.: The boundedness of penalty parameters in an augmented Lagrangian method with constrained subproblems. Optim. Methods Softw. doi:10.1080/10556788.2011.556634
10. Conn, A.R., Gould, N.I.M., Toint, P.L.: A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. SIAM J. Numer. Anal. **28**(2), 545–572 (1991)
11. Conn, A.R., Gould, N.I.M., Toint, P.L.: LANCELOT: a Fortran Package for Large-scale Nonlinear Optimization (Release A). In: Springer Series in Computational Mathematics, vol. 17. Springer, Berlin (1992)

12. Conn, A.R., Gould, N., Sartenaer, A., Toint, P.L.: Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints. SIAM J. Optim. **6**(3), 674–703 (1996)
13. Contesse-Becker, L.: Extended convergence results for the method of multipliers for non-strictly binding inequality constraints. J. Optim. Theory Appl. **79**(2), 273–310 (1993)
14. CUTEr: http://www.cuter.rl.ac.uk/index.html
15. Fernández, D.: A quasi-newton strategy for the sSQP method for variational inequality and optimization problems. Math. Program., Ser. A. doi:10.1007/s10107-011-0493-8
16. Fernández, D., Solodov, M.: Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program., Ser. A **125**(1), 47–73 (2010)
17. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program., Ser. A **94**(1), 91–124 (2002)
18. Fischer, A., Friedlander, A.: A new line search inexact restoration approach for nonlinear programming. Comput. Optim. Appl. **46**(2), 333–346 (2010)
19. Goldfarb, D., Idnani, A.: A numerically stable dual method for solving strictly convex quadratic programs. Math. Program. **27**(1), 1–33 (1983)
20. Hager, W.W.: Stabilized sequential quadratic programming. Comput. Optim. Appl. **12**(1–3), 253–273 (1999)
21. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
22. Izmailov, A., Solodov, M.: Stabilized SQP revisited. Math. Program., Ser. A. doi:10.1007/s10107-010-0413-3
23. LANCELOT: http://www.cse.scitech.ac.uk/nag/lancelot/lancelot.shtml
24. Luo, H.Z., Sun, X.L., Li, D.: On the convergence of augmented Lagrangian methods for constrained global optimization. SIAM J. Optim. **18**(4), 1209–1230 (2007)
25. Martínez, J.M.: Inexact-restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. J. Optim. Theory Appl. **111**(1), 39–58 (2001)
26. Martínez, J.M., Pilotta, E.A.: Inexact-restoration algorithm for constrained optimization. J. Optim. Theory Appl. **104**(1), 135–163 (2000)
27. Martínez, J.M., Pilotta, E.A.: Inexact restoration methods for nonlinear programming: advances and perspectives. In: Optimization and Control with Applications. Appl. Optim., vol. 96, pp. 271–291. Springer, New York (2005)
28. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operations Research. Springer, New York (1999)
29. Pennanen, T.: Local convergence of the proximal point algorithm and multiplier methods without monotonicity. Math. Oper. Res. **27**(1), 170–191 (2002)
30. Polyak, V.T., Tret'yakov, N.V.: The method of penalty estimates for conditional extremum problems. USSR Comput. Math. Math. Phys. **13**(1), 42–58 (1974)
31. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Optimization, Sympos., Univ. Keele, Keele, 1968, pp. 283–298. Academic Press, London (1969)
32. QUADPROG: http://cran.r-project.org/web/packages/quadprog/
33. Wright, S.J.: Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl. **11**(3), 253–275 (1998)
34. Wright, S.J.: Modifying SQP for degenerate problems. SIAM J. Optim. **13**(2), 470–497 (2002) (electronic)
35. Wright, S.J.: Constraint identification and algorithm stabilization for degenerate nonlinear programs. Math. Program., Ser. B **95**(1), 137–160 (2003). ISMP 2000, Part 3 (Atlanta, GA)