UNIVERSIDAD NACIONAL DE CÓRDOBA

FACULTAD DE MATEMÁTICA, ASTRONOMÍA Y FÍSICA

SERIE "A"

TRABAJOS DE MATEMÁTICA

Nº 103/2011

An inexact restoration strategy for the globalization of the sSQP method

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

Editores: Jorge G. Adrover – Gastón García

CIUDAD UNIVERSITARIA – 5000 CÓRDOBA
REPÚBLICA ARGENTINA

Preprint

An inexact restoration strategy for the globalization of the sSQP method

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

Received: date / Accepted: date

Abstract An algorithm to solve equality constrained optimization problems based on stabilized sequential quadratic programming, augmented Lagrangian and inexact restoration methods is presented. This formulation has attractive features in the sense that no constraint qualifications are needed at the limit point, and that it overcomes ill-conditioning of the subproblems when the penalty parameter is large. Well-definition of the algorithm is shown, and also it is proved that any limit point of the sequence generated by the algorithm is a KKT point or a stationary point of the problem that minimizes the infeasibility. Under suitable hypotheses the sequence generated by the algorithm converges Q-linearly. Numerical experiments on a set of problems from the Cuter collection are given to confirm theoretical results.

Keywords Augmented Lagrangian · nonlinear programming · global convergence.

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 equality constrained nonlinear program,

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

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, Universidad Nacional de Córdoba, CIEM (CONICET), Medina Allende s/n, Ciudad Universitaria (5000) Córdoba, Argentina.
Tel.: +54-351-4334051
Fax: +54-351-4334054
E-mail: {dfernandez, pilotta, torres}@famaf.unc.edu.ar

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

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

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

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

where $\rho_k > 0$ is a 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 qualification is satisfied at this solution (see [21, 18, 17]). For equality constrained problems, local convergence has recently been studied in [23] for solutions with noncritical Lagrange multipliers (weaker than SOSC) and without any constraint qualification assumptions. A quasi-Newton strategy was studied in [16], 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. First, note that the primal-dual subproblem (3) can be obtained by performing an iteration of the SQP method in the primal-dual variable at the point (x^k, λ^k) to the problem

$$\begin{aligned} \underset{(x,\lambda)}{\text{minimize}}\ & F_k(x,\lambda) = f(x) + \tfrac{1}{2\rho_k}\|\lambda\|^2 \\ \text{subject to}\ & H_k(x,\lambda) = h(x) - \tfrac{1}{\rho_k}(\lambda - \lambda^k) = 0, \\ & x \in \Omega. \end{aligned} \tag{4}$$

Thus, local convergence of the sSQP method follows by solving problem (4) inexactly at each iteration, where the inexactness comes from the fact that just one SQP iteration is performed. When the current iterate (x^k, λ^k) is far from a primal-dual pair satisfying SOSC, it is not clear why a single SQP iteration is enough. Therefore, we propose to perform as many SQP iterations as necessary up to obtain a suitable inexact solution of (4) by using inexact restoration ideas.

Inexact restoration (IR) methods were introduced in [27] and modified in [26, 10, 19]. A survey on this subject can be found in [28]. The advantage of using this method to solve problem (4) is the fact that a feasible point is always known (avoiding the restoration phase) and that subproblem (3) provides a suitable tangent direction that satisfies sufficient conditions for convergence, according to [19].

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 of the augmented Lagrangian method, and overcome the ill-conditioned subproblems for large values of the penalty parameter [6]. 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 Section 2 the last results on IR methods are summarized. The proposed algorithm as well as its well-definition is described in Section 3. The main result, global convergence of the sequence generated by the algorithm, is presented in Section 4. Local convergence and penalty boundedness results are treated in Section 5. Section 6 is devoted to numerical experiments and conclusions are given in Section 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 $\mathcal{N}_\Omega(x) = \{v \in \mathbb{R}^m \mid \langle v, y - x \rangle \leq 0 \ \forall y \in \Omega\}$ if $x \in \Omega$, or $\mathcal{N}_\Omega(x) = \emptyset$ otherwise.

2 IR methods

The IR method as presented in [19] 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:

$$\begin{aligned} \text{minimize } & F(x) \\ \text{subject to } & H(x) = 0, \\ & x \in \mathcal{W}, \end{aligned} \tag{5}$$

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:

$$\|H(Y^j)\| \leq r\|H(X^j)\|, \tag{6}$$

$$F(Y^j) \leq F(X^j) + \beta\|H(X^j)\|. \tag{7}$$

Step 2*: Search direction*

Compute $D^j \in \mathbb{R}^n$ such that $Y^j + D^j \in \mathcal{W}$ and

$$F(Y^j + tD^j) \leq F(Y^j) - \eta t\|D^j\|^2, \tag{8}$$

$$\|H(Y^j + tD^j)\| \leq \|H(Y^j)\| + \bar{\eta}t^2\|D^j\|^2, \tag{9}$$

holds for all $t \in [0, \tau]$.

Step 3*: Penalty parameter*

Determine $\theta_{j+1} \in \left\{2^{-i}\theta_j : i \in \mathbb{N} \cup \{0\}\right\}$ as large as possible such that:

$$\phi(Y^j, \theta_{j+1}) - \phi(X^j, \theta_{j+1}) \leq \frac{(r-1)}{2}\left(\|H(X^j)\| - \|H(Y^j)\|\right), \tag{10}$$

where $\phi(X, \theta) = \theta F(X) + (1-\theta)\|H(X)\|$ is a merit function.

Step 4*: Line search*

Determine $t_j \in \left\{2^{-i} : i \in \mathbb{N} \cup \{0\}\right\}$ as large as possible such that:

$$\phi(Y^j + t_j D^j, \theta_{j+1}) - \phi(X^j, \theta_{j+1}) \leq \frac{(r-1)}{2}\left(\|H(X^j)\| - \|H(Y^j)\|\right). \tag{11}$$

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 [19] is that any sequence of search directions generated by Algorithm 1 tends to zero.

Theorem 1 *[19, Theorem 2] Suppose that $\mathcal{W}$ is a compact set and Step 1 of Algorithm 1 is well defined. Then,*

$$\lim_{j \to \infty} D^j = 0. \tag{12}$$

3 Description of the algorithm

We begin this section 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 α_L, $\alpha_U > 0$. For a current penalty parameter ρ_k we call

$$\Pi_k(x, \lambda) = \left(x, \max\left\{-\alpha_L \sqrt{\rho_k} e, \min\left\{\lambda, \alpha_U \sqrt{\rho_k} e\right\}\right\}\right). \tag{13}$$

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 penalty parameter, $\psi_{-1} = \sigma(X^0)$ and $k = 0$.

Step 1*: Stopping criterion*

If the condition:

$$\sigma(X^k) \leq \varepsilon \tag{14}$$

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 $D^{k,j} \in \mathbb{R}^n \times \mathbb{R}^m$ the solution of

$$\begin{aligned} \underset{D}{\text{minimize}}\ & \langle \nabla F_k(Y^{k,j}), D \rangle + \tfrac{1}{2} \left\langle \begin{bmatrix} Q_{k,j} & 0 \\ 0 & \frac{1}{\rho_k} I \end{bmatrix} D, D \right\rangle \\ \text{subject to}\ & \nabla H_k(Y^{k,j})^\top D = 0, \\ & Y^{k,j} + D \in \Omega \times \mathbb{R}^m. \end{aligned} \tag{15}$$

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:

$$\phi_k(Y^{k,j}, \theta_{j+1}) - \phi_k(X^{k,j}, \theta_{j+1}) \leq \frac{(r-1)}{2} \|H_k(X^{k,j})\|, \tag{16}$$

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:

$$\phi_k(Y^{k,j} + t_j D^{k,j}, \theta_{j+1}) - \phi_k(X^{k,j}, \theta_{j+1}) \leq \frac{(r-1)}{2} \|H_k(X^{k,j})\|. \tag{17}$$

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 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 The subproblem (15) is equivalent to solve a problem similar to (3). It can be seen that if $D^{k,j}$ is a solution of (15), then $(x, \lambda) = Y^{k,j} + D^{k,j}$ is a solution of

$$\begin{aligned} \underset{(x,\lambda)}{\text{minimize}}\ & \langle \nabla f(x^{k,j}), x - x^{k,j} \rangle + \tfrac{1}{2} \langle Q_{k,j}(x - x^{k,j}), x - x^{k,j} \rangle + \tfrac{1}{2\rho_k} \|\lambda\|^2 \\ \text{subject to}\ & h(x^{k,j}) + \nabla h(x^{k,j})^\top (x - x^{k,j}) - \tfrac{1}{\rho_k}(\lambda - \lambda^k) = 0, \\ & x \in \Omega. \end{aligned} \tag{18}$$

When k is large enough and $X^{k+1} = Y^{k,0} + D^{k,0}$ then the sequence generated by Algorithm 2 is the same as the sequence generated by the sSQP method.

The remaining part of this section is devoted to the well-definiteness of Algorithm 2, which depends on the well-definiteness of Step 2.

Let us define the function

$$\mu^k(x, y) = \lambda^k + \rho_k h(x) + \rho_k \nabla h(x)^\top (y - x), \quad x, y \in \Omega. \tag{19}$$

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 ρ_k and λ^k such that λ^k and $\mu^k(x, y)$ belongs 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 (4) is equivalent to:

$$\begin{aligned} \underset{(x,\lambda)}{\text{minimize}}\ & F_k(x, \lambda) \\ \text{subject to}\ & H_k(x, \lambda) = 0, \\ & (x, \lambda) \in \mathcal{W}_k. \end{aligned} \tag{20}$$

It can be seen that Step 2 of Algorithm 2 is a direct application of Algorithm 1 applied to the problem (20). 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 conditions (6), (7), (8) and (9) of Algorithm 1.

Lemma 1 *Let $\{X^{k,j}\}$, $\{Y^{k,j}\}$ 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

$$F_k(Y^{k,j}) \leq F_k(X^{k,j}) + \beta_k \|H_k(X^{k,j})\|, \tag{21}$$
$$\|H_k(Y^{k,j})\| \leq r\|H_k(X^{k,j})\|, \tag{22}$$

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 $j \geq 0$ 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 $D^{k,j} = \left(x - x^{k,j}, \lambda - \left(\lambda^k + \rho_k h(x^{k,j})\right)\right)$. From the equality constraint in (15) we obtain

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

Solving for λ and using (19) we get $\lambda = \mu^k(x^{k,j}, x)$. Since $x^{k,j} \in \Omega$, and $x \in \Omega$ (because $Y^{k,j} + D^{k,j} \in \Omega \times \mathbb{R}^m$) we have that $\lambda \in [\lambda_L^k, \lambda_U^k]$. 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 (22) holds since $H_k(Y^{k,j}) = 0$.

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

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

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 (21) holds. □

It remains to prove that the direction $D^{k,j}$ generated by the subproblem (15) satisfies conditions (8) and (9).

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

$$F_k(Y^{k,j} + tD^{k,j}) \leq F_k(Y^{k,j}) - \eta_k t \|D^{k,j}\|^2, \tag{23}$$

$$\|H_k(Y^{k,j} + tD^{k,j})\| \leq \|H_k(Y^{k,j})\| + \bar{\eta}_k t^2 \|D^{k,j}\|^2, \tag{24}$$

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 (15). 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

$$\langle \nabla F_k(Y^{k,j}), D^{k,j} \rangle \leq -\frac{c_k}{2} \left\| D^{k,j} \right\|^2. \tag{25}$$

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 (25) and Lipschitzianity of ∇F_k we get

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

for all $t \in [0, 1]$. Therefore, (23) is valid for all $t \in [0, \tau_k]$ with $\tau_k = \min \left\{ 1, \frac{c_k}{2L_k} \right\}$ 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

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

for all $t \in [0, 1]$. Thus, (24) holds for all $t \in [0, 1]$ with $\bar{\eta}_k = L_k/2$. Therefore, (23) and (24) 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 $\left\{ X^k \right\}$ can be generated by Algorithm 2.

Lemma 3 *Algorihm 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 $k \geq 0$. 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 $\|D^{k,j}\| \leq \epsilon_k$ is satisfied for j sufficiently large, so Step 2 of Algorithm 2 is executed only a finite number of iterations. Therefore, 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 (15) nor the success of execution of Step 2.1 of Algorithm 2. In [19, 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 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 [22,31], $\bar{L}(x, \lambda, \rho) : \mathbb{R}^n \times \mathbb{R}^m \times (0, +\infty) \to \mathbb{R}$, defined by

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

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

$$x^{k+1} \text{ is a solution of } \underset{x \in \Omega}{\text{minimize}}\, \bar{L}(x, \lambda^k, \rho_k), \tag{26}$$

$$\lambda^{k+1} = \lambda^k + \rho_k h(x^{k+1}). \tag{27}$$

The augmented Lagrangian method had been studied by many authors [1,5, 12,14,11,30,7,3,4,9,8,25], among other literature (see also [6,29]).

Numerical implementations attempt to solve (26) inexactly, by using a suitable criterion. For example, some codes based on the augmented Lagrangian method, such as LANCELOT [13] and ALGENCAN [2], define $x^{k+1} = x$ if the residual of the minimization problem in (26) at x is less than some tolerance ϵ_k, i.e.,

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

As it was noticed in [6, p. 102], the minimization problem in (26) becomes ill-conditioned for large values of the penalty parameter ρ_k. This is a typical problem of all penalty methods.

To overcome this drawback we can see that problem (26)-(27) is equivalent to problem (4), which is well-conditioned [6]. Such equivalence comes from (4) 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*

$$\left\langle \nabla f(y^k) + M_k\left(x^{k+1} - y^k\right) + \nabla h(y^k)\nu^{k+1}, y - x^{k+1}\right\rangle \geq 0, \quad \forall\, y \in \Omega, \tag{28}$$

$$h(y^k) + \nabla h(y^k)^\top\left(x^{k+1} - y^k\right) - \tfrac{1}{\rho_k}\left(\nu^{k+1} - \lambda^k\right) = 0, \tag{29}$$

$$\|x^{k+1} - y^k\|^2 + \|\nu^{k+1} - (\lambda^k + \rho_k h(y^k))\|^2 < \epsilon_k^2. \tag{30}$$

Proof Note that the optimality conditions of problem (15) are

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

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 $\left\|D^{k,j(k)}\right\| < \epsilon_k$. Let us call $y^k = x^{k,j(k)}$, the primal component of $Y^{k,j(k)}$, ν^{k+1} the dual component of $Y^{k,j(k)} + D^{k,j(k)}$ 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 (13) 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

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

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

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

Notice that from the second relation we obtain $\nu^{k+1} = \xi^k$. Therefore, $D^{k,j(k)}$ is a solution of (15) if and only if

$$\left\langle \nabla f(y^k) + M_k\left(x^{k+1} - y^k\right) + \nabla h(y^k)\nu^{k+1}, y - x^{k+1}\right\rangle \geq 0, \quad \forall\, y \in \Omega, \tag{31}$$

$$h(y^k) + \nabla h(y^k)^\top\left(x^{k+1} - y^k\right) - \tfrac{1}{\rho_k}\left(\nu^{k+1} - \lambda^k\right) = 0. \tag{32}$$

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

$$\|x^{k+1} - y^k\|^2 + \|\nu^{k+1} - (\lambda^k + \rho_k h(y^k))\|^2 < \epsilon_k^2. \tag{33}$$

□

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

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

Proof From the definition of Π_k (13), we have that λ^{k+1} belongs to the close 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

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

In 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* $\left\{x^k\right\}$ *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*

$$\underset{x\in\Omega}{\text{minimize}}\ \tfrac{1}{2}\left\|h(x)\right\|^2. \tag{34}$$

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

$$\lim_{k\in\mathcal{K}} x^{k+1} = \bar{x}. \tag{35}$$

Since ϵ_k tends to zero, and using (35) and (30) we get

$$\lim_{k\in\mathcal{K}} y^k = \bar{x}. \tag{36}$$

Proof of 1. Let us consider 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 close set $[-\alpha_L\sqrt{\bar{\rho}}e, \alpha_U\sqrt{\bar{\rho}}e]$ for all $k \geq k_0$, that is, $\left\{\lambda^{k+1}\right\}$ 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

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

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 (35) we have that $h(\bar{x}) = 0$. Passing onto a sebsequence if necessary, assume that $\lim_{k\in\mathcal{K}} \lambda^k = \bar{\lambda}$ (because of the boundedness of $\{\lambda^k\}$). Taking limits in (30) for $k \in \mathcal{K}$, using (36) and the facts that $h(\bar{x}) = 0$ and $\rho_k = \bar{\rho}$ for k large enough, we deduce that

$$\lim_{k\in\mathcal{K}} \nu^{k+1} = \bar{\lambda}. \tag{37}$$

From (35), (36), (37) and the fact that $\{M_k\}$ are uniformly bounded, taking limits in (28) for $k \in \mathcal{K}$, we conclude that

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

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

Proof of 2. Let us consider the case when $\{\rho_k\}$ tends to infinity. Taking limits in (29) we obtain

$$\lim_{k\in\mathcal{K}} \frac{\nu^{k+1}}{\rho_k} = h(\bar{x}). \tag{38}$$

where we have used (35), (36) and the fact that $\{\lambda^k/\rho_k\}$ converges to zero by Proposition 2.

Dividing (28) by ρ_k, and using (35), (36), (38) and the fact that $\{M_k\}$ are uniformly bounded, taking limits for $k \in \mathcal{K}$, we conclude that

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

Hence, we conclude that $\bar{x}$ is a stationary point of problem (34). □

5 Penalty boundedness results

From now on we will prove that, under suitable conditions, the sequence of penalty parameter $\{\rho_k\}$ generated by Algorithm 2 remains bounded. Let us assume 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 $\bar{\lambda} \in (-\alpha_L\sqrt{\rho_k}e, \alpha_U\sqrt{\rho_k}e)$ for all $k \geq k_0$.

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

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})d, d \right\rangle > 0 \quad \forall d \in \mathcal{C} \setminus \{0\}. \tag{39}$$

where

$$\mathcal{C} = \left\{ d \in \mathbb{R}^n \,\middle|\, \begin{array}{l} \langle \nabla f(\bar{x}), d\rangle = 0,\ \nabla h(\bar{x})^\top d = 0, \\ d_i \leq 0 \text{ if } \bar{x}_i = b_i, d_i \geq 0 \text{ if } \bar{x}_i = a_i, i = 1, \ldots, n \end{array} \right\} \tag{40}$$

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

$$\epsilon_k \leq \chi(\sigma(x^k, \lambda^k)) \tag{41}$$

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 (30), 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 (28) by $\|\nu^{k+1}\|$ and taking limits, we obtain that

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

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.

Let $\bar{\nu}$ be a limit point of $\{\nu^{k+1}\}$. Passing onto subsequences if necessary and taking limits in (28) we get

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

This means that $\bar{\nu}$ is a Lagrange multiplier associated to $\bar{x}$. From A2 we conclude that $\bar{\nu} = \bar{\lambda}$. Hence, $\{\nu^{k+1}\}$ 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. Therefore $\lambda^{k+1} = \nu^{k+1}$ 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* $k_0 \in \mathbb{N}$, β_1, $\beta_2 > 0$ *such that for all* $k \geq k_0$,

$$\beta_1\|(x^k, \lambda^k) - (\bar{x}, \bar{\lambda})\| \leq \sigma(x^k, \lambda^k) \leq \beta_2\|(x^k, \lambda^k) - (\bar{x}, \bar{\lambda})\|. \tag{42}$$

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 [18, Lemma 5, Theorem 2] there exists $\beta_1 > 0$ such that $\sigma(x, \lambda) \geq \beta_1 \|(x, \lambda) - (\bar{x}, \bar{\lambda})\|$ 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*

$$\left(1 - \frac{c_2}{\rho_k}\right) \sigma(x^{k+1}, \lambda^{k+1}) \leq \left(c_1 \eta_k + \frac{c_2}{\rho_k}\right) \sigma(x^k, \lambda^k), \tag{43}$$

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

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

and therefore

$$\begin{aligned}
\frac{\partial L}{\partial x}(x^{k+1}, \lambda^{k+1}) &- \left[\frac{\partial L}{\partial x}(y^k, \nu^{k+1}) + M_k \left(x^{k+1} - y^k\right)\right] \\
&= \left(\frac{\partial^2 L}{\partial x^2}(y^k, \lambda^{k+1}) - M_k\right)(x^{k+1} - y^k) + o(\|x^{k+1} - y^k\|) \\
&= O(\|x^{k+1} - y^k\|),
\end{aligned} \tag{44}$$

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 $\{M_k\}$ are uniformly bounded.

By definition of projection and (28) we have that

$$x^{k+1} = \Pi_\Omega \left(x^{k+1} - \left[\frac{\partial L}{\partial x}(y^k, \nu^{k+1}) + M_k \left(x^{k+1} - y^k\right)\right]\right). \tag{45}$$

Since Π_Ω is nonexpansive, using (44) and (45) we obtain

$$\left\| \Pi_\Omega \left(x^{k+1} - \frac{\partial L}{\partial x}(x^{k+1}, \lambda^{k+1})\right) - x^{k+1} \right\| \leq O(\|x^{k+1} - y^k\|). \tag{46}$$

On the other hand, by using (29) and the fact that $\lambda^{k+1} = \nu^{k+1}$ for k large enough, we get

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

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

$$\begin{aligned}\sigma(x^{k+1}, \lambda^{k+1}) &\leq O(\|x^{k+1} - y^k\|) + \sqrt{2}\|h(x^{k+1})\| \\ &\leq O(\|x^{k+1} - y^k\|) + \frac{\sqrt{2}}{\rho_k}\|\lambda^{k+1} - \lambda^k\| \\ &\leq c_1\epsilon_k + \frac{\sqrt{2}}{\rho_k}\|\lambda^{k+1} - \bar{\lambda}\| + \frac{\sqrt{2}}{\rho_k}\|\lambda^k - \bar{\lambda}\| \\ &\leq c_1\epsilon_k + \frac{c_2}{\rho_k}\sigma(x^{k+1}, \lambda^{k+1}) + \frac{c_2}{\rho_k}\sigma(x^k, \lambda^k),\end{aligned}$$

where in the third inequality we use (30) and for the last inequality we use (42).

Thus, by using A5, we conclude that

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

□

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 parameter $\{\rho_k\}$ 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 (43),

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

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 Algorihm 2 we conclude that $\rho_{k+1} = \rho_k$ for k large enough, in contradiction with the initial assumption. □

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 (42), 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

$$\frac{c_2}{\rho_k} \leq \frac{q\beta_1}{q\beta_1 + \beta_2} \quad \text{and} \quad \left(1 - \frac{c_2}{\rho_k}\right)^{-1} \leq \frac{q\beta_1 + \beta_2}{\beta_2}. \tag{47}$$

Hence,

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

where for the first inequality we use the left-hand side of (42), the second inequality comes from (43), for the third inequality we use (47) and the last inequality follows from the right-hand side relation in (42).

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 with the Algorithm 2. We have considered a set of nonlinear equality constrained problems from the Cuter collection [15]. All tests were performed on a PC running Linux, the algorithm was written in Fortran 2003 and compiled with the Intel Compiler 12.0.

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

- Algorithmic parameters: $\gamma = 0.99$, $\varepsilon = 10^{-6}$, $\epsilon_k = 1/(k+1)^2$ for all $k \geq 0$, $r = 0.99$ 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 $k \geq 0$.
- 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 (15) we used an implementation of the Goldfarb-Idnani algorithm [20] written by B. Turlach [32].

We remark that we only want to show viability of the approach proposed. An optimal choice of the parameters is out of the scope of this paper.

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

Table 1 Numerical experiments from the Cuter collection

Name	n	m	$\bar{\rho}$	$f(\bar{x})$	$\sigma(\bar{x}, \bar{\lambda})$	qp calls
BT1	2	1	1.0E+02	-1.000003	0.3492216E-07	5,2,1
BT2	3	1	1.0E+09	0.3256820E-01	0.1455329E-07	1,1,5
BT3	5	3	1.0E+01	4.093022	0.4014316E-06	1,3,2
BT4	3	2	1.0E+07	-3.704768	0.4203023E-11	1,1,3
BT5	2	2	1.0E-01	961.7152	0.4927916E-06	1,1,1
BT6	5	2	1.0E+06	0.2770448	0.1040323E-06	1,1,4
BT9	4	2	1.0E+00	-1.000000	0.4629197E-06	1,1,1
BT10	2	2	1.0E+00	-1.000000	0.4827694E-06	1,1,1
BT11	5	3	1.0E+03	0.8248918	0.2520146E-06	2,1,3
BT12	5	3	1.0E+00	6.188119	0.4814203E-06	1,1,1
BYRDSPHR	3	2	1.0E-02	-4.683300	0.4981593E-06	1,1,1
DIXCHLNG	10	5	1.0E+04	0.2920977E-11	0.1259454E-06	12,2,11
HS6	2	1	1.0E+01	0.5149135E-13	0.4110476E-06	1,1,1
HS7	2	1	1.0E+02	-1.732051	0.1332312E-07	1,3,1
HS8	2	2	1.0E+00	-1.000000	0.4417480E-06	1,1,1
HS9	2	1	1.0E+00	0.5000000	0.4878523E-06	1,1,1
HS27	3	1	1.0E+06	0.4000011E-01	0.1687748E-09	1,1,6
HS28	3	1	1.0E+04	0.3543495E-13	0.3856428E-06	1,2,1
HS39	4	2	1.0E+02	-1.000000	0.4629197E-06	1,1,1
HS42	4	2	1.0E+03	13.85786	0.7307294E-08	1,1,2
HS48	5	2	1.0E+12	0.2933576E-29	0.2884765E-13	1,2,5
HS49	5	2	1.0E-02	0.3503371E-11	0.1038659E-07	2,1,18
HS50	5	3	1.0E+06	0.6883040E-15	0.7299110E-07	1,1,2
HS51	5	3	1.0E+13	0.5053640E-30	0.1055938E-13	1,3,4
HS52	5	3	1.0E+02	5.326644	0.2636263E-07	3,2,4
HS61	3	2	1.0E+02	-143.6461	0.3309869E-07	1,2,1
HS77	5	2	1.0E+06	0.2415051	0.1130220E-07	1,1,3
HS79	5	3	1.0E+05	0.7877682E-01	0.2686601E-08	1,1,2
MARATOS	2	1	1.0E-01	-1.000000	0.3845145E-06	1,1,1
ORTHREGB	27	6	1.0E-01	0.3004344E-13	0.3513701E-06	1,1,1
S316-322	2	1	1.0E+03	334.3146	0.3563555E-07	2,1,4

penalty parameter $\bar{\rho}$, the optimal objective function $f(\bar{x})$, the natural residual $\sigma(\bar{x}, \bar{\lambda})$ and number of calls to the quadratic solver in the last three iterations.

Our implementation found the same solutions obtained by LANCELOT [24] in the whole set of test problems. In case of problems BT3 and HS49 we used different starting points.

We observe that the penalty parameter ρ remains bounded ($\rho \leq 10^6$) in 90% of the problems, which confirms that, according to Theorem 2, the solution found by the Algorithm 2 is a KKT point.

In most of the problems we perform 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.

This preliminary numerical results are promising and show the robustness of our algorithm.

7 Conclusions

In this paper we present a new hybrid method for solving equality constrained optimization problems. The proposed method combines: (a) the sSQP method, which has good local behavior; (b) the augmented Lagrangian method, which has global convergence properties; and (c) the IR method, which is an appropriate strategy to inexactly solve the subproblems.

In our method, the ill-conditioned subproblems, due to large values of the penalty parameters, are overcome. Moreover, no constraint qualifications are needed. This features makes 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 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 some constraint qualifications hold (strict Mangasarian-Fromovitz and second order sufficient optimality conditions), 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 problem from the Cuter collection. The algorithm found the same solutions obtained by LANCELOT and confirmed the theoretical results.

References

1. The method of penalty estimates for conditional extremum problems. USSR Computational Mathematics and Mathematical Physics **13**(1), 42 – 58 (1974)
2. ALGENCAN: URL `http://www.ime.usp.br/~egbirgin/tango/`
3. 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)
4. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. Math. Program. **111**(1-2, Ser. B), 5–32 (2008)
5. Bertsekas, D.P.: Multiplier methods: a survey. Automatica—J. IFAC **12**(2), 133–145 (1976)
6. Bertsekas, D.P.: Constrained optimization and Lagrange multiplier methods. Computer Science and Applied Mathematics. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York (1982)
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., Fernández, D., Martínez, J.M.: On the boundedness of penalty parameters in an augmented lagrangian method with lower level constraints. Optimization Methods and Software (to appear in 2011)
9. Birgin, E.G., Floudas, C.A., Martínez, J.M.: Global minimization using an augmented Lagrangian method with variable lower-level constraints. Math. Program. **125**(1, Ser. A), 139–162 (2010)
10. 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)

11. 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)
12. 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)
13. Conn, A.R., Gould, N.I.M., Toint, P.L.: LANCELOT, *Springer Series in Computational Mathematics*, vol. 17. Springer-Verlag, Berlin (1992). A Fortran package for large-scale nonlinear optimization (release A)
14. 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)
15. CUTER: URL `http://www.cuter.rl.ac.uk/index.html`
16. Fernández, D.: A quasi-newton strategy for the sSQP method for variational inequality and optimization problems. URL `http://www.optimization-online.org/DB_HTML/2010/10/2764.html`
17. Fernández, D., Solodov, M.: Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program. **125**(1, Ser. A), 47–73 (2010)
18. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**(1, Ser. A), 91–124 (2002)
19. Fischer, A., Friedlander, A.: A new line search inexact restoration approach for nonlinear programming. Comput. Optim. Appl. **46**(2), 333–346 (2010)
20. Goldfarb, D., Idnani, A.: A numerically stable dual method for solving strictly convex quadratic programs. Math. Programming **27**(1), 1–33 (1983)
21. Hager, W.W.: Stabilized sequential quadratic programming. Comput. Optim. Appl. **12**(1-3), 253–273 (1999). Computational optimization—a tribute to Olvi Mangasarian, Part I
22. Hestenes, M.R.: Multiplier and gradient methods. J. Optimization Theory Appl. **4**, 303–320 (1969)
23. Izmailov, A., Solodov, M.: Stabilized SQP revisited. Mathematical Programming pp. 1–28
24. LANCELOT: URL `http://www.cse.scitech.ac.uk/nag/lancelot/lancelot.shtml`
25. 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)
26. Martinez, 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)
27. Martínez, J.M., Pilotta, E.A.: Inexact-restoration algorithm for constrained optimization. J. Optim. Theory Appl. **104**(1), 135–163 (2000)
28. 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)
29. Nocedal, J., Wright, S.J.: Numerical optimization. Springer Series in Operations Research. Springer-Verlag, New York (1999)
30. Pennanen, T.: Local convergence of the proximal point algorithm and multiplier methods without monotonicity. Math. Oper. Res. **27**(1), 170–191 (2002)
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: URL `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 (electronic) (2002)
35. Wright, S.J.: Constraint identification and algorithm stabilization for degenerate nonlinear programs. Math. Program. **95**(1, Ser. B), 137–160 (2003). ISMP 2000, Part 3 (Atlanta, GA)