A Flexible Inexact-Restoration Method for Constrained Optimization

L. F. Bueno · G. Haeser · J. M. Martínez

Abstract We introduce a new flexible inexact-restoration algorithm for constrained optimization problems. In inexact-restoration methods, each iteration has two phases. The first phase aims at improving feasibility and the second phase aims to minimize a suitable objective function. In the second phase, we also impose bounded deterioration of the feasibility, obtained in the first phase. Here, we combine the basic ideas of the Fischer-Friedlander approach for inexact-restoration with the use of approximations of the Lagrange multipliers. We present a new option to obtain a range of search directions in the optimization phase, and we employ the sharp Lagrangian as merit function. Furthermore, we introduce a flexible way to handle sufficient decrease requirements and an efficient way to deal with the penalty parameter. Global convergence of the new inexact-restoration method to KKT points is proved under weak constraint qualifications.

L. F. Bueno (✉)
Institute of Science and Technology, Federal University of São Paulo, São José dos Campos, SP, Brazil
e-mail: lfelipebueno@gmail.com

G. Haeser
Department of Applied Mathematics, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, SP, Brazil
e-mail: ghaeser@ime.usp.br

J. M. Martínez
Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas, SP, Brazil
e-mail: martinez@ime.unicamp.br

1 Introduction

Practical methods for solving nonlinearly constrained optimization problems are iterative. Given an iterate, one tries to find a better approximation to the solution taking into account feasibility and optimality criteria. Several traditional nonlinear programming methods preserve feasibility of all the iterates, so that the improvement criterion relies only on objective function values [1–9]. These methods are very effective in many cases, but they tend to be slow in the presence of constraints with high curvature. Modern inexact-restoration (I-R) methods, on the other hand, admit infeasibility of iterates and employ inexact nonspecific restoration procedures, which may be adapted to different classes of problems. See [10–23].

The common features of the I-R methods mentioned above are the following:

- Given the current iterate, a sufficiently more feasible point (inexactly restored point) is computed by any arbitrary method. This is the Feasibility Phase of the I-R iteration.
- At the Optimality Phase of the I-R iteration, one minimizes (approximately) a suitable objective function (the Lagrangian or some approximation) for obtaining a trial point. The trial point is compared with the current iterate in terms of feasibility and optimality (using merit functions as in [10,11,19], or filters, as in [12,18]). If the trial point is acceptable, we define it as the new iterate. Otherwise, we find a new trial point "closer" to the inexactly restored point. The fact that the new trial point should be "closer" to the inexactly restored point, instead of closer to the current iterate, is crucial in the philosophy of I-R methods, because ultimately the inexactly restored point is considered to be better than the current iterate.

The first I-R methods [10,11] motivated the introduction of a sequential optimality condition called AGP (approximate gradient projection) [24]. For several years, it was thought that AGP was equivalent to the approximate KKT condition (AKKT) implicitly used to define stopping criteria in nonlinear programming algorithms. Surprisingly, it was recently shown [25] that AGP is stronger than AKKT; moreover, one of the AGP variations, called linear AGP (L-AGP), is stronger than AGP. These theoretical facts support one of the decisions taken in our I-R method: When linear constraints are present in the problem, feasibility with respect to these constraints should be satisfied at all the iterations.

Fischer and Friedlander [19] introduced a new line search by means of which an I-R method exhibits interesting global convergence properties. The resulting approach turns out to be simpler than more classical approaches based on trust regions or filters. The main convergence result of [19] states that under proper assumptions, the search direction, used at the optimality phase, tends to zero. This result complements the suggestion of Gomes-Ruggiero, Martínez and Santos [17] who claim that by means of the spectral gradient choice [26–30] of the first optimality-phase trial point, I-R turns out to be the natural generalization to nonconvex and constrained optimization of the spectral projected gradient (SPG) method [31–33], which is currently used for large-scale convex and constrained minimization.

In [17], spectral gradient tangent directions and the trust-region globalization approach of [10] were invoked for proving convergence. With the aim of getting

closer to quadratic convergence [13], we use here a search direction that comes from solving a strictly convex Newton-related quadratic minimization problem. The resulting method is the natural generalization of the variable metric method [34] for nonlinear programming in the same sense that the method [17] is a generalization of SPG.

In this work, we introduce an improved line search I-R algorithm based on the Fischer–Friedlander approach. The proposed method is more flexible than the approaches in [10,11,19], when dealing with sufficient decrease. Several typical I-R algorithmic requirements are relaxed and, therefore, larger steps are more likely to be accepted by the algorithm. We also employ approximations of Lagrange multipliers in the merit function in order to add relevant information when solving the problem. An optimization procedure for general constrained optimization problems is introduced.

This paper is organized as follows. In Sect. 2, we introduce the new I-R method and we prove that it is well defined. In Sect. 3, we state the assumptions that will allow us to prove convergence starting from arbitrary initial points, and we prove feasibility and optimality of the limit points. The plausibility of the assumptions used to prove convergence is discussed in Sect. 4. Conclusions are given in Sect. 5.

2 Basic Method

In this section, we consider the constrained optimization problem in the form:

$$\min f(x) \quad \text{s.t.} \quad h(x) = 0, \ x \in \Omega, \tag{1}$$

where $\Omega \subset \mathbb{R}^n$ is a polytope, $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$ and both f and h are smooth. Every smooth, finite dimensional, and constrained optimization problem may be written in the form (1) using slack variables, if necessary.

For all $x \in \Omega$ and $\lambda \in \mathbb{R}^m$, we define the Lagrangian $L(x, \lambda)$ by

$$L(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i h_i(x).$$

Given a penalty parameter $\theta \in [0, 1]$, we consider, for all $x \in \Omega, \lambda \in \mathbb{R}^m$, the following merit function that consists of a convex combination of the Lagrangian function and the infeasibility measure:

$$\Phi(x, \lambda, \theta) := \theta L(x, \lambda) + (1 - \theta)\|h(x)\|. \tag{2}$$

Note that, when $\theta = 1$, the merit function coincides with the Lagrangian function and, when $\theta = 0$, it coincides with the infeasibility measure. In these cases, we are neglecting either the objective function or the feasibility at the point x. In our method θ is always smaller than one, and we will prove that it is also bounded away from zero. In this paper, the symbol $\| \cdot \|$ denotes the Euclidean norm, although many times it can be replaced by an arbitrary norm.

The merit function (2) is essentially the sharp Lagrangian defined in [35] and employed in [36–38], in the context of nonsmooth optimization, and in [11] in the I-R framework.

Algorithm 2.1 below is a flexible form of the I-R method with the line search procedure of [19]. One improvement here with respect to [19] is that we are able to use Lagrange multipliers in the algorithm (thus employing (2)) preserving convergence. Moreover, sufficient descent requirements are relaxed so that acceptance of "Pure Newtonian" steps will be more frequent.

Algorithm 2.1 Flexible inexact-restoration with sharp Lagrangian.

Step 0 Initialization
   As initial approximation we choose, arbitrarily, $x^0 \in \Omega$ and $\lambda^0 \in \mathbb{R}^m$. We initialize $\theta_{-1} \in]0, 1[$ and $k \leftarrow 0$.

Step 1 Restoration step
   If $\|h(x^k)\| = 0$, define $y^{k+1} := x^k$. Otherwise, try to compute $y^{k+1} \in \Omega$ such that

$$\|h(y^{k+1})\| < \|h(x^k)\|. \tag{3}$$

If this is not possible, we say that Step 1 is deemed unsuccessful and the algorithm stops.

Step 2 Estimation of Lagrange Multipliers
   Compute $\lambda^{k+1} \in \mathbb{R}^m$. (Practical effective ways for this choice will be given later.)

Step 3 Descent direction
   Compute $d^k \in \mathbb{R}^n$ such that $y^{k+1} + d^k \in \Omega$,

$$\nabla L(y^{k+1}, \lambda^{k+1})^T d^k < 0, \quad \text{if} \quad d^k \neq 0, \tag{4}$$

and, for some $c_k > 0$ and all small enough $t > 0$:

$$\|h(y^{k+1} + td^k)\| \leq \|h(y^{k+1})\| + c_k t^2 \|d^k\|^2. \tag{5}$$

Step 4 Globalization

Step 4.1 Sufficient decrease parameters
   If $\|h(x^k)\| = \|h(y^{k+1})\| = 0$, take $r_k \in]0, 1[$. Else, define $r_k \neq 0$ such that

$$r_k \in \left[\frac{\|h(y^{k+1})\|}{\|h(x^k)\|}, 1 \right[. \tag{6}$$

Set

$$s_k \in [0, r_k[. \tag{7}$$

Step 4.2 Penalty parameter computation

Compute θ_k as the supremum of the values of $\theta \in [0, \theta_{k-1}]$ such that

$$\Phi\big(y^{k+1}, \lambda^{k+1}, \theta\big) \leq \Phi\big(x^k, \lambda^k, \theta\big) + \frac{1 - s_k}{2} \big(\|h\big(y^{k+1}\big)\| - \|h(x^k)\|\big). \qquad (8)$$

Step 4.3 Line search

If $\|d^k\| = 0$ we define $t_k := 1$.

Otherwise, by means of some backtracking procedure, compute $t_k \in [0, 1]$ such that

$$L\big(y^{k+1} + t_k d^k, \lambda^{k+1}\big) < L\big(y^{k+1}, \lambda^{k+1}\big) \qquad (9)$$

and

$$\Phi\big(y^{k+1} + t_k d^k, \lambda^{k+1}, \theta_k\big) \leq \Phi\big(x^k, \lambda^k, \theta_k\big) + \frac{1 - r_k}{2} \big(\|h\big(y^{k+1}\big)\| - \|h(x^k)\|\big). \qquad (10)$$

Moreover, the backtracking procedure should be such that either $t_k = 1$ or there exists $\bar{t}_k \in [t_k, 10 t_k]$ such that

$$L\big(y^{k+1} + \bar{t}_k d^k, \lambda^{k+1}\big) \geq L\big(y^{k+1}, \lambda^{k+1}\big) \qquad (11)$$

or

$$\Phi\big(y^{k+1} + \bar{t}_k d^k, \lambda^{k+1}, \theta_k\big) > \Phi\big(x^k, \lambda^k, \theta_k\big) + \frac{1 - r_k}{2} \big(\|h\big(y^{k+1}\big)\| - \|h(x^k)\|\big). \qquad (12)$$

Step 4.4 Iteration update

Set

$$x^{k+1} = y^{k+1} + t_k d^k, \qquad (13)$$

update $k \leftarrow k + 1$ and go to Step 1.

We have defined Algorithm 2.1 without an explicit stopping criterion, which, of course, is necessary in practical implementations. This means that, in principle, the algorithm generates an infinite sequence whose theoretical properties are proved in Sects. 2 and 3. In Sect. 5, we define suitable criteria for stopping the execution of the algorithm.

Algorithm 2.1 is a simple inexact-restoration algorithm, where even the condition of d^k being in the tangent subspace is relaxed. However, it is interesting to show that this simplified method is well defined in the sense that an iteration can be completed if the inexactly restored point y^{k+1} can be computed at every iteration. In the following theorem, we do not make use of Lipschitz conditions at all, neither on the objective function, nor on the constraints.

Theorem 2.1 *For all $x^k \in \Omega$, if the point y^{k+1} at Step 1 of Algorithm 2.1 is successfully computed, then the iterate x^{k+1} is well defined.*

Proof By the hypothesis, we have that Step 1 is well defined. Steps 2 and 3 can be completed observing that $\lambda^{k+1} = 0 \in \mathbb{R}^m$ and $d^k = 0 \in \mathbb{R}^n$ are always acceptable.

Although these options will not be our choice, this ensures that both steps are well defined. We will show now that the other computations required in the iteration are possible, whenever $\|h(y^{k+1})\| < \|h(x^k)\|$ or $y^{k+1} = x^k$.

Computation of θ_k: If $y^{k+1} = x^k$, we have that $\|h(y^{k+1})\| = \|h(x^k)\| = 0$ and, thus, $L(x^k, \lambda) = L(y^{k+1}, \mu)$ for all $\lambda, \mu \in \mathbb{R}^m$. This means that (8) is obvious for all θ and so $\theta_k = \theta_{k-1}$.

If $y^{k+1} \neq x^k$, by (3), we have that $\|h(y^{k+1})\| < \|h(x^k)\|$ and, since $0 \leq s_k < 1$,

$$\|h(y^{k+1})\| - \|h(x^k)\| < \frac{1 - s_k}{2}\left(\|h(y^{k+1})\| - \|h(x^k)\|\right).$$

Therefore, for small enough θ, we have that

$$\begin{aligned}
\|h(y^{k+1})\| - \|h(x^k)\| &\leq \frac{\theta}{1 - \theta}\left[L(x^k, \lambda^k) - L(y^{k+1}, \lambda^{k+1})\right] \\
&\quad + \frac{1 - s_k}{2(1 - \theta)}\left(\|h(y^{k+1})\| - \|h(x^k)\|\right).
\end{aligned}$$

This implies that, for small enough $\theta > 0$, (8) takes place. By direct calculations, we note that the parameter θ_k can be computed in the following way: If (8) is verified with $\theta = \theta_{k-1}$, then take $\theta_k = \theta_{k-1}$. Otherwise, we compute

$$\theta_k = \frac{(1 + s_k)\left(\|h(x^k)\| - \|h(y^{k+1})\|\right)}{2\left[L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) + \|h(x^k)\| - \|h(y^{k+1})\|\right]}. \tag{14}$$

Computation of t_k: We only need to consider the case in which $\|d^k\| \neq 0$.

By (4), we have that (9) necessarily holds for small enough t. Let us prove now that (10) also holds for small t.

Assume first that $y^{k+1} = x^k$, then, $\|h(y^{k+1})\| = \|h(x^k)\| = 0$ and $L(x^k, \lambda) = f(x^k)$ for all $\lambda \in \mathbb{R}^m$. Thus

$$\Phi(x^k, \lambda^k, \theta_k) + \frac{1 - r_k}{2}\left(\|h(y^{k+1})\| - \|h(x^k)\|\right) = \theta_k L(x^k, \lambda^k) = \theta_k f(y^{k+1}), \tag{15}$$

and, by (5),
$$\|h(y^{k+1} + td^k)\| \leq c_k t^2 \|d^k\|^2. \tag{16}$$

Now, by (4), for small enough t, we have that

$$\begin{aligned}
L(y^{k+1} + td^k, \lambda^{k+1}) &\leq L(y^{k+1}, \lambda^{k+1}) + \frac{t}{2}\nabla L(y^{k+1}, \lambda^{k+1})^T d^k \\
&= f(y^{k+1}) + \frac{t}{2}\nabla L(y^{k+1}, \lambda^{k+1})^T d^k
\end{aligned} \tag{17}$$

and

$$\frac{\theta_k}{2}\nabla L(x^k, \lambda^{k+1})^T d^k + (1 - \theta_k)c_k t\|d^k\|^2 < 0. \tag{18}$$

By (15), (16), (17), and (18), for small enough t, we have that

$$
\begin{aligned}
\Phi\big(y^{k+1} + t d^k, \lambda^{k+1}, \theta_k\big) &= \theta_k L\big(y^{k+1} + t d^k, \lambda^{k+1}\big) + (1 - \theta_k) \,\big\|h\big(y^{k+1} + t d^k\big)\big\| \\
&\leq \theta_k\big(f(y^{k+1}) + \frac{t}{2} \nabla L\big(y^{k+1}, \lambda^{k+1}\big)^T d^k\big) + (1 - \theta_k) c_k t^2 \|d^k\|^2 \\
&\leq \Phi\big(x^k, \lambda^k, \theta_k\big) + \frac{1 - r_k}{2}\big(\|h\big(y^{k+1}\big)\| - \|h\big(x^k\big)\|\big).
\end{aligned}
$$

So we conclude that (10) holds for small enough t when $y^{k+1} = x^k$.

Now, let us show that t_k is well defined in the case that $\|h(y^{k+1})\| < \|h(x^k)\|$. In this case, by (7), we have that

$$
\frac{1 - s_k}{2}\big(\|h(y^{k+1})\| - \|h(x^k)\|\big) < \frac{1 - r_k}{2}\big(\|h(y^{k+1})\| - \|h(x^k)\|\big). \qquad (19)
$$

By (8) and (19),

$$
\Phi\big(y^{k+1}, \lambda^{k+1}, \theta_k\big) < \Phi\big(x^k, \lambda^k, \theta_k\big) + \frac{1 - r_k}{2}\big(\|h(y^{k+1})\| - \|h(x^k)\|\big).
$$

By continuity,

$$
\Phi\big(y^{k+1} + t d^k, \lambda^{k+1}, \theta_k\big) \leq \Phi\big(x^k, \lambda^k, \theta_k\big) + \frac{1 - r_k}{2}\big(\|h(y^{k+1})\| - \|h(x^k)\|\big)
$$

for small enough t. This ensures that t_k can be computed in a finite time by means of a backtracking procedure. This completes the proof. $\qquad\blacksquare$

Observe that (10) is easier to be fulfilled, if r_k is chosen to be close to 1 in (6). This means that the chances of accepting large steps t_k increase if we choose $r_k \approx 1$. Moreover, the gap in (19) is larger as s_k is closer to zero, which also favors the acceptance of large steps. On the other hand, θ_k, the weight of the Lagrangian in the merit function, is larger as s_k is closer to r_k. Since the direction d^k is a descent direction for the Lagrangian, it is natural that large values of θ_k increase the chances to accept large steps. We believe that the best choice of s_k strongly depends on the nonlinearity of the constraints in (1). For simplicity in our implementation, we decide to take the largest gap in (19) setting $s_k = 0$ for all k. This flexible way of handling sufficient decrease allows the algorithm to take larger steps than the ones in [19], where r_k and s_k are equal and fixed for every k.

Theorem 2.1 showed that, for well-definedness, the following conditions are essential: (a) simple descent of $\|h(y^{k+1})\|$ with respect to $\|h(x^k)\|$; (b) the direction d^k should be quasi-tangent (5); and (c) d^k must be a descent direction for the Lagrangian L.

3 Assumptions and Global Convergence

The global convergence proof for Algorithm 2.1 requires a problem assumption (3.1) and several algorithmic assumptions (3.2–3.6).

Assumption 3.1 *The problem should be such that Lipschitz conditions hold both on the gradients of f and h. Namely, there exists $\eta > 0$ such that, for all $x, y \in \Omega$,*

$$\|\nabla f(y) - \nabla f(x)\| \leq \eta \|y - x\| \text{ and } \|\nabla h(y) - \nabla h(x)\| \leq \eta \|y - x\|. \quad (20)$$

The algorithmic assumptions involve conditions that should be satisfied by the different steps of the algorithm in order to guarantee global convergence. For the introduction of these assumptions, we will denote the Euclidean projection of x on a non-empty, closed, and convex set Ω by $P_\Omega(x)$ and $\{x^k\}, \{y^k\}, \{\lambda^k\}$ will be the sequences generated by Algorithm 2.1. Note that these sequences have infinitely many terms, except in the case when the restoration step fails and (3) cannot be obtained for some k.

Assumption 3.2 *For all $k \in \mathbb{N}$, Step 1 of the Algorithm is successful and there exists $r \in [0, 1]$ and $\beta > 0$ such that*

$$r_k \leq r \quad (21)$$

and

$$L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) \leq \beta \|h(x^k)\|. \quad (22)$$

Conditions (21) and (22) involve the restoration step. Condition (21) states that sufficient uniform feasibility improvement should be obtained by the restoration procedure. In most I-R implementations, r is an algorithmic parameter and the algorithm stops declaring "restoration failure" if the restoration procedure fails to satisfy (21) after reasonable computer time. Here we adopt the practical point of view that even conservative algorithmic parameters r (say $r = 0.99$) could be excessively strict at some iterations of the algorithm, which, however, could converge smoothly under single descent requirements such as (3). On the other hand, the requirement (21) for some unknown value of r is usually satisfied under regularity assumptions on the constraints. Since $y^{k+1} = x^k$ if $\|h(x^k)\| = 0$, and r_k is defined as in (6) if $y^{k+1} \neq x^k$, condition (21) implies that

$$\|h(y^{k+1})\| - \|h(x^k)\| \leq r_k \|h(x^k)\| - \|h(x^k)\| \leq -(1 - r)\|h(x^k)\|. \quad (23)$$

Condition (22) requires that the deterioration of the Lagrangian at (y^{k+1}, λ^{k+1}) should be smaller than a multiple of the infeasibility $\|h(x^k)\|$. If the Lagrange multipliers estimates are bounded, and under suitable Lipschitz assumptions, condition (22) is implied by $\|y^{k+1} - x^k\| = O(\|h(x^k)\|)$. This means that the distance between x^k and the restored point y^{k+1} should be proportional to the infeasibility measure at x^k. Providing suitable safeguarding parameters β is even harder than in the case of (21), because (22) is scale-dependent. In practical terms, Assumption 3.2 says that we believe that the restoration algorithm employed is reasonable enough so that sufficient improvement and bounded-distance requirements are automatically satisfied.

Assumption 3.3 *For all $k \in \mathbb{N}$, we choose d^k such that $y^{k+1} + d^k \in \Omega$ and $\nabla h(y^{k+1})^T d^k = 0$. Moreover, we assume that there exists $\sigma > 0$ such that*

$$\nabla L(y^{k+1}, \lambda^{k+1})^T d^k \leq -\sigma \|d^k\|^2 \qquad (24)$$

and

$$\sigma \| P_k(y^{k+1} - \nabla L(y^{k+1}, \lambda^{k+1})) - y^{k+1} \| \leq \|d^k\| \qquad (25)$$

for all $k \in \mathbb{N}$, where P_k denotes the Euclidean projection on the polytope defined by $y \in \Omega$ and $\nabla h(y^{k+1})^T (y - y^{k+1}) = 0$. As we will see in Sect. 4, there is no loss of generality in the use of the same σ in (24) and (25).

In practice, the direction d^k will be obtained as the solution of the following problem:

$$\min \frac{1}{2} d^T H_k d + \nabla L(y^{k+1}, \lambda^{k+1})^T d \qquad (26)$$

$$\text{s.t. } \nabla h(y^{k+1})^T d = 0, \quad y^{k+1} + d \in \Omega, \qquad (27)$$

where H_k is an approximation to the Hessian of the Lagrangian. Since Ω is a polytope, we have that (26)–(27) is a quadratic programming problem. Note that solving (26)–(27) is equivalent to minimize $\frac{1}{2} d^t H_k d + \nabla f(y^{k+1})^T d$ subject to (27).

Note that the requirement (24) is not necessarily satisfied by the solution of (26)–(27), unless we impose some additional conditions on H_k. We will show in Sect. 4 that Assumption 3.3 holds if the eigenvalues of the matrix of $Z_k^T H_k Z_k$ lie in a positive interval $[\sigma_{\min}, \sigma_{\max}]$, where the columns of Z_k form an orthonormal basis of the null-space of $\nabla h(y^{k+1})^T$. In implementations, we will define H_k as the Hessian of the approximate Lagrangian, testing further the descent condition and switching to a safe positive definite matrix, if necessary.

Assumption 3.4 *There exists $\gamma > 0$ such that, for all $k \in \mathbb{N}$,*

$$L(y^{k+1} + t_k d^k, \lambda^{k+1}) \leq L(y^{k+1}, \lambda^{k+1}) - \gamma t_k \|d^k\|^2. \qquad (28)$$

Note that (28) states that the Lagrangian, whose quadratic approximation is minimized at (26)–(27), should decrease along the direction d^k. The sufficient decrease condition, which depends on γ, holds for small enough t_k under the choice (26)–(27) with safeguards on the eigenvalues of H. As we saw in Theorem 2.1, condition (10) also holds for small enough t_k. Therefore, Assumption 3.4 suggests a safeguarding backtracking procedure that aims to satisfy, simultaneously, (10) and (28).

Assumption 3.5 *The Lagrange multiplier estimates $\{\lambda^k\}$ lie in a compact set.*

By Assumption 3.1, we have that $\nabla L(\cdot, \lambda) : \Omega \rightarrow \mathbb{R}^n$ is Lipschitz continuous in Ω for all fixed $\lambda \in \mathbb{R}^m$. By Assumption 3.5, we can assume that the Lipschitz constant for $\nabla L(\cdot, \lambda^k)$ is the same for all $k \in \mathbb{N}$. Without any loss of generality, we will also denote by η this Lipschitz constant, that is, for all $x, y \in \Omega$ and $k \in \mathbb{N}$,

$$\|\nabla L(y, \lambda^k) - \nabla L(x, \lambda^k)\| \leq \eta \|y - x\|. \qquad (29)$$

Assumption 3.6 *The Lagrange multiplier estimates* $\{\lambda^k\}$ *are such that*

$$\lim_{k\to\infty} P_\Omega(y^k - \nabla f(y^k) - \nabla h(y^k)\lambda^{k+1}) - y^k = 0. \tag{30}$$

As we will see in Sect. 4, this assumption can be fulfilled, if we take λ^{k+1} as the Lagrange multipliers associated with $\nabla h(y^k)$ in the KKT conditions of the problem (26)–(27).

In next lemma, we will prove that the sequence of the penalty parameters is bounded away from zero. This means that the objective function always has a significant weight in the merit criterion.

Lemma 3.1 *Suppose that Assumption 3.2 holds. The sequence* $\{\theta_k\}$ *is non-increasing and bounded away from zero.*

Proof By direct calculations, (8) is equivalent to

$$\theta\big[L\big(y^{k+1}, \lambda^{k+1}\big) - L\big(x^k, \lambda^k\big) + \|h(x^k)\| - \|h\big(y^{k+1}\big)\|\big]$$
$$\leq \frac{(1 + s_k)}{2}\big(\|h(x^k)\| - \|h(y^{k+1})\|\big). \tag{31}$$

If $\big[L\big(y^{k+1}, \lambda^{k+1}\big) - L\big(x^k, \lambda^k\big) + \|h(x^k)\| - \|h\big(y^{k+1}\big)\|\big] \leq 0$, then (31) holds for all $\theta \geq 0$, thus $\theta_k = \theta_{k-1}$.

If $[L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) + \|h(x^k)\| - \|h(y^{k+1})\|] > 0$, then $y^{k+1} \neq x^k$ and, consequently,

$$\frac{(1 + s_k)\big(\|h(x^k)\| - \|h(y^{k+1})\|\big)}{2\big[L(y^{k+1}, \lambda^{k+1}) - L\big(x^k, \lambda^k\big) + \|h(x^k)\| - \|h\big(y^{k+1}\big)\|\big]} > 0.$$

In this case, we have that

$$\theta_k = \min\left\{\theta_{k-1}, \frac{(1 + s_k)(\|h(x^k)\| - \|h(y^{k+1})\|)}{2\big[L\big(y^{k+1}, \lambda^{k+1}\big) - L\big(x^k, \lambda^k\big) + \|h(x^k)\| - \|h\big(y^{k+1}\big)\|\big]}\right\}. \tag{32}$$

By the updating rule of the penalty parameter, we have that the sequence $\{\theta_k\}$ is non-increasing. It remains to be proved that $\{\theta_k\}$ is bounded away from zero. For this purpose, it is sufficient to show that θ_k is greater than a fixed positive number when $\theta_k \neq \theta_{k-1}$. In this case, we have

$$\frac{1}{\theta_k} = \frac{2\big[L\big(y^{k+1}, \lambda^{k+1}\big) - L\big(x^k, \lambda^k\big) + \|h(x^k)\| - \|h\big(y^{k+1}\big)\|\big]}{(1 + s_k)\big(\|h(x^k)\| - \|h(y^{k+1})\|\big)}$$
$$= \frac{2}{1 + s_k}\left[\frac{L\big(y^{k+1}, \lambda^{k+1}\big) - L\big(x^k, \lambda^k\big)}{\|h(x^k)\| - \|h(y^{k+1})\|} + 1\right].$$

Thus, by Assumption 3.2, (7), and (23),

$$\frac{1}{\theta_k} \le 2\left[\frac{\beta \|h(x^k)\|}{(1-r)\|h(x^k)\|} + 1\right] = 2\left[\frac{\beta}{1-r} + 1\right]. \tag{33}$$

This implies that the sequence $\{1/\theta_k\}$ is bounded. Therefore, the sequence $\{\theta_k\}$ is non-increasing and bounded away from zero. □

The following lemma ensures that any limit point of the sequence generated by Algorithm 2.1 is feasible.

Lemma 3.2 *Suppose that Assumptions 3.2 and 3.5 hold. Then, the sum $\sum_{k=0}^{\infty} \|h(x^k)\|$ is convergent.*

Proof By condition (10) and Assumption 3.2, for all $k \in \mathbb{N}$ one has that

$$\Phi(x^{k+1}, \lambda^{k+1}, \theta_k) \le \Phi(x^k, \lambda^k, \theta_k) + \frac{1-r}{2}\left(\|h(y^{k+1})\| - \|h(x^k)\|\right).$$

Therefore, by (23),

$$\Phi(x^{k+1}, \lambda^{k+1}, \theta_k) \le \Phi(x^k, \lambda^k, \theta_k) - \frac{(1-r)^2}{2}\|h(x^k)\|. \tag{34}$$

Let us define $\rho_k := (1-\theta_k)/\theta_k$ for all $k \in \{-1, 0, 1, 2 \ldots\}$. By Lemma 3.1, there exists $\bar{\theta} > 0$ such that $\theta_k \ge \bar{\theta}$ for all $k \in \mathbb{N}$. This implies that $\rho_k \le 1/\bar{\theta} - 1$ for all $k \in \mathbb{N}$. Since $\{\rho_k\}$ is bounded and non-decreasing, it follows that

$$\sum_{k=0}^{\infty} (\rho_k - \rho_{k-1}) = \lim_{k \to \infty} \rho_k - \rho_{-1} < \infty. \tag{35}$$

By compactness, the sequence $\{\|h(x^k)\|\}$ is bounded. Therefore, by (35), there exists $c > 0$ such that

$$\sum_{k=0}^{\infty} (\rho_k - \rho_{k-1})\|h(x^k)\| \le c < \infty. \tag{36}$$

Now, by (34),

$$L(x^{k+1}, \lambda^{k+1}) + \frac{1-\theta_k}{\theta_k}\|h(x^{k+1})\| \le L(x^k, \lambda^k) + \frac{1-\theta_k}{\theta_k}\|h(x^k)\|$$
$$- \frac{(1-r)^2}{2\theta_k}\|h(x^k)\|.$$

Since $0 < \theta_k < 1$, we have that $\frac{(1-r)^2}{2} < \frac{(1-r)^2}{2\theta_k}$, so

$$L(x^{k+1}, \lambda^{k+1}) + \rho_k\|h(x^{k+1})\| \le L(x^k, \lambda^k) + \rho_k\|h(x^k)\| - \frac{(1-r)^2}{2}\|h(x^k)\|.$$

Therefore, for all $k \in \mathbb{N}$,

$$L(x^{k+1}, \lambda^{k+1}) + \rho_k \|h(x^{k+1})\| \le L(x^k, \lambda^k) + \rho_{k-1} \|h(x^k)\|$$
$$+ (\rho_k - \rho_{k-1}) \|h(x^k)\| - \frac{(1-r)^2}{2} \|h(x^k)\|.$$

Thus, for all $k \in \mathbb{N}$, we have

$$L(x^{k+1}, \lambda^{k+1}) + \rho_k \|h(x^{k+1})\| \le L(x^0, \lambda^0) + \rho_{-1} \|h(x^0)\| + \sum_{j=0}^{k} (\rho_j - \rho_{j-1}) \|h(x^j)\|$$
$$- \frac{(1-r)^2}{2} \sum_{j=0}^{k} \|h(x^j)\|.$$

Therefore, by (36),

$$L(x^{k+1}, \lambda^{k+1}) + \rho_k \|h(x^{k+1})\| \le L(x^0, \lambda^0) + \rho_{-1} \|h(x^0)\|$$
$$+ c - \frac{(1-r)^2}{2} \sum_{j=0}^{k} \|h(x^j)\|.$$

Thus,

$$\frac{(1-r)^2}{2} \sum_{j=0}^{k} \|h(x^j)\| \le -\left[L(x^{k+1}, \lambda^{k+1}) + \rho_k \|h(x^{k+1})\| \right] + L(x^0, \lambda^0)$$
$$+ \rho_{-1} \|h(x^0)\| + c.$$

Since the functions L and h are continuous, by Assumption 3.5 and the compactness of Ω, we have that the sequences $\{L(x^k, \lambda^k)\}$ and $\{\|h(x^k)\|\}$ are bounded. Since $\{\rho_k\}$ is also bounded, it follows that the series $\sum_{k=0}^{\infty} \|h(x^k)\|$ is convergent. □

The following lemma shows that, if the problem is sufficiently smooth, then the direction on the tangent subspace ensures a uniform- bounded deterioration of the feasibility. The lemma also guarantees that the step size will not tend to zero. This means that Algorithm 2.1 will not produce excessively small steps.

Lemma 3.3 *Suppose that Assumptions 3.1–3.3 and 3.5 hold. Define c_k as in (5) and t_k as in Step 4.3 of Algorithm 2.1. Then, there exist $c > 0$ and $\bar{t} > 0$ such that $c_k > c$ and $t_k \ge \bar{t}$ for all $k \in \mathbb{N}$.*

Proof Given any continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^p$, by the Fundamental Theorem of Calculus, we have that

$$F(y + td) = F(y) + t\nabla F(y)^T d + t \int_0^1 (\nabla F(y + t\xi d) - \nabla F(y))^T d \, d\xi. \quad (37)$$

Applying (37) with $F(y + td) = h(y^{k+1} + td^k)$, by Assumptions 3.1 and 3.2, the triangle inequality, and the Cauchy-Schwarz inequality, we have that, for $c = \frac{\eta}{2}$,

$$\|h(y^{k+1} + td^k)\| \leq \|h(y^{k+1})\| + t \int_0^1 \eta \|y^{k+1} + t\xi d^k - y^{k+1}\| \|d^k\| \, d\xi$$

$$\leq \|h(y^{k+1})\| + ct^2 \|d^k\|^2. \tag{38}$$

Thus (5) holds for $c_k = \frac{\eta}{2}$, independently of k and for all $t \geq 0$.

Now, applying (37) with $F(y + td) = L(y^{k+1} + td^k, \lambda^k)$, by Assumptions 3.1, 3.2, 3.5, and (29),

$$L(y^{k+1} + td^k, \lambda^{k+1}) - L(y^{k+1}, \lambda^{k+1}) = t \nabla L(y^{k+1}, \lambda^{k+1})^T d^k$$

$$+ t \int_0^1 \left(\nabla L(y^{k+1} + t\xi d^k, \lambda^{k+1}) - \nabla L(y^{k+1}, \lambda^{k+1}) \right)^T d^k \, d\xi$$

$$\leq -\sigma t \|d^k\|^2 + t \int_0^1 \eta \|y^{k+1} + t\xi d^k - y^{k+1}\| \|d^k\| \, d\xi$$

$$\leq -\left(\sigma - \frac{\eta t}{2} \right) t \|d^k\|^2.$$

Therefore, if $\gamma_1 \leq \frac{\sigma}{2}$ and $\tau = \frac{\sigma}{2\eta}$, we have that

$$L(y^{k+1} + td^k, \lambda^{k+1}) \leq L(y^{k+1}, \lambda^{k+1}) - \gamma_1 t \|d^k\|^2, \tag{39}$$

for all $t \leq \tau$.

If $\|d^k\| \neq 0$, condition (39) ensures that (9) holds whenever $t_k \leq \tau$. Let us prove that we can also satisfy (10) for any t_k in a specific interval. In fact, we will prove that any $t_k \leq \min \left\{ \frac{\bar{\theta} \gamma_1}{c(1-\bar{\theta})}, \tau \right\}$, where $\bar{\theta}$ is defined in Lemma 3.2, is acceptable in Step 4.3 of Algorithm 2.1. Since $r_k \geq 0$, by (8), (19), (38), (39), and Lemma 3.1, we have that

$$\Phi(y^{k+1} + td^k, \lambda^{k+1}, \theta_k) - \Phi(x^k, \lambda^k, \theta_k)$$

$$= \Phi(y^{k+1} + td^k, \lambda^{k+1}, \theta_k) - \Phi(y^{k+1}, \lambda^{k+1}, \theta_k) + \Phi(y^{k+1}, \lambda^{k+1}, \theta_k)$$

$$- \Phi(x^k, \lambda^k, \theta_k)$$

$$\leq \theta_k \left(L(y^{k+1} + td^k, \lambda^{k+1}) - L(y^{k+1}, \lambda^{k+1}) \right)$$

$$+ (1 - \theta_k) \left(\|h(y^{k+1} + td^k)\| - \|h(y^{k+1})\| \right) + \frac{1 - s_k}{2} \left(\|h(y^{k+1})\| - \|h(x^k)\| \right)$$

$$\leq -\theta_k \gamma_1 t \|d^k\|^2 + (1 - \theta_k) c t^2 \|d^k\|^2 + \frac{1 - r_k}{2} \left(\|h(y^{k+1})\| - \|h(x^k)\| \right)$$

$$\leq -\bar{\theta} \gamma_1 t \|d^k\|^2 + (1 - \bar{\theta}) c t^2 \|d^k\|^2 + \frac{1 - r_k}{2} \left(\|h(y^{k+1})\| - \|h(x^k)\| \right)$$

$$\leq \left((1 - \bar{\theta}) c t - \bar{\theta} \gamma_1 \right) t \|d^k\|^2 + \frac{1 - r_k}{2} \left(\|h(y^{k+1})\| - \|h(x^k)\| \right)$$

$$\leq \frac{1 - r_k}{2} \left(\|h(y^{k+1})\| - \|h(x^k)\| \right).$$

Thus, due to the backtracking procedure, we have that $t_k \geq \bar{t} := \min \left\{ \frac{\bar{\theta} \gamma_1}{10 c (1 - \bar{\theta})}, \frac{\tau}{10}, 1 \right\}$ for all $k \in \mathbb{N}$. □

Lemma 3.4 *Suppose that Assumptions 3.1–3.5 hold. Then* $\lim_{k \to \infty} \|d^k\| = 0$.

Proof By Lemma 3.3, there exists $\bar{t} > 0$ such that $t_k \geq \bar{t}$ for all $k \in \mathbb{N}$. By (22) and (28),

$$L(x^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) = L(x^{k+1}, \lambda^{k+1}) - L(y^{k+1}, \lambda^{k+1})$$
$$+ L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k)$$
$$\leq -\gamma t_k \|d^k\|^2 + \beta \|h(x^k)\| \leq -\gamma \bar{t} \|d^k\|^2 + \beta \|h(x^k)\|.$$

By Lemma 3.2 there exists c such that $\sum_{k=0}^{\infty} \|h(x^k)\| = c$. Thus

$$L(x^{l+1}, \lambda^{l+1}) - L(x^0, \lambda^0) = \sum_{k=0}^{l} \left(L(x^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) \right) \leq -\gamma \bar{t} \sum_{k=0}^{l} \|d^k\|^2$$

$$+ \beta \sum_{k=0}^{l} \|h(x^k)\| \leq -\gamma \bar{t} \sum_{k=0}^{l} \|d^k\|^2 + \beta c.$$

Since f and h are bounded below on Ω, and $\{\lambda^{k+1}\}$ remains in a compact set, we have that the series $\sum_{k=0}^{\infty} \|d^k\|^2$ is convergent, and thus, $\{\|d^k\|\}$ converges to zero. □

Next Lemma and its corollary are used to prove the optimality of every limit point of the sequence generated by Algorithm 2.1.

Lemma 3.5 *Consider P_k as in Assumption 3.3. Given $y^{k+1} \in \Omega$, then*

$$P_k \left(y^{k+1} - \nabla L(y^{k+1}, \lambda) \right) = P_k \left(y^{k+1} - \nabla L(y^{k+1}, \mu) \right),$$

for all $\lambda, \mu \in \mathbb{R}^m$.

Proof Defining $S_k := \{d \ : \ \nabla h(y^{k+1})^T d = 0\}$, we have that $\nabla h(y^{k+1})(\lambda - \mu) \in S_k^{\perp}$. Since $\nabla L(y^{k+1}, \lambda) = \nabla L(y^{k+1}, \mu) + \nabla h(y^{k+1})(\lambda - \mu)$, the result follows from the Projection Theorem (see [39], Proposition B.11, item (b), p 704). □

Corollary 3.1 *Let x^* be a feasible point of ([1](#)) such that*

$$\lim_{k \to \infty} y^{k+1} = x^* \text{ and } \lim_{k \to \infty} P_k(y^{k+1} - \nabla L(y^{k+1}, \lambda^{k+1})) - y^{k+1} = 0,$$

for some $\{y^{k+1}\} \subset \Omega$ and $\{\lambda^{k+1}\} \subset \mathbb{R}^m$. Then x^* satisfies the L-AGP optimality condition [25].

Proof By hypothesis, x^* is feasible. Moreover, by Lemma [3.5](#),

$$\lim_{k \to \infty} P_k(y^{k+1} - \nabla f(y^{k+1})) - y^{k+1} = \lim_{k \to \infty} P_k(y^{k+1} - \nabla L(y^{k+1}, 0)) - y^{k+1}$$
$$= \lim_{k \to \infty} P_k(y^{k+1} - \nabla L(y^{k+1}, \lambda^{k+1})) - y^{k+1} = 0.$$

Thus, x^* satisfies the L-AGP optimality condition.   □

Next lemma will be used to prove the convergence of the Lagrange multiplier estimates $\{\lambda^k\}$.

Lemma 3.6 *Let x^* be a feasible point such that*

$$\lim_{k \to \infty} y^k = x^* \text{ and } \lim_{k \to \infty} P_\Omega(y^k - \nabla L(y^k, \lambda^{k+1})) - y^k = 0,$$

for some $\{y^k\} \subset \Omega$ and $\{\lambda^{k+1}\} \subset \mathbb{R}^m$. Then, if x^* satisfies the Mangasarian–Fromovitz constraint qualification, we have that the KKT conditions hold in x^*. Moreover, all the limit points of $\{\lambda^{k+1}\}$ are Lagrange multipliers associated with the equality constraints in x^*.

Proof Let $z^k := P_\Omega(y^k - \nabla L(y^k, \lambda^{k+1}))$; hence $\lim_{k \to \infty}(z^k - y^k) = 0$ and z^k is the solution of the problem:

$$\min \frac{1}{2} \|z - (y^k - \nabla L(y^k, \lambda^{k+1}))\|^2 \text{ s.t. } z \in \Omega.$$

Define $\Omega := \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$. Then, for all $k = 0, 1, 2, \ldots$, there exist α^k and β^k such that

$$\beta^k \geq 0, \quad (z^k)^T \beta^k = 0, \tag{40}$$

and

$$z^k - y^k + \nabla L(y^k, \lambda^{k+1}) + A^T \alpha^k - \beta^k = 0. \tag{41}$$

Defining $M_k := \max\{\|\lambda^{k+1}\|_\infty, \|\alpha^k\|_\infty, \|\beta^k\|_\infty\}$, we have that $\{M_k\}$ is bounded. Otherwise, we could divide ([41](#)) by M_k and take the limit on a suitable subsequence to obtain a contradiction with the Mangasarian-Fromovitz condition.

Since $\{M_k\}$ is bounded, if λ^* is a limit point of $\{\lambda^{k+1}\}$, there exists $K \subseteq \mathbb{N}$ such that $\lim_{k \in K} \lambda^{k+1} = \lambda^*$, $\lim_{k \in K} \alpha^k = \alpha^*$, and $\lim_{k \in K} \beta^k = \beta^*$. Taking limits for $k \in K$ in ([40](#)) and ([41](#)), we verify that λ^* is a Lagrange multiplier associated with the constraints $h(x) = 0$ in x^*.   □

The following theorem summarizes the convergence properties of Algorithm 2.1 under Assumptions 3.1–3.6.

Theorem 3.1 *Suppose that Assumptions 3.1–3.6 hold. Then*

1. *For all $k \in \mathbb{N}$, x^k is well defined.*
2. *There exists $\bar{\theta} > 0$ such that $\theta_k \geq \bar{\theta}$ for all $k \in \mathbb{N}$.*
3. *$\lim_{k \to \infty} \|h(x^k)\| = \lim_{k \to \infty} \|h(y^k)\| = 0$ and any cluster point of $\{x^k\}$ or $\{y^k\}$ is feasible.*
4. *$\lim_{k \to \infty} d^k = 0$.*
5. *$\lim_{k \to \infty} \|y^k - x^k\| = 0$.*
6. *The sequences $\{x^k\}$ and $\{y^k\}$ admit the same cluster points.*
7. *Limit points of $\{x^k\}$ satisfy the L-AGP optimality condition.*
8. *If a limit point x^* satisfies the Constant Positive Generators (CPG) constraint qualification [40], then the KKT conditions hold at x^*.*
9. *If a limit point x^* satisfies the Mangasarian–Fromovitz constraint qualification, then the sequence $\{\lambda^k\}$ admits a limit point λ^*, which is a Lagrange multiplier associated with $\nabla h(x^*)$.*

Proof Well-definedness of x^k has been proved in Theorem 2.1. The existence of $\bar{\theta}$ follows from Lemma 3.1.

By Lemma 3.2 and (3),

$$\lim_{k \to \infty} \|h(y^{k+1})\| \leq \lim_{k \to \infty} \|h(x^k)\| = 0,$$

and thus, by the continuity of h, any cluster point of $\{x^k\}$ or $\{y^k\}$ is feasible.

The fact that $\lim_{k \to \infty} d^k = 0$ follows from Lemma 3.4.

By item 4 and the triangle inequality, we have that

$$\lim_{k \to \infty} \|y^{k+1} - x^{k+1}\| = \lim_{k \to \infty} \|y^{k+1} - (y^{k+1} + t_k d^k)\| \leq \lim_{k \to \infty} \|d^k\| = 0.$$

By item 5 and the triangle inequality,

$$\lim_{k \in K \subseteq \mathbb{N}} \|x^k - z\| = 0 \Leftrightarrow \lim_{k \in K \subseteq \mathbb{N}} \|y^k - z\| = 0.$$

Thus, the sequences $\{x^k\}$ and $\{y^k\}$ admit the same cluster points.

Let x^* be a limit point of $\{x^k\}$, therefore, by item 6, there is $K \subseteq \mathbb{N}$ such that $\lim_{k \in K} y^{k+1} = x^*$. By Assumption 3.3 and item 4,

$$\lim_{k \in K} \| P_k(y^{k+1} - \nabla L(y^{k+1}, \lambda^{k+1})) - y^{k+1} \| = 0.$$

By item 3, we have that x^* is feasible. Thus we conclude, by Corollary 3.1, that x^* satisfies the L-AGP optimality condition.

The statement 8 is a consequence of item 7 and the fact that sequential optimality conditions such as L-AGP imply KKT under the CPG constraint qualification [25,40].

Finally, suppose that $\lim_{k \in K} y^k = x^*$ and the Mangasarian–Fromovitz constraint qualification holds in x^*. Since the sequence $\{\lambda^{k+1}\}$ lies in a compact set, we have that this sequence admits at least one limit point. Then, by Lemma 3.6, we conclude that any limit point of $\{\lambda^{k+1}\}$ with $k \in K$ is a Lagrange Multiplier associated with $\nabla h(x^*)$.   □

Remark 3.1 The CPG constraint qualification is weaker than CRSC [40], (R)CPLD [41–43], (R)CRCQ [44], and Mangasarian–Fromovitz constraint qualification.

4 Plausibility of Assumptions

One of our main assumptions is that, given an infeasible point x^k, a less infeasible point can always be computed at the Restoration phase. Clearly, this assumption is not always true. For example, if the feasible set is empty and x^k is a minimizer of the infeasibility, then it would not be possible to be successful at the Restoration phase.

However, if a good algorithm is used at the Restoration phase, we judge that the improvement in feasibility will be achieved, or the problem is probably infeasible.

Lipschitz conditions on the gradients of f and h, assumed in Assumption 3.1, are standard in smooth optimization. Convexity conditions on the problem (26)–(27) are also usual in quadratic programming analyses. After minimizing $\frac{1}{2} d^T H_{k-1} d + \nabla f(y^k)^T d$ subject to $y^k + d \in \Omega$, $\nabla h(y^k)^T d = 0$, a natural way to estimate the Lagrange multipliers is taking λ^{k+1} as the Lagrange multipliers associated with the constraints $\nabla h(y^k)^T d = 0$. In this case, it is plausible to assume that λ^{k+1} lies in a compact set if the Mangasarian–Fromovitz constraint qualification holds at y^k. Thus, if the sequence $\{\|H_k d^k\|\}$ is uniformly bounded, we have that Assumption 3.5 holds under the Mangasarian–Fromovitz constraint qualification on Ω. An easy alternative to guarantee that Assumption 3.5 holds is to use safeguards on the Lagrange multipliers in the implementation.

Assumption 3.2 asks for a sufficient improvement of feasibility with a bounded deterioration in the Lagrangian. This assumption can be satisfied with a typical Lipschitz hypothesis and regularity conditions on the feasible set. In Lemma 2 of [19], it is proved that if the Mangasarian Fromovitz-constraint qualification holds at all points of Ω, then the closest feasible point to x^k fulfills the bounded deterioration for the objective function f, and so any $r_k > 0$ is a possible choice for all k. A practical restoration procedure is shown in Lemma 6.1 of [11], if the Linear Independence constraint qualification holds on Ω. This result asserts that, given $r \in \,]0, 1[$, there is a neighborhood N of the feasible set such that, if $x^k \in N$, then the linearization of the constraints at x_k, $T_k := \{y \in \Omega : h(x^k) + \nabla h(x^k)^T (y - x^k) = 0\}$, is not empty, and denoting by y^{k+1} the point of T_k closest to x^k, we have that $h(y^{k+1}) \leq r h(x^k)$. Moreover, under Lipschitz conditions, bounded deterioration for the objective function holds. In both restoration procedures, if the Lagrange multipliers λ^k remain in a compact set (Assumption 3.5), then the bounded deterioration also holds for the Lagrangian.

The plausibility of Assumption 3.3 is strictly linked with our proposal in the present paper of choosing d^k as the solution of the quadratic programming problem:

$$\min \nabla L\big(y^{k+1}, \lambda^{k+1}\big)^T d + \frac{1}{2} d^T H_k d \ \text{ s.t. } y^{k+1} + d \in \Omega, \ \nabla h\big(y^{k+1}\big)^T d = 0, \quad (42)$$

where H_k is symmetric, $Z_k^T H_k Z_k$ is positive definite, and the columns of Z_k form an orthonormal basis of the null-space of $\nabla h(y^{k+1})^T$. In addition, we assume that the eigenvalues of $Z_k^T H_k Z_k$ lie in a positive interval $[\sigma_{\min}, \sigma_{\max}]$.

By selecting d^k in this way, since $d = 0$ is feasible to (42), we have that

$$\nabla L\big(y^{k+1}, \lambda^{k+1}\big)^T d^k + \frac{1}{2} (d^k)^T H_k d^k \leq 0.$$

Therefore,

$$\nabla L\big(y^{k+1}, \lambda^{k+1}\big)^T d^k \leq -\frac{\sigma_{\min}}{2} \|d^k\|^2.$$

In order to prove condition (25), let us define

$$D_k := \{d \in \mathbb{R}^n \ : \ \nabla h\big(y^{k+1}\big)^T d = 0 \text{ and } y^{k+1} + d \in \Omega\}. \quad (43)$$

Since d^k is the solution of the linearly constrained problem (42),

$$P_{D_k}\big(d^k - H_k d^k - \nabla L\big(y^{k+1}, \lambda^{k+1}\big)\big) - d^k = 0.$$

Changing variables ($y = y^{k+1} + d$) and using the fact that the projections are non-expansive, we have

$$
\begin{aligned}
\|P_k\big(y^{k+1} &- \nabla L\big(y^{k+1}, \lambda^{k+1}\big)\big) - y^{k+1}\| = \|P_{D_k}\big(-\nabla L\big(y^{k+1}, \lambda^{k+1}\big)\big)\| \\
&= \|P_{D_k}\big(-\nabla L\big(y^{k+1}, \lambda^{k+1}\big)\big) - P_{D_k}\big(d^k - H_k d^k - \nabla L\big(y^{k+1}, \lambda^{k+1}\big)\big) \\
&\quad + P_{D_k}\big(d^k - H_k d^k - \nabla L\big(y^{k+1}, \lambda^{k+1}\big)\big)\| \leq \|P_{D_k}\big(-\nabla L\big(y^{k+1}, \lambda^{k+1}\big)\big) \\
&\quad - P_{D_k}\big(d^k - H_k d^k - \nabla L\big(y^{k+1}, \lambda^{k+1}\big)\big)\| + \|P_{D_k}\big(d^k - H_k d^k - \nabla L\big(y^{k+1}, \lambda^{k+1}\big)\big)\| \\
&\leq \|d^k - H_k d^k\| + \|d^k\| \leq (2 + \sigma_{\max}) \|d^k\|.
\end{aligned}
$$

So, defining $\sigma := \min\left\{\frac{\sigma_{\min}}{2}, \frac{1}{2+\sigma_{\max}}\right\}$, we have that Assumption 3.3 holds.

From Lemma 3.3, we can see that condition (39) holds for $\gamma_1 = \frac{\sigma_{\min}}{4}$ and sufficiently small t. Thus, if $\sigma_{\min}$ is known, then we can ensure that Assumption 3.4 holds testing the sufficient decrease at Step 5 of Algorithm 2.1. Moreover, even if $\sigma_{\min}$ is not known, it is possible to do a double backtracking, in t and σ, to ensure that Assumption 3.4 holds [45]. However, this does not seem to be either efficient or necessary in practice, so we only test simple decrease in Step 5 of Algorithm 2.1.

Finally, to prove the plausibility of Assumption 3.6, we consider that λ^{k+1} is chosen as the vector of Lagrange multipliers associated with the constraints $\nabla h(y^k)^T d = 0$ in problem:

$$\min \frac{1}{2} d^T H_{k-1} d + \nabla f(y^k)^T d \ \text{ s.t. } \ y^k + d \in \Omega, \ \nabla h(y^k)^T d = 0.$$

In this case, denoting $D_k := \{d \ : \ y^k + d \in \Omega\}$, we have that

$$P_{D_k}\left(d^{k-1} - H_{k-1}d^{k-1} - \nabla f\left(y^k\right) - \nabla h\left(y^k\right)\lambda^{k+1}\right) - d^{k-1} = 0.$$

Therefore,

$$
\begin{aligned}
\|P_\Omega\left(y^k - \nabla f(y^k) - \nabla h(y^k)\lambda^{k+1}\right) - y^k\| &= \|P_{D_k}\left(-\nabla f\left(y^k\right) - \nabla h(y^k)\lambda^{k+1}\right)\| \\
&\leq \|P_{D_k}\left(-\nabla f\left(y^k\right) - \nabla h(y^k)\lambda^{k+1}\right) - P_{D_k}\left(d^{k-1} - H_{k-1}d^{k-1} - \nabla f(y^k)\right. \\
&\quad \left. -\nabla h(y^k)\lambda^{k+1}\right)\| \\
&\quad + \|P_{D_k}\left(d^{k-1} - H_{k-1}d^{k-1} - \nabla f(y^k) - \nabla h(y^k)\lambda^{k+1}\right)\| \\
&\leq \|d^{k-1} - H_{k-1}d^{k-1}\| + \|d^{k-1}\|.
\end{aligned}
$$

Thus, considering that the sequence $\{\|H_k\|\}$ is uniformly bounded and that Assumptions 3.1–3.5 hold, by Lemma 3.4, we have that

$$\lim_{k \to \infty} P_\Omega(y^k - \nabla f(y^k) - \nabla h(y^k)\lambda^{k+1}) - y^k = 0,$$

which means that Assumption 3.6 holds.

5 Conclusions

The Inexact-Restoration method introduced in this paper employs an improved version of the Fischer–Friedlander line search approach combined with the use of Lagrange multipliers estimates both in the Optimization Phase subproblem as in the merit function. The new method has enhancements in several theoretical aspects and it is appropriate to tackle many problems in which there is a natural way to restore feasibility.

It remains to be investigated its application to general constrained optimization problems. In the general case, some standard (perhaps under-determined Newtonian) method should be defined for the restoration phase and convergence should be proved taking into account the characteristics of that method.

Other possible approach for general problems could be to introduce objective function information in the restoration phase, aiming to provide better initial approximations for the second-phase subproblems. Much theoretical and practical research can be expected along these lines.

Future research will also include the application of alternative I-R approaches to multiobjective optimization.

References

1. Abadie, J., Carpentier, J.: Generalization of the Wolfe reduced-gradient method to the case of nonlinear constraints. In: Fletcher, R. (ed.) Optimization, pp. 37–47. Academic Press, New York (1968)
2. Lasdon, L.S.: Reduced gradient methods. In: Powell, M.J.D. (ed.) Nonlinear Optimization, pp. 235–242. Academic Press, New York (1982)
3. Miele, A., Huang, H.Y., Heideman, J.C.: Sequential gradient-restoration algorithm for the minimization of constrained functions, ordinary and conjugate gradient version. J. Optim. Theory Appl. **4**, 213–246 (1969)
4. Miele, A., Levy, A.V., Cragg, E.E.: Modifications and extensions of the conjugate-gradient restoration algorithm for mathematical programming problems. J. Optim. Theory Appl. **7**, 450–472 (1971)
5. Miele, A., Sims, E.M., Basapur, V.K.: Sequential gradient-restoration algorithm for mathematical programming problem with inequality constraints, Part 1, Theory. Aero-Astronautics Report No. 168, Rice University (1983).
6. Rom, M., Avriel, M.: Properties of the sequential gradient-restoration algorithm (SGRA), Part 1: introduction and comparison with related methods. J. Optim. Theory Appl. **62**, 77–98 (1989)
7. Rom, M., Avriel, M.: Properties of the sequential gradient-restoration algorithm (SGRA), Part 2: convergence analysis. J. Optim. Theory Appl. **62**, 99–126 (1989)
8. Rosen, J.B.: The gradient projection method for nonlinear programming, Part 1: linear constraints. SIAM J. Appl. Math. **8**, 181–217 (1960)
9. Rosen, J.B.: The gradient projection method for nonlinear programming, Part 2: nonlinear constraints. SIAM J. Appl. Math. **9**, 514–532 (1961)
10. Martínez, J.M., Pilotta, E.A.: Inexact restoration algorithms for constrained optimization. J. Optim. Theory Appl. **104**, 135–163 (2000)
11. Martínez, J.M.: Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. J. Optim. Theory Appl. **111**, 39–58 (2001)
12. Gonzaga, C.C., Karas, E.W., Vanti, M.: A globally convergent filter method for nonlinear programming. SIAM J. Optim. **14**, 646–669 (2003)
13. Birgin, E.G., Martínez, J.M.: Local convergence of an Inexact-Restoration method and numerical experiments. J. Optim. Theory Appl. **127**, 229–247 (2005)
14. Martínez, J.M., Pilotta, E.A.: Inexact restoration methods for nonlinear programming: advances and perspectives. In: Qi, L.Q., Teo, K., Yang, X.Q. (eds.) Optimization and Control with Applications, pp. 271–292. Springer, New York (2005)
15. Kaya, C.Y., Martínez, J.M.: Euler discretization and inexact restoration for optimal control. J. Optim. Theory Appl. **134**, 191–206 (2007)
16. Andreani, R., Castro, S.L., Chela, J.L., Friedlander, A., Santos, S.A.: An inexact-restoration method for nonlinear bilevel programming problems. Comput. Optim. Appl. **43**, 307–328 (2009)
17. Gomes-Ruggiero, M.A., Martínez, J.M., Santos, S.A.: Spectral projected gradient method with inexact restoration for minimization with nonconvex constraints. SIAM J. Sci. Comput. **31**, 1628–1652 (2009)
18. Karas, E.W., Pilotta, E.A., Ribeiro, A.A.: Numerical comparison of merit function with filter criterion in inexact restoration algorithms using Hard-Spheres Problems. Comput. Optim. Appl. **44**, 427–441 (2009)
19. Fischer, A., Friedlander, A.: A new line search inexact restoration approach for nonlinear programming. Comput. Optim. Appl. **46**, 333–346 (2010)
20. Karas, E.W., Gonzaga, C.C., Ribeiro, A.A.: Local convergence of filter methods for equality constrained nonlinear programming. Optimization **59**, 1153–1171 (2010)
21. Kaya, C.Y.: Inexact Restoration for Runge–Kutta discretization of optimal control problems. SIAM J. Numer. Anal. **48**, 1492–1517 (2010)

22. Francisco, J.B., Martínez, J.M., Martínez, L., Pisnitchenko, F.I.: Inexact Restoration method for minimization problems arising in electronic structure calculations. Comput. Optim. Appl. **50**, 555–590 (2011)
23. Bueno, L.F., Friedlander, A., Martínez, J.M., Sobral, F.: Inexact restoration method for derivative-free optimization with smooth constraints. SIAM J. Optim. **23**, 1189–1213 (2013)
24. Martínez, J.M., Svaiter, B.F.: A practical optimality condition without constraint qualifications for nonlinear programming. J. Optim. Theory Appl. **118**, 117–133 (2003)
25. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**, 627–641 (2011)
26. Barzilai, J., Borwein, J.M.: Two point step size gradient method. IMA J. Numer. Anal. **8**, 141–148 (1988)
27. Dai, Y.H., Fletcher, R.: On the asymptotic behaviour of some new gradient methods. Math. Prog. **103**, 541–559 (2005)
28. Friedlander, A., Martínez, J.M., Molina, B., Raydan, M.: Gradient method with retards and generalizations. SIAM J. Numer. Anal. **36**, 275–289 (1998)
29. Raydan, M.: On the Barzilai and Borwein choice of steplength for the gradient method. IMA J. Numer. Anal. **13**, 321–326 (1993)
30. Raydan, M.: The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM J. Optim. **7**, 26–33 (1997)
31. Birgin, E.G., Martínez, J.M., Raydan, M.: Nonmonotone spectral projected gradient methods on convex sets. SIAM J. Optim. **10**, 1196–1211 (2000)
32. Birgin, E.G., Martínez, J.M., Raydan, M.: Algorithm 813: SPG Software for convex-constrained optimization. ACM Trans. Math. Softw. **27**, 340–349 (2001)
33. Birgin, E.G., Martínez, J.M., Raydan, M.: Inexact spectral projected gradient methods on convex sets. IMA J. Numer. Anal. **23**, 539–559 (2003)
34. Andreani, R., Birgin, E.G., Martínez, J.M., Yuan, J.-Y.: Spectral projected gradient and variable metric methods for optimization with linear inequalities. IMA J. Numer. Anal. **25**, 221–252 (2005)
35. Rockafellar, R.T.: Lagrange multipliers and optimality. SIAM Rev. **35**, 183–238 (1993)
36. Burachik, R.S., Gasimov, R.N., Ismayilova, A.N., Kaya, C.Y.: On a modified subgradient algorithm for dual problems via sharp augmented Lagrangian. J. Global Optim. **34**, 55–78 (2006)
37. Burachik, R.S., Iusem, A.N., Melo, J.G.: A primal dual modified subgradient algorithm with sharp Lagrangian. J. Global Optim. **46**, 347–361 (2010)
38. Burachik, R.S., Kaya, C.Y., Mammadov, M.: An inexact modified subgradient algorithm for nonconvex optimization. Comput. Optim. Appl. **45**, 1–24 (2010)
39. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmont (1999)
40. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new constraint qualifications and applications. SIAM J. Optim. **22**, 1109–1135 (2012)
41. Andreani, R., Martínez, J.M., Schuverdt, M.L.: On the relation between the Constant Positive Linear Dependence condition and quasinormality constraint qualification. J. Optim. Theory Appl. **125**, 473–485 (2005)
42. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Prog. **135**, 255–273 (2012)
43. Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. SIAM J. Optim. **10**, 963–981 (2000)
44. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**, 429–440 (2011)
45. Bueno, L.F.: Otimização com restries LOVO, Restauração Inexata e o Equilíbrio Inverso de Nash. Ph.D. dissertation, Departamento de Matemática Aplicada, Universidade Estadual de Campinas (2011)