On the complexity of an Inexact Restoration method for constrained optimization *

L. F. Bueno † J. M. Martínez ‡

September 18, 2018

Abstract

Recent papers indicate that some algorithms for constrained optimization may exhibit worst-case complexity bounds that are very similar to those of unconstrained optimization algorithms. A natural question is whether well established practical algorithms, perhaps with small variations, may enjoy analogous complexity results. In the present paper we show that the answer is positive with respect to Inexact Restoration algorithms in which first-order methods are employed for approximating the solution of subproblems.

Key words: Complexity, Continuous Optimization, Constrained Optimization, Inexact Restoration Methods, Regularization.

1 Introduction

Inexact Restoration (IR) algorithms were introduced with the aim of solving constrained optimization problems [33, 35]. Each iteration of an IR algorithm consists of two phases. In the first phase one improves feasibility and in the second case optimality is improved onto a linear tangent approximation of the constraints. When a sufficient descent criterion does not hold the trial point is modified in such a way that, eventually, acceptance occurs at a point that may be close to the solution of the restoration (first) phase. The acceptance criterion may use merit functions [33, 35] or filters [30]. The IR approach proved to be quite useful for electronic structure calculations [28] and other matricial problems for which natural restoration procedures exist.

Constrained optimization algorithms exhibiting unconstrained-like complexity results were given in [9, 22]. These algorithms are not reliable for practical calculations because they generate short-step sequences, due to the necessity of maintaining approximate feasibility at every iteration. However, the two-phase nature of those algorithms motivated the present research. Unlike

*This work was supported by CEPID-CEMeai FAPESP Grants 2015/02528-8 and 2013/05475-7 and CNPq.

†Department of Science and Technology, Federal University of São Paulo, São José dos Campos, SP, Brazil. E-mail: lfelipebueno@gmail.com

‡Professor, Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas, SP, Brazil. E-mail: martinez@ime.unicamp.br

the algorithms [9, 22], which employ an initial restoration phase followed by a second phase of short iterations that preserve approximate feasibility, IR algorithms alternate restoration and optimization phases, not limited by feasibility tolerances but subject to a global criterion of quality. In such a way, large steps are possible when the current approximation is far from the solution. In this paper we will show that pleasant complexity results hold for an IR algorithm that is similar to the ones presented in [33] and other IR papers.

The analysis of the worst-case function-evaluation complexity of continuous optimization algorithms became relevant in the last 12 years. Given a stopping criterion based on a tolerance $\varepsilon > 0$, one tries to find a (sharp) upper bound for the number of evaluations that a given algorithm needs for satisfying the stopping requirement. In unconstrained optimization, if $f(x)$ is the objective function, the generally accepted first-order stopping criterion corresponds to the approximate annihilation of the gradient, $\|\nabla f(x)\| \leq \varepsilon$. For this case, relevant complexity results, if the derivatives of the objective function satisfy Lipschitz-continuity requirements, are (see [6, 7, 12, 17, 18, 19, 21, 25, 31, 38]):

1. Suitable gradient-related or quasi-Newton algorithms stop employing, at most, $O(\varepsilon^{-2})$ function and gradient evaluations.

2. Suitable Newton-like algorithms stop employing, at most, $O(\varepsilon^{-3/2})$ function, gradient, and Hessian evaluations.

The "$O(\varepsilon^{-q})$ statement" means that the maximal number of evaluations is smaller than a constant times ε^{-q}, where the constant depends on the initial approximation, parameters of the algorithm and characteristics of the problem.

Essentially, the results above are also true for constrained problems in which the feasible set is simple enough [23, 34]. Roughly speaking, simplicity of the feasible set means that the minimization of a quadratic function onto such a set is relatively easy to perform.

The papers [9, 22] proved that, for minimization problems with general constraints, $O(\varepsilon^{-q})$ results can be proved that are similar to those in the unconstrained case, for some algorithms based on a single feasibility phase, followed by a short-step optimization phase.

However, the short-step characteristic of the algorithms [9, 22] makes them unsuitable for practical computations. Very short steps are necessary to guarantee pleasant worst-case behavior but the number of short steps that are necessary to prove complexity is close to the number of short steps that one would employ in a sensible implementation of the algorithms.

On the other hand, Cartis, Gould, and Toint [20] considered the complexity of nonconvex equality-constrained optimization employing a first-order exact penalty method. Under the assumption that penalty parameters are bounded, the complexity of finding and approximate KKT point was proved to be $O(\varepsilon^{-2})$.

This state of facts motivated us to study well-established constrained optimization algorithms from the point of view of worst-case complexity. The analogy (feasibility and optimality phases) between the algorithms [9, 22] and the framework of Inexact Restoration (IR) methods [14, 27, 33, 35] led us to define a suitable regularized form of IR and to study its complexity properties. Inexact Restoration techniques for constrained optimization were introduced in [33, 35], improved, extended and analyzed in [27] and [14], among others, and used successfully in relevant applications [28].

In this paper we analyze a first-order version of IR (no second-derivative information will be used) and we prove that the computer work necessary to achieve suitable stopping criteria is smaller than a constant times $(\epsilon_{feas}^{-1} + \epsilon_{opt}^{-2})$, where ϵ_{feas} is the feasibility tolerance, ϵ_{opt} is the optimality tolerance and the constant depends on algorithmic parameters and characteristics of the optimization problem such as function bounds and Lipschitz constants.

In Section 2 we present the main algorithm and we state some basic results. In Section 3 we prove the complexity results and in Section 4 we state conclusions and lines for future research.

Notation

The symbol $\|\cdot\|$ will denote the Euclidean norm on $\mathbb{R}^n$.

$P_D(z)$ denotes the Euclidean projection of z onto the convex set D.

We denote $\mathbb{N} = \{0, 1, 2, \ldots\}$.

2 Inexact Restoration Algorithm

The problem considered in this paper is:

$$\text{Minimize } f(x) \quad \text{subject to} \quad h(x) = 0 \quad \text{and} \quad x \in \Omega, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, and Ω is a nonempty and compact polytope. Every nonlinear programming problem including inequality constraints may be formulated in the form (1) since a constraint $h_j(x) \leq 0$ may be replaced with $h_j(x) + z = 0, z \geq 0$. In most practical cases Ω is an n-dimensional box $\ell \leq x \leq u$. The generalization of our results for a general compact and convex Ω is also possible subject to constraint-qualification assumptions on Ω.

We define, for all $x \in \mathbb{R}^n$,

$$c(x) = \frac{1}{2}\|h(x)\|^2. \tag{2}$$

All along this paper we will assume that the functions f and h are continuously differentiable, and there exist nonnegative constants L_f, C_f, L_h, C_h, $C_{\nabla h}$, and L_c such that, for all $x, z \in \Omega$:

$$|f(x) - f(z)| \leq L_f\|x - z\|, \tag{3}$$

$$\|\nabla f(x) - \nabla f(z)\| \leq L_f\|x - z\|, \tag{4}$$

$$f(z) \leq f(x) + \nabla f(x)^T(z - x) + L_f\|z - x\|^2, \tag{5}$$

$$f(x) \leq C_f, \tag{6}$$

$$\|h(x) - h(z)\| \leq L_h\|x - z\|, \tag{7}$$

$$\|\nabla h(x) - \nabla h(z)\| \leq L_h\|x - z\|, \tag{8}$$

$$\|h(z)\| \leq \|h(x) + \nabla h(x)^T(z-x)\| + L_h\|z-x\|^2, \tag{9}$$

$$\|h(x)\| \leq C_h, \tag{10}$$

$$\|\nabla h(x)\| \leq C_{\nabla h}, \tag{11}$$

$$\|\nabla c(x) - \nabla c(z)\| \leq L_c\|x-z\|, \tag{12}$$

and

$$c(z) \leq c(x) + \nabla c(x)^T(z-x) + L_c\|z-x\|^2. \tag{13}$$

The (unknown) constants L_f, C_f, L_h, C_h, $C_{\nabla h}$, and L_c will be called *characteristics of the problem* (1). The complexity results to be proved in this paper will be of the form

$$\text{Computer Work} \;\leq\; \text{Constant} \;\times (\epsilon_{feas}^{-1} + \epsilon_{opt}^{-2}),$$

where "Constant" only depends on characteristics of the problem and algorithmic parameters, defined below. Moreover, ϵ_{feas} is a tolerance related to infeasibility, and ϵ_{opt} is a tolerance for optimality. All along the paper, expressions of the form $a = O(b)$ or $a \leq O(b)$ will mean that the nonnegative quantity a is not bigger than a constant times the quantity b where the constant only depends on characteristics of the problem and algorithmic parameters. In particular, $a \geq 1/O(1)$ indicates that the quantity a is not smaller than a positive quantity that only depends on characteristics of the problem and algorithmic parameters.

For all $x \in \Omega$, and $\theta \in (0,1)$, we define the merit function $\Phi(x,\theta)$ by

$$\Phi(x,\theta) = \theta f(x) + (1-\theta)\|h(x)\|. \tag{14}$$

The main algorithm considered in this paper is presented below. Unlike other optimization algorithms for which complexity has been analyzed [9, 22], the description of Algorithm 2.1 does not depend of the possible stopping parameters ϵ_{feas} and ϵ_{opt}. In particular, we describe the algorithm without a stopping criterion regarding ϵ_{feas} and ϵ_{opt}. This makes it easy to show that the asymptotic convergence results follow as trivial consequences of the complexity ones. If restoration breakdown does not occur, the algorithm generates an infinite sequence and complexity results will follow from bounds on the number of iterations at which desired precisions are not achieved.

Algorithm 2.1 - Inexact Restoration

Let $\gamma > 0$, $M \geq 1$, $\kappa > 0$, $\mu_{max} \geq \mu_{min} > 0$, $\theta_0 \in (0,1)$, $r \in (0,1)$, and $r_{feas} \in (0,r)$ be *algorithmic parameters*. Let $x^0 \in \Omega$. Set $k \leftarrow 0$.

Step 1 Restoration phase.

Compute $y^k \in \Omega$ using Algorithm 2.2 below with parameters M, r, r_{feas}, μ_{min}, and μ_{max}.

Test the inequality

$$\|h(y^k)\| \le r\|h(x^k)\|. \tag{15}$$

If (15) does not hold, stop Algorithm 2.1 declaring *restoration failure*.

Step 2 Penalty parameter

If

$$\Phi(y^k, \theta_k) - \Phi(x^k, \theta_k) \le \frac{1}{2}(1-r)(\|h(y^k)\| - \|h(x^k)\|) \tag{16}$$

set $\theta_{k+1} = \theta_k$.

Else, compute

$$\theta_{k+1} = \frac{(1+r)(\|h(x^k)\| - \|h(y^k)\|)}{2[f(y^k) - f(x^k) + \|h(x^k)\| - \|h(y^k)\|]}. \tag{17}$$

Step 3 Optimization phase.

Choose $\mu \in [\mu_{min}, \mu_{max}]$ and $H_k \in \mathbb{R}^{n\times n}$, symmetric, such that $\|H_k\| \le M$.

Step 3.1 Tangent set minimization

Compute $x \in \mathbb{R}^n$ an approximate solution of the following subproblem:

$$\begin{array}{ll} \text{Minimize} & \nabla f(y^k)^T(x-y^k) + \frac{1}{2}(x-y^k)^T H_k(x-y^k) + \mu\|x-y^k\|^2 \\ \text{subject to} & \nabla h(y^k)^T(x-y^k) = 0 \\ & x \in \Omega. \end{array} \tag{18}$$

(The sense in which the solution of (18) must be approximated by x will be given later in Assumptions 3.1 and 3.2. Assumption 3.2 depends on the algorithmic parameter κ.)

Step 3.2 Descent conditions

Test the conditions

$$f(x) \le f(y^k) - \gamma\|x-y^k\|^2 \tag{19}$$

and

$$\Phi(x, \theta_{k+1}) \le \Phi(x^k, \theta_{k+1}) + \frac{1}{2}(1-r)(\|h(y^k)\| - \|h(x^k)\|). \tag{20}$$

If both (19) and (20) are fulfilled, define $\mu_k = \mu$, $x^{k+1} = x$, update $k \leftarrow k+1$, and go to Step 1.

Else, update

$$\mu \in [2\mu, 10\mu] \tag{21}$$

and go to Step 3.1.

Algorithm 2.2 below describes the restoration procedure. This algorithm is similar to Algorithm 2.1 of [13] applied to the function $c(y) = \frac{1}{2}\|h(y)\|^2$ with appropriate choices of the target and the precision.

Algorithm 2.2 - Restoration procedure

Step 1 If $\|h(x^k)\| = 0$ set $y^k = x^k$ and return to Algorithm 2.1.

Step 2 Compute
$$c_{target} = \frac{1}{2} r^2 \|h(x^k)\|^2 \text{ and } \epsilon_c = r_{feas} \|h(x^k)\|. \tag{22}$$

Step 3 Initialize $z^0 = x^k$, $\sigma \leftarrow 0$, and $\ell \leftarrow 0$.

Step 4 If $c(z^\ell) \le c_{target}$ or $\|P_\Omega(z^\ell - \nabla c(z^\ell)) - z^\ell\| \le \epsilon_c$ set $y^k = z^\ell$, $\sigma_k = \sigma$, and return to Algorithm 2.1. Else, choose $\sigma \in [\mu_{min}, \mu_{max}]$ and $B_\ell \in \mathbb{R}^{n \times n}$, symmetric and positive definite, such that $\|B_\ell\| \le M$ and $\|B_\ell^{-1}\| \le M$.

Step 4.1 Find the solution $z \in \mathbb{R}^n$ of the following subproblem:

$$\begin{array}{ll} \text{Minimize} & \nabla c(z^\ell)^T (z - z^\ell) + \frac{1}{2}(z - z^\ell)^T B_\ell (z - z^\ell) + \frac{\sigma}{2} \|z - z^\ell\|^2 \\ \text{subject to} & z \in \Omega. \end{array} \tag{23}$$

Step 4.2 Test the condition

$$c(z) \le c(z^\ell) - \gamma \|z - z^\ell\|^2. \tag{24}$$

If (24) is fulfilled, define $z^{\ell+1} = z$, $\sigma_{k,\ell} = \sigma$, update $\ell \leftarrow \ell + 1$, and go to Step 4. Else, update

$$\sigma \in [2\sigma, 10\sigma] \tag{25}$$

and go to Step 4.1.

Remark

Algorithm 2.1 stops, employing a finite number of iterations, only when restoration fails. Failure of restoration is declared when a point y is found such that $\|P_\Omega(y^k - \nabla c(y^k)) - y^k\| \le r_{feas} \|h(x^k)\|$ but $\|h(y^k)\| > r \|h(x^k)\|$. Therefore,

$$\|P_\Omega(y^k - \nabla c(y^k)) - y^k\| \le r_{feas} \|h(x^k)\| \tag{26}$$

but

$$\|h(x^k)\| < \frac{1}{r} \|h(y^k)\|. \tag{27}$$

By (26) and (27),

$$\|P_\Omega(y^k - \nabla h(y^k) h(y^k) - y^k\| \le \frac{r_{feas}}{r} \|h(y^k)\| \tag{28}$$

If $r_{feas} << r$ this means that the infeasibility of y^k is considerably bigger than the projected gradient of the sum of infeasibility squares at y^k. If $\|h(y^k)\|$ is not small (28) probably indicates proximity to a local minimizer of $\|h(y)\|^2$ at which $\|h(y)\|$ does not vanish. On the other hand, if $\|h(y^k)\|$ is small, the fulfillment of (28) reflects the non-fulfillment of a desirable constraint qualification. For example, if y^k is interior to Ω, (28) implies that there exists $v \in \mathbb{R}^n$ such that $\|v\| = 1$ and

$$\|\nabla h(y^k) v\| \le \frac{r_{feas}}{r}.$$

This property probably indicates the proximity of a feasible point at which the gradients $\nabla h_1(y), \ldots, \nabla h_m(y)$ are not linearly independent.

Lemma 2.1 *Algorithm 2.2 finishes finding $y^k \in \Omega$ that satisfies*

$$c(y^k) \leq c_{target} \tag{29}$$

or

$$\|P_\Omega(y^k - \nabla c(y^k)) - y^k\| \leq \epsilon_c \tag{30}$$

employing, at most, $O\left(\frac{c(x^k)-c_{target}}{\epsilon_c^2}\right)$ *iterations and evaluations of* c*. Moreover, at each iteration of Algorithm 2.2, the descent condition (24) is tested* $O(1)$ *times. Finally, one has that* $\sigma_k \leq O(1)$*.*

Proof. By (13), if z is a solution of (23), we have that

$$c(z) - c(z^\ell) \leq \nabla c(z^\ell)^T(z - z^\ell) + L_c\|z - z^\ell\|^2$$

$$= \nabla c(z^\ell)^T(z-z^\ell)+\frac{1}{2}(z-z^\ell)^T B_\ell(z-z^\ell)+\frac{\sigma}{2}\|z-z^\ell\|^2-\frac{\sigma}{2}\|z-z^\ell\|^2-\frac{1}{2}(z-z^\ell)^T B_\ell(z-z^\ell)+L_c\|z-z^\ell\|^2$$

$$\leq (-\frac{\sigma}{2} + L_c)\|z - z^\ell\|^2.$$

Therefore, if $\frac{\sigma}{2} \geq L_c + \gamma$, we have that (24) holds.

On the other hand, by the contraction property of projections and (12),

$$\|P_\Omega(z^{\ell+1} - \nabla c(z^{\ell+1})) - z^{\ell+1} - P_\Omega(z^{\ell+1} - [\nabla c(z^\ell) + B_\ell(z^{\ell+1} - z^\ell)]) + z^{\ell+1}\|$$

$$\leq \|\nabla c(z^{\ell+1}) - \nabla c(z^\ell)\| + M\|z^{\ell+1} - z^\ell\| \leq (L_c + M)\|z^{\ell+1} - z^\ell\|.$$

Thus,

$$\|P_\Omega(z^{\ell+1}-\nabla c(z^{\ell+1}))-z^{\ell+1}\| \leq \|P_\Omega(z^{\ell+1}-[\nabla c(z^\ell)+B_k(z^{\ell+1}-z^\ell)])-z^{\ell+1}\|+(L_c+M)\|z^{\ell+1}-z^\ell\|.$$

Now, again by the contraction property of projections,

$$\|P_\Omega(z^{\ell+1}-[\nabla c(z^\ell)+B_k(z^{\ell+1}-z^\ell)+\sigma(z^{\ell+1}-z^\ell)])-P_\Omega(z^{\ell+1}-[\nabla c(z^\ell)+B_k(z^{\ell+1}-z^\ell)])\| \leq \sigma\|z^{\ell+1}-z^\ell\| \tag{31}$$

Thus, since $P_\Omega(z^{\ell+1} - [\nabla c(z^\ell) + B_k(z^{\ell+1} - z^\ell) + \sigma(z^{\ell+1} - z^\ell)]) - z^{\ell+1} = 0$, (31) implies that

$$\|P_\Omega(z^{\ell+1} - \nabla c(z^{\ell+1})) - z^{\ell+1}\| \leq (\sigma + L_c + M)\|z^{\ell+1} - z^\ell\|. \tag{32}$$

Thus, the desired result follows from (24) and (32). □

Corollary 2.1 *The number of iterations and evaluations of* h *at each call of Algorithm 2.2 is bounded by a constant* $O(1)$*.*

Proof. By (22), the result follows from Lemma 2.1. □

The objective of the following results is to show that the distance between x^k and y^k is bounded by a multiple of the infeasibility measure.

Lemma 2.2 *There exists $C_s \leq O(1)$ such that, for all k and l, the iterates generated by Algorithm 2.2 satisfy*

$$\|z^{\ell+1} - z^{\ell}\| \leq C_s \|h(x^k)\|.$$

Proof. The result is obvious if $z^{\ell+1} = z^{\ell}$. Otherwise, define

$$v = \frac{z^{\ell+1} - z^{\ell}}{\|z^{\ell+1} - z^{\ell}\|}.$$

Note that $\nabla c(z^{\ell})^T v < 0$. Consider the function $\varphi(t) = t \nabla c(z^{\ell})^T v + \frac{t^2}{2} v^T (B_{\ell} + \sigma_k I) v$, for $t \geq 0$. The unconstrained minimizer of this parabola is

$$t_* = -\frac{\nabla c(z^{\ell})^T v}{v^T (B_{\ell} + \sigma_k I) v} \leq -\frac{\nabla c(z^{\ell})^T v}{v^T B_{\ell} v}.$$

Therefore,

$$t_* \leq \frac{\|\nabla c(z^{\ell})\|}{\lambda_1(B_{\ell})},$$

where $\lambda_1(B_{\ell}) > 0$ is the smallest eigenvalue of B_{ℓ}. Therefore, $\lambda_1(B_{\ell}) = 1/\|B_{\ell}^{-1}\| \geq 1/M$. So,

$$t_* \leq M \|\nabla c(z^{\ell})\|. \tag{33}$$

Let $\bar{t}$ be the minimizer of $\varphi(t)$ subject to $z^{\ell} + tv \in \Omega$. By the convexity of Ω and the form of $\varphi(t)$ we have that $\bar{t} \leq t_*$. But, by construction, $z^{\ell} + \bar{t}v = z^{\ell+1}$. So, by (33),

$$\|z^{\ell+1} - z^{\ell}\| \leq M \|\nabla c(z^{\ell})\| \leq M \|\nabla h(z^{\ell}) h(z^{\ell})\| \leq M C_{\nabla h} \|h(z^{\ell})\|. \tag{34}$$

Since $c(z^{\ell}) \leq c(x^k)$ we have that $\|h(z^{\ell})\| \leq \|h(x^k)\|$, so the thesis follows from (34) with $C_s = M C_{\nabla h}$. □

Lemma 2.3 *There exists $\beta \leq O(1)$ such that, for all y^k computed at Step 1 of Algorithm 2.1, we have that*

$$\|y^k - x^k\| \leq \beta \|h(x^k)\|. \tag{35}$$

Proof. Let N_{Rk} be the number of iterations performed by Algorithm 2.2 at iteration k of Algorithm 2.1. By Corollary 2.1 we have that there exists N_R such that $N_{Rk} \leq N_R \leq O(1)$.

Then, by Lemma 2.2, we have that

$$\|y^k - x^k\| = \left\| \sum_{l=1}^{N_{Rk}} (z^{\ell} - z^{\ell-1}) \right\| \leq N_R C_s \|h(x^k)\|.$$

Defining $\beta = N_R C_s$ we have the desired result. □

3 Complexity and convergence

All along this section we consider sequences $\{x^k\}$ and $\{y^k\}$ generated by Algorithm 2.1. These sequences are defined for all $k \in \mathbb{N}$, except if stopping occurs at some y^k with the diagnostic of failure of restoration. The main results of this section say that, given an arbitrary $\varepsilon > 0$, the number of iterations such that the norm of infeasibility is bigger than ε is smaller than $O(\varepsilon^{-1})$ and that the number of iterations such that the norm of a vector that represents optimality is bigger than ε is smaller than $O(\varepsilon^{-2})$. As a consequence, we will obtain global convergence of the algorithm and suitable stopping criteria.

The first technical lemma states that the penalty parameters $\{\theta_k\}$ are bounded away from zero.

Lemma 3.1 *Given x^k and y^k satisfying* (15)*, Step 2 of Algorithm 2.1 is well defined. Moreover, $\theta_{k+1} \leq \theta_k$, the inequality*

$$\Phi(y^k, \theta_{k+1}) - \Phi(x^k, \theta_{k+1}) \leq \frac{1}{2}(1-r)(\|h(y^k)\| - \|h(x^k)\|) \tag{36}$$

is fulfilled, and there exits $C_\theta \leq 1/O(1)$ such that

$$\theta_k \geq C_\theta. \tag{37}$$

Proof. We will first prove that Step 2 is well defined and that $0 < \theta_{k+1} \leq \theta_k$. If $\|h(y^k)\| - \|h(x^k)\| = 0$, then, by (15), we have that $\|h(y^k)\| = \|h(x^k)\| = 0$ and $\Phi(x^k, \theta_k) = \Phi(y^k, \theta_k)$. Thus, (16) holds in this case and, consequently, $\theta_{k+1} = \theta_k > 0$.

Therefore, it remains to consider only the case in which $\|h(y^k)\| < \|h(x^k)\|$. In this case, we obtain that

$$\begin{aligned} &\|h(x^k)\| - \|h(y^k)\| + \frac{1-r}{2}(\|h(y^k)\| - \|h(x^k)\|) \\ &= \frac{1+r}{2}(\|h(x^k)\| - \|h(y^k)\|) > 0. \end{aligned} \tag{38}$$

By direct calculations, the inequality (16) is equivalent to

$$\begin{aligned} &\theta_k[f(y^k) - f(x^k) + \|h(x^k)\| - \|h(y^k)\|] \\ &\leq \|h(x^k)\| - \|h(y^k)\| + \frac{1-r}{2}(\|h(y^k)\| - \|h(x^k)\|). \end{aligned} \tag{39}$$

Thus, by (38) and the fact that $\theta_k > 0$, the requirement (16) is fulfilled whenever $f(y^k) - f(x^k) + \|h(x^k)\| - \|h(y^k)\| \leq 0$. In this case, the algorithm also chooses $\theta_{k+1} = \theta_k > 0$.

Therefore, we only need to consider the case in which

$$f(y^k) - f(x^k) + \|h(x^k)\| - \|h(y^k)\| > 0.$$

In this case, both the numerator and the denominator of (17) are positive. It turns out that $\theta_{k+1} > 0$ whenever θ_{k+1} is equal to θ_k or it is defined by (17). Moreover, if (16) does not hold, then, by (39), we have that

$$\Phi(y^k, \theta) - \Phi(x^k, \theta) > \frac{1-r}{2}(\|h(y^k)\| - \|h(x^k)\|)$$

for all $\theta \geq \theta_k$. Now, since the choice (17) obviously implies that

$$\begin{aligned} &\theta_{k+1}[f(y^k) - f(x^k) + \|h(x^k)\| - \|h(y^k)\|] \\ &= \|h(x^k)\| - \|h(y^k)\| + \frac{1-r}{2}(\|h(y^k)\| - \|h(x^k)\|), \end{aligned} \tag{40}$$

we conclude that $0 < \theta_{k+1} \leq \theta_k$ in all cases. Thus, since $\theta_0 \in (0, 1)$, the sequence $\{\theta_k\}$ is positive and non-increasing. Furthermore, by (16), (17) and (40), we have that

$$\Phi(y^k, \theta_{k+1}) - \Phi(x^k, \theta_{k+1}) \leq \frac{1-r}{2}(\|h(y^k)\| - \|h(x^k)\|).$$

It remains to prove that the sequence $\{\theta_k\}$ is bounded away from zero. For this purpose, it suffices to show that θ_{k+1} is greater than a fixed positive number when it is defined by (17). In this case, we have that:

$$\begin{aligned} \frac{1}{\theta_{k+1}} &= \frac{2[f(y^k) - f(x^k) + \|h(x^k)\| - \|h(y^k)\|]}{(1+r)[\|h(x^k)\| - \|h(y^k)\|]} \\ &\leq \frac{2}{1+r}\left[\frac{|f(y^k) - f(x^k)|}{\|h(x^k)\| - \|h(y^k)\|} + 1\right]. \end{aligned}$$

Thus, by (3), (15), and Lemma 2.4,

$$\frac{1}{\theta_{k+1}} \leq \frac{2}{1+r}\left[\frac{L_f\|y^k - x^k\|}{(1-r)\|h(x^k)\|} + 1\right] \leq \frac{2}{1+r}\left[\frac{L_f\beta}{1-r} + 1\right].$$

This implies that the sequence $\{1/\theta_{k+1}\}$ is bounded. Therefore, the monotone sequence $\{\theta_k\}$ is bounded away from zero.

Then, defining $C_\theta = \left[\frac{2}{1+r}\left(\frac{L_f\beta}{1-r} + 1\right)\right]^{-1}$ we have that $\theta_k \geq C_\theta$, as we wanted to prove. □

Assumption 3.1 *For every iteration k, the approximate solution of the quadratic programming problem* (18) *satisfies*

$$\nabla f(y^k)^T(x - y^k) + \frac{1}{2}(x - y^k)^T H_k(x - y^k) + \mu\|x - y^k\|^2 \leq 0. \tag{41}$$

Lemma 3.2 below states that the criteria for stopping an optimality phase are necessarily satisfied after $O(1)$ iterations.

Lemma 3.2 *Suppose that the approximate solutions of (18) satisfy Assumption 3.1. Then, there exists $n_{reg} \in \mathbb{N}$, such that $n_{reg} \le O(1)$, and after at most n_{reg} updates (21), conditions (19) and (20) are satisfied. (Thus, Step 3 of Algorithm 2.1 ends in finite time.) Moreover, there exists $\bar{\mu} \le O(1)$ such that $\mu_k \le \bar{\mu}$ for all k.*

Proof. By Lemma 3.1, there exists $\bar{\theta} \ge 1/O(1)$ such that

$$\theta_k \ge \bar{\theta} \text{ for all } k. \tag{42}$$

Define

$$\alpha = \max\left\{\gamma, \frac{1-\bar{\theta}}{\bar{\theta}} L_h\right\}. \tag{43}$$

If $\mu \ge M + L_f + \alpha$ then, by (5) and (41),

$$\begin{aligned} f(x) &\le f(y^k) + \nabla f(y^k)^T (x - y^k) + L_f \|x - y^k\|^2 \\ &\le f(y^k) + \nabla f(y^k)^T (x - y^k) + \frac{1}{2}(x - y^k)^T H_k (x - y^k) + \left(\frac{M}{2} + L_f + \alpha\right) \|x - y^k\|^2 - \alpha \|x - y^k\|^2 \\ &\le f(y^k) + \nabla f(y^k)^T (x - y^k) + \frac{1}{2}(x - y^k)^T H_k (x - y^k) + \mu \|x - y^k\|^2 - \alpha \|x - y^k\|^2 \\ &\le f(y^k) - \alpha \|x - y^k\|^2. \end{aligned}$$

Since $\gamma \le \alpha$, (19) holds. Moreover, by the definition of α,

$$f(x) - f(y^k) \le -\frac{1-\bar{\theta}}{\bar{\theta}} L_h \|x - y^k\|^2. \tag{44}$$

Let us show now that condition (20) also holds for $\mu \ge M + L_f + \alpha$. By (9), (36), (44), (42) and the definition of Φ, we have:

$$\begin{aligned} &\Phi(x, \theta_{k+1}) - \Phi(x^k, \theta_{k+1}) \\ &= \Phi(x, \theta_{k+1}) - \Phi(y^k, \theta_{k+1}) + \Phi(y^k, \theta_{k+1}) - \Phi(x^k, \theta_{k+1}) \\ &\le \theta_{k+1}[f(x) - f(y^k)] + (1 - \theta_{k+1})(\|h(x)\| - \|h(y^k)\|) \\ &\quad + \frac{1-r}{2}(\|h(y^k)\| - \|h(x^k)\|) \\ &\le -\theta_{k+1}\frac{1-\bar{\theta}}{\bar{\theta}} L_h \|x - y^k\|^2 + (1 - \theta_{k+1}) L_h \|x - y^k\|^2 + \frac{1-r}{2}(\|h(y^k)\| - \|h(x^k)\|) \\ &\le -\bar{\theta}\frac{1-\bar{\theta}}{\bar{\theta}} L_h \|x - y^k\|^2 + (1 - \bar{\theta}) L_h \|x - y^k\|^2 + \frac{1-r}{2}(\|h(y^k)\| - \|h(x^k)\|) \\ &\le \frac{1-r}{2}(\|h(y^k)\| - \|h(x^k)\|). \end{aligned}$$

Thus, after at most $n_{reg} \equiv \lfloor \log_2(M + L_f + \alpha) - \log_2(\mu_{min}) \rfloor + 1$ updates (21), both (19) and (20) are satisfied. Note that by (37) we have that $\frac{1-\bar{\theta}}{\bar{\theta}} \le \frac{1}{C_\theta} - 1$. Thus, we have that

$$\alpha \le O(L_h).$$

Moreover, by the boundedness of the initial μ and the update rule (21), there exists $\bar{\mu} > 0$ such that the whole sequence $\{\mu_k\}$ is bounded by $\lfloor M + L_f + \alpha \rfloor + 1 \le \bar{\mu}$ independently of k. □

In Theorem 3.1 we prove that the sum of the infeasibilities of all the iterates is bounded and that the same happens with the sum of all the squared increments computed in the optimality phases.

Theorem 3.1 *Suppose that Assumption 3.1 holds. Then, there exist* $\bar{h} \le O(1)$ *and* $C_d \le O(1)$ *such that for all* $k \in \mathbb{N}$ *and* $j \le k$*, if the sequences* $\{x^j\}$ *and* $\{y^j\}$ *are generated by Algorithm 2.1, we have that*

$$\sum_{j=1}^{k} \|h(x^j)\| \le \bar{h}, \tag{45}$$

$$\sum_{j=1}^{k} \|h(y^j)\| \le r\bar{h}, \tag{46}$$

$$f(x^k) \le f(x^0) + L_f \beta \bar{h} - \gamma \sum_{j=0}^{k-1} \|x^{j+1} - y^j\|^2, \tag{47}$$

and

$$\sum_{j=0}^{k} \|x^{j+1} - y^j\|^2 \le C_d. \tag{48}$$

Proof. By condition (20), for all j one has that

$$\Phi(x^{j+1}, \theta_{j+1}) \le \Phi(x^j, \theta_{j+1}) + \frac{1-r}{2}(\|h(y^j)\| - \|h(x^j)\|).$$

Therefore, by (15),

$$\Phi(x^{j+1}, \theta_{j+1}) \le \Phi(x^j, \theta_{j+1}) - \frac{(1-r)^2}{2}\|h(x^j)\|. \tag{49}$$

Let us define $\rho_j = (1-\theta_j)/\theta_j$ for all j. By Lemma 3.1, we have that $\theta_j \ge C_\theta$ for all j. This implies that $\rho_j \le \frac{1}{C_\theta} - 1$ for all j. Since $\{\rho_j\}$ is bounded, and nondecreasing and $\rho_0 > 0$, it follows that, for all k,

$$\sum_{j=0}^{k-1} (\rho_{j+1} - \rho_j) = \rho_k - \rho_0 < \frac{1}{C_\theta}. \tag{50}$$

By (10) and (50) we have that there exists $C_\rho \le O(1)$, such that

$$\sum_{j=0}^{k-1} (\rho_{j+1} - \rho_j)\|h(x^j)\| \le C_\rho. \tag{51}$$

Now, by (49), for all $j \leq k-1$,

$$f(x^{j+1}) + \frac{1-\theta_{j+1}}{\theta_{j+1}}\|h(x^{j+1})\| \leq f(x^j) + \frac{1-\theta_{j+1}}{\theta_{j+1}}\|h(x^j)\| - \frac{(1-r)^2}{2\theta_{j+1}}\|h(x^j)\|.$$

Since $\theta_{j+1} < 1$, this implies that, for all $j \leq k-1$,

$$f(x^{j+1}) + \rho_{j+1}\|h(x^{j+1})\| \leq f(x^j) + \rho_{j+1}\|h(x^j)\| - \frac{(1-r)^2}{2}\|h(x^j)\|.$$

Therefore, for all $j \leq k-1$,

$$f(x^{j+1}) + \rho_{j+1}\|h(x^{j+1})\| \leq f(x^j) + \rho_j\|h(x^j)\| + (\rho_{j+1} - \rho_j)\|h(x^j)\| - \frac{(1-r)^2}{2}\|h(x^j)\|.$$

Thus, we have that

$$f(x^k) + \rho_k\|h(x^k)\| \leq f(x^0) + \rho_0\|h(x^0)\| + \sum_{j=0}^{k-1}(\rho_{j+1} - \rho_j)\|h(x^j)\| - \frac{(1-r)^2}{2}\sum_{j=0}^{k-1}\|h(x^j)\|.$$

Therefore, by (51),

$$f(x^k) + \rho_k\|h(x^k)\| \leq f(x^0) + \rho_0\|h(x^0)\| + C_\rho - \frac{(1-r)^2}{2}\sum_{j=0}^{k-1}\|h(x^j)\|.$$

Thus,

$$\frac{(1-r)^2}{2}\sum_{j=0}^{k}\|h(x^j)\| \leq \frac{(1-r)^2}{2}\|h(x^k)\| - [f(x^k) + \rho_k\|h(x^k)\|] + f(x^0) + \rho_0\|h(x^0)\| + C_\rho.$$

Since $\{\rho_k\} \geq 0$, by (6) and (10), it follows that there exists $\bar{h} \leq O(1)$ such that $\sum_{j=0}^{k}\|h(x^j)\| \leq \bar{h}$. Thus, (45) is proved. Then, (46) follows from (15) and (45).

Now, by (19), for all $j \leq k-1$ we have:

$$f(x^{j+1}) - f(x^j) \leq f(x^{j+1}) - f(y^j) + f(y^j) - f(x^j) \leq -\gamma\|x^{j+1} - y^j\|^2 + f(y^j) - f(x^j).$$

Then, by (3) and (35),

$$f(x^{j+1}) - f(x^j) \leq -\gamma\|x^{j+1} - y^j\|^2 + \beta L_f\|h(x^j)\|$$

for all $j \leq k-1$. Therefore,

$$f(x^k) \leq f(x^0) - \gamma\sum_{j=0}^{k-1}\|x^{j+1} - y^j\|^2 + \beta L_f\sum_{j=0}^{k-1}\|h(x^j)\|.$$

Therefore, (47) follows from (45) and (48) follows from (6) and (45). □

Assumption 3.2 means that we should solve the quadratic subproblem of Algorithm 2.1 with precision of the order of $\|x^{k+1} - y^k\|$.

From now on, we define, for all $k \in \mathbb{N}$,

$$D_{k+1} = \{x \in \Omega \mid \nabla h(y^k)^T(x - y^k) = 0\}. \quad (52)$$

Assumption 3.2 *Assumption 3.1 holds and there exists* $\kappa > 0$ *such that, for every iteration* k*, the approximate solution of the quadratic programming problem* (18) *satisfies (41) and*

$$\|P_{D_{k+1}}(x^{k+1} - \nabla f(y^k) - H_k(x^{k+1} - y^k) - 2\mu_k(x^{k+1} - y^k)) - x^{k+1}\| \le \kappa\|x^{k+1} - y^k\|. \quad (53)$$

In the following lemma we prove that the sum of the optimality measures for all the iterates generated by Algorithm 2.1 is bounded.

Lemma 3.3 *Suppose that Assumption 3.2 holds. Then, for every iteration* k*, if* x^{k+1} *is generated by Algorithm 2.1, we have:*

$$\|P_{D_{k+1}}(x^{k+1} - \nabla f(x^{k+1})) - x^{k+1}\| \le (L_f + 2\bar{\mu} + M + \kappa)\|x^{k+1} - y^k\| \quad (54)$$

and

$$\|P_{D_{k+1}}(y^k - \nabla f(y^k)) - y^k\| \le (2 + 2L_f + 2\bar{\mu} + M + \kappa)\|x^{k+1} - y^k\|, \quad (55)$$

where $\bar{\mu}$ *is defined in Lemma 3.2.*

Moreover, for every iteration k *we have that*

$$\sum_{j=0}^{k} \|P_{D_{j+1}}(x^{j+1} - \nabla f(x^{j+1})) - x^{j+1}\|^2 \le (L_f + 2\bar{\mu} + M + \kappa)^2 C_d \quad (56)$$

and

$$\sum_{j=0}^{k} \|P_{D_{j+1}}(y^j - \nabla f(y^j)) - y^j\|^2 \le (2 + 2L_f + 2\bar{\mu} + M + \kappa)^2 C_d, \quad (57)$$

where C_d *is given in (48).*

Proof. By (53), the contraction property of projections, and Lemma 3.2 (boundedness of μ_k) we have that:

$$\begin{aligned}
&\|P_{D_{k+1}}(x^{k+1} - \nabla f(x^{k+1})) - x^{k+1}\| \\
&\le \|P_{D_{k+1}}(x^{k+1} - \nabla f(x^{k+1})) - P_{D_{k+1}}(x^{k+1} - \nabla f(y^k) - H_k(x^{k+1} - y^k) - 2\mu_k(x^{k+1} - y^k))\| \\
&\quad + \|P_{D_{k+1}}(x^{k+1} - \nabla f(y^k) - H_k(x^{k+1} - y^k) - 2\mu_k(x^{k+1} - y^k)) - x^{k+1}\| \\
&\le \|\nabla f(x^{k+1}) - \nabla f(y^k)\| + 2(\mu_k + M)\|x^{k+1} - y^k\| + \kappa\|x^{k+1} - y^k\| \\
&\le (L_f + 2\bar{\mu} + M + \kappa)\|x^{k+1} - y^k\|.
\end{aligned}$$

So, (54) is proved.

Let us now prove (55). By contraction of projections and (4) we have that

$$\begin{aligned}
&\|P_{D_{k+1}}(x^{k+1} - \nabla f(x^{k+1})) - x^{k+1} - [P_{D_{k+1}}(y^k - \nabla f(y^k)) - y^k]\| \\
&\le \|P_{D_{k+1}}(x^{k+1} - \nabla f(x^{k+1})) - P_{D_{k+1}}(y^k - \nabla f(y^k))\| + \|x^{k+1} - y^k\| \\
&\le \|\nabla f(x^{k+1}) - \nabla f(y^k)\| + 2\|x^{k+1} - y^k\| \le (L_f + 2)\|x^{k+1} - y^k\|.
\end{aligned}$$

Then, (55) follows form (54).

Finally, (56) and (57) follow from (48), (54), and (55). □

Lemma 3.4 states that the number of iterates such that the infeasibility is bigger than a given ϵ_{feas} is, at most, proportional to $1/\epsilon_{feas}$.

Lemma 3.4 *Suppose that Assumption 3.2 holds and $\epsilon_{feas} > 0$. Let N_{infeas} be the number of iterations of Algorithm 2.1 such that*

$$\|h(x^k)\| > \epsilon_{feas}.$$

Then,

$$N_{infeas} \leq \frac{\bar{h}}{\epsilon_{feas}}. \tag{58}$$

Moreover, the number of iterations such that $\|h(y^k)\| > \epsilon_{feas}$ is not bigger than $r\frac{\bar{h}}{\epsilon_{feas}}$.

Proof. By (45) we have that

$$\sum_{j=1}^{k} \|h(x^j)\| \leq \bar{h}.$$

Therefore, the number of iterations at which $\|h(x^j)\| > \epsilon_{feas}$ cannot be bigger than $\bar{h}/\epsilon_{feas}$. Analogously, by (46), the number of iterations of Algorithm 2.1 such that $\|h(y^k)\| > \epsilon_{feas}$ cannot be bigger than $r\frac{\bar{h}}{\epsilon_{feas}}$. □

In the following lemma we prove that the number of iterations for which the optimality measure is given than an arbitrarily given ϵ_{opt} is, at most, proportional to $1/\epsilon_{opt}^2$.

Lemma 3.5 *Suppose that Assumption 3.2 holds and $\epsilon_{opt} > 0$. Let N_{opt} be the number of iterations j of Algorithm 2.1 such that*

$$\|P_{D_{j+1}}(x^{j+1} - \nabla f(x^{j+1})) - x^{j+1}\| > \epsilon_{opt}. \tag{59}$$

Then,

$$N_{opt} \leq (L_f + 2\bar{\mu} + M + \kappa)^2 C_d \epsilon_{opt}^{-2}. \tag{60}$$

Moreover, if N_{opty} is the number of iterations such that

$$\|P_{D_{j+1}}(y^j - \nabla f(y^j)) - y^j\| > \epsilon_{opt}, \tag{61}$$

we have that

$$N_{opty} \leq (2 + 2L_f + 2\bar{\mu} + M + \kappa)^2 C_d \epsilon_{opt}^{-2}. \tag{62}$$

Proof. Assume that (59) holds for more than N_{opt} iterations. Then, there exists k such that among the k first iterations there are N_{opt} iterations that satisfy (60). Therefore,

$$\sum_{j=0}^{k} \|P_{D_{j+1}}(x^{j+1} - \nabla f(x^{j+1})) - x^{j+1}\|^2 > N_{opt}\epsilon_{opt}^2. \tag{63}$$

Therefore, by (56),

$$N_{opt}\epsilon_{opt}^2 < (L_f + 2\bar{\mu} + M + \kappa)^2 C_d.$$

This implies (60). Analogously, (62) follows from (57). □

The following theorem is the complexity result that says that the number of iterates with infeasibility bigger than ϵ_{feas} or projected gradient of f bigger than ϵ_{opt} cannot be bigger than a quantity proportional to $1/\epsilon_{feas} + 1/\epsilon_{opt}^2$.

Theorem 3.2 *Suppose that Assumption 3.2 holds, $\epsilon_{feas} > 0$ and $\epsilon_{opt} > 0$. Then:*

- *The number of iterations such that*

$$\|h(x^{j+1})\| > \epsilon_{feas} \text{ or } \|P_{D_{j+1}}(x^{j+1} - \nabla f(x^{j+1})) - x^{j+1}\| > \epsilon_{opt} \tag{64}$$

  is bounded by

$$\bar{h}\epsilon_{feas}^{-1} + (L_f + 2\bar{\mu} + M + \kappa)^2 C_d \epsilon_{opt}^{-2} + 1$$

  and there exists $c_1 \le O(1)$ such that the number of evaluations of h and the number of evaluations of f are bounded by $c_1[\bar{h}\epsilon_{feas}^{-1} + (L_f + 2\bar{\mu} + M + \kappa)^2 C_d \epsilon_{opt}^{-2} + 1]$.

- *The number of iterations such that*

$$\|h(y^j)\| > \epsilon_{feas} \text{ or } \|P_{D_{j+1}}(y^j - \nabla f(y^j)) - y^j\| > \epsilon_{opt} \tag{65}$$

  is bounded by

$$r\bar{h}\epsilon_{feas}^{-1} + (2 + 2L_f + 2\bar{\mu} + M + \kappa)^2 C_d \epsilon_{opt}^{-2} + 1$$

  and there exists $c_2 \le O(1)$ such that the number of evaluations of h and the number of evaluations of f are bounded by $c_2[\bar{h}\epsilon_{feas}^{-1} + (2 + 2L_f + 2\bar{\mu} + M + \kappa)^2 C_d \epsilon_{opt}^{-2} + 1]$.

Proof. Assume that k_1 iterations of Algorithm 2.1 are executed satisfying (64) and

$$k_1 > \bar{h}\epsilon_{feas}^{-1} + (L_f + 2\bar{\mu} + M + \kappa)^2 C_d \epsilon_{opt}^{-2}.$$

Therefore, for all $j = 0, 1, \ldots, k_1$ the iterate x^{j+1} was defined and at least one of the following two statements held:

$$\|h(x^{j+1})\| > \epsilon_{feas} \tag{66}$$

or

$$\|P_{D_{j+1}}(x^{j+1} - \nabla f(x^{j+1})) - x^{j+1}\| > \epsilon_{opt}. \tag{67}$$

By Lemma 3.4, (66) can occur at most in $\bar{h}\epsilon_{feas}^{-1}$ iterations. Moreover, by Lemma 3.5, we have that (67) can occur at most in $(L_f + 2\bar{\mu} + M + \kappa)^2 C_d \epsilon_{opt}^{-2}$ iterations. Therefore, the number of iterations at which at least one of the statements (66) or (67) hold cannot exceed $\bar{h}\epsilon_{feas}^{-1} + (L_f + 2\bar{\mu} + M + \kappa)^2 C_d \epsilon_{opt}^{-2} + 1$. This completes the first part of the proof. The bound on the number of evaluations of h follows from the number of iterations and Lemma 2.1 whereas the bound on the number of evaluations of f follows from the number of iterations and Lemma 3.2.

The second part of the thesis is proved in an entirely analogous way. □

The convergence theorem below is an immediate consequence of the complexity results.

Theorem 3.3 *Suppose that Assumption 3.2 holds and that Algorithm 2.1 does not stop by restoration failure. Then,*

$$\lim_{k\to\infty} \|h(x^k)\| = 0, \lim_{k\to\infty} \|P_{D_k}(x^k - \nabla f(x^k)) - x^k\| = 0. \tag{68}$$

$$\lim_{k\to\infty} \|h(y^k)\| = 0, \text{ and } \lim_{k\to\infty} \|P_{D_{k+1}}(y^k - \nabla f(y^k)) - y^k\| = 0. \tag{69}$$

Proof. Assume that the algorithm generates infinitely many iterations and that $\|h(x^k)\|$ does not converge to zero. This implies that there exists $\varepsilon > 0$ and an infinite set of indices K such that $\|h(x^k)\| > \varepsilon$ for all $k \in K$. Therefore (64) occurs infinitely many times with $\epsilon_{feas} = \varepsilon$. This is impossible by Theorem 3.2. In a similar way, we prove that $\|P_{D_k}(x^k - \nabla f(x^k)) - x^k\|$ converges to zero. Analogously, (69) is proved using (65). □

From Theorems 3.2 and 3.3 we see that a reasonable stopping criterion for the practical application of Algorithm 2.1 is

$$\|h(y^k)\| \leq \epsilon_{feas} \text{ and } \|P_{D_{k+1}}(y^k - \nabla f(y^k)) - y^k\| \leq \epsilon_{opt}. \tag{70}$$

The condition (70) is the natural stopping criterion associated to the Approximate Gradient Projection (AGP) optimality condition introduced in [36] and analyzed in [2] and [3].

The last theorem establishes that every limit point generated by Algorithm 2.1 satisfies the AGP optimality condition and, under AGP-regularity, satisfies the KKT optimality conditions.

Theorem 3.4 *Suppose that Assumption 3.2 holds and that Algorithm 2.1 does not stop by restoration failure. Let $y^* \in \Omega$ be a limit point of the sequence $\{y^k\}$ generated by the algorithm. Then, y^* is feasible ($\|h(y^*)\| = 0$) and satisfies the AGP optimality condition. Finally, if y^* satisfies the AGP-regularity condition, the KKT conditions hold at y^*.*

Proof. The first statement is a consequence of continuity and (69). KKT-fulfillment under AGP-regularity follows from [3]. □

4 Future research

Inexact Restoration methods have proved to be very efficient for solving problems in which there exist natural ways to restore feasibility, induced by the structure of constraints [1, 5, 10, 14, 15, 16, 24, 26, 28, 32, 37]. Different general-purpose implementations have been given in [8, 11, 29]. The fact that, according to the results of the present paper, our version of the IR algorithm enjoys pleasant complexity properties is quite stimulating for the search of more competitive general-purpose computer implementations. As frequently occurs, theoretical developments inspire algorithmic ideas. Clearly, the properties (15), (35), (19) and (20) play a main role in the proofs of complexity and convergence. This suggests that new developments could be possible focusing on those properties with the aim of their fulfillment by means of a single quadratic programming subproblem, opening the possibility of obtaining sequential quadratic programming (SQP) algorithms with satisfactory complexity properties.

References

[1] R. Andreani, S. L. C. Castro, J. L. Chela, A. Friedlander, and S. A. Santos, An inexact-restoration method for nonlinear bilevel programming problems, Computational Optimization and Applications 43, pp. 307-328, 2009.

[2] R. Andreani, G. Haeser, and J. M. Martínez, On sequential optimality conditions for smooth constrained optimization. Optimization 60, pp. 627-641, 2011.

[3] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva, A Cone-Continuity Constraint Qualification and algorithmic consequences, SIAM Journal on Optimization. 26, 96–110, 2016.

[4] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva, Strict Constraint Qualifications and Sequential Optimality Conditions for Constrained Optimization, Mathematics of Operations Research 43, No. 3, Published Online, 2018.

[5] R. Andreani, Ramírez, S. A. Santos, and L. D. Secchin, Bilevel optimization with a multiobjective problem in the lower level, Numerical Algorithms, 2018, online first https://doi.org/10.1007/s11075-018-0576-1

[6] E. Bergou, Y. Diouane, and S. Gratton, A line-search algorithm inspired by the adaptive cubic regularization framework, with a worst-case complexity, Computational Optimization and Applications 68, pp. 533–554, 2017.

[7] E. Bergou, Y. Diouane, and S. Gratton, A line-search algorithm inspired by the adaptive cubic regularization framework, with a worst-case complexity $O(\varepsilon^{-3/2})$, Technical Report, 2017.

[8] E. G. Birgin, L. F. Bueno, and J. M. Martínez, Assessing the reliability of general-purpose Inexact Restoration methods, Journal of Computational and Applied Mathematics 282, pp. 1–16, 2015.

[9] E. G. Birgin, J. L. Gardenghi, J. M. Martínez, S. A. Santos, and Ph. L. Toint, Evaluation complexity for nonlinear constrained optimization using unscaled KKT conditions and high-order models, SIAM Journal on Optimization 26, pp. 951-967, 2016.

[10] E. G. Birgin, N. Krejić, and J. M. Martínez, On the employment of Inexact Restoration for the minimization of functions whose evaluation is subject to errors, Mathematics of Computation 87, pp. 1307–1326, 2018.

[11] E. Birgin e J. M. Martínez, Local Convergence of an Inexact-Restoration Method and Numerical Experiments, Journal of Optimization Theory and Applications 127, pp. 229–247, 2005.

[12] E. G. Birgin and J. M. Martínez, The use of quadratic regularization with a cubic descent condition for unconstrained optimization, SIAM Journal on Optimization 27, pp. 1049–1074, 2017.

[13] E. G. Birgin and J. M. Martínez, On regularization and active-set methods with complexity for constrained optimization, SIAM Journal on Optimization 28, pp. 1367–1395, 2018.

[14] L. F. Bueno, A. Friedlander, J. M. Martínez, and F. N. Sobral, Inexact-Restoration method for Derivative-Free Optimization with smooth constraints, SIAM Journal on Optimization 23, pp. 1189-1213, 2013.

[15] L. F. Bueno, G. Haeser, and J. M. Martínez, A Flexible Inexact-Restoration Method for Constrained Optimization, Journal of Optimization Theory and Applications 165, pp. 188–208, 2015.

[16] L. F. Bueno, G. Haeser, and J. M. Martínez, An Inexact Restoration Approach to Optimization Problems with Multiobjective Constraints under Weighted-Sum Scalarization, Optimization Letters 10, pp. 1315-1325, 2016.

[17] C. Cartis, N. I. M. Gould, and Ph. L. Toint, On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization, SIAM Journal on Optimization 20, pp. 2833–2852, 2010.

[18] C. Cartis, N. I. M. Gould, and Ph. L. Toint, Adaptive cubic overestimation methods for unconstrained optimization. Part I: motivation, convergence and numerical results, Mathematical Programming 127, pp. 245–295, 2011.

[19] C. Cartis, N. I. M. Gould, and Ph. L. Toint, Adaptive cubic overestimation methods for unconstrained optimization. Part II: worst-case function-evaluation complexity, Mathematical Programming 130, pp. 295–319, 2011.

[20] C. Cartis, N. I. M. Gould, and Ph. L. Toint, On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming, SIAM Journal on Optimization pp. 1721–1739, 2011.

[21] C. Cartis, N. I. M. Gould, and Ph. L. Toint, Complexity bounds for second-order optimality in unconstrained optimization, Journal of Complexity, pp. 28–93, 2012.

[22] C. Cartis, N. I. M. Gould, and Ph. L. Toint, On the complexity of finding first-order critical points in constrained nonlinear optimization, Mathematical Programming 144, pp. 93–106, 2014.

[23] C. Cartis, N. I. M. Gould, and Ph. L. Toint, Universal regularization methods - varying the power, the smoothness and the accuracy, https://pdfs.semanticscholar.org/fd8e/c51dfd2eabd636eb8080b164aeda1deca59c.pdf.

[24] S. L. C. Castro, E. A. Fancello, and A. Friedlander, Inexact Restoration technique to solve bilevel programming problems, Revista del Instituto Chileno de investigación Operativa 1, pp. 23-34, 2010.

[25] F. E. Curtis, D. P. Robinson, and M. Samadi, A trust-region algorithm with a worst-case iteration complexity of $O(\varepsilon^{-3/2})$, Mathematical Programming 162, pp. 1–32, 2017.

[26] P. S. Ferreira, E. W. Karas, M. Sachine, and F. N. C. Sobral, Global convergence of a derivative-free inexact restoration filter algorithm for nonlinear programming, Optimization 66, pp. 271-292, 2017.

[27] A. Fischer and A. Friedlander, A new line search inexact restoration approach for nonlinear programming, Computational Optimization and Applications 46, pp. 333-346, 2010.

[28] J. B. Francisco, J. M. Martínez, L. Martínez, and F. I. Pisnitchenko, Inexact Restoration method for minimization problems arising in electronic structure calculations, Computational Optimization and Applications 50, pp. 555-590, 2011.

[29] M. A. Gomes-Ruggiero, J. M. Martínez, and S. A. Santos, Spectral projected gradient method with inexact restoration for minimization with nonconvex constraints, SIAM Journal on Scientific Computing 31, pp. 1628-1652, 2009.

[30] C. C. Gonzaga, E. W. Karas, and M. Vanti, A globally convergent filter method for Nonlinear Programming, SIAM Journal on Optimization 14, pp. 646–669, 2004.

[31] G. N. Grapiglia, J. Y. Yuan, and Y. X. Yuan, On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization, Mathematical Programming 152, pp. 491–520, 2015.

[32] E. W. Karas, E. A. Pilotta, and A. A. Ribeiro, Numerical comparison of merit function with filter criterion in inexact restoration algorithms using hard-spheres problems, Computational Optimization and Applications 44, pp. 427-441, 2009.

[33] J. M. Martínez, Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming, Journal of Optimization Theory and Applications 111, pp. 39-58, 2001.

[34] J. M. Martínez, On High-Order Model Regularization for Constrained Optimization, SIAM Journal on Optimization 27, pp. 2447-2458, 2017.

[35] J. M. Martínez and E. A. Pilotta, Inexact restoration algorithms for constrained optimization, Journal of Optimization Theory and Applications 104, pp. 135-163, 2000.

[36] J. M. Martínez and B. F. Svaiter, A practical optimality condition without constraint qualifications for nonlinear programming, Journal of Optimization Theory and Applications 118, pp. 117–133, 2003.

[37] L. M. Mito, O problema de cobertura via geometria algébrica convexa, M.Sc. Thesis, Department of Applied Mathematics, University of São Paulo, 2018.

[38] Yu. Nesterov and B. T. Polyak, Cubic regularization of Newton method and its global performance, Mathematical Programming 108, pp. 177–205, 2006.