SIAM J. OPTIM.
Vol. 26, No. 2, pp. 951–967

EVALUATION COMPLEXITY FOR NONLINEAR CONSTRAINED OPTIMIZATION USING UNSCALED KKT CONDITIONS AND HIGH-ORDER MODELS*

E. G. BIRGIN†, J. L. GARDENGHI†, J. M. MARTÍNEZ‡, S. A. SANTOS‡, AND PH. L. TOINT§

Abstract. The evaluation complexity of general nonlinear, possibly nonconvex, constrained optimization is analyzed. It is shown that, under suitable smoothness conditions, an ϵ-approximate first-order critical point of the problem can be computed in order $O(\epsilon^{1-2(p+1)/p})$ evaluations of the problem's functions and their first p derivatives. This is achieved by using a two-phase algorithm inspired by Cartis, Gould, and Toint [*SIAM J. Optim.*, 21 (2011), pp. 1721–1739; *SIAM J. Optim.*, 23 (2013), pp. 1553–1574]. It is also shown that strong guarantees (in terms of handling degeneracies) on the possible limit points of the sequence of iterates generated by this algorithm can be obtained at the cost of increased complexity. At variance with previous results, the ϵ-approximate first-order criticality is defined by satisfying a version of the KKT conditions with an accuracy that does not depend on the size of the Lagrange multipliers.

Key words. nonlinear programming, complexity, approximate KKT point

AMS subject classifications. 90C30, 65K05, 49M05, 49M37, 90C60, 68Q25

DOI. 10.1137/15M1031631

1. Introduction.

Complexity analysis of numerical nonlinear optimization is currently an active research area (see, for instance, [19, 20, 3, 10, 7, 9, 14, 17, 23, 16]). In this domain, the worst-case (first-order) evaluation complexity of general smooth nonlinear optimization (that is, the maximal number of evaluations of the problem's objective function, constraints, and their derivatives that is needed for obtaining an approximate first-order critical point) has been the subject of recent papers by Cartis, Gould, and Toint [9, 14, 13]. In the first two of these contributions, it is shown that a two-phase trust-region–based algorithm needs at most $O(\epsilon^{-2})$ evaluations of these functions (and their gradients) to compute an ϵ-approximate first-order solution, that is, either an ϵ-approximate scaled first-order critical point of the problem or, as expected barring global optimization, an infeasible approximate critical point of the constraints' violation. By "ϵ-approximate scaled first-order critical point" we mean a point satisfying the first-order Karush–Kuhn–Tucker (KKT) condition on Lagrange multipliers up to an accuracy which is proportional to the size of the multipliers. The use of second derivatives was subsequently investigated in [13], where it was shown that a similar two-phase algorithm needs at most $O(\epsilon^{-3/2})$ evaluations of the problem

*Received by the editors July 20, 2015; accepted for publication (in revised form) January 19, 2016; published electronically April 12, 2016. This work has been partially supported by the Brazilian agencies FAPESP (grants 2010/10133-0, 2013/03447-6, 2013/05475-7, 2013/07375-0, and 2013/23494-9) and CNPq (grants 304032/2010-7, 309517/2014-1, 303750/2014-6, and 490326/2013-7) and by the Belgian National Fund for Scientific Research (FNRS).

http://www.siam.org/journals/siopt/26-2/M103163.html

†Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, São Paulo, SP, Brazil (egbirgin@ime.usp.br, john@ime.usp.br).

‡Department of Applied Mathematics, Institute of Mathematics, Statistics, and Scientific Computing, University of Campinas, Campinas, SP, Brazil (martinez@ime.unicamp.br, sandra@ime.unicamp.br).

§Department of Mathematics, The University of Namur, 61, rue de Bruxelles, B-5000, Namur, Belgium (philippe.toint@unamur.be).

functions, gradients, and Hessians to compute a point satisfying similar conditions.

The purpose of this paper is to extend these results in two different ways. The first is to consider unscaled KKT conditions (i.e., where the size of the Lagrange multipliers does not appear explicitly in the accuracy of the approximate criticality condition), and the second is to examine what can be achieved if evaluation of derivatives up to order $p > 2$ is allowed. We show below that a two-phase algorithm needs a maximum number of evaluations of the problem's functions and derivatives up to order p ranging from $O(\epsilon^{1-2(p+1)/p})$ to $O(\epsilon^{1-3(p+1)/p})$ to produce an ϵ-approximate *unscaled* first-order critical point of the problem (or an infeasible approximate critical point of the constraints' violation), depending on the identified degeneracy level. The extension of the theory to arbitrary p finds its basis in a proposal [4] by the authors of the present paper which extends to high order the available evaluation complexity results for unconstrained optimization (see [12, 19, 20]).

The paper is organized as follows. Section 2 states the problem more formally, describes a class of algorithms for its solution, and discusses the proposed termination criteria. The convergence and worst-case evaluation complexity analysis is presented in section 3, and the complexity results are further discussed in section 4. Conclusions are finally outlined in section 5.

Notation: In what follows, $\|\cdot\|$ denotes the Euclidean norm and, if $v(x)$ is a vector function, $v(x)_+ \stackrel{\rm def}{=} \max[v(x), 0]$, where the maximum is taken componentwise. $\nabla v(x)$ will denote the gradient of a function v defined on $\mathbb{R}^n$ with respect to its variable x. The notation $[x]_j$ denotes the jth component of a vector x whenever the simpler notation x_j might lead to confusion.

2. The problem and a class of algorithms for its solution. We consider the optimization problem given by

$$\begin{array}{rl} \min\limits_{x\in\mathbb{R}^n} & f(x) \\ \text{s.t.} & c_{\mathcal{E}}(x) = 0, \\ & c_{\mathcal{I}}(x) \leq 0, \end{array} \tag{2.1}$$

where f is a function from $\mathbb{R}^n$ into $\mathbb{R}$ and $c_{\mathcal{E}}$ and $c_{\mathcal{I}}$ are functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $\mathbb{R}^q$, respectively. We will assume that all these functions are p times continuously differentiable. We now define, for all $x \in \mathbb{R}^n$, the infeasibility measure

$$\theta(x) \stackrel{\rm def}{=} \|c_{\mathcal{E}}(x)\|^2 + \|c_{\mathcal{I}}(x)_+\|^2, \tag{2.2}$$

and, for given $\upsilon > 0$, its associated level set

$$\mathcal{L}(\upsilon) \stackrel{\rm def}{=} \{x \in \mathbb{R}^n \mid \theta(x) \leq \upsilon\}.$$

Moreover, given $t \in \mathbb{R}$, we define

$$\Phi(x, t) \stackrel{\rm def}{=} \theta(x) + [f(x) - t]_+^2 \quad \text{for all } x \in \mathbb{R}^n,$$

where the scalar t is the *target*. We use the notation $\nabla\Phi(x, t)$ to denote the gradient of $\Phi(x, t)$ with respect to x.

The FTarget (feasibility and target following) algorithm defined on the following page is inspired by that proposed in [13] and computes a sequence of iterates x_k by means of *outer iterations* indexed by $k = 0, 1, 2, \ldots$. For obtaining the outer iterate x_{k+1}, the FTarget algorithm uses a given monotonic (inner) unconstrained minimization (UM) algorithm which computes *inner iterations* indexed by $j = 0, 1, 2, \ldots$

by using derivatives of its objective function up to order p. Consequently, inner iterates will be denoted by $x_{k,j}$.

ALGORITHM 2.1 (the FTARGET algorithm).

Input: Let $\epsilon_{\text{P}} \in (0,1]$, a primal accuracy threshold, and $\epsilon_{\text{D}} \in (0,1]$, a dual one, be given. Let $\rho \in (0,1)$, an $x_{\text{-}1} \in \mathbb{R}^n$. Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous and nondecreasing function such that $\psi(0) = 0$. Let p be the order of available derivatives of the functions f, $c_{\mathcal{E}}$, and $c_{\mathcal{I}}$.

PHASE 1: *Computing an approximately feasible initial guess.*

Step F1. Minimize $\theta(x)$ using the UM algorithm starting from $x_{\text{-}1,0} = x_{\text{-}1}$, to compute an iterate $x_{\text{-}1,j}$, $j \in \{0,1,2,\dots\}$, such that $\theta(x_{\text{-}1,j}) \le \theta(x_{\text{-}1,0})$ and such that $x_{\text{-}1,j}$ satisfies at least one of the following conditions:

$$\theta(x_{\text{-}1,j}) \le 0.99\,\epsilon_{\text{P}}^2, \tag{2.3}$$

$$\theta(x_{\text{-}1,j}) > 0.99\,\epsilon_{\text{P}}^2, \quad \text{and} \quad \|\nabla\theta(x_{\text{-}1,j})\| \le \psi(\epsilon_{\text{D}}). \tag{2.4}$$

Define $x_0 = x_{\text{-}1,j}$.

Step F2. If $\theta(x_0) > 0.99\,\epsilon_{\text{P}}^2$ and $\|\nabla\theta(x_0)\| \le \psi(\epsilon_{\text{D}})$, **stop** Algorithm 2.1, returning $x_\epsilon = x_0$.

PHASE 2: *Improving dual feasibility (target following).*

Step T0. Initialize $k \leftarrow 0$.

Step T1. Compute $t_k = f(x_k) - \sqrt{\epsilon_{\text{P}}^2 - \theta(x_k)}$.

Step T2. Minimize $\Phi(x, t_k)$ using the UM algorithm starting from $x_{k,0} = x_k$, to compute an iterate $x_{k,j}$, $j \in \{0,1,2,\dots\}$, such that $\Phi(x_{k,j}, t_k) \le \Phi(x_{k,0}, t_k) = \epsilon_{\text{P}}^2$ and such that $x_{k,j}$ satisfies at least one of the following conditions:

$$f(x_{k,j}) \le t_k + \rho(f(x_k) - t_k) \quad \text{and} \quad \theta(x_{k,j}) \le 0.99\,\epsilon_{\text{P}}^2, \tag{2.5}$$

$$f(x_{k,j}) > t_k \quad \text{and} \quad \|\nabla\Phi(x_{k,j}, t_k)\| \le 2\epsilon_{\text{D}}[f(x_{k,j}) - t_k]_+, \tag{2.6}$$

$$\theta(x_{k,j}) > 0.99\,\epsilon_{\text{P}}^2 \quad \text{and} \quad \|\nabla\theta(x_{k,j})\| \le \psi(\epsilon_{\text{D}}). \tag{2.7}$$

Define $x_{k+1} = x_{k,j}$.

Step T3. If $\theta(x_{k+1}) > 0.99\,\epsilon_{\text{P}}^2$ and $\|\nabla\theta(x_{k+1})\| \le \psi(\epsilon_{\text{D}})$, **stop** Algorithm 2.1, returning $x_\epsilon = x_{k+1}$.

Step T4. If $f(x_{k+1}) > t_k$ and $\|\nabla\Phi(x_{k+1}, t_k)\| \le 2\epsilon_{\text{D}}[f(x_{k+1}) - t_k]_+$, **stop** Algorithm 2.1, returning $x_\epsilon = x_{k+1}$.

Step T5. Set $k \leftarrow k+1$, and go to Step T1.

It is clear that the FTARGET algorithm is in fact a class of algorithms depending on the specific choices of the derivative order $p \ge 1$ and on the UM minimizer adapted to this choice.

2.1. The meaning of the stopping criteria. We now discuss the nature of the point returned by the FTARGET algorithm as a function of the stopping criterion activated. We start by outlining the main points, leaving a detailed discussion for the following subsections.

Stopping at Step T4 means that, in the limit, an *approximate KKT* (AKKT) point has been found. Stopping at Step F2 has two possible meanings: (i) an infeasible stationary point of the infeasibility may be identified in the limit; or (ii) a situation analogous to that represented by stopping at Step T3 has happened. The interpretation of stopping at Step T3 depends on a weak condition on the tolerances

ϵ_{P} and ϵ_{D} in the limit and, more significantly, on the choice of the function ψ. For three different choices of ψ, it will be shown that stopping at Step T3 means that (i) an ϵ_{P}-feasible point z has been found such that the gradients of active constraints at z are not uniformly linear independent (with nonnegative coefficients for the inequality constraints); (ii) a feasible point which does not satisfy the Mangasarian–Fromowitz constraint qualification exists in the limit; or (iii) a feasible point which does not satisfy the Łojasiewicz inequality exists in the limit. (These claims are precisely discussed below.) We may therefore conclude globally that, for the three considered choices of ψ, FTarget always finds an unscaled approximate KKT point under suitable "nondegeneracy" assumptions.

2.1.1. Terminating at Step T4. When the FTarget algorithm stops at iteration k because the stopping criterion is satisfied at Step T4, it returns $x_\epsilon = x_{k+1}$ such that

$$\begin{aligned} \sqrt{\theta(x_{k+1})} &\leq \epsilon_{\text{P}}, \\ \left\| \nabla f(x_{k+1}) + \sum_{j=1}^{m} \lambda_j [\nabla c_{\mathcal{E}}(x_{k+1})]_j + \sum_{j=1}^{q} \mu_j [\nabla c_{\mathcal{I}}(x_{k+1})]_j \right\| &\leq \epsilon_{\text{D}}, \end{aligned} \tag{2.8}$$

where

$$\lambda_j = \frac{[c_{\mathcal{E}}(x_{k+1})]_j}{f(x_{k+1}) - t_k} \text{ for } j = 1, \ldots, m \quad \text{and} \quad \mu_j = \frac{[c_{\mathcal{I}}(x_{k+1})_+]_j}{f(x_{k+1}) - t_k} \text{ for } j = 1, \ldots, q.$$

Note also that $f(x_{k+1}) - t_k > 0$ in the two expressions above and, hence, that the multipliers λ_j and μ_j are well defined. We thus have that if $[c_{\mathcal{I}}(x_{k+1})]_j \leq 0$, then $\mu_j = 0$ and, hence,

$$\min [\mu_j, -[c_{\mathcal{I}}(x_{k+1})]_j] = \mu_j = 0.$$

On the other hand, if $[c_{\mathcal{I}}(x_{k+1})]_j > 0$, then

$$\min [\mu_j, -[c_{\mathcal{I}}(x_{k+1})]_j] = -[c_{\mathcal{I}}(x_{k+1})]_j \leq \epsilon_{\text{P}}. \tag{2.9}$$

We may then conclude that complementarity (as measured with the min function) is satisfied with precision ϵ_{P}. Note that the accuracies in the right-hand side of the second inequality in (2.8) and in (2.9) do not involve the size of the Lagrange multipliers, in contrast with the termination rule used in [13, 14].

In asymptotic terms, if we assume that the FTarget algorithm is run infinitely many times with $\epsilon_{\text{P}} = \epsilon_{\text{P},\ell} \to 0$ and $\epsilon_{\text{D}} = \epsilon_{\text{D},\ell} \to 0$, and that it stops infinitely many times $(\ell_1, \ell_2, \ldots)$ returning a point $x_\epsilon(\ell_i)$ $(i = 1, 2, \ldots)$ that satisfies the stopping criterion at Step T4, then we have that, for any accumulation point z of the sequence $\{x_\epsilon(\ell_i)\}$,

$$\lim_{i_s \to \infty} x_\epsilon(\ell_{i_s}) = z \tag{2.10}$$

for some subsequence $\{i_s\} \subseteq \{i\}$ and

$$\lim_{i_s \to \infty} \theta(x_\epsilon(\ell_{i_s})) = 0, \tag{2.11}$$

$$\lim_{i_s \to \infty} \left(\nabla f(x_\epsilon(\ell_{i_s})) + \sum_{j=1}^{m} \lambda_j [\nabla c_{\mathcal{E}}(x_\epsilon(\ell_{i_s}))]_j + \sum_{j=1}^{q} \mu_j [\nabla c_{\mathcal{I}}(x_\epsilon(\ell_{i_s}))]_j \right) = 0, \tag{2.12}$$

and

$$\lim_{i_s \to \infty} \min\left[\mu_j, -[c_{\mathcal{I}}(x_\epsilon(\ell_{i_s}))]_j\right] = 0. \tag{2.13}$$

Since relations (2.10)–(2.13) define an AKKT point (see [1, 2, 21] or [5, pp. 16–17]), z is therefore such a point.

2.1.2. Terminating at Step F2 or Step T3. Some explanations regarding the choice and interpretation of the function ψ are now in order. Note that ψ is used in the algorithm to stop the execution (at Steps F2 or T3) when $\theta(x)$ is "large" (i.e., $\theta(x) \geq 0.99\,\epsilon_{\text{P}}^2$) and $\nabla\theta(x)$ is "small" (i.e., $\|\nabla\theta(x)\| \leq \psi(\epsilon_{\text{D}})$). At a first glance, these occurrences seem to indicate that x is an "approximate infeasible stationary point of $\theta(x)$." However, a more careful analysis of particular cases reveals some subtleties.

When the FTARGET algorithm stops at Step F2, it returns a point x_ϵ such that

$$\theta(x_\epsilon) > 0.99\,\epsilon_{\text{P}}^2 \quad \text{and} \quad \|\nabla\theta(x_\epsilon)\| \leq \psi(\epsilon_{\text{D}}). \tag{2.14}$$

The analysis of these conditions is better done in asymptotic terms. Therefore, assume that the FTARGET algorithm is executed infinitely many times with $\epsilon_{\text{P}} = \epsilon_{\text{P},\ell} \to 0$ and $\epsilon_{\text{D}} = \epsilon_{\text{D},\ell} \to 0$. Moreover, assume that it stops infinitely many times $\ell_1, \ell_2, \ldots$ at Step F2, returning $x_\epsilon(\ell_i)$ $(i = 1, 2, \ldots)$ such that $\lim_{i\to\infty} x_\epsilon(\ell_i) = z$. Clearly, by (2.14), we have that

$$\lim_{i \to \infty} \|\nabla\theta(x_\epsilon(\ell_i))\| = 0. \tag{2.15}$$

If, in addition, $\theta(x_\epsilon(\ell_i))$ is bounded away from zero, by (2.15), then stopping at Step F2 means that the infeasible point z is a stationary point of the infeasibility measure θ. If $\lim_{i\to\infty} \theta(x_\epsilon(\ell_i)) = 0$, then the interpretation of this fact depends on the choice of function ψ and follows exactly the same analysis that we now investigate for the case where the FTARGET algorithm stops at Step T3.

When the FTARGET algorithm stops at Step T3, it returns a point x_ϵ such that

$$0.99\,\epsilon_{\text{P}}^2 < \theta(x_\epsilon) \leq \epsilon_{\text{P}}^2 \quad \text{and} \quad \|\nabla\theta(x_\epsilon)\| \leq \psi(\epsilon_{\text{D}}), \tag{2.16}$$

whose interpretation clearly depends on the choice of ψ. We now consider functions ψ of the form

$$\psi(\epsilon_{\text{D}}) = \sigma_1 \epsilon_{\text{D}}^{\sigma_2} \tag{2.17}$$

with $\sigma_1 > 0$ and three possible choices: (a) $\sigma_2 = 1$, (b) $\sigma_2 \in (1, 2)$, and (c) $\sigma_2 = 2$.

In case (a), (2.16) implies that

$$\frac{\|\nabla\theta(x_\epsilon)\|}{\sqrt{\theta(x_\epsilon)}} < \frac{\sigma_1}{\sqrt{0.99}} \frac{\epsilon_{\text{D}}}{\epsilon_{\text{P}}}. \tag{2.18}$$

Note that, by the definition (2.2) of θ,

$$\frac{\|\nabla\theta(x_\epsilon)\|}{\sqrt{\theta(x_\epsilon)}} = 2 \left\| \sum_{j=1}^{m} \lambda_j [\nabla c_{\mathcal{E}}(x_\epsilon)]_j + \sum_{j=1}^{q} \mu_j [\nabla c_{\mathcal{I}}(x_\epsilon)]_j \right\|, \tag{2.19}$$

where

$$\lambda_j = \frac{[c_{\mathcal{E}}(x_\epsilon)]_j}{\sqrt{\|c_{\mathcal{E}}(x_\epsilon)\|^2 + \|c_{\mathcal{I}}(x_\epsilon)_+\|^2}} \quad \text{for } j = 1, \ldots, m \tag{2.20}$$

and

$$\mu_j = \frac{[c_{\mathcal{I}}(x_\epsilon)_+]_j}{\sqrt{\|c_{\mathcal{E}}(x_\epsilon)\|^2 + \|c_{\mathcal{I}}(x_\epsilon)_+\|^2}} \quad \text{for } j = 1, \ldots, q. \tag{2.21}$$

Moreover, by (2.20) and (2.21), we have that

$$\sum_{j=1}^{m} \lambda_j^2 + \sum_{j=1}^{q} \mu_j^2 = 1 \quad \text{and} \quad \mu_j = 0 \text{ whenever } [c_{\mathcal{I}}(x_\epsilon)]_j < 0. \tag{2.22}$$

A definition is now necessary.

DEFINITION 1. *Given $\xi > 0$, we say that x has ξ-uniformly positive linear independent gradients with respect to problem* (2.1) *if for all $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^q_+$ such that $\sum_{i=1}^m \lambda_i^2 + \sum_{j=1}^q \mu_j^2 = 1$ and $\mu_j = 0$ whenever $[c_{\mathcal{I}}(x)]_j < 0$ we have that*

$$\left\| \sum_{i=1}^{m} \lambda_i [\nabla c_{\mathcal{E}}(x)]_i + \sum_{j=1}^{q} \mu_j [\nabla c_{\mathcal{I}}(x)]_j \right\| \geq \xi.$$

This means that, by (2.18), the point x_ϵ returned by FTARGET has not ξ-uniformly positive linear independent gradients of active constraints with

$$\xi = \frac{\sigma_1}{2\sqrt{0.99}} \frac{\epsilon_{\text{D}}}{\epsilon_{\text{P}}}.$$

If $\sigma_1 \epsilon_{\text{D}} / \epsilon_{\text{P}}$ is small enough, this indicates that an approximate Mangasarian–Fromovitz constraint qualification with tolerance ξ is not satisfied at x_ϵ.

The analysis of cases (b) and (c) must also be done in asymptotic terms. Therefore, assume again that the FTARGET algorithm is executed infinitely many times with $\epsilon_{\text{P}} = \epsilon_{\text{P},\ell} \to 0$ and $\epsilon_{\text{D}} = \epsilon_{\text{D},\ell} \to 0$. Moreover, assume that it stops infinitely many times $\ell_1, \ell_2, \ldots$ at Step T3 returning $x_\epsilon(\ell_i)$ $(i = 1, 2, \ldots)$ such that $\lim_{i\to\infty} x_\epsilon(\ell_i) = z$. Let us assume, for the remainder of this section, that

$$\omega \stackrel{\text{def}}{=} \limsup_{i\to\infty} \frac{\epsilon_{\text{D},\ell_i}}{\epsilon_{\text{P},\ell_i}} < \infty. \tag{2.23}$$

In cases (b) and (c), by (2.16), we have that

$$\frac{\|\nabla\theta(x_\epsilon(\ell_i))\|}{\sqrt{\theta(x_\epsilon(\ell_i))}} < \frac{\sigma_1}{\sqrt{0.99}} \frac{\epsilon_{\text{D},\ell_i}^{\sigma_2}}{\epsilon_{\text{P},\ell_i}} \quad \text{for } i = 1, 2, \ldots. \tag{2.24}$$

Taking limits in (2.24), and using (2.23) and the fact that $\sigma_2 > 1$, we have, by (2.19)–(2.22), that the gradients of active constraints at z are not positively linearly independent, so the Mangasarian–Fromowitz constraint qualification does not hold at the feasible point z [22].

As a consequence, if the feasible set is compact and all the feasible points satisfy the Mangasarian–Fromowitz constraint qualification, then, for i large enough, stopping at T3 is impossible in view of (2.23). Therefore, Phase 2 of the FTARGET algorithm can only stop at T4, identifying an AKKT point in the limit.

In order to further analyze case (c), we define the Łojasiewicz [18] inequality.

DEFINITION 2. *A continuously differentiable function $v : \mathbb{R}^n \to \mathbb{R}$ satisfies the Łojasiewicz inequality at $\bar{x}$ if there exist $\delta > 0$, $\tau \in (0, 1)$, and $\kappa > 0$ such that, for all $x \in B(\bar{x}, \delta)$,*

$$|v(x) - v(\bar{x})|^\tau \leq \kappa \|\nabla v(x)\|. \tag{2.25}$$

The properties of functions that satisfy the inequality (2.25) have been studied in several recent papers in connection with minimization methods, complexity theory, asymptotic analysis of partial differential equations, tame optimization, and the fulfillment of AKKT conditions in augmented Lagrangian methods [2, 6]. Smooth functions satisfy this inequality under fairly weak conditions. For example, analytic functions satisfy the Łojasiewicz inequality (see [18]).

In case (c), by (2.16), we have that

$$\frac{\|\nabla\theta(x_\epsilon(\ell_i))\|}{\theta(x_\epsilon(\ell_i))} < \frac{\sigma_1}{0.99}\left(\frac{\epsilon_{\mathrm{D},\ell_i}}{\epsilon_{\mathrm{P},\ell_i}}\right)^2 \qquad \text{for } i = 1, 2, \ldots.$$

We therefore obtain that, for arbitrary $\tau \in (0, 1)$,

$$\theta(x_\epsilon(\ell_i))^\tau = \theta(x_\epsilon(\ell_i))^{\tau-1}\theta(x_\epsilon(\ell_i)) > \theta(x_\epsilon(\ell_i))^{\tau-1}\left[\frac{\sigma_1}{0.99}\left(\frac{\epsilon_{\mathrm{D},\ell_i}}{\epsilon_{\mathrm{P},\ell_i}}\right)^2\right]^{-1}\|\nabla\theta(x_\epsilon(\ell_i))\|.$$

Since, by (2.16), $\theta(x_\epsilon(\ell_i)) \leq \epsilon^2_{\mathrm{P},\ell_i}$, then $\theta(x_\epsilon(\ell_i)) \to 0$ when i tends to infinity. As a consequence, $\theta(x_\epsilon(\ell_i))^{\tau-1}$ tends to infinity because $\tau \in (0, 1)$, and thus, using (2.23), for every $\delta > 0$ and $\kappa > 0$ there exists i sufficiently large such that

$$x_\epsilon(\ell_i) \in \mathcal{B}(z, \delta) \quad \text{and} \quad \theta(x_\epsilon(\ell_i))^\tau > \kappa\, \|\nabla\theta(x_\epsilon(\ell_i))\|.$$

This implies that the function $\theta(\cdot)$ does not satisfy the Łojasiewicz inequality at any possible limit point z. Therefore, if we assume that the function $\theta(\cdot)$ satisfies the Łojasiewicz inequality at every feasible point and that i is large enough, then, in view of (2.23), Phase 2 of the FTARGET algorithm can only stop at Step T4, returning an AKKT point.

3. Finite termination, convergence, and complexity. In this section we will prove that the FTARGET algorithm is well defined and terminates in a finite number of iterations, provided that the UM algorithm employed to minimize $\theta(x)$ at Phase 1 and $\Phi(x, t_k)$ at Phase 2 possesses standard convergence properties. Moreover, if the UM algorithm also enjoys suitable evaluation complexity properties, we can establish complexity bounds for the FTARGET algorithm itself.

Assumption A1. There exists a constant f_{low} such that $f(x) \geq f_{\text{low}}$ for all $x \in \mathcal{L}(\epsilon^2_{\mathrm{P}})$.

Assumption A2. There exists a constant κ_∇ such that $\|\nabla f(x)\| \leq \kappa_\nabla$ for all $x \in \mathcal{L}(\epsilon^2_{\mathrm{P}})$.

In order to simplify notation below, we also assume, without loss of generality, that

$$\kappa_\nabla \geq \max\left[1, \frac{5\sigma_1}{2\rho}\right]. \tag{3.1}$$

Assumption A1 is enough to prove a first result that is essential for further analysis of convergence and complexity. Lemma 3.1 shows that, independent of the UM

algorithm used for unconstrained minimizations, the number of outer iterations computed by the FTARGET algorithm cannot exceed a bound that depends only on ϵ_{P} and the lower bound of f.

LEMMA 3.1. *Suppose that Assumption* A1 *holds. Then the* FTARGET *algorithm stops at Phase* 1 *or it performs, at most,*

$$\left\lfloor \frac{f(x_0) - f_{\text{low}}}{0.1(1-\rho)\epsilon_P} \right\rfloor + 1$$

outer iterations at Phase 2.

Proof. Assume that the FTARGET algorithm does not stop at Phase 1. By the definition of this algorithm, an outer iteration is launched only when $\theta(x_k) \leq 0.99\,\epsilon_{\text{P}}^2$ and is followed by other outer iteration only when

$$f(x_{k+1}) \leq t_k + \rho(f(x_k) - t_k) \quad \text{and} \quad \theta(x_{k+1}) \leq 0.99\,\epsilon_{\text{P}}^2.$$

By the definition of t_k, this implies that

$$\begin{aligned} f(x_{k+1}) &\leq t_k + \rho\sqrt{\epsilon_{\text{P}}^2 - \theta(x_k)} \\ &= f(x_k) - \sqrt{\epsilon_{\text{P}}^2 - \theta(x_k)} + \rho\sqrt{\epsilon_{\text{P}}^2 - \theta(x_k)} \\ &= f(x_k) - (1-\rho)\sqrt{\epsilon_{\text{P}}^2 - \theta(x_k)}. \end{aligned}$$

Thus, since $\theta(x_k) \leq 0.99\,\epsilon_{\text{P}}^2$,

$$f(x_{k+1}) \leq f(x_k) - (1-\rho)\sqrt{\epsilon_{\text{P}}^2 - 0.99\,\epsilon_{\text{P}}^2} = f(x_k) - 0.1(1-\rho)\epsilon_{\text{P}}.$$

In other words, when an outer iteration is completed satisfying (2.5), we obtain a decrease of at least $0.1(1-\rho)\epsilon_{\text{P}}$ in the objective function. Then, since $f(x) \geq f_{\text{low}}$ for all $x \in \mathcal{L}(\epsilon_{\text{P}}^2)$ by Assumption A1, the number of outer iterations that are completed satisfying (2.5) must be smaller than or equal to

$$\left\lfloor \frac{f(x_0) - f_{\text{low}}}{0.1(1-\rho)\epsilon_{\text{P}}} \right\rfloor.$$

We therefore obtain the desired result by adding a final iteration at which (2.5) may not hold. ∎

Observe that, according to this lemma, the number of outer iterations performed by the FTARGET algorithm could be only one, a case that occurs if ϵ_{P} is sufficiently large and, consequently, the first target t_0 is very low.

Now we need to prove that, once an outer iteration is launched, it can be completed in a finite number of inner iterations. The following lemma establishes that, if $\theta(x_k) \leq 0.99\,\epsilon_{\text{P}}^2$ and the UM algorithm does not stop at $x_{k,j}$, the gradient norm $\|\nabla\Phi(x_{k,j}, t_k)\|$ is bounded away from zero by a quantity that depends only on ϵ_{P}, ϵ_{D}, and a bound on the norm of the gradient of f on $\mathcal{L}(\epsilon_{\text{P}}^2)$.

LEMMA 3.2. *Suppose that Assumption* A2 *holds, that* $\theta(x_k) \leq 0.99\,\epsilon_P^2$, *and that the* UM *algorithm used for minimizing* $\Phi(x, t_k)$ *does not stop at the inner iterate* $x_{k,j}$. *Then,*

$$\|\nabla\Phi(x_{k,j}, t_k)\| \geq \min\left[0.2\rho\,\epsilon_D\,\epsilon_P,\ \frac{\psi(\epsilon_D)}{2},\ \frac{\psi(\epsilon_D)\epsilon_D}{2\kappa_\nabla}\right]. \tag{3.2}$$

Proof. Since the UM algorithm does not stop at $x_{k,j}$, we have that none of the conditions (2.5), (2.6), and (2.7) holds at $x_{k,j}$. We will consider two cases:

$$f(x_{k,j}) > t_k + \rho(f(x_k) - t_k) \tag{3.3}$$

and

$$f(x_{k,j}) \le t_k + \rho(f(x_k) - t_k). \tag{3.4}$$

Consider the case (3.3) first. Then, $f(x_{k,j}) > t_k$. Since (2.6) does not hold, we have that

$$\|\nabla\Phi(x_{k,j}, t_k)\| > 2\,\epsilon_{\mathrm{D}}\,(f(x_{k,j}) - t_k). \tag{3.5}$$

Now, by (3.3) and the definition of t_k, since $\theta(x_k) \le 0.99\,\epsilon_{\mathrm{P}}^2$,

$$f(x_{k,j}) - t_k > \rho(f(x_k) - t_k) = \rho\sqrt{\epsilon_{\mathrm{P}}^2 - \theta(x_k)} \ge 0.1\rho\,\epsilon_{\mathrm{P}}.$$

Therefore, by (3.5),

$$\|\nabla\Phi(x_{k,j}, t_k)\| > 0.2\rho\,\epsilon_{\mathrm{D}}\,\epsilon_{\mathrm{P}}. \tag{3.6}$$

Now consider the case in which (3.4) holds. Then, $\theta(x_{k,j}) > 0.99\,\epsilon_{\mathrm{P}}^2$; otherwise (2.5) would have been satisfied. Thus, since (2.7) does not hold, we also have that

$$\begin{aligned}
\psi(\epsilon_{\mathrm{D}}) &< \|\nabla\theta(x_{k,j})\| \\
&= \|\nabla\theta(x_{k,j}) + 2(f(x_{k,j}) - t_k)_+\nabla f(x_{k,j}) - 2(f(x_{k,j}) - t_k)_+\nabla f(x_{k,j})\| \\
&\le \|\nabla\theta(x_{k,j}) + 2(f(x_{k,j}) - t_k)_+\nabla f(x_{k,j})\| + 2\|\nabla f(x_{k,j})\|(f(x_{k,j}) - t_k)_+ \\
&\le \|\nabla\Phi(x_{k,j}, t_k)\| + 2\kappa_\nabla(f(x_{k,j}) - t_k)_+,
\end{aligned}$$

where we have used the monotonicity of the UM algorithm (itself implying that $x_{k,j} \in \mathcal{L}(\epsilon_{\mathrm{P}}^2)$) and Assumption A2 to derive the last inequality. This yields that

$$\|\nabla\Phi(x_{k,j}, t_k)\| > \psi(\epsilon_{\mathrm{D}}) - 2\kappa_\nabla(f(x_{k,j}) - t_k)_+. \tag{3.7}$$

We now consider two cases. In the first one,

$$2\kappa_\nabla(f(x_{k,j}) - t_k)_+ \le \frac{\psi(\epsilon_{\mathrm{D}})}{2}. \tag{3.8}$$

Then, by (3.7),

$$\|\nabla\Phi(x_{k,j}, t_k)\| \ge \frac{\psi(\epsilon_{\mathrm{D}})}{2}.$$

In the second case, i.e., if (3.8) is not true, we have that

$$2(f(x_{k,j}) - t_k)_+ = 2(f(x_{k,j}) - t_k) > \frac{\psi(\epsilon_{\mathrm{D}})}{2\kappa_\nabla}.$$

Then, since (2.6) does not hold,

$$\|\nabla\Phi(x_{k,j}, t_k)\| \ge 2\epsilon_{\mathrm{D}}(f(x_{k,j}) - t_k)_+ > \frac{\psi(\epsilon_{\mathrm{D}})\epsilon_{\mathrm{D}}}{2\kappa_\nabla}.$$

This completes the proof. ∎

The following assumption expresses that the monotonic UM algorithm enjoys sensible first-order convergence properties in that the sequence of iterates $\{x_j\}$ that it generates when applied to the minimization of a bounded-below smooth function contains a subsequence at which the gradient of this function converges to zero. We formulate this assumption in a way suitable for its application to the convergence of the FTarget algorithm.

Assumption A3. For an arbitrary $\varepsilon > 0$ one has that, if the UM algorithm is applied to the minimization of $\theta(x)$ or $\Phi(x,t)$ with respect to x, starting from an arbitrary initial point x_0, then, in a finite number of iterations, this algorithm finds an iterate x such that either $\theta(x) \leq \theta(x_0)$ and $\|\nabla\theta(x)\| \leq \varepsilon$, or $\Phi(x,t) \leq \Phi(x_0,t)$ and $\|\nabla\Phi(x,t)\| \leq \varepsilon$, respectively.

The following lemma establishes that, under Assumptions A2–A3 and given the iterate x_k of the FTarget algorithm, the iterate x_{k+1} is well defined and is computed in a finite number of iterations.

Lemma 3.3. *Suppose that Assumptions* A2–A3 *hold. Then, for all* $k = 0, 1, 2, \ldots$, *the* UM *algorithm applied to the minimization of* $\Phi(x, t_k)$ *finds a point* $x_{k,j}$ *that satisfies at least one of the criteria* (2.5), (2.6), *and* (2.7) *in a finite number of iterations.*

Proof. Define

$$\varepsilon = \min\left[0.2\rho\,\epsilon_{\mathrm{D}}\,\epsilon_{\mathrm{P}},\ \frac{\psi(\epsilon_{\mathrm{D}})}{2},\ \frac{\psi(\epsilon_{\mathrm{D}})\,\epsilon_{\mathrm{D}}}{2\kappa_{\nabla}}\right] > 0.$$

By Assumption A3, the UM algorithm eventually finds an iteration j_0 such that x_{k,j_0} satisfies $\|\nabla\Phi(x_{k,j_0}, t_k)\| \leq \varepsilon$. By Lemma 3.2 and the definition of ε, x_{k,j_0} satisfies (2.5), (2.6), or (2.7), and the iteration k of the FTarget algorithm terminates at some $x_{k,j}$ with $j \leq j_0$. ∎

Now we are ready to prove finite termination of the FTarget algorithm.

Theorem 3.4. *Suppose that Assumptions* A1–A3 *hold. Then, the* FTarget *algorithm stops at Step* F2 *of Phase* 1 *or after at most*

$$\left\lfloor \frac{f(x_0) - f_{\mathrm{low}}}{0.1(1-\rho)\epsilon_P} \right\rfloor + 1 \tag{3.9}$$

iterations of Phase 2 *with* x_{k+1} *satisfying the criteria of Step* T3 *or* T4.

Proof. By Assumption A3, the problem at Phase 1 is solved by the UM algorithm in a finite number of iterations. By Lemma 3.1, the FTarget algorithm cannot perform more than (3.9) outer iterations. By Lemma 3.3, each outer iteration is well-defined and terminates in finite time. Therefore, the FTarget algorithm stops at Step F2 of Phase 1, or, at the last outer iteration of Phase 2, x_{k+1} satisfies the criteria of Step T3 or T4. ∎

The following assumption prescribes an additional property of the UM algorithm. Whereas Assumption A3 says that, for any $\varepsilon > 0$, the UM algorithm finds a point that verifies $\|\nabla\theta(x)\| \leq \varepsilon$ or $\|\nabla\Phi(x,t)\| \leq \varepsilon$ in a finite number of iterations, Assumption A4 aims to quantify the number of function evaluations that are necessary to achieve a sufficiently small gradient.

Assumption A4. There exist $\alpha \geq 0$ and a constant $\kappa_\theta > 0$ (depending on properties of the functions $c_{\mathcal{E}}$ and $c_{\mathcal{I}}$ and on parameters of the UM algorithm) such that, given $\varepsilon > 0$, if the UM algorithm is applied to the minimization of $\theta(x)$ starting from

an arbitrary initial point x_{-1}, the algorithm finds an iterate x such that $\theta(x) \leq \theta(x_{-1})$ and $\|\nabla\theta(x)\| \leq \varepsilon$, employing, at most,

$$\kappa_\theta \left[\frac{\theta(x_{-1})}{\varepsilon^\alpha} \right]$$

evaluations of $c_{\mathcal{E}}$, $c_{\mathcal{I}}$, and their derivatives. Moreover, if, for any $t \in \mathbb{R}$, the UM algorithm is applied to the minimization of $\Phi(x,t)$ (with respect to x) starting from an arbitrary initial point x_0, there exists a constant $\kappa_\Phi > 0$ (depending on properties of the functions f, $c_{\mathcal{E}}$, and $c_{\mathcal{I}}$ and on parameters of the UM algorithm) such that the algorithm finds an iterate x such that $\Phi(x,t) \leq \Phi(x_0,t)$ and $\|\nabla\Phi(x,t)\| \leq \varepsilon$, employing, at most,

$$\kappa_\Phi \left[\frac{\Phi(x_0,t)}{\varepsilon^\alpha} \right]$$

evaluations of f, $c_{\mathcal{E}}$, $c_{\mathcal{I}}$, and their derivatives.

Note that the derivatives used in the UM algorithm may be approximated if necessary, provided that the approximation is sufficiently accurate to ensure the desired global rate of convergence. (See [11] or [23] for examples of suitable derivative-free algorithms.)

The constants κ_θ and κ_Φ mentioned in Assumption A4 depend on algorithmic parameters of the UM algorithm and on quantities associated with the objective function (θ or Φ) of the unconstrained minimization problem being solved. For example, if the UM algorithm is a typical first-order linesearch method and L_Φ is a Lipschitz constant for $\nabla\Phi(x,t)$ (for all t), we have that $\alpha = 2$ and

$$\kappa_\Phi = L_\Phi \, \kappa_{\text{UM}},$$

where κ_{UM} depends on algorithmic parameters, such as sufficient descent tolerances and angle condition constants (see [15], for instance). Note that the assumption that the same L_Φ may be a Lipschitz constant independently of t is plausible if one assumes the Lipschitz continuity of ∇f, $\nabla c_{\mathcal{E}}$, and $\nabla c_{\mathcal{I}}$. Similar conclusions hold if one considers the first-order adaptive cubic regularization method ARC [8] instead of a first-order linesearch method.

The situation is, however, slightly more complex if one wishes to exploit derivatives of order larger than one for obtaining better worst-case complexity bounds, as we now describe. We base our argument on a recent paper by Birgin et al. [4] where unconstrained optimization using high-order models is considered. It turns out from that paper that if one wishes to minimize $u(x) = v(x) + w(x)$ over $\mathbb{R}^n$ (where $v(x)$ is at least continuously differentiable and $w(x)$ is p times continuously differentiable with a Lipschitz continuous pth derivative), and if one is ready to supply derivatives of w up to order p, then a variant of the ARC method starting from $\bar{x}$ and using high-order models can be shown to produce an approximate first-order critical point ($\|\nabla u(x)\| \leq \varepsilon$) in a number of evaluations of $w(x)$, and its derivatives at most equal

$$\kappa_A \left[\frac{u(\bar{x}) - u_{\text{low}}}{\varepsilon^{(p+1)/p}} \right],$$

where u_{low} is a global lower bound on $u(x)$ and κ_A depends on the Lipschitz constant of the pth derivative of w, on p, and on algorithmic parameters only. Notice that the number of evaluations of $v(x)$ might be higher (because it is explicitly included in

the model and has to be evaluated, possibly with its first derivative, every time the model (and its derivative) is computed). In order to apply this technique, we now reformulate our initial problem (2.1) in the equivalent form

$$\begin{array}{rl} \min\limits_{x,y,z\in\mathbb{R}^{n+q+1}} & z \\ \text{s.t.} & \overline{c_{\mathcal{E}}}(x,y,z)=0, \\ & y\le 0, \end{array} \qquad \text{with} \quad \overline{c_{\mathcal{E}}}(x,y,z) \stackrel{\text{def}}{=} \left(\begin{array}{c} f(x)-z \\ c_{\mathcal{E}}(x) \\ c_{\mathcal{I}}(x)-y \end{array}\right), \tag{3.10}$$

and construct the associated $\Phi(x,y,z,t)$ function as

$$\Phi(x,y,z,t) = \underbrace{\|\overline{c_{\mathcal{E}}}(x,y,z)\|^2}_{w(x,y,z)} + \underbrace{\|y_+\|^2 + [z-t]_+^2}_{v(y,z)}.$$

Note that $v(y,z)$ is continuously differentiable and that the differentiability properties of f, $c_{\mathcal{E}}$, and $c_{\mathcal{I}}$ are transferred to $\overline{c_{\mathcal{E}}}(x,y,z)$: the Lipschitz continuity of the pth derivative of $\Phi(x,y,z,t)$ with respect to its first three variables being ensured if all derivatives of f, $c_{\mathcal{E}}$, and $c_{\mathcal{I}}$ are bounded up to order $p-1$ and Lipschitz continuous up to order p. Note also that the (very simple) evaluation of $v(y,z)$ does not involve any of the problem's functions or derivatives, and thus that the number of these evaluations does not affect the evaluation complexity of the FTARGET algorithm. In these conditions, we may then conclude that the high-order ARC algorithm presented in [4] satisfies Assumption A4 with κ_θ and κ_Φ equal to κ_A and $\alpha=(p+1)/p$.

The following result is a reformulation of Lemma 3.3 in which the number of evaluations employed by the UM algorithm for minimizing $\Phi(x,t_k)$ is quantified.

LEMMA 3.5. *Suppose that Assumptions* A1–A2 *and* A4 *hold. Then, for all* $k\ge 0$, *the* UM *algorithm applied to the minimization of* $\Phi(x,t_k)$ *finds, using no more than*

$$\left\lfloor \kappa_\Phi \left(\frac{2\kappa_\nabla}{\sigma_1}\right)^\alpha \epsilon_{_P}^2 \epsilon_{_D}^{-\alpha} \min[\epsilon_{_D}^{\sigma_2}, \epsilon_{_P}]^{-\alpha} \right\rfloor \tag{3.11}$$

evaluations of f, $c_{\mathcal{E}}$, $c_{\mathcal{I}}$, *and their derivatives, a point* $x_{k,j}$ *that satisfies at least one of the criteria* (2.5), (2.6), *and* (2.7).

Proof. Observing that $\Phi(x_{k,0},t_k)=\Phi(x_k,t_k)=\epsilon_{\rm P}^2$ and using Lemma 3.2 and Assumption A4, we deduce that the UM algorithm needs no more than

$$\kappa_\Phi \left\lceil \frac{\epsilon_{\rm P}^2}{\min\left[0.2\rho\,\epsilon_{\rm D}\,\epsilon_{\rm P},\ \frac{\psi(\epsilon_{\rm D})}{2},\ \frac{\psi(\epsilon_{\rm D})\,\epsilon_{\rm D}}{2\kappa_\nabla}\right]^\alpha} \right\rceil$$

evaluations to find an approximate minimizer. We obtain the desired bound using the definition (2.17) of ψ, (3.1), and the inequalities $\sigma_2\ge 1$, $\epsilon_{\rm P}\le 1$, and $\epsilon_{\rm D}\le 1$. ∎

Notice that, as in Assumption A4, the constant κ_Φ in (3.11) depends only on f, $c_{\mathcal{E}}$, $c_{\mathcal{I}}$, and the parameters of the UM algorithm.

It is now possible to prove a complexity result for the FTARGET algorithm.

THEOREM 3.6. *Suppose that Assumptions* A1–A2 *and* A4 *hold and that* $f(x)\le f_{\rm up}$ *for all* $x\in\mathcal{L}(\epsilon_{_P}^2)$. *Then the* FTARGET *algorithm stops at Step* F2 *of Phase* 1, *or there exists* $k\in\{0,1,2,\dots\}$ *such that* FTARGET *stops at iteration* k *of Phase* 2 *with* x_ϵ *satisfying the criteria of Step* T3 *or* T4. *Moreover, the* FTARGET *algorithm employs, in Phase* 1, *at most*

$$\left\lfloor \kappa_\theta \left[\frac{\theta(x_{-1})}{\sigma_1^\alpha}\right] \epsilon_{_D}^{-\sigma_2\alpha} \right\rfloor \tag{3.12}$$

evaluations of $c_{\mathcal{E}}$, $c_{\mathcal{I}}$, and their derivatives, and, in Phase 2, at most

$$\left\lfloor \left[\kappa_\Phi \left(\frac{2\kappa_\nabla}{\sigma_1} \right)^\alpha \left(\frac{f_{\text{up}} - f_{\text{low}}}{0.1(1-\rho)} + 1 \right) \right] \epsilon_P \epsilon_D^{-\alpha} \min[\epsilon_D^{\sigma_2}, \epsilon_P]^{-\alpha} \right\rfloor \tag{3.13}$$

evaluations of f, $c_{\mathcal{E}}$, $c_{\mathcal{I}}$, and their derivatives.

Proof. The first part of the theorem follows from the definition (2.17) of ψ and Theorem 3.4 since Assumption A4 implies Assumption A3. The bound on the total number of evaluations of Phase 1 follows directly from Assumption A4, while the bound

$$\kappa_\Phi \left[\left(\frac{2\kappa_\nabla}{\sigma_1} \right)^\alpha \epsilon_{\text{P}}^2 \epsilon_{\text{D}}^{-\alpha} \min[\epsilon_{\text{D}}^{\sigma_2}, \epsilon_{\text{P}}]^{-\alpha} \right] \left[\left\lfloor \frac{f(x_0) - f_{\text{low}}}{0.1(1-\rho)\,\epsilon_{\text{P}}} \right\rfloor + 1 \right]$$

on the total number of evaluations of Phase 2 follows from Lemmas 3.1 and 3.5. The bound (3.13) then follows from the assumption that $f(x) \leq f_{\text{up}}$ for all $x \in \mathcal{L}(\epsilon_{\text{P}}^2)$ and the fact that $\epsilon_{\text{P}} \leq 1$. ∎

Notice that, as in Assumption A4, the constant κ_θ in (3.12) depends only on $c_{\mathcal{E}}$, $c_{\mathcal{I}}$, and parameters of the UM algorithm, whereas the constant κ_Φ in (3.13) depends only on f, $c_{\mathcal{E}}$, $c_{\mathcal{I}}$, and parameters of the UM algorithm. Note also that, by introducing the upper bound f_{up} on $f(x)$ in the neighborhood $\mathcal{L}(\epsilon_{\text{P}}^2)$ of the feasible set, we have made the complexity bound independent of x_0 (which is not a datum of the problem nor an input parameter, but the result of applying the UM algorithm to the minimization of $\theta(x)$ starting from $x_{\text{-}1}$ at Phase 1 of the FTarget algorithm).

4. Discussion. Some comments are now useful to make the result of Theorem 3.6 more explicit. Table 4.1 below summarizes our convergence and complexity results for the case where

$$\epsilon = \epsilon_{\text{D}} = \epsilon_{\text{P}} \tag{4.1}$$

is small enough and for different choices of the UM algorithm and values of σ_2 in the definition (2.17) of ψ. Observe that, if (4.1) holds, the complexity of the two phases of Algorithm 2.1 reduces to

$$\text{Phase 1: } O\big(\epsilon^{-\sigma_2\alpha}\big), \qquad \text{Phase 2: } O\Big(\epsilon^{1-(1+\sigma_2)\alpha}\Big).$$

Note that, from the definition in (2.23), $\omega = 1$ in this case, making our discussion of section 2.1 relevant. Despite its appealing symmetry, the choice (4.1) is somewhat arbitrary, and variations in the complexities displayed in the table will result from different choices. (The alternative $\epsilon_{\text{D}} = \epsilon_{\text{P}}^{2/3}$ is suggested in [13], leading to a bound in $O(\epsilon_{\text{P}}^{-3/2})$ for the case where $p = 2$, $\sigma_2 = 1$, and $\alpha = 3/2$, but implying that $\omega = \infty$.) In particular, (4.1) implicitly assumes that the problem's scaling is reasonable. Note that the complexity of Phase 1 mentioned in the table is only informative since the overall complexity of the FTarget algorithm is given by the complexity of Phase 2 which dominates that of Phase 1 in all cases.

With $\sigma_2 = 2$, we obtain the maximal quality of the algorithmic results in the sense that the FTarget algorithm stops when an approximate KKT point (without scaling) is found or when a very weak property (Łojasiewicz inequality) does not hold in the limit. With $\sigma_2 = 1$, the FTarget algorithm stops when an approximate KKT point is found or when, in the limit, a relaxed Mangasarian–Fromowitz property with tolerance $\xi = \sigma_1/(2\sqrt{0.99})$ does not hold. Between those extremes, when $1 < \sigma_2 < 2$,

TABLE 4.1
Summary of convergence and complexity results for different choices of the UM *algorithm and function* $\psi(\epsilon) = \sigma_1 \epsilon^{\sigma_2}$ *with* $\sigma_1 > 0$ *and different choices for* $\sigma_2 \in [1, 2]$.

		$\psi(\epsilon) = \sigma_1 \epsilon^{\sigma_2}$		
		$\sigma_2 = 1$	$\sigma_2 \in (1, 2)$	$\sigma_2 = 2$
Convergence results		Approximate infeasible stationary point for $\epsilon \to 0$ (Phase 1), or approximate KKT point (Phase 2), or ...		
		No ξ-uniform positive linear independence of gradients of active constraints with $\xi = \sigma_1/(2\sqrt{0.99})$	MFCQ fails for $\epsilon \to 0$	Łojasiewicz fails for $\epsilon \to 0$
Assumption A4 holds for UM with	$\alpha = 2$ $(p = 1)$	Phase 1: $O(\epsilon^{-2})$ Phase 2: $O(\epsilon^{-3})$	Phase 1: $O(\epsilon^{-2\sigma_2})$ Phase 2: $O(\epsilon^{-1-2\sigma_2})$	Phase 1: $O(\epsilon^{-4})$ Phase 2: $O(\epsilon^{-5})$
	$\alpha = 3/2$ $(p = 2)$	Phase 1: $O(\epsilon^{-1.5})$ Phase 2: $O(\epsilon^{-2})$	Phase 1: $O(\epsilon^{-1.5\sigma_2})$ Phase 2: $O(\epsilon^{-0.5-1.5\sigma_2})$	Phase 1: $O(\epsilon^{-3})$ Phase 2: $O(\epsilon^{-3.5})$
	$\alpha = (p+1)/p$ $(p \geq 1)$	Phase 1: $O\left(\epsilon^{-\frac{(p+1)}{p}}\right)$ Phase 2: $O\left(\epsilon^{1-2\frac{(p+1)}{p}}\right)$	Phase 1: $O\left(\epsilon^{-\sigma_2\frac{(p+1)}{p}}\right)$ Phase 2: $O\left(\epsilon^{1-(1+\sigma_2)\frac{(p+1)}{p}}\right)$	Phase 1: $O\left(\epsilon^{-2\frac{(p+1)}{p}}\right)$ Phase 2: $O\left(\epsilon^{1-3\frac{(p+1)}{p}}\right)$

either the final point is an approximate KKT point or the full Mangasarian–Fromowitz property fails in the limit. As expected, the complexity goes in the opposite direction. The best complexity is obtained when $\sigma_2 = 1$, and the worst when $\sigma_2 = 2$. Also as expected, when the UM algorithm satisfies Assumption A4 with $\alpha = (p+1)/p$ $(p > 2)$, complexities are better than those with $\alpha = 2$ $(p = 1)$ and $\alpha = 3/2$ $(p = 2)$. The best complexity is obtained when α approaches 1 because p grows and $\sigma_2 = 1$.

We now compare the results obtained here with those obtained in [9, 13, 14] for $p = 1$ and $p = 2$, focusing as above on the case where $\epsilon = \epsilon_{\text{P}} = \epsilon_{\text{D}}$. In these contributions, the complexity of achieving scaled KKT conditions is considered, at variance with the unscaled approach used in the present paper. By scaled KKT conditions, we mean that (2.8) is modified to take (for a general approximate first-order critical triple (x, λ, μ)) the form

$$\begin{aligned} \sqrt{\theta(x)} &\leq \epsilon, \\ \left\| \nabla f(x) + \sum_{j=1}^{m} \lambda_j [\nabla c_{\mathcal{E}}(x)]_j + \sum_{j=1}^{q} \mu_j [\nabla c_{\mathcal{I}}(x)]_j \right\| &\leq \epsilon \, \|(1, \lambda, \mu)\|, \\ \sum_{j=1}^{q} \mu_j [c_{\mathcal{I}}(x)]_j &\leq 2\epsilon \, \|(1, \lambda, \mu)\|, \end{aligned} \tag{4.2}$$

where $\mu_j \leq 0$ for $j = 1, \ldots, q$.

It is shown in [9, 14] that, if f, $c_{\mathcal{E}}$, and $c_{\mathcal{I}}$ are continuously differentiable with Lipschitz continuous gradients, then a triple (x, λ, μ) can be found satisfying (4.2) or (2.4)/(2.7) in at most

$$\left\lfloor \kappa_1 \frac{\sqrt{\theta(x_{\text{-}1})} + f_{\text{up}} - f_{\text{low}}}{\epsilon^2} + \kappa_2 |\log \epsilon| + \kappa_3 \right\rfloor$$

evaluations of f, $c_{\mathcal{E}}$, and $c_{\mathcal{I}}$ (and their first derivatives), where κ_1, κ_2, and κ_3 are constants independent of ϵ. This result is one order better than the bound of $O\left(\epsilon^{-3}\right)$ evaluations reported in Table 4.1 for the case $\alpha = 2$ and $\sigma_2 = 1$, indicating that achieving scaled KKT conditions seems easier than achieving unscaled KKT conditions when using first derivatives only. We have focused on the case where $\sigma_2 = 1$ because the implications in terms of degeneracies for $\epsilon \to 0$ are not discussed in [9, 14], meaning that it is not clear whether the algorithms being compared declare failure in satisfying scaled or unscaled KKT conditions in the same situations.

The comparison is more difficult if we now allow the use of first and second derivatives, because the results in [13] are expressed using a different first-order criticality measure $\chi(x)$, whose value is the maximal decrease that is achievable on the linearized function under consideration in the intersection of the unit sphere and the feasible domain defined by positivity constraints on slack variables for inequalities (see [13] for details). This criticality measure is also used in scaled form, in that it has to be below $\epsilon \|(1, \lambda, \mu)\|$ when applied on the Lagrangian (in the spirit of (4.2)), or below ϵ when applied to $\sqrt{\theta(\cdot)}$. This last condition is conceptually similar to (2.4)/(2.7) but is significantly stronger because it involves the gradient of $\sqrt{\theta(\cdot)}$ rather than that of $\theta(\cdot)$; indeed, strong assumptions on the singular values of the constraints' Jacobians are needed to ensure their equivalence in order. In this context, it is shown that, if f, $c_{\mathcal{E}}$, and $c_{\mathcal{I}}$ are twice continuously differentiable with Lipschitz continuous gradients and Hessians, and if a possibly restrictive assumption on the subproblem solution holds (see [13] for details), then a triple (x, λ, μ) satisfying these alternative scaled criticality conditions can be obtained in at most

$$\left\lfloor \kappa_1 \frac{\sqrt{\theta(x_{\text{-}1})} + f_{\text{up}} - f_{\text{low}}}{\epsilon^2} + \kappa_2 \right\rfloor$$

evaluations of f, $c_{\mathcal{E}}$, and $c_{\mathcal{I}}$ (and their first and second derivatives), where κ_1 and κ_2 are again constants independent of ϵ. In contrast with the case where $p = 1$, this bound is now (in order) identical to the corresponding result in Table 4.1 ($\alpha = 3/2$, $\sigma_2 = 1$). Whether it could be improved to ensure that either (4.2) or (2.4)/(2.7) holds is an interesting open question.

5. Conclusions. We have presented worst-case evaluation complexity bounds for computing approximate first-order critical points of smooth constrained optimization problems. In contrast to previous bounds, these involve the unscaled Karush–Kuhn–Tucker optimality conditions and cover cases where high-order derivatives are used. As was the case in [9, 13, 14], the complexity bounds are obtained by applying a two-phase algorithm which first enforces approximate feasibility before improving optimality without deteriorating feasibility.

At this stage, the applicable nature of the two-phase algorithm is uncertain. While the specific version presented in this paper is very unlikely to be practical because it closely follows potentially nonlinear constraints, thereby enforcing possibly very short

steps, the question of whether more efficient variants of the idea can be made to work in practice remains to be explored.

REFERENCES

[1] R. Andreani, G. Haeser, and J. M. Martínez, *On sequential optimality conditions for smooth constrained optimization*, Optimization, 60 (2011), pp. 627–641.

[2] R. Andreani, J. M. Martínez, and B. F. Svaiter, *A new sequential optimality condition for constrained optimization and algorithmic consequences*, SIAM J. Optim., 20 (2010), pp. 3533–3554.

[3] S. Bellavia, C. Cartis, N. I. M. Gould, B. Morini, and Ph. L. Toint, *Convergence of a regularized Euclidean residual algorithm for nonlinear least-squares*, SIAM J. Numer. Anal., 48 (2010), pp. 1–29.

[4] E. G. Birgin, J. L. Gardenghi, J. M. Martínez, S. A. Santos, and Ph. L. Toint, *Worst-Case Evaluation Complexity for Unconstrained Nonlinear Optimization Using High-Order Regularized Models*, Technical Report naXys-05-2015, Namur Center for Complex Systems (naXys), University of Namur, Namur, Belgium, 2015.

[5] E. G. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, SIAM, Philadelphia, 2014.

[6] J. Bolte, A. Danilidis, O. Ley, and L. Mazet, *Characterizations of Łojasiewicz inequalities: Subgradient flows, talweg, convexity*, Trans. Amer. Math. Soc., 362 (2010), pp. 3319–3363.

[7] 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 problems*, SIAM J. Optim., 20 (2010), pp. 2833–2852.

[8] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *Adaptive cubic overestimation methods for unconstrained optimization. Part* II: *Worst-case function-evaluation complexity*, Math. Program. Ser. A, 130 (2011), pp. 295–319.

[9] 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 J. Optim., 21 (2011), pp. 1721–1739.

[10] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *An adaptive cubic regularization algorithm for nonconvex optimization with convex constraints and its function-evaluation complexity*, IMA J. Numer. Anal., 32 (2012), pp. 1662–1695.

[11] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *On the oracle complexity of first-order and derivative-free algorithms for smooth nonconvex minimization*, SIAM J. Optim., 22 (2012), pp. 66–86.

[12] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *On the complexity of finding first-order critical points in constrained nonlinear optimization*, Math. Program. Ser. A, 144 (2013), pp. 93–106.

[13] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *On the evaluation complexity of cubic regularization methods for potentially rank-deficient nonlinear least-squares problems and its relevance to constrained nonlinear optimization*, SIAM J. Optim., 23 (2013), pp. 1553–1574.

[14] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *Corrigendum: On the Complexity of Finding First-Order Critical Points in Constrained Nonlinear Optimization*, Technical Report RAL-P-2014-013, Rutherford Appleton Laboratory, Chilton, Oxfordshire, England, 2014.

[15] C. Cartis, Ph. R. Sampaio, and Ph. L. Toint, *Worst-case complexity of first-order non-monotone gradient-related algorithms for unconstrained optimization*, Optimization, 64 (2015), pp. 1349–1361.

[16] J. P. Dussault, *Simple Unified Convergence Proofs for the Trust-Region and a New ARC Variant*, Technical report, University of Sherbrooke, Sherbrooke, QC, Canada, 2015.

[17] G. N. Grapiglia, J. Yuan, and Y. Yuan, *On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization*, Math. Program. Ser. A, 152 (2015), pp. 491–520.

[18] S. Łojasiewicz, *Ensembles semi-analytiques*, Technical report, Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, France, 1965; available online at http://perso.univ-rennes1.fr/michel.coste/Lojasiewicz.pdf.

[19] Yu. Nesterov, *Introductory Lectures on Convex Optimization*, Appl. Optim., Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[20] Yu. Nesterov and B. T. Polyak, *Cubic regularization of Newton method and its global performance*, Math. Program. Ser. A, 108 (2006), pp. 177–205.

[21] L. Qi and Z. Wei, *On the constant positive linear dependence condition and its application to SQP methods*, SIAM J. Optim., 10 (2000), pp. 963–981.
[22] R. T. Rockafellar, *Lagrange multipliers and optimality*, SIAM Rev., 35 (1993), pp. 183–238.
[23] L. N. Vicente, *Worst case complexity of direct search*, EURO J. Comput. Optim., 1 (2013), pp. 143–153.