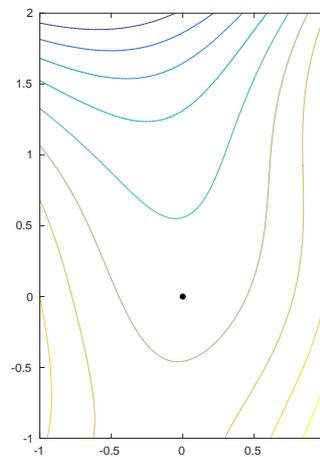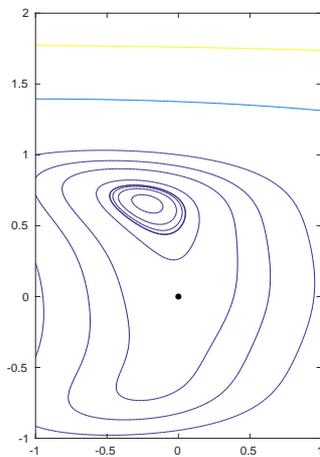Evaluation complexity bounds for
smooth constrained nonlinear optimization
using scaled KKT conditions and high-order models

by C. Cartis, N. I. M. Gould and Ph. L. Toint
Report NAXYS-11-2015(R1)   18 November 2015

University of Namur, 61, rue de Bruxelles, B5000 Namur (Belgium)

Evaluation complexity bounds
for smooth constrained nonlinear optimization
using scaled KKT conditions and high-order models

C. Cartis[*] N. I. M. Gould[†] and Ph. L. Toint[‡]

18 November 2015

Abstract

Evaluation complexity for convexly constrained optimization is considered and it is shown first that the complexity bound of $O(\epsilon^{-3/2})$ proved by Cartis, Gould and Toint (IMAJNA 32(4) 2012, pp.1662-1695) for computing an ϵ-approximate first-order critical point can be obtained under significantly weaker assumptions. Moreover, the result is generalized to the case where high-order derivatives are used, resulting in a bound of $O(\epsilon^{-(p+1)/p})$ evaluations whenever derivatives of order p are available. It is also shown that the bound of $O(\epsilon_{\rm P}^{-1/2}\epsilon_{\rm D}^{-3/2})$ evaluations ($\epsilon_{\rm P}$ and $\epsilon_{\rm D}$ being primal and dual accuracy thresholds) suggested by Cartis, Gould and Toint (SINUM, 53(2), 2015, pp.836-851) for the general nonconvex case involving both equality and inequality constraints can be generalized to yield a bound of $O(\epsilon_{\rm P}^{-1/p}\epsilon_{\rm D}^{-(p+1)/p})$ evaluations under similarly weakened assumptions.

Keywords: complexity theory, nonlinear optimization, constrained problems, high-order models, convex constraints.

1   Introduction

In [4] and [7], we examined the worst-case evaluation complexity of finding an ϵ-approximate first-order critical point for smooth nonlinear (possibly nonconvex) optimization problems for a methods using both first and second derivatives of the objective function. The case where constraints are defined by a convex set was considered in the first of these references while the general case (with equality and inequality constraints) was discussed in the second.

It was shown in [4] that at most $O(\epsilon^{-3/2})$ evaluations of the objective function and its derivatives are needed to compute such an approximate critical point. This result, which is identical in order to the best known result for the unconstrained case, comes at the price of

---

[*]Mathematical Institute, Oxford University, Oxford OX2 6GG, Great Britain. Email: coralia.cartis@maths.ox.ac.uk.

[†]Numerical Analysis Group, Rutherford Appleton Laboratory, Chilton OX11 0QX, Great Britain. Email: nick.gould@stfc.ac.uk.

[‡]Namur Center for Complex Systems (naXys) and Department of Mathematics, University of Namur, 61, rue de Bruxelles, B-5000 Namur, Belgium. Email: philippe.toint@unamur.be. Ph. Toint would like to dedicate this paper to Thaïs, his first grand-daughter.

potentially restrictive technical assumptions: it was assumed that an approximate first-order critical point of a cubic model subject to the problem's constraints can be obtained for the subproblem solution in a uniformly bounded number of descent steps that is independent of ϵ, that all iterates remains in a bounded set and that the gradient of the objective function is also Lipschitz continuous (see [4] for details). The analysis of [7] then built on the result of the convex case by first specializing it to convexly constrained nonlinear least-squares and then using the resulting complexity bound in the context of a two-phase algorithm for the problem involving general constraints. If ϵ_{P} and ϵ_{D} are the primal and the dual criticality thresholds, respectively, it was suggested that at most $O(\epsilon_{\mathrm{P}}^{-1/2}\epsilon_{\mathrm{D}}^{-3/2})$ evaluations of the objective function and its derivatives are needed to compute an approximate critical point in that case, where the Karush-Kuhn-Tucker (KKT) conditions are scaled to take the size of the Lagrange multipliers into account. Because of the proof of this result is based an the bound obtained for the convex case, it suffers from the same limitations (not to mention an additional constraint on the relative sizes of ϵ_{P} and ϵ_{D}, see [7]).

More recently, Birgin, Gardenghi, Martínez, Santos and Toint [3] provided a new regularization algorithm for the unconstrained problem with two interesting features. The first is that the model decrease condition used for the subproblem solution is weaker than that used previously, and the second is that the use of problem derivatives of order higher than two is allowed, resulting in corresponding reductions in worst-case complexity. In addition, the same authors also analyzed the worst-case evaluation complexity of the general constrained optimization problem in [2] also allowing for high-order derivatives and models in a framework inspired by that of [6,7]. At variance with the analysis of these latter references, their analysis considers unscaled approximate first-order critical points in the sense that such points satisfy the standard unscaled KKT conditions with accuracy ϵ_{P} and ϵ_{D}.

The first purpose of this paper is to explore the potential of the proposals made in [3] to overcome the limitations of [4] and to extend its scope by considering the use of high-order derivatives and models. A second objective is to use the resulting worst-case bounds to establish strengthened evaluation complexity bounds for the general nonlinearly constrained optimization problem in the framework of scaled KKT conditions, thereby improving [7]. The paper is thus organized in two main sections, Section 2 covering the convexly constrained case and Section 3 that allowing general nonlinear constraints. The results obtained are finally discussed in Section 4.

2 Convex constraints

The first problem we wish to solve is formally described as

$$\min_{x \in \mathcal{F}} f(x) \tag{2.1}$$

where we assume that $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is p-times continuously differentiable, bounded from below, and has Lipschitz continuous p-th derivatives. For the q-th derivative of a function $h : \mathbb{R}^n \to \mathbb{R}$ to be Lipschitz continuous on the set $\mathcal{S} \subseteq \mathbb{R}^n$, we require that i.e. there exists a constant $L_{h,q} \geq 0$ such that, for all $x, y \in \mathcal{S}$,

$$\|\nabla_x^q h(x) - \nabla_x^q h(y)\|_T \leq (q-1)! \, L_{h,q} \|x - y\|$$

where $\| \cdot \|_T$ is the recursively induced Euclidean norm on the space of q-th order tensors. We also assume that the feasible set $\mathcal{F}$ is closed, convex and non-empty. Note that this

formulation covers standard inequality (and linear equality) constrained optimization in its different forms: the set $\mathcal{F}$ may be defined by simple bounds, and both polyhedral and more general convex constraints. We remark though that we are tacitly assuming here that the cost of evaluating constraint functions and their derivatives is negligible.

The algorithm considered in this paper is iterative. Let $T_p(x_k, s)$ be the p-th order Taylor-series approximation to $f(x_k + s)$ at some iterate $x_k \in \mathbb{R}^n$, and define the local regularized model at x_k by

$$m_k(x_k + s) \stackrel{\text{def}}{=} T_p(x_k, s) + \frac{\sigma_k}{p+1} \|s\|^{p+1}, \tag{2.2}$$

where $\sigma_k > 0$ is the regularization parameter. Note that $m_k(x_k) = T_p(x_k, 0) = f(x_k)$. The approach used in [4] (when $p = 2$) seeks to define a new iterate x_{k+1} from the preceding one by computing an approximate solution of the subproblem

$$\min_{x \in \mathcal{F}} m_k(x_k + s) \tag{2.3}$$

using a modified version of the Adaptive Regularization with Cubics (ARC) method for unconstrained minimization. By contrast, we now examine the possibility of modifying the ARp algorithm of [3] with the aim of inheriting its interesting features. As in [4], the modification involves a suitable continuous first-order criticality measure for the constrained problem of minimizing a given function $h : \mathbb{R}^n \to \mathbb{R}$ on $\mathcal{F}$. For an arbitrary $x \in \mathcal{F}$, this criticality measure is given by

$$\pi_h(x) \stackrel{\text{def}}{=} \|P_{\mathcal{F}}[x - \nabla_x h(x)] - x\|, \tag{2.4}$$

where $P_{\mathcal{F}}$ denotes the orthogonal projection onto $\mathcal{F}$ and $\|\cdot\|$ the Euclidean norm. It is known that x is a first-order critical point of problem (2.1) if and only if $\pi_f(x) = 0$. Also note that $\pi_f(x) = \|\nabla_x h(x)\|$ whenever $\mathcal{F} = \mathbb{R}^n$.

We now describe our algorithm as the ARpCC algorithm (ARp for Convex Constraints) on the following page.

We first state a useful property of the ARpCC algorithm, which ensures that a fixed fraction of the iterations $1, 2, \ldots, k$ must be either successful or very successful.

Lemma 2.1 [3, Lem.2.4, 6, Thm.2.2]. Assume that, for some $\sigma_{\max} > 0$, $\sigma_j \leq \sigma_{\max}$ for all $0 \leq j \leq k$. Then the ARpCC algorithm ensures that

$$k \leq \kappa_u |\mathcal{S}_k|, \quad \text{where} \quad \kappa_u \stackrel{\text{def}}{=} \left(1 + \frac{|\log \gamma_1|}{\log \gamma_2}\right) + \frac{1}{\log \gamma_2} \log\left(\frac{\sigma_{\max}}{\sigma_0}\right), \tag{2.11}$$

where $\mathcal{S}_k$ is the number of successful and very successful iterations, in the sense of (2.10), up to iteration k.

We start our worst-case analysis by formalizing our assumptions

AS.1 The objective function f is p times continuously differentiable on an open set containing $\mathcal{F}$.

Algorithm 2.1: Adaptive Regularization using p-th order models for convex constraints (ARpCC)

A starting point x_{-1}, an initial and a minimal regularization parameter $\sigma_0 \geq \sigma_{\min} > 0$, algorithmic parameters $\theta > 0$, $\gamma_3 \geq \gamma_2 > 1 > \gamma_1 > 0$ and $1 > \eta_2 \geq \eta_1 > 0$, are given, as well as an accuracy threshold $\epsilon \in (0, 1)$. Compute $x_0 = P_{\mathcal{F}}[x_{-1}]$ and evaluate $f(x_0)$ and $\nabla_x f(x_0)$.

For $k = 0, 1, \ldots$, until **termination**, do:

1. Evaluate $\nabla_x f(x_k)$. If

$$\pi_f(x_k) \leq \epsilon, \tag{2.5}$$

    terminate. Otherwise compute derivatives of f of order 2 to p at x_k.

2. Compute a step s_k by approximately minimizing $m_k(x_k + s)$ over $s \in \mathcal{F}$ so that

$$x_k + s_k \in \mathcal{F}, \tag{2.6}$$

$$m_k(x_k + s_k) < m_k(x_k) \tag{2.7}$$

    and

$$\pi_{m_k}(x_k + s_k) \leq \theta \, \|s_k\|^p. \tag{2.8}$$

3. Compute $f(x_k + s_k)$ and

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{T_p(x_k, 0) - T_p(x_k, s_k)}. \tag{2.9}$$

    If $\rho_k \geq \eta_1$, set $x_{k+1} = x_k + s_k$. Otherwise set $x_{k+1} = x_k$.

4. Set

$$\sigma_{k+1} \in \begin{cases} [\max(\sigma_{\min}, \gamma_1 \sigma_k)\sigma_k] & \text{if } \rho_k > \eta_2 & \text{[very successful iteration]} \\ [\sigma_k, \gamma_2 \sigma_k] & \text{if } \eta_1 \leq \rho_k \leq \eta_2 & \text{[successful iteration]} \\ [\gamma_2 \sigma_k, \gamma_3 \sigma_k] & \text{otherwise.} & \text{[unsuccessful iteration],} \end{cases} \tag{2.10}$$

    and go to step 2 if $\rho_k < \eta_1$.

The p-th derivative of f is Lipschitz continuous on $\mathcal{F}$.

The feasible set $\mathcal{F}$ is closed, convex and non-empty.

The ARpCC algorithm is required to start from a feasible $x_0 \in \mathcal{F}$, which, together with the fact that the subproblem solution in Step 2 involves minimization over $\mathcal{F}$, leads to AS.3.

We now recall some simple results whose proof can be found in [3] in the context of the original ARp algorithm.

Lemma 2.2 Suppose that AS.1–AS.3 hold. Then, for each $k \geq 0$,

(i)

$$f(x_k + s_k) \leq T_p(x_k, s_k) + \frac{L_{f,p}}{p}\|s_k\|^{p+1} \qquad (2.12)$$

and

$$\|\nabla_x f(x_k + s_k) - \nabla_s T(x_k, s_k)\| \leq L_{f,p}\|s_k\|^p; \qquad (2.13)$$

(ii)

$$T_p(x_k, 0) - T_p(x_k, s_k) \geq \frac{\sigma_k}{p+1}\|s_k\|^{p+1}; \qquad (2.14)$$

(iii)

$$\sigma_k \leq \sigma_{\max} \overset{\text{def}}{=} \max\left[\sigma_0, \frac{\gamma_3 L_{f,p}(p+1)}{p(1-\eta_2)}\right]. \qquad (2.15)$$

Proof. See [3] for the proofs of (2.12) and (2.13), which crucially depend on AS.1 and AS.2 being valid on the segment $[x_k, x_k + s_k]$, i.e.

$$\|\nabla_x^p f(x_k + \xi s_k) - \nabla_x^p f(x_k)\|_p \leq L_{f,p}\xi\|s_k\| \quad \text{for all} \quad \xi \in [0,1]. \qquad (2.16)$$

Observe also that (2.2) and (2.7) ensure (2.14). Assume now that

$$\sigma_k \geq \frac{L_{f,p}(p+1)}{p(1-\eta_2)}. \qquad (2.17)$$

Using (2.12) and (2.14), we may then deduce that

$$|\rho_k - 1| \leq \frac{|f(x_k + s_k) - T_p(x_k, s_k)|}{|T_p(x_k, 0) - T_p(x_k, s_k)|} \leq \frac{L_{f,p}(p+1)}{p\,\sigma_k} \leq 1 - \eta_2$$

and thus that $\rho_k \geq \eta_2$. Then iteration k is very successful in that $\rho_k \geq \eta_2$ and $\sigma_{k+1} \leq \sigma_k$. As a consequence, the mechanism of the algorithm ensures that (2.15) holds. $\square$

We now prove that, at successful iterations, the step at iteration k must be bounded below by a multiple of the p-th root of the criticality measure at iteration $k+1$.

Lemma 2.3 Suppose that AS.1–AS.3 hold. Then

$$\|s_k\| \geq \left[\frac{\pi_f(x_{k+1})}{L_{f,p} + \theta + \sigma_{\max}}\right]^{\frac{1}{p}} \quad \text{for all} \quad k \in \mathcal{S}. \tag{2.18}$$

Proof. Since $k \in \mathcal{S}$ and by definition of the trial point, we have that $x_{k+1} = x_k + s_k$. Observe now that (2.13) and (2.15) imply that

$$\|\nabla f(x_{k+1}) - \nabla_x m_k(x_{k+1})\| \leq L_{f,p}\|s_k\|^p + \sigma_k\|s_k\|^p \leq (L_{f,p} + \sigma_{\max})\|s_k\|^p. \tag{2.19}$$

Combing this bound with the triangle inequality, the contractive nature of the projection and (2.8), we deduce that

$$\begin{aligned}
\pi_f(x_{k+1}) &= \|P_{\mathcal{F}}[x_{k+1} - \nabla_x f(x_{k+1})] - P_{\mathcal{F}}[x_{k+1} - \nabla_x m_k(x_{k+1})] \\
&\quad + P_{\mathcal{F}}[x_{k+1} - \nabla_x m_k(x_{k+1})] - x_{k+1}\| \\
&\leq \|P_{\mathcal{F}}[x_{k+1} - \nabla_x f(x_{k+1})] - P_{\mathcal{F}}[x_{k+1} - \nabla_x m_k(x_{k+1})]\| + \pi_{m_k}(x_{k+1}) \\
&\leq \|\nabla_x f(x_{k+1})] - \nabla_x m_k(x_{k+1})\| + \pi_{m_k}(x_{k+1}) \\
&\leq (L_{f,p} + \theta + \sigma_{\max})\|s_k\|^p
\end{aligned}$$

and (2.18) follows. $\square$

We now consolidate the previous results by deriving a lower bound on the objective function decrease at successful iterations.

Lemma 2.4 Suppose that AS.1–AS.3 hold. Then, if iteration k is successful,

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{\kappa_s^f}\,\pi_f(x_{k+1})^{\frac{p+1}{p}}$$

where

$$\kappa_s^f \overset{\text{def}}{=} \max\left[1, \frac{p+1}{\eta_1 \sigma_{\min}}\left(L_{f,p} + \theta + \sigma_{\max}\right)^{\frac{p+1}{p}}\right]. \tag{2.20}$$

Proof. If iteration k is successful, we have, using (2.9), (2.14), (2.10), (2.18) and (2.15) successively, that

$$\begin{aligned}
f(x_k) - f(x_{k+1}) &\geq \eta_1\left[T_p(x_k, 0) - T_p(x_k, s_k)\right] \\
&\geq \frac{\eta_1 \sigma_{\min}}{p+1}\,\|s_k\|^{p+1} \\
&\geq \frac{\eta_1 \sigma_{\min}}{(p+1)[L_{f,p} + \theta + \sigma_{\max}]^{\frac{p+1}{p}}}\,\pi_f(x_{k+1})^{\frac{p+1}{p}}.
\end{aligned}$$

$\square$

It is important to note that the validity of this lemma does not depend on the history of the algorithm, but is only conditional to the smoothness assumption on the objective function holding along the step from x_k to x_{k+1}. We will make use of that observation in Section 3.

Our worst-case evaluation complexity results can now be proved by combining this last result with the fact that $\pi_f(x_k)$ cannot be smaller than ϵ before termination.

Theorem 2.5 Suppose that AS.1–AS.3 hold and let f_{low} be a lower bound on f on $\mathcal{F}$. Then, given $\epsilon > 0$, the ARpCC algorithm applied to problem (2.1) needs at most

$$\left\lfloor \kappa_s^f \, \frac{f(x_0) - f_{\text{low}}}{\epsilon^{\frac{p+1}{p}}} \right\rfloor$$

successful iterations (each involving one evaluation of f and its p first derivatives) and at most

$$\kappa_u \left\lfloor \kappa_s^f \, \frac{f(x_0) - f_{\text{low}}}{\epsilon^{\frac{p+1}{p}}} \right\rfloor$$

iterations in total to produce an iterate x_ϵ such that $\pi_f(x_\epsilon) \leq \epsilon$, where κ_u is given by (2.11) with $\sigma_{\max}$ defined by (2.15).

Proof. At each successful iteration, we have, using Lemma 2.4, that

$$f(x_k) - f(x_{k+1}) \geq (\kappa_s^f)^{-1} \pi_f(x_{k+1})^{\frac{p+1}{p}} \geq (\kappa_s^f)^{-1} \epsilon^{\frac{p+1}{p}},$$

where we used the fact that $\pi_f(x_{k+1}) \geq \epsilon$ before termination to deduce the last inequality. Thus we deduce that, as long as termination does not occur,

$$f(x_0) - f(x_{k+1}) = \sum_{j \in \mathcal{S}_k} [f(x_j) - f(x_j + s_j)] \geq \frac{|\mathcal{S}_k|}{\kappa_s^f} \epsilon^{\frac{p+1}{p}},$$

from which the desired bound on the number of successful iterations follows. Lemma 2.1 is then invoked to compute the upper bound on the total number of iterations. $\square$

For what follows, it is very important to note that the Lipschitz continuity of $\nabla_x^q f$ was only used (in Lemma 2.2) to ensure that (2.16) holds for all $k \geq 0$.

3   The general constrained case

We now consider the general smooth constrained problem in the form

$$\min_{x \in \mathcal{F}} f(x) \quad \text{subject to} \quad c(x) = 0 \tag{3.1}$$

where $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is sufficiently smooth and f and $\mathcal{F}$ are as above. Note that this formulation covers the general problem involving both equality and inequality constraints,

the latter being handled using slack variables and the inclusion of the associated simple bounds in the definition of $\mathcal{F}$.

Our idea is now to first apply the ARpCC algorithm to the problem

$$\min_{x \in \mathcal{F}} \nu(x) \stackrel{\text{def}}{=} \tfrac{1}{2}\|c(x)\|^2. \tag{3.2}$$

If an approximately feasible point is found, then we may follow the spirit of [5–7] and [2] and apply the same ARpCC to approximately solve the problem

$$\min_{x \in \mathcal{F}} \mu(x, t_k) \stackrel{\text{def}}{=} \tfrac{1}{2}\|r(x, t_k)\|^2 \stackrel{\text{def}}{=} \tfrac{1}{2}\left\|\left(\begin{array}{c} c(x) \\ f(x) - t_k \end{array} \right)\right\|^2 \tag{3.3}$$

in the set for some monotonically decreasing sequence of "targets" t_k $(k = 1, \dots)$.

Observe that the recomputations of $\pi_\mu(x_{k+1}, t_{k+1})$ in Step 2.(b) do not require re-evaluating $f(x_{k+1})$ or $c(x_{k+1})$ or any of their derivatives.

We now complete our assumptions.

AS.4   All derivatives of f of order 0 to p are Lipschitz continuous in $\mathcal{F}$.

AS.5   For each $i = 1, \dots, m$, the constraint function c_i is p times continuously differentiable on an open set containing $\mathcal{F}$.

AS.6   All derivatives of order 0 to p of each c_i $(i = 1, \dots, m)$ are Lipschitz continuous in $\mathcal{F}$.

AS.7   There exists constants $\beta \geq \epsilon_{\mathrm{P}}$ and $f_{\mathrm{low}} \in \mathbb{R}$ such that $f(x) \geq f_{\mathrm{low}}$ for all $x \in \mathcal{C}_\beta \stackrel{\text{def}}{=} \{x \in \mathcal{F} \mid \|c(x)\| \leq \beta\}$.

Assume, without loss of generality, that all Lipschitz constants implied by AS.4 and AS.6 are bounded above by $L \geq 1$. Also note the problem of finding an ϵ_{P}-feasible minimizer of $f(x)$ is only meaningful if AS.7 holds.

We first verify that our assumptions are sufficient to imply that $\nu(x)$ and $\mu(x, t)$ have Lipschitz p-th derivative on all segments $[x_k, x_j + s_j]$ generated by the algorithm, allowing us to exploit the results of Section 2.

Lemma 3.1 Assume that AS.3, AS.5 and AS.6 hold. Let the iteration of the ARpCC algorithm applied to problem (3.2) be indexed by j. Then the "segment" Lipschitz condition (2.16) holds for $\nabla_x^q \nu(x)$ holds on every segment $[x_j, x_j + s_j]$ $(j \geq 0)$ generated by the ARpCC algorithm during Phase 1 and any $q \in \{1, \dots, p\}$. If, in addition, AS.1 and AS.4 also hold, then the same conclusion holds for $\nabla_x^q \mu(x, t)$ on every segment $[x_j, x_j + s_j]$ $(j \geq 0)$ generated by the ARpCC algorithm during Step 2.(a) of Phase 2 and any $q \in \{1, \dots, p\}$, the Lipschitz constant in this latter case being independent of t.

Algorithm 3.1: Adaptive Regularization using p-th order models for general constraints (ARpGC)

A constant β defining $\mathcal{C}_\beta$, a starting point x_{-1}, a minimum regularization parameter $\sigma_{\min} > 0$, an initial regularization parameter $\sigma_0 \geq \sigma_{\min}$ are given, as well as a constant $\delta \in (0,1)$. The primal and dual tolerances $0 < \epsilon_{\mathrm{P}} < 1$ and $0 < \epsilon_{\mathrm{D}} < 1$ are also given.

Phase 1:

  Starting from $x_0 = P_{\mathcal{F}}(x_{-1})$, apply the AR$p$CC algorithm to minimize $\frac{1}{2}\|c(x)\|^2$ subject to $x \in \mathcal{F}$ until a point $x_1 \in \mathcal{F}$ is found such that

$$\|c(x_1)\| < \delta\epsilon_{\mathrm{P}} \quad \text{or} \quad \pi_{\frac{1}{2}\|c\|^2}(x_1) \leq \epsilon_{\mathrm{D}}\|c(x_1)\|. \tag{3.4}$$

  If $\|c(x_1)\| \geq \delta\epsilon_{\mathrm{P}}$, terminate with $x_\epsilon = x_1$.

Phase 2:

 1. Set $t_1 = f(x_1) - \sqrt{\epsilon_{\mathrm{P}}^2 - \|c(x_1)\|^2}$.

 2. For $k = 1, 2, \ldots$, do:

  (a) Starting from x_k, apply the ARpCC algorithm to minimize $\mu(x, t_k)$ as a function of $x \in \mathcal{F}$ until an iterate $x_{k+1} \in \mathcal{F}$ is found such that

$$\|r(x_{k+1}, t_k)\| < \delta\epsilon_{\mathrm{P}} \quad \text{or} \quad f(x_{k+1}) < t_k \quad \text{or} \quad \pi_\mu(x_{k+1}, t_k) \leq \epsilon_{\mathrm{D}}\|r(x_{k+1}, t_k)\| \tag{3.5}$$

  (b) i. If $\|r(x_{k+1}, t_k)\| < \delta\epsilon_{\mathrm{P}}$, define t_{k+1} according to

$$t_{k+1} = f(x_{k+1}) - \sqrt{\epsilon_{\mathrm{P}}^2 - \|c(x_{k+1})\|^2}. \tag{3.6}$$

   and terminate with $(x_\epsilon, t_\epsilon) = (x_{k+1}, t_{k+1})$ if $\pi_\mu(x_{k+1}, t_{k+1}) \leq \epsilon_{\mathrm{D}}\|r(x_{k+1}, t_{k+1})$.

   ii. If $\|r(x_{k+1}, t_k)\| \geq \delta\epsilon_{\mathrm{P}}$ and $f(x_{k+1}) < t_k$, define t_{k+1} according to

$$t_{k+1} = 2f(x_{k+1}) - t_k \tag{3.7}$$

   and terminate with $(x_\epsilon, t_\epsilon) = (x_{k+1}, t_{k+1})$ if $\pi_\mu(x_{k+1}, t_{k+1}) \leq \epsilon_{\mathrm{D}}\|r(x_{k+1}, t_{k+1})\|$.

   iii. If $\|r(x_{k+1}, t_k)\| \geq \delta\epsilon_{\mathrm{P}}$ and $f(x_{k+1}) \geq t_k$, terminate with $(x_\epsilon, t_\epsilon) = (x_{k+1}, t_k)$

Proof. Since

$$\nabla_x^q \nu(x) = \sum_{i=1}^m \left[\sum_{\ell,j>0,\,\ell+j=q} \alpha_{\ell,j} \nabla_x^\ell c_i(x) \nabla_x^j c_i(x) + c_i(x) \nabla_x^q c_i(x) \right]$$

(where $\{\alpha_{\ell,j}\}$ are suitable positive and finite coefficients), condition (2.16) is satisfied on the segment $[x_j, x_j + s_j]$ if (i) the derivatives $\{\nabla_x^{\min[\ell,j]} c_i(x)\}_{i=1}^m$ are Lipschitz continuous on $[x_j, x_j + s_j]$, (ii) $\{\nabla_x^{\max[\ell,j]} c_i(x)\}_{i=1}^m$ are uniformly bounded on $[x_j, x_j + s_j]$, and (iii) we have that

$$\sum_{i=1}^m \|c_i(x_j + \xi s_j)\nabla_x^q c_i(x_j + \xi s_j) - c_i(x_j)\nabla_x^q c_i(x_j)\|_q \le L_1 \xi \|s_j\| \tag{3.8}$$

for some constant $L_1 > 0$. The first of these conditions is ensured by AS.6, the second by the observation that AS.6 again implies that $\|\nabla_x^\ell c_i(x)\| \le L$ for $\ell \in \{1, \ldots, q\}$ (see [11, Lem. 1.2.2, p. 21]). Moreover,

$$
\begin{aligned}
\|c_i(x_j + \xi s_j)\nabla_x^q c_i(x_j + \xi s_j) &- c_i(x_j)\nabla_x^q c_i(x_j)\| \\
&\le \ |c_i(x_j + \xi s_j) - c_i(x_j)| \, \|\nabla_x^q c_i(x_j + \xi s_j)\|_q \\
&\quad + |c_i(x_j)| \, \|\nabla_x^q c_i(x_j + \xi s_j) - \nabla_x^q c_i(x_j)\|_q
\end{aligned}
$$

and the first term on the right-hand side is bounded above by $L^2 \xi \|s_j\|$ and the second by $|c_i(x_j)| L \xi \|s_j\|$. Hence (3.8) holds with

$$L_1 = \sum_{i=1}^m \left(L^2 + |c_i(x_j)| L \right) \le m L^2 + m\|c(x_j)\| L \le m L^2 + m\|c(x_0)\| L$$

because the ARpCC algorithm ensures that $\|c(x_j)\| \le \|c(x_0)\|$ for all $j \ge 0$. As a consequence, AS.3, AS.5 and AS.6 guarantee that (2.16) holds with the Lipschitz constant

$$m \left[\left(\max_{i=1,\ldots,m} \alpha_i \right) L^2 + L^2 + \|c(x_0)\| L \right].$$

If we now assume that AS.1 and AS.4 also hold, we may repeat, for $\mu(x,t)$ (with fixed t) the same reasoning as above and obtain that condition (2.16) holds for each segment $[x_j, x_j + s_j]$ generated by the ARpCC algorithm applied in Step2.(a) of Phase 2, with Lipschitz constant

$$
\begin{aligned}
m \left[\left(\max_{i=1,\ldots,m} \alpha_i \right) L^2 + L^2 + \|c(x_{j,0})\| L \right] &+ \left(\max_{i=1,\ldots,m} \alpha_i \right) L^2 + L^2 + |f(x_{j,0}) - t_j| L \\
&\le \ (m+1) L^2 \left[1 + \left(\max_{i=1,\ldots,m} \alpha_i \right) \right] + 2mL \stackrel{\text{def}}{=} L_{\mu,p},
\end{aligned}
$$

where we have used (3.14) and $\epsilon_\mathrm{P} \le 1$ to deduce the inequality. Note that this constant is independent of t_j, as requested. $\qquad\square$

We now start our complexity analysis as such by examining the complexity of Phase 1.

Lemma 3.2 Suppose that AS.3, AS.5 and AS.6 hold. Then Phase 1 of the ARpGC algorithm terminates after at most

$$\left\lfloor \kappa_{\mathrm{CC}}^c \|c(x_0)\| \, \max\left[\epsilon_{\mathrm{P}}^{-1}, \epsilon_{\mathrm{P}}^{-\frac{1}{p}} \epsilon_{\mathrm{D}}^{-\frac{p+1}{p}} \right] \right\rfloor + 1.$$

evaluations of c and its derivatives, where $\kappa_{\mathrm{CC}}^c \overset{\text{def}}{=} \frac{1}{2}\kappa_u \kappa_s^{\frac{1}{2}\|c\|^2} \delta^{\frac{1}{p}}$ with $\kappa_s^{\frac{1}{2}\|c\|^2}$ being the problem-dependent constant defined in (2.20) for the function $\frac{1}{2}\|c(x)\|^2$ corresponding to (3.2).

Proof. Let us index the iteration of the ARpCC algorithm applied to problem (3.2) by j and assume that iteration j is successful and that

$$\|c(x_j)\| \geq \delta \epsilon_{\mathrm{P}}. \tag{3.9}$$

If $\|c(x_{j+1})\| \leq \frac{1}{2}\|c(x_j)\|$, then

$$\|c(x_j)\| - \|c(x_{j+1})\| \geq \frac{1}{2}\|c(x_j)\| \geq \frac{1}{2}\delta\,\epsilon_{\mathrm{P}} \tag{3.10}$$

By contrast, if $\|c(x_{j+1})\| > \frac{1}{2}\|c(x_j)\|$, then, using the decreasing nature of the sequence $\{\|c(x_j)\|\}$, Lemma 2.4 (which is applicable because of Lemma 3.1) and the second part of (3.4), we obtain that

$$
\begin{aligned}
\left(\|c(x_j)\| - \|c(x_{j+1})\|\right)\|c(x_j)\| \;&\geq\; \tfrac{1}{2}\|c(x_j)\|^2 - \tfrac{1}{2}\|c(x_{j+1})\|^2 \\[1ex]
&\geq\; \left(\kappa_s^{\frac{1}{2}\|c\|^2}\right)^{-1} \left(\epsilon_{\mathrm{D}}\|c(x_{j+1})\|\right)^{\frac{p+1}{p}} \\[1ex]
&\geq\; \left(\kappa_s^{\frac{1}{2}\|c\|^2}\right)^{-1} \left(\tfrac{1}{2}\epsilon_{\mathrm{D}}\|c(x_j)\|\right)^{\frac{p+1}{p}}
\end{aligned}
$$

and thus that

$$\|c(x_j)\| - \|c(x_{j+1})\| \geq \left(\kappa_s^{\frac{1}{2}\|c\|^2}\right)^{-1} 2^{-\frac{p+1}{p}} \|c(x_j)\|^{\frac{1}{p}} \epsilon_{\mathrm{D}}^{\frac{p+1}{p}} \geq \frac{1}{2}\left(\kappa_s^{\frac{1}{2}\|c\|^2}\right)^{-1} \delta^{\frac{1}{p}} \epsilon_{\mathrm{P}}^{\frac{1}{p}} \epsilon_{\mathrm{D}}^{\frac{p+1}{p}}.$$

where we have used (3.9) to derive the last inequality. Because of (2.20), we thus obtain from this last bound and (3.10) that, for all j,

$$\|c(x_j)\| - \|c(x_{j+1})\| \geq \frac{1}{2}\left(\kappa_s^{\frac{1}{2}\|c\|^2}\right)^{-1} \delta^{\frac{1}{p}} \min\left[\epsilon_{\mathrm{P}}, \epsilon_{\mathrm{P}}^{\frac{1}{p}} \epsilon_{\mathrm{D}}^{\frac{p+1}{p}}\right]$$

As in Theorem 2.5, we then deduce that the number of successful iterations required for the ARpCC algorithm to produce a point x_1 satisfying (3.4) is bounded above by

$$\frac{1}{2}\kappa_s^{\frac{1}{2}\|c\|^2} \delta^{\frac{1}{p}} \|c(x_0)\| \, \max\left[\epsilon_{\mathrm{P}}^{-1}, \epsilon_{\mathrm{P}}^{-\frac{1}{p}} \epsilon_{\mathrm{D}}^{-\frac{p+1}{p}}\right].$$

The desired conclusion then follows by using Lemma 2.1 and adding one for the final evaluation at termination. □

Note that an improved complexity bound for convexly-constrained least-squares problems, and hence for Phase 1, was given in [8]. In particular, the bound in Lemma 3.2 improves to

$$\left\lfloor \kappa_{\text{CC-1}}^c \|c(x_0)\|^{\frac{1}{p}} \max\left[\epsilon_{\text{P}}^{-\frac{1}{p}}, \epsilon_{\text{D}}^{-\frac{p+1}{p}}\right]\right\rfloor + 1$$

whenever p is a power of 2. However, we are not aware how to use the better Phase 1 result to improve the complexity of Phase 2, and so we are omitting including it here in full.
We now partition the Phase 2 outer iterations (before that where termination occurs) into two subsets whose indexes are given by

$$\mathcal{K}_+ \overset{\text{def}}{=} \{k \geq 0 \mid \|r(x_{k+1}, t_k)\| < \delta\epsilon_{\text{P}} \quad \text{and (3.6) is applied }\} \tag{3.11}$$

and

$$\mathcal{K}_- \overset{\text{def}}{=} \{k \geq 0 \mid \|r(x_{k+1}, t_k)\| \geq \delta\epsilon_{\text{P}} \quad \text{and (3.7) is applied }\} \tag{3.12}$$

This partition allows us to prove the following technical results.

Lemma 3.3 The sequence $\{t_k\}$ is monotonically decreasing. Moreover, in every Phase 2 iteration of the ARpGC algorithm of index $k \geq 1$, we have that

$$f(x_k) - t_k \geq 0, \tag{3.13}$$

$$\|r(x_{k+1}, t_{k+1})\| = \epsilon_{\text{P}} \quad \text{for} \quad k \in \mathcal{K}_+, \tag{3.14}$$

$$\|r(x_{k+1}, t_{k+1})\| = \|r(x_{k+1}, t_k)\| \leq \epsilon_{\text{P}} \quad \text{for} \quad k \in \mathcal{K}_-, \tag{3.15}$$

$$\|c(x_k)\| \leq \epsilon_{\text{P}} \quad \text{and} \quad f(x_k) - t_k \leq \epsilon_{\text{P}}, \tag{3.16}$$

$$t_k - t_{k+1} \geq (1 - \delta)\epsilon_{\text{P}} \quad \text{for} \quad k \in \mathcal{K}_+. \tag{3.17}$$

Finally, at termination of the ARpGC algorithm,

$$\|r(x_\epsilon, t_\epsilon)\| \geq \delta\epsilon_{\text{P}} \quad \text{and} \quad f(x_\epsilon) \geq t_\epsilon \quad \text{and} \quad \pi_\mu(x_\epsilon, t_\epsilon) \leq \epsilon_{\text{D}}\|r(x_\epsilon, t_\epsilon)\|. \tag{3.18}$$

Proof.   The inequality (3.13) follows from (3.6) for $k - 1 \in \mathcal{K}_+$ and from (3.7) for $k - 1 \in \mathcal{K}_-$. (3.14) is also deduced from (3.6) while (3.7) implies the equality in (3.15), the inequality in that statement resulting from the monotonically decreasing nature of $\|r(x, t_k)\|$ during inner iterations in Step 2.(a) of the ARpGC algorithm. The inequalities (3.16) then follow from (3.13), (3.14) and (3.15). We now prove (3.17), which only occurs when $\|r(x_{k+1}, t_k)\| \leq \delta\epsilon_{\text{P}}$, that is when

$$(f(x_{k+1}) - t_k)^2 + \|c(x_{k+1})\|^2 \leq \delta^2\epsilon_{\text{P}}^2. \tag{3.19}$$

From (3.6), we then have that

$$t_k - t_{k+1} = -(f(x_{k+1}) - t_k) + \sqrt{\|r(x_k, t_k)\|^2 - \|c(x_{k+1})\|^2}. \tag{3.20}$$

Now taking into account that the global minimum of the problem

$$\min_{(f,c)\in\mathbb{R}^2} \psi(f,c) \overset{\text{def}}{=} -f + \sqrt{\epsilon_{\mathrm{P}}^2 - c^2} \quad \text{subject to} \quad f^2 + c^2 \leq \omega^2,$$

for $\omega \in [0, \epsilon_{\mathrm{P}}]$ is attained at $(f_*, c_*) = (\omega, 0)$ and it is given by $\psi(f_*, c_*) = \epsilon_{\mathrm{P}} - \omega$ (see [7, Lemma 5.2]), we obtain from (3.19) and (3.20) (setting $\omega = \delta\epsilon_{\mathrm{P}}$) that

$$t_k - t_{k+1} \geq \epsilon_{\mathrm{P}} - \omega = (1-\delta)\epsilon_{\mathrm{P}} \quad \text{for} \quad k \in \mathcal{K}_+$$

for $k \in \mathcal{K}_+$, which is (3.17). Note that, if $k \in \mathcal{K}_-$, then we must have that $t_k > f(x_{k+1})$ and thus (3.7) ensures that $t_{k+1} < t_k$. This observation and (3.17) then allow us to conclude that the sequence $\{t_k\}$ is monotonically decreasing.

In order to prove (3.18), we need to consider, in turn, each of the three possible cases where termination occurs in Step 2.(b). In the first case (i), $\|r(x_{k+1}, t_k)\|$ is small (in the sense that the first inequality in (3.5) holds) and (3.6) is then used, implying that (3.14) holds and that $f(x_{k+1}) > t_{k+1}$. If termination occurs because $\pi(x_{k+1}, t_{k+1}) \leq \epsilon_{\mathrm{D}}\|r(x_{k+1}, t_{k+1})\|$, then (3.18) clearly holds at (x_{k+1}, t_{k+1}). In the second case (ii), $\|r(x_{k+1}, t_k)\|$ is large (the first inequality in (3.5) fails), but $f(x_{k+1}) < t_k$, and t_{k+1} is then defined by (3.7), ensuring that $f(x_{k+1}) > t_{k+1}$ and, because of (3.15), that $\|r(x_{k+1}, t_{k+1})\|$ is also large. As before (3.18) holds at (x_{k+1}, t_{k+1}) if termination occurs because $\pi(x_{k+1}, t_{k+1}) \leq \epsilon_{\mathrm{D}}\|r(x_{k+1}, t_{k+1})\|$. The third case (iii) is when $\|r(x_{k+1}, t_k)\|$ is sufficiently large and $f(x_{k+1}) \geq t_k$. But (3.5) then guarantees that $\pi(x_{k+1}, t_k) \leq \epsilon_{\mathrm{D}}\|r(x_{k+1}, t_k)\|$, and the inequalities (3.18) are again satisfied at (x_{k+1}, t_k). $\qquad\square$

Using the results of this lemma allows us to bound the number of outer iterations in $\mathcal{K}_+$.

Lemma 3.4 Suppose that AS.7 holds. Then

$$|\mathcal{K}_+| \leq \frac{f(x_1) - f_{\mathrm{low}} + 1}{1 - \delta}\,\epsilon_{\mathrm{P}}^{-1}.$$

Proof. We first note that (3.14) and (3.15) and AS.7 ensure that $x_k \in \mathcal{C}_\beta$ for all $k \geq 0$. The result then immediately follows from AS.7 again and the observation that, from (3.17), t_k decreases monotonically with a decrease of at least $(1-\delta)\epsilon_{\mathrm{P}}$ for $k \in \mathcal{K}_+$. $\square$

We now state a very useful consequence of Lemmas 3.1 and 3.3.

Lemma 3.5 Suppose that AS.1 and AS3–AS.6 hold. Then there exists a constant $\sigma_{\mu,\max} > \sigma_{\min}$ such that all regularization parameters arising in the ARpCC algorithm within Step 2.(a) of the ARpGC algorithm are bounded above by $\sigma_{\mu,\max}$.

Proof. AS.1, AS.4–AS.6 and Lemma 3.1 guarantee the existence of a Lipchitz constant independent of t such that the "segment-wise" Lipschitz condition (2.16) holds for each segment $[x_{j,\ell}, x_{j,\ell} + s_{j,\ell}]$. The result is then derived by introducing $L_{\mu,p}$ in (2.15) to obtain $\sigma_{\mu,\max}$. □

The main consequence of this result is that we may apply the ARpCC algorithm to the minimization of $\mu(x, t_k)$ in Step 2.(a) of the ARpGC algorithm and use all the properties of the former (as derived in the previous section) using problem constants valid for every possible t_k.

Consider now x_k for $k \in \mathcal{K}_+$ and denote by $x_{k+\ell(k)}$ the next iterate such that $k+\ell(k) \in \mathcal{K}_+$ or the algorithm terminates at $k + \ell(k)$. Two cases are then possible: either a single pass in Step 2.(a) of the ARpGC algorithm is sufficient to obtain $x_{k+\ell(k)}$ ($\ell(k) = 1$) or two or more passes are necessary, with iterations $k + 1, \ldots, k + \ell(k) - 1$ belonging to $\mathcal{K}_-$. Assume that the iterations of the ARpCC algorithm at Step 2.(a) of the outer iteration j are numbered $(j, 0), (j, 1), \ldots, (j, e_j)$ and note that the mechanism of the ARpGC algorithm ensures that iteration (j, e_j) is successful for all j. Now define, for $k \in \mathcal{K}_+$, the index set of all inner iterations necessary to deduce $x_{k+\ell(k)}$ from x_k, that is

$$\mathcal{I}_k \overset{\text{def}}{=} \{(k, 0), \ldots, (k, e_k), \ldots, (j, 0), \ldots, (j, e_j), \ldots, (k+\ell(k)-1, 0), \ldots (k+\ell(k)-1, e_{k+\ell(k)-1})\} \tag{3.21}$$

where $k < j < k + \ell(k) - 1$. Observe that, by the definitions (3.11) and (3.21), the index set of all inner iterations before termination is given by $\cup_{k \in \mathcal{K}_+} \mathcal{I}_k$, and therefore that the number of evaluations of problem's functions required to terminate in Phase 2 is bounded above by

$$|\bigcup_{k \in \mathcal{K}_+} \mathcal{I}_k| + 1 \leq \left(\frac{f(x_1) - f_{\text{low}} + 1}{1 - \delta} \epsilon_{\text{P}}^{-1} \times \max_{k \in \mathcal{K}_+} |\mathcal{I}_k| \right) + 1, \tag{3.22}$$

where we added 1 to take the final evaluation into account and where we used Lemma 3.4 to deduce the inequality. We now invoke the complexity properties of the ARpCC algorithm applied to problem (3.3) to obtain an upper bound on the cardinality of each $\mathcal{I}_k$.

Lemma 3.6 Suppose that AS.1, AS.3–AS.6 hold. Then, for each $k \in \mathcal{K}_+$ before termination,

$$|\mathcal{I}_k| \leq (1 - \delta)\kappa_{\text{CC}}^{\mu} \max\left[1, \epsilon_{\text{P}}^{\frac{p-1}{p}} \epsilon_{\text{D}}^{-\frac{p+1}{p}} \right].$$

where κ_{CC}^{μ} is independent of ϵ_{P} and ϵ_{D} and captures the problem-dependent constants associated with problem (3.3) for all values of t_k generated by the algorithm.

Proof. Observe first that, because of Lemma 3.5, we may apply the ARpCC algorithm for the minimization of $\mu(x, t_j)$ for each j such that $k \leq j < k + \ell(k)$. Observe that (3.15) and the mechanism of this algorithm guarantees the decreasing nature of the sequence $\{\|r(x_j, t_j)\|\}_{j=k}^{k+\ell(k)-1}$ and hence of the sequence $\{\|r(x_{j,s}, t_j)\|\}_{(j,s) \in \mathcal{I}_k}$. This reduction starts from the initial value $\|r(x_{k,0}, t_k)\| = \epsilon_{\text{P}}$ and is carried out for all iterations with index in $\mathcal{I}_k$ at worst until it is smaller than $\delta\epsilon_{\text{P}}$ (see the first part of (3.5)). We may then invoke

Lemmas 3.5 and 2.4 to deduce that, if $(j,s) \in \mathcal{I}_k$ is the index of a successful inner iteration and as long as the third part of (3.5) does not hold,

$$
\begin{aligned}
(\|r(x_{j,s},t_j)\| - \|r(x_{j,s+1},t_j)\|)\|r(x_{j,s},t_j)\| & \geq \tfrac{1}{2}\|r(x_{j,s},t_j)\|^2 - \tfrac{1}{2}\|r(x_{j,s+1},t_j)\|^2 \\
& \geq [\kappa_{\text{CC}}^{\mu,s}]^{-1} \left(\epsilon_{\text{D}}\|r(x_{j,s+1},t_j)\|\right)^{\frac{p+1}{p}},
\end{aligned}
\tag{3.23}
$$

for $0 \leq s < e_j$ and for some constant $\kappa_{\text{CC}}^{\mu,s} > 0$ independent of ϵ_{P}, ϵ_{D}, s and j, while

$$
\tfrac{1}{2}\|r(x_{j,e_j},t_j)\| - \tfrac{1}{2}\|r(x_{j+1,0},t_{j+1})\| = 0.
$$

As above, suppose first that $\|r(x_{j,s+1},t_j)\| \leq \tfrac{1}{2}\|r(x_{j,s},t_j)\|$. Then

$$
\|r(x_{j,s},t_j)\| - \|r(x_{j,s+1},t_j)\| \geq \tfrac{1}{2}\|r(x_{j,s},t_j)\| \geq \tfrac{1}{2}\delta\epsilon_{\text{P}}
\tag{3.24}
$$

because of the first part of (3.5). If $\|r(x_{j,s+1},t_j)\| > \tfrac{1}{2}\|r(x_{j,s},t_j)\|$ instead, then (3.23) implies that

$$
\|r(x_{j,s},t_j)\| - \|r(x_{j,s+1},t_j)\| \geq [\kappa_{\text{CC}}^{\mu,s}]^{-1} 2^{-\frac{p+1}{p}} \|r(x_{j,s},t_j)\|^{\frac{1}{p}}\epsilon_{\text{D}}^{\frac{p+1}{p}} \geq [\kappa_{\text{CC}}^{\mu,s}]^{-1} 2^{-\frac{p+1}{p}} \delta^{\frac{1}{p}} \epsilon_{\text{P}}^{\frac{1}{p}} \epsilon_{\text{D}}^{\frac{p+1}{p}}.
$$

Combining this bound with (3.24) gives that

$$
\|r(x_{j,s},t_j)\| - \|r(x_{j,s+1},t_j)\| \geq [\kappa_{\text{CC}}^{\mu,s}]^{-1} 2^{-\frac{p+1}{p}} \delta^{\frac{1}{p}} \min\left[\epsilon_{\text{P}}, \epsilon_{\text{P}}^{\frac{1}{p}} \epsilon_{\text{D}}^{\frac{p+1}{p}}\right].
$$

As a consequence, the number of successful iterations of the ARpCC algorithm needed to compute $x_{k+\ell(k)}$ from x_k cannot exceed

$$
\kappa_{\text{CC}}^{\mu,s} 2^{\frac{p+1}{p}} \delta^{-\frac{1}{p}} \left\lceil \frac{\epsilon_{\text{P}} - \delta\epsilon_{\text{P}}}{\min\left[\epsilon_{\text{P}}, \epsilon_{\text{P}}^{\frac{1}{p}} \epsilon_{\text{D}}^{\frac{p+1}{p}}\right]} \right\rceil = (1-\delta)\,\kappa_{\text{CC}}^{\mu,s} 2^{\frac{p+1}{p}} \delta^{-\frac{1}{p}} \max\left[1, \epsilon_{\text{P}}^{\frac{p-1}{p}} \epsilon_{\text{D}}^{-\frac{p+1}{p}}\right].
$$

We now use Lemma 3.5 again and invoke Lemma 2.1 to account for possible unsuccessful inner iterations, yielding that the total number of successful and unsuccessful iterations of the ARpCC algorithm necessary to deduce $x_{k+\ell(k)}$ from x_k is bounded above by

$$
\kappa_u(1-\delta) 2^{\frac{p+1}{p}} \delta^{-\frac{1}{p}} \kappa_{\text{CC}}^{\mu,s} \max\left[1, \epsilon_{\text{P}}^{\frac{p-1}{p}} \epsilon_{\text{D}}^{-\frac{p+1}{p}}\right] \stackrel{\text{def}}{=} (1-\delta)\kappa_{\text{CC}}^{\mu} \max\left[1, \epsilon_{\text{P}}^{\frac{p-1}{p}} \epsilon_{\text{D}}^{-\frac{p+1}{p}}\right].
$$

$\square$

We now state a useful property of the set $\mathcal{F}$.

Lemma 3.7 For arbitrary $x \in \mathcal{F}$, $v \in \mathbb{R}^n$ and $\tau \in \mathbb{R}$ with $\tau \geq 1$,

$$
\|P_{\mathcal{F}}[x + \tau v] - x\| \leq \tau \|P_{\mathcal{F}}[x + v] - x\|.
$$

Proof. The result follows immediately from [1, Lem.2.3.1] which states that $\|P_{\mathcal{F}}[x + \tau v] - x\|/\tau$ is a monotonically non-increasing function of $\tau > 0$ for any x in a given convex set $\mathcal{F}$. □

We finally combine our results in a final theorem stating our evaluation complexity bound for the ARpGC algorithm.

Theorem 3.8 Suppose that AS.1, and AS.3–AS.7 hold. Then, for some constants κ_{CC}^{c} and κ_{CC}^{μ} independent of ϵ_{P} and ϵ_{D}, the ARpGC algorithm applied to problem (3.1) needs at most

$$\left\lfloor \left(\kappa_{CC}^{c} \|c(x_0)\| + \kappa_{CC}^{\mu} [f(x_1) - f_{\text{low}} + 1] \right) \max\left[\epsilon_{P}^{-1}, \epsilon_{P}^{-\frac{1}{p}} \epsilon_{D}^{-\frac{p+1}{p}} \right] + 1 \right\rfloor \qquad (3.25)$$

evaluations of f, c and their derivatives up to order p to compute a point x_ϵ such that either

$$\|c(x_\epsilon)\| > \delta \epsilon_{P} \quad \text{and} \quad \pi_{\frac{1}{2}\|c\|^2}(x_\epsilon) \leq \epsilon_{D} \|c(x_\epsilon)\| \qquad (3.26)$$

or

$$\|c(x_\epsilon)\| \leq \epsilon_{P} \quad \text{and} \quad \pi_{\Lambda}(x_\epsilon, y_\epsilon) \leq \epsilon_{D} \|(y_\epsilon, 1)\| \qquad (3.27)$$

where $\Lambda(x, y) \overset{\text{def}}{=} f(x) + y^T c(x)$ is the Lagrangian with respect to the equality constraints and y_ϵ is a vector of Lagrange multipliers associated with the equality constraints.

Proof. If the ARpGC algorithm terminates in Phase 1, we immediately obtain that (3.26) holds, and Lemma 3.2 then ensures that the number of evaluations of c and its derivatives cannot exceed

$$\left\lfloor \kappa_{CC}^{c} \|c(x_0)\| \max\left[\epsilon_{P}^{-1}, \epsilon_{P}^{-\frac{1}{p}} \epsilon_{D}^{-\frac{p+1}{p}} \right] \right\rfloor + 1. \qquad (3.28)$$

The conclusions of the theorem therefore hold in this case.

Let us now assume that termination does not occur in Phase 1. Then the ARpGC algorithm must terminate after a number of evaluations of f and c and their derivatives which is bounded above by the upper bound on the number of evaluations in Phase 1 given by (3.28) plus the bound on the number of evaluations of μ given by (3.22) and Lemma 3.6. Using the fact that $\lfloor a \rfloor + \lfloor b \rfloor \leq \lfloor a + b \rfloor$ for $a, b \geq 0$ and $\lfloor a + i \rfloor = \lfloor a \rfloor + i$ for $a \geq 0$ and $i \in \mathbb{N}$, this yields the combined upper bound

$$\left\lfloor \kappa_{CC}^{c} \|c(x_0)\| \max\left[\epsilon_{P}^{-1}, \epsilon_{P}^{-\frac{1}{p}} \epsilon_{D}^{-\frac{p+1}{p}} \right] \right.$$
$$\left. + \left[(1-\delta) \kappa_{CC}^{\mu} \epsilon_{P}^{\frac{p-1}{p}} \max\left[1, \epsilon_{P}^{\frac{p-1}{p}} \epsilon_{D}^{-\frac{p+1}{p}} \right] \right] \times \left[\frac{f(x_1) - f_{\text{low}} + 1}{1 - \delta} \epsilon_{P}^{-1} \right] \right\rfloor + 2,$$

and (3.25) follows.

Remember now that (3.18) holds at termination of Phase 2, and therefore that

$$\epsilon_{P} \geq \|r(x_\epsilon, t_\epsilon)\| \geq \delta \epsilon_{P}. \qquad (3.29)$$

Moreover, we also obtain from (3.18) that

$$|P_{\mathcal{F}}[x_\epsilon - J(x_\epsilon)^T c(x_\epsilon) - (f(x_\epsilon) - t_k)\nabla_x f(x_\epsilon)] - x_\epsilon\| = \pi_\mu(x_\epsilon, t_\epsilon) \le \epsilon_{\mathrm{D}} \|r(x_\epsilon, t_\epsilon)\|. \quad (3.30)$$

Assume first that $f(x_\epsilon) = t_\epsilon$. Then, using the definition of $r(x, t)$, we deduce that

$$\pi_{\frac{1}{2}\|c\|^2}(x_\epsilon) = \|P_{\mathcal{F}}[x_\epsilon - J(x_\epsilon)^T c(x_\epsilon)] - x_\epsilon\| \le \epsilon_{\mathrm{D}} \|c(x_\epsilon)\|$$

and (3.26) is again satisfied because (3.29) gives that $\|c(x_\epsilon)\| = \|r(x_\epsilon, t_\epsilon)\| \ge \delta\epsilon_{\mathrm{P}}$.

Assume now that $f(x_\epsilon) > t_\epsilon$ (the case where $f(x_\epsilon) < t_\epsilon$ is excluded by (3.18)) and note that

$$0 < f(x_\epsilon) - t_\epsilon \le \epsilon_{\mathrm{P}} \le 1$$

because of the second bound in (3.16) and the decreasing nature of $\|r(x, t_k)\|$ during inner iterations. Defining now

$$y_\epsilon \overset{\mathrm{def}}{=} \frac{c(x_\epsilon)}{f(x_\epsilon) - t_\epsilon}.$$

and successively using Lemma 3.7 with $x = x_\epsilon$, $v = -(J(x_\epsilon)^T c(x_\epsilon) + (f(x_\epsilon) - t_\epsilon)\nabla_x f(x_\epsilon))$ and $\tau = 1/(f(x_\epsilon) - t_\epsilon) \ge 1$, the third part of (3.5), (3.29) and the definition of $r(x, t)$, we deduce that

$$
\begin{aligned}
\pi_\Lambda(x_\epsilon, y_\epsilon) &= \left\| P_{\mathcal{F}}\left[x_\epsilon - J(x_\epsilon)^T \frac{c(x_\epsilon)}{f(x_\epsilon) - t_\epsilon} - \nabla_x f(x_\epsilon)\right] - x_\epsilon \right\| \\
&\le \frac{1}{f(x_\epsilon) - t_\epsilon} \|P_{\mathcal{F}}[x_\epsilon - J(x_\epsilon)^T c(x_\epsilon) - (f(x_\epsilon) - t_\epsilon)\nabla_x f(x_\epsilon)] - x_\epsilon\| \\
&= \frac{1}{f(x_\epsilon) - t_\epsilon} \pi_\mu(x_\epsilon, t_\epsilon) \\
&\le \epsilon_{\mathrm{D}} \frac{\|r(x_\epsilon, t_\epsilon)\|}{f(x_\epsilon) - t_\epsilon} \\
&= \epsilon_{\mathrm{D}} \|(y_\epsilon, 1)\|
\end{aligned}
$$

This implies (3.27) since $\|c(x_\epsilon)\| \le \|r(x_\epsilon, t_\epsilon)\| \le \epsilon$. $\qquad\square$

Note that the bound (3.25) is $O(\epsilon^{-\frac{p+2}{p}})$ whenever $\epsilon_{\mathrm{P}} = \epsilon_{\mathrm{D}} = \epsilon$.

It is important to note that the complexity bound given by Theorem 3.8 depends on $f(x_1)$, the value of the objective function at the end of Phase 1. Giving an upper bound on this quantity is in general impossible, but can be done in some case. A trivial bound can of course be obtained if $f(x)$ is bounded above in $\mathcal{C}_\beta$. This has the advantage of providing a complexity result which is self-contained (in that it only involves problem-dependent quantities), but it is quite restrictive as it excludes, for instance, problems only involving equality constraints ($\mathcal{F} = \mathbb{R}^n$) and coercive objective functions. A bound is also readily obtained if the set $\mathcal{F}$ is itself bounded (for instance when the variables are subject to finite lower and upper bounds) or if one assumes that the iterates generated by Phase 1 remain bounded. This may for example be the case if the set $\{x \in \mathbb{R}^n \mid c(x) = 0\}$ is bounded. An ϵ_{P}-dependent bound can finally be obtained if one is ready to assume that all derivatives of order 1 to p of $c(x)$ (and thus of $\nu(x)$) are bounded by a constant in the level set $\mathcal{C}_0 \overset{\mathrm{def}}{=} \{x \in \mathcal{F} \mid \|c(x)\| \le \|c(x_0)\|\}$ because it can then be shown that $\|s_k\|$ is uniformly bounded above and hence that $\|x_1 - x_0\|$ is itself bounded above by a constant times the (ϵ_{P}-dependent) number

of iterations in Phase 1 given by Lemma 3.2. Using the boundedness of the gradient of $\nu(x)$ on the path of iterates then ensures the (extremely pessismistic) upper bound $f(x_1) = f(x_0) + O\left(\max\left[\epsilon_P^{-1}, \epsilon_P^{-\frac{1}{p}}\epsilon_D^{-\frac{p+1}{p+1-q}}\right]\right)$. Substituting this bound in (3.25) in effect squares the complexity of obtaining (x_ϵ, t_ϵ).

4 Discussion

We have first shown in Section 2 that, if derivatives of the objective function up to order p can be evaluated and if the p-th one is Lipschitz continuous, then the ARpCC algorithm applied to the convexly constrained problem (2.1) needs at most $O(\epsilon^{\frac{p+1}{p}})$ evaluations of f and its derivatives to compute an ϵ-approximate first-order critical point. This worst-case bound corresponds to that obtained in [4] when $p = 2$, but with significantly weaker assumptions. Indeed, the present proposal no longer needs any assumption on the number of descent steps in the subproblem solution, the iterates are no longer assumed to remain in a bounded set and the Lipschitz continuity of the gradient is no longer necessary. That these stronger results are obtained as the result of a considerably simpler analysis is an added bonus. While we have not developed here the case (covered for $p = 2$ in [4]) where the p-th derivative is only known approximately (in the sense that $\nabla_x^p f(x_k)$ is replaced in the model's expression by some tensor B_k such that the norm of $(\nabla_x^p f(x_k) - B_k)$ applied $p - 1$ times to s_k must be $O(\|s_k\|^p)$), the generalization of the present proposal to cover this situation is easy.

The proposed worst-case evaluation bound also generalizes that of [3] for unconstrained optimization to the case of set-constrained problems, under very weak assumptions on the feasible set. As was already the case for $p \leq 2$, it is remarkable that the complexity bound for the considered class of problems (which includes the standard bound constrained case) is, for all $p \geq 1$, identical in order to that of unconstrained problems.

The present framework for handling convex constraints is however not free of limitations, resulting from the choice to transfer difficulties associated with the original problem to the subproblem solution, thereby sparing precious evaluations of f and its derivatives. The first is that we need to compute projections onto the feasible set to obtain values of π_f and π_{m_k}. While this is straightforward and computationally inexpensive for simple convex sets such boxes, spheres, cylinders or the order-simplex, the process might be more intensive for the general case. The second limitation is that, even if the projections can be computed, the approximate solution of the subproblem may also be very expensive in terms of internal calculations (we do not consider here suitable algorithms for this purpose). Observe nevertheless that, crucially, neither the computation of the projections nor the subproblem solution involve evaluating the objective function or its derivatives: despite their potential computational drawbacks, they have therefore no impact on the evaluation complexity of the original problem. However, as the cost of evaluating any constraint function/derivative possibly necessary for computing projections is neglected by the present approach, it must therefore be seen as a suitable framework to handle "cheap inequality constraint" such as simple bounds.

We have also shown in Section 3 that the evaluation complexity of finding an approximate first-order scaled critical point for the general smooth nonlinear optimization problem involving both equality and inequality constraints is at most $O(\epsilon_P^{-1/p}\epsilon_D^{-(p+1)/p})$ evaluations of the objective function, constraints and their derivatives up to order p. We refer here to an "approximate scaled critical point" in that such a point is required to satisfy (3.26) or (3.27),

where the accuracy is scaled by the size of the constraint violation or that of the Lagrange multipliers. Because this bound now only depends on the assumptions necessary to prove the evaluation complexity bound for the ARpCC algorithm in Section 2, it therefore strengthens and generalizes that of [7] since the latter directly hinges on [4]. Moreover, it also corrects an unfortunate error[1] in [7], that allows a vector of Lagrange multipliers whose sign is arbitrary (in line with a purely first-order setting where minimization and maximization are not distinguished). The present analysis now yields the multiplier with the sign associated with minimization.

Interestingly, an $O(\epsilon_{\mathrm{P}}\epsilon_{\mathrm{D}}^{-(p+1)/p}\min[\epsilon_{\mathrm{D}},\epsilon_{\mathrm{P}}]^{-(p+1)/p})$ evaluation complexity bound was also proved by Birgin, Gardenghi, Martínez, Santos and Toint in [2] for *unscaled*, standard KKT conditions and in the least expensive of three cases depending on the degree of degeneracy identifiable by the algorithm[2]. Even if the bounds for the scaled and unscaled cases coincide in order when $\epsilon_{\mathrm{P}} \leq \epsilon_{\mathrm{D}}$, comparing the two results is however not straightforward. On one hand the scaled conditions take into account the possibly different scaling of the objective function and constraints. On the other hand the same scaled conditions may result in earlier termination with (3.27) if the Lagrange multipliers are very large, as (3.27) is then consistent with the weaker requirement of finding a John's point. But the framework discussed in the present paper also differs from that of [2] in two additional significant ways. The first is that the present one provides a potentially stronger version of the termination of the algorithm at infeasible points (in Phase 1): indeed the second part of (3.26) can be interpreted as requiring that the size of the feasible gradient of $\|c(x)\|$ is below ϵ_{D}, while [2] considers the gradient of $\|c(x)\|^2$ instead. The second is that, if termination occurs in Phase 2 for an x_ϵ such that $\pi_{\frac{1}{2}\|c\|^2}(x_\epsilon) = \|J(x_\epsilon)^T c(x_\epsilon)\|$ is itself of order $\epsilon_{\mathrm{P}}\epsilon_{\mathrm{D}}$ (thereby covering the case where $f(x_\epsilon) = t_k$ discussed in Theorem 3.8) , then Birgin *et al.* show that the Łojaciewicz inequality [10] must fail for c in the limit for ϵ_{P} and ϵ_{D} tending to zero (see [2] for details). This observation is interesting because smooth functions satisfy the Łojaciewicz inequality under relatively weak conditions, implying that termination in these circumstances is unlikely. The same information is also obtained in [2], albeit at the price of worsening the evaluation complexity bound mentioned above by an order of magnitude in ϵ_{D}. We also note that the approach of [2] requires the minimization, at each iteration, of a residual whose second derivatives are discontinuous, while all functions used in the present paper are p times continuously differentiable. A final difference between the two approaches is obviously our introduction of π_Λ and $\pi_{\frac{1}{2}\|c\|^2}$ in the expression of the criticality condition in Theorem 3.8 for taking the inequality constraints into account.

We conclude our discussion by a remark about criticality measures. At variance with [4] and [7], we have preferred, in this paper, to use the first-order criticality measure $\pi_f(x)$ rather than $\chi_f(x)$, the decrease in the linearized function in the intersection of the feasible set and the ball of radius one. While a similar result can indeed be obtained for this latter measure (in this case for general closed non-empty convex sets even in Section 3), our choice is motivated by the observation made by Gratton, Mouffe and Toint [9] that π_f is backward stable as a criticality measure (in the sense that an approximate solution of problem (2.1) yielding a small value of π_f can be interpreted as the exact solution of a neighbouring problem), while this is not the case for χ_f even when $\mathcal{F}$ is described by simple bounds on the problem's variables.

---

[1] The second equality in the first equation of Lemma 3.4 in [7] only holds if one is ready to flip the gradient's sign if necessary.

[2] This result also assumes boundedness of $f(x_1)$.

References

[1] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, Massachusetts, USA, 2nd edition, 1999.

[2] 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. Technical Report naXys-08-2015, Namur Center for Complex Systems (naXys), University of Namur, Namur, Belgium, 2015.

[3] 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.

[4] 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 Journal of Numerical Analysis*, 32(4):1662–1645, 2012.

[5] 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, Series A*, 144(1):93–106, 2013.

[6] 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 Journal on Optimization*, 23(3):1553–1574, 2013.

[7] C. Cartis, N. I. M. Gould, and Ph. L. Toint. On the evaluation complexity of constrained nonlinear least-squares and general constrained nonlinear optimization using second-order methods. *SIAM Journal on Numerical Analysis*, 53(2):836–851, 2015.

[8] C. Cartis, N. I. M. Gould, and Ph. L. Toint. Improved worst-case evaluation complexity for potentially rank-deficient nonlinear least-Euclidean-norm problems using higher-order regularized models. Technical Report NA-TR 15-17, Numerical Analysis Group, Mathematical Institute, University of Oxford, 2015.

[9] S. Gratton, M. Mouffe, and Ph. L. Toint. Stopping rules and backward error analysis for bound-constrained optimization. *Numerische Mathematik*, 119(1):163–187, 2011.

[10] 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.

[11] Yu. Nesterov. *Introductory Lectures on Convex Optimization*. Applied Optimization. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.