Science & Technology
Facilities Council

On the evaluation complexity of constrained nonlinear least-squares and general constrained nonlinear optimization using second-order methods

C Cartis, NIM Gould, PhL Toint

April 2013

On the evaluation complexity of constrained nonlinear least-squares and general constrained nonlinear optimization using second-order methods

Coralia Cartis[1,2], Nicholas I. M. Gould[2,3,4] and Philippe L. Toint[2,5]

ABSTRACT

When solving the general smooth nonlinear optimization problem involving equality and/or inequality constraints, an approximate first-order critical point of accuracy ϵ can be obtained by a second-order method using cubic regularization in at most $O(\epsilon^{-3/2})$ problem-functions evaluations, the same order bound as in the unconstrained case. This result is obtained by first showing that the same result holds for inequality constrained nonlinear least-squares. As a consequence, the presence of (possibly nonlinear) equality/inequality constraints does not affect the complexity of finding approximate first-order critical points in nonconvex optimization. This result improves on the best known ($O(\epsilon^{-2})$) evaluation-complexity bound for solving general nonconvexly constrained optimization problems.

[1] School of Mathematics, The King's Buildings, University of Edinburgh, EH9 3JZ,
Scotland, EU. Email: coralia.cartis@ed.ac.uk .
Current reports available from "http://www.maths.ed.ac.uk/ERGO/preprints.html".

[2] This work was supported by the Royal Society through the International Joint Project 14265.

[3] Computational Science and Engineering Department, Rutherford Appleton Laboratory,
Chilton, Oxfordshire, OX11 0QX, England, EU. Email: nick.gould@stfc.ac.uk .
Current reports available from "http://www.numerical.rl.ac.uk/reports/reports.shtml".

[4] This work was supported by the EPSRC grant EP/I013067/1.

[5] Namur Center for Complex Systems (naXys) and Department of Mathematics,
University of Namur, 61, rue de Bruxelles, B-5000 Namur, Belgium, EU.
Email : philippe.toint@unamur.be .
Current reports available from "http://www.fundp.ac.be/~phtoint/pht/publications.html".

1 Introduction

The past few years have seen several contributions on the worst-case complexity of solving smooth but nonlinear and possibly nonconvex optimization problems. Starting with contributions of Vavasis (1993), Nesterov (2004) and Gratton, Sartenaer and Toint (2008) on (essentially) first-order methods for the unconstrained case, a significant step was made with the proposal by Nesterov and Polyak (2006) of a second-order method including cubic regularization terms. This paper showed that solving the smooth unconstrained nonlinear optimization method can be achieved in at most $O(\epsilon^{-3/2})$ problem-functions evaluations, if one is happy to terminate the process with an approximate first-order critical point at which the Euclidean norm of the objective function's gradient is at most a user-prescribed threshold $\epsilon \in (0,1)$. This is in contrast with what can be obtained for first-order methods, which require at most $O(\epsilon^{-2})$ problem-function evaluations in a similar context. This remarkable result by Nesterov was subsequently extended by Cartis, Gould and Toint (2011a, 2011b) to a wider class of algorithms, leading to the conclusion that the class of cubic regularization (ARC) methods and its evaluation complexity of order $O(\epsilon^{-3/2})$ are optimal for the smooth unconstrained nonlinear optimization problem in terms of problem-function evaluations.

Following up on these results for the unconstrained problem, the authors of this paper then examined the smooth constrained problem and showed that, somewhat surprisingly, the complexity in $O(\epsilon^{-2})$ problem-function evaluations obtained for first-order methods is not affected at all (in order) by the presence of equality and/or inequality constraints (see Cartis, Gould and Toint, 2012a, for the convex inequality case, and Cartis, Gould and Toint, 2013a, for the general case). Moreover, the first of these papers also showed that a complexity of order $O(\epsilon^{-3/2})$ problem-function evaluations can also be achieved under some conditions for the problem involving convex inequalities, while a very similar result was presented for the general equality constrained case in Cartis, Gould and Toint (2013b), provided one is ready to solve the primal more accurately than the dual. This left open the central question of whether the general smooth problem involving both equality and inequality constraints can be solved in at most $O(\epsilon^{-3/2})$ problem-functions evaluations using a cubic regularization method under the same conditions.

The purpose of the present note is to confirm this proposition, thereby providing a complete worst-case analysis for the computation of approximate first-order critical points in smooth optimization. This is achieved by exhibiting an algorithm for which the $O(\epsilon^{-3/2})$ upper complexity bound holds, and which is inspired by the two-stages methods of Cartis et al. (2013a) and Cartis et al. (2013b), combined with the projection technique described in Cartis et al. (2012a). Broadly speaking, it consists of a first phase where a least-squares formulation of the constraint violation is minimized subject to convex constraints, resulting in either an (approximate) feasible point or an local infeasible minimizer of the constraint violation, an outcome which cannot be excluded from our analysis barring the use of global optimization techniques. If an approximate feasible point is obtained, then a short-step target-following technique is used to reduce the objective function while preserving

(approximate) feasibility.

Because both phases of our algorithm crucially depend on the solution of a convexly-constrained nonlinear least-squares problem, we start by considering this minimization problem and its complexity in Section 2. We then turn to the general constrained case, present our two-phase method and analyze its complexity in Section 3. Some conclusions and perspectives are finally discussed in Section 4.

2   Constrained nonlinear least-squares problems

We start by considering the smooth constrained nonlinear least-squares problem given by

$$\min_{x \in \mathcal{F}} \phi(x) \overset{\text{def}}{=} \tfrac{1}{2}\|r(x)\|^2 \tag{2.1}$$

where r is a twice-continuously differentiable "residual" function from $\mathrm{I\!R}^n$ into $\mathrm{I\!R}^m$, $\|\cdot\|$ is the standard Euclidean norm and $\mathcal{F}$ is a non-empty closed convex set. Our objective is to apply the cubically regularized COCARC-S method described in Cartis et al. (2012*a*) for general nonconvex optimization subject to convex constraints to this problem and to specialize the corresponding complexity results.

2.1   A review of the COCARC-S algorithm

The COCARC-S algorithm is designed for minimizing a general nonlinear twice continuously differentiable objective function $f : \mathrm{I\!R}^n \to \mathrm{I\!R}$ within the closed convex set $\mathcal{F}$, that is

$$\min_{x \in \mathcal{F}} f(x) \tag{2.2}$$

Iteration k of the COCARC-S algorithm proceeds by first checking the approximate first-order criticality of the current iterate. This is achieved by testing if

$$\chi_f(x_k) \leq \epsilon \tag{2.3}$$

where $\epsilon > 0$ is a user-specified accuracy threshold and where

$$\chi_f(x) \overset{\text{def}}{=} \left| \min_{x+d \in \mathcal{F}, \|d\| \leq 1} \langle \nabla_x f(x), d \rangle \right| \tag{2.4}$$

($\langle \cdot, \cdot \rangle$ denotes the standard Euclidean inner-product) is the linearized decrease in f achievable inside a feasible neighbourhood of diameter one (Conn, Gould, Sartenaer and Toint, 1993, see also Yuan, 1985). If (2.3) fails, a step s_k is computed from the iterate x_k by (approximately) minimizing a cubic model of f of the form

$$m_k(x_k + s) \overset{\text{def}}{=} f(x_k) + \langle \nabla_x f(x_k), s \rangle + \tfrac{1}{2}\langle s, B_k s \rangle + \tfrac{1}{3}\sigma_k\|s\|^3 \tag{2.5}$$

subject to $x_k + s \in \mathcal{F}$, for a given 'regularization weight' $\sigma_k > 0$, and where B_k is a symmetric approximation of $\nabla_{xx} f(x_k)$. What is meant by approximate minimization of

the constrained model (2.5) is not detailed here (see AS7 in Cartis et al., 2012a for the full details, involving the reduction of a model-specific first-order criticality measure). For what follows, it is important to assume that this model minimization can be seen as performing a number ℓ_k of successive (possibly incomplete) line minimizations of the model m_k between feasible points (see AS8 in Cartis et al., 2012a). Once the trial point $x_k^+ = x_k + s_k$ is computed, the achieved reduction $f(x_k) - f(x_k^+)$ is compared to the predicted one, $f(x_k) - m_k(x_k^+)$. If the ratio of the former to the latter is sufficiently positive, x_k^+ is accepted as the next iterate x_{k+1} and the regularizaton weight is (possibly) decreased (an iteration where this occurs is called a *successful* iteration). If this ratio is not sufficiently positive, i.e. is below some constant $\eta_1 \in (0, 1)$, the trial point is rejected ($x_{k+1} = x_k$) and the regularization weight is increased (by a factor at least $\gamma_1 > 1$).

Algorithm 2.1: COCARC-S Algorithm for (2.2)

A starting point $x_0 \in \mathcal{F}$, a minimum regularization parameter $\sigma_{\min} > 0$, an initial regularization parameter $\sigma_0 \geq \sigma_{\min}$, and algorithmic parameters $\gamma_2 \geq \gamma_1 > 1$ and $1 > \eta_1 > 0$, as well as the tolerance $\epsilon \in (0, 1)$ are given.

Step 0: Check for termination. If $\chi_f(x_k) \leq \epsilon$, terminate.

Step 1: Computation of the step. Starting from x_k, approximately minimize $m_k(x_k + s)$ subject to the $x_k + s \in \mathcal{F}$, yielding a trial point $x_k^+ = x_k + s_k$.

Step 2: Acceptance of the trial point. Compute $f(x_k^+)$ and define

$$\rho_k = \frac{f(x_k) - f(x_k^+)}{m_k(x_k) - m_k(x_k^+)}. \tag{2.6}$$

If $\rho_k \geq \eta_1$, then $x_{k+1} = x_k^+$, else set $x_{k+1} = x_k$.

Step 3: Regularization parameter update. If $\rho_k \geq \eta_1$, choose

$$\sigma_{k+1} \in [\sigma_{\min}, \gamma_1 \sigma_k). \tag{2.7}$$

Otherwise, choose $\sigma_{k+1} \in [\gamma_1 \sigma_k, \gamma_2 \sigma_k]$.

Note that a feasible x_0 can be obtained by projection of any user-supplied initial guess onto the convex set $\mathcal{F}$. For future reference, we define $\mathcal{S}$ to be the index set of the successful iterations, that is

$$\mathcal{S} \stackrel{\text{def}}{=} \{k \geq 0 \mid \rho_k \geq \eta_1\}. \tag{2.8}$$

Denote by $\mathcal{X}$ the closed convex hull of all iterates x_k and trial points x_k^+.

What is the maximum number of COCARC-S iterations that can be necessary before an iterate x_k is found which satifies (2.3)? In order to answer this question, we now recall the assumptions used to derive that the required complexity results for the COCARC-S algorithm and those of these results that are of interest in our context. More specifically, we assume that

AS1 $\mathcal{F}$ is closed, convex and non-empty;

AS2f f is twice continuously differentiable, its gradient is uniformly Lipschitz continuous on $\mathcal{X}$ and its Hessian is "weakly" uniformly Lipschitz-continuous on the segments $[x_k, x_k^+]$, in the sense that there exists a constant $\kappa_{\mathrm{L}} \geq 0$ such that, for all k and all $y \in [x_k, x_k^+]$,

$$\|[\nabla_{xx} f(y) - \nabla_{xx} f(x_k)] s_k\| \leq \kappa_{\mathrm{L}} \|s_k\|^2;$$

AS3f   the Hessian $\nabla_{xx} f(x_k)$ is well approximated by B_k, in the sense that there exists a constant $\kappa_{\mathrm{B}} > 0$ such that, for all k,

$$\|[B_k - \nabla_{xx} f(x_k)] s_k\| \leq \kappa_{\mathrm{B}} \|s_k\|^2.$$

AS4 ℓ_k is bounded above by a constant independent of k and ϵ;

AS5 $\mathcal{X}$ is bounded.

Assumptions ASA, AS2f, AS3f and AS5 are relatively standard, even if AS5 may be judged to be somewhat restrictive. AS4 is really an assumption on the solution of the subproblem of minimizing the model (2.5) subject to convex constraints, which raises the question of a practical algorithm to perform this task. However, since this minimization involves the model and not the objective function itself, the complexity of the task does not affect the function-evaluation complexity of the COCARC-S algorithm. See Cartis et al. (2012a) for further discussion.

The above conditions guarantee that the following holds.

Lemma 2.1. Suppose that AS1, AS2f, AS3f, AS4 and AS5 hold. Then there exists a positive constant $\kappa_{\mathrm{C}} \in (0, 1)$ independent of k and ϵ such that

$$f(x_k) - f(x_{k+1}) \geq \eta_1 \kappa_{\mathrm{C}} \chi_f(x_{k+1})^{\frac{3}{2}}, \quad \text{for all } k \in \mathcal{S}. \tag{2.9}$$

Moreover, there exists a constant $\sigma_{\mathrm{succ}} > 0$ independent of k and ϵ such that, if $\sigma_k \geq \sigma_{\mathrm{succ}}$, then iteration k of the COCARC-S algorithm is successful ($k \in \mathcal{S}$). Finally, there exists a constant $\sigma_{\max} \geq \sigma_{\mathrm{succ}}$ such that $\sigma_k \leq \sigma_{\max}$ for all $k \geq 0$.

Proof. Relation (2.9) follows from Theorem 4.7 in Cartis et al. (2012*a*). In order to prove the second statement, we may follow the line of thought of Lemma 5.2 in Cartis et al. (2011*a*): we note that, using (2.5), AS2f and AS3f,

$$
\begin{aligned}
f(x_k^+) - m_k(x_k^+) &\leq \tfrac{1}{2}\|H(\xi_k) - H(x_k)\| \, \|s\|^2 + \tfrac{1}{2}\|(H(x_k) - B_k)s_k\| \, \|s_k\| - \tfrac{1}{3}\sigma_k\|s_k\|^3 \\
&\leq \left[\tfrac{1}{2}(\kappa_L + \kappa_B) - \tfrac{1}{3}\sigma_k\right]\|s_k\|^3,
\end{aligned}
$$

where ξ_k belongs to the segment $[x_k, x_k^+]$ and thus $\|\xi_k - x_k\| \leq \|s_k\|$. This relation shows that $f(x_k^+) \leq m_k(x_k^+)$, and hence that $k \in \mathcal{S}$, provided $\sigma_k \geq \kappa_{\text{succ}} = \tfrac{3}{2}(\kappa_L + \kappa_B)$. The final statement of the Lemma follows from Lemma 4.3 in Cartis et al. (2012*a*). □

2.2 Application to constrained nonlinear least-squares

We now return to considering the complexity of problem (2.1). In the spirit of Cartis et al. (2013*b*), this involves redefining a suitable termination criterion that exploits the particularity of the least-squares problem. In the latter paper, we have indeed argued that, for the unconstrained case ($\mathcal{F} = \mathrm{I\!R}^n$), it is advisable to replace the standard rule where the algorithm is terminated as soon as an iterate x_k is found such that

$$
\|\nabla_x \phi(x_k)\| = \|J(x_k)^T r(x_k)\| \leq \epsilon \tag{2.10}
$$

(where $J(x)$ denotes the Jacobian of r at x and $\epsilon \in (0,1)$ is a user defined accuracy threshold) by the rule that iterations are instead terminated as soon as an iterate x_k is found such that, for some user-defined accuracy thresholds $\epsilon_p \in (0,1)$ and $\epsilon_d \in (0,1)$,

$$
\|r(x_k)\| \leq \epsilon_p \quad \text{or} \quad \|\,\nabla_x\|r(x_k)\|\,\| \leq \epsilon_d, \tag{2.11}
$$

where

$$
\nabla_x\|r(x)\| \overset{\text{def}}{=} \begin{cases} \dfrac{J(x)^T r(x)}{\|r(x)\|} & \text{when} \quad r(x) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}
$$

The case corresponding to the first condition in (2.11) is the situation where an ϵ_p-approximate optimal point is found with 'zero' residual, while the second corresponds to the case where the residual at the approximate first-order critical point is nonzero.

If we now consider the constrained case ($\mathcal{F} \subset \mathrm{I\!R}^n$), making the gradient of the objective function small, as requested in (2.10), is no longer appropriate, because the solution might lie on the boundary of $\mathcal{F}$. In the spirit of (2.3), an alternative to (2.10) in the constrained case is to stop the COCARC-S algorithm when applied to (2.1) as soon as a point x_k is found such that

$$
\chi_\phi(x_k) \leq \epsilon, \tag{2.13}
$$

where

$$
\chi_\phi(x) = \left| \min_{x+d \in \mathcal{F}, \|d\| \leq 1} \langle J(x)^T r(x), d \rangle \right|. \tag{2.14}
$$

(see (2.4)). However, since we are considering a least-squares problem, we may apply the same reasoning as in the unconstrained case, and we therefore suggest terminating the COCARC-S algorithm as soon as an x_k is found in $\mathcal{F}$ such that

$$\|r(x_k)\| \leq \epsilon_p \quad \text{or} \quad \chi_{\|r\|}(x_k) = \left| \min_{x_k + d \in \mathcal{F}, \|d\| \leq 1} \langle \nabla_x \|r(x_k)\|, d \rangle \right| \leq \epsilon_d. \qquad (2.15)$$

Note that $\chi_{\|r\|}(x)$ is continuous as a function of x for $r(x) \neq 0$ and is zero if and only if x is a first-order critical point of problem (2.1). Condition (2.15) therefore replaces (2.11) in the constrained case.

In order to adapt this framework of Section 2.1 to the constrained nonlinear least-squares case (2.1), we first note that AS1, AS4 and AS5 need not to be modified[1]. AS2f and AS3f must however be reformulated in terms of the residual function r.

> **AS2**   Each r_i $(i = 1 \ldots, m)$ is twice continuously differentiable, uniformly Lipschitz continuous on $\mathcal{X}$ and its Hessian is "weakly" uniformly Lipschitz-continuous on the segments $[x_k, x_k^+]$, in the sense that there exists a constant $\kappa_{\mathrm{L}} \geq 0$ such that, for all k, all $y \in [x_k, x_k^+]$ and all $i = 1, \ldots, m$,
>
> $$\|[\nabla_{xx} r_i(y) - \nabla_{xx} r_i(x_k)]s_k\| \leq \kappa_{\mathrm{L}} \|s_k\|^2;$$
>
> Moreover, the Jacobian $J(x)$ is Lipschitz continuous on $\mathcal{X}$, in the sense that there exists a constant $\kappa_{\mathrm{J}} \geq 0$ such that, for all $x, y \in \mathcal{X}$,
>
> $$\|J(x) - J(y)\| \leq \kappa_{\mathrm{J}} \|x - y\|.$$

> **AS3**   The Hessian $\nabla_{xx}\phi(x_k)$ is well approximated by B_k, in the sense that there exists a constant $\kappa_{\mathrm{B}} > 0$ such that, for all k,
>
> $$\|[B_k - \nabla_{xx}\phi(x_k)]s_k\| \leq \kappa_{\mathrm{B}} \|s_k\|^2.$$

We refer the reader to the discussion in Cartis et al. (2013*b*) showing that these assumptions ensure AS2f with ϕ playing the role of f. Observe that, strictly speaking, the fact that the residuals $r_i(x)$ are twice continuously differentiable (AS.2) on a bounded set $\mathcal{X}$ (AS.5) is enough to ensure that r_i and $\nabla_x r_i$ are Lipschitz continuous on $\mathcal{X}$, but we prefer to require these properties explicitly for clarity.

Having reformulated our assumptions, we are now entitled to deduce that (2.9) (with $f = \phi$) holds as AS1–AS5 are satisfied. The next step is then to modify this lower bound in the spirit of Lemma 3.1 in Cartis et al. (2013*b*). Assume first that, for a given $\beta \in (0, 1)$, $\|r(x_{k+1})\| \leq \beta \|r(x_k)\|$. Then $k \in \mathcal{S}$ and

$$\|r(x_k)\| - \|r(x_{k+1})\| \geq (1 - \beta)\|r(x_k)\| \qquad (2.16)$$

---

[1]Except for the obvious change from ϵ to ϵ_p and ϵ_d in AS4.

and

$$\|r(x_k)\|^{\frac{1}{2}} - \|r(x_{k+1})\|^{\frac{1}{2}} \geq (1 - \sqrt{\beta})\|r(x_k)\|^{\frac{1}{2}} \geq \frac{(1 - \sqrt{\beta})}{\sqrt{\beta}}\|r(x_{k+1})\|^{\frac{1}{2}}. \tag{2.17}$$

If, on the other hand, $\|r(x_{k+1})\| > \beta\|r(x_k)\|$, we nevertheless know that, for $k \in \mathcal{S}$, $\|r(x_{k+1})\| = \|r(x_k^+)\| < \|r(x_k)\|$, and thus that

$$
\begin{aligned}
\|r(x_k)\| - \|r(x_{k+1})\| &= \frac{\|r(x_k)\|^2 - \|r(x_{k+1})\|^2}{\|r(x_k)\| + \|r(x_{k+1})\|} \\
&\geq \frac{\phi(x_k) - \phi(x_{k+1})}{\|r(x_k)\|} \\
&\geq \eta_1 \kappa_{\mathrm{C}} \left(\frac{\|r(x_{k+1})\|}{\|r(x_k)\|}\right)^{\frac{3}{2}} \|r(x_k)\|^{\frac{1}{2}} \left(\frac{\chi_\phi(x_{k+1})}{\|r(x_{k+1})\|}\right)^{\frac{3}{2}} \\
&> \eta_1 \kappa_{\mathrm{C}} \beta^{\frac{3}{2}} \|r(x_k)\|^{\frac{1}{2}} \left(\chi_{\|r\|}(x_{k+1})\right)^{\frac{3}{2}},
\end{aligned} \tag{2.18}
$$

where we have used (2.9) to obtain the penultimate inequality and (2.12), (2.4) and the inequality $\|r(x_{k+1})\| > \beta\|r(x_k)\|$ to obtain the last. Hence, using this latter inequality again, we have that

$$\sqrt{\|r(x_k)\|} - \sqrt{\|r(x_{k+1})\|} = \frac{\|r(x_k)\| - \|r(x_{k+1})\|}{\sqrt{\|r(x_k)\|} + \sqrt{\|r(x_{k+1})\|}} \geq \tfrac{1}{2}\eta_1 \kappa_{\mathrm{C}} \beta^{\frac{3}{2}} \left(\chi_{\|r\|}(x_{k+1})\right)^{\frac{3}{2}} \tag{2.19}$$

As a consequence, we conclude from (2.16) and (2.18) that

$$\|r(x_k)\| - \|r(x_{k+1})\| \geq \min[\eta_1 \kappa_{\mathrm{C}} \beta^{\frac{3}{2}}, (1-\beta)] \min[\|r(x_k)\|^{\frac{1}{2}} \left(\chi_{\|r\|}(x_{k+1})\right)^{\frac{3}{2}}, \|r(x_k)\|] \tag{2.20}$$

Similarly, we deduce from (2.17) and (2.19) that

$$\|r(x_k)\|^{\frac{1}{2}} - \|r(x_{k+1})\|^{\frac{1}{2}} \geq \kappa_\phi \min[\left(\chi_{\|r\|}(x_{k+1})\right)^{\frac{3}{2}}, \|r(x_{k+1})\|^{\frac{1}{2}}],$$

where $\kappa_\phi \stackrel{\text{def}}{=} \min[\tfrac{1}{2}\eta_1 \kappa_{\mathrm{C}} \beta^{\frac{3}{2}}, \beta^{-\frac{1}{2}} - 1]$. Thus, as long as the COCARC-S algorithm applied to problem (2.1) is not terminated, i.e. as long as (2.15) is violated, we have that, for $k \in \mathcal{S}$,

$$\|r(x_k)\|^{\frac{1}{2}} - \|r(x_{k+1})\|^{\frac{1}{2}} \geq \kappa_\phi \min[\epsilon_d^{\frac{3}{2}}, \epsilon_p^{\frac{1}{2}}]. \tag{2.21}$$

Because, obviously, $0 \leq \|r(x_k)\|^{\frac{1}{2}} \leq \|r(x_0)\|^{\frac{1}{2}}$ for all k, we deduce that, provided AS1–AS5 hold, there are at most

$$\left\lfloor \frac{\|r(x_0)\|^{\frac{1}{2}}}{\kappa_\phi \min[\epsilon_d^{\frac{3}{2}}, \epsilon_p^{\frac{1}{2}}]} \right\rfloor$$

successful iterations until the COCARC-S algorithm applied to the constrained nonlinear least-squares problem (2.1) finds an iterate x_k such that (2.15) holds at x_{k+1}. We also observe that the final statement of Lemma 2.1 and Theorem 2.1 in Cartis et al. (2011a)

then allow us to conclude that the total number of (successful and unsuccessful) iterations required by algorithm COCARC-S to find x_j is bounded above by

$$\left\lfloor \frac{\kappa_{\mathrm{CNLS}}}{\min[\epsilon_d^{\frac{3}{2}}, \epsilon_p^{\frac{1}{2}}]} \right\rfloor \tag{2.22}$$

with

$$\kappa_{\mathrm{CNLS}} \overset{\mathrm{def}}{=} \frac{\|r(x_0)\|^{\frac{1}{2}}}{\kappa_\phi} + \left(1 + \frac{\|r(x_0)\|^{\frac{1}{2}}}{\kappa_\phi}\right) \frac{\log(\sigma_{\max}/\sigma_{\min})}{\log \gamma_1}. \tag{2.23}$$

We summarize our findings in the form of the following theorem.

Theorem 2.2. Assume that AS1–AS5 hold. Consider $\epsilon_p, \epsilon_d \in (0, 1)$. Then there is a constant $\kappa_{\mathrm{CNLS}} > 0$ whose expression is given by (2.23) such that the COCARC-S algorithm applied to problem (2.1) requires at most

$$\left\lfloor \kappa_{\mathrm{CNLS}} \max[\epsilon_d^{-\frac{3}{2}}, \epsilon_p^{-\frac{1}{2}}] \right\rfloor \tag{2.24}$$

iterations (and evaluations of r and possibly its derivatives) to find an iterate x_k such that

$$\|r(x_k)\| \leq \epsilon_p \quad \text{or} \quad \chi_{\|r\|}(x_k) \leq \epsilon_d. \tag{2.25}$$

Note that this bound is identical in order to that obtained by Cartis et al. (2013*b*) (Theorem 3.2) for the *unconstrained* nonlinear least-squares problem ($\mathcal{F} = \mathrm{IR}^n$). Moreover, provided $\epsilon_p \geq \epsilon_d^3$, then the number of COCARC-S iterations is bounded above by $O(\epsilon_d^{-3/2})$, which is the same complexity as that of solving the general *unconstrained* nonlinear optimization problem with the ARC cubic regularization algorithm (see Nesterov and Polyak, 2006, and Cartis, Gould and Toint, 2012*b*).

3   The general nonlinear optimization problem

Having considered the constrained nonlinear least-squares case, we now turn to the general nonlinear optimization problem. Without loss of generality, we consider it in the form

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

where $\mathcal{F}$ is again a closed non-empty convex set. Of particular interested is the case where $\mathcal{F} \overset{\mathrm{def}}{=} \{x \in \mathrm{IR}^n \mid x \geq x_{\mathrm{low}}\}$, where x_{low} is a vector of lower bounds with some components being possible equal to $-\infty$ and the inequality in this definition being understood componentwise. Indeed, it is well-known that this more specific formulation covers that involving

explicit inequality constraints, and can be obtained from the former by the incorporation of slack variables.

Both phases of our proposed analysis critically depend on applying the COCARC-S algorithm, first to the squared norm of the constraint violation

$$\theta(x) \stackrel{\text{def}}{=} \tfrac{1}{2}\|c(x)\|^2 \quad \text{for} \quad x \in \mathcal{F}, \tag{3.2}$$

terminating the computation as soon as a point $x_1 \in \mathcal{F}$ is found such that, for some user-defined accuracy thresholds $\epsilon_p \in (0,1)$ and $\epsilon_d \in (0,1)$,

$$\|c(x_1)\| \leq \delta\epsilon_p \quad \text{or} \quad \chi_{\|c\|}(x_1) \leq \epsilon_d, \tag{3.3}$$

(for some $\delta \in (0,1)$), and subsequently to a sequence of suitably defined least-squares problems whose objective function is denoted by

$$\mu(x, t_k) \stackrel{\text{def}}{=} \tfrac{1}{2}\|r(x, t_k)\|^2 \stackrel{\text{def}}{=} \tfrac{1}{2}\left\|\begin{pmatrix} c(x) \\ f(x) - t_k \end{pmatrix}\right\|^2 \quad \text{for} \quad x \in \mathcal{F}, \tag{3.4}$$

for some monotonically decreasing sequence of "targets" t_k ($k = 1, \ldots$).

We now describe our two-phases algorithm as Algorithm 3.1, where we use the symbol $\mathcal{P}$ to denote the projection onto the set $\mathcal{F}$. Note that the iterations in Step 2(a) of Phase 2 correspond to applying the COCARC-S algorithm (ignoring the termination test in Step 0) for each new value of the target t_k until the first successful iteration occurs. Also observe that $\chi_{\|r(\cdot,t)\|}(x_{k+1}, t_k)$ (in Step 2(b)) is well-defined when $\|r(x_{k+1}, t_k)\| > \delta\epsilon_p > 0$.

Analyzing the worst-case behaviour of this algorithm once more requires to specify the necessary assumptions. As above, we denote by $\mathcal{X} \subseteq \mathcal{F}$ the closed convex hull of all Phase 1 and Phase 2 iterates and trials points, and by $\mathcal{X}_2 \subseteq \mathcal{X}$ that of all Phase 2 iterates and trials points.

AS6 The function c is twice continuously differentiable on an open neighbourhood of $\mathcal{X}$ and f is twice continuously differentiable in an open neighbourhood of $\mathcal{X}_2$.

AS7 The components c_i ($i = 1, \ldots, m$) and the Jacobian $J(x)$ is globally Lipschitz continuous in $\mathcal{X}$ with Lipschitz constants $L_{c_i} > 0$ and $L_J > 0$, respectively. The components $\nabla^2 c_i(x)$ are weakly Lipschitz continuous on the segments $[x_k, x_k^+]$ (for both phases) with Lipschitz constant L_{H,c_i}, for $i \in \{1, \ldots, m\}$.

AS8 $f(x)$, and $g(x)$ are Lipschitz continuous in $\mathcal{X}_2$ with Lipschitz constants L_f and $L_{g,f} > 0$, respectively. Moreover, $\nabla^2 f(x)$ is weakly Lipschitz continuous on all Phase 2 segments $[x_k, x_k^+]$ with Lipschitz constant $L_{H,f}$.

AS.9 The objective $f(x)$ is bounded above and below in a neighbourhood of the feasible set, that is there exist constants $\alpha > 0$, f_{low} and $f_{\text{up}} \geq f_{\text{low}} + 1$ such that

$$f_{\text{low}} \leq f(x) \leq f_{\text{up}} \quad \text{for all} \quad x \in \mathcal{F} \cap \mathcal{C}_\alpha.$$

Algorithm 3.1: Short-Step ARC Algorithm for (3.1)

A starting point x_0, a minimum regularization parameter $\sigma_{\min} > 0$, an initial regularization parameter $\sigma_0 \geq \sigma_{\min}$, a parameter $\delta \in (0, 1)$, as well as the tolerances $\epsilon_p \in (0, 1)$ and $\epsilon_d \in (0, 1)$, are given.

Phase 1:

   Starting from $\mathcal{P}(x_0)$, apply the COCARC-S algorithm to minimize $\theta(x)$ subject to $x \in \mathcal{F}$ until a point $x_1 \in \mathcal{F}$ is found at which (3.3) holds. If $\|c(x_1)\| > \delta\epsilon_p$, terminate.

Phase 2:

   1. Set $t_1 = f(x_1) - \sqrt{\epsilon_p^2 - \|c(x_1)\|^2}$ and $k = 1$.

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

      2a. Loop on Steps 1 to 3 of the COCARC-S algorithm to minimize $\mu(x, t_k)$ as a function of $x \in \mathcal{F}$ until a successful iteration is obtained, yielding a new iterate $x_{k+1} \in \mathcal{F}$ and a new value of the regularization parameter $\sigma_{k+1} \geq \sigma_{\min}$.

      2b. If $\|r(x_{k+1}, t_k)\| > \delta\epsilon_p$, then terminate if $\chi_{\|r(\cdot,t)\|}(x_{k+1}, t_k) \leq \epsilon_d$. Otherwise (i.e., if either $\|r(x_{k+1}, t_k)\| \leq \delta\epsilon_p$ or $\chi_{\|r(\cdot,t)\|}(x_{k+1}, t_k) > \epsilon_d$ and $\|r(x_{k+1}, t_k)\| > \delta\epsilon_p$), set

$$t_{k+1} = f(x_{k+1}) - \sqrt{\|r(x_k, t_k)\|^2 - \|r(x_{k+1}, t_k)\|^2 + (f(x_{k+1}) - t_k)^2}.$$
$$(3.5)$$

   where

$$\mathcal{C}_\alpha = \{x \in \mathbb{R}^n \mid \|c(x)\| \leq \alpha\}. \qquad (3.6)$$

AS10 ℓ_k is bounded above by a constant independent of k and ϵ in all constrained cubic model minimizations (in both phases of the algorithm),

AS11 $\mathcal{X}$ is bounded.

   From here on, our analysis is nearly identical to that presented in Cartis et al. (2013*b*). We know from the previous section that Phase 1 of Algorithm 3.1 will terminate in a number of iterations (and function evaluations) as given by (2.24). Let us now consider Phase 2, and exploit again the least-squares structure of the minimizations carried on in Step 2. We obtain the following properties.

Lemma 3.1. Suppose that $\epsilon_p \leq \alpha$. Then in every Phase 2 iteration $k \geq 1$ of Algorithm 3.1 we have that

$$t_k \geq t_{k+1}, \tag{3.7}$$

$$f(x_k) - t_k \geq 0, \tag{3.8}$$

$$\|r(x_k, t_k)\| = \epsilon_p, \tag{3.9}$$

$$\|c(x_k)\| \leq \epsilon_p \quad \text{and} \quad |f(x_k) - t_k| \leq \epsilon_p, \tag{3.10}$$

and so $x_k \in \mathcal{F} \cap \mathcal{C}_\alpha$. In addition, if AS6–AS8 hold, each Phase 2 iteration requires at most $\kappa_\mathcal{S}$ problem function (and derivatives) evaluations, where $\kappa_\mathcal{S} \geq 1$ is a constant independent of k.

Proof. See Lemma 4.1 in Cartis et al. (2013b) for the proof of (3.7)–(3.10) and the conclusion that $x_k \in \mathcal{C}_\alpha$. But $x_k \in \mathcal{F}$ for all k, and thus $x_k \in \mathcal{F} \cap \mathcal{C}_\alpha$. Consider now a Phase 2 iteration in Algorithm 3.1. In each such iteration, Steps 1 to 3 of the COCARC-S Algorithm 2.1 are iterated upon until an (inner) successful iteration is obtained in Step 2. Because the regularization parameter is increased by a factor at least $\gamma_1 > 1$ at unsuccessful (inner) iterations, because $\sigma_k \geq \sigma_{\min}$ and because of the first part of Lemma 2.1, we know that at most

$$\kappa_\mathcal{S} \stackrel{\text{def}}{=} \max\left[\frac{\log(\sigma_{\text{succ}}/\sigma_{\min})}{\log\gamma_1}, 1\right]$$

inner iterations of this algorithm are necessary to compute x_{k+1}. Since each of these iterations involves a single problem function evaluation (in Step 2), we obtain the desired result. □

We may the pursue our analysis exactly as in Cartis et al. (2013b) and deduce the following important result on the decrease in t_k, directly inspired from Lemma 5.3 in that reference.

Lemma 3.2. Suppose that AS3 (with $\phi = r(\cdot, t_k)$) and AS6-AS11 hold and that

$$\epsilon_p \leq \alpha \quad \text{and} \quad \epsilon_d \leq \epsilon_p^{1/3}. \tag{3.11}$$

Then, for every Phase 2 iteration k of Algorithm 3.1, we have that,

$$t_k - t_{k+1} \geq \kappa_t \epsilon_d^{3/2} \epsilon_p^{1/2} \tag{3.12}$$

for some constant $\kappa_t \in (0, 1)$ independent of k, ϵ_d and ϵ_p.

Proof. We consider two cases. The first is when $\|r(x_{k+1}, t_k)\| \leq \delta\epsilon_p$. In this case, we have, from (3.5), (3.4) and (3.9) that

$$
\begin{aligned}
t_k - t_{k+1} &= -(f(x_{k+1}) - t_k) + \sqrt{\|r(x_k, t_k)\|^2 - \|c(x_{k+1})\|^2} \\
&= -(f(x_{k+1}) - t_k) + \sqrt{\epsilon_p^2 - \|c(x_{k+1})\|^2}.
\end{aligned}
\tag{3.13}
$$

It also follows from (3.4) that

$$
(f(x_{k+1}) - t_k)^2 + \|c(x_{k+1})\|^2 = \|r(x_{k+1}, t_k)\|^2 \leq \delta^2 \epsilon_p^2.
\tag{3.14}
$$

Minimizing (3.13) subject to (3.14) (see Lemma 5.2 in Cartis et al., 2013*b*) then yields that

$$
t_k - t_{k+1} \geq -\delta\epsilon_p + \epsilon_p = (1 - \delta)\epsilon_p \geq (1 - \delta)\epsilon_d^{\frac{3}{2}}\epsilon_p^{\frac{1}{2}},
\tag{3.15}
$$

where we have used the second part of (3.11) to deduce the last inequality. The second case is when $\|r(x_{k+1}, t_k)\| > \delta\epsilon_d$ and $\chi_{\|r(\cdot,t)\|}(x_{k+1}, t_k) > \epsilon_d$. Then, from (2.20) and the second part of (3.11), we have that

$$
t_k - t_{k+1} \geq \kappa_1 \epsilon_d^{3/2} \epsilon_p^{1/2}
$$

for some $\kappa_1 \in (0, 1)$. Combining this last bound with (3.15) then gives (3.12) with $\kappa_t = \min[\kappa_1, (1 - \delta)]$. □

It is then easy to combine the complexity analysis we already mentioned for Phase 1 of Algorithm 3.1 with the second part of (3.10) and AS9.

Lemma 3.3. Suppose that AS6-AS11 and (3.11) hold. Then Algorithm 3.1 generates an iterate $x_j \in \mathcal{F}$ such that either

$$
\chi_{\|c\|}(x_j) \leq \epsilon_d \text{ and } \|c(x_j)\| > \delta\epsilon_p
\tag{3.16}
$$

or

$$
\|r(x_j, t_{j-1})\| > \delta\epsilon_p, \quad \chi_{\|r(\cdot,t)\|}(x_j, t_{j-1}) \leq \epsilon_d \text{ and } \|c(x_j)\| \leq \epsilon_p
\tag{3.17}
$$

in at most

$$
\left\lfloor \kappa_{\text{NLO}} \, \epsilon_p^{-\frac{1}{2}} \epsilon_d^{-\frac{3}{2}} \right\rfloor
\tag{3.18}
$$

evaluations of f and c (and their derivatives), where $\kappa_{\text{NLO}} > 0$ is a problem dependent constant independent from ϵ_p, ϵ_d and x_0.

Proof. We have already discussed above (see Theorem 2.2) the fact that Phase 1 of Algorithm 3.1 will terminate in at most

$$
\left\lfloor \kappa_{\text{CNLS}} \max[(\delta\epsilon_p)^{-\frac{1}{2}}\epsilon_d^{-\frac{3}{2}}, (\delta\epsilon_p)^{-1}] \right\rfloor \leq \left\lfloor \kappa_{\text{CNLS}} \delta^{-1} \max[\epsilon_p^{-\frac{1}{2}}\epsilon_d^{-\frac{3}{2}}, \epsilon_p^{-1}] \right\rfloor \leq \left\lfloor \kappa_{\text{CNLS}} \delta^{-1} \epsilon_p^{-\frac{1}{2}} \epsilon_d^{-\frac{3}{2}} \right\rfloor
\tag{3.19}
$$

iterations (and problem functions evaluations), where we have taken our change of primal accuracy from ϵ_p to $\delta\epsilon_p$ specified in (3.3) and (3.11) into account. If the algorithm terminates at this stage, then (3.16) most hold, as required. Assume now that Phase 2 of Algorithm 3.1 is entered. We start by observing that AS9 implies that, for every k,

$$f_{\text{low}} \leq f(x_k) \leq t_k + \epsilon_p \leq t_1 - k\,\kappa_t\epsilon_d^{\frac{3}{2}}\epsilon_p^{\frac{1}{2}} + \epsilon_p \leq f(x_1) - k\,\kappa_t\epsilon_d^{\frac{3}{2}}\epsilon_p^{\frac{1}{2}} + \epsilon_p$$

where we have also used (3.12) and the definition of t_1 in Algorithm 3.1. Hence, we obtain from the inequality $f(x_1) \leq f_{\text{up}}$ (itself implied by AS9 again), the second part of (3.10) and $\epsilon_p \in (0,1)$ that at most

$$\frac{f_{\text{up}} - f_{\text{low}} + 1}{\kappa_t\epsilon_d^{\frac{3}{2}}\epsilon_p^{\frac{1}{2}}}$$

Phase 2 iterations may occur before $\chi_{\|r(\cdot,t)\|}(x_{k+1},t_k) \leq \epsilon_d$ with $\|r(x_{k+1},t_k)\| > \delta\epsilon_p$. Since the last part of Lemma 3.1 states that at most $\kappa_{\mathcal{S}}$ evaluations of f and c (and their derivatives) occur for each such iteration, we therefore deduce that at most

$$\kappa_{\text{phase2}} \stackrel{\text{def}}{=} \left\lceil \frac{\kappa_S(f_{\text{up}} - f_{\text{low}} + 1)}{\kappa_t\epsilon_d^{\frac{3}{2}}\epsilon_p^{\frac{1}{2}}} \right\rceil$$

Phase 2 iterations are needed to satisfy the termination test in Step 2(b) of Algorithm 3.1. Combining this result with (3.19) and the first part of (3.10), we obtain the desired conclusion with $\kappa_{\text{NLO}} = \max[\kappa_{\text{CNLS}}\delta^{-1}, \kappa_{\text{phase2}}]$. $\square$

To complete our analysis, we now comment on the meaning of the termination test (3.17); we already discussed (3.16) in Section 2.2. This meaning is best expressed by using

$$\ell(x,y) \stackrel{\text{def}}{=} f(x) + \langle y, c(x) \rangle, \tag{3.20}$$

the Lagrangian of the original problem where only equality constraints are kept, that is

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{such that} \quad c(x) = 0. \tag{3.21}$$

Lemma 3.4. Assume that, at iteration k in Phase 2 of Algorithm 3.1,

$$\|r(x_{k+1},t_k)\| > \delta\epsilon_p \quad \text{and} \quad \chi_{\|r(\cdot,t)\|}(x_{k+1},t_k) \leq \epsilon_d. \tag{3.22}$$

Then, either, for some vector $y_{k+1} \in \mathbb{R}^m$,

$$\chi_\ell(x_{k+1},y_{k+1}) \leq \epsilon_d\|(y_{k+1},1)\| \quad \text{and} \quad \|c(x_{k+1})\| \leq \epsilon_p \tag{3.23}$$

where $\ell(x,y)$ is the Lagrangian of the equality constrained problem given by (3.20) , or

$$\chi_{\|c\|}(x_{k+1}) \leq \epsilon_d \quad \text{and} \quad \|c(x_{k+1})\| \in (\delta\epsilon_p, \epsilon_p) \tag{3.24}$$

Proof.   Assume that $f(x_{k+1}) \neq t_k$. Then the second part of (3.22) can be rewritten as

$$\epsilon_d \geq \left| \min_{x+d \in \mathcal{F}, \|d\| \leq 1} \left\langle \frac{J(x_{k+1})^T c(x_{k+1}) + (f(x_{k+1}) - t_k) g_{k+1}}{\|r(x_{k+1}, t_k)\|}, d \right\rangle \right|$$

$$= \left| \min_{x+d \in \mathcal{F}, \|d\| \leq 1} \left\langle \frac{J(x_{k+1})^T y_{k+1} + g_{k+1}}{\|(y_{k+1}, 1)\|}, d \right\rangle \right|$$

where $y_{k+1} = c(x_{k+1})/(f(x_{k+1}) - t_k)$. Thus, given the definition of $\ell(x, y)$ in (3.20) and that of χ in (2.4), we obtain that the first part of (3.23) holds. The second part of this statement results from the first inequality in (3.10). Suppose now that $f(x_{k+1}) = t_k$. The second part of (3.24) is easily deduced from the observation that, in this case,

$$\|c(x_{k+1})\| = \|r(x_{k+1}, t_k)\| \in (\delta\epsilon_p, \|r(x_k, t_k)\|) = (\delta\epsilon_p, \epsilon_p), \qquad (3.25)$$

where we successively used (3.4), the first part of (3.22), the monotonic nature of the COCARC-S algorithm and (3.9). The first part of (3.24) then follows directly from (3.22) and the relation

$$\chi_{\|r(\cdot,t)\|}(x_{k+1}, t_k) = \left| \min_{x+d \in \mathcal{F}, \|d\| \leq 1} \left\langle \frac{J(x_{k+1})^T c(x_{k+1})}{\|r(x_{k+1}, t_k)\|}, d \right\rangle \right| = \chi_{\|c\|}(x_{k+1}),$$

where we used (3.25) to deduce the second equality.   □

Condition (3.23) expresses the approximate first-order criticality of (x_{k+1}, y_{k+1}) by assessing that the maximum feasible linearized decrease in the Lagrangian corresponding to the problem (3.21) only involving equalities is small compared to the size of the multiplier. The use of a scaled measure of criticality of this type was already argued in Cartis et al. (2013b).

Combining Lemmas 3.3 and 3.4 then gives our final complexity result.

Theorem 3.5. Suppose that AS3 (with $\phi = r(\cdot, t_k)$) and AS6-AS11 and (3.11) hold. Then Algorithm 3.1 generates an iterate $x_k \in \mathcal{F}$ such that either

$$\chi_\ell(x_k, y_k) \leq \epsilon_d \|(y_k, 1)\| \quad \text{and} \quad \|c(x_k)\| \leq \delta\epsilon_p \qquad (3.26)$$

where $\ell(x, y)$ is the Lagrangian of the equality constrained problem (3.20) and $y_k \in \mathbb{R}^m$ is an approximate Lagrange multiplier for problem (3.21), or

$$\chi_{\|c\|}(x_k) \leq \epsilon_p \quad \text{and} \quad \|c(x_k)\| > \delta\epsilon_p \qquad (3.27)$$

in at most

$$\left\lfloor \kappa_{\text{NLO}} \, \epsilon_p^{-\frac{1}{2}} \epsilon_d^{-\frac{3}{2}} \right\rfloor \qquad (3.28)$$

evaluations of f and c (and their derivatives), where $\kappa_{\text{NLO}} > 0$ is a problem dependent constant independent from ϵ_p, ϵ_d and x_0.

Again, we note that if $\epsilon_d = \epsilon_p^{2/3}$, Lemma 3.3 implies an overall complexity bound of $O(\epsilon_p^{-3/2})$ iterations and problem function evaluations for applying Algorithm 3.1 to the general nonlinear optimization problem (3.1).

4   Conclusions and perspectives

We have examined the worst-case complexity of finding approximate first-order critical points for the nonlinear least-squares problem with inequality constraints and the general nonlinear optimization problem (involving both equality and inequality constraints). We have shown that, under acceptable assumptions, both these problems can be approximately solved using a second-order method with cubic regularization in a number of problem functions (objective, constraints and derivatives) that is at most $O(\epsilon_d^{-3/2})$ (for the contrained nonlinear least-squares) and $O(\epsilon_d^{-3/2}\epsilon_p^{-1/2})$ (for the general problem), where ϵ_d and ϵ_p are the dual and primal accuracy thresholds, respectively. The latter bound reduces to $O(\epsilon_p^{-3/2})$ problem function evaluations if the dual threshold is chosen such that $\epsilon_d = \epsilon_p^{2/3}$. It is also known that this last bound is sharp for methods using cubic regularization (see Cartis, Gould and Toint, 2010) and optimal in a large class of second-order methods (see Cartis et al., 2011b).

This result remains surprising because it shows that the inclusion of nonlinear (in)equality constraints in the problem does not affect its worst-case evaluation analysis. Indeed, the worst-case complexity of the general nonlinear optimization problem is identical (in order) to that of the unconstrained case.

References

C. Cartis, N. I. M. Gould, and Ph. L. Toint. On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization. *SIAM Journal on Optimization*, **20**(6), 2833–2852, 2010.

C. Cartis, N. I. M. Gould, and Ph. L. Toint. Adaptive cubic overestimation methods for unconstrained optimization. Part II: worst-case function-evaluation complexity. *Mathematical Programming, Series A*, **130**(2), 295–319, 2011a.

C. Cartis, N. I. M. Gould, and Ph. L. Toint. Optimal Newton-type methods for nonconvex optimization. Technical Report naXys-17-2011, Namur Centre for Complex Systems (naXys), FUNDP-University of Namur, Namur, Belgium, 2011b.

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, 2012a.

C. Cartis, N. I. M. Gould, and Ph. L. Toint. Complexity bounds for second-order optimality in unconstrained optimization. *Journal of Complexity*, **28**, 93–108, 2012b.

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*, (to appear), 2013*a*. DOI: 10.1007/s10107-012-0617-9.

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*, (submitted), 2013*b*.

A. R. Conn, N. I. M. Gould, A. Sartenaer, and Ph. L. Toint. Global convergence of a class of trust region algorithms for optimization using inexact projections on convex constraints. *SIAM Journal on Optimization*, **3**(1), 164–221, 1993.

S. Gratton, A. Sartenaer, and Ph. L. Toint. Recursive trust-region methods for multiscale nonlinear optimization. *SIAM Journal on Optimization*, **19**(1), 414–444, 2008.

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

Yu. Nesterov and B. T. Polyak. Cubic regularization of Newton method and its global performance. *Mathematical Programming, Series A*, **108**(1), 177–205, 2006.

S. A. Vavasis. Black-box complexity of local minimization. *SIAM Journal on Optimization*, **3**(1), 60–80, 1993.

Y. Yuan. Conditions for convergence of trust region algorithms for nonsmooth optimization. *Mathematical Programming*, **31**(2), 220–228, 1985.