Corrigendum: On the complexity of finding first-order critical points in constrained nonlinear optimization

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

11th November, 2014; Revised 16th July, 2015

Abstract

In a recent paper (Cartis, Gould and Toint, Math. Prog. A 144(1-2) 93–106, 2014), the evaluation complexity of an algorithm to find an approximate first-order critical point for the general smooth constrained optimization problem was examined. Unfortunately, the proof of Lemma 3.5 in that paper uses a result from an earlier paper in an incorrect way, and indeed the result of the lemma is false. The purpose of this corrigendum is to provide a modification of the previous analysis that allows us to restore the complexity bound for a different, scaled measure of first-order criticality.

Keywords: evaluation complexity, worst-case analysis, constrained nonlinear optimization.

1 Introduction

In a recent paper [4], we aimed to show that the complexity of finding ϵ-approximate first-order critical points for the general smooth constrained optimization problem requires no more than $O(\epsilon^{-2})$ function and constraint evaluations. The analysis involved examining the worst-case behaviour of a short-step homotopy algorithm in which a sequence of approximately feasible points are tracked downhill. The entire framework relies on the $O(\epsilon^{-2})$ iteration complexity bound of a general first-order method for non-smooth composite minimization [2]. Unfortunately, the given proof of [4, Lem. 3.5] invokes [2, Thm. 3.1] incorrectly, and indeed the result of the lemma is false. Furthermore, the claimed generalization to inequality constraints [4, §4] fails to account for complementary slackness, and is thus incomplete.

Our aim here is to correct our previous analysis. To do so, we need first to re-examine what we believe it means to be approximately first-order critical, and this leads to an alternative stopping rule for our homotopy method. Armed with that, we then use a different merit function for the second phase of our homotopy method compared to that we considered in [4] to establish a variant of [4, Lem. 3.5], and this reveals a worst-case evaluation complexity bound of $O(\epsilon^{-2})$ for the revised ϵ-criticality measure.

2 Corrigendum

2.1 Stopping criteria for constrained optimization

In [4], we consider the general nonlinearly constrained optimization problem

$$\text{minimize } f(x) \text{ such that } c_E(x) = 0, \text{ and } c_I(x) \geq 0, \tag{2.1}$$

*Mathematical Institute, University of Oxford, Woodstock Road, Oxford, OX2 6GG, England, UK. Email: coralia.cartis@maths.ox.ac.uk

†Computational Science and Engineering Department, Rutherford Appleton Laboratory, Chilton, Oxfordshire, OX11 0QX, England, UK. Email: nick.gould@sftc.ac.uk

‡Namur Center for Complex Systems (naXys) and Department of Mathematics, FUNDP-University of Namur, 61, rue de Bruxelles, B-5000 Namur, Belgium. Email: philippe.toint@fundp.ac.be

where c_E and c_I are continuously differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $\mathbb{R}^p$, respectively, having Lipschitz continuous Jacobians, and $\mathcal{E}$ and $\mathcal{I}$ are non-intersecting index sets of equality and inequality constraints, respectively. Ideally, we would like to find a point x_*, and corresponding Lagrange multiplier estimates y_*, that satisfy the first-order criticality—or Karush–Kuhn–Tucker (KKT)—conditions [8, 9]

$$\begin{aligned} g(x_*) + J^T(x_*)y_* &= 0, & (2.2a)\\ c_i(x_*) &= 0 \ \text{ for all } \ i \in \mathcal{E}, & (2.2b)\\ c_i(x_*) \geq 0 \ \text{ and } \ [y_*]_i &\leq 0 \ \text{ for all } \ i \in \mathcal{I}, & (2.2c)\\ \text{and } \ c_i(x_*)[y_*]_i &= 0 \ \text{ for all } \ i \in \mathcal{I}, & (2.2d) \end{aligned}$$

where $g(x) := \nabla f(x)$, $J(x) := \nabla c(x)$ and $c(x) := (c_E^T(x), c_I^T(x))^T$. Of course, there might be no feasible point for the problem, or in the absence of a suitable constraint qualification, it might be that we may have to be satisfied with the John condition [7]

$$\nu_* g(x_*) + J^T(x_*)y_* = 0, \tag{2.3}$$

instead of (2.2a), for which there is an extra, possibly zero, multiplier ν_* associated with the objective function and at least one multiplier is nonzero. The last of the KKT conditions, (2.2d), is known as the complementarity condition and in conjunction with (2.2b) is often written as

$$\langle c(x_*), y_* \rangle = 0, \tag{2.4}$$

while the first, (2.2a), requires that the gradient of the Lagrangian

$$\ell(x, y) = f(x) + \langle c(x), y \rangle,$$

taken with respect to the variables x, vanish at a KKT point; here and elsewhere $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. Since it is very unlikely that we can find (x_*, y_*) exactly, our goal is to find suitable approximations that satisfy a perturbation of these criticality conditions.

While proper scaling of the objective and constraint functions is to a large extent the responsibility of the problem formulator—and ideally they should be scaled so that unit changes in x in regions of interest result in similar changes in f and c—the values of the optimal Lagrange multipliers y_* are essentially controlled by (2.2a), and should be taken into account when deriving stopping criteria. Consider perturbations $x = x_* + \delta x$ and $y = y_* + \delta y$ to some KKT point x_* and to a corresponding multiplier y_*. Then supposing for argument's sake that f and $c \in C^2$, a Taylor expansion and the KKT condition $g(x_*) + J^T(x_*)y_* = 0$ give that the perturbed dual feasibility residual

$$g(x) + J^T(x)y \cong \left[H(x_*) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} [y_*]_i H_i(x_*) \right] \delta x + J^T(x_*)\delta y$$

to first order, where $H(x) \stackrel{\text{def}}{=} \nabla_{xx} f(x)$ and $H_i(x) \stackrel{\text{def}}{=} \nabla_{xx} c_i(x)$. The presence of the multiplier y_* here illustrates that the size of the multiplier should not be ignored when measuring KKT equation residuals. Similarly, the complementary slackness condition (2.4) is

$$\langle y, c(x) \rangle \cong \langle \delta x, J^T(x_*)y_* \rangle + \langle \delta y, c(x_*) \rangle$$

to first order, and the value of y_* is once again relevant.

Thus when trying to solve (2.1), we pick primal and dual feasibility and complementarity tolerances $\epsilon_p, \epsilon_d, \epsilon_c > 0$, and aim to find x_ϵ along with Lagrange multiplier estimates y_ϵ such that

$$\left\| \begin{pmatrix} c_E(x_\epsilon) \\ \min[0, c_I(x_\epsilon)] \end{pmatrix} \right\| \leq \epsilon_p, \ \frac{\| g(x_\epsilon) + J^T(x_\epsilon)y_\epsilon \|}{\|(y_\epsilon, 1)\|_{\text{D}}} \leq \epsilon_d, \ \frac{\langle c(x_\epsilon), y_\epsilon \rangle}{\|(y_\epsilon, 1)\|_{\text{D}}} \leq \epsilon_c \ \text{ and } \ [y_\epsilon]_{\mathcal{I}} \leq 0 \tag{2.5}$$

as a reasonable goal when trying to satisfy (2.2); here $\|\cdot\|_{\rm D}$ is the dual norm to the chosen norm $\|\cdot\|$ induced by the given inner product $\langle\cdot,\cdot\rangle$. We have previously used this scaled dual-feasibility rule for equality-constrained problems [3], while the requirement on approximate complementarity is an obvious generalization.

Notice that the stopping rules (2.5) are consistent with the John conditions (2.2b)–(2.2d) and (2.3) in which $1/\|(y_\epsilon,1)\|_{\rm D}$ and $y_\epsilon/\|(y_\epsilon,1)\|_{\rm D}$ approximate ν_* and y_* respectively, and that if additionally y_ϵ remains bounded, they give an approximate KKT point (2.2a)–(2.2d) in the sense that

$$\left\|\begin{pmatrix} c_E(x_\epsilon) \\ \min[\,0, c_I(x_\epsilon)] \end{pmatrix}\right\| \leq \epsilon_p,\ \left\|g(x_\epsilon)+J^T(x_\epsilon)y_\epsilon\right\| \leq \epsilon_d\|(y_\epsilon,1)\|_{\rm D},\ \langle c(x_\epsilon),y_\epsilon\rangle \leq \epsilon_c\|(y_\epsilon,1)\|_{\rm D} \ \text{ and } \ [y_\epsilon]_{\mathcal{I}} \leq 0,$$

where now y_ϵ approximates y_*. Thus no constraint qualification will be presumed or required in the definition of our algorithm or in its analysis.

Having defined our problem, and what we will be looking for with our algorithm, we turn now to the description and analysis of the algorithm itself.

2.2 Composite-nonsmooth optimization

The analysis of [4, Alg. 2.1], which was intended for problems that only involve equality constraints, and the extension for mixed equality-inequality problems that we shall shortly describe, depends on basic properties of critical points of the *composite, nonsmooth* function

$$\Phi(x) := h(r(x)), \tag{2.6}$$

in which $r : \mathbb{R}^n \to \mathbb{R}^m$ is smooth and $h : \mathbb{R}^m \to \mathbb{R}$ is convex and continuous but may be nonsmooth. We say that x_* is a *first-order critical point* of Φ if

$$J_r^T(x_*)y = 0 \ \text{ for some } \ y \in \partial h(r(x_*)) \tag{2.7}$$

holds, where ∂h denotes the subdifferential of h and $J_r(x) := \nabla r(x)$. It is well known [10] that x_* is a first-order critical point of Φ if and only if

$$\chi_\Phi(x_*) = 0, \tag{2.8}$$

where the predicted reduction of a linear model of Φ in a unit ball,

$$\chi_\Phi(x) := l_\Phi(x,0) - \min_{\|d\|\leq 1} l_\Phi(x,d) \tag{2.9}$$

and

$$l_\Phi(x,d) := h\left(r(x)+J_r(x)d\right), \quad d \in \mathbb{R}^n, \tag{2.10}$$

and that $\chi_\Phi(x)$ is a continuous criticality measure for Φ [10]. Our updated analysis hinges on what can be deduced when $\chi_\Phi(x)$ is small. Theorem 2.1 is a generalization of [2, Thm. 3.1].

Theorem 2.1. Suppose that $r \in C^1$, and that $h \in C^0$ is convex. Given $\epsilon > 0$, suppose that

$$\chi_\Phi(x_\epsilon) \leq \epsilon, \tag{2.11}$$

for some x_ϵ. Then

$$\|J_r^T(x_\epsilon)y_\epsilon\| \leq \epsilon, \tag{2.12}$$

where $y_\epsilon \in \partial h(r(x_\epsilon)+J_r(x_\epsilon)d_\epsilon)$ and

$$d_\epsilon = \arg\min_{\|d\|\leq 1} l_\Phi(x_\epsilon,d). \tag{2.13}$$

Proof. Let d_ϵ satisfy (2.13). Suppose that $\|d_\epsilon\| < 1$. Then since (2.13) is unconstrained and $l_\Phi(x_\epsilon, d)$ is convex, applying [6, (14.2.16)] to $l_\Phi(x_\epsilon, d)$ shows that there is a $y_\epsilon \in \partial h(r(x_\epsilon) + J_r(x_\epsilon)d_\epsilon)$ for which $J_r^T(x_\epsilon)y_\epsilon = 0$, and thus (2.12) holds trivially. So it remains to consider $\|d_\epsilon\| = 1$. In this case, first-order conditions for (2.13) imply that there exists $y_\epsilon \in \partial h(r(x_\epsilon) + J_r(x_\epsilon)d_\epsilon)$ and $\lambda_* \geq 0$ such that

$$J_r^T(x_\epsilon)y_\epsilon + \lambda_* z_\epsilon = 0, \tag{2.14}$$

where $z_\epsilon \in \partial\|d_\epsilon\| = \{z \mid \|z\|_{\rm D} = 1 \text{ and } \langle z, d_\epsilon\rangle = \|d_\epsilon\|\}$. It follows from the definition (2.9) of $\chi_\Phi(x)$, (2.14), the definition of $\partial\|d_\epsilon\|$ and $\|d_\epsilon\| = 1$ that

$$\begin{aligned} \chi_\Phi(x_\epsilon) &= [h(r(x_\epsilon) - h(J_r(x_\epsilon)d_\epsilon + r(x_\epsilon))] \\ &= \left[h(r(x_\epsilon)) - h(r(x_\epsilon) + J_r(x_\epsilon)d_\epsilon) + \langle d_\epsilon, J_r^T(x_\epsilon)y_\epsilon\rangle\right] + \lambda_*\langle d_\epsilon, z_\epsilon\rangle \\ &= \left[h(r(x_\epsilon)) - h(r(x_\epsilon) + J_r(x_\epsilon)d_\epsilon) + \langle d_\epsilon, J_r^T(x_\epsilon)y_\epsilon\rangle\right] + \lambda_*. \end{aligned} \tag{2.15}$$

Since $l_\Phi(x_\epsilon, d)$ is convex, the subgradient inequality implies that $l_\Phi(x_\epsilon, 0) - l_\Phi(x_\epsilon, d_\epsilon) \geq \langle y, -J_r(x_\epsilon)d_\epsilon\rangle = -\langle d_\epsilon, J_r^T(x_\epsilon)y\rangle$, for any $y \in \partial h(r(x_\epsilon) + J_r(x_\epsilon)d_\epsilon)$. Letting $y = y_\epsilon$, we deduce

$$h(r(x_\epsilon)) - h(r(x_\epsilon) + J_r(x_\epsilon)d_\epsilon) + \langle d_\epsilon, J_r^T(x_\epsilon)y_\epsilon\rangle \geq 0,$$

and so, from (2.11) and (2.15), it follows that

$$\epsilon \geq \chi_\Phi(x_\epsilon) \geq \lambda_*. \tag{2.16}$$

From (2.14) and the definition of $\partial\|d_\epsilon\|$, we deduce

$$\lambda_* = \lambda_*\|z_\epsilon\| = \|J_r^T(x_\epsilon)y_\epsilon\|. \tag{2.17}$$

and this together with (2.16) yields (2.12). □

2.3 A short-step steepest descent algorithm for constrained optimization

Both our original Algorithm 2.1 [4] and the extension to allow inequality constraints that we shall analyse here work in two phases. The first aims to reduce the infeasibility

$$\|c^-(x)\|, \quad \text{where } \; c^-(x) = \begin{pmatrix} c_E(x) \\ \min(c_I(x), 0) \end{pmatrix}, \tag{2.18}$$

to an acceptable level using [2, Alg. 2.1], and terminates when the criticality measure,

$$\psi(x) := l_c(x, 0) - \min_{\|d\|\leq 1} l_c(x, d) \;\text{ where }\; l_c(x, d) := \|[c(x) + J(x)d]^-\|, \tag{2.19}$$

for the infeasibility at the terminating point x_1 is smaller than ϵ_d. If the infeasibility is itself smaller than a fraction $\delta \in (0, 1)$ of ϵ_p, a second phase is performed in which the penalty function

$$\phi(x, t) = \max(f(x) - t, 0) + \|c^-(x)\| \tag{2.20}$$

is reduced for a sequence of decreasing parameters $t = t_j$, $j \geq 1$. This second phase terminates when the criticality measure for the penalty function (the predicted reduction of a linear model of ϕ in a unit ball),

$$\chi(x, t) := l_\phi(x, 0; t) - \min_{\|d\|\leq 1} l_\phi(x, d; t), \;\text{ where }\; l_\phi(x, d; t) := l_c(x, d) + \max(f(x) + \langle g(x), d\rangle - t, 0), \tag{2.21}$$

at x_k is smaller that ϵ_d.

We now formally present our idealised short-step algorithm for (2.1). This is simply a restatement of [4, Alg.2.1], with the obvious extensions to cope with inequality constraints, the modified merit functions (2.18) and (2.20) and the replacement criticality measures (2.19) and (2.21) that lie at the heart of the algorithm.

Algorithm 2.1: The short-step steepest-descent algorithm.

Let $\delta \in (0,1)$, $\epsilon_p, \epsilon_d \in (0,1]$ and $\Delta_1 > 0$ be given, together with a starting point x_0.

Phase 1:

Starting from x_0, minimize $\|c^-(x)\|$ using the trust-region method of [2] until a point x_1 is found such that

$$\psi(x_1) \le \epsilon_d.$$

If $\|c^-(x_1)\| > \delta\epsilon_p$, terminate [locally infeasible].

Phase 2:

1. Set $t_1 = \|c^-(x_1)\| + f(x_1) - \epsilon_p$ and $k = 1$.
2. While $\chi(x_k, t_k) \ge \epsilon_d$,

   2a. Compute a first-order step s_k by solving

   $$\underset{s\in\mathbb{R}^n}{\text{minimize}} \;\; l_\phi(x_k, s; t_k) \;\; \text{such that} \;\; \|s\| \le \Delta_k. \tag{2.22}$$

   2b. Compute $\phi(x_k + s_k; t_k)$ and define

   $$\rho_k = \frac{\phi(x_k;t_k) - \phi(x_k+s_k;t_k)}{l_\phi(x_k,0;t_k) - l_\phi(x_k,s_k;t_k)}. \tag{2.23}$$

   If $\rho_k \ge \eta$, then $x_{k+1} = x_k + s_k$; else $x_{k+1} = x_k$. Set

   $$\Delta_{k+1} = \begin{cases} \Delta_k & \text{if } \rho_k \ge \eta \quad [k \text{ successful}] \\ \gamma\Delta_k & \text{if } \rho_k < \eta \quad [k \text{ unsuccessful}] \end{cases} \tag{2.24}$$

   2c. If $\rho_k \ge \eta$, set

   $$t_{k+1} = \begin{cases} t_k - \phi(x_k;t_k) + \phi(x_{k+1};t_k) & \text{if } f(x_{k+1}) \ge t_k, \\ 2f(x_{k+1}) - t_k - \phi(x_k;t_k) + \phi(x_{k+1};t_k) & \text{if } f(x_{k+1}) < t_k. \end{cases} \tag{2.25}$$

   Otherwise, set $t_{k+1} = t_k$.

   2d. Increment k by one and return to Step 2.
3. Terminate [(approximately) first-order critical]

The introductory results [4, Lem. 2.1–2.2 & 3.1–3.4] were established for the equality-constrained problem,

$$\text{minimize} \;\; f(x) \;\; \text{such that} \;\; c_E(x) = 0,$$

measuring constraint violation by $\|c_E(x)\|$, and used the penalty function $|f(x)-t| + \|c_E(x)\|$ rather than[1] (2.20), but generalize without difficulty for the inequality problem (2.1) and the infeasibility measures (2.18) and (2.20) needed here. For completeness, we now establish modified versions of [4, Lem. 2.1–2.2 & 3.1–3.4] for the new merit functions (2.18) and (2.20); the only significant difference is that [4, Lem. 2.2 eq.(2.17)] becomes $f(x_k) - t_k \le \epsilon_p$, which combines with [4, Lem. 2.2 eq.(2.15)] to give

$$0 < f(x_k) - t_k \le \epsilon_p; \tag{2.26}$$

see (2.29)–(2.30) in the upcoming Lemma 2.3.

[1]We may derive similar complexity results for the *equality* problem with the original penalty function $|f(x) - t| + \|c_E(x)\|$.

To do so, we make the following assumptions. Consider the slightly extended neighbourhood $\mathcal{C}_{\Delta_1} := \mathcal{C}_1 + \mathcal{B}(0, \beta\Delta_1)$ of the feasible region (if there is one), where $\mathcal{C}_1 := \{x : \|c^-(x)\| \leq \kappa_{\mathcal{C}_1}\}$, $\kappa_{\mathcal{C}_1} > \epsilon_p$ $\mathcal{B}(0, \beta\Delta_1)$ is a unit ball of radius $\beta\Delta_1$ for some β slightly larger than 1, and Δ_1 is the initial trust-region radius in Algorithm 2.1.

A1. The function c is continuously differentiable on $\mathbb{R}^n$ and f is continuously differentiable in $\mathcal{C}_{\Delta_1}$.

A2. The Jacobian $J(x)$ is globally Lipschitz continuous (with Lipschitz constant L_J) on $\mathbb{R}^n$, and the gradient $g(x)$ is Lipschitz continuous in $\mathcal{C}_{\Delta_1}$ (with constant $L_g \geq 1$).

A3. $f_{\text{low}} \leq f(x) \leq f_{\text{lup}}$ for all x in $\mathcal{C}_1$, where without loss of generality $f_{\text{lup}} \geq f_{\text{low}} + 1$.

Assumptions **A1** and **A2** ensure that suitable Taylor approximations hold at points required by our analysis to establish our main result, and are simply extensions of those in [4] to allow for inequality constraints. The assumptions on f are only needed if Phase 2 of the algorithm is required.

To show that Phase 2 of Algorithm 2.1, most especially (2.23), is well-defined, we use the following result.

Lemma 2.2. (cf. [4, Lem. 2.1]) Suppose that **A1** holds. If $x_k \in \mathcal{C}_1$, then the model decrease satisfies

$$l_\phi(x_k, 0; t_k) - l_\phi(x_k, s_k; t_k) \geq \min(\Delta_k, 1)\,\chi(x_k, t_k). \tag{2.27}$$

Proof. Apply [2, Lem.2.1] with $h \stackrel{\text{def}}{=} \|\cdot\| + \max(\cdot, 0)$ and $\Phi_h(x) \stackrel{\text{def}}{=} \chi(x, t_k)$ considered as a function of x only. □

Our next result shows that x_k not only belongs to $\mathcal{C}_1$ so that Phase 2 is well-defined, but it remains approximately feasible for all Phase 2 iterations, and, additionally, successive objective function values stay close to their targets.

Lemma 2.3. (cf. [4, Lem. 2.2]) Suppose that **A1** holds. On each Phase 2 iteration $k \geq 1$ of Algorithm 2.1, we have

$$\phi(x_k; t_k) = \epsilon_p, \tag{2.28}$$

$$f(x_k) > t_k, \tag{2.29}$$

$$f(x_k) - t_k \leq \epsilon_p, \tag{2.30}$$

$$\|c^-(x_k)\| \leq \epsilon_p, \tag{2.31}$$

and $x_k \in \mathcal{C}_1$, for ϕ defined in (2.20)

Proof. Firstly, note that (2.20) and (2.28) imply (2.30) and (2.31); the latter implies $x_k \in \mathcal{C}_1$ since $\epsilon_p < \kappa_{\mathcal{C}_1}$. Thus it remains to prove (2.28) and (2.29). The proof of these relations is by induction on k. For $k = 1$, recall that we only enter Phase 2 of the algorithm if $\|c^-(x_1)\| \leq \delta\epsilon_p < \epsilon_p$, which gives (2.29) and (2.28) for $k = 1$, due to the particular choice of t_1. Also, (2.27) holds at $k = 1$ and ρ_1 in (2.23) is well-defined.

Now let $k > 1$ and assume that (2.28) and (2.29) are satisfied, and so

$$\phi(x_k; t_k) = \epsilon_p. \tag{2.32}$$

If k is an unsuccessful iteration, $x_{k+1} = x_k$ and $t_{k+1} = t_k$ and so (2.29) and (2.28) continue to hold at x_{k+1}. It remains to consider the case when k is successful. Recall that (2.32) implies $\|c^-(x_k)\| \leq \epsilon_p$

and $x_k \in \mathcal{C}_1$ since $\epsilon_p < \kappa_{\mathcal{C}_1}$, and so (2.27) holds. Thus, since we have not terminated, Lemma 2.2 shows that (2.23) has a positive denominator, which together with k being successful so that $\rho_k \geq \eta$, implies

$$\phi(x_k; t_k) > \phi(x_{k+1}; t_k).$$

This and (2.25) immediately give that $f(x_{k+1}) - t_{k+1} > 0$ so that (2.29) holds at $k+1$. Using the latter and (2.20), we deduce

$$\phi(x_{k+1}; t_{k+1}) = \|c^-(x_{k+1})\| + f(x_{k+1}) - t_k + (t_k - t_{k+1}). \tag{2.33}$$

Consider first the case when $f(x_{k+1}) \geq t_k$. Then, using (2.33) and (2.25), we obtain that

$$\phi(x_{k+1}; t_{k+1}) = \phi(x_{k+1}; t_k) + \phi(x_k; t_k) - \phi(x_{k+1}; t_k) = \phi(x_k; t_k).$$

If $f(x_{k+1}) < t_k$, we have that

$$\begin{aligned} \phi(x_{k+1}; t_{k+1}) &= \|c^-(x_{k+1})\| - f(x_{k+1}) + t_k + \phi(x_k; t_k) - \phi(x_{k+1}; t_k) \\ &= \phi(x_{k+1}; t_k) + \phi(x_k; t_k) - \phi(x_{k+1}; t_k) \\ &= \phi(x_k; t_k), \end{aligned}$$

where we again use (2.33) and (2.25). Combining the two cases and using (2.32), we then deduce that

$$\phi(x_{k+1}; t_{k+1}) = \phi(x_k; t_k) = \epsilon_p,$$

and thus (2.28) holds at $k+1$. This concludes the inductive step. □

Our evaluation-complexity analysis requires that we bound the number of Phase 1 evaluations.

Lemma 2.4. (cf. [4, Lem. 3.1]) Suppose that **A1-A2** hold. Then at most

$$\left\lfloor \kappa^{\rm a}_{\sf TRNS1} \frac{\|c^-(x_0)\|}{\epsilon_d^2} + \kappa^{\rm b}_{\sf TRNS1} |\log \epsilon_d| + \kappa^{\rm c}_{\sf TRNS1} \right\rfloor \tag{2.34}$$

evaluations of $c(x)$ and its derivatives are needed to complete Phase 1 of Algorithm 2.1, for some $\kappa^{\rm a}_{\sf TRNS1}$, $\kappa^{\rm b}_{\sf TRNS1}$ and $\kappa^{\rm c}_{\sf TRNS1} > 0$ independent of ϵ_d and x_0.

Proof. This is a direct application of [2, Th,.2.4] with $h \stackrel{\rm def}{=} \|\cdot\|$, $\Phi_h(x) \stackrel{\rm def}{=} \|c^-(x)\|$, $L_h = 1$, $\eta_1 = \eta_2 \stackrel{\rm def}{=} \eta$ and $\gamma_1 = \gamma_2 \stackrel{\rm def}{=} \gamma$. □

We next use Lemma 2.2 to provide a lower bound on the trust-region radius computed during Phase 2.

Lemma 2.5. (cf. [4, Lem. 3.2]) Suppose that **A1-A2** hold. Then any Phase 2 iteration $k \geq 1$ of Algorithm 2.1 satisfying $\chi(x_k, t_k) \geq \epsilon_d$ and

$$\Delta_k \leq \frac{(1-\eta)\epsilon_d}{L_g + \frac{1}{2} L_J} \tag{2.35}$$

is successful in the sense of (2.24). Furthermore, while $\chi(x_k, t_k) \geq \epsilon_d$, we have

$$\Delta_k \geq \kappa_\Delta \epsilon_d, \quad \text{for all Phase 2 iterations } k \geq 1, \tag{2.36}$$

where

$$\kappa_\Delta \stackrel{\rm def}{=} \min\left(\Delta_1, \frac{(1-\eta)\gamma}{L_g + \frac{1}{2} L_J}\right). \tag{2.37}$$

Proof. From (2.23) and (2.20), and using the fact that $|\max(a,0) - \max(b,0)| \leq |a-b|$ for all $a, b \in \mathbb{R}$, we have

$$
\begin{aligned}
|\rho_k - 1| &= \frac{|\phi(x_k+s_k;t_k) - l_\phi(x_k;t_k,s_k)|}{l_\phi(x_k,0;t_k) - l_\phi(x_k,s_k;t_k)} \\
&= \frac{|\|c^-(x_k+s_k)\| - \|[c(x_k)+J(x_k)s_k]^-\| + \max(f(x_k+s_k)-t_k,0) - \max(f(x_k)+\langle g(x_k),s_k\rangle - t_k,0)|}{l_\phi(x_k,0;t_k) - l_\phi(x_k,s_k;t_k)} \\
&\leq \frac{|\|c^-(x_k+s_k)\| - \|[c(x_k)+J(x_k)s_k]^-\|| + |f(x_k+s_k) - f(x_k) - \langle g(x_k),s_k\rangle|}{l_\phi(x_k,0;t_k) - l_\phi(x_k,s_k;t_k)}.
\end{aligned} \tag{2.38}
$$

Standard Taylor expansions give that

$$
f(x_k+s_k) = f(x_k) + g(\xi_k)^T s_k \quad \text{for some} \quad \xi_k \in [x_k, x_k+s_k],
$$

and

$$
c(x_k+s_k) = c(x_k) + \int_0^1 J(x_k+ts_k)s_k\,dt.
$$

Observe that $x_k \in \mathcal{C}_1$ because of Lemma 2.3, and $\|\xi_k - x_k\| \leq \|s_k\| \leq \Delta_k \leq \Delta_1$ (as the radius is never increased in Phase 2) then implies that ξ_k, $x_k+s_k \in \mathcal{C}_\Delta$. Thus **A1**–**A2** apply at these points, and together with the Taylor expansions, gives that

$$
|f(x_k+s_k) - f(x_k) - \langle g(x_k),s_k\rangle| \leq L_g\|s_k\|^2
$$

and

$$
\big|\|c^-(x_k+s_k)\| - \|[c(x_k)+J(x_k)s_k]^-\|\big| \leq \tfrac{1}{2}L_J\|s_k\|^2.
$$

Thus, from (2.27), (2.38) and $\|s_k\| \leq \Delta_k$, we deduce

$$
|\rho_k - 1| \leq \frac{(L_g + \frac{1}{2}L_J)\,\Delta_k^2}{\min(\Delta_k,1)\,\chi(x_k,t_k)} \leq \frac{(L_g+\frac{1}{2}L_J)}{\epsilon_d}\Delta_k,
$$

where to obtain the second inequality, we use $\chi(x_k,t_k) \geq \epsilon_d$ and $\Delta_k \leq 1$, where the latter follows from (2.35), $L_g \geq 1$ and $\epsilon_d \in (0,1]$. Finally, (2.35) implies $|\rho_k - 1| \leq 1-\eta$, which gives that k is successful due to (2.24).

Now whenever (2.35) holds, (2.24) sets $\Delta_{k+1} = \Delta_k$. This implies that when $\Delta_1 \geq \gamma(1-\eta)\epsilon_d/(L_g + \frac{1}{2}L_J)$, we have $\Delta_k \geq \gamma(1-\eta)\epsilon_d/(L_g+\frac{1}{2}L_J)$ for all k, where the factor γ is introduced for the case when Δ_k is greater than $(1-\eta)\epsilon_d/(L_g+\frac{1}{2}L_J)$ and iteration k is unsuccessful. Applying again the implication resulting from (2.35) and (2.24) for $k=1$, we deduce (2.36) when $\Delta_1 < \gamma(1-\eta)\epsilon_d/(L_g+\frac{1}{2}L_J)$ since $\gamma \in (0,1)$ and $\epsilon \in (0,1]$. □

We now bound the total number of unsuccessful iterations in the course of Phase 2.

Lemma 2.6. (cf. [4, Lem. 3.3]) There are at most

$$
\left\lfloor \frac{1}{|\log\gamma|}\left|\log\epsilon_d + \log\left(\frac{(1-\eta)}{\Delta_1(L_g+\frac{1}{2}L_J)}\right)\right|\right\rfloor \tag{2.39}
$$

unsuccessful iterations in Phase 2 of Algorithm 2.1.

Proof. Note that (2.24) implies that the trust-region radius is never increased, and therefore Lemma 2.5 guarantees that all iterations must be successful once Δ_1 has been reduced (by a factor γ) enough times to ensure (2.35). Hence there are at most (2.39) unsuccessful iterations during the complete execution of the Phase 2. □

The final auxiliary lemma establishes that the targets t_k decrease by a quantity bounded below by a multiple of ϵ_d^2 at every successful iteration.

Lemma 2.7. (cf. [4, Lem. 3.4]) Suppose that **A1-A2** hold. Then on each successful Phase 2 iteration $k \geq 1$ of Algorithm 2.1, we have

$$\phi(x_k + s_k; t_k) \leq \phi(x_k; t_k) - \kappa_{\rm C}\epsilon_d^2 \tag{2.40}$$

and

$$t_k - t_{k+1} \geq \kappa_{\rm C}\epsilon_d^2 \tag{2.41}$$

where

$$\kappa_{\rm C} \stackrel{\rm def}{=} \eta\kappa_\Delta \tag{2.42}$$

and κ_Δ is defined in (2.37), independently of ϵ_d.

Proof. From (2.23) and k being successful, we deduce

$$\phi(x_k; t_k) - \phi(x_k + s_k; t_k) \geq \eta\,[l_\phi(x_k, 0; t_k) - l_\phi(x_k, s_k; t_k)] \geq \eta \min(\Delta_k, 1)\,\epsilon_d,$$

where to obtain the second inequality, we use (2.27) and $\chi(x_k, t_k) \geq \epsilon_d$. Further, we employ the bound (2.36) and obtain

$$\phi(x_k; t_k) - \phi(x_k + s_k; t_k) \geq \eta \min(\kappa_\Delta\epsilon_d, 1)\,\epsilon_d = \eta\kappa_\Delta\epsilon_d^2,$$

where we also use $\epsilon_d \in (0, 1]$ and $\kappa_\Delta \leq 1$ due to $L_g \geq 1$, η, $\gamma \in (0, 1)$; this gives (2.40). Finally, (2.41) results from (2.25) and (2.40). □

2.4 Corrected results

Our flawed version of [4, Lem. 3.5] aimed to connect approximate critical points of the merit functions of Phases 1 and 2 of [4, Alg. 2.1] to those in (2.5) for our original problem (2.1). Here is our correction.

Lemma 2.8. [Correction to [4, Lem. 3.5]] Given $\epsilon_p, \epsilon_d, \epsilon_c > 0$ for which $\epsilon_d < \epsilon_p$ and $\epsilon_p + \epsilon_d \leq \epsilon_c$, suppose that $\|c^-(x_k)\| \leq \epsilon_p$ and $\chi(x_k, t_k) \leq \epsilon_d$. Then either x_k is an approximate critical point of (2.1) in the sense that $x_\epsilon = x_k$ and $y_\epsilon = y_k$ satisfy (2.5) for some vector of Lagrange multiplier estimates $y_k \in \mathbb{R}^m$, or x_k is an almost-feasible approximate critical point of $\|c^-(x)\|$ in the sense that x_k and z_k satisfy

$$\|J^T(x_k)z_k\| \leq \epsilon_d,\ [z_k]_{\mathcal{I}} \leq 0 \ \text{ and } \ \|z_k\|_{\rm D} = 1 \tag{2.43}$$

as well as $\|c^-(x_k)\| \leq \epsilon_p$ for another vector of Lagrange multiplier estimates $z_k \in \mathbb{R}^m$. Similarly, suppose that $\psi(x_1) \leq \epsilon_d$ and $\|c^-(x_1)\| > \delta\epsilon_p$, where $\delta\epsilon_p \leq \epsilon_d$ and $\delta \in (0, 1)$. Then (2.43) holds with $k = 1$ for some vector of multipliers $z_1 \in \mathbb{R}^m$.

Proof. Applying Theorem 2.1 to ϕ when $\chi(x_k, t_k) \leq \epsilon_d$, we have that

$$\|\nu_k g(x_k) + J^T(x_k)z_k\| \leq \epsilon_d, \tag{2.44}$$

where $(\nu_k, z_k) \in \partial l_\phi(x_k, d_k; t_k)$ for some d_k with $\|d_k\| \leq 1$. Now suppose that $l_\phi(x_k, d_k; t_k) = 0$. In this case

$$\chi(x_k, t_k) = l_\phi(x_k, 0; t_k) = \phi(x_k, t_k). \tag{2.45}$$

But (2.28) ensures that $\phi(x_k;t_k) = \epsilon_p$, in which case (2.45) contradicts the requirement $\chi(x_k,t_k) \leq \epsilon_d < \epsilon_p$. Thus

$$l_\phi(x_k,d_k;t_k) > 0. \tag{2.46}$$

Standard convex analysis (see for example, [5, Thm. 11.4.1 & Cor. 11.4.2], and use (2.46) to ensure that $\|(\nu, z^T)^T)\|_{\rm D} = 1$) gives that

$$\partial l_\phi(x_k,d_k;t_k) = \left\{ \left(\begin{array}{c}\nu\\ z\end{array}\right) = \left(\begin{array}{c}\nu\\ z_{\mathcal{E}}\\ z_{\mathcal{I}}\end{array}\right) \left| \begin{array}{l} \nu[f(x_k) - t_k + \langle g(x_k), d_k\rangle] + \langle z, c(x_k) + J(x_k)d_k\rangle \\ = l_\phi(x_k,d_k;t_k), \;\; \nu \geq 0, \;\; z_{\mathcal{I}} \leq 0 \;\text{ and }\; \left\| \left(\begin{array}{c}\nu\\ z\end{array}\right)\right\|_{\rm D} = 1 \end{array}\right. \right\}. \tag{2.47}$$

But since $(\nu_k, z_k) \in \partial l_\phi(x_k,d_k;t_k)$, we deduce

$$\begin{array}{l} l_\phi(x_k,d_k;t_k) \\ \quad = \nu_k[f(x_k) - t_k + \langle g(x_k), d_k\rangle] + \langle z_k, c(x_k) + J(x_k)d_k\rangle \\ \quad = \nu_k[f(x_k) - t_k] + \langle z_k, c(x_k)\rangle + \langle d_k, \nu_k g(x_k) + J^T(x_k)z_k\rangle, \end{array} \tag{2.48}$$

and the definition of l_ϕ, together with the fact that d_k minimizes $l_\phi(x_k,d;t_k)$ when $\|d\| \leq 1$, gives

$$0 \leq l_\phi(x_k,d_k;t_k) \leq l_\phi(x_k,0;t_k) = \phi(x_k;t_k). \tag{2.49}$$

It follows from the definition of the subgradient, (2.49), (2.48), the Cauchy-Schwarz inequality, $0 < \nu_k \leq 1$, (2.26), Lemma 2.3, $\|d_k\| \leq 1$ and (2.44) that

$$\begin{array}{rl} \langle z_k, c(x_k)\rangle & \leq -\nu_k(f(x_k) - t_k) - \langle d_k, \nu_k g(x_k) + J^T(x_k)z_k\rangle + \phi(x_k;t_k) \\ & \leq \|d_k\|\|\nu_k g(x_k) + J^T(x_k)z_k\| + \phi(x_k;t_k) \\ & \leq \epsilon_p + \epsilon_d. \end{array}$$

Similarly

$$\begin{array}{rl} \langle z_k, c(x_k)\rangle & \geq -\nu_k|f(x_k) - t_k| - \langle d_k, \nu_k g(x_k) + J^T(x_k)z_k\rangle \\ & \geq -|f(x_k) - t_k| - \|d_k\|\|\nu_k g(x_k) + J^T(x_k)z_k\| \\ & \geq -\epsilon_p - \epsilon_d. \end{array}$$

Thus since $\epsilon_p + \epsilon_d \leq \epsilon_c$, we have

$$|\langle z_k, c(x_k)\rangle| \leq \epsilon_c. \tag{2.50}$$

Now suppose that $\nu_k \neq 0$, so that $\nu_k > 0$. In this case, define $y_k = z_k/\nu_k$. Then (2.44) and (2.50) become

$$\nu_k\|g(x_k) + J^T(x_k)y_k\| \leq \epsilon_d \;\text{ and }\; \nu_k|\langle y_k, c(x_k)\rangle| \leq \epsilon_c,$$

while $\|(\nu_k, z_k)\|_{\rm D} = 1$ gives $\nu_k = 1/\|(1, y_k)\|_{\rm D}$. Combining these, and using the assumption $\|c^-(x_k)\| \leq \epsilon_p$ and the deduction $[y_k]_{\mathcal{I}} \leq 0$ from (2.47), it follows that $x_\epsilon = x_k$ and $y_\epsilon = y_k$ satisfy (2.5). If by contrast $\nu_k = 0$, then (2.44), (2.47) and (2.50) directly give (2.43).

The proof of (2.43) when $\psi(x_1) \leq \epsilon_d$ follows in essentially the same way. Applying Theorem 2.1 to $\|c^-(x)\|$ when $\psi(x_1) \leq \epsilon_d$, we have that

$$\|J^T(x_1)z_1\| \leq \epsilon_d, \tag{2.51}$$

where $z_1 \in \partial l_c(x_1,d_1)$ for some d_1 with $\|d_1\| \leq 1$. Now suppose that $l_c(x_1,d_1) = 0$. In this case

$$\psi(x_1) = l_c(x_1,0) = \|c^-(x_1)\|. \tag{2.52}$$

But this contradicts $\psi(x_1) \leq \epsilon_d$ and $\|c^-(x_1)\| > \delta\epsilon_p$ since $\delta\epsilon_p \leq \epsilon_d$. Thus

$$l_c(x_1,d_1) > 0, \tag{2.53}$$

and thus standard convex analysis (see for example, [5, Cor. 11.4.2] , and using (2.53) to ensure that the dual norm of z is one) gives that

$$\partial l_c(x_1,d_1) = \left\{ z = \begin{pmatrix} z_{\mathcal{E}} \\ z_{\mathcal{I}} \end{pmatrix} \;\middle|\; \begin{array}{l} \langle z, c(x_1)+J(x_1)d_1\rangle = l_c(x_1,d_1), \\ z_{\mathcal{I}} \leq 0 \;\text{ and }\; \|z\|_{\rm D} = 1 \end{array} \right\},$$

Hence, as $z_1 \in \partial l_c(x_1,d_1)$, it follows immediately that $[z_1]_{\mathcal{I}} \leq 0$ and $\|z_1\|_{\rm D} = 1$, and thus (2.51) gives (2.43). □

In passing, we note that the requirement $\epsilon_d < \epsilon_p$ in Lemma 2.8 may be removed provided we change Algorithm 2.1 to allow it to take the step $s_k = d_k = \arg\min_{||d||\leq 1} l_\phi(x_k,d,t_k)$ that results from calculating the optimality measure $\chi(x_k,t_k)$ whenever $l_\phi(x_k,d_k,t_k) = 0$.

This leads directly to our desired complexity result.

Theorem 2.9. [Correction to [4, Thm. 3.6]] Suppose that **A1**–**A3** hold. Then there are positive constants $\kappa^{\rm a}_{\sf TR1GC}$, $\kappa^{\rm b}_{\sf TR1GC}$ and $\kappa^{\rm c}_{\sf TR1GC}$ such that, for any $\epsilon_p \in (0,\kappa_{\mathcal{C}_1}]$, $\epsilon_d \in (0,\min(1,\epsilon_p))$ and $\epsilon_p+\epsilon_d \leq \epsilon_c$, Algorithm 2.1 problem (2.1) requires at most

$$\left\lfloor \kappa^{\rm a}_{\sf TR1GC}\,\frac{\|c^-(x_0)\| + f_{\rm lup} - f_{\rm low}}{\epsilon_d^2} + \kappa^{\rm b}_{\sf TR1GC}|\log \epsilon_d| + \kappa^{\rm c}_{\sf TR1GC}\right\rfloor \tag{2.54}$$

evaluations of c and f and their derivatives before an iterate x_k is computed for which either

(i)
$$\left\|\begin{pmatrix} c_E(x_k) \\ \min[\,0, c_I(x_k)] \end{pmatrix}\right\| \leq \epsilon_p, \quad \frac{\|g(x_k)+J^T(x_k)y_k\|}{\|(y_k,1)\|_{\rm D}} \leq \epsilon_d, \quad \frac{\langle c(x_k), y_k\rangle}{\|(y_k,1)\|_{\rm D}} \leq \epsilon_c \quad\text{and}\quad [y_k]_{\mathcal{I}} \leq 0$$

for some vector $y_k \in {\rm I\!R}^m$, or

(ii)
$$\left\|\begin{pmatrix} c_E(x_k) \\ \min[\,0, c_I(x_k)] \end{pmatrix}\right\| \geq \delta\epsilon_p, \quad \|J^T(x_k)z_k\| \leq \epsilon_d, \quad [z_k]_{\mathcal{I}} \leq 0 \quad\text{and}\quad \|z_k\|_{\rm D} = 1$$

for some vector $z_k \in {\rm I\!R}^m$.

Proof. We have from Lemma 2.4 that the number of evaluations required to find x_1 is bounded above by

$$\kappa_1\|c^-(x_0)\|\epsilon_d^{-2} \tag{2.55}$$

for some constant $\kappa_1 > 0$. Thus, as $\psi(x_1) \leq \epsilon_d$, Lemma 2.8 ensures that (2.43) holds. If the algorithm terminates at this stage, then both (2.43) and $\|c^-(x_k)\| > \delta\epsilon_p$ hold, and thus Lemma 2.8 and $\epsilon_d \leq 1 \leq f_{\rm lup} - f_{\rm low}$ yield alternative (ii) provided $\kappa^{\rm a}_{\sf TR1GC} \geq \kappa_1$. So now suppose that Phase 2 of the algorithm is entered. We then observe that Lemma 2.5 implies that successful iterations must happen as long as $\chi(x_k,t_k) \geq \epsilon_d$. Moreover, we have that

$$\begin{array}{rcl} f_{\rm low} & \leq & f(x_k) \leq t_k + \epsilon_p \leq t_1 - i_k\kappa_{\rm C}\epsilon_d^2 + \epsilon_p = f(x_1) + \|c^-(x_1)\| - i_k\kappa_{\rm C}\epsilon_d^2 \\ & \leq & f(x_1) + \|c^-(x_0)\| - i_k\kappa_{\rm C}\epsilon_d^2, \end{array} \tag{2.56}$$

where i_k is the number of these successful iterations from iterations 1 to k of Phase 2, and where we use successively **A3**, (2.26), the fact that $t_j \leq t_{j-1} - i_k\kappa_{\rm C}\epsilon_d^2$ on each successful iteration $j-1$, cf. (2.41), the definition of t_1 in the algorithm, and the fact that Phase 1 decreases $\|c^-(x)\|$. Hence, we obtain from the inequality $f(x_1) \leq f_{\rm lup}$ (itself implied by **A3** again) that

$$i_k \leq \left\lfloor \frac{f_{\rm lup} - f_{\rm low} + \|c^-(x_0)\|}{\kappa_{\rm C}\epsilon_d^2} \right\rfloor. \tag{2.57}$$

The number of Phase 2 iterations satisfying $\chi(x_k, t_k) \geq \epsilon_d$ is therefore bounded above, and the algorithm must terminate after (2.57) such iterations at most, yielding, because of Lemma 2.8, an ϵ-first-order critical point satisfying one of the alternatives (i) or (ii). Remembering that only one evaluation of c and f (and their derivatives, if successful) occurs per iteration, we therefore conclude from (2.57) and Lemma 2.6 that the total number of such evaluations in Phase 2 is bounded above by

$$\left\lfloor \frac{f_{\rm lup} - f_{\rm low} + \|c^-(x_0)\|}{\kappa_{\rm C}\epsilon_d^2} \right\rfloor + \kappa_2 |\log \epsilon| + \kappa_3$$

for some positive constants κ_2 and κ_3.

Summing this upper bound with that for the number of iterations in Phase 1 given by (2.55) and using also that $\epsilon_d \leq 1 \leq f_{\rm lup} - f_{\rm low}$, then yields (2.54) with

$$\kappa^{\rm a}_{\sf TR1GC} = \kappa_1 + \frac{1}{\kappa_{\rm C}}, \quad \kappa^{\rm b}_{\sf TR1GC} = \kappa_2 \;\text{ and }\; \kappa^{\rm c}_{\sf TR1GC} = \kappa_3.$$

□

3 Conclusion

We have corrected two errors that appeared in our paper [4] on the complexity of finding ϵ-approximate first-order critical points for the general smooth constrained optimization problem. We did so by re-defining both the stopping rules and merit function that we used. Our new algorithm and analysis reveals a worst-case complexity bound of $O(\epsilon^{-2})$ function and derivative evaluations to find either an approximate (scaled) KKT point or an approximate stationary point of the violation of the constraints.

As was the case in [4], we make no claim that this is an effective method in practice, merely that we are able to find an algorithm whose worst-case evaluation complexity matches that which is known for first-order methods for non-convex unconstrained minimization [1].

Acknowledgement

The authors wish to thank Ernesto Birgin, John Gardenghi, Jose-Mario Martinez and Sandra Santos for discussing the issue with them while this correction was being polished, and Sven Leyffer, an anonymous associate editor and two referees for their helpful comments on a earlier draft of this work.

References

[1] 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.

[2] C. Cartis, N. I. M. Gould, and Ph. L. Toint. On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming. *SIAM Journal on Optimization*, 21(4):1721–1739, 2011.

[3] 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.

[4] 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(2):93–106, 2014.

[5] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods.* Number 01 in MPS-SIAM Series on Optimization. SIAM, Philadelphia, USA, 2000.

[6] R. Fletcher. *Practical Methods of Optimization.* J. Wiley and Sons, New York and Chichester, second edition, 1987.

[7] F. John. Extreme problems with inequalities as subsidiary conditions. In O. Neugebauer K. O. Friedrichs and J. J. Stoker, editors, *Studies and Essays Presented to R. Courant on His 60th Birthday, Jan. 8, 1948*, page 187204, New York and Chichester, 1948. J. Wiley and Sons.

[8] W. Karush. Minima of functions of several variables with inequalities as side conditions. Master's thesis, Department of Mathematics, University of Chicago, Illinois, USA, 1939.

[9] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In J. Neyman, editor, *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, California, USA, 1951. University of Berkeley Press.

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