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

by C. Cartis, N. I. M. Gould and Ph. L. Toint

Report NAXYS-13-2011 13 April 2011

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

http://www.fundp.ac.be/sciences/naxys

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

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

13 April 2011

Abstract

The complexity of finding ϵ-approximate first-order critical points for the general smooth constrained optimization problem is shown to be no worse that $O(\epsilon^{-2})$ in terms of function and constraints evaluations. This result is obtained by analyzing the worst-case behaviour of a first-order short-step homotopy algorithm consisting of a feasibility phase followed by an optimization phase, and requires minimal assumptions on the objective function. Since a bound of the same order is known to be valid for the unconstrained case, this leads to the conclusion that the presence of possibly nonlinear/nonconvex inequality/equality constraints is irrelevant for this bound to apply.

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

1 Introduction

Evaluation complexity analysis for nonconvex smooth optimization problems has recently been a very active area of research and has covered both standard methods for the unconstrained case, such as steepest-descent (see Vavasis, 1993, Nesterov, 2004, Cartis, Gould and Toint, 2010*b*), trust-region methods (see Gratton, Sartenaer and Toint, 2008), Newton's algorithm (see Cartis et al., 2010*b*) or finite-difference and derivative-free approaches (see Vicente, 2010, Cartis, Gould and Toint, 2010*c*), along with newer approaches involving regularization (see Nesterov and Polyak, 2006, Cartis, Gould and Toint, 2010*a*). The issue considered in this paper is to bound the number of objective function (and gradient) evaluations that are necessary to find an approximate first-order critical point for the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{such that} & x \in \mathbb{R}^n, \end{array}$$

where f is a continuously differentiable possibly nonconvex function from $\mathbb{R}^n$ to $\mathbb{R}$ with Lipschitz continuous gradient. Such an approximate critical point is defined as a point x such that

$$\|g(x)\| \leq \epsilon, \tag{1.1}$$

where $\epsilon \in (0,1)$ is a user-specified accuracy, $\|\cdot\|$ is the Euclidean norm and $g(x) \stackrel{\text{def}}{=} \nabla_x f(x)$. For first-order methods, i.e., for steepest-descent and trust-region algorithm with linear models, it has been shown that this maximum number of objective function (and gradient) evaluations is bounded above by

$$\left\lceil \frac{\kappa}{\epsilon^2} \right\rceil \tag{1.2}$$

for some constant $\kappa > 0$ independent of n and ϵ (Nesterov, 2004, Gratton et al., 2008). Moreover, Cartis et al. (2010*b*) proved that this order in ϵ is sharp. A first extension of this kind of results to constrained

*School of Mathematics, University of Edinburgh, The King's Buildings, Edinburgh, EH9 3JZ, Scotland, UK. Email: coralia.cartis@ed.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

problems was provided by Cartis, Gould and Toint (2009), where it is shown that (1.2) also holds for a first-order projection-based method for the more general problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{such that} & x \in \mathcal{C}, \end{array} \tag{1.3}$$

where $\mathcal{C}$ is a convex set and where (1.1) is suitably adapted to define an ϵ-approximate first-order critical point for constrained problem (1.3). Alternatively, Cartis, Gould and Toint (2011*b*) considered a first-order exact penalty function algorithm for solving the completely general nonlinearly constrained nonconvex optimization problem

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{such that} & c_E(x) = 0, \\ \text{and} & c_I(x) \geq 0, \end{array} \tag{1.4}$$

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. They proved that the complexity of finding an ϵ-approximate first-order critical point for (2.1) is given by an appropriate variant of (1.2) when the penalty parameters remain finite and is bounded above by $O(\epsilon^{-4})$ otherwise. Unfortunately, this last result requires the undesirable assumption that the objective function f is bounded below on the whole of $\mathbb{R}^n$.

In this paper, we provide a theoretical approach to the same problem and improve the results of Cartis et al. (2011*b*) by showing that the complexity of achieving ϵ-approximate first-order criticality for (1.4) is bounded by (1.2) for a first-order algorithm. Morever, this stronger result only require f to be bounded in an ϵ-neighbourhood of the feasible set, which is considerably weaker than assuming this property on the whole space.

The paper is organized as follows. The SSSD Algorithm is introduced in Section 2 for approximately solving the equality constrained problem and its complexity is shown in Section 3 to be bounded above by (1.2). Section 4 briefly covers the simple extension of this result to the general problem (1.4). Some conclusions and perspectives are finally proposed in Section 5.

2 The SSSD Algorithm for the equality constrained problem

For the sake of simplicity, we start by considering the equality constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{such that} & c(x) = 0, \end{array} \tag{2.1}$$

where c is a continuously differentiable function from $\mathbb{R}^n$ to $\mathbb{R}^m$ with Lipschitz continuous Jacobian. The algorithm we now describe consists of two phases. In the first, a first-order algorithm is applied to minimize $\|c(x)\|$ (independently of the objective function f), resulting in a point which is either (approximately) feasible, or is an approximate infeasible stationary point of $\|c(x)\|$. This last outcome is not desirable if one wishes to solve (2.1), but cannot be avoided by any algorithm not relying on global minimization. If an (approximate) feasible point has been found, Phase 2 of the algorithm then performs short steps along generalized steepest-descent directions so long as first-order criticality is not satisfied. These steps are computed by attempting to preserve feasibility of the iterates while producing values of the objective function that are close to a sequence of decreasing "targets".

Both phases rely on the first-order trust-region algorithm[(1)] proposed in Cartis et al. (2011*b*), which can be used to solve the problem

$$\begin{array}{ll} \text{minimize} & \theta\big(u(x)\big) \\ \text{such that} & x \in \mathbb{R}^n, \end{array} \tag{2.2}$$

where θ is a (potentially nonsmooth) convex and globally Lipschitz function from $\mathbb{R}^p$ into $\mathbb{R}$ and $u(x)$ is a continuously differentiable function from $\mathbb{R}^n$ into $\mathbb{R}^p$ with Lipschitz continuous Jacobian $A(x)$. In this

[(1)] We make this choice for simplicity of exposition, but other methods can be considered with similar results. In particular, the quadratic regularization technique of Cartis et al. (2011*b*) or the trust-region technique proposed by Byrd, Gould, Nocedal and Waltz (2005) are also adequate.

algorithm, a "Cauchy step" s_k is obtained from the iterate x_k by solving the linearized model problem

$$\begin{array}{ll} \text{minimize} & \theta\big(u(x_k) + A(x_k)s\big) \\ \text{such that} & \|s\| \le \Delta_k, \end{array} \tag{2.3}$$

where Δ_k is a trust-region radius. Because θ is convex and its argument in (2.3) linear, this problem is computationally tractable. The rest of the algorithm specification follows standard trust-region technology.

We now return to the solution of problem (2.1) proper, and define the merit function

$$\phi(x,t) \stackrel{\text{def}}{=} \|c(x)\| + |f(x) - t|, \tag{2.4}$$

where t is meant as a "target" for $f(x)$. We also define the local linearizations of $\|c(x)\|$ and $\phi(x,t)$ given by

$$\ell_c(x,s) \stackrel{\text{def}}{=} \|c(x) + J(x)s\| \quad \text{and} \quad \ell_\phi(x,t,s) \stackrel{\text{def}}{=} \ell_c(x,s) + |f(x) + \langle g(x), s\rangle|,$$

(where $\langle\cdot,\cdot\rangle$ is the Euclidean inner product). The value of the decrease of the linearized model in a ball of unit radius may then be considered as a first-order criticality measure for the problems of minimizing $\|c(x)\|$ and $\phi(x,t)$, yielding the measures

$$\psi(x) \stackrel{\text{def}}{=} \ell_c(x,0) - \min_{\|d\|\le 1} \ell_c(x,d) \quad \text{and} \quad \chi(x,t) \stackrel{\text{def}}{=} \ell_\phi(x,t,0) - \min_{\|d\|\le 1} \ell_\phi(x,t,d).$$

Note that $\psi(x)$ is zero if and only if x is first-order critical for the problem of minimizing $\|c(x)\|$, while $\chi(x,t)$ is zero if and only if (x,t) is a first-order critical point for the problem

$$\begin{array}{ll} \text{minimize} & \phi(x,t) \\ \text{such that} & x \in \mathbb{R}^n, \end{array} \tag{2.5}$$

(t fixed). In Phase 1 of the SSSD algorithm (aiming for feasiblility), we apply the first-order trust-region algorithm of Cartis et al. (2011*b*) by identifying

$$p = m, \quad u(x) = c(x), \quad \text{and} \quad \theta(\cdot) = \|\cdot\|, \tag{2.6}$$

in (2.2), yieding $\theta\big(u(y)\big) = \|c(x)\|$. For Phase 2 (the optimality phase), we choose in (2.2), for t fixed,

$$p = m+1, \quad u(x) = (c(x), f(x) - t) \quad \text{and} \quad \theta(\cdot) = \|\cdot\| + |\cdot|, \tag{2.7}$$

which gives $\theta\big(u(x)\big) = \phi(x,t)$. Note that $\theta(\cdot)$ is clearly convex with global Lipschitz constant equal to one in both cases.

We are now ready to formalize our Short Step Steepest-Descent (SSSD) Algorithm as presented on the following page.

Since the SSSD algorithm makes no pretense of being practical, we have written Steps 2.2.b and 2.2.c by only using the constants

$$0 < \eta < 1, \quad \text{and} \quad 0 < \gamma < 1,$$

instead of the more usual $\eta_1 \le \eta_2$ and $\gamma_1 \le \gamma_2$, a simplified choice which is allowed in the standard trust-region case, including that studied in Cartis et al. (2011*b*)[(2)]. Note that the SSSD algorithm requires one evaluation of the objective function and its gradient and one evaluation of the constraint's function and its Jacobian per iteration.

Note also that one could also consider using the ARC algorithm (see Cartis, Gould and Toint, 2011*a*) to minimize $\|c(x)\|^2$ to find x_1 such that $\|J(x_1)^T c(x_1)\| \le \epsilon$. We do not consider this (potentially more efficient) possibility here because it would require stronger assumptions on the constraint function c.

[(2)] By selecting $\eta_1 = \eta_2$ and $\gamma_1 = \gamma_2$ in this reference.

Algorithm 2.1: The SSSD algorithm

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

Phase 1:

Starting from x_0, minimize $\|c(x)\|$ (using (2.2) and (2.6) and the trust-region method of Cartis et al., 2011*b*) until a point x_1 is found such that

$$\psi(x_1) \leq \epsilon.$$

If $\|c(x_1)\| > \kappa_f \epsilon$, terminate [locally infeasible].

Phase 2:

1. Set $t_1 = \|c(x_1)\| + f(x_1) - \epsilon$ and $k = 1$.
2. While $\chi(x_k, t_k) \geq \epsilon$,
   (a) Compute a first-order step s_k by solving

$$\begin{array}{ll} \text{minimize} & \ell_\phi(x_k, t_k, s) \\ \text{such that} & \|s\| \leq \Delta_k, \end{array} \tag{2.8}$$

   (b) Compute $\phi(x_k + s_k, t_k)$ and define

$$\rho_k = \frac{\phi(x_k, t_k) - \phi(x_k + s_k, t_k)}{\ell_\phi(x_k, t_k, 0) - \ell_\phi(x_k, t_k, s_k)}. \tag{2.9}$$

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

$$\Delta_{k+1} = \begin{cases} \Delta_k & \text{if } \rho_k \geq \eta \quad [k \text{ successful}] \\ \gamma \Delta_k & \text{if } \rho_k < \eta, \quad [k \text{ unsuccessful}] \end{cases} \tag{2.10}$$

   (d) If $\rho_k \geq \eta$, set

$$t_{k+1} = \begin{cases} t_k - \phi(x_k, t_k) + \phi(x_{k+1}, t_k) & \text{if } f(x_{k+1}) \geq t_k, \\ 2f(x_{k+1}) - t_k - \phi(x_k, t_k) + \phi(x_{k+1}, t_k) & \text{if } f(x_{k+1}) < t_k. \end{cases} \tag{2.11}$$

   Otherwise, set $t_{k+1} = t_k$.
   (e) Increment k by one and return to 2.
3. Terminate [(approximately) first-order critical]

3 Complexity of the SSSD Algorithm for the equality constrained problem

Before analyzing the complexity of Algorithm SSSD, we state our assumptions formally.

A.1: The function c is continuously differentiable on $\mathbb{R}^n$ and f is continuously differentiable in an open neighbourhood of

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

A.2: $J(x)$ is globally Lipschitz continuous in $\mathbb{R}^n$ with Lipschitz constant bounded above by $L_J > 0$, and $g(x)$ is Lispchitz continuous in $\mathcal{C}_\epsilon$ with Lipschitz constant bounded above by $L_g > 0$.

A.3: The objective function is bounded above and below in the neighbourhood of the feasible set, that is there exist constants f_{low} and $f_{\text{up}} \geq f_{\text{low}} - 1$ such that

$$f_{\text{low}} \leq f(x) \leq f_{\text{up}} \qquad \text{for all } x \in \mathcal{C}_\epsilon.$$

We start our analysis by exploiting the results of Cartis et al. (2011*b*) and bounding the number of Phase 1 iterations.

Lemma 3.1 *Suppose that A.1 and A.2 hold. Then, at most*

$$\left\lceil \|c(x_0)\| \frac{\kappa_1}{\epsilon^2} \right\rceil \tag{3.12}$$

evaluations of $c(x)$ and its derivatives are needed to complete Phase 1, for some $\kappa_1 > 0$ independent of n, ϵ and x_0.

Proof. See Theorem 2.4 in Cartis et al. (2011*b*). □

We next extract from the same reference a property which is crucial for Phase 2.

Lemma 3.2 *Suppose that A.1 and A.2 hold. Suppose also that $\chi(x_k, t_k) \geq \epsilon$ and that*

$$\Delta_k \leq \frac{(1-\eta)\epsilon}{L_g + \frac{1}{2}L_J}. \tag{3.13}$$

Then iteration k is successful and

$$\phi(x_k + s_k, t_k) \leq \phi(x_k, t_k) - \kappa_{\rm C}\epsilon^2, \tag{3.14}$$

where

$$\kappa_{\rm C} \stackrel{\rm def}{=} \eta \min\left[\Delta_1, \frac{(1-\eta)\gamma}{L_g + \frac{1}{2}L_J}\right]. \tag{3.15}$$

is a constant independent of n and ϵ.

Proof. This follows by applying Lemmas 2.1 and 2.3 in Cartis et al. (2011*b*) to the objective $\phi(x, t_k)$ considered as a function of x only. □

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

Lemma 3.3 *There are at most $O(|\log \epsilon|)$ unsuccessful iterations in Phase 2 of the SSSD algorithm.*

Proof. Note that (2.10) implies that the trust-region radius is never increased, and therefore Lemma 3.2 guarantees that all iterations must be successful once Δ_1 has been reduced (by a factor γ) enough times to ensure (3.13). Hence there are at most

$$\left\lceil \frac{1}{|\log \gamma|} \left|\log \epsilon + \log(1-\eta) - \log \Delta_1 - \log(L_g + \tfrac{1}{2}L_J)\right| \right\rceil = O(|\log \epsilon|) \tag{3.16}$$

unsuccessful iterations during the complete execution of the Phase 2. □

The next lemma proves the crucial observation that all Phase 2 iterates remain (approximately) feasible, and that the targets t_k decrease by a quantity bounded below by a multiple of ϵ^{-2} at every successful iteration.

Lemma 3.4 *For every $k \geq 1$, we have that*

$$f(x_{k+1}) - t_{k+1} > 0, \tag{3.17}$$

$$\phi(x_k, t_k) = \epsilon \tag{3.18}$$

$$\|c(x_k)\| \leq \epsilon, \tag{3.19}$$

$$|f(x_k) - t_k| \leq \epsilon. \tag{3.20}$$

Moreover, if iteration k is successful, then

$$t_k - t_{k+1} \geq \kappa_{\rm C}\epsilon^2 \tag{3.21}$$

where $\kappa_{\rm C}$ is defined in (3.15).

Proof. We start by observing that (3.17) immediately follows from (2.11) and (3.14) and the fact that $f(x_k)$ and t_k remain unchanged on unsuccessful iterations. We now prove (3.18) by induction on k. We first note that this inequality holds by construcion for $k = 1$. We now show that this bound remains valid for $k > 1$. Assume that iteration k is successful and that

$$\phi(x_k, t_k) = \epsilon. \tag{3.22}$$

Using (2.4) and (3.17), we observe that

$$\phi(x_{k+1}, t_{k+1}) = \|c(x_{k+1})\| + f(x_{k+1}) - t_k + (t_k - t_{k+1}). \tag{3.23}$$

Consider the case where $f(x_{k+1}) \geq t_k$ first. Then, using (3.23) and (2.11), 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 used (3.23) and (2.11). Combining the two cases and using (3.14) and (3.22), we then deduce that

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

By induction, and since t_k and $f(x_k)$ (and hence ϕ) are unmodified at unsuccessul iterations, (3.18) therefore holds for all $k \geq 1$. Relations (3.19) and (3.20) immediately follow from (2.4). Finally, (3.21) results from (2.11) and (3.14) at successful iterations. □

Lemma 3.5 *Assume that $\|c(x_k)\| \leq \epsilon$ and $\chi(x_k, t_k) \leq \epsilon$. Then x_k is an approximate critical point in the sense that*

$$\|c(x_k)\| \leq \epsilon \quad \text{and} \quad \|J(x_k)^T y - g(x_k)\| \leq \epsilon \tag{3.24}$$

for some vector of multipliers $y \in I\!R^m$. Similarly, assume that $\psi(x) \leq \epsilon$. Then

$$\|J(x)^T z\| \leq \epsilon \tag{3.25}$$

for some vector of multipliers $z \in I\!R^m$.

Proof. See Theorem 3.1 in Cartis et al. (2011*b*) and the comments thereafter. □

Theorem 3.6 *Assume A.1-A.3 hold. Then the SSSD algorithm generates an ϵ-first-oder critical point for problem (2.1), that is an iterate x_k satisfying either*

$$\text{(3.24)} \quad \text{or} \quad \big[\text{(3.25) with } \|c(x_k)\| > \kappa_f \epsilon\big],$$

in at most

$$\left\lceil \left(\|c(x_0)\| + f_{\text{up}} - f_{\text{low}} \right) \frac{\kappa_2}{\epsilon^2} \right\rceil \tag{3.26}$$

evaluations of c and f (and their derivatives), where $\kappa_2 > 0$ is a constant independent of n, ϵ and x_0.

Proof. We have seen in Lemma 3.1 that the complexity of obtaining x_1 is bounded above by $O(\lceil \|c(x_0)\| \epsilon^{-2} \rceil)$. Thus, as $\psi(x_1) \leq \epsilon$, Lemma 3.5 ensures that (3.25) holds. If the algorithm terminates at this stage, then both (3.25) and $\|c(x_k)\| > \kappa_f \epsilon$ hold, as requested. Assume now that Phase 2 of the algorithm is entered. We then observe that Lemma 3.2 implies that successful iterations must happen as long as $\chi(x_k, t_k) \geq \epsilon$. Moreover, we have that

$$f_{\text{low}} \leq f(x_k) \leq t_k + \epsilon \leq t_1 - \sigma_k \kappa_{\text{C}} \epsilon^2 + \epsilon \leq f(x_1) - \sigma_k \kappa_{\text{C}} \epsilon^2 + \epsilon$$

where σ_k is the number of these successful iterations from iterations 1 to k of Phase 2, and where we have successively used A.3, (3.20) and (3.21). Hence, we obtain from the inequality $f(x_1) \leq f_{\rm up}$ (itself implied by A.3 again) that

$$\sigma_k \leq \left\lceil \frac{f_{\rm up} - f_{\rm low} + \epsilon}{\kappa_{\rm C}\epsilon^2} \right\rceil . \tag{3.27}$$

The number of Phase 2 iterations satisfying $\chi(x_k, t_k) \geq \epsilon$ is therefore bounded above, and the algorithm must terminate after (3.27) such iterations at most, yielding, because of Lemma 3.5, an ϵ-first-order critical point satisfying (3.24). Remembering that only one evaluation of c and f (and their derivatives, if successful) occurs per iteration, we therefore conclude from (3.27) and Lemma 3.3 that the total number of such evaluations in Phase 2 is bounded above by

$$\left\lceil \frac{f_{\rm up} - f_{\rm low} + \epsilon}{\kappa_{\rm C}\epsilon^2} \right\rceil + O(|\log \epsilon|)$$

Summing this upper bound with that for the number of iterations in Phase 1 given by Lemma 3.1 then yields (3.26). □

4 Including general inequality constraints

If we now return to the solution of problem (1.4), we may consider defining

$$c(x) = \begin{pmatrix} c_E(x) \\ \min[\,0, c_I(x)\,] \end{pmatrix}$$

in the above. The quantity $\|c(x)\|$ can again be considered as the composition of a nonsmooth convex function with the smooth function $(c_E(x)^T, c_I(x)^T)^T$ and the theory developed above applies without modification, except that Lemma 3.5 must be adapted for the presence of inequality constraints. If an inequality constraint is active at an approximate critical point, then its multiplier has to be non-negative because $y \in \partial(\|\min[0, \cdot\,]\|)$ implies that $y \geq 0$. If it is inactive, then it may as well be absent from the problem (and its multiplier must be zero). Hence Lemma 3.5 generalizes to the inequality constraints case (1.4) without difficulty.

5 Conclusions

We have shown that the complexity to achieve either a ϵ-first-order critical point or an infeasible ϵ-critical point for the infeasibilities of the general smooth nonlinear optimization problem (1.4) is $O(\epsilon^{-2})$, where the constant involved is independent of problem dimension. This is a marked improvement over the results presented in Cartis et al. (2011*b*), where the same complexity was achieved only if the penalty parameter of the minimization scheme discussed in this reference remains bounded, the complexity being $O(\epsilon^{-4})$ othrewise. Moreover, the results obtained in the present paper only assume boundedness of the objective function on an ϵ-neighbourhood of the feasible set, rather than on the whole space.

Since Cartis et al. (2010*b*) have shown that the $\mathrm{O}(\epsilon^{-2})$ bound can be effectively achieved by steepest descent in the unconstrained case, improving the same bound in the constrained case is also impossible for methods of the same type.

We fully accept that the SSSD algorithm discussed in Section 2 is most likely to be extremely inefficient in practice, because it amounts to following the constraints manifold with very small steps. "Long steps" variants may be considered in which the setting of the target t_k is more aggressively geared towards minimizing the objective function. Whether such variants can be numerically effective remains to be seen, but their complexity will be difficult to guarantee with the kind of technique used here, as this would rely on global optimization of the constraint violation.

That we expect Algorithm SSSD to be outperformed in practice is to be welcomed, indicating that the $O(\epsilon^{-2})$ evaluation bound may be as pessimistic for the contrained case as it is for the unconstrained one. But it remains remarkable that this pessimistic bound is unaffected by the presence of possibly nonlinear and nonconvex constraints.

Acknowledgements

The work of the second author is funded by EPSRC Grant EP/E053351/1. All three authors are grateful to the Royal Society for its support through the International Joint Project 14265.

References

R. H. Byrd, N. I. M. Gould, J. Nocedal, and R. A. Waltz. On the convergence of successive linear-quadratic programming algorithms. *SIAM Journal on Optimization*, **16**(2), 471–489, 2005.

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. Technical Report 08/05R, Department of Mathematics, FUNDP - University of Namur, Namur, Belgium, 2009.

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*, 2010*a*. DOI: 10.1007/s10107-009-0337-y.

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*b*.

C. Cartis, N. I. M. Gould, and Ph. L. Toint. On the oracle complexity of first-order and derivative-free algorithms for smooth nonconvex minimization. Technical Report naXys-03-2010, Namur Centre for Complex Systems (naXys), FUNDP-University of Namur, Namur, Belgium, 2010*c*.

C. Cartis, N. I. M. Gould, and Ph. L. Toint. Adaptive cubic overestimation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Mathematical Programming, Series A*, **127**(2), 245–295, 2011*a*.

C. Cartis, N. I. M. Gould, and Ph. L. Toint. On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming. Technical Report naXys-06-2011, Namur Centre for Complex Systems (naXys), FUNDP-University of Namur, Namur, Belgium, 2011*b*.

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.

L. N. Vicente. Worst case complexity of direct search. Technical report, Department of Mathematics, University of Coimbra, Coimbra, Portugal, May 2010.