Optimization Methods and Software

ISSN: 1055-6788 (Print) 1029-4937 (Online) Journal homepage: http://china.tandfonline.com/loi/goms20

On the worst-case complexity of nonlinear stepsize control algorithms for convex unconstrained optimization

G.N. Grapiglia, J. Yuan & Y. Yuan

Optimization Methods & Software, 2016
Vol. 31, No. 3, 591–604, http://dx.doi.org/10.1080/10556788.2015.1130129

On the worst-case complexity of nonlinear stepsize control algorithms for convex unconstrained optimization

G.N. Grapiglia[a*], J. Yuan[a] and Y. Yuan[b]

[a]*Department of Mathematics, Federal University of Paraná, Curitiba, Brazil;* [b]*State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*

(*Received 11 February 2015; accepted 7 December 2015*)

A Nonlinear Stepsize Control (NSC) framework has been proposed by Toint [*Nonlinear stepsize control, trust regions and regularizations for unconstrained optimization*, Optim.Methods Softw. 28 (2013), pp. 82–95] for unconstrained optimization, generalizing many trust-region and regularization algorithms. More recently, worst-case complexity bounds for the generic NSC framework were proved by Grapiglia *et al.* [*On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization*, Math. Program. 152 (2015), pp. 491–520] in the context of non-convex problems. In this paper, improved complexity bounds are obtained for convex and strongly convex objectives.

Keywords: worst-case complexity; trust-region methods; regularization methods; unconstrained optimization

AMS Subject Classification: 90C30; 65K05; 49M37; 49M15; 90C29

1. Introduction

We are interested in the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is continuously differentiable and bounded below by f_*. Practical numerical methods for problem (1) are iterative. Given an initial point $x_1\in\mathbb{R}^n$, at the kth iteration a new iterate point x_{k+1} is obtained by using the information at the current iterate point x_k and, possibly, the information at the previous points. It is expected that at least one subsequence of the sequence $\{x_k\}$ generated by the method will converge to a critical point of f. The goal of the worst-case complexity analysis is the estimation of upper bounds on the number of iterations (or function/gradient evaluations) required in the worst-case to generate an iterate x_k such that $\|\nabla f(x_k)\|\leq\epsilon$ or $f(x_k)-f_*\leq\epsilon$, for a given $\epsilon>0$ (see [5]).

When the objective function f is possibly non-convex, it was shown by Nesterov [18] that the steepest descent method takes at most $\mathcal{O}(\epsilon^{-2})$ iterations to generate x_k such that $\|\nabla f(x_k)\|\leq\epsilon$,

*Corresponding author. Email: grapiglia@ufpr.br; geovani_mat@outlook.com

for $\epsilon \in (0,1)$. Similar upper bounds of $\mathcal{O}(\epsilon^{-2})$ have been proved by Gratton *et al.* [16] for trust-region methods, by Cartis *et al.* [4] for the basic ARC algorithm, by Cartis *et al.* [8] for a non-monotone lineasearch algorithm, and by Ueda and Yamashita [23] for the Levenberg–Marquardt method when $f(x) = (\frac{1}{2})\|F(x)\|^2$ with $F : \mathbb{R}^n \to \mathbb{R}^m$ continuously differentiable. Improved complexity bounds of $\mathcal{O}(\epsilon^{-3/2})$ have been proved by Nesterov and Polyak [20] for the cubic regularization of the Newton's method, by Cartis *et al.* [4] for second-order variants of the ARC algorithm and, recently, by Curtis *et al.* [10] for a modified trust-region method. Furthermore, Cartis *et al.* [2] have proved that the bound of $\mathcal{O}(\epsilon^{-3/2})$ is sharp for the ARC algorithm, and the bounds of $\mathcal{O}(\epsilon^{-2})$ are sharp for the steepest descent method, the Newton's method and the classical trust-region method. In the derivative-free setting, Vicente [24] has proved a complexity bound of $\mathcal{O}(n^2\epsilon^{-2})$ function evaluations for direct-search methods, while Cartis *et al.* [7] have proved a bound of $\mathcal{O}((n^2+5n)[1+|\log(\epsilon)|+\epsilon^{-3/2}])$ function evaluations for the ARC algorithm with derivatives approximated by finite-differences.

When the objective function f is convex (or strongly convex), Nesterov [18] has shown that the steepest descent method takes at most $\mathcal{O}(\epsilon^{-1})$ (resp. $\mathcal{O}(\log(\epsilon^{-1}))$ iterations to generate x_k such that $f(x_k) - f_* \leq \epsilon$, where f_* is the global minimum of f. Similar bounds of $\mathcal{O}(\epsilon^{-1})$ and $\mathcal{O}(\log(\epsilon^{-1}))$ have been proved by Cartis *et al.* [6] for the basic ARC when it is applied to the same classes of objectives. Improved complexity bounds of $\mathcal{O}(\epsilon^{-1/2})$ and $\mathcal{O}(C + \log\log(\epsilon^{-1}))$ (for convex and strongly convex objectives, respectively) have been proved by Nesterov and Polyak [20] for the cubic regularization of Newton's method and by Cartis *et al.* [6] for second-order variants of the ARC algorithm.[1] Furthermore, in the context of derivative-free optimization, Dodangeh and Vicente [12] also have obtained an improved complexity bound of $\mathcal{O}(n^2\epsilon^{-1})$ function evaluations for direct-search methods under the convexity assumption.

In this context, the focus of the current paper is the worst-case complexity analysis of the Nonlinear Stepsize Control (NSC) framework when the objective f in (1) is convex. This framework has been proposed by Toint [22] as a generalization of many trust-region and regularization methods. As pointed in [22], the NSC covers the classical trust-region method [9,21], the basic ARC algorithm of Cartis *et al.* [3,4], the quadratic regularization method for $f(x) = \|F(x)\|$ proposed by Nesterov [19] (as extended by Bellavia *et al.* [1]), the modified trust-region method of Fan and Yuan [14], and the quadratic regularization methods for $f(x) = (1/2)\|F(x)\|^2$ proposed by Zhang and Wang [25] and by Fan [13].

To describe the NSC in a compact way, we shall consider functions $\omega, \phi, \psi, \chi : \mathbb{R}^n \to \mathbb{R}$ satisfying the following conditions:

A1 ω is a continuous, bounded and non-negative function, such that

$$\omega(x) = 0 \Longrightarrow \|\nabla f(x)\| = 0. \tag{2}$$

A2 ϕ, ψ, χ are continuous non-negative functions, such that

$$\omega(x) > 0 \quad \text{and} \quad \min\{\phi(x), \psi(x), \chi(x)\} = 0 \Longrightarrow \|\nabla f(x)\| = 0. \tag{3}$$

A3 There exists $\kappa_\chi > 0$ such that

$$\chi(x) \leq \kappa_\chi \quad \text{for all } x. \tag{4}$$

These functions act as stationarity measures. By convention, from here, we denote

$$\phi_k = \phi(x_k), \quad \psi_k = \psi(x_k), \quad \chi_k = \chi(x_k) \quad \text{and} \quad \omega_k = \omega(x_k).$$

Now, the generic NSC framework can be summarized as follows.

Algorithm 1 NSC Algorithm [23]

Step 0 Given $x_1 \in \mathbb{R}^n$, $H_1 \in \mathbb{R}^{n\times n}$, $\delta_1 > 0$, $0 < \gamma_1 < \gamma_2 < \gamma_3 < 1 < \gamma_4$ and $0 < \eta_1 \leq \eta_2 < 1$, set $k := 1$.

Step 1 Choose a model $m_k(x_k + s) : \mathbb{R}^n \to \mathbb{R}$ such that

$$m_k(x_k) = f(x_k) \quad \text{and} \quad f(x_k + s) - m_k(x_k + s) \leq \kappa_m \|s\|^2 \quad \forall s \in \mathbb{R}^n, \tag{5}$$

for some constant $\kappa_m > 0$. Then, compute a step $s_k \in \mathbb{R}^n$ such that

$$\|s_k\| \leq \kappa_s \Delta(\delta_k, \chi_k) \quad \text{whenever } \delta_k \leq \kappa_\delta \chi_k, \tag{6}$$

for some constants $\kappa_s \geq 1$ and $\kappa_\delta > 0$, and

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_c \psi_k \min\left\{\frac{\phi_k}{1 + \|H_k\|}, \Delta(\delta_k, \chi_k)\right\}, \tag{7}$$

for some constant $\kappa_c \in (0, 1)$, where $\Delta(\delta_k, \chi_k) = \delta_k^\alpha \chi_k^\beta$ with powers $\alpha \in (0, 1]$ and $\beta \in [0, 1]$.

Step 2 Compute the ratio

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}, \tag{8}$$

set the next iterate

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{if } \rho_k \geq \eta_1, \\ x_k, & \text{otherwise,} \end{cases} \tag{9}$$

and choose the stepsize parameter δ_{k+1} satisfying

$$\delta_{k+1} \in \begin{cases} [\gamma_1\delta_k, \gamma_2\delta_k], & \text{if } \rho_k < \eta_1, \\ [\gamma_2\delta_k, \gamma_3\delta_k], & \text{if } \rho_k \in [\eta_1, \eta_2), \\ [\delta_k, \gamma_4\delta_k], & \text{if } \rho_k \geq \eta_2, \end{cases} \tag{10}$$

Step 3 Compute $H_{k+1} \in \mathbb{R}^{n\times n}$, set $k := k + 1$ and go to Step 1.

Remark 1 Usually, H_k is an $n \times n$ symmetric matrix approximating the second-order behaviour of f in a neighbourhood of x_k. Moreover, it is worth to mention that δ_{k+1} can be chosen arbitrarily from the intervals specified by (10).

Regarding the worst-case complexity of Algorithm 1, Grapiglia *et al.* [15] have proved upper bounds on the number of iterations required to ensure $F_k = \min\{\omega_k, \phi_k, \psi_k, \chi_k\} \leq \epsilon$, for a given $\epsilon \in (0, 1]$. Before to recall the complexity bounds given in [15], consider the following additional conditions:

A4 There exists a constant $\kappa_H > 0$ such that $\|H_k\| \leq \kappa_H$ for all k.
A5 The powers α and β satisfy the inequalities $\alpha + \beta \leq 1$ and $2\alpha + \beta \geq 1$.
A6 For all k, $\phi_k \geq \chi_k$ and $\psi_k \geq \chi_k$.

Theorem 1.1 (Theorem 7 in [15]) *Suppose that* A1–A4 *hold, and let f be bounded below by f_*. Then, to reduce the criticality measure* $F_k = \min\{\omega_k, \phi_k, \psi_k, \chi_k\}$ *below* $\epsilon \in (0, 1]$, *Algorithm* 1 *takes at most $\mathcal{O}(\epsilon^{-(2+\beta)})$ iterations. If additionally,* A5–A6 *hold, then this worst-case complexity bound is reduced to $\mathcal{O}(\epsilon^{-2})$ iterations.*

In this paper, a class of NSC algorithms is shown to have improved worst-case complexity when applied to convex and strongly convex objectives. Specifically, by generalizing the analysis of Cartis *et al.* [6] for the ARC algorithm, we prove that if $\alpha+\beta=1$, $\chi_k=\|\nabla f(x_k)\|$ and $\phi_k,\psi_k\geq\chi_k$, then Algorithm 1 can take at most $\mathcal{O}(\epsilon^{-1})$ or $\mathcal{O}(\log(\epsilon^{-1}))$ iterations to generate x_k such that

$$D_k\equiv f(x_k)-f_*\leq\epsilon, \tag{11}$$

where f_* is the global minimum of a convex or strongly convex objective f, respectively. As expected, these bounds match in order those obtained by Cartis *et al.* [6] for the basic ARC algorithm, and by Nesterov [19] for the steepest descent method, on the same classes of objectives.

The paper is organized as follows. In Section 2, worst-case complexity bounds for Algorithm 1 are proved in a general setting. In Section 3, the complexity results are specialized for problems where the objective is a gradient dominated function, which includes convex and strongly convex functions. Finally, in Section 4, the results obtained are summarized and some of its consequences are discussed.

2. Worst-case complexity analysis: General case

Throughout this section, we say that iteration k is successful whenever $\rho_k\geq\eta_1$, very successful whenever $\rho_k\geq\eta_2$ and unsuccessful whenever $\rho_k<\eta_1$. From this naming, we shall consider the following notation:

$$S=\{k\geq 1\,|\,k\text{ successful}\}, \tag{12}$$

$$S_j=\{k\leq j\,|\,k\in S\},\ \text{ for each } j\geq 1, \tag{13}$$

$$U_j=\{k\leq j\,|\,k\notin S\}\text{ for each } j\geq 1, \tag{14}$$

where S_j and U_j form a partition of $\{1,\ldots,j\}$, and $|S_j|$ and $|U_j|$ denote the cardinality of these sets.

We initiate our analysis by recalling the following useful property.

Lemma 2.1 *Let Conditions* A1–A4 *hold. Also, assume that*

$$\left(\frac{1}{\delta_k}\right)^{\alpha}\min\{\chi_k^{\alpha},\chi_k^{-\beta}\phi_k,\chi_k^{-\beta}\psi_k\}>\max\left\{\frac{\kappa_m\kappa_s^2}{(1-\eta_2)\kappa_c},\frac{1+\kappa_H}{\kappa_\delta^{\alpha}},1+\kappa_H\right\}\equiv\kappa_{HB}. \tag{15}$$

Then, iteration k is very successful and consequently

$$\delta_{k+1}\geq\delta_k. \tag{16}$$

Proof See Lemma 6 in [15]. ■

From here, we shall consider problems for which the following condition is satisfied:

H1 There exists $\kappa_d>0$ and $p\in[1,2]$ such that $D_k\leq\kappa_d\min\{\chi_k^{\alpha},\chi_k^{-\beta}\phi_k,\chi_k^{-\beta}\psi_k\}^{p/\alpha}$ for all k.

The lemma below gives a lower bound on δ_k when D_k is bounded away from zero. Its proof is a direct adaptation of the proof of Lemma 7 in [15].

LEMMA 2.2 *Let Conditions* A1–A4 *and* H1 *hold. Also, let* $\epsilon \in (0,1]$ *such that* $D_k > \epsilon$ *for all* $k = 1, \ldots, j$, *where* $j \leq +\infty$. *Then, there exists* $\tau > 0$ *independent of* k *and* ϵ *such that*

$$\delta_k \geq \tau \epsilon^{1/p}.$$

Proof First, by induction over k, we shall prove that

$$\delta_k \geq \min\left\{\delta_1, \frac{\gamma_1}{\kappa_{HB}^{1/\alpha} \kappa_d^{1/p}} \epsilon^{1/p}\right\} \tag{17}$$

for $k = 1, \ldots, j+1$. Clearly, (17) is true for $k = 1$. We assume that (17) is true for $k \in \{1, \ldots, j\}$ and prove it is also true for $k+1$. In fact, from H1 and the inequality $D_k > \epsilon$, it follows that

$$\epsilon < D_k \leq \kappa_d \min\{\chi_k^\alpha, \chi_k^{-\beta}\phi_k, \chi_k^{-\beta}\psi_k\}^{p/\alpha}$$
$$\Longrightarrow \min\{\chi_k^\alpha, \chi_k^{-\beta}\phi_k, \chi_k^{-\beta}\psi_k\} > \left(\frac{1}{\kappa_d}\right)^{\alpha/p} \epsilon^{\alpha/p}.$$

Therefore, by Lemma 2.1 and the induction assumption, if

$$\left(\frac{1}{\delta_k}\right)^{\alpha} \left(\frac{1}{\kappa_d}\right)^{\alpha/p} \epsilon^{\alpha/p} > \kappa_{HB}, \tag{18}$$

then

$$\delta_{k+1} \geq \delta_k \geq \min\left\{\delta_1, \frac{\gamma_1}{\kappa_{HB}^{1/\alpha} \kappa_d^{1/p}} \epsilon^{1/p}\right\}$$

and so, (17) is true for $k+1$.

Now, suppose that (18) is not true. Then

$$\left(\frac{1}{\delta_k}\right)^{\alpha} \left(\frac{1}{\kappa_d}\right)^{\alpha/p} \epsilon^{\alpha/p} \leq \kappa_{HB} \Longrightarrow \frac{1}{\delta_k}\left(\frac{1}{\kappa_d}\right)^{1/p} \epsilon^{1/p} \leq \kappa_{HB}^{1/\alpha} \Longrightarrow \delta_k \geq \frac{1}{\kappa_{HB}^{1/\alpha} \kappa_d^{1/p}} \epsilon^{1/p}$$

and so, by rule (10) we see that

$$\delta_{k+1} \geq \gamma_1 \delta_k \geq \frac{\gamma_1}{\kappa_{HB}^{1/\alpha} \kappa_d^{1/p}} \epsilon^{1/p} \geq \min\left\{\delta_1, \frac{\gamma_1}{\kappa_{HB}^{1/\alpha} \kappa_d^{1/p}} \epsilon^{1/p}\right\},$$

that is, (17) is true for $k+1$. It completes the induction argument.

Finally, since $\epsilon^{1/p} \leq 1$, by (17) we conclude that, for $k = 1, \ldots, j+1$,

$$\delta_k \geq \min\left\{\delta_1, \frac{\gamma_1}{\kappa_{HB}^{1/\alpha} \kappa_d^{1/p}}\right\} \epsilon^{1/p} = \tau \epsilon^{1/p},$$

where $\tau = \min\{\delta_1, \gamma_1/\kappa_H^{1/\alpha}, \kappa_d^{1/p}\}$. ■

The next lemma provides a lower bound for the reduction in $m_k(x_k + s)$ produced by s_k. Its proof is based on the Proof of Lemma 2.3 in [6].

LEMMA 2.3 *Suppose that* A1–A4 *and* H1 *hold. Then, there exists a constant* $\kappa_u > 0$ *such that, for each* $k \geq 1$,

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_u D_k^{2(\alpha+\beta)/p}. \tag{19}$$

Proof First, by induction, we shall prove that

$$\left(\frac{1}{\delta_k}\right) D_k^{1/p} \leq \max\left\{\left(\frac{1}{\delta_1}\right) D_1^{1/p}, \gamma_1^{-1}\kappa_d^{1/p}\kappa_{HB}^{1/\alpha}\right\} \equiv \kappa_G, \tag{20}$$

for all k. Clearly, (20) is true for $k = 1$. We assume that (20) is true for k and prove it is also true for $k + 1$. In fact, suppose that

$$\left(\frac{1}{\delta_k}\right) D_k^{1/p} > \kappa_d^{1/p}\kappa_{HB}^{1/\alpha}. \tag{21}$$

In this case, by H1 and Lemma 2.1 we have $\delta_{k+1} \geq \delta_k$. Note that, by (7)–(9) we have

$$D_k - D_{k+1} = f(x_k) - f(x_{k+1}) \geq 0, \quad \text{for all } k.$$

Thus, the sequence $\{D_k\}$ is monotonically non-increasing. Hence, it follows from the induction assumption that

$$\left(\frac{1}{\delta_{k+1}}\right) D_{k+1}^{1/p} \leq \left(\frac{1}{\delta_k}\right) D_k^{\frac{1}{p}} \leq \max\left\{\left(\frac{1}{\delta_1}\right) D_1^{1/p}, \gamma_1^{-1}\kappa_d^{1/p}\kappa_{HB}^{1/\alpha}\right\},$$

and so (20) holds for $k + 1$.

On the other hand, suppose that (21) is not true. Then, from $D_{k+1} \leq D_k$ and (10), we see that (20) is true for $k + 1$:

$$\begin{aligned}\left(\frac{1}{\delta_{k+1}}\right) D_{k+1}^{1/p} &\leq \gamma_1^{-1}\left(\frac{1}{\delta_k}\right) D_k^{1/p} \\ &\leq \gamma_1^{-1}\kappa_d^{1/p}\kappa_{HB}^{1/\alpha} \\ &\leq \max\left\{\left(\frac{1}{\delta_1}\right) D_1^{1/p}, \gamma_1^{-1}\kappa_d^{1/p}\kappa_{HB}^{1/\alpha}\right\}.\end{aligned}$$

Now we deduce from H1 that

$$\chi_k \geq c_1 D_k^{1/p} \quad \text{and, consequently, } \phi_k, \psi_k \geq c_2 D_k^{(\alpha+\beta)/p}, \tag{22}$$

where $c_1 = \kappa_d^{-1/p}$ and $c_2 = \kappa_d^{-(\alpha+\beta)/p}$. Thus, combining (7), (22), A4 and (20), we conclude that

$$\begin{aligned} m_k(x_k) - m_k(x_k + s_k) &\geq \kappa_c \psi_k \min\left\{\frac{\phi_k}{1 + \|H_k\|}, \delta_k^\alpha \chi_k^\beta\right\} \\ &\geq \kappa_c c_2 D_k^{(\alpha+\beta)/p} \min\left\{\frac{c_2 D_k^{(\alpha+\beta)/p}}{1 + \kappa_H}, \delta_k^\alpha c_1^\beta D_k^{\beta/p}\right\} \\ &= \kappa_c c_2 D_k^{(\alpha+\beta)/p} \min\left\{\frac{c_2 D_k^{(\alpha+\beta)/p}}{1 + \kappa_H}, \frac{c_1^\beta D_k^{(\alpha+\beta)/p}}{\left(\frac{1}{\delta_k}\right)^\alpha D_k^{\alpha/p}}\right\}\end{aligned}$$

$$\geq \kappa_c c_2 D_k^{(\alpha+\beta)/p} \min \left\{ \frac{c_2 D_k^{(\alpha+\beta)/p}}{1+\kappa_H}, \frac{c_1^{\beta} D_k^{(\alpha+\beta)/p}}{\kappa_G^{\alpha}} \right\}$$

$$= \kappa_u D_k^{2(\alpha+\beta)/p}, \tag{23}$$

where $\kappa_u = \kappa_c c_2 \min\{c_2/(1+\kappa_H), c_1^{\beta}/\kappa_G^{\alpha}\}$. ■

The next theorem provides an upper bound of $\mathcal{O}(\epsilon^{-1})$ iterations for Algorithm 1 when α, β and p satisfy

$$\frac{\alpha+\beta}{p} = 1. \tag{24}$$

The proof of this result is based on the Proof of Theorem 2.5 in [6].

THEOREM 2.4 *Let Conditions* A1–A4, H1 *and* (24) *hold. Also, given* $\epsilon \in (0, \mathrm{e}^{-p})$, *assume that* $D_1 > \epsilon$ *and let* $j_1 \leq +\infty$ *be the first iteration such that* $D_{j_1+1} \leq \epsilon$. *Then, Algorithm* 1 *takes at most* $O(\epsilon^{-1})$ *iterations to achieve* $D_k \leq \epsilon$.

Proof By (8) we have

$$f(x_k) - f(x_{k+1}) \geq \eta_1 (m_k(x_k) - m_k(x_k + s_k)), \quad k \in S. \tag{25}$$

Then, from Lemma 2.3 and (24) it follows that

$$f(x_k) - f(x_{k+1}) \geq \eta_1 \kappa_u D_k^2, \quad k \in S. \tag{26}$$

Thus, by the definition of D_k in (11), we have

$$D_k - D_{k+1} \geq \eta_1 \kappa_u D_k^2, \quad k \in S, \tag{27}$$

and consequently

$$\frac{1}{D_{k+1}} - \frac{1}{D_k} = \frac{D_k - D_{k+1}}{D_k D_{k+1}} \geq \eta_1 \kappa_u \frac{D_k}{D_{k+1}} \geq \eta_1 \kappa_u, \quad k \in S, \tag{28}$$

where in the last inequality we used $D_k \geq D_{k+1}$. Since $D_k = D_{k+1}$ for any $k \notin S$, summing up the above inequalities up to j_1 gives

$$\frac{1}{D_{j_1}} \geq \frac{1}{D_1} + |S_{j_1}| \eta_1 \kappa_u \geq |S_{j_1}| \eta_1 \kappa_u.$$

As $D_{j_1} > \epsilon$, it follows that

$$|S_{j_1}| \leq (\kappa_u \eta_1)^{-1} \epsilon^{-1}. \tag{29}$$

On the other hand, by rule (10) and Lemma 2.2,

$$\begin{aligned} \delta_{k+1} &\leq \gamma_4 \delta_k, \quad \text{if } k \in S_{j_1}, \\ \delta_{k+1} &\leq \gamma_2 \delta_k, \quad \text{if } k \in U_{j_1}, \\ \delta_k &\geq \tau \epsilon^{1/p}, \quad \text{for } k = 1, \ldots, j_1 + 1. \end{aligned}$$

Thus, considering $\nu_k \equiv 1/\delta_k$, we have

$$\alpha_4 \nu_k \leq \nu_{k+1}, \quad \text{if } k \in S_{j_1}, \tag{30}$$

$$\alpha_2 \nu_k \leq \nu_{k+1}, \quad \text{if } k \in U_{j_1}, \tag{31}$$

$$\nu_k \leq \bar{\nu}\epsilon^{-(1/p)}, \quad \text{for } k = 1, \ldots, j_1 + 1, \tag{32}$$

where $\alpha_4 = \gamma_4^{-1} \in (0, 1)$, $\alpha_2 = \gamma_2^{-1} > 1$ and $\bar{\nu} = \tau^{-1}$. From (30) and (31) we deduce inductively

$$\nu_1 \alpha_4^{|S_{j_1}|} \alpha_2^{|U_{j_1}|} \leq \nu_{j_1+1}.$$

Hence, from (32) it follows that

$$\begin{aligned}
&\nu_1 \alpha_4^{|S_{j_1}|} \alpha_2^{|U_{j_1}|} \leq \bar{\nu}\epsilon^{-(1/p)} \\
&\quad \Longrightarrow \alpha_4^{|S_{j_1}|} \alpha_2^{|U_{j_1}|} \leq \frac{\bar{\nu}\epsilon^{-(1/p)}}{\nu_1} \\
&\quad \Longrightarrow \log(\alpha_4^{|S_{j_1}|} \alpha_2^{|U_{j_1}|}) \leq \log\left(\frac{\bar{\nu}\epsilon^{-(1/p)}}{\nu_1}\right) = \log\left(\frac{\bar{\nu}}{\nu_1}\right) + \log(\epsilon^{-(1/p)}) \\
&\quad \Longrightarrow |S_{j_1}| \log(\alpha_4) + |U_{j_1}| \log(\alpha_2) \leq C + \log(\epsilon^{-(1/p)}) \\
&\quad \Longrightarrow |U_{j_1}| \leq \left[-\frac{\log(\alpha_4)}{\log(\alpha_2)}|S_{j_1}| + \frac{C}{\log(\alpha_2)} + \frac{1}{\log(\alpha_2)} \log(\epsilon^{-(1/p)})\right]. \\
&\quad \Longrightarrow |U_{j_1}| \leq \left[-\frac{\log(\alpha_4)}{\log(\alpha_2)}|S_{j_1}| + \frac{C+1}{\log(\alpha_2)} \log(\epsilon^{-(1/p)})\right],
\end{aligned} \tag{33}$$

where $C = \log(\bar{\nu}/\nu_1)$ and the last implication is due to the assumption $\epsilon \leq \mathrm{e}^{-p}$ (which gives $\log(\epsilon^{-(1/p)}) \geq 1$). Finally, as $j_1 = |S_{j_1}| + |U_{j_1}|$ and $\epsilon^{-1} \geq \log(\epsilon^{-(1/p)})$ (due to $p \in [1, 2]$), it follows from (29) and (33) that $j_1 \leq \mathcal{O}(\epsilon^{-1})$. ■

Below we show that the complexity bound given above can be reduced to $\mathcal{O}(\log(\epsilon^{-1}))$ iterations when α, β and p satisfy

$$\frac{\alpha + \beta}{p} = \frac{1}{2}. \tag{34}$$

The proof of this result is based on the Proof of Theorem 2.7 in [6].

THEOREM 2.5 *Let conditions* A1–A4, *H1 and* (34) *hold. Also, given* $\epsilon \in (0, \mathrm{e}^{-p})$, *assume that* $D_1 > \epsilon$ *and let* $j_1 \leq +\infty$ *be the first iteration such that* $D_{j_1+1} \leq \epsilon$. *Then, Algorithm* 1 (*with* $\eta_1 < \kappa_u^{-1}$) *takes at most* $\mathcal{O}(\log(\epsilon^{-1}))$ *iterations to achieve* $D_k \leq \epsilon$.

Proof By (8) we have

$$f(x_k) - f(x_{k+1}) \geq \eta_1(m_k(x_k) - m_k(x_k + s_k)), \quad k \in S. \tag{35}$$

Then, from Lemma 2.3 and (34) it follows that

$$f(x_k) - f(x_{k+1}) \geq \eta_1 \kappa_u D_k. \tag{36}$$

Thus, by the definition of D_k in (11), we have

$$D_k - D_{k+1} \geq \eta_1 \kappa_u D_k, \quad k \in S. \tag{37}$$

Since $D_k = D_{k+1}$ for any $k \notin S$, (37) implies the inequality

$$D_j = f(x_j) - f(x_*) \le (1 - \eta_1 \kappa_u)^{|S_j|} D_1, \quad j \ge 1. \tag{38}$$

As $D_{j_1} > \epsilon$ and $\eta_1 \kappa_u < 1$, it follows from (38) that

$$\begin{aligned}
\epsilon &< (1 - \eta_1 \kappa_u)^{|S_{j_1}|} D_1 \\
&\Longrightarrow \log(\epsilon) < |S_{j_1}| \log(1 - \eta_1 \kappa_u) + \log(D_1) \\
&\Longrightarrow -|S_{j_1}| \log(1 - \eta_1 \kappa_u) < \log\left(\frac{D_1}{\epsilon}\right) \\
&\Longrightarrow |S_{j_1}| < \frac{1}{-\log(1 - \eta_1 \kappa_u)} \left[\log(\epsilon^{-1}) + \log(D_1)\right] \\
&\Longrightarrow |S_{j_1}| < \left[\frac{1}{|\log(1 - \eta_1 \kappa_u)|}(1 + \log(D_1))\right] \log(\epsilon^{-1}),
\end{aligned} \tag{39}$$

where the last implication is due to the assumption $\epsilon \le \mathrm{e}^{-p}$ (which gives $\log(\epsilon^{-1}) \ge \log(\epsilon^{-(1/p)}) \ge 1$). As we saw in the Proof of Theorem 2.4,

$$|U_{j_1}| \le \left[-\frac{\log(\alpha_4)}{\log(\alpha_2)} |S_{j_1}| + \frac{C+1}{\log(\alpha_2)} \log(\epsilon^{-(1/p)})\right]. \tag{40}$$

Since $j_1 = |S_j| + |U_j|$ and $\log(\epsilon^{-1}) \ge \log(\epsilon^{-1/p})$ for $p \in [1, 2]$, it follows from (39) and (40) that $j_1 \le \mathcal{O}(\log(\epsilon^{-1}))$. ■

3. Worst-case complexity analysis: Particular cases

In this section we shall specialize the complexity results above for a class of NSC algorithms. Specifically, we consider the class of NSC algorithms determined by conditions:

A7 The powers α and β satisfy the equality $\alpha + \beta = 1$.
A8 For all k, $\chi_k = \|\nabla f(x_k)\|$, $\phi_k \ge \chi_k$ and $\psi_k \ge \chi_k$.

Note that this class of algorithms is included in the class specified by A5–A6. Our analysis will focus on functions with the following property.

Definition 1 (Nesterov and Polyak [20]) *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called gradient dominated of degree $p \in [1, 2]$ if it attains a global minimum f_* at some x^* and for any $x \in \Omega \subset \mathbb{R}^n$ we have*

$$f(x) - f_* \le \kappa_g \|\nabla f(x)\|^p,$$

where κ_g is a positive constant.

Remark 2 Examples of gradient dominated functions are convex functions (with $p = 1$), and strongly convex functions and sum of squares (with $p = 2$). For details, see Section 4.2 in [20] and Lemmas 2.4 e 2.6 in [6].

For NSC algorithms satisfying A7–A8, it is easy to see that Condition H1 holds if the objective f is a gradient dominated function of degree p. Thus, as a consequence of Theorems 2.4 and 2.5, we have the following complexity result.

THEOREM 3.1 *Suppose that* A1–A4 *and* A7–A8 *hold and let* $\epsilon \in (0, \mathrm{e}^{-p})$. *If f is a gradient dominated function of degree* $p = 1$ *(e.g. convex), then Algorithm* 1 *takes at most* $\mathcal{O}(\epsilon^{-1})$ *iterations to achieve*

$$f(x_k) - f_* \leq \epsilon.$$

Additionally if f is a gradient dominated function of degree $p = 2$ *(e.g. strongly convex) and* $\eta_1 < \kappa_u^{-1}$, *then this worst-case complexity bound is reduced to* $\mathcal{O}(\log(\epsilon^{-1}))$ *iterations.*

Remark 3 Theorem 3.1 generalizes to the NSC framework the complexity results presented by Cartis *et al.* [6] for the basic ARC framework (which is a particular case of NSC with $\phi_k = \psi_k = \chi_k = \|\nabla f(x_k)\|$ and $\alpha = \beta = \frac{1}{2}$).

In order to evaluate the implications of the above result on the complexity of Algorithm 1 for achieving $\|\nabla f(x_k)\| \leq \epsilon$, let us consider first the following property.

LEMMA 3.2 *Let f be continuously differentiable and assume that its gradient* ∇f *is Lipschitz continuous with Lipschitz constant* $\kappa_L \geq 1$. *Also, suppose that f is bounded below by* f_*. *Then, when Algorithm* 1 *is applied to minimizing f, we have*

$$f(x_k) - f_* \geq \frac{1}{2\kappa_L}\|\nabla f(x_k)\|^2, \quad \forall k.$$

Then, for any $\epsilon > 0$, $\|\nabla f(x_k)\| \leq \epsilon$ *holds whenever*

$$f(x_k) - f_* \leq \frac{\epsilon^2}{2\kappa_L}.$$

Proof See Lemma 2.8 in [6]. ■

Now, from Theorem 3.1 and Lemma 3.2 we have the following result.

THEOREM 3.3 *Suppose that* A1–A4, A7–A8 *and the assumptions of Lemma* 3.2 *hold. Then, if f is a gradient dominated function of degree* $p = 1$, *Algorithm 1 will take at most* $O(\epsilon^{-2})$ *iterations to ensure* $\|\nabla f(x_k)\| \leq \epsilon$. *Additionally, if f is a gradient dominated function of degree* $p = 2$ *and* $\eta_1 < \kappa_u^{-1}$, *then this worst-case complexity bound is reduced to* $O(\log(\epsilon^{-1}))$ *iterations.*

Table 1 summarizes the complexity results described in Theorems 1.1 and 3.3 according to the choice of the powers $\alpha \in (0, 1]$ and $\beta \in [0, 1]$.

Finally, to justify our analysis, we shall see important examples of Algorithm 1 for which conditions A7–A8 are satisfied. As pointed out by Toint [22], under mild conditions,[2] Algorithm 1 covers the following algorithms:

- the classical trust-region algorithm [9, 21], specified by:

$$m_k(x_k + s) \equiv f(x_k) + \nabla f(x_k)^{\mathrm{T}} s + \tfrac{1}{2} s^{\mathrm{T}} H_k s,$$
$$\omega(x) = 1, \quad \phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|,$$
$$\delta_k = \Delta_k, \quad \alpha = 1, \quad \beta = 0;$$

Table 1. Complexity of NSC on gradient dominated objectives of degree p.

p	$2\alpha + \beta \geq 1$ and $\alpha + \beta < 1$	$\alpha + \beta = 1$	Otherwise
1	$\mathcal{O}(\epsilon^{-2})$	$\mathcal{O}(\epsilon^{-2})$	$\mathcal{O}(\epsilon^{-(2+\beta)})$
2	$\mathcal{O}(\epsilon^{-2})$	$\mathcal{O}(\log(\epsilon^{-1}))$	$\mathcal{O}(\epsilon^{-(2+\beta)})$

- the basic ARC algorithm of Cartis *et al.* [4], specified by:

$$m_k(x_k + s) \equiv f(x_k) + \nabla f(x_k)^{\mathrm{T}} s + \frac{1}{2} s^{\mathrm{T}} H_k s + \frac{1}{3}\sigma_k \|s\|^3,$$
$$\omega(x) = 1, \quad \phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|,$$
$$\delta_k = \frac{1}{\sigma_k}, \quad \alpha = \frac{1}{2}, \quad \beta = \frac{1}{2};$$

- the nonlinear trust region method suggested by Toint [22], specified by:

$$m_k(x_k + s) \equiv f(x_k) + \nabla f(x_k)^{\mathrm{T}} s + \tfrac{1}{2} s^{\mathrm{T}} H_k s,$$
$$\omega(x) = 1, \quad \phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|,$$
$$\alpha = \tfrac{1}{2}, \quad \beta = \tfrac{1}{2}.$$

Clearly, all these algorithms satisfy A7–A8. Furthermore, we claim that the conic trust-region method in [17] can also be viewed as particular cases of Algorithm 1 satisfying A7 and A8 with the choices:

$$m_k(x_k + s) \equiv f(x_k) + \frac{\nabla f(x_k)^{\mathrm{T}} s}{1 - h_k^{\mathrm{T}} s} + \frac{s^{\mathrm{T}} H_k s}{2(1 - h_k^{\mathrm{T}} s)^2}, \tag{41}$$
$$\omega(x) = 1, \quad \phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|, \tag{42}$$
$$\delta_k = \Delta_k, \quad \alpha = 1, \quad \beta = 0, \tag{43}$$

where the *horizontal vector* $h_k \in \mathbb{R}^n$ is chosen such that m_k and ∇m_k satisfy suitable interpolation conditions (see, e.g. [11] and references therein). In fact, regarding the conic trust-region method specified by (41)–(42), consider the following assumptions:

C1 The objective function $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.
C2 There exists a compact and convex set $\Omega \subset \mathbb{R}^n$ such that x_k and $x_k + s_k$ belong to Ω for all k.
C3 There exists a constant $\kappa_H > 0$ such that, $\|H_k\| \leq \kappa_H$ for all k.
C4 There exists a constant $\kappa_h > 0$ such that $\|h_k\| \leq \kappa_h$ for all k.

Let us now check all the conditions that define Algorithm 1. Conditions A1 and A2 follow directly from (42). Condition A3 follows from C1 and C2. Clearly, by (41) we have $m_k(x_k) = f(x_k)$. Moreover, it follows from C1–C4 and Lemma 3.4 in [26] that

$$f(x_k + s) - m_k(x_k + s) \leq \kappa_m \|s\|^2 \quad \forall s \in \mathbb{R}^n,$$

for some constant $\kappa_m > 0$. Thus, the conic model $m_k(x_k + s)$ given in (41) satisfies (5). As in the conic trust-region method, s_k is computed by minimizing $m_k(x_k + s)$ restricted to the set $\{s \in \mathbb{R}^n \mid \|s\| \leq \Delta_k, \ |1 - h_k^{\mathrm{T}} s| \geq \epsilon_0\}$, for some $\epsilon_0 \in (0, 1)$, it follows that (6) is automatically satisfied with $\kappa_s = 1$ and $\kappa_\delta = +\infty$. If we impose $\Delta_k \leq \bar{\Delta}$ for all k, it follows from Theorem 2.5 in [17] that, for all k,

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_c \|\nabla f(x_k)\| \min\left\{\frac{\|\nabla f(x_k)\|}{1 + \|H_k\|}, \Delta_k\right\},$$

for some constant $\kappa_c > 0$, that is, (7) is satisfied. Therefore, the conic trust-region method specified by (41)–(42) is a particular case of Algorithm 1 satisfying A7 and A8.

Thus, in the light of Theorem 3.1, all the algorithms mentioned above (seen as particular cases of Algorithm 1) take at most $\mathcal{O}(\epsilon^{-1})$ and $\mathcal{O}(\log(\epsilon^{-1}))$ iterations to generate x_k such that $f(x_k) - f_* \leq \epsilon$, when applied to a convex and strongly convex objective f, respectively. Furthermore, if the gradient function ∇f is Lipschitz continuous, Theorem 3.3 provides upper bounds of $\mathcal{O}(\epsilon^{-2})$ and $\mathcal{O}(\log(\epsilon^{-1}))$ iterations for achieving $\|\nabla f(x_k)\| \leq \epsilon$, when f is convex and strongly convex, respectively.

4. Conclusion

In this paper, we investigate the worst-case complexity of the NSC framework on convex and strongly convex objectives. By generalizing the analysis of Cartis *et al.* [6] for the basic ARC algorithm, we proved that if $\alpha + \beta = 1$, $\chi_k = \|\nabla f(x_k)\|$ and $\phi_k, \psi_k \geq \chi_k$, then the NSC algorithm can take at most $\mathcal{O}(\epsilon^{-1})$ or $\mathcal{O}(\log(\epsilon^{-1}))$ iterations to achieve $f(x_k) - f_* \leq \epsilon$, where f_* is the global minimum of a convex or strongly convex objective f, respectively. Unsurprinsingly, these bounds match in order those obtained by Cartis *et al.* [6] for the basic ARC algorithm, and by Nesterov [19] for the steepest descent method, on the same classes of objectives. A natural topic for future research is to investigate whether these improved complexity bounds can be established for NSC algorithms in which $\alpha + \beta \neq 1$.

In [15], we have extended the NSC framework to some algorithms for composite non-smooth optimization (NSO) problems of the form

$$\min_{x\in\mathbb{R}^n} f(x) \equiv g(x) + h(c(x)),$$

where $h : \mathbb{R}^m \to \mathbb{R}$ is convex but may be non-smooth, and $g : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. We also have generalized the NSC framework to include algorithms for unconstrained multiobjective optimization (MOO) problems of the form

$$\min_{x\in\mathbb{R}^n} f(x) \equiv (f_1(x), \ldots, f_m(x))^{\mathrm{T}},$$

where each $f_i : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. It is worth to mention that all the results in Section 2 (namely, Lemmas 2.1–2.3 and Theorems 2.4–2.5) remain true if we consider these extensions of the NSC algorithm.[3] Thus, as another topic for future research, it would be interesting to search examples of NSC algorithms for these problems and identify classes of objectives for which Condition H1 is satisfied. Note that, this would allow us to obtain improved complexity bounds for composite NSO problems and unconstrained MOO problems.

Acknowledgements

The authors are grateful to the anonymous referees, whose comments helped to greatly improve this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

G.N. Grapiglia was supported by the Coordination for the Improvement of Higher Education Personnel (CAPES), Brazil. J. Yuan was partially supported by the Coordination for the Improvement of Higher Education Personnel (CAPES) and by the National Council for Scientific and Technological Development (CNPq), Brazil. Y. Yuan was partially supported by the Natural Science Foundation of China (NSFC), China.

Notes

1. In the bound for strongly convex functions, C is a problem-dependent constant. For details, see [[6, p. 214], [20, pp. 203–204]].
2. For example, if f is twice continuously differentiable, the norms of H_k and $\nabla^2 f(x)$ are bounded from above, and the sequence $\{x_k\}$ is contained in a compact set $\Omega \subset \mathbb{R}^n$
3. For NSO and MOO problems, $\|\nabla f(x)\|$ in A1 and A2 must be replaced by the appropriated criticality measures. Furthermore, in the case of MOO problems, $f(x)$ must be replaced by $\Phi(x) = \max_{i=1,\ldots,m}\{f_i(x)\}$. For details, see Section 4 in [15].

ORCID

G.N. Grapiglia http://orcid.org/0000-0003-3284-3371

References

[1] S. Bellavia, C. Cartis, N.I.M. Gould, B. Morini, and Ph.L. Toint, *Convergence of a regularized euclidean residual algorithm for nonlinear least-squares*, SIAM J. Numer. Anal. 48 (2010), pp. 1–29.
[2] C. Cartis, N.I.M. Gould, and Ph.L. Toint, *On the complexity of steepest descent, Newton's and regularized Newton's method for nonconvex unconstrained optimization problems*, SIAM J. Optim. 20 (2010), pp. 2833–2852.
[3] C. Cartis, N.I.M. Gould, and Ph.L. Toint, *Adaptive cubicregularisation methods for unconstrained optimization. Part I: Motivation, convergence and numerical results*, Math. Program. 127 (2011), pp. 245–295.
[4] C. Cartis, N.I.M. Gould, and Ph.L. Toint, *Adaptive cubic regularisation methods for unconstrained optimization. Part II: Worst-casefunction- and derivative-evaluation complexity*, Math. Program. 130 (2011), pp. 295–319.
[5] C. Cartis, N.I.M. Gould, and Ph.L. Toint, *How much patience do you have? A worst-case perspective on smooth nonconvex optimization*, Optimization 88 (2012), pp. 1–10.
[6] C. Cartis, N.I.M. Gould, and Ph.L. Toint, *Evaluation complexity of adaptive cubic regularization methods for convex unconstrained optimization*, Optim. Methods Softw. 27 (2012), pp. 197–219.
[7] C. Cartis, N.I.M. Gould, and Ph.L. Toint, *On the oracle complexity of first-order and derivative-free algorithms for smooth nonconvex minimization*, SIAM J. Optim. 22 (2012), pp. 66–86.
[8] C. Cartis, Ph.R. Sampaio, and Ph.L. Toint, *Worst-case evaluation complexity of non-monotone gradient-related algorithms for unconstrained optimization*, Optimization 64 (2015), pp. 1349–1361.
[9] A.R. Conn, N.I.M. Gould, and Ph.L. Toint, *Trust-Region Methods*, SIAM, Piladelphia, PA, 2000.
[10] F.E. Curtis, D.P. Robinson, and M. Samadi, *A trust region algorithm with a worst-case iteration complexity of $O(\epsilon^{-3/2})$ for nonconvex optimization*, Tech. Rep. 14T-009, COR@L Laboratory, Department of ISE, Lehigh University, 2014.
[11] S. Di and W. Sun, *A trust region method for conic model to solve unconstrained optimizations*, Optim. Methods Softw. 6 (1996), pp. 237–263.
[12] M. Dodangeh and L.N. Vicente, *Worst case complexity of direct search under convexity*, Math. Program. doi:10.1007/s10107-014-0847-0.
[13] J. Fan, *Convergence rate of the trust region method for nonlinear equations under local error bound condition*, Comput. Optim. Appl. 34 (2006), pp. 215–227.
[14] J. Fan and Y. Yuan, *A new trust region algorithm with trust region radius converging to zero*, in *Proceeding of the 5th International Conference on Optimization: Techiniques and Applications*, D. Li, ed., ICOTA, Hong Kong, 2001, pp. 786–794.
[15] G.N. Grapiglia, J. Yuan, and Y.-X. Yuan, *On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization*, Math. Program. 152 (2015), pp. 491–520.
[16] S. Gratton, A. Sartenaer, and Ph.L. Toint, *Recursive trust-region methods for multiscale nonlinear optimization*, SIAM J. Optim. 19 (2008), pp. 414–444.
[17] X. Lu and Q. Ni, *A quasi-Newtontrust region method with a new conic model for the unconstrained optimization*, Appl. Math. Comput. 204 (2008), pp. 373–384.
[18] Yu. Nesterov, *Nonlinear Stepsize Control, Trust Regions and Regularizations for Unconstrained Optimization*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.
[19] Yu. Nesterov, *Modified Gauss–Newton scheme with worst case guarantees for global performance*, Optim. Methods Softw. 22 (2007), pp. 469–483.
[20] Yu. Nesterov and B.T. Polyak, *Cubic regularization of Newton method and its global performance*, Math. Program. 108 (2006), pp. 177–205.
[21] M.J.D. Powell, *Convergence properties of a class of minimization algorithms*, in *Nonlinear Programming 2*, O.L. Mangasarian, R.R. Meyer, and S.M. Robinson, eds., Academic Press, New York, 1975, pp. 1–27.
[22] Ph.L. Toint, *Nonlinear stepsize control, trust regions and regularizations for unconstrained optimization*, Optim. Methods Softw. 28 (2013), pp. 82–95.
[23] K. Ueda and N. Yamashita, *On a global complexity bound of the Levenberg–Marquardt method*, J. Optim. Theory Appl. 147 (2010), pp. 443–453.

[24] L.N. Vicente, *Worst case complexity of direct search*, EURO J. Comput. Optim. 1 (2013), pp. 143–153.

[25] J.-L . Zhang and Y. Wang, *A new trust region method for nonlinear equations*, Math. Methods Oper. Res. 58 (2003), pp. 283–298.

[26] L. Zhao, *Non-monotone adaptive filter conic trust region method for unconstrained optimization*, J. Math. Comput. Sci. 2 (2012), pp. 1874–1893.