A note on the worst-case complexity of nonlinear stepsize control methods for convex smooth unconstrained optimization

R. Garmanjani[a]

[a]Department of Mathematics, FCT-UNL-CMA, Campus de Caparica, 2829-516 Caparica, Portugal

ARTICLE HISTORY
Compiled May 6, 2020

ABSTRACT
In this paper, we analyze the worst-case complexity (WCC) of Nonlinear Stepsize Control (NSC) algorithms for solving convex smooth unconstrained optimization problems. We show that, to drive the norm of the gradient below some given positive ϵ, such methods take at most $\mathcal{O}(\epsilon^{-1})$ iterations, which shows that the complexity bound for these methods is in parity with that of gradient descent methods for the same class of problems. As NSC algorithm is a generalization of several methods such as trust-region and adaptive cubic with regularization methods, such bound holds automatically for these methods as well.

KEYWORDS
nonlinear stepsize control algorithms, worst-case complexity, convex smooth unconstrained optimization,

AMS CLASSIFICATION
90C25, 90C30, 65K05, 49M37

1. Introduction

In this paper, we consider unconstrained optimization problems of the form

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex continuously differentiable function. For this class of optimization problems, we are interested in evaluating the worst-case complexity (WCC) of the Nonlinear Stepsize Control (NSC) framework for driving the norm of the gradient of the objective function below some given positive threshold. In other words, we are interested in measuring the effort needed for finding a point $\bar{x} \in \mathbb{R}^n$ such that $\|\nabla f(\bar{x})\| \leq \epsilon$.

Over the past couple of decades, there have been some interests in measuring the WCC (in terms of the number of iterations, functions or gradient evaluations) of various optimization algorithms for bringing the norm of the gradient of the objective function (as a measure of criticality) below some given positive threshold. In the

CONTACT R. Garmanjani Email: nima@mat.uc.pt

following we mostly review those works that measure the effort needed for finding a point $\bar{x} \in \mathbb{R}^n$ such that $\|\nabla f(\bar{x})\| \leq \epsilon$. There are also numerous research studies in the literature that establish complexity bounds for finding a point $\bar{x} \in \mathbb{R}^n$ such that $f(\bar{x}) - f_* \leq \epsilon$, where f_* is the optimal function value.

In the context of nonconvex smooth unconstrained optimization, a WCC bound of $\mathcal{O}(\epsilon^{-2})$ has been derived by Nesterov for gradient methods [16]; Cartis et al. [3] for basic adaptive regularization with cubic (ARC); Vicente for direct-search methods [20]; Gratton et al. [12] for direct-search methods based on probabilistic descent; Konecnỳ and Richtárik for restricted direct-search methods where no stepsize increases are allowed [14]. An improved WCC bound of $\mathcal{O}(\epsilon^{-1.5})$ has been derived by Nesterov and Polyak for Cubic Newton Method [18]; Cartis et al. for ARC [3], and a derivative-free version of it [5]; Curtis et al. for a novel trust-region algorithm [7]; Martínez and Raydan for Cubic-regularization counterpart of a variable norm trust-region method [15]; Birgin and Martínez for quadratic regularization with a cubic descent condition [2].

When the objective function is convex, a WCC bound of $\mathcal{O}(\epsilon^{-1})$ has been established by Nesterov [16, 17] for gradient methods; Dodangeh and Vicente [8] for direct-search methods. Nesterov has derived a bound of $\mathcal{O}(\epsilon^{-0.5} \log \epsilon^{-1})$ for fast gradient methods [17].

With regard to the WCC of trust-region methods for nonconvex smooth unconstrained optimization problems, Gratton et al. [13] have shown a WCC bound of $\mathcal{O}(\epsilon^{-2})$. Garmanjani et al. [9] have obtained a bound of $\mathcal{O}(n^2\epsilon^{-2})$ for derivative-free trust-region methods.

Some research works have developed the WCC analysis of the class of nonconvex nonsmooth optimization problems using a smoothing approach. A WCC bound of $\mathcal{O}(\epsilon^{-3} \log \epsilon^{-1})$ has been developed by Garmanjani and Vicente for direct search methods of directional type [10]. Bian and Chen [1] have proposed a smoothing quadratic regularization algorithm and have derived a WCC of $\mathcal{O}(\epsilon^{-2})$ and $\mathcal{O}(\epsilon^{-3})$ for reaching an ϵ scaled stationary and ϵ-Clarke stationary point, respectively.

For problem (1), Cartis et al. [4], have established a WCC bound of $\mathcal{O}(\epsilon^{-2})$ (resp. $\mathcal{O}(\epsilon^{-1})$) for finding a point $\bar{x} \in \mathbb{R}^n$ such that $\|\nabla f(\bar{x})\| \leq \epsilon$ (resp. $f(\bar{x}) - f_* \leq \epsilon$). Similar bounds have been derived by Grapiglia et al. [11] for Nonlinear Stepsize Control (NSC) algorithms, which is a framework proposed by Toint [19] that generalizes trust-region and regularization methods.

When the WCC analysis of deriving the norm of the gradient below some given positive threshold is of interest, to the best of our knowledge, no bound of $\mathcal{O}(\epsilon^{-1})$ has been derived in the literature for NSC algorithm, trust-region methods, and ARC algorithm when applied to the class of convex smooth unconstrained optimization problems. Therefore, inspired by the above-mentioned studies, in particular [4, 16, 17], we analyze the WCC of NSC methods for such class of optimization problems. We show that the WCC bound of the NSC algorithm is of $\mathcal{O}(\epsilon^{-1})$ for minimizing convex objective functions. As a byproduct, since NCS framework is a generalization of several methods such as trust-region methods and adaptive cubic with regularization [4], a similar bound holds for those methods as well.

The rest of the paper is organized as follows. In Section 2 we present the NSC algorithm and some preliminary results. The WCC analysis of the NSC is established in Section 3. The paper ends in Section 4 with some conclusions.

Notation. In this paper, for simplicity we assume all the norms are Euclidean norm. Notation $|\cdot|$ represents the cardinality of a set. By $B = \mathcal{O}(A)$ we mean $B \leq MA$, where constant $M > 0$ does not depend on the iteration counter. We represent the global solution of (1) with x_*, and the corresponding optimal function value with f_*.

2. The NSC framework and some preliminary results

In this section, we will use the NSC framework presented in [11], which was based originally in [19]. Similarly to [11], for compact presentation of the NSC framework, we will use auxiliary functions $\phi, \psi, \chi : \mathbb{R}^n \to \mathbb{R}$, which satisfy the following assumptions:

Assumption 2.1. *Let ϕ, ψ, χ be continuous non-negative functions, such that*

$$\min\{\phi(x), \phi(x), \chi(x)\} \;=\; 0 \Longrightarrow \|\nabla f(x)\| \;=\; 0.$$

Assumption 2.2. *There exists $\kappa_\chi > 0$ such that*

$$\chi(x) \;\leq\; \kappa_\chi \quad \textit{for all } x.$$

Hereafter, for the sake of conciseness we use the following notations:

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

Algorithm 2.1 (NSC Algorithm [11, 19]).
Step 0 Given $x_0 \in \mathbb{R}^n$, $H_0 \in \mathbb{R}^{n\times n}$, $\delta_1 > 0$, $0 < \gamma_1 < \gamma_2 < \gamma_3 < 1 < \gamma_4$, and $0 < \eta_1 \leq \eta_2 < 1$, set $k = 0$.
Step 1 Choose a model $m_k(x_k + s) : \mathbb{R}^n \to \mathbb{R}$ such that

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

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

$$\|s_k\| \leq \kappa_s \Delta(\delta_k, \chi_k), \quad \text{whenever} \quad \delta_k \leq \kappa_\delta \chi_k, \tag{3}$$

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

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_c \psi_k \min\left\{\frac{\phi_k}{1 + \|H_k\|}, \Delta(\delta_k, \chi_k)\right\}, \tag{4}$$

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

$$\rho_k \;=\; \frac{f(x_k) - f(x_k + s_k)}{m(x_k) - m(x_k + s_k)}. \tag{5}$$

If $\rho_k \geq \eta_1$ then $x_{k+1} = x_k + s_k$; else, $x_{k+1} = x_k$, and update δ_{k+1} as follows

$$\delta_{k+1} \in \begin{cases} [\gamma_1\delta_k, \gamma_2\delta_k) & \text{if } \rho_k < \eta_1, \\ [\gamma_2\delta_k, \gamma_3\delta_k] & \text{if } \rho_k \in [\eta_1, \eta_2), \\ [\delta_k, \gamma_4\delta_k] & \text{if } \rho_k \geq \eta_2. \end{cases} \tag{6}$$

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

We let $\mathcal{S}$ be the set of all successful (i.e. $\rho_k \geq \eta_1$ in (6)) or very successful (i.e. $\rho_k \geq \eta_2$ in (6)) iterations:

$$\mathcal{S} \ = \ \{k \geq 0 : \text{iteration } k \text{ is successful or very successful}\}, \tag{7}$$

and $\mathcal{S}_\ell$ be the set of all such iterations up to iteration ℓ, i.e.,

$$\mathcal{S}_\ell \ = \ \{k \leq \ell : k \in \mathcal{S}\}. \tag{8}$$

We also denote the set of all unsuccessful iterations (i.e. $\rho_k < \eta_1$ in (6)) with $\mathcal{U}$ and the ones up to iteration ℓ with $\mathcal{U}_\ell$. Henceforth, we will call all successful and very successful iterations as successful iterations.

The following assumptions will be required in the continuation.

Assumption 2.3. *The objective function f is continuously differentiable and bounded from below on $\mathbb{R}^n$.*

Assumption 2.4. *The gradient of the function f is Lipschitz continuous with constant L_g.*

Assumption 2.5. *There is a constant $\kappa_H > 0$ such that $\|H_k\| \leq \kappa_H$, for all $k \geq 0$.*

Assumption 2.6. *The powers α and β satisfy $\alpha + \beta = 1$.*

Assumption 2.7. *For all k, $\phi_k \geq \chi_k$ and $\psi_k \geq \chi_k$.*

The following lemma, which is a slightly modified version of Lemma 2.2 in [11], will be central in establishing the WCC results in this manuscript.

Lemma 2.1. *Let Assumptions 2.1–2.5 hold. Also, let $\epsilon \in (0,1]$ such that $\min\left\{\chi_k^\alpha, \chi_k^{-\beta}\phi_k, \chi_k^{-\beta}\psi_k\right\}^{1/\alpha} > \epsilon$, for all $k = 0, \cdots, j$, where $j \leq +\infty$. Then,*

$$\delta_k \geq C_1 \epsilon, \tag{9}$$

where $C_1 = \gamma_1 / \kappa_{HB}^{1/\alpha}$ and $\kappa_{HB} = \max\left\{\frac{\kappa_m \kappa_s^2}{(1-\eta_2)\kappa_c}, \frac{1+\kappa_H}{\kappa_\delta^\alpha}, 1 + \kappa_H\right\}$.

Remark 1. In [11], the inequality (9) is of the form

$$\delta_k \geq \min\{\delta_0, C_1\}\epsilon. \tag{10}$$

Here, without loss of generality, we assume that δ_0 chosen in a way that $\delta_0 \geq C_1$. Thus, inequality (10) becomes of the form (9).

Before moving to the next section, we state a result that shows that the number of unsuccessful iterations is a function of the number of successful iterations. Hence, in order to count the total number of iterations, it suffices to count the number of successful iterations. This is a typical approach that is taken for establishing the WCC analysis of those class of methods for which iterations are of two types or more. The proof is similar to the corresponding results in [3, 13, 20], but we bring it for the sake of completeness.

Theorem 2.2. *Given any* $\epsilon \in (0,1)$. *Let Assumptions 2.1–2.5 hold. Let* $\min\{\chi_k, \chi_k^{(1-\beta)/\alpha}\} > \epsilon$ *for* $k = 0, 1, \cdots, \ell$. *Then the set of all unsuccessful iterations up to iteration* ℓ *satisfies*

$$|\mathcal{U}_\ell| \;\leq\; -\log_{\gamma_2}(\gamma_3)|\mathcal{S}_\ell| + \log_{\gamma_2}\frac{C_1\epsilon}{\Delta_0},$$

for having $\min\{\chi_{\ell+1}, \chi_{\ell+1}^{(1-\beta)/\alpha}\} \leq \epsilon$.

Proof. If $k \in \mathcal{S}_\ell$ then $\delta_{k+1} \leq \gamma_4\delta_k$, and if $k \in \mathcal{U}_\ell$ then $\delta_{k+1} \leq \gamma_2\delta_k$. Thus by induction we have

$$\delta_\ell \;\leq\; \gamma_2^{|\mathcal{U}_\ell|}\gamma_4^{|\mathcal{S}_\ell|}\Delta_0.$$

Hence, by taking logarithm and considering that $\log\gamma_2 < 0$, we have

$$|\mathcal{U}_\ell| \;\leq\; -\log_{\gamma_2}(\gamma_4)|\mathcal{S}_\ell| + \log_{\gamma_2}\frac{\delta_\ell}{\delta_0}.$$

Now, since $\delta_\ell \geq C_1\epsilon$ in view of Lemma 2.1, we thus derive the desired upper bound on the number of unsuccessful iterations. □

3. WCC of NSC framework for minimizing convex functions

In this section we analyze the WCC of the NSC framework for minimization of convex smooth unconstrained optimization problems.

Similarly to [11], hereafter we will assume that the following condition holds for the class of problems under consideration.

Assumption 3.1. *There exists* $\kappa_d > 0$ *such that*

$$D_k \;\equiv\; f(x_\ell) - f_* \;\leq\; \kappa_d \min\{\chi_k^\alpha, \chi_k^{-\beta}\phi_k, \chi_k^{-\beta}\psi_k\}^{1/\alpha},$$

for all k.

Assumption 3.2. *We assume that function* f *is convex and there exists* $R \geq 1$ *such that* $\{x|f(x) \leq f(x_0)\} \subseteq \{x| \|x - x_*\| \leq R\}$

The following lemma shows that, under Assumptions 2.3 and 3.2, and by setting $\phi(x) = \psi(x) = \chi(x) = \nabla f(x)$, Assumption 3.1 hold.

Lemma 3.1. *[4, Lemma 2.4] Let Assumptions 2.3 and 3.2 hold. Then, for any iteration* x_k *produced by Algorithm 2.1, we have*

$$f(x_k) - f(x_*) \;\leq\; R\|\nabla f(x_k)\|.$$

In [11], the authors have shown (cf. Theorem 3.3 in [11]) that the NSC framework takes at most $\mathcal{O}(\epsilon^{-2})$ iterations for bringing the norm of the gradient below some $\epsilon > 0$. In this section, we will show a bound of $\mathcal{O}(\epsilon^{-1})$ could be established for NSC framework, which is in parity with that of gradient descent methods for the problem (1).

Since the NSC framework is a generalization of several methods, as a byproduct similar upper bound is automatically established for those methods as well. We will need the following key lemma, which could be proved similarly to the Theorem 2.4 in [11].

Lemma 3.2. *Let Assumptions 2.1–2.6, and 3.1 hold. Then, by applying Algorithm 2.1, we have*

$$D_\ell \;\leq\; \frac{1}{|\mathcal{S}_\ell|\eta_1 C_2}, \qquad \forall \ell \geq 0,$$

where

$$C_2 \;=\; \frac{1}{2}\kappa_c c_2 \min\left\{\frac{c_2}{1+\kappa_H}, \frac{c_1^\beta}{\kappa_G^\alpha}\right\},$$

and

$$\kappa_G \;=\; \max\left\{\frac{D_0}{\delta_0}, \gamma_1^{-1}\kappa_d \kappa_{HB}^{1/\alpha}\right\}.$$

Now, we are ready to present the main result of this paper. In the continuation, without loss of generality, we assume that $N \geq 1$ can be chosen such that $|\mathcal{S}_N| \geq 2$.

Theorem 3.3. *Let Assumptions 2.1–2.7, and 3.1 hold. Then, for any $\epsilon \in (0,1)$, the NSC algorithm takes at most $\mathcal{O}(\epsilon^{-\max\{1,\frac{\alpha}{1-\beta}\}})$ successful iterations to find a point $\bar{x} \in \mathbb{R}^n$ such that $\min\{\chi(\bar{x}), \chi(\bar{x})^{(1-\beta)/\alpha}\} \leq \epsilon$.*

Proof. Let $0 \leq \ell < N$ be chosen in a way that $|\mathcal{S}_\ell| \geq 1$, and $|\mathcal{S}_N| \geq 2|\mathcal{S}_\ell|$. In view of Lemma 3.2, we have

$$f(x_\ell) - f_* \;=\; D_\ell \;\leq\; \frac{1}{|\mathcal{S}_\ell|\eta_1 C_2}, \qquad \forall \ell \geq 0.$$

Hence, in view of (4) and (5), we have

$$\begin{aligned}
\frac{1}{|\mathcal{S}_\ell|\eta_1 C_2} \;&\geq\; f(x_\ell) - f_* \\
&\geq\; f(x_{N+1}) - f_* + f(x_{\ell+1}) - f(x_{N+1}) \\
&=\; f(x_{N+1}) - f_* + \sum_{\substack{k=\ell+1 \\ k\in\mathcal{S}}}^{N} f(x_k) - f(x_{k+1}) \\
&\geq\; f(x_{N+1}) - f_* + \frac{\eta_1\kappa_c}{2}\sum_{\substack{k=\ell+1 \\ k\in\mathcal{S}}}^{N} \|\psi_k\| \min\left\{\frac{\|\phi_k\|}{1+\kappa_H}, \delta_k^\alpha \chi_k^\beta\right\} \\
&\geq\; \frac{\eta_1}{2}|\mathcal{S}_\ell| \min_{\substack{0\leq k\leq N \\ k\in\mathcal{S}}} \left\{\|\psi_k\| \min\left\{\frac{\|\phi_k\|}{1+\kappa_H}, \delta_k^\alpha \chi_k^\beta\right\}\right\}.
\end{aligned}$$

Thus, there exists $1 \le k_0 \le N$ such that

$$\frac{2}{|\mathcal{S}_\ell|^2 \eta_1^2 C_2} \;\geq\; \|\psi_{k_0}\| \min\left\{ \frac{\|\phi_{k_0}\|}{1+\kappa_H}, \delta_{k_0}^{\alpha} \chi_{k_0}^{\beta} \right\}, \tag{11}$$

which, in view of Assumption 2.7, leads to

$$\frac{2}{|\mathcal{S}_\ell|^2 \eta_1^2 C_2} \;\geq\; \chi_{k_0} \min\left\{ \frac{\chi_{k_0}}{1+\kappa_H}, \delta_{k_0}^{\alpha} \chi_{k_0}^{\beta} \right\}. \tag{12}$$

Now, we consider two cases:

a) If $\frac{\chi_{k_0}}{1+\kappa_H} \le \delta_{k_0}^{\alpha}\chi_{k_0}^{\beta}$, then (12) leads to

$$\frac{2(1+\kappa_H)}{|\mathcal{S}_\ell|^2 \eta_1^2 C_2} \;\geq\; \chi_{k_0}^2. \tag{13}$$

Hence, in order to drive $\chi_{k_0} \le \epsilon$, for some $0 \le k_0 \le N$, Algorithm 2.1 takes at most

$$|\mathcal{S}_N| = 2|\mathcal{S}_\ell| = 2\left\lceil \frac{\sqrt{2(1+\kappa_H)}}{\eta_1\sqrt{C_2}}\epsilon^{-1} \right\rceil$$

successful iterations.

b) If $\frac{\chi_{k_0}}{1+\kappa_H} \ge \delta_{k_0}^{\alpha}\chi_{k_0}^{\beta}$, then we have $\chi_{k_0} \ge \left((1+\kappa_H)\delta_{k_0}^{\alpha}\right)^{\frac{1}{1-\beta}}$, which in view of (12) we obtain

$$\begin{aligned}
\frac{2}{|\mathcal{S}_N|^2 \eta_1^2 C_2} \;&\geq\; \delta_{k_0}^{\alpha}\chi_{k_0}^{1+\beta} \\
&\geq\; \delta_{k_0}^{\alpha}\left((1+\kappa_H)\delta_{k_0}^{\alpha}\right)^{\frac{1+\beta}{1-\beta}} \\
&\geq\; (1+\kappa_H)^{\frac{1+\beta}{1-\beta}}\delta_{k_0}^{\frac{2\alpha}{1-\beta}}.
\end{aligned}$$

Thus, we have

$$\left(\frac{2}{|\mathcal{S}_\ell|^2 \eta_1^2 C_2 (1+\kappa_H)^{\frac{1+\beta}{1-\beta}}} \right)^{\frac{1-\beta}{2\alpha}} \;\geq\; \delta_{k_0}. \tag{14}$$

Hence, in order to obtain $\delta_{k_0} < C_1\epsilon$, for some $0 \le k_0 \le N$, Algorithm 2.1 takes at most

$$|\mathcal{S}_N| = 2|\mathcal{S}_\ell| = 2\left\lceil \frac{\sqrt{2}}{C_1^{\frac{\alpha}{1-\beta}}\eta_1\sqrt{C_2(1+\kappa_H)^{\frac{1+\beta}{1-\beta}}}}\epsilon^{-\frac{\alpha}{1-\beta}} \right\rceil$$

successful iterations. On the other hand, Lemma 2.1 implies that if $\delta_{k_0} < C_1\epsilon$ then, $\min\left\{\chi_{k_1}^{\alpha}, \chi_{k_1}^{-\beta}\phi_{k_1}, \chi_{k_1}^{-\beta}\psi_{k_1}\right\}^{1/\alpha} \le \epsilon$ for some $0 \le k_1 \le N$, which in view of Assumption 2.7 leads to $\min\left\{\chi_{k_1}, \chi_{k_1}^{\alpha/(1-\beta)}\right\} \le \epsilon$ for some $0 \le k_1 \le N$.

Therefore, by combining both cases, the number of successful iterations for finding a point $\bar{x} \in \mathbb{R}^n$ such that $\min\left\{\chi(\bar{x}), \chi(\bar{x})^{\alpha/(1-\beta)}\right\} \le \epsilon$ is of $\mathcal{O}(\epsilon^{-\max\{1, \frac{\alpha}{1-\beta}\}})$, and the proof is completed. □

The following corollary is readily resulted from Theorem 3.3.

Corollary 3.4. *Let Assumptions 2.1–2.6, and 3.2 hold. Let $\epsilon \in (0,1)$ and $\phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|$. Then, the NSC framework takes at most $\mathcal{O}(\epsilon^{-1})$ iterations for driving the norm of gradient below ϵ.*

Proof. In view of Lemma 3.1, Assumption 3.1 holds. Now, by applying Theorem 3.3, the proof is completed. □

As it has been mentioned in [19] (see also [11]), the following methods are resulted from NSC framework:

- the classical trust-region method [6], by setting

$$m_k(x_k + s) \;:=\; f(x_k) + \nabla f(x_k)^\top s + \frac{1}{2} s^\top H_k s,$$

$$\phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|,$$

$$\delta_k = \Delta_k, \quad \alpha = 1, \quad \beta = 0,$$

  Thus, in view of Corollary 3.4, the WCC of this class of method for deriving the norm of the gradient below some given positive threshold is of $\mathcal{O}(\epsilon^{-1})$.
- the ARC algorithm [4], by setting

$$m_k(x_k + s) \;:=\; f(x_k) + \nabla f(x_k)^\top s + \frac{1}{2} s^\top B_k s + \frac{1}{3}\sigma_k \|s\|^3,$$

$$\phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|,$$

$$\delta_k = \frac{1}{\sigma_k}, \quad \alpha = \frac{1}{2}, \quad \beta = \frac{1}{2}.$$

  Thus in view of Corollary 3.4, the WCC bound of $\mathcal{O}(\epsilon^{-1})$ is established for the ARC algorithm too.
- the nonlinear trust-region method [19], by setting

$$m_k(x_k + s) \;:=\; f(x_k) + \nabla f(x_k)^\top s + \frac{1}{2} s^\top H_k s,$$

$$\phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|,$$

$$\alpha = \frac{1}{2}, \quad \beta = \frac{1}{2}.$$

Thus, in view of Corollary 3.4 the WCC bound of $\mathcal{O}(\epsilon^{-1})$ is also established for this method when deriving the norm of the gradient below some given positive threshold is required.

Finally, we should add that there are other methods which are covered by the NSC framework (see [11] and the references therein). For those methods the WCC bound of $\mathcal{O}(\epsilon^{-1})$ is easily derived as well.

4. Conclusions

In this paper, we analyzed the WCC of NSC framework for minimization of convex smooth unconstrained problems. We showed that, in order to drive the norm of the gradient of the objective function below some predefined positive ϵ, the required number of iterations this framework takes is at most of $\mathcal{O}(\epsilon^{-1})$. As the NSC framework is a generalization of several methods, such bound is automatically established for those methods. To the best of our knowledge, it is for the first time that is shown that these class of methods enjoys the same WCC bound as that of the gradient descent methods. We should add that the bound $\mathcal{O}(\epsilon^{-1})$ for both NSC and ARC, when the goal is to drive the norm of the gradient below ϵ, has not been previously established in the literature.

Funding

Support for the author was provided by FCT (Fundação para a Ciência e a Tecnologia) under the scholarship SFRH/BPD/89903/2012 and by the Center for Mathematics of the University of Coimbra - UID/MAT/00324/2013, funded by the Portuguese Government through FCT/MEC and co-funded by the European Regional Development Fund through the Partnership Agreement PT2020, until March 2019. Also, from April 01, 2019 by FCT under the projects PTDC/MAT-APL/28400/2017 and UID/MAT/00297/2019 (Centro de Matemática e Aplicações, FCT-UNL).

References

[1] W. Bian and X. Chen. Worst-case complexity of smoothing quadratic regularization methods for non-Lipschitzian optimization. *SIAM J. Optim.*, 23:1718–1741, 2013.

[2] E. G. Birgin and J. M. Martínez. The use of quadratic regularization with a cubic descent condition for unconstrained optimization. *SIAM J. Optim.*, 27(2):1049–1074, 2017.

[3] C. Cartis, N. I. M. Gould, and Ph. L. Toint. Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function-evaluation complexity. *Math. Program.*, 130:295–319, 2011.

[4] C. Cartis, N. I. M. Gould, and Ph. L. Toint. Evaluation complexity of adaptive cubic regularization methods for convex unconstrained optimization. *Optim. Methods Softw.*, 27:197–219, 2012.

[5] 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.*, 22:66–86, 2012.

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

[7] F. E. Curtis, P. D. Robinson, and M. Samadi. A trust region algorithm with a worst-case iteration complexity of for $\mathcal{O}(\epsilon^{-3/2})$ nonconvex optimization. *Math. Program.*, 162(1-2):1–32, 2017.

[8] M. Dodangeh and L. N. Vicente. Worst case complexity for direct search under convexity. *Math. Program.*, 155:307–332, 2016.

[9] R. Garmanjani, D. Júdice, and L. N. Vicente. Trust-region methods without using derivatives: Worst case complexity and the non-smooth case. *SIAM J. Optim.*, 26:1987–2011, 2016.

[10] R. Garmanjani and L. N. Vicente. Smoothing and worst case complexity for direct-search methods in non-smooth optimization. *IMA J. Numer. Anal.*, 33:1008–1028, 2013.

[11] G.N. Grapiglia, J. Yuan, and Y. Yuan. On the worst-case complexity of nonlinear stepsize control algorithms for convex unconstrained optimization. *Optim. Methods Softw.*, 31:591–604, 2016.

[12] S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang. Direct search based on probabilistic descent. *SIAM J. Optim.*, 25:1515–1541, 2015.

[13] S. Gratton, A. Sartenaer, and Ph. L. Toint. Recursive trust-region methods for multiscale nonlinear optimization. *SIAM J. Optim.*, 19:414–444, 2008.

[14] J. Konecnỳ and P. Richtárik. Simple complexity analysis of direct search. Technical report, 2014.

[15] J. M. Martínez and M. Raydan. Cubic-regularization counterpart of a variable-norm trust-region method for unconstrained minimization. *J. Global Optim.*, 68(2):367–385, 2017.

[16] Y. Nesterov. *Introductory Lectures on Convex Optimization.* Kluwer Academic Publishers, Dordrecht, 2004.

[17] Y. Nesterov. How to make the gradients small. *Optima*, 88:10–11, 2012.

[18] Y. Nesterov and B. Polyak. Cubic regularization of Newton's method and its global performance. *Math. Program.*, 108:177–205, 2006.

[19] Ph. L. Toint. Nonlinear stepsize control, trust regions and regularizations for unconstrained optimization. *Optim. Methods Softw.*, 28:82–95, 2013.

[20] L. N. Vicente. Worst case complexity of direct search. *EURO Journal on Computational Optimization*, 1:143–153, 2013.