Nonlinear Stepsize Control Algorithms: Complexity Bounds for First- and Second-Order Optimality

Geovani Nunes Grapiglia, Jinyun Yuan & Ya-xiang Yuan

Journal of Optimization Theory and Applications

ISSN 0022-3239

J Optim Theory Appl
DOI 10.1007/s10957-016-1007-x

Springer

J Optim Theory Appl
DOI 10.1007/s10957-016-1007-x

Nonlinear Stepsize Control Algorithms: Complexity Bounds for First- and Second-Order Optimality

Geovani Nunes Grapiglia[1] · Jinyun Yuan[1] · Ya-xiang Yuan[2]

Received: 14 September 2015 / Accepted: 29 August 2016

Abstract A nonlinear stepsize control (NSC) framework has been proposed by Toint (Optim Methods Softw 28:82–95, 2013) for unconstrained optimization, generalizing several trust-region and regularization algorithms. More recently, worst-case complexity bounds to achieve approximate first-order optimality were proved by Grapiglia, Yuan and Yuan (Math Program 152:491–520, 2015) for the generic NSC framework. In this paper, improved complexity bounds for first-order optimality are obtained. Furthermore, complexity bounds for second-order optimality are also provided.

Keywords Worst-case complexity · Trust-region methods · Regularization methods · Unconstrained optimization

Mathematics Subject Classification 90C30 · 65K05 · 49M37 · 49M15 · 90C29 · 90C60 · 68Q25

Communicated by Johannes O. Royset.

✉ Geovani Nunes Grapiglia
grapiglia@ufpr.br

Jinyun Yuan
jin@ufpr.br

Ya-xiang Yuan
yyx@lsec.cc.ac.cn

1 Departamento de Matemática, Universidade Federal do Paraná, Centro Politécnico, Cx. Postal 19.081, Curitiba, Paraná 81531-980, Brazil

2 State Key Laboratory of Scientific/Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Zhong Guan Cun Donglu 55, Beijing 100190, People's Republic of China

Published online: 08 September 2016

1 Introduction

The focus of this work is the theoretical analysis of a class of algorithms for unconstrained optimization of a possibly non-convex function. Notable iterative schemes for this type of problem are line-search, trust-region, and regularization algorithms. As observed by Shultz et al. [1] and by Toint [2], line-search algorithms can be seen as particular instances of trust-region algorithms. Recently, a nonlinear stepsize control (NSC) framework has been proposed by Toint [3] as a generalization of many trust-region and regularization algorithms. This framework provides a unified setting in which theoretical results can be proved for a wide range of algorithms. In particular, the NSC framework covers the classical trust-region method [4,5], the ARC algorithm proposed by Cartis et al. [6,7], the quadratic regularization method proposed by Nesterov [8] (as extended by Bellavia et al. [9]), the modified trust-region method proposed by Fan and Yuan [10], the quadratic regularization methods proposed by Zhang and Wang [11] and by Fan [12], respectively, and the conic trust-region method proposed by Lu and Ni [13]. For details, see [3] and [14].

Under the assumption that the Hessians of the models are uniformly bounded, Toint [3] has proved lim-type and liminf-type global convergence results for the class of NSC algorithms. For the more general case in which the norm of the Hessians can grow by a constant amount at each iteration, Grapiglia, Yuan and Yuan [15] have proved a liminf-type convergence result by adapting the seminal analysis done by Powell [16] for the classical trust-region method. Regarding the worst-case complexity of NSC algorithms, Grapiglia, Yuan and Yuan [15] have also provided upper bounds on the number of iterations required in the worst-case to reduce a certain first-order criticality measure below a given threshold.

In this paper, we investigate the worst-case complexity of the NSC algorithms to achieve approximate first- and second-order optimality. More specifically, we can distinguish three main contributions of this work. Firstly, by modifying one basic assumption, we further generalize the NSC framework in order to include the regularization algorithms described in [17]. Then, by using a different proof technique, we are able to improve the complexity results established in [15] to ensure approximate first-order optimality. Finally, we propose a second-order NSC framework, and we estimate an upper bound on the number of iterations required in the worst case to ensure approximate second-order optimality. This result generalizes the complexity bound proved by Cartis, Gould, and Toint [18] for a second-order version of the classical trust-region method.

The paper is organized as follows. In Sect. 2, the NSC framework is reviewed. In Sect. 3, the improved complexity bounds for first-order optimality are proved. In Sect. 4, complexity bounds for second-order optimality are obtained. Finally, in Sect. 5, the contributions of the paper are summarized.

2 Preliminaries on the NSC Framework

As mentioned above, 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 smooth and bounded below. In order to describe the NSC in a compact way, let us consider first the following conditions:

A1 There exists a continuous, bounded, and nonnegative function $\omega : \mathbb{R}^n \to \mathbb{R}$ such that $\omega(x) = 0$ only if $\nabla f(x) = 0$.
A2 There exist three continuous nonnegative functions $\phi, \psi, \chi : \mathbb{R}^n \to \mathbb{R}$, such that, provided $\omega(x) > 0$, we have $\min\{\phi(x), \psi(x), \chi(x)\} = 0$ only if $\nabla f(x) = 0$.
A3 There exists $\kappa_\chi > 0$ such that

$$\chi(x) \leq \kappa_\chi \text{ for all } x. \tag{2}$$

By convention, from here, we denote

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

A4 The step s_k satisfies the bound

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

for some constants $\kappa_s \geq 1$ and $\kappa_\delta > 0$, where the function Δ is of the form

$$\Delta(\delta, \chi) = \delta^\alpha \chi^\beta, \tag{4}$$

for some powers $\alpha \in [0, 1]$, $\alpha \neq 0$, and $\beta \in [0, 1]$.
A5 The step s_k produces a decrease in the model, which is sufficient in the sense that

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_c \psi_k \min\left\{\frac{\phi_k}{1 + \|H_k\|}, \Delta(\delta_k, \chi_k)\right\}, \tag{5}$$

for some positive constant $\kappa_c < 1$, and where H_k is the Hessian matrix of f at x_k or an approximation thereof.
A6 For all $k \geq 1$, the model $m_k(x_k + s) : \mathbb{R}^n \to \mathbb{R}$ satisfies[1]

$$m_k(x_k) = f(x_k) \quad \text{and} \quad f(x_k + s) - m_k(x_k + s) \leq \kappa_m \|s\|^2 \ \forall s \in \mathbb{R}^n, \tag{6}$$

for some constant $\kappa_m > 0$.

Considering A1-A6, we can summarize the generic NSC framework as follows.

Algorithm 1. (Nonlinear Stepsize Control Algorithm (first order) [3])

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 $k := 1$.

[1] The inequality in (6) typically results from an error bound on Taylor series and a bounded H_k.

Step 1 Choose a model $m_k(x_k+s)$ satisfying A6 and find a step s_k which sufficiently reduces the model in the sense of A5 for which $\|s_k\|$ satisfies A4.

Step 2 Compute the ratio

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}, \tag{7}$$

set the next iterate

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{if} \quad \rho_k \geq \eta_1, \\ x_k, & \text{otherwise}, \end{cases} \tag{8}$$

and choose the stepsize parameter δ_{k+1} by the update rule

$$\delta_{k+1} \in \begin{cases} [\gamma_1\delta_k, \gamma_2\delta_k], & \text{if} \quad \rho_k < \eta_1, \\ [\gamma_2\delta_k, \gamma_3\delta_k], & \text{if} \quad \eta_1 \leq \rho_k < \eta_2, \\ [\delta_k, \gamma_4\delta_k], & \text{if} \quad \rho_k \geq \eta_2. \end{cases} \tag{9}$$

Step 3 Compute H_{k+1}, set $k := k + 1$, and go to Step 1.

Just to mention a few examples, under suitable assumptions, it can be shown that Algorithm 1 covers the following algorithms[2]:

- the classical trust-region algorithm [4,5]:

$$\begin{aligned} &m_k(x_k + s) := f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T H_k s, \\ &\omega(x) = 1, \quad \phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|, \\ &\delta_k = \Delta_k, \quad \alpha = 1, \quad \beta = 0, \end{aligned}$$

- the ARC algorithm of Cartis et al. [6]:

$$\begin{aligned} &m_k(x_k + s) := f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T H_k s + \frac{\sigma_k}{3}\|s\|^3, \\ &\omega(x) = 1, \quad \phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|, \\ &\delta_k = \frac{1}{\sigma_k}, \quad \alpha = 1/2, \quad \beta = 1/2, \end{aligned}$$

- the trust-region algorithm of Fan and Yuan [10]:

$$\begin{aligned} &m_k(x_k + s) := f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T H_k s, \\ &\omega(x) = 1, \quad \phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|, \\ &\delta_k = \mu_k, \quad \alpha = 1, \quad \beta = 1, \end{aligned}$$

2 For details, see Section 2 in [3].

Table 1 Worst-case complexity bounds obtained from Theorem 2.1

Algorithm	A8	A9	(α, β)	Complexity bound
Classical trust-region [4,5]	Yes	Yes	$(1, 0)$	$O(\varepsilon^{-2})$
Trust-region of Fan and Yuan [10]	Yes	No	$(1, 1)$	$O(\varepsilon^{-3})$
ARC algorithm [6]	Yes	Yes	$\left(\frac{1}{2}, \frac{1}{2}\right)$	$O(\varepsilon^{-2})$
Quadratic regularization [11,12]	No	No	$(1, 1)$	$O(\varepsilon^{-3})$

- the quadratic regularization algorithms for $f(x) = (1/2)\|F(x)\|^2$ proposed by Zhang and Wang [11] and by Fan [12]:

$$
\begin{aligned}
&m_k(x_k + s) := \frac{1}{2}\|F(x_k) + J_F(x_k)s\|^2, \\
&\omega(x) = 1, \quad \phi(x) = \psi(x) = \|J_F(x)^T F(x)\|, \\
&\chi(x) = \|F(x)\|^\gamma, \quad \delta_k = \nu^j, \quad \alpha = 1, \quad \beta = 1,
\end{aligned}
$$

  where $J_F(x)$ is the Jacobian of F at x, $1/2 < \gamma < 1$, $0 < \nu < 1$ and j is reset to zero when a new iterate is accepted and incremented by one otherwise.

Before to recall the complexity bounds given in [15], consider the following additional conditions:

A7 There exists a constant $\kappa_H > 0$ such that $\|H_k\| \leq \kappa_H$ for all k.
A8 For all k, $\phi_k \geq \chi_k$ and $\psi_k \geq \chi_k$.
A9 The powers α and β satisfy the inequalities $\alpha + \beta \leq 1$ and $2\alpha + \beta \geq 1$.

Theorem 2.1 (Theorems 2 and 3 in [15]) *Suppose that A1–A7 hold. Let $\{f(x_k)\}$ be bounded below by f_{low}. Then, to reduce the criticality measure $F_k = \min\{\omega_k, \phi_k, \psi_k, \chi_k\}$ below ε, $0 < \varepsilon < 1$, Algorithm 1 takes at most $O(\varepsilon^{-(2+\beta)})$ iterations. If additionally, A8–A9 hold, then this worst-case complexity bound is reduced to $O(\varepsilon^{-2})$ iterations.*

Table 1 above summarizes the complexity bounds obtained from Theorem 2.1 for the NSC algorithms described above. In the next section, by using a proof technique different from that in [15], we shall establish a complexity bound of $O(\varepsilon^{-2})$ for NSC algorithms satisfying A8, but with (α, β) outside the region defined by A9.

3 Complexity Bounds for First-Order Optimality

In what follows, 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 consider the following notation:

$$S = \{k \geq 1 \mid k \text{ successful}\}, \tag{10}$$
$$S_j = \{k \leq j \mid k \in S\}, \text{ for each } j \geq 1, \tag{11}$$
$$U_j = \{k \leq j \mid k \notin S\} \text{ for each } j \geq 1, \tag{12}$$

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. Moreover, we shall replace A4 by the following condition:

A4' The step s_k satisfies either
(a) $\|s_k\| \leq \kappa_s \Delta(\delta_k, \chi_k)$ for all k; or
(b) if $\beta \neq 0$, $\|s_k\| \leq \kappa_s \Delta(\delta_k, \chi_k)$ whenever $\delta_k \leq \kappa_\delta \chi_k^{\frac{(1-\beta)}{\beta}}$.
for some constants $\kappa_s \geq 1$ and $\kappa_\delta > 0$, where the function Δ is defined as in A4.

Remark 3.1 The new condition A4'(b) allows us to include in the NSC framework the class of regularization algorithms described in [17]. Specifically, this class of algorithms is characterized by the choices:

$$m_k(x_k + s) := f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T H_k s + \frac{\sigma_k}{p} \|s\|^p \quad (p > 2), \tag{13}$$
$$\omega(x) = 1, \quad \phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|, \tag{14}$$
$$\delta_k = \frac{1}{\sigma_k} \quad \text{and} \quad \alpha = \beta = \frac{1}{p-1}. \tag{15}$$

Indeed, by Lemma 3.3 in [17],

$$\|s_k\| \leq \left(\frac{2p}{\sigma_k}\right)^{\frac{1}{p-1}} \|\nabla f(x_k)\|^{\frac{1}{p-1}} \quad \text{whenever} \quad \sigma_k \geq \frac{(p\kappa_H)^{p-1}}{(2p\|\nabla f(x_k)\|)^{p-2}}. \tag{16}$$

Then, using (14) and (15), we can see that (16) is equivalent to the statement

$$\|s_k\| \leq \kappa_s \Delta(\delta_k, \chi_k) \quad \text{whenever} \quad \delta_k \leq \kappa_\delta \chi_k^{\frac{(1-\beta)}{\beta}},$$

where $\kappa_s = (2p)^{\frac{1}{p-1}}$ and $\kappa_\delta = (1/p\kappa_H)^{p-1}(2p)^{p-2}$. That is, the class of algorithms specified by (13)–(15) satisfies A4'(b). Note that this class includes the ARC when $p = 3$. Moreover, for the ARC algorithm ($\alpha = \beta = 1/2$), condition A4'(b) reduces to condition A4.

Remark 3.2 For convenience, in the rest of the paper, when we refer to A4, we actually mean conditions A4'.

3.1 Worst-Case Complexity Analysis

The next lemma gives a lower bound on δ_k^α when χ_k is bounded away from zero and $A4(a)$ holds.

Lemma 3.1 *Suppose that A1–A8 hold and let* $0 < \varepsilon < 1$. *If A4(a) holds and*

$$\chi_k \geq \varepsilon \quad \textit{for} \quad k = 1, \ldots, j, \tag{17}$$

then there exists $\bar{\kappa}_f > 0$, *independent of* k *and* ε, *such that*

$$\delta_k^\alpha \geq \bar{\kappa}_f \varepsilon^{(1-\beta)} \quad \textit{for} \quad k = 1, \ldots, j+1. \tag{18}$$

Proof We show by induction that (18) holds with

$$\bar{\kappa}_f = \min\left\{ \delta_1^\alpha, \frac{\gamma_1^\alpha}{1+\kappa_H}, \frac{\gamma_1^\alpha \kappa_c (1-\eta_2)}{\kappa_m \kappa_s^2} \right\}. \tag{19}$$

Clearly, (18) holds for $k = 1$. Assuming that (18) holds for some $k \in \{1, \ldots, j\}$, we shall prove that (18) also holds for $k+1$. Indeed, by A4(a),

$$\|s_k\| \leq \kappa_s \Delta(\delta_k, \chi_k). \tag{20}$$

Thus, by A5, A6, (20), and A8, we have

$$1 - \rho_k = \frac{f(x_k + s_k) - m_k(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)} \leq \frac{\kappa_m \kappa_s^2 \Delta(\delta_k, \chi_k)^2}{\kappa_c \chi_k \min\left\{ \frac{\chi_k}{1+\kappa_H}, \Delta(\delta_k, \chi_k) \right\}}. \tag{21}$$

Suppose that

$$\delta_k^\alpha < \min\left\{ \frac{\chi_k^{(1-\beta)}}{1+\kappa_H}, \frac{\chi_k^{(1-\beta)} \kappa_c (1-\eta_2)}{\kappa_m \kappa_s^2} \right\}. \tag{22}$$

In this case, it follows that

$$\Delta(\delta_k, \chi_k) = \delta_k^\alpha \chi_k^\beta \leq \min\left\{ \frac{\chi_k}{1+\kappa_H}, \frac{\chi_k \kappa_c (1-\eta_2)}{\kappa_m \kappa_s^2} \right\}. \tag{23}$$

Then, by (21) and (23), we obtain

$$1 - \rho_k \leq \frac{\kappa_m \kappa_s^2 \Delta(\delta_k, \chi_k)^2}{\kappa_c \chi_k \Delta(\delta_k, \chi_k)} = \frac{\kappa_m \kappa_s^2 \Delta(\delta_k, \chi_k)}{\kappa_c \chi_k} \leq 1 - \eta_2,$$

and so $\rho_k \geq \eta_2$. Thus, from rule (9) and the induction assumption, it follows that

$$\delta_{k+1}^\alpha \geq \delta_k^\alpha \geq \bar{\kappa}_f \varepsilon^{(1-\beta)},$$

that is, (18) holds for $k+1$. Now, suppose that (22) is not true. Then, from (9), (17), and (19), we have

$$\begin{aligned}\delta_{k+1}^{\alpha} \geq \gamma_1^{\alpha}\delta_k^{\alpha} &\geq \gamma_1^{\alpha} \min\left\{\frac{\chi_k^{(1-\beta)}}{1+\kappa_H}, \frac{\chi_k^{(1-\beta)}\kappa_c(1-\eta_2)}{\kappa_m\kappa_s^2}\right\}\\ &\geq \min\left\{\frac{\gamma_1^{\alpha}}{1+\kappa_H}, \frac{\gamma_1^{\alpha}\kappa_c(1-\eta_2)}{\kappa_m\kappa_s^2}\right\}\varepsilon^{(1-\beta)}\\ &\geq \bar{\kappa}_f\varepsilon^{(1-\beta)}\end{aligned}$$

that is, (18) holds for $k+1$. This completes the induction argument. □

The next lemma gives a lower bound on δ_k^{α} when χ_k is bounded away from zero and A4(b) holds.

Lemma 3.2 *Suppose that A1–A8 hold and let* $0<\varepsilon<1$. *If A4(b) holds and*

$$\chi_k \geq \varepsilon \quad \textit{for} \quad k=1,\ldots,j, \tag{24}$$

then there exists $\kappa_f>0$, *independent of k and* ε, *such that*

$$\delta_k^{\alpha} \geq \kappa_f\varepsilon^{\max\left\{(1-\beta),\frac{\alpha(1-\beta)}{\beta}\right\}} \quad \textit{for} \quad k=1,\ldots,j+1. \tag{25}$$

Proof We show by induction that (25) holds with

$$\kappa_f = \min\left\{\delta_1^{\alpha}, \gamma_1^{\alpha}\kappa_\delta^{\alpha}, \bar{\kappa}_f\right\} \tag{26}$$

where $\bar{\kappa}_f$ is defined in (19). Clearly, (25) holds for $k=1$. Assuming that (25) holds for some $k\in\{1,\ldots,j\}$, we shall prove that it also holds for $k+1$. Indeed, suppose that

$$\|s_k\| > \kappa_s\Delta(\delta_k,\chi_k). \tag{27}$$

Then, by A4(b), we must have $\delta_k > \kappa_\delta\chi_k^{\frac{(1-\beta)}{\beta}}$. In this case, it follows from (9), (24), and (26) that

$$\begin{aligned}\delta_{k+1}^{\alpha} \geq \gamma_1^{\alpha}\delta_k^{\alpha} > \gamma_1^{\alpha}\kappa_\delta^{\alpha}\chi_k^{\frac{\alpha(1-\beta)}{\beta}} &\geq \gamma_1^{\alpha}\kappa_\delta^{\alpha}\varepsilon^{\frac{\alpha(1-\beta)}{\beta}}\\ &\geq \kappa_f\varepsilon^{\max\left\{(1-\beta),\frac{\alpha(1-\beta)}{\beta}\right\}},\end{aligned}$$

that is, (25) holds for $k+1$. On the other hand, if

$$\|s_k\| \leq \kappa_s\Delta(\delta_k,\chi_k)$$

by following the same argument as in the proof of Lemma 3.1, we also get the bound

$$\delta_{k+1}^{\alpha} \geq \kappa_f \varepsilon^{\max\left\{(1-\beta), \frac{\alpha(1-\beta)}{\beta}\right\}},$$

that is, (25) also holds for $k+1$. This completes the induction argument. □

The next result combines the two previous lemmas.

Lemma 3.3 *Suppose that A1–A8 hold and let* $0 < \varepsilon < 1$. *If*

$$\chi_k \geq \varepsilon \quad \textit{for} \quad k = 1, \ldots, j,$$

then

$$\delta_k^{\alpha} \geq \kappa_f \varepsilon^{g(\alpha,\beta)} \quad \textit{for} \quad k = 1, \ldots, j+1,$$

where κ_f *is the positive constant defined in (26) and where the function g is given by*

$$g(\alpha, \beta) = \begin{cases} 1-\beta, & \textit{if A4(a) holds,} \\ \max\left\{1-\beta, \frac{\alpha(1-\beta)}{\beta}\right\}, & \textit{if A4(b) holds.} \end{cases}$$

The theorem below provides an iteration complexity bound for Algorithm 1 to achieve approximate first-order optimality. Its proof is based on the proof of Theorem 2.1 and Corollary 3.4 in [7].

Theorem 3.1 *Suppose that A1–A8 hold and that* $\{f(x_k)\}$ *is bounded below by* f_{low}. *Then, given* $0 < \varepsilon < 1$, *Algorithm 1 takes at most* $O\left(\varepsilon^{-\max\{2, 1+g(\alpha,\beta)+\beta\}}\right)$ *iterations to ensure* $F_k = \min\{\omega_k, \phi_k, \psi_k, \chi_k\} \leq \varepsilon$.

Proof Let $j_1 \leq +\infty$ be the first iteration such that $F_{j_1+1} \leq \varepsilon$. Then, $F_k > \varepsilon$ for $k = 1, \ldots, j_1$. Thus, by A5, A7, and Lemma 3.3, we have

$$\begin{aligned} m_k(x_k) - m_k(x_k + s_k) &> \kappa_c \varepsilon \min\left\{\frac{\varepsilon}{1+\kappa_H}, \kappa_f \varepsilon^{g(\alpha,\beta)} \varepsilon^{\beta}\right\} \\ &\geq \kappa_c \min\left\{\frac{1}{1+\kappa_H}, \kappa_f\right\} \min\left\{\varepsilon^2, \varepsilon^{1+g(\alpha,\beta)+\beta}\right\} \\ &= \bar{\kappa}_c \varepsilon^{\max\{2, 1+g(\alpha,\beta)+\beta\}}, \quad \text{for } k = 1, \ldots, j_1, \end{aligned}$$

where $\bar{\kappa}_c = \kappa_c \min\left\{1/(1+\kappa_H), \kappa_f\right\}$. Thus, as $f(x_k) \geq f_{\text{low}}$ for all k and the sequence $\{f(x_k)\}$ is monotonically non-increasing, it follows that

$$\begin{aligned} f(x_1) - f_{\text{low}} &\geq \sum_{k=1}^{\infty} [f(x_k) - f(x_{k+1})] \geq \sum_{k=1,\, k \in S_{j_1}} [f(x_k) - f(x_{k+1})] \\ &\geq \sum_{k=1,\, k \in S_{j_1}} \eta_1 \bar{\kappa}_c \varepsilon^{\max\{2, 1+g(\alpha,\beta)+\beta\}} \\ &= \eta_1 \bar{\kappa}_c |S_{j_1}| \varepsilon^{\max\{2, 1+g(\alpha,\beta)+\beta\}}. \end{aligned}$$

Therefore,

$$|S_{j_1}| \leq \frac{(f(x_1) - f_{\text{low}})}{\eta_1 \bar{\kappa}_c} \varepsilon^{-\max\{2, 1+g(\alpha,\beta)+\beta\}}. \tag{28}$$

On the other hand, from (9) and Lemma 3.3, it follows that

$$\begin{aligned} \delta_{k+1}^{\alpha} &\leq \gamma_4^{\alpha} \delta_k^{\alpha}, \quad \text{if } k \in S_{j_1}, \\ \delta_{k+1}^{\alpha} &\leq \gamma_2^{\alpha} \delta_k^{\alpha}, \quad \text{if } k \in U_{j_1}, \\ \delta_k^{\alpha} &\geq \kappa_f \varepsilon^{g(\alpha,\beta)}, \quad \text{for } k = 1, \ldots, j_1 + 1. \end{aligned}$$

Thus, considering $u_k \equiv 1/\delta_k^{\alpha}$, we have

$$\alpha_4 u_k \leq u_{k+1}, \quad \text{if} \quad k \in S_{j_1}, \tag{29}$$

$$\alpha_2 u_k \leq u_{k+1}, \quad \text{if} \quad k \in U_{j_1}, \tag{30}$$

$$u_k \leq \bar{u} \varepsilon^{-g(\alpha,\beta)}, \quad \text{for} \quad k = 1, \ldots, j_1 + 1, \tag{31}$$

where $\alpha_4 = \gamma_4^{-\alpha} \in (0, 1)$, $\alpha_2 = \gamma_2^{-\alpha} > 1$, and $\bar{u} = \kappa_f^{-1}$. From (29) and (30), we deduce inductively

$$u_1 \alpha_4^{|S_{j_1}|} \alpha_2^{|U_{j_1}|} \leq u_{j_1+1}.$$

From (31), it follows that

$$\alpha_4^{|S_{j_1}|} \alpha_2^{|U_{j_1}|} \leq \frac{\bar{u}}{u_1} \varepsilon^{-g(\alpha,\beta)}.$$

Then, taking logarithm on both sides, we get

$$|U_{j_1}| \leq \left[-\frac{\log(\alpha_4)}{\log(\alpha_2)} |S_{j_1}| + \frac{\bar{u}}{u_1 \log(\alpha_2)} \varepsilon^{-g(\alpha,\beta)} \right]. \tag{32}$$

Finally, since $j_1 = |S_{j_1}| + |U_{j_1}|$ and

$$\varepsilon^{-g(\alpha,\beta)} \leq \varepsilon^{-\max\{2, 1+g(\alpha,\beta)+\beta\}},$$

we conclude from (28) and (32) that $j_1 \leq O\left(\varepsilon^{-\max\{2, 1+g(\alpha,\beta)+\beta\}}\right)$. □

Remark 3.3 If A4(a) holds, then $g(\alpha, \beta) = 1 - \beta$ and so Theorem 3.1 provides an upper bound of $O(\varepsilon^{-2})$ iterations. This result is better than the corresponding result in [15] (see Theorem 2.1), as it was established without assumption A9. For example, if we consider the trust-region algorithm of Fan and Yuan [10], Theorem 2.1 gives a bound of $O(\varepsilon^{-3})$ iterations, while Theorem 3.1 gives a bound of $O(\varepsilon^{-2})$, which is a significant improvement. On the other hand, when A4(b) holds, we have two cases. If $\beta \geq \alpha$, then

$$\max\{2, 1 + g(\alpha, \beta) + \beta\} = 2,$$

and we also get a bound of $O(\varepsilon^{-2})$ iterations. If $\beta < \alpha$, then

$$\max\{2, 1 + g(\alpha, \beta) + \beta\} = 1 + \frac{\alpha(1-\beta)}{\beta} + \beta > 2,$$

and so Theorem 3.1 provides an upper bound of $O\left(\varepsilon^{-\left(1+\frac{\alpha(1-\beta)}{\beta}+\beta\right)}\right)$ iterations. However, under assumptions A4(b), the authors are only aware of algorithms for which $\beta = \alpha$. Specifically, we have in mind the regularization algorithms specified by (13)–(15). In this case, Theorem 3.1 recovers results proved in [17]. Table 2 below shows a comparison between the complexity bounds obtained from Theorems 2.1 and 3.1 for some of the algorithms covered by Algorithm 1 and that satisfy A1–A8.

Remark 3.4 It is worth to mention that similar upper bounds of $O(\varepsilon^{-2})$ have been proved by Nesterov [19] for the steepest descent method, by Gratton, Sartenaer, and Toint [20] for trust-region methods, by Cartis, Gould, and Toint [7] for the basic ARC algorithm, by Cartis, Sampaio, and Toint [21] for a non-monotone line-search algorithm, and by Ueda and Yamashita [22] for the Levenberg–Marquardt method when $f(x) = (1/2)\|F(x)\|^2$ with $F : \mathbb{R}^n \to \mathbb{R}^m$ continuously differentiable. With additional second-order information, improved complexity bounds of $O(\varepsilon^{-3/2})$ have been proved by Nesterov and Polyak [23] for the cubic regularization of the Newton's method, by Cartis, Gould, and Toint [7] for second-order variants of the ARC algorithm, by Curtis, Robinson, and Samadi [24] for a modified trust-region method, and by Martínez and Raydan [25] for a cubic regularization version of a variable-norm trust-region method. Furthermore, a complexity bound of $O(\varepsilon^{-(p+1)/p})$ has been proved by Birgin et al. [26] for a regularization method based on the minimization of $(p+1)$-rst order models.

Table 2 Comparison between the complexity bounds obtained from Theorems 2.1 and 3.1

Algorithm	Assumption	(α, β)	Theorem 2.1	Theorem 3.1
Classical trust-region [4,5]	A4(a)	$(1, 0)$	$O(\varepsilon^{-2})$	$O(\varepsilon^{-2})$
Trust-region of Fan and Yuan [10]	A4(a)	$(1, 1)$	$O(\varepsilon^{-3})$	$O(\varepsilon^{-2})$
ARC algorithm [6]	A4(b)	$\left(\frac{1}{2}, \frac{1}{2}\right)$	$O(\varepsilon^{-2})$	$O(\varepsilon^{-2})$
Regularization algorithms [17]	A4(b)	$\left(\frac{1}{p-1}, \frac{1}{p-1}\right)$	Does not apply	$O(\varepsilon^{-2})$

Remark 3.5 In [15], the NSC framework was extended to include some algorithms for composite non-smooth optimization (NSO) problems and for unconstrained multiobjective optimization (MOO) problems. As expected, with minor changes in the proof, Theorem 3.1 remains true if we consider these extensions of the NSC algorithm. Specifically, for NSO and MOO problems, $\|\nabla f(x)\|$ in A1 and A2 must be replaced by the appropriate 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].

4 Complexity Bounds for Second-Order Optimality

In this section, we investigate the worst-case complexity of a second-order variant of Algorithm 1. We shall denote by $\tau_k = \lambda_{min}(H_k)$ the smallest eigenvalue of the symmetric matrix H_k. Let us consider the following conditions:

C1–C3 Same as A1–A3.
C4 The step s_k satisfies

$$\|s_k\| \leq \kappa_s \Delta(\delta_k, \upsilon_k) \quad \text{for all } k,$$

for some constant $\kappa_s \geq 1$, where the function Δ is of the form

$$\Delta(\delta, \upsilon) = \delta^\alpha \upsilon^\beta$$

for some powers $\alpha \in [0, 1]$, $\alpha \neq 0$, and $\beta \in [0, 1]$. From here, we shall denote $\upsilon_k = \max\{\chi_k, -\tau_k\}$.
C5 The step s_k produces a decrease in the model, which is sufficient in the sense that

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_c \max\left\{\psi_k \min\left\{\frac{\phi_k}{1 + \|H_k\|}, \Delta(\delta_k, \upsilon_k)\right\}, -\tau_k \Delta(\delta_k, \upsilon_k)^2\right\}$$

for some positive constant $\kappa_c < 1$, and where H_k is the Hessian matrix of f or an approximation thereof.
C6 For all $k \geq 1$, the model $m_k(x_k + s) : \mathbb{R}^n \to \mathbb{R}$ satisfies

$$m_k(x_k) = f(x_k) \quad \text{and} \quad f(x_k + s) - m_k(x_k + s) \leq \kappa_m \min\left\{\|s_k\|^2, \|s_k\|^3\right\},$$

for some constant $\kappa_m > 0$.
C7–C8 Same as A7–A8.

Now, we can state a generic second-order NSC algorithm as follows.

Algorithm 2. (Nonlinear Stepsize Control Algorithm (second order))

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 $m_k(x_k+s)$ satisfying C6 and find a step s_k which sufficiently reduces the model in the sense of C5 for which $\|s_k\|$ satisfies C4.

Step 2 Compute the ratio ρ_k by (7);
Set the next iterate x_{k+1} by (8); and
Choose the stepsize parameter δ_{k+1} by the update rule (9).

Step 3 Compute H_{k+1}, set $k := k+1$ and go to Step 1.

We claim that, under suitable assumptions, Algorithm 2 covers the class of nonlinear trust-region algorithms specified by the choices:

$$m_k(x_k+s) := f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T H_k s, \tag{33}$$

$$\omega(x) = 1, \quad \phi(x) = \psi(x) = \chi(x) = \|\nabla f(x)\|, \tag{34}$$

$$\Delta_k = \delta_k^\alpha \max\{\chi_k, -\tau_k\}^\beta, \quad \alpha \in [0,1]\ (\alpha \neq 0) \quad \text{and} \quad \beta \in [0,1], \tag{35}$$

where Δ_k is the trust-region radius. Note that this class includes the second-order version of the classical trust-region algorithm described in [4,18] for which $(\alpha, \beta) = (1, 0)$. Moreover, if we consider $(\alpha, \beta) = (1, 1)$, this class also covers a novel second-order variant of the trust-region algorithm of Fan and Yuan [10]. Specifically, let us consider the assumptions below.

H1 The function $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable with Lipschitz continuous gradient and Hessian.

H2 For all $k \geq 1$,

$$\|H_k\| \leq \kappa_H \quad \text{and} \quad \left\| \left(\nabla^2 f(x_k) - H_k\right) s_k \right\| \leq \kappa_{hess} \|s_k\|^2,$$

for some constants $\kappa_H > 1$ and $\kappa_{hess} > 0$.

H3 There exists a bounded set $\Omega \subset \mathbb{R}^n$ such that $x_k \in \Omega$ for all k.

H4 For all $k \geq 1$, the step s_k produces a decrease in the model, which is sufficient in the sense that

(a) $m_k(x_k+s_k) \leq m_k(x_k+s_k^C)$, where $s_k^C = -\alpha_k^C \nabla f(x_k)$ and

$$\alpha_k^C = \text{argmin}_{\alpha>0} \{m_k(x_k - \alpha\nabla f(x_k)) : \|-\alpha\nabla f(x_k)\| \leq \Delta_k\}.$$

(b) $m_k(x_k+s_k) \leq m_k(x_k+s_k^E)$ whenever $\tau_k < 0$, where $s_k^E = \alpha_k^E z_k$,

$$\alpha_k^E = \text{argmin}_{0<\alpha\leq 1} \{m_k(x_k + \alpha z_k)\}$$

and v_k satisfies

$$z_k^T \nabla f(x_k) \leq 0, \quad \|z_k\| = \Delta_k \quad \text{and} \quad z_k^T H_k z_k \leq \kappa_\nu \tau_k \Delta_k^2.$$

for some constant $0 < \kappa_\nu \leq 1$.

Let us now justify our claim. Conditions C1 and C2 follow directly from (34), while C3 is satisfied due to H3 and the continuity of $\nabla f(x)$ (guaranteed by H1). As in trust-region algorithms we have $\|s_k\| \leq \Delta_k$ for all k, it follows from (35) that C4 is satisfied. On its turn, Condition C5 follows from Theorems 6.3.1 and 6.6.1 in Conn, Gould, and Toint [4]. Finally, Condition C6 follows from Lemma 4.1 in Cartis, Gould, and Toint [18].

4.1 Worst-Case Complexity Analysis

The next lemma gives a lower bound on δ_k^α when χ_k or $-\tau_k$ is bounded away from zero.

Lemma 4.1 *Suppose that C1–C8 hold and let* $0 < \varepsilon_F, \varepsilon_H < 1$. *If, for* $k = 1, \ldots, j$,

$$\chi_k \geq \varepsilon_F \quad or \quad \tau_k \leq -\varepsilon_H, \tag{36}$$

then there exists a constant $\kappa_w > 0$, *independent of* k, ε_F *and* ε_H, *such that, for* $k = 1, \ldots, j+1$,

$$\delta_k^\alpha \geq \kappa_w \min\{\varepsilon_F, \varepsilon_H\}^{(1-\beta)}. \tag{37}$$

Proof We show by induction that (37) holds with

$$\kappa_w = \min\left\{\delta_1^\alpha, \frac{\gamma_1^\alpha}{1+\kappa_H}, \frac{\gamma_1^\alpha \kappa_c(1-\eta_2)}{\kappa_m \kappa_s^2}, \frac{\gamma_1^\alpha \kappa_c(1-\eta_2)}{\kappa_m \kappa_s^3}\right\}. \tag{38}$$

Clearly, (37) holds for $k = 1$. Assuming that (37) holds for some $k \in \{1, \ldots, j\}$, we shall prove that (37) also holds for $k+1$. Let us divide the proof in two cases.
Case I: $\chi_k \geq -\tau_k$.
In this case, by (36) and C4, we have

$$\chi_k \geq \min\{\varepsilon_F, \varepsilon_H\} \quad \text{and} \quad \|s_k\| \leq \kappa_s \delta_k^\alpha \chi_k^\beta. \tag{39}$$

Thus, (7), C5–C8, and (39) imply that

$$1 - \rho_k = \frac{f(x_k + s_k) - m_k(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)} \leq \frac{\kappa_m \kappa_s^2 \left(\delta_k^\alpha \chi_k^\beta\right)^2}{\kappa_c \chi_k \min\left\{\frac{\chi_k}{1+\kappa_H}, \delta_k^\alpha \chi_k^\beta\right\}}. \tag{40}$$

Suppose that

$$\delta_k^\alpha < \min\left\{\frac{\chi_k^{(1-\beta)}}{1+\kappa_H}, \frac{\chi_k^{(1-\beta)} \kappa_c(1-\eta_2)}{\kappa_m \kappa_s^2}\right\}. \tag{41}$$

In this case, it follows that

$$\delta_k^\alpha \chi_k^\beta < \min\left\{\frac{\chi_k}{1+\kappa_H}, \frac{\chi_k \kappa_c (1-\eta_2)}{\kappa_m \kappa_s^2}\right\}. \tag{42}$$

Then, by (40) and (42), we obtain

$$1-\rho_k \le \frac{\kappa_m \kappa_s^2 \left(\delta_k^\alpha \chi_k^\beta\right)^2}{\kappa_c \chi_k \left(\delta_k^\alpha \chi_k^\beta\right)} = \frac{\kappa_m \kappa_s^2 \left(\delta_k^\alpha \chi_k^\beta\right)}{\kappa_c \chi_k} \le 1-\eta_2,$$

and so, $\rho_k \ge \eta_2$. Thus, from rule (9) and the induction assumption, it follows that

$$\delta_{k+1}^\alpha \ge \delta_k^\alpha \ge \kappa_w \min\{\varepsilon_F, \varepsilon_H\}^{(1-\beta)},$$

that is, (37) holds for $k+1$.

On the other hand, if (41) is not true, then from (9), (39), and (38), it follows that

$$\begin{aligned}\delta_{k+1}^\alpha \ge \gamma_1^\alpha \delta_k^\alpha &\ge \min\left\{\frac{\gamma_1^\alpha}{1+\kappa_H}, \frac{\gamma_1^\alpha \kappa_c (1-\eta_2)}{\kappa_m \kappa_s^2}\right\} \min\{\varepsilon_F, \varepsilon_H\}^{(1-\beta)} \\ &\ge \kappa_w \min\{\varepsilon_F, \varepsilon_H\}^{(1-\beta)},\end{aligned}$$

that is, (37) holds for $k+1$.

Case II: $-\tau_k > \chi_k$.

In this case, by (36) and C4, we have

$$-\tau_k \ge \min\{\varepsilon_F, \varepsilon_H\} \quad \text{and} \quad \|s_k\| \le \kappa_s \delta_k^\alpha |\tau_k|^\beta. \tag{43}$$

Thus, (7), C5–C8, and (43) imply that

$$1-\rho_k = \frac{f(x_k+s_k) - m_k(x_k+s_k)}{m_k(x_k) - m_k(x_k+s_k)} \le \frac{\kappa_m \kappa_s^3 \left(\delta_k^\alpha |\tau_k|^\beta\right)^3}{\kappa_c |\tau_k| \left(\delta_k^\alpha |\tau_k|^\beta\right)^2} = \frac{\kappa_m \kappa_s^3}{\kappa_c |\tau_k|} \left(\delta_k^\alpha |\tau_k|^\beta\right). \tag{44}$$

Suppose that

$$\delta_k^\alpha < \frac{|\tau_k|^{(1-\beta)} \kappa_c (1-\eta_2)}{\kappa_m \kappa_s^3}. \tag{45}$$

In this case, it follows that

$$\delta_k^\alpha |\tau_k|^\beta < \frac{|\tau_k| \kappa_c (1-\eta_2)}{\kappa_m \kappa_s^3}. \tag{46}$$

Then, combining (44) and (46), we obtain $1 - \rho_k \leq 1 - \eta_2$, and so $\rho_k \geq \eta_2$. Thus, from rule (9) and the induction assumption, it follows that

$$\delta_{k+1}^{\alpha} \geq \delta_k^{\alpha} \geq \kappa_w \min \{\varepsilon_F, \varepsilon_H\}^{(1-\beta)},$$

that is, (37) holds for $k + 1$. On the other hand, if (45) is not true, then from (9), (43), and (38), it follows that

$$\delta_{k+1}^{\alpha} \geq \gamma_1^{\alpha} \delta_k^{\alpha} \geq \frac{\gamma_1^{\alpha} \kappa_c (1 - \eta_2)}{\kappa_m \kappa_s^3} |\tau_k|^{(1-\beta)} \geq \kappa_w \min \{\varepsilon_F, \varepsilon_H\}^{(1-\beta)},$$

that is, (37) holds for $k + 1$. This completes the induction argument. □

We are now ready to obtain an iteration complexity bound for Algorithm 2 to achieve approximate second-order optimality. The proof of this result is based on the proofs of Lemmas 4.5 and 4.6 in [18].

Theorem 4.1 *Suppose that C1–C8 hold and that $\{f(x_k)\}$ is bounded below by f_{low}. Then, given $0 < \varepsilon_F, \varepsilon_H < 1$, Algorithm 2 takes at most $O(\max \left\{\varepsilon_F^{-3}, \varepsilon_H^{-3}\right\})$ iterations to ensure*

$$F_k \leq \varepsilon_F \quad \textit{and} \quad \tau_k \geq -\varepsilon_H. \tag{47}$$

Proof Let $j_1 + 1 \leq +\infty$ by the first iteration such that (47) holds. Then, for $k = 1, \ldots, j_1$, either $F_k > \varepsilon_F$ or $\tau_k < -\varepsilon_H$. If $F_k > \varepsilon_F$, it follows from C5, C8, C4, Lemma 4.1 and $\upsilon_k \geq \min \{\varepsilon_F, \varepsilon_H\}$ that

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_c \varepsilon_F \min \left\{ \frac{\varepsilon_F}{1 + \kappa_H}, \kappa_w \min \{\varepsilon_F, \varepsilon_H\} \right\} \geq \bar{\kappa}_c \min \{\varepsilon_F, \varepsilon_H\}^3,$$

where $\bar{\kappa}_c = \kappa_c \min \{1/(1 + \kappa_H), \kappa_w\}$.

On the other hand, if $\tau_k < -\varepsilon_H$, it follows from C5, C4, Lemma 4.1 and $\upsilon_k \geq \min \{\varepsilon_F, \varepsilon_H\}$ that

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_c |\tau_k| \Delta(\delta_k, \upsilon_k)^2 \geq \tilde{\kappa}_c \min \{\varepsilon_F, \varepsilon_H\}^3,$$

where $\tilde{\kappa}_c = \kappa_c \kappa_w^2$. Thus, as $f(x_k) \geq f_{\text{low}}$ for all k and $\{f(x_k)\}$ is monotonically non-increasing, it follows that

$$\begin{aligned} f(x_1) - f_{\text{low}} &\geq \sum_{k=1}^{\infty} [f(x_k) - f(x_{k+1})] \geq \sum_{k=1, k \in S_{j_1}} \eta_1 \min \{\bar{\kappa}_c, \tilde{\kappa}_c\} \min \{\varepsilon_F, \varepsilon_H\}^3 \\ &= \eta_1 \alpha_c |S_{j_1}| \min \{\varepsilon_F, \varepsilon_H\}^3, \end{aligned}$$

where $\alpha_c = \min \{\bar{\kappa}_c, \tilde{\kappa}_c\}$. Therefore,

$$|S_{j_1}| \leq \bar{\alpha}_c \max \left\{\varepsilon_F^{-3}, \varepsilon_H^{-3}\right\}, \tag{48}$$

where $\bar{\alpha}_c = (f(x_1) - f_{\text{low}})/\eta_1\alpha_c$. Finally, as in the proof of Theorem 3.1, we also can prove that

$$|U_{j_1}| \leq O\left(\max\left\{\varepsilon_F^{-3}, \varepsilon_H^{-3}\right\}\right). \tag{49}$$

Thus, since $j_1 = |S_{j_1}| + |U_{j_1}|$, by (48) and (49), we obtain the desired upper bound. □

Remark 4.1 If $\varepsilon_F = O(\varepsilon_H)$, Theorem 4.1 provides an upper bound of $O(\varepsilon_H^{-3})$ iterations. This bound recovers the result of Theorem 4.7 in [18] for the second-order variant of the classical trust-region algorithm, which correspond to the choice $(\alpha, \beta) = (1, 0)$. In fact, Theorem 4.1 generalizes the referred result in [18] for the whole domain $(\alpha, \beta) \in [0, 1] \times [0, 1]$ $(\alpha \neq 0)$ of stepsize parameters, and thus cover all the nonlinear trust-region algorithms in the class specified by (33)–(35). Moreover, as expected, our bound of $O(\varepsilon_H^{-3})$ matches in order the bounds obtained by Nesterov and Polyak [23] for the cubic regularization of the Newton's method, and by Cartis, Gould, and Toint [18] for the ARC framework.

5 Conclusions

In this paper, we have investigated the worst-case complexity of the nonlinear stepsize control (NSC) framework recently proposed by Toint [3] for unconstrained optimization. Firstly, by modifying one basic assumption, we further generalize the NSC framework. In particular, we were able to include in the framework the regularization algorithms described in [17]. Then, under suitable conditions, we have proved a worst-case complexity bound of $O(\varepsilon^{-2})$ iterations for the generic NSC framework to achieve first-order optimality within ε. This bound improves the results obtained by Grapiglia, Yuan, and Yuan [15]. Finally, we have studied the worst-case complexity of a class of second-order variants of the NSC framework. Specifically, we have proved a worst-case complexity bound of $O(\varepsilon^{-3})$iterations for the generic algorithm to achieve second-order optimality within ε. This bound matches in order the bounds obtained by Nesterov and Polyak [23] for the cubic regularization of Newton's method, and by Cartis, Gould, and Toint [18] for the ARC framework and for the classical trust-region method.

Acknowledgments The authors are grateful to two anonymous referees for their valuable comments and suggestions to improve the paper. G.N. Grapiglia was supported by the Coordination for Improvement of Higher Education Personnel (CAPES), Brazil. J. Yuan was partially supported by the Coordination of 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.

References

1. Shultz, G.A., Schnabel, R.B., Byrd, R.H.: A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties. SIAM J. Numer. Anal. **22**, 47–67 (1985)
2. Toint, PhL: Global convergence of a class of trust region methods for nonconvex minimization in Hilbert space. IMA J. Numer. Anal. **8**, 231–252 (1988)

3. Toint, PhL: Nonlinear stepsize control, trust regions and regularizations for unconstrained optimization. Optim. Methods Softw. **28**, 82–95 (2013)
4. Conn, A.R., Gould, N.I.M., Toint, PhL: Trust-Region Methods. SIAM, Philadelphia (2000)
5. Powell, M.J.D.: Convergence properties of a class of minimization algorithms. In: Mangasarian, O.L., Meyer, R.R., Robinson, S.M. (eds.) Nonlinear Programming 2. Academic Press, New York (1975)
6. Cartis, C., Gould, N.I.M., Toint, PhL: Adaptive cubic overestimation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. Math. Program. **127**, 245–295 (2011)
7. Cartis, C., Gould, N.I.M., Toint, PhL: Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function-and derivative-evaluation complexity. Math. Program. **130**, 295–319 (2011)
8. Nesterov, Yu.: Modified Gauss–Newton scheme with worst-case guarantees for global performance. Optim. Methods Softw. **22**, 469–483 (2007)
9. Bellavia, S., Cartis, C., Gould, N.I.M., Morini, B., Toint, PhL: Convergence of a regularized euclidean residual algorithm for nonlinear least-squares. SIAM J. Numer. Anal. **48**, 1–29 (2010)
10. Fan, J., Yuan, Y.: A new trust region algorithm with trust region radius converging to zero. In: Li, D. (ed.) Proceeding of the 5th International Conference on Optimization: Techiniques and Applications, pp. 786–794. Hong Kong (2001)
11. Zhang, J., Wang, Y.: A new trust region method for nonlinear equations. Math. Methods Oper. Res. **58**, 283–298 (2003)
12. Fan, J.: Convergence rate of the trust region method for nonlinear equations under local error bound condition. Comput. Optim. Appl. **34**, 215–227 (2006)
13. Lu, X., Ni, Q.: A quasi-Newton trust-region method with a new conic model for the unconstrained optimization. Appl. Math. Comput. **204**, 373–384 (2008)
14. Grapiglia, G.N., Yuan, J., Yuan, Y.: On the worst-case complexity of nonlinear stepsize control algorithms for convex unconstrained optimization. Optim. Methods Softw. **31**, 591–604 (2016)
15. Grapiglia, G.N., Yuan, J., Yuan, Y.: On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization. Math. Program. **152**, 491–520 (2015)
16. Powell, M.J.D.: On the global convergence of trust region algorithms for unconstrained minimization. Math. Program. **29**, 297–303 (1984)
17. Cartis, C., Gould, N.I.M., Toint, P.L.: Worst-case evaluation complexity of regularization methods for smooth unconstrained optimization using Hölder continuous gradients. Technical Report NA-14-21, Mathematical Institute, University of Oxford (2014)
18. Cartis, C., Gould, N.I.M., Toint, PhL: Complexity bounds for second-order optimality in unconstrained optimization. J. Complex. **28**, 93–108 (2012)
19. Nesterov, Yu.: Introductory Lectures on Convex Optimization. Applied Optimization. Kluwer, Dordrecht (2004)
20. Gratton, S., Sartenaer, A., Toint, Ph L.: Recursive trust-region methods for multiscale nonlinear optimization. SIAM J. Optim. **19**, 414–444 (2008)
21. Cartis, C., Sampaio, PhR, Toint, PhL: Worst-case evaluation complexity of non-monotone gradient-related algorithms for unconstrained optimization. Optimization **64**, 1349–1361 (2015)
22. Ueda, K., Yamashita, N.: On a global complexity bound of the Levenberg–Marquardt method. J. Optim. Theory Appl. **147**, 443–453 (2010)
23. Nesterov, Yu., Polyak, B.T.: Cubic regularization of Newton method and its global performance. Math. Program. **108**, 177–205 (2006)
24. Curtis, F.E., Robinson, D.P., Samadi, M.: A trust region algorithm with a worst-case iteration complexity of $O(\epsilon^{-3/2})$ for nonconvex optimization. Math. Program. doi:10.1007/s10107-016-1026-2
25. Martínez, J.M., Raydan, M.: Cubic-regularization counterpart of a variable-norm trust-region method for unconstrained minimization (2015). http://www.optimization-online.org/DB_FILE/2015/11/5203.pdf
26. Birgin, E.G., Gardenghi, J.L., Martínez, J.M., Toint, Ph.L.: Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models. Technical Report naXys-05-2015, Namur Center for Complex Systems (naXys), University of Namur, Namur, Belgium (2015)