Math. Program., Ser. A
DOI 10.1007/s10107-014-0794-9

FULL LENGTH PAPER

On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization

Geovani N. Grapiglia · Jinyun Yuan · Ya-xiang Yuan

Received: 23 September 2013 / Accepted: 10 July 2014

Abstract A nonlinear stepsize control framework for unconstrained optimization was recently proposed by Toint (Optim Methods Softw 28:82–95, 2013), providing a unified setting in which the global convergence can be proved for trust-region algorithms and regularization schemes. The original analysis assumes that the Hessians of the models are uniformly bounded. In this paper, the global convergence of the nonlinear stepsize control algorithm is proved under the assumption that the norm of the Hessians can grow by a constant amount at each iteration. The worst-case complexity is also investigated. The results obtained for unconstrained smooth optimization are extended

---

G. N. Grapiglia was supported by CAPES, Brazil (Grant PGCI 12347/12-4). J. Yuan was partially supported by CAPES and CNPq, Brazil. Y. Yuan was partially supported by NSFC, China (Grant 11331012).

G. N. Grapiglia (✉) · J. Yuan
Departamento de Matemática, Universidade Federal do Paraná,
Centro Politécnico, Cx. Postal 19.081, Curitiba, Paraná 81531-980, Brazil
e-mail: geovani_mat@outlook.com

J. Yuan
e-mail: jin@ufpr.br

G. N. Grapiglia
The Capes Foundation, Ministry of Education of Brazil,
Cx. Postal 250, Brasília, Distrito Federal 70.040-020, Brazil

Y. Yuan
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
e-mail: yyx@lsec.cc.ac.cn

Published online: 22 July 2014

to some algorithms for composite nonsmooth optimization and unconstrained multi-objective optimization as well.

Keywords Global convergence · Worst-case complexity · Trust-region methods · Regularization methods · Unconstrained Optimization · Composite nonsmooth optimization · Multiobjective optimization

Mathematics Subject Classification (2010) 90C30 · 65K05 · 49M37 · 49M15 · 90C29 · 90C60 · 68Q25

1 Introduction

Consider the unconstrained optimization problem

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

where $f:\mathbb{R}^n\rightarrow\mathbb{R}$ is continuously differentiable and bounded below. Traditional iterative methods for solving (1) are trust-region [5] and line-search [7] algorithms. As observed independently by Shultz et al. [17] and by Toint [20], line-search methods can also be reinterpreted as trust-region methods, which ensures the derivation of a common convergence theory for both classes of methods (for details, see Section 10.3 in Conn et al. [5]).

Recently, an adaptive regularization approach with cubics (ARC) has been proposed by Cartis et al. [2] as a new globalization technique for unconstrained optimization. However, with the development of the ARC methods, as well non-standard trust-region algorithms [9] and other regularization schemes [1,8,13,24], the unified setting for convergence analysis was lost, since for each one of these methods the global convergence is proved in a different way.

With the purpose to obtain a unifying framework to prove the global convergence of trust-region algorithms and regularization schemes, Toint [19] has proposed the class of nonlinear stepsize control algorithms. In order to describe this class, it is convenient to consider in advance the following conditions:

A1 There exists a continuous, bounded and non-negative function $\omega:\mathbb{R}^n\rightarrow\mathbb{R}$ such that

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

A2 There exist three continuous non-negative functions $\phi,\psi,\chi:\mathbb{R}^n\rightarrow\mathbb{R}$, possibly undefined at roots of ω, 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 $\chi(x)\leq\kappa_\chi$ for all x. 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 function $\Delta : [0, +\infty) \times [0, +\infty) \to \mathbb{R}$ defining the trust-region radius is of the form

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

for some powers $\alpha \in (0, 1]$ 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 constant $\kappa_c \in (0, 1)$, and where H_k is an $n \times n$ symmetric matrix approximating the second order behavior of f in a neighbourhood of x_k.

A6 The step s_k satisfies the bound

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

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

A7 For all $k \geq 1$, the model $m_k(x_k + s) : \mathbb{R}^n \to \mathbb{R}$ satisfies

$$m_k(x_k) = f(x_k) \quad \text{and} \quad f(x_k + s) - m_k(x_k + s) \leq \kappa_m \|s\|^2 \; \forall s \in \mathbb{R}^n, \tag{7}$$

for some constant $\kappa_m > 0$.

Now, the algorithm proposed by Toint [19] can be summarized as follows.

Algorithm 1. (Nonlinear Stepsize Control Algorithm)

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

Step 1 Choose a model $m_k(x_k + s)$ satisfying A7 and find a step s_k which sufficiently reduces the model in the sense of A5 for which $\|s_k\|$ satisfies A6.

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} by the update rule

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

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

It is shown in Toint [19] that Algorithm 1 covers the following algorithms:

- the classical trust-region algorithm [5, 14]:

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

- the ARC algorithm of Cartis et al. [2]:

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

- the quadratic regularization algorithm for $f(x) = \|F(x)\|$ proposed by Nesterov [13] (as extended in [1]):

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

  where $J_F(x)$ is the Jacobian of F at x,
- the trust region algorithm of Fan and Yuan [9]:

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

- the quadratic regularization algorithms for $f(x) = (1/2)\|F(x)\|^2$ proposed by Zhang and Wang [24] and Fan [8]:

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

  where $\gamma \in (\frac{1}{2}, 1)$, $\nu \in (0, 1)$, and j is reset to zero when a new iterate is accepted and incremented by one otherwise.

Toint [19] also provides a global convergence analysis for Algorithm 1, showing the weak and strong convergence under conditions A1–A7 and the assumption that the sequence $\{\|H_k\|\}$ is bounded above. However, for suitable quasi-Newton update formulas one can prove that

$$\|H_k\| \leq c_1 + c_2 k \tag{11}$$

(see Section 8.4 in Conn et al. [5]), even when the boundedness of $\{\|H_k\|\}$ is not explicit. On the other hand, regarding the worst-case complexity of Algorithm 1, Toint [19] argues that the structure of the algorithm and assumptions A1–A7 suggest an upper bound of $O(\varepsilon^{-3})$ iterations for the algorithm to reduce the size of a first-order criticality measure below ε. But, no proof was given for this iteration bound.

Motivated by these observations, in this paper we show the convergence of a slight modification of Algorithm 1, assuming that (11) is satisfied for all k. Furthermore, if the matrices H_k are uniformly bounded, we prove that this algorithm requires at most $O(\varepsilon^{-(2+\beta)})$ iterations to reduce the size of a first-order criticality measure below ε, which is a complexity bound less pessimistic than that one discussed by Toint [19]. For the particular case in which $\alpha + \beta \leq 1$, $2\alpha + \beta \geq 1$ and $\phi_k, \psi_k \geq \chi_k$ (which includes the ARC algorithm), this estimate is even improved to $O(\varepsilon^{-2})$ iterations. These results are then extended to some algorithms for composite nonsmooth optimization (NSO) and unconstrained multiobjective optimization (MOO).

The paper is organized as follows. In Sect. 2, the global convergence results are given. The worst-case complexity is investigated in Sect. 3. Finally, in Sect. 4, the results of the two previous sections are extended for composite NSO and unconstrained MOO.

2 Global convergence analysis

In this section, the global convergence is shown for a slight modification of Algorithm 1 with H_k satisfying (11). Specifically, we replace (10) by a slightly more restrictive rule:

$$\delta_{k+1} \in \begin{cases} \left[\gamma_1 \delta_k, \gamma_2 \delta_k\right], & \text{if } \rho_k < \eta_1, \\ \left[\gamma_2 \delta_k, \gamma_3 \delta_k\right], & \text{if } \rho_k \in [\eta_1, \eta_2), \\ \left[\delta_k, \gamma_4 \delta_k\right], & \text{if } \rho_k \geq \eta_2, \end{cases} \tag{12}$$

where $0 < \gamma_1 < \gamma_2 < \gamma_3 < 1 < \gamma_4$. Moreover, we can replace A3, A6 and A7 by the slightly weaker conditions:

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

$$\chi_k \leq \kappa_\chi, \text{ for all } k.$$

A6′′ The step s_k satisfies the bound

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

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

$$M_k = 1 + \max_{1\leq i\leq k} \|H_i\|.$$

A7$''$ For all $k \geq 1$, the model $m_k(x_k + s) : \mathbb{R}^n \to \mathbb{R}$ satisfies

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

for some constant $\kappa_m > 0$.

Remark 1 For convenience, for the rest of the paper, when we refer to A3, A6 and A7 we actually mean conditions A3$''$, A6$''$ and A7$''$, respectively.

Remark 2 Except by the ARC algorithm, for all the other algorithms mentioned in Sect. 1, there exists $\kappa_s \geq 1$ such that $\|s_k\| \leq \kappa_s \Delta(\delta_k, \chi_k)$ for all k, and so A6 is naturally satisfied. For the ARC algorithm, recall that $\delta_k = 1/\sigma_k$, $\chi_k = \|\nabla f(x_k)\|$ and $\alpha = \beta = 1/2$. Then, by Lemma 2.2 in Cartis et al. [2] (see the proof),

$$\|s_k\| \leq 3 \max\left\{\delta_k \|H_k\|, \Delta(\delta_k, \chi_k)\right\}.$$

Note that, in this case,

$$\begin{aligned} \delta_k M_k^{\frac{1}{\alpha}} \leq \chi_k &\Longrightarrow \delta_k \|H_k\|^2 \leq \chi_k \Longrightarrow \delta_k^{\frac{1}{2}} \|H_k\| \leq \chi_k^{\frac{1}{2}} \\ &\Longrightarrow \delta_k \|H_k\| = \delta_k^{\frac{1}{2}} \delta_k^{\frac{1}{2}} \|H_k\| \leq \delta_k^{\frac{1}{2}} \chi_k^{\frac{1}{2}} = \Delta(\delta_k, \chi_k), \end{aligned}$$

and so, $\|s_k\| \leq 3\Delta(\delta_k, \chi_k)$. Hence, the ARC algorithm satisfies A6 with $\kappa_s = 3$ and $\kappa_\delta = 1$.

The lemma below provides a lower bound for δ_k^α. Its proof is based on the proof of the lemma on page 299 of Powell [16].

Lemma 1 *Suppose that A1–A7 hold. If there exists $\varepsilon > 0$ such that*

$$\min\left\{\phi_k, \psi_k, \chi_k\right\} \geq \varepsilon \quad \textit{for all } k, \tag{13}$$

then, there exists a constant $\tau > 0$ such that

$$\delta_k^\alpha \geq \frac{\tau}{M_k} \quad \textit{for all } k, \tag{14}$$

where M_k is defined by

$$M_k = 1 + \max_{1\leq i\leq k} \|H_i\|. \tag{15}$$

Proof We show by induction that (14) holds with

$$\tau = \min\left\{\delta_1^\alpha M_1, (\gamma_1 \kappa_\delta \varepsilon)^\alpha, \gamma_1^\alpha \varepsilon/\kappa_\chi^\beta, \gamma_1^\alpha \kappa_c \varepsilon (1-\eta_2)/\kappa_\chi^\beta \kappa_m \kappa_s^2\right\}. \tag{16}$$

By the definition of τ, clearly (14) holds for $k = 1$. Assuming that (14) is true for k, we prove that (14) is also true for $k + 1$. If $\delta_k M_k^{\frac{1}{\alpha}} > \kappa_\delta \chi_k$, it follows from (13), (15), (12) and (16) that

$$\delta_k > \frac{\kappa_\delta \varepsilon}{M_k^{\frac{1}{\alpha}}} \geq \frac{\kappa_\delta \varepsilon}{M_{k+1}^{\frac{1}{\alpha}}}, \tag{17}$$

$$\implies \delta_{k+1}^\alpha \geq \gamma_1^\alpha \delta_k^\alpha > \frac{(\gamma_1 \kappa_\delta \varepsilon)^\alpha}{M_{k+1}} \geq \frac{\tau}{M_{k+1}}, \tag{18}$$

that is, (14) holds for $k + 1$. Therefore, for the remainder of the proof we assume $\delta_k M_k^{\frac{1}{\alpha}} \leq \kappa_\delta \chi_k$, which by A6 provides us the bound

$$\|s_k\| \leq \kappa_s \Delta(\delta_k, \chi_k). \tag{19}$$

From A5 and (13) it follows that

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_c \varepsilon \min\left\{\frac{\varepsilon}{1 + \|H_k\|}, \Delta(\delta_k, \chi_k)\right\}. \tag{20}$$

Then, by (8), A7, (20) and (19),

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

Suppose that

$$\delta_k^\alpha \kappa_\chi^\beta < \min\left\{\frac{\varepsilon}{1 + \|H_k\|}, \frac{\kappa_c \varepsilon (1 - \eta_2)}{\kappa_m \kappa_s^2}\right\}. \tag{22}$$

In this case, by A4 and A3 we have

$$\Delta(\delta_k, \chi_k) = \delta_k^\alpha \chi_k^\beta \leq \delta_k^\alpha \kappa_\chi^\beta < \min\left\{\frac{\varepsilon}{1 + \|H_k\|}, \frac{\kappa_c \varepsilon (1 - \eta_2)}{\kappa_m \kappa_s^2}\right\}, \tag{23}$$

which implies in (21) that

$$\begin{aligned} 1 - \rho_k &\leq \frac{\kappa_m \kappa_s^2 \Delta(\delta_k, \chi_k)^2}{\kappa_c \varepsilon \Delta(\delta_k, \chi_k)} = \frac{\kappa_m \kappa_s^2 \Delta(\delta_k, \chi_k)}{\kappa_c \varepsilon} < 1 - \eta_2 \\ &\implies \rho_k > \eta_2. \end{aligned} \tag{24}$$

Thus, from rule (12), the induction assumption and the inequality $M_{k+1} \geq M_k$ it follows that

$$\delta_{k+1} \geq \delta_k \Longrightarrow \delta_{k+1}^{\alpha} \geq \delta_k^{\alpha} \geq \frac{\tau}{M_k} \geq \frac{\tau}{M_{k+1}}, \tag{25}$$

and so (14) holds for $k+1$.

Finally, suppose that (22) is not true. Then, from (12), (15) and (16), it follows that

$$\begin{aligned} \delta_{k+1}^{\alpha} \geq \gamma_1^{\alpha}\delta_k^{\alpha} &\geq \min\left\{\frac{\gamma_1^{\alpha}\varepsilon}{\kappa_{\chi}^{\beta}(1+\|H_k\|)}, \frac{\gamma_1^{\alpha}\kappa_c\varepsilon(1-\eta_2)}{\kappa_{\chi}^{\beta}\kappa_m\kappa_s^2}\right\} \\ &\geq \frac{\min\left\{\gamma_1^{\alpha}\varepsilon/\kappa_{\chi}^{\beta}, \gamma_1^{\alpha}\kappa_c\varepsilon(1-\eta_2)/\kappa_{\chi}^{\beta}\kappa_m\kappa_s^2\right\}}{M_{k+1}} \\ &\geq \frac{\tau}{M_{k+1}}. \end{aligned} \tag{26}$$

This shows that (14) holds for $k+1$ and completes the proof. □

In what follows, we consider the lemma below given by Yuan [22], which and whose proof are due to Powell [16].

Lemma 2 *Let $\{\mu_k\}$ and $\{M_k\}$ be two sequences of real numbers such that $\mu_k \geq \frac{\tau}{M_k} > 0$, where τ is a positive constant. Let J be a subset of $\{1, 2, 3, \ldots\}$ and assume that*

$$\mu_{k+1} \leq c_3\mu_k, \quad k \in J, \tag{27}$$

$$\mu_{k+1} \leq c_4\mu_k, \quad k \notin J, \tag{28}$$

$$M_{k+1} \geq M_k, \quad k \geq 1, \tag{29}$$

$$\sum_{k\in J}\frac{1}{M_k} < +\infty, \tag{30}$$

where $c_3 > 1$ and $c_4 < 1$ are positive constants. Then

$$\sum_{k=1}^{\infty}\frac{1}{M_k} < +\infty. \tag{31}$$

Proof See Lemma 3.4 in Yuan [22]. □

The proof of the next lemma is based on the proof of Lemma 3.3 in Dai and Xu [6].

Lemma 3 *The conditions of Lemma 1, including bound (13), imply that (31) holds for M_k defined by (15).*

Proof Consider the set $J_2 = \{k \mid \rho_k \geq \eta_2\}$. From update rule (12) and definition (15) of M_k, it follows that

$$\delta_{k+1}^{\alpha} \leq \gamma_4^{\alpha}\delta_k^{\alpha}, \quad k \in J_2, \tag{32}$$

$$\delta_{k+1}^{\alpha} \leq \gamma_3^{\alpha}\delta_k^{\alpha}, \quad k \notin J_2, \tag{33}$$

$$M_{k+1} \geq M_k, \quad k \geq 1, \tag{34}$$

where $\gamma_4^\alpha > 1$ and $\gamma_3^\alpha < 1$ are positive constants. Moreover, since $\{f(x_k)\}$ is non-increasing and bounded below, definition of J_2, Conditions A5 and A4, bound (13), Eq. (15) and Lemma 1 imply that

$$\begin{aligned}
+\infty &> \sum_{k=1}^{\infty}(f(x_k) - f(x_{k+1})) \geq \sum_{k\in J_2}(f(x_k) - f(x_{k+1}))\\
&\geq \eta_2 \sum_{k\in J_2}[m_k(x_k) - m_k(x_k + s_k)]\\
&\geq \eta_2 \sum_{k\in J_2}\kappa_c \varepsilon \min\left\{\frac{\varepsilon}{1+\|H_k\|}, \delta_k^\alpha \varepsilon^\beta\right\}\\
&\geq \eta_2 \kappa_c \varepsilon \sum_{k\in J_2}\min\left\{\frac{\varepsilon}{M_k}, \frac{\tau}{M_k}\varepsilon^\beta\right\}\\
&= \eta_2 \kappa_c \varepsilon \min\left\{\varepsilon, \tau\varepsilon^\beta\right\}\sum_{k\in J_2}\frac{1}{M_k}.
\end{aligned} \tag{35}$$

Therefore, $\sum_{k\in J_2} 1/M_k < +\infty$ and, by Lemma 2 (with $\mu_k = \delta_k^\alpha$, $J = J_2$, $c_3 = \gamma_4^\alpha$ and $c_4 = \gamma_3^\alpha$), we conclude that this lemma is true. □

Now, we are ready to give the global convergence results. The proof of the next theorem is based on the proofs of Theorem 3.4 in Toint [19] and Theorem 3.5 in Yuan [22].

Theorem 1 *Suppose that A1–A7 hold. Moreover, assume that*

$$\sum_{k=1}^{\infty} 1/M_k = +\infty, \tag{36}$$

with M_k defined by (15). Then,

$$\liminf_{k\to+\infty} \omega_k = 0, \tag{37}$$

or

$$\liminf_{k\to+\infty} \min\{\phi_k, \psi_k, \chi_k\} = 0. \tag{38}$$

Therefore, at least one limit point of the sequence $\{x_k\}$ (if any exists) is a stationary point of f.

Proof If (37) holds, then the conclusion follows from A1. Otherwise, there exist $\varepsilon_\omega > 0$ such that

$$\omega_k \geq \varepsilon_\omega \text{ for all } k. \tag{39}$$

In this case, if (38) holds, the conclusion follows from A2. Thus, suppose by contradiction that the bound (13) is true for some $\varepsilon > 0$. Then, by Lemma 3, we have that (31) holds for M_k defined by (15), which contradicts assumption (36). This contradiction shows that this theorem is true. □

Corollary 1 *Suppose that A1–A7 hold. If all the matrices H_k satisfy (11), then at least one limit point of the sequence $\{x_k\}$ (if any exists) is a stationary point of f.*

Proof Indeed, it follows from (15) and (11) that

$$M_k \leq 1 + \max_{1\leq i\leq k} \{c_1 + c_2 i\} \leq (1 + c_1 + c_2)\, k \Longrightarrow \frac{1}{(1 + c_1 + c_2)} \frac{1}{k} \leq \frac{1}{M_k}.$$

Since the harmonic series $\sum_{k=1}^{\infty} 1/k$ is divergent, from the comparison test for numeric series it follows that $\sum_{k=1}^{\infty} 1/M_k = +\infty$. Hence, the conclusion follows from Theorem 1. □

3 Worst-case complexity analysis

This section is divided in two parts. In the first subsection, an iteration complexity bound is obtained for Algorithm 1 with update rule (12). Then, under additional conditions, an improved complexity bound is given in the second subsection. Although the convergence of Algorithm 1 has been proved under bound (11), to obtain the complexity bounds we shall consider the stronger Condition:

A8 There exists a constant $\kappa_H > 0$ such that $\|H_k\| \leq \kappa_H$ for all k.

By convenience, 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:

$$S = \{k \geq 1 \,|\, k \text{ successful}\}, \tag{40}$$
$$S_j = \{k \leq j \,|\, k \in S\}, \text{ for each } j \geq 1, \tag{41}$$
$$U_j = \{k \leq j \,|\, k \notin S\} \text{ for each } j \geq 1, \tag{42}$$
$$F_k = \min\{\omega_k, \phi_k, \psi_k, \chi_k\}, \; k \geq 1, \text{ and} \tag{43}$$
$$S_F^{\varepsilon} = \{k \in S \,|\, F_k > \varepsilon\}, \; \varepsilon > 0, \tag{44}$$

where S_j and U_j form a partition of $\{1, \ldots, j\}$, and $|S_j|$, $|U_j|$ and $|S_F^{\varepsilon}|$ will denote the cardinality of these sets. Furthermore, let S_0 be a generic index set such that

$$S_0 \subseteq S_F^{\varepsilon}, \tag{45}$$

and whose cardinality is denoted by $|S_0|$.

3.1 General case

We start considering the lemma below, which provides an upper bound on $|S_0|$.

Lemma 4 *Let $\{f(x_k)\}$ be bounded below by f_{low}. Given any $\varepsilon > 0$, let S_F^ε and S_0 be defined in (44) and (45), respectively. Suppose that the successful iterates x_k generated by Algorithm 1 have the property that*

$$m_k(x_k) - m_k(x_k + s_k) \geq \alpha_c \varepsilon^p, \text{ for all } k \in S_0, \tag{46}$$

where α_c is a positive constant independent of k and ε, and $p > 0$. Then,

$$|S_0| \leq \left\lceil \kappa_p \varepsilon^{-p} \right\rceil, \tag{47}$$

where $\kappa_p \equiv (f(x_1) - f_{low}) / (\eta_1 \alpha_c)$.

Proof See Theorem 2.2 in Cartis et al. [3]. □

Remark 3 As pointed by Cartis et al. [3], if (46) holds with $S_0 = S_F^\varepsilon$, then (47) implies that Algorithm 1 takes at most $\lceil \kappa_p \varepsilon^{-p} \rceil$ successful iterations to generate an iterate k such that $F_{k+1} \leq \varepsilon$.

The next result gives a lower bound for δ_k^α and is crucial in the further analysis.

Lemma 5 *Suppose that A1–A8 hold and $\varepsilon \in (0, 1]$. If*

$$\min\{\phi_k, \psi_k, \chi_k\} \geq \varepsilon, \quad \text{for } k = 1, \ldots, j, \tag{48}$$

then there exists a constant $\bar{\tau} > 0$ independent of k and ε such that

$$\delta_k^\alpha \geq \frac{\bar{\tau}}{1 + \kappa_H}\varepsilon, \quad \text{for } k = 1, \ldots, j + 1. \tag{49}$$

Proof By the same argument used to prove Lemma 1, we can see that

$$\delta_k^\alpha \geq \frac{\tau}{M_k}, \quad \text{for } k = 1, \ldots, j + 1, \tag{50}$$

where M_k and τ are given by (15) and (16), respectively. Due to Condition A8, we have the inequality $M_k \leq 1 + \kappa_H$ for all k. On the other hand, the assumptions $\varepsilon \in (0, 1]$ and $\alpha \in (0, 1]$ imply that $\varepsilon^\alpha \geq \varepsilon$. Then, by (16) we obtain the inequality $\tau \geq \bar{\tau}\varepsilon$, where

$$\bar{\tau} = \min\left\{\delta_1^\alpha M_1, (\gamma_1 \kappa_\delta)^\alpha, \gamma_1^\alpha / \kappa_\chi^\beta, \gamma_1^\alpha \kappa_c (1 - \eta_2)/\kappa_\chi^\beta \kappa_m \kappa_s^2\right\}$$

is independent of k and ε. Hence, combining these two observations with (50) we conclude that

$$\delta_k^\alpha \geq \frac{\bar{\tau}}{1 + \kappa_H}\varepsilon, \quad \text{for } k = 1, \ldots, j + 1. \tag{51}$$

□

The next theorem provides an iteration complexity bound for Algorithm 1 with rule (12). Its proof is based on the proofs of Theorem 2.1 and Corollary 3.4 in Cartis et al. [3], and on the proof of Theorem 2.4 in Cartis et al. [4].

Theorem 2 *Let Conditions A1–A8 hold and* $\{f(x_k)\}$ *be bounded below by* f_{low}. *Given any* $\varepsilon \in (0, 1]$, *assume that* $F_1 > \varepsilon$ *and let* $j_1 \leq +\infty$ *be the first iteration such that* $F_{j_1+1} \leq \varepsilon$. *Then, Algorithm 1 with update rule (12) takes at most*

$$L_1^s \equiv \left\lceil \kappa_c^s \varepsilon^{-(2+\beta)} \right\rceil \tag{52}$$

successful iterations to generate $F_{j_1+1} \leq \varepsilon$, *where*

$$\kappa_c^s \equiv \left(f(x_1) - f_{low}\right) / \left(\eta_1 \alpha_c\right), \;\; \alpha_c \equiv \left(\kappa_c \min\{1, \bar{\tau}\}\right) / (1 + \kappa_H). \tag{53}$$

Furthermore,

$$j_1 \leq \left\lceil \kappa_d \varepsilon^{-(2+\beta)} \right\rceil \equiv L_1, \tag{54}$$

and so Algorithm 1 takes at most L_1 *(successful and unsuccessful) iterations to generate* $F_{j_1+1} \leq \varepsilon$, *where*

$$\kappa_d \equiv \left(1 - \frac{log(\gamma_4^{-\alpha})}{log(\gamma_2^{-\alpha})}\right)\kappa_c^s + \frac{(1+\kappa_H)\delta_1^{\alpha}}{\bar{\tau} log(\gamma_2^{-\alpha})}.$$

Proof The definition of j_1 in the statement of the Theorem implies that

$$\min\{\phi_k, \psi_k, \chi_k\} > \varepsilon, \quad \text{for} \quad k = 1, \ldots, j_1. \tag{55}$$

Thus, by A5, A4, (55), A8, Lemma 5 and the inequality $\varepsilon^{\beta} \leq 1$, we have

$$\begin{aligned} m_k(x_k) - m_k(x_k + s_k) &\geq \kappa_c \varepsilon \min\left\{\frac{\varepsilon}{1+\kappa_H}, \frac{\varepsilon \bar{\tau}}{1+\kappa_H}\varepsilon^{\beta}\right\} \\ &= \frac{\kappa_c \min\left\{1, \bar{\tau}\varepsilon^{\beta}\right\}}{1+\kappa_H}\varepsilon^2 \\ &\geq \frac{\kappa_c \min\{1, \bar{\tau}\}}{1+\kappa_H}\varepsilon^{2+\beta} \\ &= \alpha_c \varepsilon^{2+\beta}, \quad \text{for } k = 1, \ldots, j_1, \end{aligned} \tag{56}$$

where α_c is defined by (53). Now, with $j = j_1$ in (41) and (42), Lemma 4 with $S_0 = S_{j_1}$ and $p = 2 + \beta$ provides the complexity bound

$$|S_{j_1}| \leq L_1^s, \tag{57}$$

where L_1^s is defined by (52).

On the other hand, from rule (12) and Lemma 5 it follows that

$$\begin{aligned} \delta_{k+1}^{\alpha} &\leq \gamma_4^{\alpha}\delta_k^{\alpha}, \quad \text{if } k \in S_{j_1}, \\ \delta_{k+1}^{\alpha} &\leq \gamma_2^{\alpha}\delta_k^{\alpha}, \quad \text{if } k \in U_{j_1}, \\ \delta_k^{\alpha} &\geq \frac{\bar{\tau}}{1+\kappa_H}\varepsilon, \quad \text{for } k = 1, \ldots, j_1 + 1. \end{aligned}$$

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

$$\alpha_4 \nu_k \leq \nu_{k+1}, \quad \text{if} \quad k \in S_{j_1}, \tag{58}$$

$$\alpha_2 \nu_k \leq \nu_{k+1}, \quad \text{if} \quad k \in U_{j_1}, \tag{59}$$

$$\nu_k \leq \bar{\nu}\varepsilon^{-1}, \quad \text{for} \quad k = 1, \ldots, j_1 + 1, \tag{60}$$

where $\alpha_4 = \gamma_4^{-\alpha} \in (0, 1)$, $\alpha_2 = \gamma_2^{-\alpha} > 1$ and $\bar{\nu} = (1 + \kappa_H)/\bar{\tau}$. From (58) and (59) we deduce inductively

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

Hence, from (60) it follows that

$$\alpha_4^{|S_{j_1}|} \alpha_2^{|U_{j_1}|} \leq \frac{\bar{\nu}}{\nu_1}\varepsilon^{-1},$$

and so, taking logarithm on both sides, we get

$$|U_{j_1}| \leq \left[-\frac{log(\alpha_4)}{log(\alpha_2)}|S_{j_1}| + \frac{\bar{\nu}}{\nu_1 log(\alpha_2)}\varepsilon^{-1} \right]. \tag{61}$$

Finally, since $j_1 = |S_{j_1}| + |U_{j_1}|$ and $\varepsilon^{-(2+\beta)} \geq \varepsilon^{-1}$, the bound (54) is the sum of the upper bounds (57) and (61). $\square$

Remark 4 In words, Theorem 2 says that Algorithm 1 [with rule (12)] requires at most $O(\varepsilon^{-(2+\beta)})$ iterations to drive the optimality measure F_k below the desired accuracy ε, which is less pessimistic than the bound of $O(\varepsilon^{-3})$ iterations discussed by Toint [19]. This improved complexity result is due to the bound $\Delta(\delta_k, \chi_k) \geq O(\varepsilon^{1+\beta})$ derived from Lemma 5, which is sharper than the bound $\Delta(\delta_k, \chi_k) \geq O(\varepsilon^2)$ given by Lemma 3.2 in [19]. Table 1 below summarizes the complexity bounds obtained from Theorem 2 for the algorithms mentioned in Sect. 1.

3.2 Particular case

The generality of the nonlinear stepsize control framework and of our analysis may lead to pessimistic complexity results. An example is the ARC algorithm, for which the complexity bound of $O(\varepsilon^{-5/2})$ derived from Theorem 2 is worse than the bound

Table 1 Worst-case complexity bounds

Algorithm	β	Complexity bound
Classical trust-region [5,14]	0	$O(\varepsilon^{-2})$
ARC algorithm [2,3]	1/2	$O(\varepsilon^{-5/2})$
Quadratic regularization [1,13]	1	$O(\varepsilon^{-3})$
Trust-region of Fan and Yuan [9]	1	$O(\varepsilon^{-3})$
Quadratic regularization [8,24]	1	$O(\varepsilon^{-3})$

of $O(\varepsilon^{-2})$ obtained by Cartis et al. [3]. In this subsection, we shall refine the analysis to prove a complexity bound of $O(\varepsilon^{-2})$ for a subclass of methods represented by Algorithm 1, including the ARC algorithm. For that, we consider the additional Conditions:

A9 The powers α and β satisfy the inequality $\alpha+\beta\leq 1$.
A10 For all k, $\phi_k\geq\chi_k$ and $\psi_k\geq\chi_k$.
A11 The powers α and β satisfy the inequality $2\alpha+\beta\geq 1$.

The proof of the next lemma is based on the proof of Lemma 3.2 in Cartis et al. [3].

Lemma 6 *Let Conditions A1–A8 hold. Also, assume that*

$$\left(\frac{1}{\delta_k}\right)^{\alpha}\min\left\{\chi_k^{\alpha},\,\chi_k^{-\beta}\phi_k,\,\chi_k^{-\beta}\psi_k\right\}$$
$$>\max\left\{\frac{\kappa_m\kappa_s^2}{(1-\eta_2)\kappa_c},\,\frac{1+\kappa_H}{\kappa_\delta^{\alpha}},\,1+\kappa_H\right\}\equiv\kappa_{HB}. \tag{62}$$

Then, iteration k is very successful and consequently

$$\delta_{k+1}\geq\delta_k. \tag{63}$$

Proof Inequality (62) imply that $\min\{\phi_k,\psi_k,\chi_k\}>0$, and so, by Condition A5, we have $m_k(x_k)-m_k(x_k+s_k)>0$. Hence, it follows from (8) that

$$\rho_k>\eta_2\Longleftrightarrow r_k\equiv f(x_k+s_k)-f(x_k)-\eta_2\left[m_k(x_k+s_k)-m_k(x_k)\right]<0. \tag{64}$$

In order to prove (63), we shall derive a negative upper bound on r_k. First, from the equality $m_k(x_k)=f(x_k)$, note that

$$r_k=\left[f(x_k+s_k)-m_k(x_k+s_k)\right]+(1-\eta_2)\left[m_k(x_k+s_k)-f(x_k)\right]. \tag{65}$$

A bound for the first term in (65) is given by Condition A7:

$$f(x_k+s_k)-m_k(x_k+s_k)\leq\kappa_m\|s_k\|^2. \tag{66}$$

On the other hand, by (62), Condition A8 and (15), we have

$$\left(\frac{1}{\delta_k}\right)^{\alpha}\chi_k^{\alpha} > \frac{1+\kappa_H}{\kappa_\delta^{\alpha}} \Longrightarrow \kappa_\delta^{\alpha}\chi_k^{\alpha} > \delta_k^{\alpha}(1+\kappa_H) \geq \delta_k^{\alpha}M_k$$
$$\Longrightarrow \delta_k M_k^{\frac{1}{\alpha}} < \kappa_\delta\chi_k.$$

Hence, Condition A6 implies that $\|s_k\| \leq \kappa_s\Delta(\delta_k, \chi_k)$, which together with (66) and Condition A4 gives

$$f(x_k+s_k) - m_k(x_k+s_k) \leq \kappa_m\kappa_s^2\delta_k^{2\alpha}\chi_k^{2\beta}. \tag{67}$$

Regarding the second difference in (65), from (62) and A8 note that

$$\left(\frac{1}{\delta_k}\right)^{\alpha}\chi_k^{-\beta}\phi_k > 1+\kappa_H \geq 1+\|H_k\|,$$

and so

$$\Delta(\delta_k, \chi_k) = \delta_k^{\alpha}\chi_k^{\beta} = \frac{\phi_k}{\left(\frac{1}{\delta_k}\right)^{\alpha}\chi_k^{-\beta}\phi_k} < \frac{\phi_k}{1+\|H_k\|}.$$

Consequently, by Conditions A5 and A7, we obtain

$$f(x_k) - m_k(x_k+s_k) \geq \kappa_c\psi_k\Delta(\delta_k, \chi_k) = \kappa_c\psi_k\delta_k^{\alpha}\chi_k^{\beta}$$
$$\Longrightarrow m_k(x_k+s_k) - f(x_k) \leq -\kappa_c\delta_k^{\alpha}\chi_k^{\beta}\psi_k. \tag{68}$$

Now, (65), (67) and (68) provide the following upper bound for r_k, namely,

$$r_k \leq \kappa_m\kappa_s^2\delta_k^{2\alpha}\chi_k^{2\beta} - (1-\eta_2)\kappa_c\delta_k^{\alpha}\chi_k^{\beta}\psi_k$$
$$= \delta_k^{2\alpha}\chi_k^{2\beta}\left[\kappa_m\kappa_s^2 - (1-\eta_2)\kappa_c\left(\frac{1}{\delta_k}\right)^{\alpha}\chi_k^{-\beta}\psi_k\right].$$

But, it follows from (62) that

$$\left(\frac{1}{\delta_k}\right)^{\alpha}\chi_k^{-\beta}\psi_k > \frac{\kappa_m\kappa_s^2}{(1-\eta_2)\kappa_c} \Longrightarrow (1-\eta_2)\kappa_c\left(\frac{1}{\delta_k}\right)^{\alpha}\chi_k^{-\beta}\psi_k > \kappa_m\kappa_s^2$$
$$\Longrightarrow \kappa_m\kappa_s^2 - (1-\eta_2)\kappa_c\left(\frac{1}{\delta_k}\right)^{\alpha}\chi_k^{-\beta}\psi_k < 0.$$

Thus, $r_k < 0$, which means that k is very successful. Therefore, (63) follows from (12). □

The next lemma gives a lower bound on δ_k when F_k is bounded away from zero. Its proof is based on the proof of Lemma 3.3 in Cartis et al. [3].

Lemma 7 *Let Conditions A1–A10 hold. Also, let* $\varepsilon \in (0, 1]$ *such that* $F_k > \varepsilon$ *for all* $k = 1, \dots, j$*, where* $j \leq +\infty$*. Then, there exists* $\bar{\tau} > 0$ *independent of* k *and* ε *such that*

$$\delta_k \geq \bar{\tau}\varepsilon^{(1-\beta)/\alpha} \quad \text{for } k = 1, \dots, j+1. \tag{69}$$

Proof First, by induction, we shall prove that

$$\delta_k \geq \min\left\{\delta_1, \frac{\gamma_1}{\kappa_{HB}^{1/\alpha}}\varepsilon^{(1-\beta)/\alpha}\right\}, \tag{70}$$

for $k = 1, \dots, j+1$, where κ_{HB} is defined as the constant in the right hand side of (62). Clearly, (70) is true for $k = 1$. We assume that (70) is true for $k \in \{1, \dots, j\}$ and prove it is also true for $k+1$. From inequalities $\phi_k, \psi_k \geq \chi_k$ (due to A10), $F_k > \varepsilon \in (0, 1]$ and $0 < \alpha \leq (1-\beta)$ (due to A9), it follows that

$$\begin{aligned}\min\left\{\chi_k^{\alpha}, \chi_k^{-\beta}\phi_k, \chi_k^{-\beta}\psi_k\right\} &\geq \min\left\{\chi_k^{\alpha}, \chi_k^{(1-\beta)}\right\} \\ &\geq \min\left\{\varepsilon^{\alpha}, \varepsilon^{(1-\beta)}\right\} \\ &= \varepsilon^{(1-\beta)}\end{aligned} \tag{71}$$

Therefore, by (62), Lemma 6 and the induction assumption, if

$$\left(\frac{1}{\delta_k}\right)^{\alpha}\varepsilon^{(1-\beta)} > \kappa_{HB}, \tag{72}$$

then

$$\delta_{k+1} \geq \delta_k \geq \min\left\{\delta_1, \frac{\gamma_1}{\kappa_{HB}^{1/\alpha}}\varepsilon^{(1-\beta)/\alpha}\right\}, \tag{73}$$

and so, (70) is true for $k+1$.

Now, suppose that (72) is not true. Then

$$\begin{aligned}\left(\frac{1}{\delta_k}\right)^{\alpha}\varepsilon^{(1-\beta)} \leq \kappa_{HB} &\Longrightarrow \frac{1}{\delta_k}\varepsilon^{(1-\beta)/\alpha} \leq \kappa_{HB}^{1/\alpha} \\ &\Longrightarrow \delta_k \geq \frac{1}{\kappa_{HB}^{1/\alpha}}\varepsilon^{(1-\beta)/\alpha}\end{aligned}$$

and by rule (12) we see that (70) is true for $k+1$:

$$\delta_{k+1} \geq \gamma_1\delta_k \geq \frac{\gamma_1}{\kappa_{HB}^{1/\alpha}}\varepsilon^{(1-\beta)/\alpha} \geq \min\left\{\delta_1, \frac{\gamma_1}{\kappa_{HB}^{1/\alpha}}\varepsilon^{(1-\beta)/\alpha}\right\}.$$

Finally, since $\varepsilon^{(1-\beta)/\alpha} \leq 1$, by (70) we conclude that, for $k = 1, \dots, j+1$,

$$\delta_k \geq \min\left\{\delta_1, \frac{\gamma_1}{\kappa_{HB}^{1/\alpha}}\right\}\varepsilon^{(1-\beta)/\alpha} = \bar{\tau}\varepsilon^{(1-\beta)/\alpha}, \tag{74}$$

where $\bar{\tau}$ is independent of k and ε. □

We are now ready to obtain an iteration complexity bound in this particular case. The proof is a direct adaptation of the proof of Theorem 2.

Theorem 3 *Let Conditions A1–A10 hold and* $\{f(x_k)\}$ *be bounded below by* f_{low}. *Given any* $\varepsilon \in (0, 1]$, *assume that* $F_1 > \varepsilon$ *and let* $j_1 \leq +\infty$ *be the first iteration such that* $F_{j_1+1} \leq \varepsilon$. *Then, Algorithm 1 with update rule (12) takes at most*

$$L_1^s \equiv \left\lceil \kappa_c^s \varepsilon^{-2} \right\rceil \tag{75}$$

successful iterations to generate $F_{j_1+1} \leq \varepsilon$, *where*

$$\kappa_c^s \equiv (f(x_1) - f_{low}) / (\eta_1 \alpha_c), \;\; \alpha_c = \kappa_c \min\left\{1/(1+\kappa_H), \bar{\tau}^{\alpha}\right\}. \tag{76}$$

Additionally, assume that Condition A11 holds. Then,

$$j_1 \leq \left\lceil \kappa_d \varepsilon^{-2} \right\rceil \equiv L_1, \tag{77}$$

and so Algorithm 1 takes at most L_1 *(successful and unsuccessful) iterations to generate* $F_{j_1+1} \leq \varepsilon$, *where*

$$\kappa_d \equiv \left(1 - \frac{log(\gamma_4^{-1})}{log(\gamma_2^{-1})}\right)\kappa_c^s + \frac{\delta_1}{\bar{\tau} log(\gamma_2^{-1})}.$$

Proof The definition of j_1 in the statement of the Theorem implies that

$$\min\{\phi_k, \psi_k, \chi_k\} > \varepsilon, \quad \text{for} \quad k = 1, \ldots, j_1. \tag{78}$$

Thus, by A5, A4, (78), A8 and Lemma 7,

$$\begin{aligned} m_k(x_k) - m_k(x_k + s_k) &\geq \kappa_c \varepsilon \min\left\{\frac{\varepsilon}{1+\kappa_H}, \left(\bar{\tau}\varepsilon^{(1-\beta)/\alpha}\right)^{\alpha} \varepsilon^{\beta}\right\} \\ &= \kappa_c \varepsilon \min\left\{\frac{\varepsilon}{1+\kappa_H}, \bar{\tau}^{\alpha}\varepsilon\right\} \\ &= \kappa_c \min\left\{\frac{1}{1+\kappa_H}, \bar{\tau}^{\alpha}\right\}\varepsilon^2 \\ &= \alpha_c \varepsilon^2, \quad \text{for } k = 1, \ldots, j_1, \end{aligned} \tag{79}$$

where α_c is defined by (76). Now, with $j = j_1$ in (41) and (42), Lemma 4 with $S_0 = S_{j_1}$ and $p = 2$ provides the complexity bound

$$|S_{j_1}| \leq L_1^s, \tag{80}$$

where L_1^s is defined by (75).

On the other hand, from rule (12) and Lemma 7 it follows that

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

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

$$\alpha_4 \nu_k \leq \nu_{k+1}, \quad \text{if} \quad k \in S_{j_1}, \tag{81}$$

$$\alpha_2 \nu_k \leq \nu_{k+1}, \quad \text{if} \quad k \in U_{j_1}, \tag{82}$$

$$\nu_k \leq \bar{\nu} \varepsilon^{-(1-\beta)/\alpha}, \quad \text{for} \quad k = 1, \ldots, j_1 + 1, \tag{83}$$

where $\alpha_4 = \gamma_4^{-1} \in (0, 1)$, $\alpha_2 = \gamma_2^{-1} > 1$ and $\bar{\nu} = \bar{\tau}^{-1}$. From (81) and (82) we deduce inductively

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

Hence, from (83) it follows that

$$\alpha_4^{|S_{j_1}|} \alpha_2^{|U_{j_1}|} \leq \frac{\bar{\nu}}{\nu_1} \varepsilon^{-(1-\beta)/\alpha},$$

and so, taking logarithm on both sides, we get

$$|U_{j_1}| \leq \left[-\frac{log(\alpha_4)}{log(\alpha_2)} |S_{j_1}| + \frac{\bar{\nu}}{\nu_1 log(\alpha_2)} \varepsilon^{-(1-\beta)/\alpha} \right]. \tag{84}$$

Finally, since $j_1 = |S_{j_1}| + |U_{j_1}|$ and $\varepsilon^{-2} \geq \varepsilon^{-(1-\beta)/\alpha}$ (due to Condition A11), the bound (77) is the sum of the upper bounds (80) and (84). □

Remark 5 Note that Conditions A9–A11 are satisfied for the ARC algorithm [2,3], where $\alpha = \beta = 1/2$ and $\phi = \psi = \chi$. Hence, Theorem 3 provides the known complexity bound of $O(\varepsilon^{-2})$ iterations for this algorithm.

The complexity result given by Theorem 3 is due to the bound $\Delta(\delta_k, \chi_k) \geq O(\varepsilon)$ derived from Lemma 7. It is worth notice that, with a specific analysis, such bound can also be obtained for the quadratic regularization algorithm in [1]. In fact, as we saw in Sect. 1, for this algorithm we have $\alpha = 1$, $\beta = 1$ and $\delta_k = 1/\sigma_k$, where σ_k is the quadratic regularization parameter. Suppose that $\chi_k \geq \varepsilon$, for some $\varepsilon \in (0, 1]$. By Lemma 4.7 in [1], there exists $\sigma_{max} > 0$, independent of k and ε, such that $\sigma_k \leq \sigma_{max}$. Consequently, $\delta_k \geq 1/\sigma_{max}$, and so $\Delta(\delta_k, \chi_k) \geq O(\varepsilon)$. Thus, by the same argument used above, we can prove a complexity bound of $O(\varepsilon^{-2})$ iterations for the quadratic regularization algorithm in [1], which is better than the bound of $O(\varepsilon^{-3})$ derived from Theorem 2 (see Table 1). This is another example of the lack of sharpness in the complexity bound given by Theorem 2. However, it is unclear whether a specific analysis can be done to yield less pessimistic complexity estimates for the trust region algorithm of Fan and Yuan [9] or for the quadratic regularization algorithms in [8,24].

4 Extensions

As pointed by Toint [19], the nonlinear stepsize control framework is not limited to unconstrained optimization problems. It can be extended, for example, to projection-based trust-region algorithms for optimization of a function f (possibly non-convex) over a convex set C. In this section, we extend the nonlinear stepsize control approach and the results of the previous sections to algorithms for composite nonsmooth optimization and unconstrained multiobjective optimization.

Throughout this section, given a point $x \in \mathbb{R}^n$ and a number $r > 0$, we shall consider the following notation:

$$B[x,r] \equiv \left\{y \in \mathbb{R}^n \mid \|y-x\| \leq r\right\} \quad \text{and} \quad B(x,r) \equiv \left\{y \in \mathbb{R}^n \mid \|y-x\| < r\right\}.$$

4.1 Composite nonsmooth optimization

Consider the composite nonsmooth optimization (NSO) problem

$$\min_{x \in \mathbb{R}^n} f(x) \equiv g(x) + h(c(x)), \tag{85}$$

where $h : \mathbb{R}^m \to \mathbb{R}$ is convex but may be nonsmooth, and $g : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable.

Definition 1 (Yuan [23], page 271) A point x^* is said to be a stationary point of f if

$$g(x^*) + h(c(x^*)) \leq g(x^*) + \nabla g(x^*)^T s + h(c(x^*) + J_c(x^*)s), \ \forall s \in \mathbb{R}^n, \tag{86}$$

where J_c denotes the Jacobian of c.

For each $x \in \mathbb{R}^n$, define

$$l(x,s) \equiv g(x) + \nabla g(x)^T s + h(c(x) + J_c(x)s), \ \forall s \in \mathbb{R}^n. \tag{87}$$

Then, for all $r > 0$, let

$$\xi_r(x) \equiv l(x,0) - \min_{\|s\| \leq r} l(x,s). \tag{88}$$

Following Cartis et al. [4], as a stationarity measure for f, we shall use the quantity

$$\xi_1(x) \equiv l(x,0) - \min_{\|s\| \leq 1} l(x,s). \tag{89}$$

This choice is justified by the lemma below.

Lemma 8 *Let $\xi_1 : \mathbb{R}^n \to \mathbb{R}$ be defined by (89), and let S be a bounded subset of $\mathbb{R}^n$. Suppose that $h : \mathbb{R}^m \to \mathbb{R}$ is convex, and that $g : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. Then:*

(a) ξ_1 is continuous on S;
(b) $\xi_1(x) \geq 0$ for all $x \in \mathbb{R}^n$;
(c) x^* is a stationary point of $f \Longleftrightarrow \xi_1(x^*) = 0$.

Proof See Lemma 2.1 in [22]. □

Now, let us consider the following trust-region algorithm, which is a modification of the model algorithm proposed by Fletcher [10].

Algorithm 2. (Trust-Region Algorithm for Composite NSO)

Step 0 Given $x_1 \in \mathbb{R}^n$, $H_1 \in \mathbb{R}^{n\times n}$ symmetric, $\bar{\Delta} > 0$, $\Delta_1 \in \left(0, \bar{\Delta}\right]$, $0 < \gamma_1 < \gamma_2 < \gamma_3 < 1 < \gamma_4$ and $0 < \eta_1 \leq \eta_2 < 1$, set $k := 1$.

Step 1 Let s_k^* be a solution of the subproblem

$$\min_{s\in\mathbb{R}^n} \; m_k(x_k+s) \equiv g(x_k)+\nabla g(x_k)^T s+h(c(x_k)+J_c(x_k)s)+\frac{1}{2}s^T H_k s, \tag{90}$$

$$\text{s. t.} \quad \|s\| \leq \Delta_k. \tag{91}$$

Compute a step s_k for which $\|s_k\| \leq \Delta_k$ and

$$m_k(x_k) - m_k(x_k + s_k) \geq \gamma_0 \left[m_k(x_k) - m_k(x_k + s_k^*)\right], \tag{92}$$

where $\gamma_0 \in (0, 1)$ is a constant independent of k.

Step 2 Compute the ratio

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}, \tag{93}$$

set the next iterate

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{if } \rho_k \geq \eta_1, \\ x_k, & \text{otherwise,} \end{cases} \tag{94}$$

and choose the trust-region radius Δ_{k+1} by the update rule

$$\Delta_{k+1} \in \begin{cases} \left[\gamma_1\Delta_k, \gamma_2\Delta_k\right], & \text{if } \rho_k < \eta_1, \\ \left[\gamma_2\Delta_k, \gamma_3\Delta_k\right], & \text{if } \rho_k \in [\eta_1, \eta_2), \\ \left[\Delta_k, \min\left\{\gamma_4\Delta_k, \bar{\Delta}\right\}\right], & \text{if } \; \rho_k \geq \eta_2. \end{cases} \tag{95}$$

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

Remark 6 The matrix H_k is an $n \times n$ symmetric matrix approximating the second order behavior of g and c in a neighbourhood of x_k [for example, see equation (1.5) in [23]]. If $H_k = 0$ for all k, and the step s_k is a solution of the subproblem (90)–(91), then Algorihm 2 reduces to an instance of the first-order trust-region algorithm proposed by Cartis et al. [4]. On the other hand, when $g = 0$, Algorithm 2 reduces to an instance of the trust-region algorithms of Powell [15] and Yuan [22] for the problem $\min_{x\in\mathbb{R}^n} h(c(x))$. Finally, note that (95) is a particular case of (12).

We claim that, under suitable conditions and replacing the stationarity measure $\|\nabla f(x)\|$ by $\xi_1(x)$ in A1 and A2, Algorithm 2 is a particular case of Algorithm 1 with the choices

$$\begin{gathered} m_k(x_k+s) \equiv g(x_k) + \nabla g(x_k)^T s + h(c(x_k) + J_c(x_k)s) + \frac{1}{2}s^T H_k s, \\ \omega(x) = 1, \quad \phi(x) = \psi(x) = \chi(x) = \xi_1(x), \\ \delta_k = \Delta_k, \quad \alpha = 1, \quad \beta = 0. \end{gathered} \tag{96}$$

Specifically, we assume the conditions below:

C1 The function $h : \mathbb{R}^m \to \mathbb{R}$ is convex and globally Lipschitz continuous, with Lipschitz constant L_h;
C2 The functions $g : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable;
C3 The gradient function of g, $\nabla g : \mathbb{R}^n \to \mathbb{R}^n$, and the Jacobian function of c, $J_c : \mathbb{R}^n \to \mathbb{R}^{m\times n}$, are Lipschitz continuous on $[x_k, x_k + s_k]$ for all k, with constants $L_g \geq 1$ and L_J, respectively;
C4 There exists a constant $\kappa_H > 0$ such that $\|H_k\| \leq \kappa_H$ for all k;
C5 There exists a bounded set $S \subset \mathbb{R}^n$ such that x_k and $x_k + s_k$ belong to S for all k.

Let us now justify our claim. Conditions A1 and A2 follow from Lemma 8(c), while Condition A3 is satisfied due C5 and the continuity of ξ_1 [Lemma 8(a)]. On the other hand, by Step 1 in Algorithm 2, we have $\|s_k\| \leq \Delta_k$ for all k. Hence, A4 and A6 are naturally satisfied. Regarding A7, from (90) it follows that $m_k(x_k) = f(x_k)$. The second part of A7 is provided by the following result.

Lemma 9 *Suppose that C1–C5 hold. Then, there exists a constant $\kappa_m > 0$ such that for all k,*

$$f(x_k + s_k) - m_k(x_k + s_k) \leq \kappa_m \|s_k\|^2. \tag{97}$$

Proof By C1 and C4, we have

$$\begin{aligned} f(x_k+s_k) - m_k(x_k+s_k) \leq & \; |g(x_k+s_k) - g(x_k) - \nabla g(x_k)^T s_k| + \frac{\kappa_H}{2}\|s_k\|^2 \\ & + L_h\|c(x_k+s_k) - c(x_k) - J_c(x_k)s_k\|. \end{aligned} \tag{98}$$

On the other hand, from C2 and C3 it follows that[1]

$$|g(x_k+s_k) - g(x_k) - \nabla g(x_k)^T s_k| \leq \frac{L_g}{2}\|s_k\|^2 \tag{99}$$

[1] See Theorem 1.2.22 in Sun and Yuan [18].

and

$$\|c(x_k+s_k)-c(x_k)-J_c(x_k)s_k\| \leq \frac{L_J}{2}\|s_k\|^2. \tag{100}$$

Now, combining (98)–(100), we obtain (97) with $\kappa_m = \left(L_g + L_h L_J + \kappa_H\right)/2$. □

It remains to prove that A5 is satisfied. For that, we consider the following lemma.

Lemma 10 *Suppose that C1 holds and let $r > 0$. Then, for all x*

$$\xi_r(x) \geq \min\{1, r\}\,\xi_1(x). \tag{101}$$

Proof See Lemma 2.1 in Cartis et al. [4]. □

Now, Condition A5 follows from the theorem below. Its proof is based on the proof of Lemma 2.2 in Yuan [22].

Lemma 11 *Suppose that C1 holds. Then, there exists a constant $\kappa_c \in (0, 1)$ such that, for all k,*

$$m_k(x_k) - m_k(x_k+s_k) \geq \kappa_c \xi_1(x_k) \min\left\{\frac{\xi_1(x_k)}{1+\|H_k\|}, \Delta_k\right\}. \tag{102}$$

Proof Let s_k^* be a solution of subproblem (90)–(91). Then, for all $s \in B[0, \Delta_k]$,

$$m_k(x_k) - m_k(x_k+s_k^*) \geq m_k(x_k) - m_k(x_k+s). \tag{103}$$

Since h is continuous (by C1), $l(x, .)$ is also continuous. Then, by the Weierstrass Theorem, there exists $\tilde{s}_k \in B[0, \Delta_k]$ such that

$$\min_{\|s\|\leq \Delta_k} l(x_k, s) = l(x_k, \tilde{s}_k). \tag{104}$$

Now, using (103), (104), the convexity of h and the Cauchy-Schwarz inequality, for all $\theta \in [0, 1]$ we obtain:

$$\begin{aligned} m_k(x_k) - m_k(x_k+s_k^*) &\geq m_k(x_k) - m_k(x_k+\theta\tilde{s}_k) \\ &\geq \theta\left[l(x_k, 0) - l(x_k, \tilde{s}_k)\right] - \frac{1}{2}\|H_k\|\Delta_k^2\theta^2 \\ &\geq \theta\xi_{\Delta_k}(x_k) - \frac{1}{2}(1+\|H_k\|)\Delta_k^2\theta^2. \end{aligned} \tag{105}$$

As inequality (105) holds for all $\theta \in [0, 1]$, we conclude that

$$\begin{aligned} m_k(x_k) - m_k(x_k+s_k^*) &\geq \sup_{0\leq\theta\leq 1}\left\{\theta\xi_{\Delta_k}(x_k) - \frac{1}{2}(1+\|H_k\|)\Delta_k^2\theta^2\right\} \\ &\geq \frac{1}{2}\min\left\{\xi_{\Delta_k}(x_k), \frac{\left[\xi_{\Delta_k}(x_k)\right]^2}{(1+\|H_k\|)\Delta_k^2}\right\}. \end{aligned} \tag{106}$$

From (92), (106) and Lemma 10, it follows that

$$m_k(x_k) - m_k(x_k + s_k) \geq \frac{\gamma_0}{2} \min\{1, \Delta_k\} \xi_1(x_k) \min\left\{1, \frac{\min\{1, \Delta_k\} \xi_1(x_k)}{(1 + \|H_k\|)\Delta_k^2}\right\}. \tag{107}$$

If $\Delta_k \leq 1$, then (107) reduces to (102) with $\kappa_c = \gamma_0/2$. Without loss of generality, assume $\bar{\Delta} > 1$. In this case, if $\Delta_k \geq 1$, it follows from (107) and $\Delta_k \leq \bar{\Delta} \leq \bar{\Delta}^2$ that

$$m_k(x_k) - m_k(x_k + s_k) \geq \frac{\gamma_0}{2\bar{\Delta}^2} \xi_1(x_k) \min\left\{\Delta_k, \frac{\xi_1(x_k)}{1 + \|H_k\|}\right\},$$

which gives (102) with $\kappa_c = \gamma_0/2\bar{\Delta}^2$. Therefore, we obtain (102) with the constant $\kappa_c = \min\left\{\gamma_0, \gamma_0/\bar{\Delta}^2\right\}/2$. □

Hence, Algorithm 2 is covered by Algorithm 1 [with update rule (12)]. By the theory presented in Sect. 3, we have the following worst-case complexity result for the composite NSO problem.

Theorem 4 *Let C1–C5 hold and $\{f(x_k)\}$ be bounded below by f_{low}. Then, to reduce the stationarity measure $\xi_1(x)$ below $\varepsilon \in (0, 1]$, Algorithm 2 takes at most $O(\varepsilon^{-2})$ iterations.*

Remark 7 The order of the complexity bound given above is the same as that proved by Cartis et al. [4] for a first-order trust-region method and a first-order quadratic regularization method, which require the exact solution of the subproblem on each iteration. However, the result presented here is more general, in the sense that Algorithm 2 may employ second-order information in the models $m_k(x_k + s)$ and requires only an approximate solution of the subproblem on each iteration.

As discussed by Toint [19], the nonlinear stepsize control framework also can be used to design new algorithms. In the case of the composite NSO problem, a generalization of Algorithm 2 is obtained from the relaxations $\alpha \in (0, 1]$ and $\beta \in [0, 1]$ in (96). By Theorems 2 and 3, such nonlinear stepsize control trust-region algorithm takes at most $O(\varepsilon^{-(2+\beta)})$ iterations to reduce $\xi_1(x)$ below $\varepsilon \in (0, 1]$, and this bound is reduced to $O(\varepsilon^{-2})$ when $\alpha + \beta \leq 1$ and $2\alpha + \beta \geq 1$.

4.2 Unconstrained multiobjective optimization

Let $\mathbb{R}^m_+ = \{z \in \mathbb{R}^m \mid z_i \geq 0, i = 1, \ldots, m\}$, $\mathbb{R}^m_{++} = \{z \in \mathbb{R}^m \mid z_i > 0, i = 1, \ldots, m\}$, and consider the relations $\succ$ and $\succ_w$ given, respectively by

$$y \succ x \iff y - x \in \mathbb{R}^m_+ - \{0\} \quad \text{and} \quad y \succ_w x \iff y - x \in \mathbb{R}^m_{++}.$$

In this subsection, we shall extend the nonlinear stepsize control framework to deal with the unconstrained multiobjective optimization (MOO) problem

$$\min_{x \in \mathbb{R}^n} f(x) \equiv (f_1(x), \ldots, f_m(x))^T. \tag{108}$$

Definition 2 (Guerraggio and Luc [12], page 619) Given a point $x^* \in \mathbb{R}^n$,

(a) x^* is said to be an efficient solution of (108) when there is no $y \in \mathbb{R}^n$ such that $f(x^*) \succ f(y)$;
(b) x^* is said to be a weakly efficient solution of (108) when there is no $y \in \mathbb{R}^n$ such that $f(x^*) \succ_w f(y)$; and
(c) x^* is said to be a local (or local weakly) efficient solution of (108) when there exists a neighborhood $N(x^*)$ of x^* for which there is no $y \in N(x^*)$ such that $f(x^*) \succ f(y)$ (or, respectively, $f(x^*) \succ_w f(y)$).

The theorem below gives a necessary condition for a point $x^* \in \mathbb{R}^n$ to be a local weakly efficient solution of (108).

Theorem 5 *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable function. If x^* is a local weakly efficient solution of (108), then*

$$range(J_f(x^*)) \cap (-\mathbb{R}^m_{++}) = \emptyset, \tag{109}$$

where $J_f(x^*) = \left[\nabla f_1(x^*) \ldots \nabla f_m(x^*)\right]^T \in \mathbb{R}^{m \times n}$.

Proof See Theorem 5.1 [item (ii)–(a)] in Guerraggio and Luc [12]. □

Definition 3 (Fliege and Svaiter [11], page 481) A point $x^* \in \mathbb{R}^n$ is said to be a Pareto critical point of f if it satisfies condition (109).

Remark 8 Note that when $m = 1$, the MOO problem (108) reduces to problem (1), local and weakly local efficient solutions correspond to local minimizers of f, and the Pareto criticality condition (109) implies the stationarity condition $\|\nabla f(x)\| = 0$.

In order to extend Algorithm 1 to solve unconstrained multiobjective optimization problems, we consider the conditions below:

A1′ There exists a continuous, bounded and non-negative function $\omega : \mathbb{R}^n \to \mathbb{R}$ such that, if $\omega(x) = 0$ then x is a Pareto critical point of f.
A2′ There exist three continuous non-negative functions $\phi, \psi, \chi : \mathbb{R}^n \to \mathbb{R}$, possibly undefined at roots of ω, such that, provided $\omega(x) > 0$, if $\min\{\phi(x), \psi(x), \chi(x)\} = 0$ then x is a Pareto critical point of f.
A3′ There exists $\kappa_\chi > 0$ such that $\chi_k \leq \kappa_\chi$ for all k.
A4′ The function $\Delta : [0, +\infty) \times [0, +\infty) \to \mathbb{R}$ defining the trust-region radius is of the form

$$\Delta(\delta, \chi) = \delta^\alpha \chi^\beta, \tag{110}$$

for some powers $\alpha \in (0, 1]$ and $\beta \in [0, 1]$;

A5′ The step s_k produces a decrease in the model, which is sufficient in the sense that

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_c \psi_k \min\left\{\frac{\phi_k}{1 + \|H_k\|}, \Delta(\delta_k, \chi_k)\right\}, \tag{111}$$

for some constant $\kappa_c \in (0, 1)$, and where H_k is an $n \times n$ symmetric matrix approximating the second order behavior of the functions f_i in a neighbourhood of x_k (the connection between H_k and the Hessians of the functions f_i will depend of the model m_k);

A6′ The step s_k satisfies the bound

$$\|s_k\| \leq \kappa_s \Delta(\delta_k, \chi_k), \text{ whenever } \delta_k M_k^{\frac{1}{\alpha}} \leq \kappa_\delta \chi_k, \tag{112}$$

for some constants $\kappa_s \geq 1$ and $\kappa_\delta > 0$, where M_k is defined by (15).

A7′ For all $k \geq 1$, the model $m_k(x_k + s) : \mathbb{R}^n \to \mathbb{R}$ and the merit function $\Phi : \mathbb{R}^n \to \mathbb{R}$ satisty

$$m_k(x_k) = \Phi(x_k) \quad \text{and} \quad \Phi(x_k + s_k) - m_k(x_k + s_k) \leq \kappa_m \|s_k\|^2, \tag{113}$$

where $\kappa_m > 0$ is a constant.

Clearly, when $m = 1$, conditions A1′–A7′ are reduced to conditions A1–A7 with $\Phi(x) = f(x)$ (where the correspondence between A1′–A2′ and A1–A2 follows from Remark 10). Hence, Algorithm 1 can be generalized to MOO problems in the following way.

Algorithm 3. (Nonlinear Stepsize Control Algorithm for unconstrained MOO)

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

Step 1 Choose a model $m_k(x_k + s)$ and a function Φ satisfying A7′ and find a step s_k which sufficiently reduces the model in the sense of A5′ for which $\|s_k\|$ satisfies A6′.

Step 2 Compute the ratio

$$\rho_k = \frac{\Phi(x_k) - \Phi(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}, \tag{114}$$

set the next iterate

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{if } \rho_k \geq \eta_1, \\ x_k, & \text{otherwise,} \end{cases} \tag{115}$$

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

$$\delta_{k+1} \in \begin{cases} \left[\gamma_1 \delta_k, \gamma_2 \delta_k\right], & \text{if } \rho_k < \eta_1, \\ \left[\gamma_2 \delta_k, \delta_k\right], & \text{if } \rho_k \in [\eta_1, \eta_2), \\ [\delta_k, +\infty], & \text{if } \rho_k \geq \eta_2. \end{cases} \tag{116}$$

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

Remark 9 Note that Algorithm 3 is nothing else than Algorithm 1 applied to minimize the merit function Φ. This function is used as a scalar representation of the vector function f, and the relation between Φ and f is given implicitly by conditions A7′, A5′ and A2′.

Proceeding as in Sects. 2 and 3 (with Φ in place of f), we obtain the following results, which generalize Corollary 1 and Theorems 2 and 3.

Theorem 6 *Suppose that A1′–A7′ hold and let $\{x_k\}$ be a sequence generated by Algorithm 3 with update rule (12). If $\{\Phi(x_k)\}$ is bounded below and all the matrices H_k satisfy (11), then at least one limit point of $\{x_k\}$ (if any exists) is a Pareto critical point of f.*

Theorem 7 *Suppose that A1′–A7′ and A8 hold, and let $\{\Phi(x_k)\}$ be bounded below by Φ_{low}. Then, to reduce the Pareto criticality measure $F_k = \min\{\omega_k, \phi_k, \psi_k, \chi_k\}$ below $\varepsilon \in (0, 1]$, Algorithm 3 with update rule (12) takes at most $O(\varepsilon^{-(2+\beta)})$ iterations. If, additionaly, A9–A11 are satisfied, then this worst-case complexity bound is reduced to $O(\varepsilon^{-2})$ iterations.*

To justify our generalization of the nonlinear stepsize control algorithm, we need provide at least one non-trivial special case of Algorithm 3. For this purpose, we shall consider the trust-region method for unconstrained MOO recently proposed by Villacorta et al. [21], which is called TRMP algorithm. First, let $I = \{1, \ldots, m\}$ and define the function $\mu : \mathbb{R}^n \to \mathbb{R}$ by

$$\mu(x) \equiv -\min_{\|d\|\leq 1}\left(\max_{i\in I}\left\{\nabla f_i(x)^T d\right\}\right). \tag{117}$$

The next result, due to Fliege and Svaiter [11], provides some useful properties of the function μ and establishes its relation with the concept of Pareto critical points.

Lemma 12 *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable and $\mu : \mathbb{R}^n \to \mathbb{R}$ be defined by (117). Then,*

(a) μ is continuous;
(b) $\mu(x) \geq 0$ for all $x \in \mathbb{R}^n$;
(c) x^* is a Pareto critical point of $f \Longleftrightarrow \mu(x^*) = 0$.

Proof See Lemma 3 in Fliege and Svaiter [11]. □

Now, the TRMP Algorithm can be summarized in the following way.

Algorithm 4. (TRMP Algorithm for unconstrained MOO)

Step 0 Given $x_1 \in \mathbb{R}^n$, $H_1 \in \mathbb{R}^{n\times n}$ symmetric, $\Delta_1 > 0$, $0 < \gamma_1 < \gamma_2 < 1$ and $0 < \eta_1 \leq \eta_2 < 1$, set $k := 1$.

Step 1 Let the model $m_k(x_k + s) : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$m_k(x_k + s) \equiv \max_{i \in I} \left\{ f_i(x_k) + \nabla f_i(x_k)^T s \right\} + \frac{1}{2} s^T H_k s. \tag{118}$$

Compute a step s_k for which $\|s_k\| \leq \Delta_k$ and

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa_c \mu(x_k) \min \left\{ \frac{\mu(x_k)}{1 + \|H_k\|}, \Delta_k \right\}, \tag{119}$$

where $\kappa_c \in (0, 1)$ is a constant independent of k and where $\mu(x)$ is defined by (117).

Step 2 Compute the ratio

$$\rho_k = \frac{\max_{i \in I} \{f_i(x_k)\} - \max_{i \in I} \{f_i(x_k + s_k)\}}{m_k(x_k) - m_k(x_k + s_k)}, \tag{120}$$

set the next iterate

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{if } \rho_k \geq \eta_1, \\ x_k, & \text{otherwise}, \end{cases} \tag{121}$$

and choose the trust-region radius Δ_{k+1} by the update rule

$$\Delta_{k+1} \in \begin{cases} \left[\gamma_1 \Delta_k, \gamma_2 \Delta_k\right], & \text{if } \rho_k < \eta_1, \\ \left[\gamma_2 \Delta_k, \Delta_k\right], & \text{if } \rho_k \in [\eta_1, \eta_2), \\ [\Delta_k, +\infty], & \text{if } \rho_k \geq \eta_2. \end{cases} \tag{122}$$

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

Remark 10 Note that the model m_k given by (118) can be rewrited as

$$m_k(x_k + s) = h(f(x_k) + J_f(x_k)s) + \frac{1}{2} s^T H_k s,$$

where $h(f) = max_{i \in I} f_i$. Thus, the matrix H_k can be chosen according to equation (1.5) in [23].

We claim that, under suitable conditions, the TRMP algorithm is a particular case of Algorithm 3 with the choices

$$\begin{aligned} m_k(x_k + s) &= \max_{i \in I} \left\{ f_i(x_k) + \nabla f_i(x_k)^T s \right\} + \frac{1}{2} s^T H_k s, \\ \Phi(x) &= \max_{i \in I} \{f_i(x)\}, \quad \omega(x) = 1, \quad \phi(x) = \psi(x) = \chi(x) = \mu(x), \\ \delta_k &= \Delta_k, \quad \alpha = 1 \quad \text{and} \quad \beta = 0. \end{aligned} \tag{123}$$

Indeed, conditions A1′, A2′, A4′, A5′ and A6′ are naturally satisfied. The possibility of obtain s_k satisfying A5′ is guaranteed by Lemma 4.1, Corollary 4.1 and Lemma 4.2 in Villacorta et al. [21]. On the other hand, if we assume that $x_k \in S \subset \mathbb{R}^n$ for all k, with S bounded, then condition A3′ is satisfied due to the continuity of $\chi(=\mu)$. Finally, it is direct the fact that $m_k(x_k) = \max_{i \in I}\{f_i(x_k)\} = \Phi(x_k)$, while the second part of A7′ is given by the theorem below (regarding Algorithm 4).

Theorem 8 *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be twice continuously differentiable. Suppose that there exist constants $\kappa_f > 0$ and $\kappa_H \geq 1$ such that $\|\nabla^2 f_i(x)\| \leq \kappa_f$ for all $i \in I$ and $x \in \mathbb{R}^n$, and $\|H_k\| \leq \kappa_H - 1$ for all k. Then,*

$$\Phi(x_k + s_k) - m_k(x_k + s_k) \leq \kappa_m \|s_k\|^2,$$

where $\kappa_m = \max\left\{\kappa_f, \kappa_H\right\}$.

Proof See Proposition 5.1 in Villacorta et al. [21]. □

Therefore, the TRMP algorithm is a particular case of Algorithm 3. As a consequence of Theorem 7, we have the following worst-case complexity result.

Corollary 2 *Assume that the conditions of Theorem 8 are satisfied. Moreover, suppose that $x_k \in S \subset \mathbb{R}^n$ for all k, with S bounded, and let $\{\Phi(x_k)\}$ be bounded below by Φ_{low}. Then, to reduce the Pareto criticality measure $\mu(x)$ below $\varepsilon \in (0, 1]$, the TRMP algorithm [21] with update rule (12) (where $\delta_k = \Delta_k$) takes at most $O(\varepsilon^{-2})$ iterations.*

Remark 11 As far we know, the result above is the first iteration complexity bound of this kind for unconstrained multiobjective optimization in which the coordinate functions f_i are allowed to be nonlinear and nonconvex.

We finish this section noting that, as in the case of composite nonsmooth optimization, the nonlinear stepsize control framework also can be used to design new algorithms for unconstrained multiobjective optimization. For example, a generalization of the TRMP algorithm is obtained from the relaxations $\alpha \in (0, 1]$ and $\beta \in [0, 1]$ in (123). By Theorem 7, such nonlinear stepsize control trust-region algorithm takes at most $O(\varepsilon^{-(2+\beta)})$ iterations to reduce the Pareto criticality measure $\mu(x)$ below $\varepsilon \in (0, 1]$, and this bound is reduced to $O(\varepsilon^{-2})$ when $\alpha + \beta \leq 1$ and $2\alpha + \beta \geq 1$.

5 Conclusion

In this paper, we investigate the global convergence and the worst-case complexity of the nonlinear stepsize control algorithm recently proposed by Toint [19] for unconstrained optimization. Using a slightly more restrictive update rule for the stepsize parameter, we proved that the algorithm of Toint [19] still remains globally convergent if we assume that the norm of the Hessians H_k of the models can grow by a constant amount at each iteration. In this sense, our results are a generalization of the results of Powell [14,16] for trust region algorithms. In particular, they provide a convergence guarantee when the matrices H_k are updated by standard quasi-Newton methods. Furthermore, assuming that the matrices H_k are uniformly bounded, we have proved a worst-case complexity of $O(\varepsilon^{-(2+\beta)})$ iterations for the algorithm to achieve the first order criticality within ε, which is less pessimistic than the bound of $O(\varepsilon^{-3})$ discussed by Toint [19]. For the particular case in which $\alpha+\beta \leq 1$, $2\alpha+\beta \geq 1$ and $\phi_k, \psi_k \geq \chi_k$ (including the ARC algorithm), this estimate was even improved to $O(\varepsilon^{-2})$. Finally, we have extended the nonlinear stepsize control framework to some algorithms for composite nonsmooth optimization and unconstrained multiobjective optimization, which allowed us to obtain new complexity results. As a topic for future research, it would be interesting to investigate whether the complexity results presented here can be obtained under weaker assumptions over H_k.

Acknowledgments This work was done while the first author was visiting the Institute of Computational Mathematics and Scientific/Engineering Computing of the Chinese Academy of Sciences. He would like to express his deep gratitude to Professor Ya-xiang Yuan, Professor Yu-hong Dai, Dr. Xin Liu and Dr. Ya-feng Liu for their warm hospitality. The authors also are grateful to two anonymous referees, whose comments helped a lot to improve the first version of this paper.

References

1. Bellavia, S., Cartis, C., Gould, N.I.M., Morini, B., Toint, Ph.L.: Convergence of a regularized euclidean residual algorithm for nonlinear least-squares. SIAM J. Numer. Anal. **48**, 1–29 (2010)
2. Cartis, C., Gould, N.I.M., Toint, Ph.L.: Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. Math. Program. **127**, 245–295 (2011)
3. Cartis, C., Gould, N.I.M., Toint, Ph.L.: Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function—and derivative—evaluation complexity. Math. Program. **130**, 295–319 (2011)
4. Cartis, C., Gould, N.I.M., Toint, Ph.L.: On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming. SIAM J. Optim. **21**, 1721–1739 (2011)
5. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust-Region Methods. SIAM, Philadelphia (2000)
6. Dai, Y.H., Xu, D.C.: A new family of trust region algorithms for unconstrained optimization. J. Comp. Math. **21**, 221–228 (2003)
7. Dennis, J.E., Schnabel, R.B.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, Englewood Cliffs (1983)
8. Fan, J.: Convergence rate of the trust region method for nonlinear equations under local error bound condition. Comput. Optim. Appl. **34**, 215–227 (2006)
9. 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)
10. Fletcher, R.: A model algorithm for composite nondifferentiable optimization problems. Math. Program. **17**, 67–76 (1982)
11. Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. Math. Meth. Oper. Res. **51**, 479–494 (2000)
12. Guerraggio, A., Luc, D.T.: Optimality conditions for $C^{1,1}$ vector optimization problems. J. Optim. Theory Appl. **109**, 615–629 (2001)
13. Nesterov, Yu.: Modified Gauss–Newton scheme with worst-case guarantees for global performance. Optim. Methodos Softw. **22**, 469–483 (2007)
14. 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, vol. 2. Academic Press, New York (1975)
15. Powell, M.J.D.: General Algorithms for Discrete Nonlinear Approximation Calculations. Report DAMTP 1983/NA2, Department of Applied Mathematics and Theoretical Physics. University of Cambridge, Cambridge, England (1983)

16. Powell, M.J.D.: On the global convergence of trust region algorithms for unconstrained minimization. Math. Program. **29**, 297–303 (1984)
17. 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)
18. Sun, W., Yuan, Y.: Optimization Theory and Methods: Nonlinear Programming. Springer, Berlin (2006)
19. Toint, Ph.L.: Nonlinear stepsize control, trust regions and regularizations for unconstrained optimization. Optim. Methods Softw. **28**, 82–95 (2013)
20. Toint, Ph.L.: Global convergence of a class of trust region methods for nonconvex minimization in Hilbert space. IMA J. Numer. Anal. **8**, 231–252 (1988)
21. Villacorta, K.D.V., Oliveira, P.R., Soubeyran, A.: A trust-region method for unconstrained multiobjective problems with applications in satisficing processes. J. Optim. Theory Appl. **160**, 865–889 (2014)
22. Yuan, Y.: Conditions for convergence of trust region algorithm for nonsmooth optimization. Math. Program. **31**, 220–228 (1985)
23. Yuan, Y.: On the superlinear convergence of a trust region algorithm for nonsmooth optimization. Math. Program. **31**, 269–285 (1985)
24. Zhang, J., Wang, Y.: A new trust region method for nonlinear equations. Math. Meth. Oper. Res. **58**, 283–298 (2003)