IMA Journal of Numerical Analysis (2018) **00**, 1–32
doi:10.1093/imanum/dry022

An inexact regularized Newton framework with a worst-case iteration complexity of $\mathscr{O}(\varepsilon^{-3/2})$ for nonconvex optimization

Frank E. Curtis*
Department of Industrial and Systems Engineering, Lehigh University
*Corresponding author: frank.e.curtis@gmail.com

Daniel P. Robinson
Department of Applied Mathematics and Statistics, Johns Hopkins University
daniel.p.robinson@gmail.com

and

Mohammadreza Samadi
Department of Industrial and Systems Engineering, Lehigh University
mohammadreza.samadi@lehigh.edu

[Received on 27 September 2017; revised on 14 March 2018]

An algorithm for solving smooth nonconvex optimization problems is proposed that, in the worst-case, takes $\mathscr{O}(\varepsilon^{-3/2})$ iterations to drive the norm of the gradient of the objective function below a prescribed positive real number ε and can take $\mathscr{O}(\varepsilon^{-3})$ iterations to drive the leftmost eigenvalue of the Hessian of the objective above $-\varepsilon$. The proposed algorithm is a general framework that covers a wide range of techniques including quadratically and cubically regularized Newton methods, such as the Adaptive Regularization using Cubics (ARC) method and the recently proposed Trust-Region Algorithm with Contractions and Expansions (TRACE). The generality of our method is achieved through the introduction of generic conditions that each trial step is required to satisfy, which in particular allows for inexact regularized Newton steps to be used. These conditions center around a new subproblem that can be approximately solved to obtain trial steps that satisfy the conditions. A new instance of the framework, distinct from ARC and TRACE, is described that may be viewed as a hybrid between quadratically and cubically regularized Newton methods. Numerical results demonstrate that our hybrid algorithm outperforms a cubically regularized Newton method.

Keywords: unconstrained optimization; nonlinear optimization; nonconvex optimization; inexact Newton methods; worst-case iteration complexity; worst-case evaluation complexity.

1. Introduction

This paper proposes an algorithm for solving

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1.1}$$

where the (possibly nonconvex) objective function $f : \mathbb{R}^n \to \mathbb{R}$ is assumed to be twice continuously differentiable. The optimization problem (1.1) has been widely studied, as evidenced by its appearance as the focal point of numerous textbooks; e.g., see Bertsekas (1999), Conn *et al.* (2000), Bazaraa *et al.* (2006), Nocedal & Wright (2006), Ruszczynski (2006), and Griva *et al.* (2008).

For many years the most popular methods for solving (1.1) were in classes known as line search and trust region methods. Recently, however, cubic regularization methods have become popular, which are based on the pioneering works by Griewank (1981) and Nesterov & Polyak (2006). Their rise in popularity is due to increased interest in algorithms with improved complexity properties, which stems from the impact of so-called optimal algorithms for solving convex optimization problems. For problem (1.1) by complexity properties, we mean a guaranteed bound on the number of iterations (or function evaluations or derivative evaluations) needed by an algorithm before the norm of the gradient of the objective must fall below a positive threshold $\varepsilon > 0$. In other words if x_k denotes the kth iteration of an algorithm, one seeks a bound on the number of iterations until it is guaranteed that

$$\|\nabla f(x_k)\| \leqslant \varepsilon.$$

The complexity of a traditional trust region method (e.g., see Algorithm 6.1.1 in Conn *et al.*, 2000) is $\mathscr{O}(\varepsilon^{-2})$ (see Cartis *et al.*, 2010) that falls short of the $\mathscr{O}(\varepsilon^{-3/2})$ complexity for cubic regularization methods (e.g., see the Adaptive Regularization using Cubics [ARC] method by Cartis *et al.*, 2011a,b). This latter complexity is optimal among a certain broad class of second-order methods when employed to minimize a broad class of objective functions; see Cartis *et al.* (2011c). That said, one can obtain even better complexity properties if higher-order derivatives are used; see Birgin *et al.* (2017) and Cartis *et al.* (2017).

The better complexity properties of regularization methods such as ARC have been a major point of motivation for discovering other methods that attain the same worst-case iteration complexity bounds. For example, the recently introduced (nontraditional) trust region method known as Trust Region Algorithm with Contractions and Expansions (TRACE) (see Curtis *et al.*, 2017) has the same optimal $\mathscr{O}(\varepsilon^{-3/2})$ complexity, while at the same time allowing traditional trust region trial steps to be computed and used. A key aspect of the TRACE framework is that a solution to an implicit trust region problem is obtained by varying a regularization parameter instead of a trust region radius. This key idea has been adopted and advanced further by Birgin & Martínez (2017); in particular, they propose an algorithm that has optimal iteration complexity by solving quadratic subproblems that have a carefully chosen quadratic regularization parameter.

CONTRIBUTIONS The main contributions of this paper relate to advancing the understanding of optimal complexity algorithms for solving the smooth optimization problem (1.1). Our proposed framework is intentionally very general; it is not a trust region method, a quadratic regularization method or a cubic regularization method. Rather, we propose a generic set of conditions that each trial step must satisfy that still allow us to establish an optimal first-order complexity result as well as a second-order complexity bound similar to the methods above. Our framework contains as special cases other optimal complexity algorithms such as ARC and TRACE. To highlight this generality of our contribution, we describe one particular instance of our framework that appears to be new to the literature.

During the final preparation of this article we came across the work in Dussault (2017) and Dussault & Orban (2017). This work shares certain commonalities with our own and appears to have been developed at the same time. Although there are numerous differences we shall only point out three of them. First, the precise conditions that they require for each trial step are different from ours. In particular, the condition stated as (3.1c) in Dussault & Orban (2017) requires that regularization is used to compute every trial step, a property not shared by our method (which can employ Newton steps). Secondly, they do not consider second-order convergence or complexity properties, although they might be able to do so by incorporating second-order conditions similar to ours. Thirdly, they focus

on strategies for identifying an appropriate value for the regularization parameter. An implementation of our method might consider their proposals, but could employ other strategies as well. In any case, overall, we believe that our papers are quite distinct and in some ways are complementary.

ORGANIZATION In Section 2 we present our general framework that is formally stated as Algorithm 1. In Section 3 we prove that our framework enjoys first-order convergence (see Section 3.1), an optimal first-order complexity (see Section 3.2) and certain second-order convergence and complexity guarantees (see Section 3.3). In Section 4 we show that ARC and TRACE can be viewed as special cases of our framework and present yet another instance that is distinct from these methods. In Section 5 we present details of implementations of a cubic regularization method and our newly proposed instance of our framework, and provide the results of numerical experiments with both. Finally, in Section 6 we present final comments.

NOTATION We use $\mathbb{R}_+$ to denote the set of non-negative scalars, $\mathbb{R}_{++}$ to denote the set of positive scalars and $\mathbb{N}_+$ to denote the set of non-negative integers. Given a real symmetric matrix A we write $A \succeq 0$ (respectively, $A \succ 0$) to indicate that A is positive semidefinite (respectively, positive definite). Given a pair of scalars $(a, b) \in \mathbb{R} \times \mathbb{R}$ we write $a \perp b$ to indicate that $ab = 0$. Similarly, given such a pair we denote their maximum as $\max\{a, b\}$ and their minimum as $\min\{a, b\}$. Given a vector v we denote its (Euclidean) ℓ_2-norm as $\|v\|$. Finally, given a discrete set $\mathscr{S}$ we denote its cardinality by $|\mathscr{S}|$.

Corresponding to the objective $f : \mathbb{R}^n \to \mathbb{R}$ we define the gradient function $g := \nabla f : \mathbb{R}^n \to \mathbb{R}^n$ and the Hessian function $H := \nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n \times n}$. Given an iterate x_k in an algorithm for solving (1.1) we define $f_k := f(x_k)$, $g_k := g(x_k) := \nabla f(x_k)$, and $H_k := H(x_k) := \nabla^2 f(x_k)$. Similarly, we apply a subscript to other algorithmic quantities whose definition depends on the iteration number k.

2. Algorithm description

Our algorithm involves generic conditions that a trial step toward solving problem (1.1) must satisfy. One can obtain a step satisfying these conditions by computing—for appropriate positive lower and upper bounds σ_k^{L} and σ_k^{U}, respectively, on the ratio between a regularization variable $\lambda \geqslant 0$ and the norm of the trial step—an approximate solution of the subproblem

$$\mathscr{P}_k\left(\sigma_k^{\mathrm{L}}, \sigma_k^{\mathrm{U}}\right) \; : \; \min_{(s,\lambda) \in \mathbb{R}^n \times \mathbb{R}_+} f_k + g_k^T s + \tfrac{1}{2} s^T (H_k + \lambda I) s$$
$$\text{s.t.} \left(\sigma_k^{\mathrm{L}}\right)^2 \|s\|^2 \leqslant \lambda^2 \leqslant \left(\sigma_k^{\mathrm{U}}\right)^2 \|s\|^2. \tag{2.1}$$

For a given value of the regularization variable λ, this problem involves a quadratic objective function and an upper bound on the norm of the trial step, just as in a trust region method. However, it also includes a lower bound on the norm of the trial step, and, in general, with λ as a variable, it encapsulates other types of subproblems as well, including those present in a cubic regularization framework. For additional details on the properties of this subproblem and its solutions see Appendices A and B.

The conditions that the kth trial step and regularization pair, i.e., (s_k, λ_k), must satisfy are stated in Assumption 2.1 below, wherein we invoke the following (unregularized) quadratic model of f at x_k:

$$q_k(s) := f_k + g_k^T s + \tfrac{1}{2} s^T H_k s.$$

ASSUMPTION 2.1 The pair (s_k, λ_k) is computed such that it is feasible for problem (2.1), and with

$$\Delta_k(s_k, \lambda_k) := \begin{cases} \|s_k\| & \text{if } \lambda_k = 0 \\ \frac{1}{\sqrt{6}}\sqrt{\frac{\|g_k\|\|s_k\|}{\lambda_k}} & \text{if } \lambda_k > 0 \end{cases} \tag{2.2}$$

and constants $(\kappa_1, \kappa_2, \kappa_3) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$ the following hold:

$$f_k - q_k(s_k) \geqslant \frac{\|g_k\|}{6\sqrt{2}} \min\left\{\frac{\|g_k\|}{1 + \|H_k\|}, \Delta_k(s_k, \lambda_k)\right\}; \tag{2.3a}$$

$$s_k^T(g_k + (H_k + \lambda_k I)s_k) \leqslant \min\left\{\kappa_1\|s_k\|^2, \tfrac{1}{2}s_k^T(H_k + \lambda_k I)s_k + \tfrac{1}{2}\kappa_2\|s_k\|^3\right\}; \text{ and} \tag{2.3b}$$

$$\|g_k + (H_k + \lambda_k I)s_k\| \leqslant \lambda_k\|s_k\| + \kappa_3\|s_k\|^2. \tag{2.3c}$$

To see that Assumption 2.1 is well posed and consistent with problem (2.1) we refer the reader to Theorem B6 in Appendix B, wherein we prove that any solution of problem (2.1) with s restricted to a sufficiently large dimensional subspace of $\mathbb{R}^n$ satisfies all of the conditions in Assumption 2.1. We also claim that one can obtain a pair satisfying Assumption 2.1 in either of the following two ways:

- Choose $\sigma \in [\sigma_k^{\mathrm{L}}, \sigma_k^{\mathrm{U}}]$, compute s_k by minimizing the cubic function

$$c_k(s; \sigma) := q_k(s) + \tfrac{1}{2}\sigma\|s\|^3 = f_k + g_k^T s + \tfrac{1}{2}s^T H_k s + \tfrac{1}{2}\sigma\|s\|^3 \tag{2.4}$$

  over a sufficiently large dimensional subspace of $\mathbb{R}^n$ (assuming, when $\sigma = \sigma_k^{\mathrm{L}} = 0$, that this function is bounded below), then set $\lambda_k \leftarrow \sigma\|s_k\|$. This is essentially the strategy employed in cubic regularization methods such as ARC.

- Choose $\lambda_k \geqslant 0$, then compute s_k by minimizing the objective of (2.1) with $\lambda = \lambda_k$ over a sufficiently large dimensional subspace of $\mathbb{R}^n$ (assuming that the function is bounded below). The resulting pair (s_k, λ_k) satisfies Assumption 2.1 as long as it is feasible for (2.1). This is essentially the strategy employed in Birgin & Martínez (2017) and partly employed in TRACE.

One can imagine other approaches as well. Overall, we state problem (2.1) as a guide for various techniques for computing the pair (s_k, λ_k). Our theory simply relies on the fact that any such computed pair satisfies the conditions in Assumption 2.1.

Our algorithm, stated as Algorithm 1, employs the following ratio (also employed, e.g., in TRACE) to determine whether a given trial step is accepted or rejected:

$$\rho_k := \frac{f_k - f(x_k + s_k)}{\|s_k\|^3}.$$

One potential drawback of employing this ratio is that the ratio is not invariant to scaling of the objective function. However, the use of this ratio can still be justified. For example, if one were to compute s_k by minimizing the cubic model (2.4) for some $\sigma > 0$, then the reduction in this model yielded by s_k is bounded below by a fraction of $\sigma\|s_k\|^3$ (see Cartis *et al.*, 2011b, Lemma 4.2), meaning that $\rho_k \geqslant \eta$ holds when $\sigma \geqslant \eta$ and the actual reduction in f is proportional to the reduction in the cubic model. For further justification for this choice—such as how it allows the algorithm to accept Newton steps

when the norm of the trial step is small (and, indeed, the norms of accepted steps vanish asymptotically as shown in Lemma 3.9 later on)—we refer the reader to Birgin & Martínez (2017) and Curtis *et al.* (2017).

Algorithm 1 Inexact Regularized Newton Framework

Require: an acceptance constant $\eta \in \mathbb{R}_{++}$ with $0 < \eta < 1$

Require: bound update constants $\{\gamma_1, \gamma_2\} \subset \mathbb{R}_{++}$ with $1 < \gamma_1 \leqslant \gamma_2$

Require: ratio lower- and upper-bound constants $\{\underline{\sigma}, \overline{\sigma}\} \subset \mathbb{R}_{++}$ such that $\overline{\sigma} \geqslant \underline{\sigma}$

procedure INEXACT REGULARIZED NEWTON
  set $x_0 \in \mathbb{R}^n$
  set $\sigma_0^{\mathrm{L}} \leftarrow 0$ and $\sigma_0^{\mathrm{U}} \in [\underline{\sigma}, \overline{\sigma}]$
  for $k \in \mathbb{N}_+$ **do**
    set (s_k, λ_k) satisfying Assumption 2.1
    if $\rho_k \geqslant \eta$ **then** [accept step]
      set $x_{k+1} \leftarrow x_k + s_k$
      set $\sigma_{k+1}^{\mathrm{L}} \leftarrow 0$ and $\sigma_{k+1}^{\mathrm{U}} \leftarrow \sigma_k^{\mathrm{U}}$
    else (i.e., $\rho_k < \eta$) [reject step]
      set $x_{k+1} \leftarrow x_k$
      if $\lambda_k < \underline{\sigma}\|s_k\|$ **then**
        set $\sigma_{k+1}^{\mathrm{L}} \in [\underline{\sigma}, \overline{\sigma}]$ and $\sigma_{k+1}^{\mathrm{U}} \in [\sigma_{k+1}^{\mathrm{L}}, \overline{\sigma}]$
      else
        set $\sigma_{k+1}^{\mathrm{L}} \leftarrow \gamma_1 \frac{\lambda_k}{\|s_k\|}$ and $\sigma_{k+1}^{\mathrm{U}} \leftarrow \gamma_2 \frac{\lambda_k}{\|s_k\|}$

3. Convergence analysis

In this section we prove global convergence guarantees for Algorithm 1. In particular we prove under common assumptions that, from remote starting points, the algorithm converges to first-order stationarity has a worst-case iteration complexity to approximate first-order stationarity that is on par with the methods in Cartis *et al.* (2011b), Birgin & Martínez (2017) and Curtis *et al.* (2017), and—at least in a subspace determined by the search path of the algorithm—converges to second-order stationarity with a complexity on par with the methods in Cartis *et al.* (2011b) and Curtis *et al.* (2017).

3.1 *First-order global convergence*

Our goal in this subsection is to prove that the sequence of objective gradients vanishes. We make the following assumption about the objective function, which is assumed to hold throughout this section.

ASSUMPTION 3.1 The objective function $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and bounded below by a scalar $f_{\inf} \in \mathbb{R}$ on $\mathbb{R}^n$.

We also make the following assumption related to the sequence of iterates.

ASSUMPTION 3.2 The gradient function $g : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with Lipschitz constant $g_{Lip} \in \mathbb{R}_{++}$ in an open convex set containing the sequences $\{x_k\}$ and $\{x_k + s_k\}$. Furthermore, the gradient sequence $\{g_k\}$ has $g_k \neq 0$ for all $k \in \mathbb{N}_+$ and is bounded in that there exists a scalar constant $g_{\max} \in \mathbb{R}_{++}$ such that $\|g_k\| \leqslant g_{\max}$ for all $k \in \mathbb{N}_+$.

It is worthwhile to note in passing that our complexity bounds for first- and second-order stationarity remain true even if one were to consider the possibility that $g_k = 0$ for some $k \in \mathbb{N}_+$, in which case one would have the algorithm terminate finitely or, if $H_k \not\succeq 0$, compute an improving direction of negative curvature for H_k. However, allowing this possibility—which is typically unlikely ever to occur in practice—would only serve to obscure certain aspects of our analysis. We refer the reader, e.g., to Cartis *et al.* (2011b) (specifically, to the discussions at the ends of Sections 2.1, 4 and 5 in that work) for commentary about why zero gradient values do not ruin complexity guarantees such as we present.

We begin with two lemmas, each revealing an important consequence of Assumptions 3.1 and 3.2.

LEMMA 3.3 For all $k \in \mathbb{N}_+$, it follows that $s_k \neq 0$.

Proof. The result follows by combining that $g_k \neq 0$ for all $k \in \mathbb{N}_+$ (see Assumption 3.2) with (2.3c). □

LEMMA 3.4 The Hessian sequence $\{H_k\}$ is bounded in norm in that there exists a scalar constant $H_{\max} \in \mathbb{R}_{++}$ such that $\|H_k\| \leqslant H_{\max}$ for all $k \in \mathbb{N}_+$.

Proof. The result follows by Assumption 3.1, the Lipschitz continuity of g in Assumption 3.2 and Lemma 1.2.2 in Nesterov (2004). □

In our next lemma we prove an upper bound for the regularization variable λ_k.

LEMMA 3.5 For all $k \in \mathbb{N}_+$ the pair (s_k, λ_k) satisfies

$$\lambda_k \leqslant 2\frac{\|g_k\|}{\|s_k\|} + \tfrac{3}{2}H_{\max} + \kappa_1.$$

Proof. Since (2.3a) ensures $q_k(s_k) - f_k \leqslant 0$ it follows with (2.3b) and Lemma 3.4 that

$$\begin{aligned} 0 \geqslant q_k(s_k) - f_k &= g_k^T s_k + \tfrac{1}{2}s_k^T H_k s_k \\ &\geqslant g_k^T s_k + \tfrac{1}{2}s_k^T H_k s_k + s_k^T(g_k + (H_k + \lambda_k I)s_k) - \kappa_1\|s_k\|^2 \\ &= 2g_k^T s_k + \tfrac{3}{2}s_k^T H_k s_k + \lambda_k\|s_k\|^2 - \kappa_1\|s_k\|^2 \\ &\geqslant -2\|g_k\|\|s_k\| - \tfrac{3}{2}H_{\max}\|s_k\|^2 + \lambda_k\|s_k\|^2 - \kappa_1\|s_k\|^2. \end{aligned}$$

After rearrangement and dividing by $\|s_k\|^2 \neq 0$ (see Lemma 3.3) the desired result follows. □

Using Lemma 3.5 we now prove a lower bound for the reduction in q_k yielded by s_k.

LEMMA 3.6 For all $k \in \mathbb{N}_+$ the step s_k satisfies

$$f_k - q_k(s_k) \geqslant \frac{\|g_k\|}{6\sqrt{2}} \min\left\{ \frac{\|g_k\|}{1 + H_{\max}}, \frac{\|s_k\|}{\sqrt{6}}\sqrt{\frac{\|g_k\|}{2\|g_k\| + \|s_k\|\left(\frac{3}{2}H_{\max} + \kappa_1\right)}} \right\}.$$

Proof. If $\lambda_k = 0$ then by (2.3a) and Lemma 3.4, it follows that

$$f_k - q_k(s_k) \geqslant \frac{\|g_k\|}{6\sqrt{2}} \min\left\{\frac{\|g_k\|}{1+\|H_k\|}, \|s_k\|\right\} \geqslant \frac{\|g_k\|}{6\sqrt{2}} \min\left\{\frac{\|g_k\|}{1+H_{\max}}, \|s_k\|\right\}.$$

On the other hand, if $\lambda_k > 0$ then (2.3a), Lemmas 3.4 and 3.5 imply that

$$\begin{aligned} f_k - q_k(s_k) &\geqslant \frac{\|g_k\|}{6\sqrt{2}} \min\left\{\frac{\|g_k\|}{1+\|H_k\|}, \frac{1}{\sqrt{6}}\sqrt{\frac{\|g_k\|\|s_k\|}{\lambda_k}}\right\} \\ &\geqslant \frac{\|g_k\|}{6\sqrt{2}} \min\left\{\frac{\|g_k\|}{1+H_{\max}}, \frac{\|s_k\|}{\sqrt{6}}\sqrt{\frac{\|g_k\|}{2\|g_k\| + \|s_k\|\left(\frac{3}{2}H_{\max}+\kappa_1\right)}}\right\}. \end{aligned}$$

Combining the inequalities from these two cases the desired result follows. □

Going forward for ease of reference, we respectively define sets of indices corresponding to accepted and rejected steps throughout a run of the algorithm as

$$\mathscr{A} := \left\{k \in \mathbb{N}_+ : \rho_k \geqslant \eta\right\} \text{ and } \mathscr{R} := \left\{k \in \mathbb{N}_+ : \rho_k < \eta\right\}.$$

We now show that if the algorithm were only to compute rejected steps from some iteration onward, then the sequence $\{\lambda_k/\|s_k\|\}$ diverges to infinity.

LEMMA 3.7 If $k \in \mathscr{R}$ for all sufficiently large $k \in \mathbb{N}_+$ then $\{\lambda_k/\|s_k\|\} \to \infty$.

Proof. Without loss of generality assume that $\mathscr{R} = \mathbb{N}_+$. We now prove that the condition in Step 11 cannot be true more than once. Suppose, in iteration $\hat{k} \in \mathbb{N}_+$, Step 12 is reached, which means that $\lambda_{\hat{k}+1}/\|s_{\hat{k}+1}\| \geqslant \underline{\sigma}$ since $(s_{\hat{k}+1}, \lambda_{\hat{k}+1})$ is required to be feasible for $\mathscr{P}_{\hat{k}+1}(\sigma^{\mathrm{L}}_{\hat{k}+1}, \sigma^{\mathrm{U}}_{\hat{k}+1})$ in Step 5 where $\sigma^{\mathrm{L}}_{\hat{k}+1} \geqslant \underline{\sigma}$. Therefore, the condition in Step 11 tests false in iteration $(\hat{k}+1)$. Then from Step 5, Step 14 and the fact that $\gamma_1 > 1$, it follows that $\{\lambda_k/\|s_k\|\}$ is monotonically increasing for all $k \geqslant \hat{k}$. Therefore, the condition in Step 11 cannot test true for any $k \geqslant \hat{k}+1$. Now, to see that the sequence diverges notice from this fact, Step 5 and Step 14, it follows that for all $k \geqslant \hat{k}+1$ we have $\lambda_{k+1}/\|s_{k+1}\| \geqslant \gamma_1(\lambda_k/\|s_k\|)$ where $\gamma_1 > 1$. Thus, $\{\lambda_k/\|s_k\|\} \to \infty$, as claimed. □

We now prove that if the gradients are bounded away from zero and the sequence of ratios $\{\lambda_k/\|s_k\|\}$ diverges, then $\rho_k \geqslant \eta$ for all sufficiently large $k \in \mathbb{N}_+$, meaning that the steps are accepted.

LEMMA 3.8 Suppose that $\mathscr{I} \subseteq \mathbb{N}_+$ is an infinite index set such that for $\varepsilon \in \mathbb{R}_{++}$ independent of k, one finds that $\|g_k\| \geqslant \varepsilon$ for all $k \in \mathscr{I}$ and $\{\lambda_k/\|s_k\|\}_{k\in\mathscr{I}} \to \infty$. Then for all sufficiently large $k \in \mathscr{I}$, it follows that $\rho_k \geqslant \eta$, meaning $k \in \mathscr{A}$.

Proof. From the Mean Value Theorem, there exists $\bar{x}_k \in [x_k, x_k + s_k]$ such that

$$\begin{aligned} q_k(s_k) - f(x_k + s_k) &= (g_k - g(\bar{x}_k))^T s_k + \tfrac{1}{2} s_k^T H_k s_k \\ &\geqslant -\|g_k - g(\bar{x}_k)\|\,\|s_k\| - \tfrac{1}{2}\|H_k\|\|s_k\|^2. \end{aligned} \tag{3.1}$$

From this, Lemma 3.6 and Assumption 3.2, it follows that for all $k \in \mathscr{I}$

$$\begin{aligned} f_k - f(x_k + s_k) &= f_k - q_k(s_k) + q_k(s_k) - f(x_k + s_k) \\ &\geqslant \frac{\|g_k\|}{6\sqrt{2}} \min\left\{ \frac{\|g_k\|}{1+H_{\max}}, \frac{\|s_k\|}{\sqrt{6}} \sqrt{\frac{\|g_k\|}{2\|g_k\| + \|s_k\|\left(\frac{3}{2}H_{\max} + \kappa_1\right)}} \right\} - \left(g_{Lip} + \tfrac{1}{2}H_{\max}\right)\|s_k\|^2 \\ &\geqslant \frac{\varepsilon}{6\sqrt{2}} \min\left\{ \frac{\varepsilon}{1+H_{\max}}, \frac{\|s_k\|}{\sqrt{6}} \sqrt{\frac{\varepsilon}{2g_{\max} + \|s_k\|\left(\frac{3}{2}H_{\max} + \kappa_1\right)}} \right\} - \left(g_{Lip} + \tfrac{1}{2}H_{\max}\right)\|s_k\|^2. \end{aligned}$$

This shows that there exists a threshold $s_{\text{thresh}} > 0$ such that

$$f_k - f(x_k + s_k) \geqslant \eta \|s_k\|^3 \ \text{ whenever } \ k \in \mathscr{I} \ \text{ and } \ \|s_k\| \leqslant s_{\text{thresh}}.$$

We now claim that $\{\|s_k\|\}_{k\in\mathscr{I}} \to 0$. To prove this claim, suppose by contradiction that there exists an infinite subsequence $\mathscr{I}_s \subseteq \mathscr{I}$ and scalar $\varepsilon_s \in \mathbb{R}_{++}$ such that $\|s_k\| \geqslant \varepsilon_s$ for all $k \in \mathscr{I}_s$. It then follows from the boundedness of $\{\|g_k\|\}$ (see Assumption 3.2) and Lemma 3.5 that $\{\lambda_k\}_{k\in\mathscr{I}_s}$ is bounded. This allows us to conclude that $\{\lambda_k/\|s_k\|\}_{k\in\mathscr{I}_s}$ is bounded, which contradicts the assumptions of the lemma. Thus, $\{\|s_k\|\}_{k\in\mathscr{I}} \to 0$. Hence, there exists $k_s \in \mathscr{I}$ such that for all $k \in \mathscr{I}$ with $k \geqslant k_s$ one finds $\|s_k\| \leqslant s_{\text{thresh}}$. Therefore, for all $k \in \mathscr{I}$ with $k \geqslant k_s$, it follows that $\rho_k \geqslant \eta$, as claimed. □

Next, we prove that the algorithm produces infinitely many accepted steps.

LEMMA 3.9 It holds that $|\mathscr{A}| = \infty$ and $\{s_k\}_{k\in\mathscr{A}} \to 0$.

Proof. To derive a contradiction suppose that $|\mathscr{A}| < \infty$. This implies that there exists k_0 such that for all $k \geqslant k_0$ one has $k \in \mathscr{R}$ and $(x_k, g_k, H_k) = (x_{k_0}, g_{k_0}, H_{k_0})$. From this fact and Assumption 3.2, it follows that $\|g_k\| \geqslant \varepsilon$ for all $k \geqslant k_0$ for some $\varepsilon \in \mathbb{R}_{++}$. From the fact that $k \in \mathscr{R}$ for all $k \geqslant k_0$ and Lemma 3.7, it follows that $\{\lambda_k/\|s_k\|\} \to \infty$. This fact and $\|g_k\| \geqslant \varepsilon$ for all $k \geqslant k_0$ imply that all the conditions of Lemma 3.8 are satisfied for $\mathscr{I} := \{k \in \mathbb{N}_+ : k \geqslant k_0\}$; therefore, Lemma 3.8 implies that for all sufficiently large $k \in \mathscr{I}$ one finds $\rho_k \geqslant \eta$ so that $k \in \mathscr{A}$, a contradiction.

To complete the proof notice that the objective function values are monotonically decreasing. Combining this with the condition in Step 6 the fact that f is bounded below by f_{inf} (see Assumption 3.1), and $|\mathscr{A}| = \infty$, one deduces that $\{s_k\}_{k\in\mathscr{A}} \to 0$, as claimed. □

We now prove that there exists an infinite subsequence of iterates such that the sequence of gradients computed at those points converges to zero.

LEMMA 3.10 It holds that

$$\liminf_{k\in\mathbb{N}_+, k\to\infty} \|g_k\| = 0.$$

Proof. To derive a contradiction suppose that $\liminf_{k\in\mathbb{N}_+,k\to\infty} \|g_k\| > 0$, which along with the fact that $g_{k+1} = g_k$ for any $k \in \mathbb{N}_+ \setminus \mathscr{A}$ means $\liminf_{k\in\mathscr{A},k\to\infty} \|g_k\| > 0$. Thus, there exists $\varepsilon \in \mathbb{R}_{++}$ such that

$$\|g_k\| \geqslant \varepsilon \ \text{ for all sufficiently large } \ k \in \mathscr{A}. \tag{3.2}$$

Under (3.2) let us prove that $\{\lambda_k\}_{k\in\mathscr{A}} \to \infty$. To derive a contradiction suppose there exists an infinite $\mathscr{A}_\lambda \subseteq \mathscr{A}$ such that $\lambda_k \leqslant \lambda_{\max}$ for some $\lambda_{\max} \in \mathbb{R}_{++}$. On the other hand, by $\{s_k\}_{k\in\mathscr{A}} \to 0$ (see Lemma 3.9) and (2.3c), it follows that $\{g_k + (H_k + \lambda_k I)s_k\}_{k\in\mathscr{A}_\lambda} \to 0$. Combining the upper bound on $\{\lambda_k\}_{k\in\mathscr{A}_\lambda}$ the fact that $\{s_k\}_{k\in\mathscr{A}} \to 0$ and $\|H_k\| \leqslant H_{\max}$ (see Lemma 3.4), it follows that $\{g_k\}_{k\in\mathscr{A}_\lambda} \to 0$, which violates (3.2). Therefore, $\{\lambda_k\}_{k\in\mathscr{A}} \to \infty$.

Our next goal is to prove, still under (3.2), that $k \in \mathscr{A}$ for all sufficiently large $k \in \mathbb{N}_+$. To prove this our strategy is to show that the sets of iterations involving a rejected step followed by an accepted step are finite. In particular, let us define the index sets

$$\mathscr{R}_1 := \left\{k \in \mathscr{R} : \text{the condition in Step 11 tests true and } (k+1) \in \mathscr{A}\right\} \text{ and}$$
$$\mathscr{R}_2 := \left\{k \in \mathscr{R} : \text{the condition in Step 11 tests false and } (k+1) \in \mathscr{A}\right\}.$$

We aim to prove that these are finite. First, consider $\mathscr{R}_1$. To derive a contradiction suppose that $|\mathscr{R}_1| = \infty$. By definition for all $k \in \mathscr{R}_1$ the condition in Step 11 tests true, meaning (s_{k+1}, λ_{k+1}) is found in Step 5 satisfying $\lambda_{k+1}/\|s_{k+1}\| \leqslant \overline{\sigma}$. On the other hand, since $(k+1) \in \mathscr{A}$ for all $k \in \mathscr{R}_1$, it follows from Lemma 3.9 that $\{s_{k+1}\}_{k\in\mathscr{R}_1} \to 0$. Combining the conclusions of these last two sentences shows that $\{\lambda_{k+1}\}_{k\in\mathscr{R}_1} \to 0$. However, this contradicts the conclusion of the previous paragraph, which showed that $\{\lambda_k\}_{k\in\mathscr{A}} \to \infty$. Hence, we may conclude that $|\mathscr{R}_1| < \infty$. Now consider $\mathscr{R}_2$. To derive a contradiction suppose that $|\mathscr{R}_2| = \infty$. The fact that the condition in Step 11 tests false for $k \in \mathscr{R}_2$ implies that (s_{k+1}, λ_{k+1}) is found in Step 5 satisfying $\lambda_{k+1}/\|s_{k+1}\| \leqslant \gamma_2\lambda_k/\|s_k\|$. However, since $\{s_{k+1}\}_{k\in\mathscr{R}_2} \to 0$ (see Lemma 3.9) and $\{\lambda_{k+1}\}_{k\in\mathscr{R}_2} \to \infty$ (established in the previous paragraph), it follows that $\{\lambda_{k+1}/\|s_{k+1}\|\}_{k\in\mathscr{R}_2} \to \infty$, which combined with the previously established inequality $\lambda_{k+1}/\|s_{k+1}\| \leqslant \gamma_2\lambda_k/\|s_k\|$ shows that $\{\lambda_k/\|s_k\|\}_{k\in\mathscr{R}_2} \to \infty$. Therefore, with (3.2), the conditions in Lemma 3.8 hold for $\mathscr{I} = \mathscr{R}_2$, meaning that, for all sufficiently large $k \in \mathscr{R}_2$, the inequality $\rho_k \geqslant \eta$ holds. This contradicts the fact that $\mathscr{R}_2 \subseteq \mathscr{R}$; hence, we conclude that $\mathscr{R}_2$ is finite. Since $\mathscr{R}_1$ and $\mathscr{R}_2$ are finite, it follows from the logic of Algorithm 1 that either $k \in \mathscr{A}$ for all sufficiently large k or $k \in \mathscr{R}$ for all sufficiently large k. By Lemma 3.9, it follows that $k \in \mathscr{A}$ for all sufficiently large k.

Thus far we have proved under (3.2) that $\{\lambda_k\}_{k\in\mathscr{A}} \to \infty$ and that $k \in \mathscr{A}$ for all large $k \in \mathbb{N}_+$. From this latter fact, it follows that there exists k_σ such that $\sigma_k^{\mathrm{U}} = \sigma_{k_\sigma}^{\mathrm{U}} \in \mathbb{R}_{++}$ for all $k \geqslant k_\sigma$. In addition from Step 5 it follows that for $k \geqslant k_\sigma$ one finds $\lambda_k/\|s_k\| \leqslant \sigma_k^{\mathrm{U}} = \sigma_{k_\sigma}^{\mathrm{U}} < \infty$. However, this leads to a contradiction to the facts that $\{\lambda_k\}_{k\in\mathscr{A}} \to \infty$ and $\{s_k\}_{k\in\mathscr{A}} \to 0$ (see Lemma 3.9). Overall, we have shown that (3.2) cannot be true, which proves the desired result. □

We close with our main global convergence result of this subsection, the proof of which borrows much from that of Theorem 3.14 in Curtis *et al.* (2017).

THEOREM 3.11 Under Assumptions 2.1, 3.1 and 3.2, it follows that

$$\lim_{k\in\mathbb{N}_+, k\to\infty} \|g_k\| = 0. \tag{3.3}$$

Proof. For the purpose of reaching a contradiction suppose that (3.3) does not hold. Combining this with the fact that $|\mathscr{A}| = \infty$ (see Lemma 3.9), it follows that there exist an infinite subsequence $\{t_i\} \subseteq \mathscr{A}$ (indexed over $i \in \mathbb{N}_+$) and a scalar $\varepsilon > 0$ such that, for all $i \in \mathbb{N}_+$, one finds $\|g_{t_i}\| \geqslant 2\varepsilon > 0$. Also, the

fact that $|\mathscr{A}| = \infty$ and Lemma 3.10 imply that there exists an infinite subsequence $\{\ell_i\} \subseteq \mathscr{A}$ (indexed over $i \in \mathbb{N}_+$) such that, for all $i \in \mathbb{N}_+$ and $k \in \mathbb{N}_+$ with $t_i \leqslant k < \ell_i$, one finds

$$\|g_k\| \geqslant \varepsilon \;\text{ and }\; \|g_{\ell_i}\| < \varepsilon. \tag{3.4}$$

Let us now restrict our attention to indices in the infinite index set

$$\mathscr{K} := \{k \in \mathscr{A} : t_i \leqslant k < \ell_i \text{ for some } i \in \mathbb{N}_+\}.$$

Observe from (3.4) that, for all $k \in \mathscr{K}$, it follows that $\|g_k\| \geqslant \varepsilon$. Also, from the definition of $\mathscr{A}$,

$$f_k - f_{k+1} \geqslant \eta \|s_k\|^3 \;\text{ for all }\; k \in \mathscr{K} \subseteq \mathscr{A}. \tag{3.5}$$

Since $\{f_k\}$ is monotonically decreasing and bounded below one finds that $\{f_k\} \to \underline{f}$ for some $\underline{f} \in \mathbb{R}$, which when combined with (3.5) shows that

$$\lim_{k \in \mathscr{K}, k \to \infty} \|s_k\| = 0. \tag{3.6}$$

Using this fact, Lemma 3.6, Assumption 3.2 and the Mean Value Theorem (as it is used in the proof of Lemma 3.8 to yield (3.1)), it follows that, for all sufficiently large $k \in \mathscr{K}$, one has

$$\begin{aligned} f_k - f_{k+1} &= f_k - q_k(s_k) + q_k(s_k) - f(x_k + s_k) \\ &\geqslant \frac{\|g_k\|}{6\sqrt{2}} \min\left\{ \frac{\|g_k\|}{1 + H_{\max}}, \frac{\|s_k\|}{\sqrt{6}} \sqrt{\frac{\|g_k\|}{2\|g_k\| + \|s_k\|\left(\frac{3}{2}H_{\max} + \kappa_1\right)}} \right\} - \left(g_{Lip} + \tfrac{1}{2}H_{\max}\right)\|s_k\|^2 \\ &\geqslant \frac{\varepsilon}{6\sqrt{2}} \frac{\|s_k\|}{\sqrt{6}} \sqrt{\frac{\|g_k\|}{2\|g_k\| + \|s_k\|\left(\frac{3}{2}H_{\max} + \kappa_1\right)}} - \left(g_{Lip} + \tfrac{1}{2}H_{\max}\right)\|s_k\|^2. \end{aligned}$$

It now follows from (3.4) and (3.6) that, as $k \to \infty$ over $k \in \mathscr{K}$, the square root term in the previous inequality converges to $1/\sqrt{2}$. Since the second term in the previous inequality is of order $\|s_k\|^2$ the first term is of order $\|s_k\|$, and $1/\sqrt{2} > 1/\sqrt{3}$, one can thus conclude that $f_k - f_{k+1} \geqslant \varepsilon\|s_k\|/36$ for all sufficiently large $k \in \mathscr{K}$. Consequently, it follows that, for all sufficiently large $i \in \mathbb{N}_+$, one finds

$$\begin{aligned} \left\|x_{t_i} - x_{\ell_i}\right\| &\leqslant \sum_{k \in \mathscr{K}, k = t_i}^{\ell_i - 1} \|x_k - x_{k+1}\| \\ &= \sum_{k \in \mathscr{K}, k = t_i}^{\ell_i - 1} \|s_k\| \leqslant \sum_{k \in \mathscr{K}, k = t_i}^{\ell_i - 1} \frac{36}{\varepsilon}\left(f_k - f_{k+1}\right) = \frac{36}{\varepsilon}\left(f_{t_i} - f_{\ell_i}\right). \end{aligned}$$

Since $\{f_{t_i} - f_{\ell_i}\} \to 0$ (recall that $\{f_k\} \to \underline{f}$ monotonically) this implies that $\{\|x_{t_i} - x_{\ell_i}\|\} \to 0$, which, in turn, implies that $\{\|g_{t_i} - g_{\ell_i}\|\} \to 0$ because of the continuity of g. However, this is a contradiction

since, for any $i \in \mathbb{N}_+$, we have $\|g_{t_i} - g_{\ell_i}\| \geqslant \varepsilon$ by the definitions of $\{t_i\}$ and $\{\ell_i\}$. Overall, we conclude that our initial supposition must be false implying that (3.3) holds. □

3.2 *First-order complexity*

Our next goal is to prove, with respect to a prescribed positive threshold, a worst-case upper bound on the number of iterations required for our algorithm to reduce the norm of the gradient below the threshold. In this subsection, along with Assumptions 2.1, 3.1 and 3.2, we add the following.

ASSUMPTION 3.12 The Hessian function H is Lipschitz continuous on a path defined by the sequence of iterates and trial steps; in particular, it is Lipschitz continuous with a scalar Lipschitz constant $H_{\text{Lip}} > 0$ on the set $\{x_k + \tau s_k : k \in \mathbb{N}_+, \tau \in [0, 1]\}$.

We begin our analysis in this subsection by providing a lemma that shows that successful steps always result if λ_k is sufficiently large relative to the size of the step.

LEMMA 3.13 For any $k \in \mathbb{N}_+$ if the pair (s_k, λ_k) satisfies

$$\lambda_k \geqslant (H_{\text{Lip}} + \kappa_2 + 2\eta)\|s_k\| \tag{3.7}$$

then $\rho_k \geqslant \eta$.

Proof. It follows from Assumption 3.12 and Taylor's expansion with Lagrange remainder that there exists $\bar{x}_k$ on the line segment $[x_k, x_k + s_k]$ such that

$$q_k(s_k) - f(x_k + s_k) = \tfrac{1}{2} s_k^T (H_k - H(\bar{x}_k)) s_k \geqslant -\tfrac{1}{2} H_{\text{Lip}} \|s_k\|^3. \tag{3.8}$$

Also, it follows from (2.3b) that

$$\begin{aligned} f_k - q_k(s_k) &= -g_k^T s_k - \tfrac{1}{2} s_k^T H_k s_k \\ &= -s_k^T (g_k + (H_k + \lambda_k I) s_k) + \tfrac{1}{2}\lambda_k \|s_k\|^2 + \tfrac{1}{2} s_k^T (H_k + \lambda_k I) s_k \\ &\geqslant -\tfrac{1}{2} s_k^T (H_k + \lambda_k I) s_k - \tfrac{1}{2}\kappa_2 \|s_k\|^3 + \tfrac{1}{2}\lambda_k \|s_k\|^2 + \tfrac{1}{2} s_k^T (H_k + \lambda_k I) s_k \\ &= -\tfrac{1}{2}\kappa_2 \|s_k\|^3 + \tfrac{1}{2}\lambda_k \|s_k\|^2. \end{aligned} \tag{3.9}$$

From (3.8) and (3.9), it follows that

$$\begin{aligned} f_k - f(x_k + s_k) &= f_k - q_k(s_k) + q_k(s_k) - f(x_k + s_k) \\ &\geqslant \tfrac{1}{2}\lambda_k \|s_k\|^2 - \tfrac{1}{2}\kappa_2 \|s_k\|^3 - \tfrac{1}{2} H_{\text{Lip}} \|s_k\|^3, \end{aligned}$$

which together with (3.7) implies that $\rho_k \geqslant \eta$, as claimed. □

We now prove that the sequence $\left\{\sigma_k^{\text{U}}\right\}$ is bounded above.

LEMMA 3.14 There exists a scalar constant $\sigma_{\max} \in \mathbb{R}_{++}$ such that, for all $k \in \mathbb{N}_+$,

$$\sigma_k^{\text{U}} \leqslant \sigma_{\max}.$$

Proof. Consider any $k \in \mathbb{N}_+$. If s_k is accepted (i.e., $k \in \mathscr{A}$) then $\sigma_{k+1}^{\mathrm{U}} \leftarrow \sigma_k^{\mathrm{U}}$. On the other hand, if s_k is rejected (i.e., $k \in \mathscr{R}$), then it follows from Step 12 and Step 14 that $\sigma_{k+1}^{\mathrm{U}} \leqslant \max\{\overline{\sigma}, \gamma_2\lambda_k/\|s_k\|\}$. Moreover, since $k \in \mathscr{R}$, meaning that $\rho_k < \eta$, it follows from Lemma 3.13 that $\lambda_k/\|s_k\|$ is bounded above by $(H_{\mathrm{Lip}} + \kappa_2 + 2\eta)$. Thus, it follows that $\sigma_{k+1}^{\mathrm{U}} \leqslant \max\{\overline{\sigma}, \gamma_2(H_{\mathrm{Lip}} + \kappa_2 + 2\eta)\}$ for all $k \in \mathscr{R}$. Overall, the desired result follows for any $\sigma_{\max} \geqslant \max\{\overline{\sigma}, \gamma_2(H_{\mathrm{Lip}} + \kappa_2 + 2\eta)\}$. □

We now establish a lower bound on the norm of any accepted trial step.

LEMMA 3.15 For all $k \in \mathscr{A}$, it follows that

$$\|s_k\| \geqslant \left(\tfrac{1}{2}H_{\mathrm{Lip}} + 2\sigma_{\max} + \kappa_3\right)^{-1/2} \|g_{k+1}\|^{1/2}.$$

Proof. Let $k \in \mathscr{A}$. It follows that

$$\begin{aligned}\|g_{k+1}\| &\leqslant \|g_{k+1} - (g_k + (H_k + \lambda_k I)s_k)\| + \|g_k + (H_k + \lambda_k I)s_k\| \\ &\leqslant \|g_{k+1} - (g_k + H_k s_k)\| + \lambda_k\|s_k\| + \|g_k + (H_k + \lambda_k I)s_k\|. \end{aligned} \tag{3.10}$$

By Taylor's theorem and Assumption 3.12 the first term on the right-hand side of this inequality satisfies

$$\begin{aligned}\|g_{k+1} - (g_k + H_k s_k)\| &\leqslant \left\| \int_0^1 (H(x_k + \tau s_k) - H_k)s_k \,\mathrm{d}\tau \right\| \\ &\leqslant \int_0^1 \|H(x_k + \tau s_k) - H_k\| \,\mathrm{d}\tau \cdot \|s_k\| \\ &\leqslant \int_0^1 \tau \,\mathrm{d}\tau \cdot H_{\mathrm{Lip}}\|s_k\|^2 = \tfrac{1}{2}H_{\mathrm{Lip}}\|s_k\|^2.\end{aligned}$$

Combining this with (3.10) and observing Step 5, (2.3c) and Lemma 3.14, it follows that

$$\begin{aligned}\|g_{k+1}\| &\leqslant \tfrac{1}{2}H_{\mathrm{Lip}}\|s_k\|^2 + 2\frac{\lambda_k}{\|s_k\|}\|s_k\|^2 + \kappa_3\|s_k\|^2 \\ &\leqslant \tfrac{1}{2}H_{\mathrm{Lip}}\|s_k\|^2 + 2\sigma_{\max}\|s_k\|^2 + \kappa_3\|s_k\|^2,\end{aligned}$$

which, after rearrangement, completes the proof. □

We are now prepared to prove a worst-case upper bound on the total number of accepted steps that may occur for iterations, in which the norm of the gradient of the objective is above a positive threshold.

LEMMA 3.16 For any $\varepsilon \in \mathbb{R}_{++}$ the total number of elements in the index set

$$\mathscr{K}_\varepsilon := \left\{k \in \mathbb{N}_+ : k \geqslant 1,\ (k-1) \in \mathscr{A},\ \|g_k\| > \varepsilon\right\}$$

is at most

$$\left\lfloor \left(\frac{f_0 - f_{\inf}}{\eta \left(\frac{1}{2} H_{\text{Lip}} + 2\sigma_{\max} + \kappa_3 \right)^{-3/2}} \right) \varepsilon^{-3/2} \right\rfloor =: N_{\mathscr{A}}(\varepsilon) \geqslant 0. \tag{3.11}$$

Proof. The proof follows in a similar manner as that of Lemma 3.20 in Curtis *et al.* (2017). By Lemma 3.15, it follows that for all $k \in \mathscr{K}_\varepsilon$ one finds

$$\begin{aligned} f_{k-1} - f_k &\geqslant \eta \|s_{k-1}\|^3 \\ &\geqslant \eta \left(\tfrac{1}{2} H_{\text{Lip}} + 2\sigma_{\max} + \kappa_3 \right)^{-3/2} \|g_k\|^{3/2} \\ &\geqslant \eta \left(\tfrac{1}{2} H_{\text{Lip}} + 2\sigma_{\max} + \kappa_3 \right)^{-3/2} \varepsilon^{3/2}. \end{aligned}$$

In addition, it follows from Theorem 3.11 that $|\mathscr{K}_\varepsilon| < \infty$. Hence, the reduction in f obtained up to the largest index in $\mathscr{K}_\varepsilon$, call it $\bar{k}_\varepsilon$, satisfies

$$f_0 - f_{\bar{k}_\varepsilon} = \sum_{k=1}^{\bar{k}_\varepsilon} (f_{k-1} - f_k) \geqslant \sum_{k \in \mathscr{K}_\varepsilon} (f_{k-1} - f_k) \geqslant |\mathscr{K}_\varepsilon| \eta \left(\tfrac{1}{2} H_{\text{Lip}} + 2\sigma_{\max} + \kappa_3 \right)^{-3/2} \varepsilon^{3/2}.$$

Rearranging this inequality to yield an upper bound for $|\mathscr{K}_\varepsilon|$ and using the fact that $f_0 - f_{\inf} \geqslant f_0 - f_{\bar{k}_\varepsilon}$, one obtains the desired result. □

In order to prove a result similar to Lemma 3.16 for the *total* number of iterations with $\|g_k\| > \varepsilon$, we require an upper bound on the total number of trial steps that may be rejected between accepted steps. To this end let us define, for a given $\hat{k} \in \mathscr{A} \cup \{0\}$, the iteration number and corresponding set

$$\begin{aligned} k_{\mathscr{A}}(\hat{k}) &:= \min \left\{ k \in \mathscr{A} : k > \hat{k} \right\} \\ \text{and } \mathscr{I}(\hat{k}) &:= \left\{ k \in \mathbb{N}_+ : \hat{k} < k < k_{\mathscr{A}}(\hat{k}) \right\}, \end{aligned}$$

i.e., we let $k_{\mathscr{A}}(\hat{k})$ be the smallest of all iteration numbers in $\mathscr{A}$ that is strictly larger than $\hat{k}$, and we let $\mathscr{I}(\hat{k})$ be the set of iteration numbers between $\hat{k}$ and $k_{\mathscr{A}}(\hat{k})$.

We now show that the number of rejected steps between the first iteration and the first accepted step or between consecutive accepted steps is bounded above.

LEMMA 3.17 For any $\hat{k} \in \mathscr{A} \cup \{0\}$, it follows that

$$\left| \mathscr{I}(\hat{k}) \right| \leqslant 1 + \left\lfloor \frac{1}{\log(\gamma_1)} \log \left(\frac{\sigma_{\max}}{\underline{\sigma}} \right) \right\rfloor =: N_{\mathscr{R}} \geqslant 0.$$

Proof. The proof follows in a similar manner as for Lemma 3.24 in Curtis *et al.* (2017). First, the result holds trivially if $|\mathscr{I}(\hat{k})| = 0$. Thus, we may assume that $|\mathscr{I}(\hat{k})| \geqslant 1$. Since $(\hat{k}+1) \in \mathscr{R}$ by construction,

it follows from Steps 11–14 and Step 5 that $\lambda_{\hat{k}+2}/\|s_{\hat{k}+2}\| \geqslant \underline{\sigma}$, which, due to the lower bound on $\lambda_{k+1}/\|s_{k+1}\|$ in Step 14 and Step 5, leads to

$$\lambda_{k_{\mathscr{A}}(\hat{k})} \geqslant \underline{\sigma}\,(\gamma_1)^{k_{\mathscr{A}}(\hat{k})-\hat{k}-2}\left\|s_{k_{\mathscr{A}}(\hat{k})}\right\|.$$

Combining this with Step 5 and Lemma 3.14 shows that

$$\sigma_{\max} \geqslant \sigma^{\mathrm{U}}_{k_{\mathscr{A}}(\hat{k})} \geqslant \lambda_{k_{\mathscr{A}}(\hat{k})}/\left\|s_{k_{\mathscr{A}}(\hat{k})}\right\| \geqslant \underline{\sigma}\,(\gamma_1)^{k_{\mathscr{A}}(\hat{k})-\hat{k}-2}.$$

After rearrangement it now follows that

$$k_{\mathscr{A}}(\hat{k}) - \hat{k} - 2 \leqslant \frac{1}{\log(\gamma_1)}\log\left(\frac{\sigma_{\max}}{\underline{\sigma}}\right).$$

The desired result follows from this inequality since $|\mathscr{I}(\hat{k})| = k_{\mathscr{A}}(\hat{k}) - \hat{k} - 1$. □

We are now prepared to prove our main complexity result of this subsection.

THEOREM 3.18 Under Assumptions 2.1, 3.1, 3.2 and 3.12 for a scalar $\varepsilon \in \mathbb{R}_{++}$, the total number of elements in the index set $\{k \in \mathbb{N}_+ : \|g_k\| > \varepsilon\}$ is at most

$$N(\varepsilon) := 1 + N_{\mathscr{R}} N_{\mathscr{A}}(\varepsilon), \tag{3.12}$$

where $N_{\mathscr{A}}(\varepsilon)$ and $N_{\mathscr{R}}$ are defined in Lemmas 3.16 and 3.17, respectively. Consequently, for any $\overline{\varepsilon} \in \mathbb{R}_{++}$, it follows that $N(\varepsilon) = \mathscr{O}(\varepsilon^{-3/2})$ for all $\varepsilon \in (0, \overline{\varepsilon}]$.

Proof. Without loss of generality we may assume that at least one iteration is performed. Lemma 3.16 guarantees that the total number of elements in the index set $\{k \in \mathscr{A} : k \geqslant 1,\ \|g_k\| > \varepsilon\}$ is at most $N_{\mathscr{A}}(\varepsilon)$, where, immediately prior to each of the corresponding accepted steps, Lemma 3.17 guarantees that at most $N_{\mathscr{R}}$ trial steps are rejected. Accounting for the first iteration the desired result follows. □

3.3 *Second-order global convergence and complexity*

Our goal in this subsection is to prove results showing that, in some sense, the algorithm converges to second-order stationarity and does so with a worst-case iteration complexity on par with the methods in Cartis *et al.* (2011b) and Curtis *et al.* (2017). In particular, our results show that if the algorithm computes each search direction to satisfy a curvature condition over a subspace, then second-order stationarity is reached in a manner that depends on the subspaces.

In this subsection we make the following additional assumption about the subproblem solver.

ASSUMPTION 3.19 For all $k \in \mathbb{N}_+$ let $\mathscr{L}_k \subseteq \mathbb{R}^n$ denote a subspace with an orthonormal basis formed from the columns of a matrix R_k. The step s_k satisfies

$$\xi\left(R_k^T H_k R_k\right) \geqslant -\kappa_4\|s_k\| \tag{3.13}$$

for some $\kappa_4 \in \mathbb{R}_+$, where $\xi(R_k^T H_k R_k)$ indicates the smallest eigenvalue of $R_k^T H_k R_k$.

This assumption is reasonable, e.g., in cases when s_k is computed by solving problem 2.1 with the component s restricted to a subspace of $\mathbb{R}^n$. We refer the reader to Theorem B6 for a proof of this fact, which also reveals that this assumption is congruous with Assumption 2.1.

Under this assumption we have the following second-order convergence result.

THEOREM 3.20 Suppose Assumptions 2.1, 3.1, 3.2, 3.12 and 3.19 hold. It follows that

$$\liminf_{k\in\mathscr{A},k\to\infty} \xi\left(R_k^T H_k R_k\right) \geqslant 0.$$

Proof. The result follows from (3.13) since $\{s_k\}_{k\in\mathscr{A}} \to 0$ (see Lemma 3.9). □

As a consequence of Theorem 3.20 if the sequence $\{R_k\}_{k\in\mathscr{A}}$ tends toward full dimensionality as $k \to \infty$, then any limit point x_* of $\{x_k\}$ must have $H(x_*) \succeq 0$.

Our next goal is to prove a worst-case iteration complexity result for achieving second-order stationarity in a sense similar to that in Theorem 3.20. Toward this end we first prove the following lemma, which is similar to Lemma 3.16.

LEMMA 3.21 For any $\varepsilon \in \mathbb{R}_{++}$ the total number of elements in the index set

$$\mathscr{K}_{\varepsilon,\xi} := \left\{k \in \mathbb{N}_+ : k \geqslant 1,\ (k-1) \in \mathscr{A},\ \xi\left(R_k^T H_k R_k\right) < -\varepsilon\right\}$$

is at most

$$\left\lfloor \left(\frac{f_0 - f_{\inf}}{\eta\kappa_4^{-3}}\right)\varepsilon^{-3} \right\rfloor =: N_{\mathscr{A},\xi}(\varepsilon) \geqslant 0. \tag{3.14}$$

Proof. Under Assumption 3.19, it follows that, for all $k \in \mathscr{K}_{\varepsilon,\xi}$, one finds

$$f_{k-1} - f_k \geqslant \eta\|s_{k-1}\|^3 \geqslant \eta\left(\frac{-\xi\left(R_k^T H_k R_k\right)}{\kappa_4}\right)^3 \geqslant \eta\kappa_4^{-3}\varepsilon^3.$$

It follows from this inequality, the fact that f is monotonically decreasing over the sequence of iterates and Assumption 3.1 that

$$f_0 - f_{\inf} \geqslant \sum_{k\in\mathscr{K}_{\varepsilon,\xi}} \left(f_{k-1} - f_k\right) \geqslant |\mathscr{K}_{\varepsilon,\xi}|\eta\kappa_4^{-3}\varepsilon^3.$$

Rearranging this inequality to yield an upper bound for $|\mathscr{K}_{\varepsilon,\xi}|$ gives the result. □

We close with the following second-order complexity result.

THEOREM 3.22 Under Assumptions 2.1, 3.1, 3.2, 3.12 and 3.19 for any pair of scalars $(\varepsilon_1, \varepsilon_2) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$, the number of elements in the index set

$$\left\{k \in \mathbb{N}_+ : \|g_k\| > \varepsilon_1 \ \vee\ \xi\left(R_k^T H_k R_k\right) < -\varepsilon_2\right\}$$

is at most

$$N(\varepsilon_1, \varepsilon_2) := 1 + N_{\mathscr{R}} \max\left\{N_{\mathscr{A}}(\varepsilon_1), N_{\mathscr{A},\xi}(\varepsilon_2)\right\}, \tag{3.15}$$

where $N_{\mathscr{A}}(\cdot)$, $N_{\mathscr{R}}$ and $N_{\mathscr{A},\xi}(\cdot)$ are defined in Lemmas 3.16, 3.17 and 3.21, respectively. Consequently, for any pair of scalars $(\overline{\varepsilon}_1, \overline{\varepsilon}_2) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$, it follows that

$$N(\varepsilon_1, \varepsilon_2) = \mathscr{O}\left(\max\left\{\varepsilon_1^{-3/2}, \varepsilon_2^{-3}\right\}\right) \text{ for all } (\varepsilon_1, \varepsilon_2) \in (0, \overline{\varepsilon}_1] \times (0, \overline{\varepsilon}_2].$$

Proof. The proof follows in a similar manner as that of Theorem 3.18 by additionally incorporating the bound proved in Lemma 3.21. □

4. Algorithm instances

Algorithm 1 is a broad framework containing, among other algorithms, ARC and TRACE. Indeed the proposed framework and its supporting analyses cover a wide range of algorithms as long as the pairs in the sequence $\{(s_k, \lambda_k)\}$ satisfy Assumption 2.1.

In this section we show that ARC and TRACE are special cases of our proposed framework in that the steps these algorithms accept would also be acceptable for our framework, and that the procedures followed by these methods after a step is rejected are consistent with our framework. We then introduce an instance of our framework that is new to the literature. (If desired for the guarantees in Section 3.3, one could also mind whether the elements in the sequence $\{(s_k, \lambda_k)\}$ satisfy Assumption 3.19. However, for brevity in this section let us suppose that one is interested only in Assumption 2.1.)

4.1 ARC *as a special case*

The ARC method, which was inspired by the work in Griewank (1981) and Nesterov & Polyak (2006), was first proposed and analyzed in Cartis *et al.* (2011a,b). In these papers various sets of step computation conditions are considered involving exact and inexact subproblem solutions yielding different types of convergence and worst-case complexity guarantees. For our purposes here we consider the more recent variant of ARC stated and analyzed as '$\text{AR}p$' with $p = 2$ in Birgin *et al.* (2017). (For ease of comparison, we consider this algorithm when their regularization parameter update—see Step 4 in their algorithm—uses $\eta_1 = \eta_2$. Our algorithm is easily extended to employ a two-tier acceptance condition, involving two thresholds η_1 and η_2, as is used in (Birgin *et al.* 2017) and (Cartis *et al.* 2011a,b).)

Suppose that a trial step s_k is computed by this version of ARC. In particular, let us make the reasonable assumption that the subproblem for which s_k is an approximate solution is defined by some regularization value $\sigma_k \in [\sigma_k^{\text{L}}, \sigma_k^{\text{U}}]$ (with $\sigma_k^{\text{L}} \geqslant \sigma_{\min}$ since ARC ensures that $\sigma_k \geqslant \sigma_{\min} \in \mathbb{R}_{++}$ for all $k \in \mathbb{N}$), and that this subproblem is minimized over a subspace $\mathscr{L}_k$ such that $g_k \in \mathscr{L}_k$ (see Appendix B). As is shown using a similar argument as in the proof of our Theorem B6(b), one can show under these conditions that (s_k, λ_k) with $\lambda_k = \sigma_k \|s_k\|$ satisfies (2.3a). In addition, considering the algorithm statement in Birgin *et al.* (2017), but using our notation, one is required to have

$$g_k^T s_k + \tfrac{1}{2} s_k^T H_k s_k + \lambda_k \|s_k\|^2 < 0 \text{ and } \|g_k + (H_k + \lambda_k I) s_k\| \leqslant \theta \|s_k\|^2 \text{ for some } \theta \in \mathbb{R}_{++}.$$

It is easily seen that (s_k, λ_k) satisfying these conditions also satisfies (2.3b) and (2.3c) for any $(\kappa_1, \kappa_2, \kappa_3)$ such that $\kappa_1 \geqslant \frac{1}{2}H_{\max}$ and $\kappa_3 \geqslant \theta$. Overall, we have shown that a trial step s_k computed by this version of ARC satisfies Assumption 2.1, meaning that it satisfies the condition in Step 5 in Algorithm 1. If this trial step is accepted by ARC, then this means that $f_k - f(x_k + s_k) \geqslant \eta_1(f_k - q_k(s_k))$. Along with Birgin *et al.* (2017, Lemma 2.1) this implies that $f_k - f(x_k + s_k) \geqslant \frac{1}{3}\eta\sigma_k\|s_k\|^3$, meaning that $\rho_k \geqslant \frac{1}{3}\eta_1\sigma_{\min}$. Hence, this trial step would also be accepted in Algorithm 1 under the assumption that $\eta \in \left(0, \frac{1}{3}\eta_1\sigma_{\min}\right]$.

Finally, if a trial step is rejected in this version of ARC, then σ_{k+1} is set to a positive multiple of σ_k. This is consistent with the procedure after a step rejection in Algorithm 1, where it is clear that, with appropriate parameter choices, one would find $\sigma_{k+1} \in [\sigma_{k+1}^{\mathrm{L}}, \sigma_{k+1}^{\mathrm{U}}]$.

4.2 TRACE *as a special case*

TRACE is proposed and analyzed in Curtis *et al.* (2017). Our goal in this subsection is to show that, with certain parameter settings, a trial step that is computed and accepted by TRACE could also be one that is computed and accepted by Algorithm 1, and that the procedures for rejecting a step in TRACE are consistent with those in Algorithm 1. Among other procedures TRACE involves dynamic updates for two sequences, $\{\delta_k\}$ and $\{\Delta_k\}$. The elements of $\{\delta_k\}$ are the trust region radii, while $\{\Delta_k\}$ is a monotonically nondecreasing sequence of upper bounds for the trust region radii; consequently, $\|s_k\| \leqslant \delta_k \leqslant \Delta_k$ with $\Delta_{k+1} \geqslant \Delta_k$ for all $k \in \mathbb{N}$. For simplicity in our discussion here let us assume that $\|s_k\| < \Delta_k$ for all $k \in \mathbb{N}$. This is a fair assumption since, as shown in Curtis *et al.* (2017, Lemma 3.11), the manner in which $\{\Delta_k\}$ is set ensures that $\|s_k\| = \Delta_k$ only a finite number of times in any run.

In TRACE during iteration $k \in \mathbb{N}$, a trust region radius $\delta_k \in \mathbb{R}_{++}$ is given and a trial step s_k and regularization value λ_k are computed, satisfying the standard trust region subproblem optimality conditions

$$g_k + (H_k + \lambda_k I)s_k = 0, \quad H_k + \lambda_k I \succeq 0, \quad \text{and} \quad \lambda_k(\delta_k - \|s_k\|) = 0, \quad \text{where} \quad (\lambda_k, \delta_k - \|s_k\|) \geqslant 0.$$

By the first of these conditions the pair (s_k, λ_k) clearly satisfies (2.3b) and (2.3c). In addition, one can use standard trust region theory, in particular, related to Cauchy decrease (see Conn *et al.*, 2000 or Nocedal & Wright, 2006), to show that the pair also satisfies (2.3a). Overall, assuming that the pair $\left(\sigma_k^{\mathrm{L}}, \sigma_k^{\mathrm{U}}\right)$ is set such that $\lambda_k/\|s_k\| \in [\sigma_k^{\mathrm{L}}, \sigma_k^{\mathrm{U}}]$, it follows that Assumption 2.1 is satisfied, meaning that TRACE offers the condition in Step 5 in Algorithm 1. If the trial step s_k is subsequently accepted by TRACE, then it would also be accepted by Algorithm 1 since both algorithms use the same step acceptance condition.

Now suppose that a trial step is not accepted in TRACE. This can occur in two circumstances. It can occur if $\rho_k \geqslant \eta$ while $\lambda_k > \sigma_k\|s_k\|$, in which case the trust region radius is *expanded* and a new subproblem is solved. By the proof of Curtis *et al.* (2017, Lemma 3.7), the solution of this new subproblem yields (in iteration $k+1$ in TRACE) the relationship that $\lambda_{k+1}/\|s_{k+1}\| \leqslant \sigma_{k+1} = \sigma_k$. Hence, under the same assumption as above that the pair $(\sigma_k^{\mathrm{L}}, \sigma_k^{\mathrm{U}})$ is set such that $\lambda_k/\|s_k\| \in [\sigma_k^{\mathrm{L}}, \sigma_k^{\mathrm{U}}]$, this shows that the procedure in TRACE involving *an expansion of the trust region radius and the computation of the subsequent trial step* yields a trial step that would be offered in a *single iteration* in Algorithm 1. The other circumstance in which a trial step is rejected in TRACE is when $\rho_k < \eta$, in which case the trust region radius is contracted. In this case one can see that the outcome of the CONTRACT subroutine in TRACE is consistent with Steps 11–14 of Algorithm 1, in the sense that the solution of the subsequent subproblem in TRACE will have $\lambda_{k+1}/\|s_{k+1}\| \in [\underline{\sigma}, \overline{\sigma}]$ (if $\lambda_k < \underline{\sigma}\|s_k\|$) or $\lambda_{k+1}/\|s_{k+1}\|$ within a range defined by positive multiples of $\lambda_k/\|s_k\|$; see Lemmas 3.17 and 3.23 in Curtis *et al.* (2017).

4.3 *A hybrid algorithm*

The primary distinguishing feature of our algorithm instance is the manner in which we compute the pair (s_k, λ_k) in Step 5 of Algorithm 1. Our newly proposed hybrid algorithm considers two cases.

Case 1: $\sigma_k^{\mathrm{L}} > 0$. In this case we find a pair (s_k, λ_k) by solving problem (A.3) over a sequence of increasingly higher-dimensional Krylov subspaces as described in Cartis *et al.* (2011a) until (2.3) and (3.13) are satisfied. The reason we know that (2.3) and (3.13) will eventually be satisfied can be seen as follows. Solving problem (A.3) over a Krylov subspace is equivalent to solving problem (B.3), with an appropriate choice of R_k as a basis for that Krylov subspace, then setting $s_k = R_k v_k$. Then it follows from Theorem B4(i) that solving (B.3) is equivalent to solving (B.2), which in turn is equivalent to solving (B.1) in the sense that if $(v_k, \lambda_k, \beta_k^{\mathrm{L}}, \beta_k^{\mathrm{U}}, \beta_k^{\mathrm{N}})$ is a first-order primal-dual solution of problem (B.2), then $(s_k, \lambda_k, \beta_k^{\mathrm{L}}, \beta_k^{\mathrm{U}}, \beta_k^{\mathrm{N}})$ with $s_k = R_k v_k$ is a solution of problem (B.1). Finally, we need only note from Theorem B6 that solutions to problem (B.1) satisfy (2.3a) for all Krylov subspaces $\mathscr{L}_k$ (recall that g_k is contained in all Krylov subspaces), (2.3b) for all Krylov subspaces, (2.3c) if the Krylov subspace $\mathscr{L}_k$ includes enough of the space (in the worst case, $\mathscr{L}_k = \mathbb{R}^n$) and (3.13) for all Krylov subspaces.

Case 2: $\sigma_k^{\mathrm{L}} = 0$. In this case we begin by applying the linear conjugate gradient (CG) method in an attempt to solve the linear system $H_k s = -g_k$, which iteratively solves

$$\min_{s \in \mathbb{R}^n} \; q_k(s) \tag{4.1}$$

over a sequence of expanding Krylov subspaces. One of two outcomes is possible. First, the CG algorithm may ultimately identify a vector s_k such that (s_k, λ_k) with $\lambda_k = 0$ satisfies (2.3) and (3.13). Secondly, the CG algorithm may never identify a vector s_k such that (s_k, λ_k) with $\lambda_k = 0$ satisfies (2.3) and (3.13). Indeed this might occur if CG encounters a direction of negative curvature—in which case we terminate CG immediately—or if CG solves (4.1) accurately or reaches an iteration limit, and yet at least one condition in (2.3) or (3.13) is not satisfied. In such a case we choose to reset $\sigma_k^{\mathrm{L}} \in (0, \sigma_k^{\mathrm{U}}]$ then solve problem (A.3) over a sequence of expanding Krylov subspaces as described in Case 1. In this manner we are guaranteed to identify a pair (s_k, λ_k) satisfying (2.3) and (3.13) as required.

5. Implementation and numerical results

We implemented two algorithms in `MATLAB`, one following the strategy in Section 4.3 and, for comparison purposes, one following the ARC algorithm in Cartis *et al.* (2011b) with ideas from Birgin *et al.* (2017). We refer to our implementation of the former as `iR_Newton`, for inexact Regularized Newton, and to our implementation of the latter as `iARC`, for inexact ARC. In this section we describe our approach for computing the pairs $\{(s_k, \lambda_k)\}$ in `iR_Newton` and `iARC`, as well as other implementation details, and discuss the results of numerical experiments on a standard set of nonlinear optimization test problems.

5.1 *Implementation details*

Let us begin by noting that the implemented algorithms terminate in iteration $k \in \mathbb{N}_+$ if

$$\|g_k\|_\infty \leqslant 10^{-6} \max\{\|g_0\|_\infty, 1\}.$$

TABLE 1 *Input parameters for* `iARC` *and* `iR_Newton`

η_1	`1.0e-16`	γ_0	`2.0e-01`	κ_1	`1.0e+00`	$\underline{\sigma}$	`1.0e-10`
η_2	`1.0e-01`	γ_1	`1.0e+01`	κ_2	`1.0e+00`	$\overline{\sigma}$	`1.0e+20`
		γ_2	`2.0e+02`	κ_3	`1.0e+00`		

We chose not to employ a termination test based on a second-order stationarity condition. Correspondingly, neither of the algorithms check a second-order condition when computing a trial step; e.g., in `iR_Newton`, we are satisfied with a step satisfying (2.3) and do not check (3.13). In addition, for practical purposes we set a maximum iteration limit of 10^6, a time limit of 4 hours and a minimum step norm limit of 10^{-20}. For reference the input parameter values we used are given in Table 1. We chose these values as ones that worked well on our test set for both implemented algorithms.

For both implemented algorithms we employ a sequence $\{\sigma_k\}$ that is updated dynamically. In `iARC` this sequence is handled as described in Cartis *et al.* (2011b), namely,

$$\sigma_{k+1} \leftarrow \begin{cases} \max\{\underline{\sigma}, \gamma_0\sigma_k\} & \text{if } \frac{f_k - f(x_k+s_k)}{f_k - c_k(s_k;\sigma_k)} \geqslant \eta_2 \\ \sigma_k & \text{if } \frac{f_k - f(x_k+s_k)}{f_k - c_k(s_k;\sigma_k)} \in [\eta_1, \eta_2) \\ \gamma_1\sigma_k & \text{if } \frac{f_k - f(x_k+s_k)}{f_k - c_k(s_k;\sigma_k)} < \eta_1. \end{cases}$$

The value σ_k is used in defining $c_k(\cdot;\sigma_k)$ (recall (2.4)) that is minimized approximately to compute the trial step s_k for all $k \in \mathbb{N}_+$. In particular, the implementation iteratively constructs Krylov subspaces of increasing dimension using the Lanczos process, where for each subspace we employ the `RQS` function from the `GALAHAD` software library (see Gould *et al.*, 2003 and Gould *et al.*, 2010) to minimize $c_k(\cdot;\sigma_k)$ over the subspace. If the subspace is full-dimensional or the resulting step s_k satisfies

$$\|g_k + (H_k + \sigma_k\|s_k\|I)s_k\| \leqslant \kappa_3\|s_k\|^2 \tag{5.1}$$

then it is used as the trial step. Otherwise, the process continues with a larger subspace. We remark that condition (5.1) is more restrictive than our condition (2.3c), but we use it since it is one that has been proposed for cubic regularization methods; e.g., see (2.13) in Birgin *et al.* (2017).

One could employ more sophisticated techniques for setting the elements of the sequence $\{\sigma_k\}$ in `iARC` that attempt to reduce the number of rejected steps; e.g., see Gould *et al.* (2012). Such improvements might aid `iR_Newton` as well. However, for simplicity and to avoid the need for additional parameter tuning, we did not include such enhancements in our implemented algorithms.

As for `iR_Newton`, for consistency between the two implementations, we do not explicitly compute the sequence $\{\lambda_k\}$, but rather employ $\{\sigma_k^{\text{L}}\|s_k\|\}$ in its place. For example, whenever an acceptable step is computed with $\sigma_k^{\text{L}} = 0$, then, as described in **Case 2** in Section 4.3, we effectively use $\lambda_k = 0$. On the other hand, when $\sigma_k^{\text{L}} > 0$ we employ the same iterative approach as used for `iARC` to compute the trial step s_k as an approximate minimizer of $c_k(\cdot;\sigma_k^{\text{L}})$, where in place of λ_k in (2.3) we employ $\sigma_k^{\text{L}}\|s_k\|$.

Then in either case in the remainder of iteration $k \in \mathbb{N}_+$, specifically for setting $\sigma_{k+1}^{\mathrm{L}}$ and $\sigma_{k+1}^{\mathrm{U}}$, we use $\sigma_k^{\mathrm{L}}\|s_k\|$ in place of λ_k in Steps 11 and 14. We also define an auxiliary sequence $\{\sigma_k\}$ using the update

$$\sigma_{k+1} \leftarrow \begin{cases} \max\left\{\underline{\sigma}, \gamma_0\sigma_k\right\} & \text{if } \rho_k \geqslant \eta_1 \text{ and } \sigma_k^{\mathrm{L}} > 0 \\ \sigma_k & \text{if } \sigma_k^{\mathrm{L}} = 0 \\ \min\left\{\gamma_1\sigma_k, \overline{\sigma}\right\} & \text{if } \rho_k < \eta_1 \text{ and } \sigma_k^{\mathrm{L}} > 0. \end{cases}$$

This update is similar to the one employed for `iARC` with the added assurance that $\{\sigma_k\} \subset [\underline{\sigma}, \overline{\sigma}]$. The elements of this sequence are used in two circumstances. First, if, as described in **Case 2** in Section 4.3, CG fails to produce a trial step s_k satisfying (2.3) (with $\lambda_k = 0$), then we reset $\sigma_k^{\mathrm{L}} \leftarrow \sigma_k$ and revert to the same scheme as above to compute the trial step when $\sigma_k^{\mathrm{L}} > 0$. Secondly, if a step is rejected and $\sigma_k^{\mathrm{L}} < \underline{\sigma}$ (equivalently, $\lambda_k < \underline{\sigma}\|s_k\|_2$ as in Step 12 in Algorithm 1), then we set $\sigma_{k+1}^{\mathrm{L}} \leftarrow \sigma_{k+1}$. Finally, we note that if CG ever performs n iterations and the resulting solution (due to numerical error) does not satisfy (2.3) and no negative curvature is detected, then the resulting approximate solution s_k is used as the trial step.

5.2 *Results on the `CUTEst` test set*

We employed our implemented algorithms, `iARC` and `iR_Newton`, to solve unconstrained problems in the `CUTEst` test set; see Gould *et al.* (2013). Among 171 unconstrained problems in the set one (`FLETCBV2`) was removed since the algorithms terminated at the initial point, five (`ARGLINC`, `DECONVU`, `FLETCHBV`, `INDEFM` and `POWER`) were removed due to a function evaluation error or our memory limitation of 8GB and nine (`EIGENBLS`, `EIGENCLS`, `FMINSURF`, `NONMSQRT`, `SBRYBND`, `SCURLY10`, `SCURLY20`, `SCURLY30` and `SSCOSINE`) were removed since neither algorithms terminated within our time limit. In addition, four were removed since neither of the algorithms terminated successfully: for `HIELOW`, `iARC` reached our maximum iteration limit; for `CURLY20` and `SCOSINE`, `iARC` reached the time limit; for `INDEF`, `iARC` terminated due to a subproblem solver error; and for all of these four problems, `iR_Newton` terminated due to our minimum step norm limit. The remaining set consisted of 152 test problems with number of variables ranging from 2 to 100,000. For additional details on the problems used and their sizes see Appendix C.

To compare the performance of the implemented algorithms we generated performance profiles for the number of iterations and number of Hessian-vector products required before termination. These are shown in Fig. 1. A performance profile graph of an algorithm at point α shows the fraction of the test set for which the algorithm is able to solve within a factor of 2^α of the best algorithm for the given measure; see Dolan & Moré (2002). When generating the profiles we did not include three of the test problems—`CURLY10`, `CURLY30` and `MODBEALE`—on which `iARC` was unsuccessful while `iR_Newton` was successful. (In particular, `iARC` reached the time limit for all problems.) We feel that this gives a fairer comparison with respect to the problems on which both algorithms were successful.

As seen in Fig. 1 the algorithms performed relatively comparably when it came to the number of iterations required, though clearly `iR_Newton` had an edge in terms of requiring fewer iterations on various problems. The difference in terms of numbers of Hessian-vector products required was more drastic, and indeed we point to this as the main measure of improved performance for `iR_Newton` versus `iARC`. One reason for this discrepancy is that `iR_Newton` required fewer iterations on some problems. However, more significantly, the difference was due in part to `iR_Newton`'s ability to

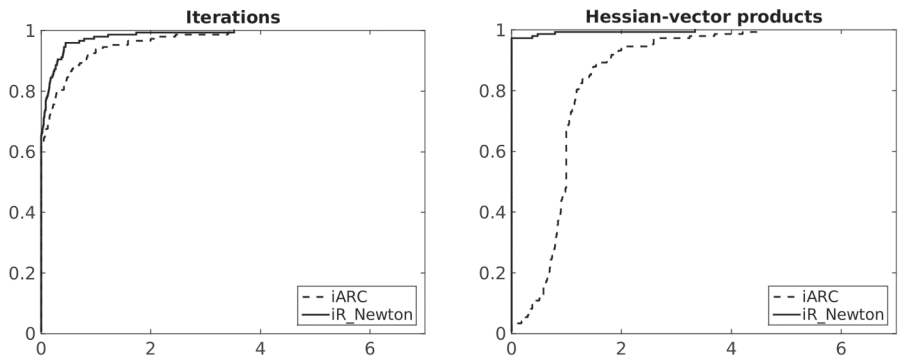

FIG. 1. Performance profiles for `iARC` and `iR_Newton`.

employ and accept inexact Newton steps (with $\lambda_k = 0$) on many iterations. This is due to the fact that, in CG, one is able to compute the Hessian-vector product $H_k s_k$, needed to check the termination conditions for the computation of s_k, by taking a linear combination of Hessian-vector products already computed in CG; i.e., if $\{p_{k,i}\}$ are the search directions computed in CG such that $s_k = \sum_i \alpha_{k,i} p_{k,i}$, then CG involves computing $H_k p_{k,i}$ for each i and can compute $H_k s_k = \sum_i \alpha_{k,i}(H_k p_{k,i})$. By contrast, one is unable to retrieve this product via a linear combination when the step is computed from the minimization of a cubic function, as is needed in `iARC` and in `iR_Newton` whenever $\sigma_k^{\mathrm{L}} > 0$. Overall, we claim that the primary strength of `iR_Newton` as compared to `iARC` is its ability to employ inexact Newton steps.

For further details of our numerical results see Appendix C. In these results we also indicate the number of tridiagonal factorizations required; at least one is needed involving a tridiagonal matrix of size $m \times m$ every time an algorithm solves a cubic subproblem over an m-dimensional subspace.

6. Conclusion

We have proposed a general framework for solving smooth nonconvex optimization problems and proceeded to prove worst-case iteration complexity bounds for it. In fact, for a certain class of second-order methods employed to minimize a certain class of nonconvex functions, our first-order complexity result for our method is known to be optimal; see Cartis *et al.* (2011c). Our framework is flexible enough to cover a wide range of popular algorithms, an achievement made possible by the use of generic conditions that each trial step is required to satisfy. The use of such conditions allows for the calculation of inexact Newton steps, for example by performing minimization over expanding Krylov subspaces. Although we have presented a particular instance of our framework motivated by subproblem (2.1), additional instances can easily be derived by applying other optimization strategies for solving (2.1). Numerical experiments with an instance of our algorithm showed that it can lead to improved performance on a broad test set, as compared to an implementation of a straightforward cubic regularization approach.

Acknowledgements

We thank the anonymous referees for their valuable comments, suggestions and corrections which helped to improve the paper. We are also grateful to the Associate Editor for handling the paper.

Funding

U.S. Department of Energy, Office of Science, Applied Mathematics, Early Career Research Program (DE–SC0010615); U.S. National Science Foundation, Division of Mathematical Sciences, Computational Mathematics Program (DMS–1016291).

References

Bazaraa, M. S., Sherali, H. D. & Shetty, C. M. (2006) *Nonlinear Programming: Theory and Algorithms*, 3rd edn. Hoboken, NJ: John Wiley, Inc.

Bertsekas, D. P. (1999) *Nonlinear Programming*, 2nd edn. Nashua, NH: Athena Scientific.

Birgin, E. G., Gardenghi, J., Martínez, J. M., Santos, S. A. & Toint, P. L. (2017) Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models. *Math. Program.*, **163**, 359–368.

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

Cartis, C., Gould, N. I. M. & Toint, P. L. (2010) On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization problems. *SIAM J. Optim.*, **20**, 2833–2852.

Cartis, C., Gould, N. I. M. & Toint, P. L. (2011a) Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Math. Program.*, **127**, 245–295.

Cartis, C., Gould, N. I. M. & Toint, P. L. (2011b) Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function- and derivative-evaluation complexity. *Math. Program.*, **130**, 295–319.

Cartis, C., Gould, N. I. M. & Toint, P. L. (2011c) Optimal Newton-type methods for nonconvex smooth optimization problems. *Technical Report ERGO Technical Report 11–009*. Edinburgh, Scotland: School of Mathematics, University of Edinburgh.

Cartis, C., Gould, N. & Toint, P. L. (2017) Evaluation complexity bounds for smooth constrained nonlinear optimisation using scaled kkt conditions, high-order models and the criticality measure χ. arXiv preprint arXiv:1705.04895 .

Conn, A. R., Gould, N. I. M. & Toint, P. L. (2000) *Trust-Region Methods*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).

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

Curtis, F. E., Robinson, D. P. & Samadi, M. (2018) An inexact regularized Newton framework with a worst-case iteration complexity of $\mathcal{O}(\varepsilon^{-3/2})$ for nonconvex optimization. arXiv 1708.00475.

Dolan, E. D. & Moré, J. J. (2002) Benchmarking optimization software with performance profiles. *Math. Program.*, **91**, 201–213.

Dussault, J.-P. (2018) ARCq: a new adaptive regularization by cubics. *Optim. Methods Softw.* **33**, 322–335. doi:10.1080/10556788.2017.1322080.

Dussault, J.-P. & Orban, D. (2017) Scalable adaptive cubic regularization methods. *Technical Report G-2015–109*. GERAD. Montreal, Canada: HEC Montreal, University of Montreal.

Gould, N. I. M., Orban, D. & Toint, P. L. (2003) Galahad, a library of thread-safe fortran 90 packages for large-scale nonlinear optimization. *ACM Trans. Math. Softw.*, **29**, 353–372.

Gould, N. I. M., Orban, D. & Toint, P. L. (2013) CUTEst: a constrained and unconstrained testing environment with safe threads. *Technical Report RAL-TR-2013–005*. Chilton, Oxfordshirt, England: Rutherford Appleton Laboratory.

Gould, N. I. M., Porcelli, M. & Toint, P. L. (2012) Updating the regularization parameter in the adaptive cubic regularization algorithm. *Comput. Optim. Appl.*, **53**, 1–22.

Gould, N. I. M., Robinson, D. P. & Thorne, H. S. (2010) On solving trust-region and other regularised subproblems in optimization. *Math. Program. Comput.*, **2**, 21–57.

Griewank, A. (1981) The modification of Newton's method for unconstrained optimization by bounding cubic terms. *Technical Report NA/12*. Department of Applied Mathematics and Theoretical Physics. Cambridge, England: University of Cambridge.

Griva, I., Nash, S. G. & Sofer, A. (2008) *Linear and Nonlinear Optimization, 2nd edn*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).

Nesterov, Y. (2004) *Introductory Lectures on Convex Optimization*, vol. 87. New York, NY: Springer Science & Business Media.

Nesterov, Y. & Polyak, B. T. (2006) Cubic regularization of Newton's method and its global performance. *Math. Program.*, **108**, 117–205.

Nocedal, J. & Wright, S. J. (2006) *Numerical Optimization*, 2nd edn. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer.

Ruszczynski, A. (2006) *Nonlinear Optimization*. Princeton, NJ: Princeton University Press.

Appendix A. Subproblem solution properties

In this appendix we explore properties of any first-order stationary solution of problem $\mathscr{P}_k(\sigma_k^{\mathrm{L}}, \sigma_k^{\mathrm{U}})$ defined as (2.1). Let us define a Lagrangian function for (2.1) as

$$\begin{aligned}\mathscr{L}\left(s, \lambda, \beta^{\mathrm{L}}, \beta^{\mathrm{U}}, \beta^{\mathrm{N}}\right) = f_k + g_k^T s + \tfrac{1}{2} s^T (H_k + \lambda I) s \\ - \tfrac{\beta^{\mathrm{L}}}{2}\left(\lambda^2 - \left(\sigma_k^{\mathrm{L}}\right)^2 \|s\|^2\right) + \tfrac{\beta^{\mathrm{U}}}{2}\left(\lambda^2 - \left(\sigma_k^{\mathrm{U}}\right)^2 \|s\|^2\right) - \beta^{\mathrm{N}}\lambda,\end{aligned}$$

where $(\beta^{\mathrm{L}}, \beta^{\mathrm{U}}) \in \mathbb{R}_+ \times \mathbb{R}_+$ are the dual variables associated with the left-hand and right-hand constraints on λ, respectively, and $\beta^{\mathrm{N}} \in \mathbb{R}_+$ is the dual variable associated with the non-negativity constraint on λ. The tuple $(s_k, \lambda_k, \beta_k^{\mathrm{L}}, \beta_k^{\mathrm{U}}, \beta_k^{\mathrm{N}})$ is a first-order primal-dual stationary solution of $\mathscr{P}_k\left(\sigma_k^{\mathrm{L}}, \sigma_k^{\mathrm{U}}\right)$ if it satisfies the following conditions:

$$g_k + (H_k + \lambda_k I) s_k + \beta_k^{\mathrm{L}}\left(\sigma_k^{\mathrm{L}}\right)^2 s_k - \beta_k^{\mathrm{U}}\left(\sigma_k^{\mathrm{U}}\right)^2 s_k = 0, \tag{A.1a}$$

$$\tfrac{1}{2}\|s_k\|^2 - \lambda_k\left(\beta_k^{\mathrm{L}} - \beta_k^{\mathrm{U}}\right) - \beta_k^{\mathrm{N}} = 0, \tag{A.1b}$$

$$0 \leqslant \beta_k^{\mathrm{L}} \perp \left(\lambda_k^2 - \left(\sigma_k^{\mathrm{L}}\right)^2 \|s_k\|^2\right) \geqslant 0, \tag{A.1c}$$

$$0 \leqslant \beta_k^{\mathrm{U}} \perp \left(\lambda_k^2 - \left(\sigma_k^{\mathrm{U}}\right)^2 \|s_k\|^2\right) \leqslant 0 \text{ and} \tag{A.1d}$$

$$0 \leqslant \beta_k^{\mathrm{N}} \perp \lambda_k \geqslant 0. \tag{A.1e}$$

We make the following assumption throughout this appendix.

Assumption A1 The vector g_k is nonzero.

Under this assumption the following lemma is a simple consequence of (A.1a).

Lemma A2 Any solution of (2.1) has $s_k \neq 0$.

We now establish conditions that must hold depending on the value of $\sigma_k^{\mathrm{L}} \in \mathbb{R}_+$.

Lemma A3 The following hold true for any solution of (A.1).

(i) If $\sigma_k^{\mathrm{L}} > 0$ then $\lambda_k > 0$, $\beta_k^{\mathrm{N}} = 0$, $\beta_k^{\mathrm{L}} > 0$ and $\lambda_k = \sigma_k^{\mathrm{L}}\|s_k\|$.

(ii) If $\sigma_k^{\mathrm{L}} = 0$ then $\lambda_k = 0$.

Proof. Consider part (i). For the sake of deriving a contradiction suppose $\sigma_k^{\mathrm{L}} > 0$ and $\lambda_k = 0$. These, along with Lemma A2, imply that $0 = \lambda_k^2 < (\sigma_k^{\mathrm{L}})^2 \|s_k\|^2$, which contradicts (A.1c). Hence, $\lambda_k > 0$, as claimed. Then it follows from (A.1e) that $\beta_k^{\mathrm{N}} = 0$, as claimed. Next, observe that from (A.1b), Lemma A2, $\beta_k^{\mathrm{N}} = 0$, $\lambda_k > 0$ and $(\beta_k^{\mathrm{L}}, \beta_k^{\mathrm{U}}) \geqslant 0$, it follows that $\beta_k^{\mathrm{L}} > 0$, as claimed. This, along with (A.1c), implies that $\lambda_k^2 = (\sigma_k^{\mathrm{L}})^2 \|s_k\|^2$. This implies that $\lambda_k = \pm(\sigma_k^{\mathrm{L}})\|s_k\|$, which combined with $\lambda_k \in \mathbb{R}^+$ means that $\lambda_k = \sigma_k^{\mathrm{L}} \|s_k\|$, as claimed.

Now consider part (ii). For the sake of deriving a contradiction suppose that $\sigma_k^{\mathrm{L}} = 0$ and $\lambda_k > 0$. Then it follows from (A.1e) that $\beta_k^{\mathrm{N}} = 0$. Moreover, combining $\sigma_k^{\mathrm{L}} = 0$ and $\lambda_k > 0$, it follows from (A.1c) that $\beta_k^{\mathrm{L}} = 0$. It now follows from $\beta_k^{\mathrm{L}} = 0$, $\beta_k^{\mathrm{N}} = 0$ and (A.1b) that

$$\tfrac{1}{2}\|s_k\|^2 = -\lambda_k \beta_k^{\mathrm{U}} \leqslant 0, \tag{A.2}$$

where the inequality follows from $\lambda_k > 0$ and $\beta_k^{\mathrm{U}} \geqslant 0$. This contradicts Lemma A2. □

Our main result is the following. In part (i) with $\sigma_k^{\mathrm{L}} > 0$ we show that solving (2.1) is equivalent to solving what may be referred to as an ARC subproblem (Cartis *et al.*, 2011a). In part (ii) with $\sigma_k^{\mathrm{L}} = 0$ we show that it is equivalent to minimizing a quadratic if a minimizer exists.

THEOREM A4 The following hold true.

(i) Suppose $\sigma_k^{\mathrm{L}} > 0$. Then (2.1) has a solution (s_k, λ_k), which can be obtained as

$$s_k \in \arg\min_{s \in \mathbb{R}^n} \left(f_k + g_k^T s + \tfrac{1}{2} s^T H_k s + \tfrac{1}{2}\sigma_k^{\mathrm{L}} \|s\|^3 \right), \tag{A.3}$$

then setting $\lambda_k = \sigma_k^{\mathrm{L}} \|s_k\| > 0$.

(ii) If $\sigma_k^{\mathrm{L}} = 0$ then a solution of problem (2.1) exists if and only if $H_k \succeq 0$ and $g_k^T u = 0$ for all $u \in \mathrm{Null}(H_k)$. In such cases computing a solution (s_k, λ_k) of problem (2.1) is equivalent to computing a solution s_k of problem (4.1) and setting $\lambda_k = 0$.

Proof. Consider part (i). Since $\sigma_k^{\mathrm{L}} > 0$, it follows from Lemma A3 that problem (2.1) is equivalent to

$$\begin{aligned} \min_{(s,\lambda) \in \mathbb{R}^n \times \mathbb{R}_+} \quad & f_k + g_k^T s + \tfrac{1}{2} s^T (H_k + \lambda I) s \\ \text{s.t.} \quad & \sigma_k^{\mathrm{L}} \|s\| = \lambda, \end{aligned} \tag{A.4}$$

where, by Lemma A2, it follows that the solution has $\lambda_k > 0$, as desired. Substituting the constraint of (A.4) into the objective of (A.4) one finds that solving it is equivalent to solving (A.3) for s_k, then setting $\lambda_k = \sigma_k^{\mathrm{L}} \|s_k\|$, as claimed. Since $\sigma_k^{\mathrm{L}} > 0$ a minimizer of problem (A.3) exists because it involves the minimization of a coercive function.

Now consider part (ii). Since $\sigma_k^{\mathrm{L}} = 0$, it follows from Lemma A3 that $\lambda_k = 0$, meaning that problem (2.1) is equivalent to (4.1). This problem has a solution if and only if the objective is bounded below, which is the case if and only if $H_k \succeq 0$ and $g_k^T u = 0$ for all $u \in \mathrm{Null}(H_k)$. □

Appendix B. Subproblem solution properties over subspaces

In this appendix we explore properties of any first-order stationary solution (when one exists) of problem $\mathscr{P}_k(\sigma_k^{\mathrm{L}}, \sigma_k^{\mathrm{U}})$ defined as (2.1), when the search space for s is restricted to a subspace of $\mathbb{R}^n$. Specifically,

for some m-dimensional subspace $\mathscr{L}_k \subseteq \mathbb{R}^n$ consider the problem

$$\begin{aligned} \min_{(s,\lambda)\in\mathscr{L}_k\times\mathbb{R}_+} \quad & f_k + g_k^T s + \tfrac{1}{2} s^T (H_k + \lambda I) s \\ \text{s.t.} \quad & \left(\sigma_k^{\mathrm{L}}\right)^2 \|s\|^2 \leqslant \lambda^2 \leqslant \left(\sigma_k^{\mathrm{U}}\right)^2 \|s\|^2. \end{aligned} \tag{B.1}$$

Given an orthogonal basis R_k for $\mathscr{L}_k$, a solution of (B.1) can be obtained from that of

$$\begin{aligned} \min_{(v,\lambda)\in\mathbb{R}^m\times\mathbb{R}_+} \quad & f_k + g_k^T R_k v + \tfrac{1}{2} (R_k v)^T (H_k + \lambda I) R_k v \\ \text{s.t.} \quad & \left(\sigma_k^{\mathrm{L}}\right)^2 \|v\|^2 \leqslant \lambda^2 \leqslant \left(\sigma_k^{\mathrm{U}}\right)^2 \|v\|^2. \end{aligned} \tag{B.2}$$

Specifically, if $(v_k, \lambda_k, \beta_k^{\mathrm{L}}, \beta_k^{\mathrm{U}}, \beta_k^{\mathrm{N}})$ is a first-order primal-dual solution of problem (B.2), then the tuple $(s_k, \lambda_k, \beta_k^{\mathrm{L}}, \beta_k^{\mathrm{U}}, \beta_k^{\mathrm{N}})$ with $s_k = R_k v_k$ is such a solution of problem (B.1).

In Appendix A we proved properties of a solution (if one exists) of a problem of the form (B.2). Let us now translate the results of that appendix to the present setting, for which we require the following assumption on the reduced gradient $R_k^T g_k$.

Assumption B1 The vector $R_k^T g_k$ is nonzero.

Lemma B2 Any solution of (B.2) has $v_k \neq 0$.

Lemma B3 The following hold for any first-order primal-dual solution of (B.1).

(i) If $\sigma_k^{\mathrm{L}} > 0$ then $\lambda_k > 0$, $\beta_k^{\mathrm{N}} = 0$, $\beta_k^{\mathrm{L}} > 0$ and $\lambda_k = \sigma_k^{\mathrm{L}} \|v_k\|$.

(ii) If $\sigma_k^{\mathrm{L}} = 0$ then $\lambda_k = 0$.

Theorem B4 The following hold true.

(i) Suppose $\sigma_k^{\mathrm{L}} > 0$. Then (B.2) has a solution (v_k, λ_k), which can be obtained as

$$v_k \in \arg\min_{v\in\mathbb{R}^m} \left(f_k + g_k^T R_k v + \tfrac{1}{2} v^T R_k^T H_k R_k v + \tfrac{1}{2} \sigma_k^{\mathrm{L}} \|v\|^3 \right), \tag{B.3}$$

then setting $\lambda_k = \sigma_k^{\mathrm{L}} \|v_k\| > 0$.

(ii) If $\sigma_k^{\mathrm{L}} = 0$ then a solution of (B.2) exists if and only if $R_k^T H_k R_k \succeq 0$ and $g_k^T R_k u = 0$ for all $u \in \mathrm{Null}(R_k^T H_k R_k)$. In such cases computing a solution (v_k, λ_k) of problem (B.2) is equivalent to computing a solution v_k of

$$\min_{v\in\mathbb{R}^m} \; f_k + g_k^T R_k v + \tfrac{1}{2} v^T R_k^T H_k R_k v \tag{B.4}$$

and setting $\lambda_k = 0$.

Considering problem (B.3) we obtain the following result from Cartis *et al.* (2011a, Lemma 3.2).

Lemma B5 If $\sigma_k^{\mathrm{L}} > 0$ then v_k from (B.3) satisfies

$$g_k^T R_k v_k + v_k^T R_k^T H_k R_k v_k + \tfrac{3}{2} \sigma_k^{\mathrm{L}} \|v_k\|^3 = 0 \tag{B.5a}$$

$$v_k^T R_k^T H_k R_k v_k + \tfrac{3}{2} \sigma_k^{\mathrm{L}} \|v_k\|^3 \geqslant 0 \tag{B.5b}$$

$$R_k^T H_k R_k + \tfrac{3}{2} \sigma_k^{\mathrm{L}} \|v_k\| I \succeq 0. \tag{B.5c}$$

We now show that, under certain reasonable assumptions, solutions of the primal-dual reduced-space subproblem (B.1) satisfy the conditions required by Assumptions 2.1 and 3.19.

THEOREM B6 The following hold true.

(a) Any solution of problem (B.1) satisfies (2.3b).

(b) Any solution of problem (B.1) satisfies (2.3a) provided $g_k \in \mathscr{L}_k$.

(c) Any solution of problem (B.1) satisfies (2.3c) provided $\mathscr{L}_k = \mathbb{R}^n$.

(d) Any solution of problem (B.1) satisfies (3.13) for any $\kappa_4 \geqslant \frac{3}{2} \sup_{k \in \mathbb{N}_+} \{\sigma_k^{\mathrm{L}}\}$.

Proof. Any first-order primal-dual solution $(s_k, \lambda_k, \beta_k^{\mathrm{L}}, \beta_k^{\mathrm{U}}, \beta_k^{\mathrm{N}})$ of problem (B.1) corresponds to such a solution $(v_k, \lambda_k, \beta_k^{\mathrm{L}}, \beta_k^{\mathrm{U}}, \beta_k^{\mathrm{N}})$ of problem (B.2) where $s_k = R_k v_k$. Hence, throughout this proof, for any solution vector s_k for problem (B.1), we may let $s_k = R_k v_k$ where v_k satisfies the properties in Lemmas B2–B3 and B5.

First, suppose $\sigma_k^{\mathrm{L}} > 0$, which by Theorem B4(i) implies that problem (B.1) has a solution. Then it follows from (B.5a), $s_k = R_k v_k$ and Lemma B3(i) that

$$0 = g_k^T s_k + s_k^T H_k s_k + \tfrac{3}{2}\sigma_k^{\mathrm{L}} \|s_k\|^3 = g_k^T s_k + s_k^T H_k s_k + \tfrac{3}{2}\lambda_k \|s_k\|^2,$$

which means that

$$s_k^T (g_k + (H_k + \lambda_k I) s_k) = -\tfrac{1}{2}\lambda_k \|s_k\|^2. \tag{B.6}$$

Meanwhile, from (B.5b), $s_k = R_k v_k$ and Lemma B3(i), it follows that

$$0 \leqslant s_k^T H_k s_k + \tfrac{3}{2}\sigma_k^{\mathrm{L}} \|s_k\|^3 = s_k^T H_k s_k + \tfrac{3}{2}\lambda_k \|s_k\|^2 = s_k^T (H_k + \lambda_k I) s_k + \tfrac{1}{2}\lambda_k \|s_k\|^2,$$

which means that

$$-\tfrac{1}{4}\lambda_k \|s_k\|^2 \leqslant \tfrac{1}{2} s_k^T (H_k + \lambda_k I) s_k. \tag{B.7}$$

It follows from (B.6), (B.7), $\lambda_k > 0$ (by Lemma B3(i)) and $(\kappa_1, \kappa_2) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$ that

$$\begin{aligned} s_k^T (g_k + (H_k + \lambda_k I) s_k) = -\tfrac{1}{2}\lambda_k \|s_k\|^2 &\leqslant \min\left\{ \tfrac{1}{2}\kappa_1 \|s_k\|^2, \tfrac{1}{2} s_k^T (H_k + \lambda_k I) s_k - \tfrac{1}{4}\lambda_k \|s_k\|^2 \right\} \\ &\leqslant \min\left\{ \tfrac{1}{2}\kappa_1 \|s_k\|^2, \tfrac{1}{2} s_k^T (H_k + \lambda_k I) s_k + \tfrac{1}{2}\kappa_2 \|s\|^3 \right\}, \end{aligned}$$

which implies (2.3b). This establishes that part (a) is true. Now consider part (b). From Theorem B4, Curtis *et al.* (2011a, Lemma 2.1) and $s_k = R_k v_k$, it follows that

$$f_k - q_k(s_k) - \tfrac{1}{2}\sigma_k^{\mathrm{L}} \|s_k\|^3 \geqslant \frac{\|R_k^T g_k\|}{6\sqrt{2}} \min\left\{ \frac{\|R_k^T g_k\|}{1 + \|R_k^T H_k R_k\|}, \frac{1}{\sqrt{6}} \sqrt{\frac{\|R_k^T g_k\|}{\sigma_k^{\mathrm{L}}}} \right\}.$$

Since, under assumption, $g_k \in \mathscr{L}_k$ so that $g_k = R_k y$ for some $y \in \mathbb{R}^m$, it follows that

$$\|R_k^T g_k\| = \|R_k^T R_k y\| = \|y\| = \|R_k y\| = \|g_k\|.$$

Combining this with $\|R_k^T H_k R_k\| \leqslant \|H_k\|$ and the previous displayed inequality shows

$$f_k - q_k(s_k) - \tfrac{1}{2}\sigma_k^{\mathrm{L}} \|s_k\|^3 \geqslant \frac{\|g_k\|}{6\sqrt{2}} \min\left\{ \frac{\|g_k\|}{1 + \|H_k\|}, \frac{1}{\sqrt{6}} \sqrt{\frac{\|g_k\|}{\sigma_k^{\mathrm{L}}}} \right\}.$$

This may now be combined with Theorem B4 (specifically $\lambda_k = \sigma_k^{\mathrm{L}}\|s_k\| > 0$) to obtain

$$f_k - q_k(s_k) \geqslant f_k - q_k(s_k) - \tfrac{1}{2}\sigma_k^{\mathrm{L}}\|s_k\|^3 \geqslant \frac{\|g_k\|}{6\sqrt{2}} \min\left\{\frac{\|g_k\|}{1+\|H_k\|}, \frac{1}{\sqrt{6}}\sqrt{\frac{\|g_k\|\|s_k\|}{\lambda_k}}\right\},$$

which means that (s_k, λ_k) satisfies (2.3a), proving part (b). Now consider part (c). It follows from Theorem A4(i) and the optimality conditions for problem (A.3) that

$$0 = g_k + H_k s_k + \tfrac{3}{2}\sigma_k^{\mathrm{L}}\|s_k\|s_k = g_k + H_k s_k + \tfrac{3}{2}\lambda_k s_k = g_k + (H_k + \lambda_k I)s_k + \tfrac{1}{2}\lambda_k s_k.$$

This and the fact that $\kappa_3 > 0$ imply that

$$\|g_k + (H_k + \lambda_k I)s_k\| = \tfrac{1}{2}\lambda_k\|s_k\| \leqslant \lambda_k\|s_k\| + \kappa_3\|s_k\|^2,$$

which completes the proof of part (c). Finally, consider part (d). From (B.5c) the fact that $\|s_k\| = \|v_k\|$ and $\kappa_4 \geqslant \frac{3}{2}\sup_{k\in\mathbb{N}_+}\{\sigma_k^{\mathrm{L}}\}$, it follows that

$$\xi\left(R_k^T H_k R_k\right) \geqslant -\tfrac{3}{2}\sigma_k^{\mathrm{L}}\|s_k\| \geqslant -\kappa_4\|s_k\|,$$

as desired to prove part (d).

Now suppose that $\sigma_k^{\mathrm{L}} = 0$. From Theorem B4(ii) a solution of problem (B.1) exists if and only if $R_k^T H_k R_k \succeq 0$ and $g_k^T R_k u = 0$ for all $u \in \mathrm{Null}(R_k^T H_k R_k)$. If this is not the case then there is nothing left to prove; hence, let us assume that these conditions hold. From these conditions, Theorem B4(ii), the optimality conditions of problem (B.4), the fact that $\lambda_k = 0$ and $s_k = R_k v_k$, it follows that

$$g_k^T s_k + s_k^T H_k s_k = 0 \ \text{ and } \ s_k^T H_k s_k \geqslant 0.$$

This shows that (2.3b) holds, proving part (a) for this case. Next, since v_k is given by the solution of problem (B.4), it follows that the reduction in the objective yielded by v_k is at least as large as the reduction obtained by minimizing the objective over the span of $-R_k^T g_k$. Hence, from standard theory on Cauchy decrease (see Conn *et al.*, 2000 or Nocedal & Wright, 2006), one can conclude that

$$f_k - q_k(s_k) \geqslant \frac{\left\|R_k^T g_k\right\|}{2} \min\left\{\frac{\left\|R_k^T g_k\right\|}{1+\left\|R_k^T H_k R_k\right\|}, \|s_k\|\right\}.$$

Thus, using the arguments in the previous paragraph under the assumption that $g_k \in \mathscr{L}_k$ one is led to the conclusion that (2.3a) holds, which proves part (b) for this case. Next, when $\mathscr{L}_k = \mathbb{R}^n$ the optimality conditions for problem (B.4) imply that $g_k + H_k s_k = 0$, which, since $\lambda_k = 0$, implies that (2.3c) holds, proving part (c). Finally, since $R_k^T H_k R_k \succeq 0$, it follows that (3.13) holds, proving part (d). □

Appendix C. Detailed numerical results

Further details of the results of our numerical experiments are shown in Table C1. (The table only includes problems with at least 1000 variables. Results for all problems, including those with less than 1000 variables, can be found in the technical report version of this paper; see Curtis *et al.*, 2018.) In the table `#Var` indicates the number of variables, `#Iter` indicates the number of iterations required (with `%Newton` indicating the percentage that were inexact Newton steps with $\lambda_k = 0$), `#Acc` indicates the number of accepted steps (again with `%Newton` indicating the percentage that were inexact Newton steps), `#Hv-prod` indicates the number of Hessian-vector products required and `#T-fact` indicates the number of tridiagonal matrix factorizations required.

TABLE C1 *Numerical results for* `iARC` *and* `iR_Newton`

Prob	#Var	Alg	#Iter	(%Newton)	#Acc	(%Newton)	#Hv-prod	#T-fact
ARWHEAD	5000	iARC	4		4		10	6
		iR_Newton	4	(%100)	4	(%100)	5	0
BDQRTIC	5000	iARC	9		9		34	33
		iR_Newton	9	(%100)	9	(%100)	17	0
BOX	10000	iARC	3		3		14	12
		iR_Newton	4	(%75)	3	(%67)	9	4
BOXPOWER	20000	iARC	3		3		10	9
		iR_Newton	7	(%100)	7	(%100)	13	0
BROYDN7D	5000	iARC	472		279		7202	12598
		iR_Newton	812	(%29)	346	(%2)	5033	10022
BRYBND	5000	iARC	19		10		240	367
		iR_Newton	17	(%53)	11	(%82)	206	81
CHAINWOO	4000	iARC	81		57		798	942
		iR_Newton	70	(%81)	59	(%88)	409	185
COSINE	10000	iARC	12		7		108	140
		iR_Newton	11	(%55)	7	(%71)	45	29
CRAGGLVY	5000	iARC	30		30		228	208
		iR_Newton	31	(%100)	31	(%100)	108	0
CURLY10	10000	iARC	—		—		—	—
		iR_Newton	328	(%95)	318	(%97)	271881	19957
CURLY30	10000	iARC	—		—		—	—
		iR_Newton	87	(%83)	77	(%91)	125639	630
DIXMAANA	3000	iARC	6		6		14	8
		iR_Newton	6	(%100)	6	(%100)	7	0
DIXMAANB	3000	iARC	7		7		16	9
		iR_Newton	7	(%100)	7	(%100)	8	0
DIXMAANC	3000	iARC	8		8		18	9
		iR_Newton	8	(%100)	8	(%100)	9	0
DIXMAAND	3000	iARC	9		9		20	10
		iR_Newton	9	(%100)	9	(%100)	10	0
DIXMAANE	3000	iARC	59		59		670	622
		iR_Newton	60	(%100)	60	(%100)	331	0

(Continued).

TABLE C1 *Continued*

Prob	#Var	Alg	#Iter (%Newton)		#Acc (%Newton)		#Hv-prod	#T-fact
DIXMAANF	3000	iARC	38		37		510	487
		iR_Newton	37	(%100)	37	(%100)	249	0
DIXMAANG	3000	iARC	39		39		532	514
		iR_Newton	40	(%100)	40	(%100)	288	0
DIXMAANH	3000	iARC	41		41		448	421
		iR_Newton	41	(%100)	41	(%100)	224	0
DIXMAANI	3000	iARC	193		193		3456	3464
		iR_Newton	249	(%100)	249	(%100)	2814	0
DIXMAANJ	3000	iARC	34		34		324	296
		iR_Newton	34	(%100)	34	(%100)	162	0
DIXMAANK	3000	iARC	30		30		248	225
		iR_Newton	30	(%100)	30	(%100)	124	0
DIXMAANL	3000	iARC	29		29		180	148
		iR_Newton	29	(%100)	29	(%100)	90	0
DIXMAANM	3000	iARC	375		375		10902	11542
		iR_Newton	398	(%100)	398	(%100)	6126	0
DIXMAANN	3000	iARC	82		82		1368	1358
		iR_Newton	87	(%100)	87	(%100)	789	0
DIXMAANO	3000	iARC	63		63		908	893
		iR_Newton	59	(%100)	59	(%100)	371	0
DIXMAANP	3000	iARC	51		51		476	432
		iR_Newton	51	(%100)	51	(%100)	238	0
DIXON3DQ	10000	iARC	2257		2256		143968	164858
		iR_Newton	2476	(%100)	2476	(%100)	81042	0
DQDRTIC	5000	iARC	6		6		34	32
		iR_Newton	4	(%100)	4	(%100)	10	0
DQRTIC	5000	iARC	15		15		30	15
		iR_Newton	11	(%100)	11	(%100)	11	0
EDENSCH	2000	iARC	15		15		44	25
		iR_Newton	15	(%100)	15	(%100)	22	0
EG2	1000	iARC	3		3		6	3
		iR_Newton	3	(%100)	3	(%100)	3	0

(Continued).

TABLE C1 *Continued*

Prob	#Var	Alg	#Iter	(%Newton)	#Acc	(%Newton)	#Hv-prod	#T-fact
EIGENALS	2550	iARC	179		134		15548	20388
		iR_Newton	173	(%84)	150	(%89)	7871	1999
ENGVAL1	5000	iARC	9		9		64	54
		iR_Newton	9	(%100)	9	(%100)	32	0
EXTROSNB	1000	iARC	179		107		2978	3586
		iR_Newton	185	(%62)	114	(%64)	1576	1553
FLETBV3M	5000	iARC	41		34		86	43
		iR_Newton	56	(%43)	32	(%41)	65	32
FLETCHCR	1000	iARC	2437		1450		66056	90373
		iR_Newton	2187	(%66)	1438	(%69)	29012	23819
FMINSRF2	5625	iARC	875		567		6528	7378
		iR_Newton	905	(%50)	448	(%40)	2666	1989
FREUROTH	5000	iARC	17		11		102	120
		iR_Newton	18	(%39)	10	(%60)	51	35
GENHUMPS	5000	iARC	14931		11710		477824	1724919
		iR_Newton	3567	(%2)	2077	(%1)	25952	85744
JIMACK	3549	iARC	54		54		36564	45769
		iR_Newton	52	(%100)	52	(%100)	16267	0
LIARWHD	5000	iARC	12		12		46	45
		iR_Newton	11	(%100)	11	(%100)	21	0
MODBEALE	20000	iARC	—		—		—	—
		iR_Newton	3317	(%99)	3304	(%100)	65293	351
MOREBV	5000	iARC	4		4		1102	2064
		iR_Newton	1	(%100)	1	(%100)	401	0
MSQRTALS	1024	iARC	39		33		9830	12602
		iR_Newton	44	(%73)	36	(%83)	4743	149
MSQRTBLS	1024	iARC	32		26		5822	7090
		iR_Newton	39	(%69)	31	(%81)	3131	156
NCB20	5010	iARC	106		70		2888	4664
		iR_Newton	65	(%32)	43	(%42)	688	614
NCB20B	5000	iARC	29		18		4286	9958
		iR_Newton	38	(%47)	19	(%42)	3297	8386

(Continued).

TABLE C1 *Continued*

Prob	#Var	Alg	#Iter (%Newton)		#Acc (%Newton)		#Hv-prod	#T-fact
NONCVXU2	5000	iARC	10302		10302		20604	10302
		iR_Newton	11094	(%100)	11094	(%100)	11094	0
NONCVXUN	5000	iARC	23771		23771		47542	23771
		iR_Newton	20913	(%100)	20913	(%100)	20913	0
NONDIA	5000	iARC	2		2		4	2
		iR_Newton	2	(%100)	2	(%100)	2	0
NONDQUAR	5000	iARC	45		37		156	126
		iR_Newton	38	(%95)	36	(%97)	70	2
OSCIGRAD	100000	iARC	13		10		190	220
		iR_Newton	15	(%40)	9	(%56)	92	61
PENALTY1	1000	iARC	14		14		28	14
		iR_Newton	12	(%100)	12	(%100)	12	0
POWELLSG	5000	iARC	17		17		98	91
		iR_Newton	17	(%100)	17	(%100)	49	0
QUARTC	5000	iARC	15		15		30	15
		iR_Newton	11	(%100)	11	(%100)	11	0
SCHMVETT	5000	iARC	5		5		142	166
		iR_Newton	6	(%100)	6	(%100)	89	0
SINQUAD	5000	iARC	16		11		64	63
		iR_Newton	15	(%33)	9	(%44)	32	26
SPARSINE	5000	iARC	153		143		15246	18485
		iR_Newton	188	(%88)	174	(%94)	10745	183
SPARSQUR	10000	iARC	15		15		64	49
		iR_Newton	15	(%100)	15	(%100)	32	0
SPMSRTLS	4999	iARC	17		15		582	761
		iR_Newton	17	(%76)	15	(%87)	275	4
SROSENBR	5000	iARC	9		7		36	35
		iR_Newton	10	(%70)	7	(%86)	20	12
SSBRYBND	5000	iARC	75		45		77075	177454
		iR_Newton	39	(%38)	23	(%52)	22010	11269
TESTQUAD	5000	iARC	162		162		16908	19812
		iR_Newton	163	(%100)	163	(%100)	8271	0

(Continued).

TABLE C1 *Continued*

Prob	#Var	Alg	#Iter (%Newton)		#Acc (%Newton)		#Hv-prod	#T-fact
TOINTGSS	5000	iARC	4		4		14	10
		iR_Newton	3	(%100)	3	(%100)	7	0
TQUARTIC	5000	iARC	11		11		44	50
		iR_Newton	1	(%100)	1	(%100)	2	0
TRIDIA	5000	iARC	16		16		2128	2630
		iR_Newton	17	(%100)	17	(%100)	1310	0
WOODS	4000	iARC	15		15		40	26
		iR_Newton	172	(%87)	157	(%92)	404	144