Journal of Complexity 28 (2012) 93–108

Contents lists available at SciVerse ScienceDirect

Journal of Complexity

journal homepage: www.elsevier.com/locate/jco

Complexity bounds for second-order optimality in unconstrained optimization

C. Cartis [a,*], N.I.M. Gould [b], Ph.L. Toint [c]

[a] School of Mathematics, University of Edinburgh, The King's Buildings, Edinburgh, EH9 3JZ, Scotland, UK

[b] Computational Science and Engineering Department, Rutherford Appleton Laboratory, Chilton, Oxfordshire, OX11 0QX, England, UK

[c] Namur Center for Complex Systems (naXys), FUNDP—University of Namur, 61, rue de Bruxelles, B-5000 Namur, Belgium

A R T I C L E I N F O

Article history:
Received 17 December 2010
Accepted 26 May 2011
Available online 14 June 2011

Keywords:
Evaluation complexity
Worst-case analysis
Nonconvex optimization
Second-order optimality conditions

A B S T R A C T

This paper examines worst-case evaluation bounds for finding weak minimizers in unconstrained optimization. For the cubic regularization algorithm, Nesterov and Polyak (2006) [15] and Cartis et al. (2010) [3] show that at most $O(\epsilon^{-3})$ iterations may have to be performed for finding an iterate which is within ϵ of satisfying second-order optimality conditions. We first show that this bound can be derived for a version of the algorithm, which only uses one-dimensional global optimization of the cubic model and that it is sharp. We next consider the standard trust-region method and show that a bound of the same type may also be derived for this method, and that it is also sharp in some cases. We conclude by showing that a comparison of the bounds on the worst-case behaviour of the cubic regularization and trust-region algorithms favours the first of these methods.

1. Introduction

We consider algorithms for the solution of the unconstrained (possibly nonconvex) optimization problem

$$\min_x f(x) \tag{1.1}$$

where we assume that $f : \mathbb{R}^n \to \mathbb{R}$ is smooth (in a sense to be specified later) and bounded below. All methods for the solution of (1.1) are iterative and, starting from some initial guess x_0, generate a sequence $\{x_k\}$ of iterates approximating a critical point of f. Many such algorithms exist, and they are

* Corresponding author.
E-mail addresses: coralia.cartis@ed.ac.uk (C. Cartis), nick.gould@sftc.ac.uk (N.I.M. Gould), philippe.toint@fundp.ac.be (Ph.L. Toint).

0885-064X/$ – see front matter
doi:10.1016/j.jco.2011.06.001

often classified according to their requirements in terms of computing derivatives of the objective function. In this paper, we focus on second-order methods, that is, methods which evaluate the objective function $f(x)$, its gradient $g(x)$ and its Hessian $H(x)$ (or an approximation thereof) at every iteration. The advantage of these methods is that they can be expected to converge to solutions x_* satisfying the second-order optimality conditions

$$\nabla_x f(x_*) = 0, \quad \text{and} \quad \lambda_{\min}(H(x_*)) \geq 0 \tag{1.2}$$

where $\lambda_{\min}(A)$ is the smallest eigenvalue of the symmetric matrix A, rather than only satisfying first-order optimality (i.e., the first of these relations). In practice, however, a second-order algorithm is typically terminated as soon as an iterate x_k is found which is within ϵ of satisfying (1.2), that is, such that

$$\|\nabla_x f(x_k)\| \leq \epsilon_g \quad \text{and} \quad \lambda_{\min}(H(x_k)) \geq -\epsilon_H, \tag{1.3}$$

for some user-specified tolerances $\epsilon_g, \epsilon_H \in (0, 1)$, where $\|\cdot\|$ denotes the Euclidean norm. It is then of interest to bound the number of iterations which may be necessary to find an iterate satisfying (1.3) as a function of the thresholds ϵ_g and ϵ_H. It is the purpose of worst-case complexity analysis to derive such bounds. Many results are available in the literature for the case where the objective function f is convex (see, for instance, [13,14,12,1]). The convergence to approximate first-order points in the nonconvex case has also been investigated for some time (see [16–18,15,10,3–5,8], or [19]).

Of particular interest here is the Adaptive Regularization with Cubics (ARC) algorithm independently proposed by Griewank [11], Weiser et al. [20] and Nesterov and Polyak [15], whose worst-case complexity was shown in the last of these references to be of $O(\epsilon_g^{-3/2})$ iterations for finding an iterate x_k satisfying the approximate first-order optimality conditions (the first relation in (1.3) only) and of $O(\epsilon_H^{-3})$ iterations for finding an iterate x_k satisfying the whole of (1.3).[1] These results were extended by Cartis et al. [3] to an algorithm no longer requiring the computation of exact second-derivatives (but merely of a suitably accurate approximation), nor an (also possibly approximate) knowledge of the objective function's Hessian's Lipschitz constant. More importantly, these authors showed that the $O(\epsilon_g^{-3/2})$ complexity bound for convergence to first-order critical points can be achieved without requiring multi-dimensional global optimization of the cubic model (see [6]). However, such a global minimization on nested Krylov subspaces of increasing dimensions was still required to obtain the $O(\epsilon_H^{-3})$ convergence to second-order critical points.

The present paper focuses on worst-case complexity bounds for convergence to second-order critical points and shows that, as in the first-order case, multi-dimensional global minimization of the cubic model is unnecessary for obtaining the mentioned $O(\epsilon_H^{-3})$ bound for the ARC algorithm. This latter bound is also shown to be sharp. We also prove that a bound of the same type holds for the standard trust-region method. Moreover, we show that it is also sharp for a range of relative values of ϵ_g and ϵ_H. We finally compare the known bounds for the ARC and trust-region algorithms and show that the ARC algorithm is always as good or better from this point of view.

The ARC algorithm is recalled in Section 2 and the associated complexity bounds are derived without multi-dimensional global minimization. Section 3 then discusses an example showing that the bound on convergence of the ARC algorithm to approximate second-order critical points is sharp. A bound of this type is derived in Section 4 for the trust-region methods, its sharpness for suitable values of ϵ_g and ϵ_H is demonstrated, and the comparison with the ARC algorithm discussed. Conclusions and perspectives are finally presented in Section 5.

2. The ARC algorithm and its worst-case complexity

The Adaptive Regularization with Cubics (ARC) algorithm is based on the approximate minimization, at iteration k, of the (possibly nonconvex) cubic model

$$m_k(s) = \langle g_k, s\rangle + \frac{1}{2}\langle s, B_k s\rangle + \frac{1}{3}\sigma_k\|s\|^3, \tag{2.1}$$

[1] It appears that this latter result is the first worst-case complexity bound for convergence to approximate second-order critical points ever proved.