Journal of Computational Mathematics, Vol. 24, No. 6, 2006, 676–692.

SECOND-ORDER CONVERGENCE PROPERTIES OF TRUST-REGION METHODS USING INCOMPLETE CURVATURE INFORMATION, WITH AN APPLICATION TO MULTIGRID OPTIMIZATION*

Serge Gratton
(*CERFACS, av. G. Coriolis, Toulouse, France*)
Annick Sartenaer and Philippe L. Toint
(*Department of Mathematics, University of Namur, 61, rue de Bruxelles, B-5000 Namur, Belgium*)

Abstract

Convergence properties of trust-region methods for unconstrained nonconvex optimization is considered in the case where information on the objective function's local curvature is incomplete, in the sense that it may be restricted to a fixed set of "test directions" and may not be available at every iteration. It is shown that convergence to local "weak" minimizers can still be obtained under some additional but algorithmically realistic conditions. These theoretical results are then applied to recursive multigrid trust-region methods, which suggests a new class of algorithms with guaranteed second-order convergence properties.

Mathematics subject classification: 90C30, 65K05, 90C26, 90C06.
Key words: Nonlinear optimization, Convergence to local minimizers, Multilevel problems.

1. Introduction

It is highly desirable that iterative algorithms for solving nonconvex unconstrained optimization problems of the form

$$\min_{x \in \Re^n} f(x) \tag{1.1}$$

converge to solutions that are local minimizers of the objective function f, rather than mere first-order critical points, or, at least, that the objective function f becomes asymptotically convex. Trust-region methods (see Conn, Gould and Toint [2] for an extensive coverage) are well-known for being able to deliver these guarantees under assumptions that are not too restrictive in general. In particular, it may be proved that every limit point of the sequence of iterates satisfies the second-order necessary optimality conditions under the assumption that the smallest eigenvalue of the objective function's Hessian is estimated at each iteration. There exist circumstances, however, where this assumption appears either costly or irrealistic. We are in particular motivated by multigrid recursive trust-region methods of the type investigated in [3]: in these methods, gradient smoothing is achieved on a given grid by successive coordinate minimization, a procedure that only explores curvature along the vectors of the canonical basis. As a consequence, some negative curvature directions on the current grid may be undetected. Moreover, these smoothing iterations are intertwined with recursive iterations which only give information on coarser grids. As a result, information on negative curvature directions at a given iteration may either be incomplete or simply missing, causing the assumption required

* Received November 7, 2005.

for second-order convergence to fail. Another interesting example is that of algorithms in which the computation of the step may require the explicit determination of the smallest eigenvalue of the objective function's Hessian in order to take negative curvature into account (see, for instance, [5]). Because of cost, one might then wish to avoid the eigenvalue calculation at every iteration, which again jeopardizes the condition ensuring second-order convergence.

Our purpose is therefore to investigate what can be said about second-order convergence of trust-region methods when negative curvature information is incomplete or missing at some iterations. We indicate that a weaker form of second-order optimality may still hold at the cost of imposing a few additional assumptions that are algorithmically realistic. Section 2 introduces the necessary modifications of the basic trust-region algorithm, whose second-order convergence properties are then investigated in Section 3. Application to the recursive multigrid trust-region methods is then discussed in more detail in Section 4. Some conclusions and extensions are finally proposed in Section 5.

2. A Trust-region Algorithm with Incomplete Curvature Information

We consider the unconstrained optimization problem (1.1), where f is a twice-continuously differentiable objective function which maps $\Re^n$ into $\Re$ and is bounded below. We are interested in using a trust-region algorithm for solving (1.1). Methods of this type are iterative and, given an initial point x_0, produce a sequence $\{x_k\}$ of iterates (hopefully) converging to a local minimizer of the problem, i.e., to a point x_* where $g(x_*) \stackrel{\text{def}}{=} \nabla_x f(x_*) = 0$ (first-order convergence) and $\nabla_{xx} f(x_*)$ is positive semi-definite (second-order convergence). At each iterate x_k, classical trust-region methods build a model $m_k(x_k + s)$ of $f(x_k + s)$. This model is then assumed to be adequate in a "trust region" $\mathcal{B}_k$, defined as a sphere of radius $\Delta_k > 0$ centered at x_k, i.e.,

$$\mathcal{B}_k = \{x_k + s \in \Re^n \mid \|s\| \le \Delta_k\},$$

where $\|\cdot\|$ is the Euclidean norm. A step s_k is then computed that "sufficiently reduces" this model in this region, which is typically achieved by (approximately) solving the subproblem

$$\min_{\|s\| \le \Delta_k} m_k(x_k + s).$$

The objective function is then computed at the trial point $x_k + s_k$ and this trial point is accepted as the next iterate if and only if the ratio

$$\rho_k \stackrel{\text{def}}{=} \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)} \tag{2.1}$$

is larger than a small positive constant η_1. The value of the radius is finally updated to ensure that it is decreased when the trial point cannot be accepted as the next iterate, and is increased or unchanged if ρ_k is sufficiently large. In many practical trust-region algorithms, the model $m_k(x_k + s)$ is quadratic and takes the form

$$m_k(x_k + s) = f(x_k) + \langle g_k, s \rangle + \tfrac{1}{2}\langle s, H_k s \rangle, \tag{2.2}$$

where

$$g_k \stackrel{\text{def}}{=} \nabla_x m_k(x_k) = \nabla_x f(x_k), \tag{2.3}$$

H_k is a symmetric $n \times n$ approximation of $\nabla_{xx} f(x_k)$ and $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. If the model is not quadratic, it is assumed that it is twice-continuously differentiable and that the