Operations Research Letters 46 (2018) 7–12

Contents lists available at ScienceDirect

Operations Research Letters

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

An adaptive truncation criterion, for linesearch-based truncated Newton methods in large scale nonconvex optimization

Andrea Caliciotti [a], Giovanni Fasano [b], Stephen G. Nash [c], Massimo Roma [a,*]

[a] Dipartimento di Ingegneria Informatica, Automatica e Gestionale "A. Ruberti" SAPIENZA, Università di Roma, via Ariosto, 25 –00185 Roma, Italy
[b] Department of Management, University Ca' Foscari of Venice, S. Giobbe, Cannaregio 873 –30121 Venice, Italy
[c] Systems Engineering & Operations Research Department, George Mason University, 4400 University Drive Fairfax –VA 22030, USA

ARTICLE INFO

Article history:
Received 6 June 2016
Received in revised form 18 October 2017
Accepted 18 October 2017
Available online 28 October 2017

Keywords:
Large scale nonconvex optimization
Linesearch-based truncated Newton methods
Krylov subspace methods
Adaptive truncation criterion

ABSTRACT

Starting from the paper by Nash and Sofer (1990), we propose a heuristic adaptive truncation criterion for the inner iterations within linesearch-based truncated Newton methods. Our aim is to possibly avoid "over-solving" of the Newton equation, based on a comparison between the predicted reduction of the objective function and the actual reduction obtained. A numerical experience on unconstrained optimization problems highlights a satisfactory effectiveness and robustness of the adaptive criterion proposed, when a residual-based truncation criterion is selected.

1. Introduction

In this paper we consider the unconstrained optimization problem

$$\min f(x), \tag{1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a real valued function possibly nonconvex, and n is large. We consider the standard assumptions that f is twice continuously differentiable, and that for a given $x_0 \in \mathbb{R}^n$ the level set $\Omega_0 = \{x \in \mathbb{R}^n \mid f(x) \le f(x_0)\}$ is compact.

Truncated Newton methods are widely used for solving such problems. They are also called Newton–Krylov methods since a Krylov subspace method is usually employed, for approximately solving the Newton equation at each iteration. A general description of the truncated Newton methods can be found in the survey paper [17]. Now we briefly recall the main features of this class of methods. In the sequel, we denote by $f_k = f(x_k)$, $g_k = \nabla f(x_k)$ and $H_k = \nabla^2 f(x_k)$ respectively the function, the gradient and the Hessian matrix of the function f at the point x_k. Moreover, we do not assume any sparsity pattern for H_k.

It is well known that, given an initial guess x_0 of a local minimizer of problem (1), a truncated Newton method is based on two nested loops: the *outer iterations* which represent the actual steps of the method, where the current estimate of the solution is updated; the *inner iterations* which carry out an iterative algorithm for computing, at each outer iteration k, a search direction d_k by approximately solving the Newton equation

$$H_k d = -g_k. \tag{2}$$

The iterative algorithm used for solving (2) is actually "truncated", i.e. terminated before the exact solution is obtained. This strategy is based on the fact that, since the benefits of using a Newton direction are local, i.e. in the neighborhood of a stationary point, an accurate solution of (2) may be unjustified when x_k is far from a local optimizer. As matter of fact, in this case, a much simpler search direction can often perform comparably well. Instead, more accuracy is required when the iterates approach a local minimizer. A good trade-off between the accuracy in solving the Newton equation (2) and the computational effort employed per outer iteration is a key point, for the overall efficiency of a truncated Newton method. Indeed, there might be significant advantages to terminating the inner iterations early, when we are still far from a solution and the problem has significant nonlinearities. On the contrary, when close to a solution, there may be disadvantages to early terminating, inasmuch as the corresponding search direction d_k might be poor.

Since the early papers [4,5] where truncated Newton methods were introduced, the importance of an efficient truncation criterion for the inner iterations was pointed out. The stopping rule proposed therein is based on controlling the magnitude of the residual. Under suitable assumptions, this allows to guarantee local convergence and a good convergence rate. Another truncation rule

* Corresponding author.
E-mail addresses: caliciotti@dis.uniroma1.it (A. Caliciotti), fasano@unive.it (G. Fasano), snash@gmu.edu (S.G. Nash), roma@dis.uniroma1.it (M. Roma).

https://doi.org/10.1016/j.orl.2017.10.014

has been proposed in [19]. It is based on a comparison between the reduction of the quadratic model of the objective function, at the current iteration, and the average reduction per iteration of the model.

The proper choice of a suitable truncation criterion, along with the necessity to handle the indefinite case (see also [8]) and the choice of an effective preconditioning strategy, are three "open questions" within truncated Newton frameworks, listed in the early survey paper [21]. In subsequent years, even if the research activity on these topics was greatly developed, actually no definitive answer has been yet provided.

Another fundamental aspect of a truncated Newton method concerns the global convergence. Indeed, as well known, a globalization strategy must be adopted to this aim, i.e. the method must be embedded within a linesearch or a trust region framework.

On the basis of these remarks, in this paper we propose a simple adaptive truncation criterion for the inner iterations within a linesearch-based truncated Newton method, and analyze its effectiveness in both the unpreconditioned and the preconditioned case. Our aim is to define an additional rule which enables to avoid "over-solving" of the Newton equation (2) in some circumstances. The latter phenomenon occurs whenever unnecessary inner iterations are performed, so that indulging in solving the Newton equation does not produce a better search direction. This possibly yields a reduction of the overall inner iterations, for both convex and nonconvex problems. Our proposal is partially inspired by trust region approach (see e.g. [3]), and is based on a comparison between the reduction of the objective function predicted by the quadratic model, and the actual reduction obtained. In particular, we consider a linesearch–based truncated Newton method where the inner iterations are performed using the Conjugate Gradient (CG) algorithm. A numerical experience was carried on for a selection of large scale (convex and nonconvex) test problems from CUTEst collection [11], showing the satisfactory effectiveness and the robustness of the adaptive rule proposed, when the residual-based criterion is adopted. For the sake of brevity, we only report a few significant summary results. The complete numerical results are detailed in the companion paper [1].

The paper is organized as follows: in Section 2 the two commonest truncation criteria used in the literature are reported and discussed. Section 3 reports some introductory theoretical motivations which are at the basis of our proposal. The novel adaptive rule is introduced in Section 4 and a summary of the numerical testing is reported in Section 5. Finally, Section 6 reports some concluding remarks. As regards the notations, $\|v\|$ denotes the 2-norm of the vector $v \in \mathbb{R}^n$.

2. Common truncation criteria

In order to briefly recall the truncation criteria commonly used in the literature, we denote by d_k an approximate solution of (2), and by $r_k = H_k d_k + g_k$ the corresponding residual.

A natural stopping criterion for the inner iterations is the *residual-based criterion*, proposed in the seminal papers [4] and [5]. Indeed, the authors propose to terminate the inner iterations whenever the residual r_k is sufficiently small, namely

$$\frac{\|r_k\|}{\|g_k\|} \le \eta_k, \qquad (3)$$

for a specified value of η_k. It is well known that criterion (3) is scale invariant and that the choice of the *forcing sequence* $\{\eta_k\}$ is crucial for controlling the convergence rate of the algorithm. A widely used choice proposed in [5] is $\eta_k = \min\{1/k, \|g_k\|^r\}$, with $0 < r \le 1$. Other forcing sequences have been proposed later in [6] and [7]. The possibility to easily control the rate of convergence of the algorithm, by means of suitable choices of the sequence $\{\eta_k\}$, is a key point for this criterion. On the other hand, some drawbacks deriving from the adoption of this rule are well known. Indeed, at the jth inner iteration of the Krylov subspace method adopted for solving the Newton equation (2), a stationary point of the quadratic model

$$q_k(d) = \frac{1}{2} d^T H_k d + g_k^T d \qquad (4)$$

over the Krylov subspace

$$\mathcal{K}_j(H_k, g_k) = span\left\{g_k, H_k g_k, H_k^2 g_k, \ldots, H_k^{j-1} g_k\right\}$$

is sought. In the case of positive definite Hessian H_k, the quadratic model (4) has a global minimizer which exactly solves the Newton equation (2). Of course, this case corresponds to a null residual. Conversely, whenever an approximate solution is sought, monitoring the magnitude of the residual might be, as discussed in [19], misleading. Indeed, the actual decrease of the objective function values can be alternatively predicted by means of the quadratic model decrease; however, the magnitude of the residual r_k and the quadratic model $q_k(d_k)$ could be significantly different. Moreover, the rounding error in computing $\|r_k\|$ could be relevant since r_k is usually computed by recurrence. In addition, whenever a CG method is used in the inner iterations and H_k is positive definite, the quadratic model monotonically decreases as the inner iterations progress, while the sequence $\{\|r_k\|\}$ is not monotone (unlike, for instance, using MINRES).

These remarks induced the authors to propose in [19] also a truncation rule based on *the decrease of the quadratic model*, rather than considering only the residual. Namely, the truncation criterion proposed is the following: the inner iterations are terminated if, for a specified value of $\eta_k \in (0, 1)$,

$$\frac{q_k(d_j) - q_k(d_{j-1})}{\frac{q_k(d_j)}{j}} \le \eta_k, \qquad (5)$$

where d_j denotes the approximate solution of (2) at the jth inner iteration. This criterion is then based on the comparison between the reduction of the quadratic model $q_k(d_j) - q_k(d_{j-1})$, and the average reduction per iteration $q_k(d_j)/j$. The criterion (5) is often considered preferable to (3) , since it gains information directly from the values of the quadratic model. Moreover, in [9] it was extended to possibly consider also an indefinite Hessian matrix H_k, providing some theoretical results, too.

However, in the framework of truncated Newton methods, in unconstrained as well as in constrained optimization, some codes currently available on the web and commonly used by the optimizers community still adopt the residual-based truncation criterion (3), both within linesearch-based and trust region-based codes (see e.g. [2,13], [15] page 9, [12,22–24] and URL https://neos-guide.org/content/truncated-newton-methods). This might be due also to the fact that, as well known, the adoption of (3) ensures theoretical superlinear convergence. Conversely, the criterion based on the quadratic model reduction (5), with the suggested value of $\eta_k = 0.5$ (constant), guarantees only the theoretical linear rate of convergence [19], even if it works very efficiently in practice.

On the basis of these remarks, in this paper we focus on the possibility to "enrich" the residual-based criterion (3) by conveying, also in this case, information gained from the behavior of the quadratic model. To this aim, we propose an adaptive rule for deciding the maximum number of inner iterations allowed at each outer iteration. The latter rule combined with the criterion (3) should enhance the overall efficiency of a truncated Newton method, by possibly avoiding the over–solving phenomenon.

3. Motivation for the truncation rule

Both the stopping criteria (3) and (5) in the previous section may not prevent over-solving of the Newton equation. For the