Applied Mathematics and Computation 203 (2008) 297–307

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

A variant of trust-region methods for unconstrained optimization ☆

Fusheng Wang [a,*], Kecun Zhang [a], Chuanlong Wang [b], Li Wang [a]

[a] Faculty of Science, Xi'an Jiaotong University, Xi'an 710049, PR China
[b] Taiyuan Normal University, Taiyuan 030012, PR China

ARTICLE INFO

Keywords:
Trust-region methods
Linear model
Quadratic model
Conic model
Unconstrained optimization

ABSTRACT

In traditional trust-region methods, one, in practice, always employs quadratic model or conic model as the local approximation of the objective function, and there are lots of theoretical results and ripe algorithms. In this paper, we develop a practical trust-region algorithm with a linear model for unconstrained optimization problems. In particular, we combine a special weighted norm with the linear model so that the subproblem contains the information of Hessian matrix of the objective function, which successfully overcome the drawbacks of linear model, and we further complete the trust-region methods with three main types of models, namely, linear model, quadratic model and conic model. We show that the new method preserves the strong global convergence. Moreover, under the linear model, it unveils independently that the line-search algorithms can be viewed as a special case of trust-region methods. Numerical results indicate that the new method is effective and practical.

1. Introduction

In this paper, we study the unconstrained optimization problems:

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

Many algorithms are available for solving (1.1) when $f(x)$ is twice continuously differentiable in R^n. At present, they can reduce to two classes of iterate methods, that is, line-search methods and trust-region methods [9].

Quasi-Newton line-search methods is particularly used whenever the analytical expression of Hessian matrix of objective function is hard to get or is expensive to compute or store. The methods firstly compute a search direction:

$$d_k = -B_k^{-1}\nabla f(x_k), \tag{1.2}$$

where B_k is an appropriately generated matrix that approximates to the Hessian via updating formula, such as BFGS, DFP, SR1 etc. Secondly, in order to achieve global convergence, it is followed by an appropriate line-search, such as Wolfe-Powell line-search [8] etc., along the direction defined by (1.2) to obtain a step-length α_k. Finally, we obtain a new approximate point $x_{k+1} = x_k + \alpha_k d_k$ until it satisfies the terminative criterion.

Trust-region methods have attracted attention from more and more researchers since their emergence (see [1–25]), because they possess strong global convergence. They produce each iterate as the approximate minimizer of a relatively simple model function within a region in which the algorithm 'trust' that the model function behaves like f. Traditionally, one

☆ This work was supported by the National Natural Science Foundation of China under Grant 10671057.
* Corresponding author.
E-mail address: fswang2005@163.com (F. Wang).

0096-3003/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.amc.2008.04.049

usually chooses the model function to be quadratic approximation around the current iterate. Thus, traditional trust-region methods (hereafter simply denoted by TTR) compute a trial step by solving the subproblem:

$$\mathrm{QSP}:\begin{cases}\min \quad m_k(s)=g_k^{\mathrm{T}}s+\frac{1}{2}s^{\mathrm{T}}B_k s, & \text{(a)}\\ \text{s.t.} \quad \|s\|_2\leqslant \varDelta_k, & \text{(b)}\end{cases} \tag{1.3}$$

where $s=x-x_k\in R^n$, $g_k=\nabla f(x_k)$, B_k is an $n\times n$ symmetric matrix which approximates to the Hessian of objective function or chosen to be exact Hessian $B_k=\nabla^2 f(x_k)$, and $\varDelta_k>0$ is a trust-region radius. However, according to the TTR methods (see Algorithm 2.1 in Section 2), the above subproblem may be resolved many times at each iteration until the trial step is acceptable. This strategy is applicable for small-scale problems, but if the number of variables is large, resolving the subproblem can considerably increase the total cost of computation (see [2]), since this requires solving one or more linear systems of form $(B_k+\lambda I)s=-g_k$ [18]. In order to overcome this weakness, [2] proposed a nonmonotone trust-region method, and some authors proposed a hybrid algorithm which combining line-search methods with trust-region methods (see [12,13]). Motivated by the idea of linear model from [3], in Section 4 we will see that the new algorithm proposed establishes a new method to overcome this drawback successfully.

In addition, instead of quadratic model (1.3a), some authors (see [4–6]) study the conic model:

$$m_k(s)=\frac{g_k^{\mathrm{T}}s}{1+b_k^{\mathrm{T}}s}+\frac{1}{2}\frac{s^{\mathrm{T}}B_k s}{(1+b_k^{\mathrm{T}}s)^2}. \tag{1.4}$$

The conic model has certain advantages over the quadratic model, especially, for the objective function with high oscillation. But obviously it is more complex than quadratic model, and resolving the subproblem can be costly too.

Comparing the Quasi-Newton line-search methods with trust-region methods, we see that the former often requires more iterations to find a minimizer of f than does a trust-region methods, but it tends to compute each iterate more quickly than does a comparable trust-region methods [13]. Thus, how to incorporate the good qualities of the two classes methods is worth further investigating. It's widely known that making things simple is a very important principle in constructing an algorithm, and the linearized model is the simplest form of approximation to the objective function. The idea of linear model that can be allowed for a local approximation to objective function was considered by many authors(see [3,10]). However, how to embed the linear model into the trust-region methods to be a practical algorithm seems not to get enough emphasis. In this paper, we investigate a new type of trust-region method with linear model motivated by the following fact:

Quasi-Newton direction is the steepest descent one in the sense of weighted norm [8], that is,

$$\begin{cases}\min \quad f_k+g_k^{\mathrm{T}}s, & \text{(a)}\\ \text{s.t.} \quad \|s\|_{B_k}=1. & \text{(b)}\end{cases} \tag{1.5}$$

where $\|s\|_{B_k}=\sqrt{s^{\mathrm{T}}B_k s}$, (1.5a) is the linearized model of objective function at iterate x_k. Replacing (1.5b) with

$$\|s\|_{B_k}\leqslant \varDelta_k, \tag{1.6}$$

then we obtain a new subproblem

$$\mathrm{LSP}:\begin{cases}\min \quad m_k(s)=f_k+g_k^{\mathrm{T}}s\\ \text{s.t.} \quad \|s\|_{B_k}\leqslant \varDelta_k.\end{cases} \tag{1.7}$$

to compute the trial step, where $\varDelta_k>0$ is the trust-region radius. After studying the above subproblem (LSP), we found that the new type of trust-region method based on the linearized subproblem (LSP) had very good properties(see Section 4). Unlike the available trust-region methods which possibly solve the subproblem many times at each iteration, it does take only one time to solve the subproblem (LSP) at each iteration. So that the cost of computation in each iteration can considerably decrease. In addition, via the linear model, it unveils independently that the line-search algorithms can be viewed as a special case of trust-region methods.

The contents of this paper is organized as follows: In Section 2, we briefly describe the traditional trust-region method with quadratic model. In Section 3, the trust-region method with conic model is recalled simply. In Section 4, the new trust-region method with linear model is presented, and the global convergence is proved under certain conditions. In Section 5, we give some numerical experiments to compare the new algorithm both with quadratic model and with conic model. Finally, we end the paper with the conclusions.

We shall use the following notation and terminology. Unless otherwise stated, the vector norm used in this paper is Euclidean vector norm on R^n, and the matrix norm is the induced operator norm on $R^{n\times n}$, the weighted norm of a vector s with respect to a positive definite matrix B is denoted by $\|s\|_B=\sqrt{s^{\mathrm{T}}Bs}$, $g(x)\in R^n$ is the gradient of f evaluated at x, $H(x)\in R^{n\times n}$ is the Hessian of f evaluated at x, $f_k=f(x_k)$, $g_k=g(x_k)$, and $H_k=H(x_k)$.

2. The traditional trust-region methods with quadratic model

For easy reference, let us recall the traditional trust-region methods (TTR) first, and for an in-depth overview of the TTR methods (see books [7–10]). The TTR methods are based on a local quadratic model of $f(x_k+s)-f(x_k)$ about the kth iterate x_k