Applied Mathematics and Computation 204 (2008) 485–489

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

A new trust region filter algorithm ☆

Shujun Li [1], Zhenhai Liu *

School of Mathematics and Computing Science, Changsha University of Science and Technology, Changsha 410076, PR China

ARTICLE INFO

Keywords:
Unconstrained optimization
Trust region
Adaptive strategy
Filter method
Global convergence

ABSTRACT

In this paper, an adaptive trust region filter algorithm for unconstrained optimization is presented. The global convergence results of the algorithm are established and the numerical results show that the new algorithm is efficient in practical computation.

1. Introduction

Consider an unconstrained optimization problem

$$\min f(x), \tag{1.1}$$

where $f : R^n \rightarrow R^1$ is twice continuously differentiable. For convenience, we introduce some notations. Throughout the paper, the notation $\|\cdot\|$ denotes the Euclidean norm on R^n. Suppose $\{x_k\}$ is a sequence of points generated by our algorithm, we denote $f_k = f(x_k), g_k = g(x_k), g_{k,i} = g_i(x_k)$ and $H_k = H(x_k)$, where $g(x_k) \in R^n$ and $H(x_k) \in R^{n\times n}$ are the gradient and Hessian of $f(x)$ evaluated at x_k, respectively, B_k is a symmetric matrix which is either H_k or an approximation to H_k. The trust region algorithms [14] usually obtained the trail step by solving the following subproblem:

$$\begin{aligned} \min \quad & \varphi_k(d) = g_k^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} B_k d, \\ \text{s.t.} \quad & \|d\| \leqslant \Delta_k, \end{aligned} \tag{1.2}$$

where Δ_k is the trust region radius. It is well known that the trust region radius Δ_k is independent of g_k and B_k in the traditional trust region filter algorithms [2,9–12]. We do not know whether the radius Δ_k is suitable to the whole algorithm. This situation would possibly increase the number of solving subproblems and decrease the efficiency of the algorithms. Furthermore, the choice of Δ_0 also affects the efficiency of these algorithms, but we do not have any general rule on choosing Δ_0. Therefore, Sartenaer [13] presented a strategy which can automatically determine an initial trust region radius. Zhang et al. [15,16] presented other two strategies of determining the trust region radius by $\Delta_k = \frac{c^p\|g_k\|}{\delta_k}$ and $\Delta_k = \frac{c^p\|g_k\|^3}{g_k^{\mathrm{T}}\widehat{B_k}g_k}$, where $c \in (0,1)$, p is a nonnegative integer, $\delta_k = \min\{\|B_k\|, 1\}$ and $\widehat{B_k} = B_k + iI$ is a positive definite matrix for some i. In this paper, we present a new adaptive trust region strategy, which is different from these in [6,15,16]. By using the information of the

☆ This work was supported by the National Nature Science Foundation of China (Grant No. 10671211) and Nature Science Foundation of Hunan province (Grant No. 07JJY3005).

* Corresponding author.
E-mail addresses: lishujunnew@163.com (S. Li), zhhliu@mail.csu.edu.cn (Z. Liu).
[1] Present Address: Shaanxi Costume and Art College, Xianyang 712046, PR China.

0096-3003/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.amc.2008.07.007

former iterates, it is not necessary to compute $\widehat{B_k}$ to obtain Δ_k, which decreases the computational complexity of our algorithm. Therefore, our new strategy has been an improvement in a sence. Furthermore, by combining with filter technique, this new strategy will avoid this situation $\Delta_k = \infty$ when $g_k = g_{k-1}$ in [6].

The rest of this paper is organized as follows: In Section 2, we introduce the filter technique. The new algorithm is presented in Section 3. Some assumptions and the global convergence are presented in Section 4. The numerical result is summarized in Section 5.

2. The multidimensional filter

In this section, we introduce the mechanism of the filter, which is proposed by Flecher and Leyffer [3]. The definition of the filter is based on the definition of dominance. For our problem, we say that a point x_1 dominates a point x_2 if and only if

$$\|g_{1,i}\| \leqslant \|g_{2,i}\| \quad \text{for all} \quad i \in \{1, 2, \ldots, n\}.$$

Thus, if iterate x_1 dominates iterate x_2 and we focus our attention on the first-order critical point only, the latter has no real interest to us since x_1 is better than x_2 for each of the components of the gradient. Therefore, all we need to do is to remember iterates that are not dominated by other iterates by using a structure called filter [2–4,10]. A multidimensional filter $\mathbb{F}$ [3] is a list of of the form $(g_{k,1}, g_{k,2}, \ldots, g_{k,n})$ such that

$$|g_{k,j}| < |g_{l,j}| \quad \text{for at least one } j \in \{1, 2, \cdots, n\}, \; g_l \in \mathbb{F} \text{ and } k \neq l.$$

By F_k we denote the filter $\mathbb{F}$ at kth iterate.

However, we do not wish to accept a new point x_k^+ if one of the components of $(g_1(x_k^+), g_2(x_k^+), \ldots, g_n(x_k^+))$ is arbitrary close to being dominated by the other points already in the filter. In order to avoid this situation, we slightly strengthen our acceptable conditions and say that a new trial point x_k^+ is acceptable for the filter F_k if and only if

$$\forall g_m \in F_k \; \exists j \in \{1, 2, \ldots, n\} \quad \text{such that } \|g_j(x_k^+)| < |g_{m,j}| - \gamma_g \|g(x_m)\|, \tag{2.1}$$

where γ_g is a small positive constant. If x_k^+ satisfies (2.1), then we add $(g_1(x_k^+), g_2(x_k^+), \ldots, g_n(x_k^+))$ to the filter and remove all the points which are dominated by x_k^+ from the filter. This operation is also called "add x_k^+ to the filter" in the sequel.

3. Algorithm

In this section, we propose the new algorithm. At the current iterate x_k, we need to solve the subproblem (1.2) to obtain the trail step d_k. If d_k satisfies $r_k = \frac{f(x_k) - f(x_k + d_k)}{\varphi_k(0) - \varphi_k(d_k)} > u$, then we call d_k an f-step, otherwise d_k is a g-step if d_k satisfies (2.1). Now the algorithm for the solution of problem (1.1) can be presented as follow:

Algorithm 3.1.

Step 0: (Initialization) Given $x_0 \in R^n$, $B_0 \in R^{n \times n}$, set the parameters $0 < u < 1$, $0 < \lambda < 1, \epsilon > 0$. Initialize the filter as $F_0 = (g_1(x_0), g_2(x_0), \ldots, g_n(x_0))$. Let $k := 0$, $i := 0$.
Step 1: If $\|g_k\| \leqslant \epsilon$, then stop.
Step 2: Compute $\theta_k = \min\{\|g_m\| | (g_1(x_m), g_2(x_m) \ldots, g_n(x_m)) \in F_k\}$.
Step 3: Let $\Delta_k = \lambda^i \min\left\{\frac{\|d_{k-1}\|}{\|g_k - g_{k-1}\|}\|g_k\|, \|g_k\|, \theta_k\right\}$, solve the subproblem (1.2) to obtain d_k, set $x_k^+ = x_k + d_k$.
Step 4: Compute r_k, if d_k is an f-step, then $x_{k+1} = x_k^+, F_{k+1} = F_k$ and go to step 6, else go to step5.
Step 5: If d_k is a g-step, then $x_{k+1} = x_k^+$, add x_k^+ to F_k, $F_{k+1} = F_k$ and go to step 6, else $i = i + 1$ and go to step 3.
Step 6: Modify B_k to obtain B_{k+1}. Set $i = 0$, $k := k + 1$ and go to step 1.

4. Convergence analysis

In the convergence analysis, we need the following assumptions

A1 The gradient of the objective function f is Lipschitz with a constant L.
A2 The iterate $\{x_k\}$ remain in a closed, bounded convex domain of R^n.
A3 $\{B_k\}$ is uniformly bounded. i.e., there exists a constant M such that $\|B_k\| \leqslant M$ for all k.

The cycle between steps 1 and 6 are called outer cycle, and the cycle between steps 3 and 5 are called inner cycle. Let d_k^i be the solution of (1.2) with respect to $\Delta_k = \Delta_k^i = \lambda^i \min\left\{\frac{\|d_{k-1}\|}{\|g_k - g_{k-1}\|}\|g_k\|, \|g_k\|, \theta_k\right\}$ and d_k^* is the solution of (1.2) when the inner cycle at iterate x_k is terminated.