Operations Research Letters 45 (2017) 565–569

Contents lists available at ScienceDirect

Operations Research Letters

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

Descent line search scheme using Geršgorin circle theorem

Suvra Kanti Chakraborty, Geetanjali Panda *

Department of Mathematics, Indian Institute of Technology, Kharagpur-721302, India

ARTICLE INFO

Article history:
Received 21 February 2017
Received in revised form 16 July 2017
Accepted 14 August 2017
Available online 30 August 2017

Keywords:
Geršgorin disc
Wolfe inexact line search
Zoutendijk theorem
Super linear convergence

ABSTRACT

In the context of the minimization of a real function, we propose a line search scheme that involves a new positive definite modification of the Hessian. In this framework, a safeguard based on Geršgorin Circle's theorem provides an approximation of the Hessian that improves with iteration count. Convergence analysis of the scheme is validated by numerical experiments.

1. Introduction

Line search method is one of the most competent, effectual and expeditious methods for solving optimization problem, that has been a significant area of fascination and concern for the researchers (see [3,9,10,12,13]). The general structure of a descent line search iterative scheme for a function $f : R^n \to R$ is $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$, where $p^{(k)}$ is a descent direction of f at $x^{(k)}$ and α_k is the step length at $x^{(k)}$ along $p^{(k)}$. For Newton's method, $p^{(k)}$ is $\nabla^2 f(x^{(k)})^{-1}\nabla f(x^{(k)})$, where $\nabla^2 f(x^{(k)})$ and $\nabla f(x^{(k)})$ represent the Hessian and gradient of f at $x^{(k)}$ respectively. In this method if the initial guess is chosen far from the solution point, then the Hessian of the objective function may not be positive definite at every iterating point. In such case, the Hessian matrix can be modified to an approximate positive definite matrix to ensure the descent property of the scheme. Several techniques on positive definite modification of Hessian matrix are summarized in Section 3.4, Chapter 3, [7], which are based on computing eigenvalues using two basic strategies. One of these strategies directly computes the eigenvalues of the Hessian matrix at the current iteration point and then suitably adds a diagonal matrix of the form τI, where $\tau = \max(0, \eta - \lambda_{\min}(\nabla^2 f))$, η and $\lambda_{\min}(\nabla^2 f)$ being a small positive number and the minimum eigenvalue of the Hessian of f respectively. This is a computationally expensive process in higher dimension. Another strategy is based on the concept of modified symmetric indefinite factorization of the Hessian matrix (see [4,8]). This method factorizes the permuted Hessian matrix into LBL^T with a lower triangular matrix L and a block diagonal matrix B of at most 2×2 blocks, which makes the process computationally easier for computing eigenvalues in comparison to the first strategy.

In this paper, we propose a new approach that does not compute eigenvalues, not even implicitly, at any stage like earlier studies. This process first allows a positive definite safeguard at every iteration, and then backtracks sequentially to the Hessian matrix using Geršgorin circle theorem. The possibility of Cholesky factorization is verified once in each iteration for investigating the positive definiteness of a matrix. A real positive sequence (converging to 1), is assigned during the iteration process for generating the backtracking step. The global convergence property of the proposed scheme is established with Wolfe inexact line search under Zoutendijk condition. Further, it is proved that the modified matrix at each iteration converges to the Hessian matrix at the solution point. This fact ensures the superlinear convergence property of the proposed scheme. The computational experience on a set of test problems is provided for numerical support. The performance profiles for the number of iterations, the number of function evaluations, and the elapsed execution time on this test set are also presented.

The following two existing results are required to proceed for the theoretical development of this paper. The first result is based on Geršgorin Circle Theorem [6,11].

Theorem 1.1. *Let A be a complex matrix of order n, with entries a_{ij}. For $j \in \{1, 2, 3, \dots, n\}$, let $R_i = \sum_{j\neq i}|a_{ij}|$ and $D(a_{ii}, R_i)$ be the closed disc centred at a_{ii} with radius R_i. Such a disc is known as Geršgorin disc. Every eigenvalue λ of A lies within at least one of the Geršgorin disc $D(a_{ii}, R_i)$, that is, $|\lambda - a_{ii}| \leq R_i$ for some i.*

Theorem 1.2 (*Zoutendijk Theorem [14], Theorem No. 3.2 of [7]*). *Consider kth iteration of an optimization algorithm for minimizing $f(x)$, $x \in \mathbb{R}^n$ in the form $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$, where $p^{(k)}$ is a descent direction and α_k satisfies Wolfe condition. Suppose f is bounded below*

* Corresponding author.
E-mail address: geetanjali@maths.iitkgp.ernet.in (G. Panda).

http://dx.doi.org/10.1016/j.orl.2017.08.010

in $\mathbb{R}^n$ and continuously differentiable in an open set containing the level set $\mathcal{L} = \{x : f(x) \le f(x^{(0)})\}$, where $x^{(0)}$ is the starting point of the iteration. Assume also that ∇f is Lipschitz continuous on $\mathcal{L}$. That is, there exists a constant $L > 0$ such that $\|\nabla f(x) - \nabla f(\tilde{x})\| < L\|x - \tilde{x}\| \ \forall x, \tilde{x} \in \mathcal{L}$. Then $\sum_{k\ge 0}\cos^2\theta_k \|\nabla f(x^{(k)})\|^2 < \infty$, where θ_k is the angle between $p^{(k)}$ and $\nabla f(x^{(k)})$.

The rest of the paper is organized as follows. Section 2 proposes the idea of the new scheme and Section 3 describes the algorithm and convergence of the scheme. A detailed computational experience is illustrated in Section 4 and some concluding remarks are provided in Section 5.

2. Proposing new line search scheme

Consider an optimization problem

$$(P): \quad \min_{x\in\mathbb{R}^n} f(x),$$

where f is twice differentiable. We propose a Newton-like scheme that modifies kth iteration point as $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$, where α_k is the step length and $p^{(k)}$ is the descent direction to move along. Generally, $p^{(k)}$ is chosen as $-{D^{(k)}}^{-1}\nabla f(x^{(k)})$, where $D^{(k)}$ is a positive definite approximation of the Hessian matrix. In this framework, $D^{(k)}$ is formed based on Geršgorin Circle theorem which is used as a positive definite safeguard to the Hessian, that improves with iteration count. The construction of $D^{(k)}$ requires the following lemma.

Lemma 2.1. *The matrix $A = (a_{ij})_{n\times n}$, where $a_{ij} = \bar{a}_{ij} + \bar{a}_{ji}$, $\delta > 0$ a small number,*

$$\bar{a}_{ij} = \begin{cases} \dfrac{\partial^2 f}{\partial x_i \partial x_j} & \text{for } j < i, \\[2ex] \dfrac{1}{2}\left(\displaystyle\sum_{t=1}^{j-1}\left|\frac{\partial^2 f}{\partial x_i \partial x_t}\right| + \sum_{s=i+1}^{n}\left|\frac{\partial^2 f}{\partial x_s \partial x_j}\right| + \delta\right) & \text{for } j = i, \\[2ex] 0 & \text{for } j > i, \end{cases} \tag{1}$$

is positive definite.

Proof. A can be expressed explicitly as (see the equation in Box I).

Let $R_i = \sum_{j|i\neq j}|a_{ij}|, \forall i, j = 1, 2, \ldots, n$. From the construction of A,

$$a_{ii} > R_i. \tag{3}$$

By Geršgorin's Theorem, the eigenvalue λ of the symmetric matrix A satisfies the condition

$$|\lambda - a_{ii}| \le R_i, \tag{4}$$

which implies $-(\lambda - a_{ii}) \le R_i$ and consequently from (3), $\lambda \ge a_{ii} - R_i > 0$. Hence A is positive definite. □

At kth stage $\nabla^2 f(x^{(k)})$ may not be positive definite. Let us consider a real positive monotonically convergent sequence $\{c_k\}$ such that $\{c_k\} \to 1$, with $0 < c_0 < 1$ and construct a matrix $B^{(k)}$ as

$$B^{(k)} = (1 - c_k)A^{(k)} + c_k \nabla^2 f(x^{(k)}), \tag{5}$$

where $A^{(k)}$ is designed using Lemma 2.1. As $c_k \to 1$, $B^{(k)}$ is close to $\nabla^2 f(x^{(k)})$ and $A^{(k)}$ can be treated as a safeguard at this stage. For large k, $B^{(k)}$ approaches towards the actual Hessian matrix at local minimum point. At every stage, a check on $B^{(k)}$ determines whether it is a positive definite matrix. The following algorithm summarizes the idea described above.

Algorithm 1: Backtracking Based Positive Definite Hessian Modification

Data: Starting point $x^{(0)}$, ϵ, c_0;
for $k = 0, 1, 2 \ldots$
  1. Find $A^{(k)}$ using (1); Compute $B^{(k)}$ by (5);
  2. if Cholesky decomposition of $B^{(k)}$ is possible
    $D^{(k)} = B^{(k)}$
  else
    $D^{(k)} = A^{(k)}$;
  3. $x^{(k+1)} = x^{(k)} - \alpha_k {D^{(k)}}^{-1}\nabla f(x^{(k)})$, α_k satisfies Wolfe conditions;
  4. if $\|\nabla f(x^{(k+1)})\| < \epsilon$
    Stop;
  else
    $k = k + 1$, c_{k+1} is computed ;
  end;
end;

3. Convergence of the scheme

Consider f to be twice continuously differentiable. First, we recall Theorem 3.6 from [7], which shows that if the search direction approximates the Newton-direction well enough, then the unit step length will satisfy the Wolfe conditions as the iterates converge to the solution. Now, in addition to the assumptions of Theorem 3.6 from [7], we consider that the condition number $\kappa(D^{(k)})$ is uniformly bounded for each k. Then from the discussion of Section 3.2 of [7], global convergence of the proposed Newton-like scheme can be established as follows.

Theorem 3.1. *Let $\kappa(D^{(k)})$ denote the condition number of $D^{(k)}$. If there exists some $C > 0$ such that $\kappa(D^{(k)}) < C$ for every k, then under Theorem 1.2, $\|\nabla f(x^k)\| \to 0$ as $k \to \infty$.*

Proof. Let the eigenvalues of $D^{(k)}$ be $0 < \lambda_1^{(k)} \le \lambda_2^{(k)} \le \cdots \le \lambda_n^{(k)}$. For any $u \in \mathbb{R}^n$,

$$u^T D^{(k)} u \ge \lambda_1^{(k)}\|u\|^2.$$

Let θ_k be the angle between $p^{(k)}$ and $\nabla f(x^{(k)})$, where $p^{(k)} = -{D^{(k)}}^{-1}\nabla f(x^{(k)})$. Hence,

$$\cos\theta_k = -\frac{{\nabla f^{(k)}}^T p^{(k)}}{\|{\nabla f^{(k)}}^T\|\|p^{(k)}\|} = \frac{{p^{(k)}}^T D^{(k)} p^{(k)}}{\|{\nabla f^{(k)}}^T\|\|p^{(k)}\|} \ge \lambda_1^{(k)}\frac{\|p^{(k)}\|}{\|\nabla f(x^{(k)})\|}. \tag{6}$$

$\|\nabla f(x^{(k)})\| = \|D^{(k)}p^{(k)}\| \le \|D^{(k)}\|\ \|p^{(k)}\| = \lambda_n^{(k)}\|p^{(k)}\|$. Using this in (6), we have

$$\cos\theta_k = -\frac{{\nabla f^{(k)}}^T p^{(k)}}{\|{\nabla f^{(k)}}^T\|\ \|p^{(k)}\|} \ge \frac{\lambda_1^{(k)}}{\lambda_n^{(k)}} = \frac{1}{\|D^{(k)}\|\|{D^{(k)}}^{-1}\|} \ge \frac{1}{C}.$$

Hence, under Zoutendijk condition, $\lim_{k\to\infty}\|\nabla f(x^{(k)})\| = 0$. □

Moreover, in Section 3.3 of [7], it has been further stated that, if the descent search direction is of the form $p^{(k)} = -{D^{(k)}}^{-1}\nabla f(x^{(k)})$, then the condition of Theorem 3.6 of [7] may be put equivalently as

$$\lim_{k\to\infty}\frac{\|(D^{(k)} - \nabla^2 f(x^*))p^{(k)}\|}{\|p^{(k)}\|} = 0. \tag{7}$$

Hence, the superlinear convergence rate can be established if one can show that $B^{(k)}$ becomes increasingly accurate approximation to $\nabla^2 f(x^{(k)})$ along the search direction $p^{(k)}$, which we intend to show in the proof of the following theorem.

Theorem 3.2. *Suppose the sequence $\{x^{(k)}\}$ generated by Algorithm 1, converges to the solution x^*. Then $\{x^{(k)}\}$ converges to x^* superlinearly.*