Research Article

Eigenvalue Decomposition-Based Modified Newton Algorithm

Wen-jun Wang, Ju-bo Zhu, and Xiao-jun Duan

Department of Mathematics and System, Science College, National University of Defense Technology, Changsha 410073, China

Correspondence should be addressed to Wen-jun Wang; cammer@tom.com

Received 2 November 2012; Revised 3 February 2013; Accepted 19 February 2013

Academic Editor: Jian Guo Zhou

Copyright © 2013 Wen-jun Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When the Hessian matrix is not positive, the Newton direction may not be the descending direction. A new method named eigenvalue decomposition-based modified Newton algorithm is presented, which first takes the eigenvalue decomposition of the Hessian matrix, then replaces the negative eigenvalues with their absolute values, and finally reconstructs the Hessian matrix and modifies the searching direction. The new searching direction is always the descending direction. The convergence of the algorithm is proven and the conclusion on convergence rate is presented qualitatively. Finally, a numerical experiment is given for comparing the convergence domains of the modified algorithm and the classical algorithm.

1. Introduction

Newton algorithm (also known as Newton's method) is commonly used in numerical analysis, especially in nonlinear optimization [1]. But, in the definition domain of the object function, the method requires that (1) object function $f(\mathbf{x})$ is twice differentiable, (2) the Hessian matrix must be positive, and (3) the initial solution $\mathbf{x}^0$ should be near the extreme point $\mathbf{x}^k$ [2]. If there is a point in the definition domain whose Hessian matrix is not positive, then the Newton direction of this point is not the descending direction. So the classical algorithm may not be convergent in the definition domain.

There are several improved forms of the classical Newton algorithm on the above problem: (1) combining Newton direction with the steepest descent direction [3]; (2) combining Newton direction with the negative curvature direction [4]; (3) adding a diagonal matrix $v_k \mathbf{I}$ to the Hessian matrix so that the sum matrix is positive, where v_k is a positive number bigger than the absolutely value of the smallest eigenvalue [5]. This paper presents a new algorithm to handle the "nonpositive problem."

2. Newton Algorithm

We discuss the extreme value problems of multi-function. Suppose object function $f(\mathbf{x})$ is a twice differentiable function, and the extreme value problem of $f(\mathbf{x})$ is

$$\min_{\mathbf{x}} f(\mathbf{x}). \quad (1)$$

If the Hessian matrix of $f(\mathbf{x})$ is positive, the Taylor expansion to the second order of $f(\mathbf{x})$ is:

$$f(\mathbf{x}) \approx f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^k)^T \nabla^2 f(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k). \quad (2)$$

As we know, $\nabla_{\mathbf{x}} f(\mathbf{x}) = 0$ is the necessary condition, so the derivative form of (2) is

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = 0 + \nabla f(\mathbf{x}^k) + (\mathbf{x} - \mathbf{x}^k)^T \nabla^2 f(\mathbf{x}^k) = 0. \quad (3)$$

Equation (3) is usually called Newton's equation. When matrix $\nabla^2 f(\mathbf{x}^k)$ is positive, and the inverse matrix $[\nabla^2 f(\mathbf{x}^k)]^{-1}$ exists, we can get an iterative equation as follows:

$$\mathbf{x} = \mathbf{x}^k + [\nabla^2 f(\mathbf{x}^k)]^{-1} [-\nabla f(\mathbf{x}^k)]. \quad (4)$$

Equation (4) is usually called Newton iterative algorithm format, where $-\nabla f(\mathbf{x}^k)$ is called negative gradient direction, and $[\nabla^2 f(\mathbf{x}^k)]^{-1} [-\nabla f(\mathbf{x}^k)]$ is called Newton direction.

Next, we must give a proper step length to ensure the value of function declining. A well-known algorithm is "Damped Newton Algorithm," which gets the step length by solving the following optimization problem:

$$\alpha = \arg \min_{\alpha} \left\{ \mathbf{x}^k + \alpha [\nabla^2 f(\mathbf{x}^k)]^{-1} [-\nabla f(\mathbf{x}^k)] \right\}. \quad (5)$$

3. Modified Newton Algorithm Based on Eigenvalue Decomposition

3.1. Modified Algorithm. First, take eigenvalue decomposition on the Hessian matrix $[\nabla^2 f(\mathbf{x}^k)]^{-1}$:

$$[\nabla^2 f(\mathbf{x}^k)]^{-1} = \mathbf{U}\Sigma\mathbf{U}^H, \quad (6)$$

where $\mathbf{U}$ is a unitary matrix; Σ is a diagonal matrix. Because matrix $[\nabla^2 f(\mathbf{x}^k)]^{-1}$ is not positive, there must be a few negative eigenvalues in Σ . We replace the negative eigenvalues with their absolutely values, so the modified searching direction is:

$$\mathbf{d}^k = \mathbf{U}|\Sigma|\mathbf{U}^H[-\nabla f(\mathbf{x}^k)]. \quad (7)$$

The inner product of the negative gradient direction with $\mathbf{d}^k$ is always positive, because the right part of the next equation is a quadratic form [6]:

$$\begin{aligned} & \langle [-\nabla f(\mathbf{x}^k)], \mathbf{U}|\Sigma|\mathbf{U}^H[-\nabla f(\mathbf{x}^k)] \rangle \\ &= [-\nabla f(\mathbf{x}^k)]^T \mathbf{U}|\Sigma|\mathbf{U}^H[-\nabla f(\mathbf{x}^k)] > 0. \end{aligned} \quad (8)$$

Equation (8) shows that the angle between the negative gradient direction and $\mathbf{d}^k$ is less than 90° , so the new direction $\mathbf{d}^k$ is always a descent direction. Then we can get the similar iterative equation by replacing the Newton direction with $\mathbf{d}^k$, that is,

$$\mathbf{x} = \mathbf{x}^k + \alpha \mathbf{d}^k, \quad (9)$$

where α is the step length obtained by solving optimization problem (5).

3.2. Convergence Conclusion. We prove that the modified Newton algorithm based on eigenvalue decomposition is convergent. First, we talk about Wolfe-Powell condition [7].

If step length $\alpha_k > 0$ satisfies the two following equations:

$$\begin{aligned} f(\mathbf{x}^k + \alpha_k \mathbf{d}^k) - f(\mathbf{x}^k) &\leq \rho \alpha_k \nabla f(\mathbf{x}^k)^T \mathbf{d}^k, \\ \nabla f(\mathbf{x}^k + \alpha_k \mathbf{d}^k)^T \mathbf{d}^k &\geq \sigma \nabla f(\mathbf{x}^k)^T \mathbf{d}^k. \end{aligned} \quad (10)$$

Equation (10) is named Wolfe-Powell condition, where $\rho \in (0, 1)$ and $\sigma \in (\rho, 1)$.

Theorem 1. Let f be a continuously differentiable function, and let $\hat{\alpha}$ be defined by

$$\begin{aligned} \hat{\alpha} &= \min \left\{ \alpha \geq 0 \mid f(\mathbf{x}^k + \alpha \mathbf{d}^k) \right. \\ &\quad \left. = f(\mathbf{x}^k) + \alpha \rho \nabla f(\mathbf{x}^k)^T \mathbf{d}^k \right\}, \end{aligned} \quad (11)$$

where $\hat{\alpha}$ has a bound. Then there must be a finite interval $(a, b) \subseteq [0, \hat{\alpha}]$, for all $\alpha_k \in (a, b)$, the Wolfe-Powell conditions are right.

Proof. Let $\phi(t) \triangleq f(\mathbf{x}^k + t\mathbf{d}^k)$, as $\hat{\alpha} > 0$ and satisfies the following equation:

$$\phi(\alpha) \triangleq \phi(0) + \hat{\alpha} \rho \phi'(0). \quad (12)$$

Then

$$\phi(\alpha) < \phi(0) + \hat{\alpha} \rho \phi'(0) \quad \forall \alpha \in (0, \hat{\alpha}). \quad (13)$$

According to the Mean Value Theorem, there must be $\bar{\alpha} \in (0, \hat{\alpha})$ satisfies the following equation:

$$\frac{\phi(\hat{\alpha}) - \phi(0)}{\hat{\alpha}} < \phi'(\bar{\alpha}) < 0. \quad (14)$$

Notice $\rho \phi'(0) = \phi'(\bar{\alpha})$, and $\rho < \sigma$, so we can get:

$$\phi'(\bar{\alpha}) > \sigma \phi'(0). \quad (15)$$

Notice ϕ is a continuously differentiable operator, so there is an open set (a, b) satisfies $(a, b) \subset (0, \hat{\alpha})$, where $\bar{\alpha} \in (a, b)$. For any α_k when $\alpha_k \in (a, b)$, the following equations are right:

$$(1) \quad \phi(\alpha_k) \leq \phi(0) + \alpha_k \phi'(0), \quad (16)$$

$$(2) \quad \phi'(\alpha_k) \geq \sigma \phi'(0).$$

Finally, we can get the Wolfe-Powell condition. $\square$

Theorem 2. Let $f: \mathbf{R}^n \rightarrow \mathbf{R}$ be a continuously differentiable function, so f has lower bounded in $\mathbf{R}^n$; ∇f is uniformly continuous on O , which contains a horizontal set $\text{Lev}_{\alpha_0} := \{\mathbf{x} \in \mathbf{R}^n \mid f(\mathbf{x}) \leq \alpha_0\}$, where $\alpha_0 = f(\mathbf{x}^0)$ and $\mathbf{x}^0 \in \mathbf{R}^n$. If $\mathbf{d}^k$ satisfy the following conditions: $\theta^k \leq \pi/2 - \mu$, where θ^k stands for the angle between $-\nabla f(\mathbf{x}^k)$ and $\mathbf{d}^k$, and $\mu \in (0, \pi/2)$ is a constant, then there must be a k satisfying $\nabla f(\mathbf{x}^k) = 0$ or $\nabla f(\mathbf{x}^k) \rightarrow 0$ by Wolfe-Powell conditions.

Proof. We proof the theorem by contradiction. Suppose $\nabla f(\mathbf{x}^k) \neq 0$ or $\nabla f(\mathbf{x}^k) \rightarrow 0$ for all k , and there is a set $\{k_i\}$ and a constant $\varepsilon > 0$ satisfy $\|\nabla f(\mathbf{x}^{k_i})\| \geq \varepsilon$ for all i . Because $\mathbf{d}^k$ satisfies the angle condition $\theta^k \leq \pi/2 - \mu$, so

$$\begin{aligned} \nabla f(\mathbf{x}^k)^T \mathbf{d}^k &= -\|\nabla f(\mathbf{x}^k)\| \|\mathbf{d}^k\| \cos \theta_k \\ &\leq -\|\nabla f(\mathbf{x}^k)\| \|\mathbf{d}^k\| \sin \mu < 0. \end{aligned} \quad (17)$$

Equation (17) shows that $\mathbf{d}^k$ is a descending direction, so $\{f(\mathbf{x}^k)\}$ is a decreasing sequence. Suppose the limit of $\{f(\mathbf{x}^k)\}$ is ξ , and $\xi > -\infty$. That is $f(\mathbf{x}^k) \rightarrow \xi$, so $f(\mathbf{x}^k) - f(\mathbf{x}^{k+1}) \rightarrow 0$, take the Wolfe-Powell condition (16)(1) on index k_i ,

$$\begin{aligned} f(\mathbf{x}^{k_i}) - f(\mathbf{x}^{k_i+1}) &\geq -\rho \alpha_{k_i} \nabla f(\mathbf{x}^{k_i})^T \mathbf{d}^{k_i} \\ &\geq \|\nabla f(\mathbf{x}^{k_i})\| \|\alpha_{k_i} \mathbf{d}^{k_i}\| \cos \theta_{k_i} \\ &\geq \varepsilon \sin \mu \|\alpha_{k_i} \mathbf{d}^{k_i}\|. \end{aligned} \quad (18)$$

FIGURE 1: Results of numerical experiment (a) objective function; (b) contour map; (c) convergence domain of Newton's algorithm; (d) convergence domain of the new algorithm.

It is easy to get $\|\alpha_{k_i} \mathbf{d}^{k_i}\| \rightarrow 0$, again by the Wolfe-Powell condition (16)(2),

$$\left[\nabla f(\mathbf{x}^{k_i+1}) - \nabla f(\mathbf{x}^{k_i}) \right]^T \mathbf{d}^{k_i} \geq -(1-\sigma) \nabla f(\mathbf{x}^{k_i})^T \mathbf{d}^{k_i}. \quad (19)$$

Further

$$\|\nabla f(\mathbf{x}^{k_i+1}) - \nabla f(\mathbf{x}^{k_i})\| \geq \varepsilon(1-\sigma) \sin \mu. \quad (20)$$

Because of $\|\mathbf{x}^{k_i+1} - \mathbf{x}^{k_i}\| = \|\alpha_{k_i} \mathbf{d}^{k_i}\| \rightarrow 0$, (20) is in contradiction with the assumption that ∇f is uniformly continuous on Lev_{α_0} . The contradiction suggests that $\nabla f(\mathbf{x}^k) \rightarrow 0$. $\square$

4. Simulations on Convergence Properties

As the new searching direction $\mathbf{d}^k$ satisfies the following equation:

$$\mathbf{d}^k = \begin{cases} \mathbf{U} \Sigma \mathbf{U}^H \left[-\nabla f(\mathbf{x}^k) \right] & \text{if } \nabla^2 f(\mathbf{x}^k) > 0 \\ \mathbf{U} |\Sigma| \mathbf{U}^H \left[-\nabla f(\mathbf{x}^k) \right] & \text{else.} \end{cases} \quad (21)$$

So, if the Hessian matrix of all points in definition domain is positive, then the convergence rate of modified algorithm equals the classical Newton algorithm. The other side, if there are any points with its Hessian matrix negative, then the convergence rate at most equals the one of classical Newton algorithm.

In addition, in order to verify that the improved algorithm has a wider convergence domain than the classical Newton algorithm, we employ a convex function: $f(x, y) = -3(1-x)^2 e^{-x^2-(y+1)^2}$ as the objective function. Let the definition domain be $(x, y) \in \{[-3, 1], [-3, 1]\}$, and we choose the grid points by 0.1 interval, then use both classical Newton's algorithm and eigenvalue decomposition-based algorithm to find the minimum value. The results are drawn in Figure 1.

In Figure 1(c), the green point means that the algorithm has found the minimum value using this point as the initial solution; and the red point means that the algorithm did not converge or has not found the minimum value. Obviously,

the convergence domain of eigenvalue decomposition algorithm is bigger than the one of the classical Newton algorithm.

5. Conclusion

A new algorithm is put forward based on eigenvalue decomposition and classical Newton algorithm. The eigenvalue decomposition based-algorithm has a wider convergence domain than the classical Newton algorithm.

Acknowledgments

This work is supported by the Natural Science Foundation of China (60974124), the Program for New Century Excellent Talents in University, and the Project sponsored by SRF for ROCS, SEM in China, the key lab open foundation for Space Flight Dynamics technique (SFDLXZ-2010-004).

References

- [1] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Mass, USA, 1999.
- [2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
- [3] A. A. Goldstein and J. F. Price, "On descent from local minima," *Mathematics of Computation*, vol. 25, no. 115, pp. 569–574, 1971.
- [4] A. V. Fiacco and G. P. McCormick, "The sequential unconstrained minimization technique for nonlinear—a primal-dual method," *Management Science*, vol. 10, no. 2, pp. 360–365, 1964.
- [5] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY, USA, 1970.
- [6] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, NY, USA, 1985.
- [7] M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The Computer Journal*, vol. 7, pp. 155–162, 1964.