Journal of Computational Mathematics
Vol.42, No.6, 2024, 1656–1687.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2306-m2022-0279

ADAPTIVE REGULARIZED QUASI-NEWTON METHOD USING INEXACT FIRST-ORDER INFORMATION*

Hongzheng Ruan[1) and Weihong Yang
School of Mathematical Sciences, Fudan University, Shanghai 200433, P.R. China
Emails: hzruan19@fudan.edu.cn, whyang@fudan.edu.cn

Abstract

Classical quasi-Newton methods are widely used to solve nonlinear problems in which the first-order information is exact. In some practical problems, we can only obtain approximate values of the objective function and its gradient. It is necessary to design optimization algorithms that can utilize inexact first-order information. In this paper, we propose an adaptive regularized quasi-Newton method to solve such problems. Under some mild conditions, we prove the global convergence and establish the convergence rate of the adaptive regularized quasi-Newton method. Detailed implementations of our method, including the subspace technique to reduce the amount of computation, are presented. Encouraging numerical results demonstrate that the adaptive regularized quasi-Newton method is a promising method, which can utilize the inexact first-order information effectively.

Mathematics subject classification: 90C30, 68Q25.
Key words: Inexact first-order information, Regularization, Quasi-Newton method.

1. Introduction

In this paper, we consider the following problem:

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

where f is a continuously differentiable function. In many applications, the first-order information can not be known exactly. There are some errors in computations of $f(x)$ and $\nabla f(x)$, which are due to stochastic noise [10], internal discretization [29, 30], and gradient approximations based on finite differencing, interpolating, or smoothing [4, 5, 23, 24, 31, 33]. Problems of this type arise in a variety of fields, such as multidimensional numerical integration optimization [12], derivative-free optimization (DFO) [15, 28], and machine learning [16].

Devolder *et al.* [21] introduce the notion of first-order inexact oracle, and study the properties of several smooth and non-smooth convex optimization algorithms relying on such oracles. Inexact oracles have been widely studied for smoothing convex optimization problems. Readers are referred to [17, 20, 22, 32].

Inexact oracles can also be applied to DFO problems. In [7], a derivative-free method based on inexact oracles is proposed. Recently there has been much interest in the proximal method for composite optimization with inexact oracles. Readers are referred to [18, 35, 36].

Classical quasi-Newton algorithms have been extended to problems with inexact first-order information. An established BFGS algorithm with inexact gradient, named implicit filtering

* Received December 12, 2022 / Revised version received March 27, 2023 / Accepted June 28, 2023 / Published online November 26, 2023 /
1) Corresponding author

method, is proposed in [26]. It is designed for the case when errors are diminishing. Quasi-Newton method with non-diminishing and bounded errors (BFGSe) is analyzed in [40]. It describes a noise tolerant modification of the BFGS method. Berahas *et al.* [3] propose implementations of the BFGS method and L-BFGS method in which gradients are computed by an appropriate finite differencing technique. The study of BFGS methods with inaccurate gradients has attracted much interest in recent years. For details readers are referred to [6].

An interesting approach to problem (1.1) is the trust region method. Analysis for trust region methods with inexact gradients is presented in [11], which establishes strong global convergence results under the relative error condition. Convergence results for truncated trust region methods are established in [27]. Trust region methods can also be applied to problems where only stochastic gradient information can be used. Chen *et al.* [14] establish almost sure global convergence under the assumption that the first-order information is sufficiently accurate with high enough probability. In [2, 8], the authors analyze trust region methods with adaptive accuracy in function and gradient evaluation.

Inspired by the idea in [13, 25], we propose an adaptive regularized quasi-Newton method, named adaQN, to solve problem (1.1). Specifically, we approximate problem (1.1) and construct a subproblem by adding a regularization term to the quasi-Newton quadratic model. One major difference between adaQN and BFGSe is that adaQN does not use the line search and utilizes a trust-region-like framework to monitor the acceptance of trial steps. In many cases, this strategy can save some computational cost. Numerical experiments demonstrate the effectiveness of the adaQN method. For large scale problems, we incorporate subspace techniques [37, 41] into our adaQN method. Such techniques can reduce the amount of computation significantly in early iterations, especially when the dimension of the problem is very large. We also study the global convergence of the adaQN method. Under mild conditions, local convergence rates of adaQN are established.

This paper is organized into six sections. In Section 2, we describe the problem and propose the adaptive regularized quasi-Newton method. The main convergence results for different kinds of errors are presented in Section 3. In Section 4, we describe practical implementations of the proposed method. Numerical experiments, summarized in Section 5, indicate that the proposed method is more effective and robust. Some conclusions are drawn in Section 6.

2. Problem and Algorithm

Our goal in this section is to design an effective quasi-Newton method, which uses inexact first-order information to solve problem (1.1). We first introduce some notations that will be used throughout the paper for our descriptions. We use $g(x)$ to denote $\nabla f(x)$ in the paper. Given $x \in \mathbb{R}^n, \delta \geq 0$ and $\eta \geq 0$, we use $f_\delta(x)$ and $g_\eta(x)$ to denote the approximate values of $f(x)$ and $g(x)$ with errors controlled by δ and η, that is

$$|f(x) - f_\delta(x)| \leq \delta, \tag{2.1}$$

$$\|g(x) - g_\eta(x)\| \leq \eta. \tag{2.2}$$

Given a sequence $\{x_k\}$, $g(x_k)$ is denoted as g_k for ease of notation. Let $\{\eta_k\}$ be a non-negative sequence. We use the notation $\tilde{g}_k := g_{\eta_k}(x_k)$, where $g_{\eta_k}(x_k)$ satisfies (2.2).

In the classical quasi-Newton method, the approximate Hessian matrix B_k is updated by using $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = g_k - g_{k-1}$, and the search direction is computed by