Journal of Computational Mathematics
Vol.41, No.6, 2023, 1192–1221.

http://www.global-sci.org/jcm
doi:10.4208/jcm.2112-m2021-0072

A STOCHASTIC NEWTON METHOD FOR NONLINEAR EQUATIONS*

Jiani Wang
School of Mathematical Sciences, Dalian University of Technology, Dalian, China
Email: jianiwang@163.com
Xiao Wang[1)]
School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, China
Peng Cheng Laboratory, Shenzhen, China
Email: wangxiao@ucas.ac.cn; wangx07@pcl.ac.cn
Liwei Zhang
School of Mathematical Sciences, Dalian University of Technology, Dalian, China
Email: lwzhang@dlut.edu.cn

Abstract

In this paper, we study a stochastic Newton method for nonlinear equations, whose exact function information is difficult to obtain while only stochastic approximations are available. At each iteration of the proposed algorithm, an inexact Newton step is first computed based on stochastic zeroth- and first-order oracles. To encourage the possible reduction of the optimality error, we then take the unit step size if it is acceptable by an inexact Armijo line search condition. Otherwise, a small step size will be taken to help induce desired good properties. Then we investigate convergence properties of the proposed algorithm and obtain the almost sure global convergence under certain conditions. We also explore the computational complexities to find an approximate solution in terms of calls to stochastic zeroth- and first-order oracles, when the proposed algorithm returns a randomly chosen output. Furthermore, we analyze the local convergence properties of the algorithm and establish the local convergence rate in high probability. At last we present preliminary numerical tests and the results demonstrate the promising performances of the proposed algorithm.

Mathematics subject classification: 49M37, 65K05, 90C30.
Key words: Nonlinear equations, Stochastic approximation, Line search, Global convergence, Computational complexity, Local convergence rate.

1. Introduction

In this paper, we consider the following system of nonlinear equations:

$$F(x) = \mathbb{E}[f(x, \xi)] = 0, \tag{1.1}$$

where $\xi : \Omega \to \mathbb{W}$ is a random variable defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $f : \mathbb{R}^n \times \mathbb{W} \to \mathbb{R}^n$. Here $\mathbb{W}$ is a measurable space, and $\mathbb{E}$ represents the expectation with respect to the random variable. We assume that (1.1) has a solution and $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. As in (1.1) the probability distribution function may not be

* Received March 9, 2021 / Revised version received June 28, 2021 / Accepted December 7, 2021 / Published online December 12, 2022 /
[1)] Corresponding author

available or the expectation with respect to ξ may be difficult to calculate, we assume in this paper that the exact function information such as function value and Jacobian matrix cannot be obtained. The problem (1.1) covers a wide range of applications, such as stochastic dynamic programming [27], and stochastic PDEs [16, 23]. Problem (1.1) can also be regarded as an extension of finding a stationary point of minimization problems. Consider the well-known expected risk minimization problem

$$\min_{x\in\mathbb{R}^n} \ H(x) = \mathbb{E}[h(x,\xi)],$$

where $H:\mathbb{R}^n\to\mathbb{R}$ is a smooth function. Referring to Theorem 7.44 [28], if the function $h(x,\xi)$ is differentiable at x with probability 1 and satisfies certain conditions, its local minimizer is also the solution of the following system of equations:

$$\mathbb{E}[\nabla_x h(x,\xi)] = 0.$$

The classic Newton's method, first presented by Newton in 1687 [34], for finding the root of $F(x)$ is to update iterates through

$$x^{k+1} = x^k - (\mathcal{J}F(x^k))^{-1}F(x^k),$$

if the Jacobian matrix $\mathcal{J}F(x^k)$ is non-singular. As is well-known, classic Newton's method owns a fast convergence rate when close to a solution of nonlinear equations. More specifically, locally q-quadratic convergence rate can be achieved under certain conditions if $\mathcal{J}F(x^*)$ is non-singular [7]. However, when solving nonlinear equations (1.1), as the exact function value $F(x)$ and its Jacobian $\mathcal{J}F(x)$ are not available, classic Newton's method is not applicable any more. Similar to [8,21,25], we assume that $F(x)$ and Jacobian $\mathcal{J}F(x)$ can be approximated via calls to stochastic zeroth-order oracles ($\mathcal{SZO}$) and stochastic first-order oracles ($\mathcal{SFO}$), respectively. In this paper, we will employ those stochastic information to propose a stochastic approximation method for solving (1.1).

Study on stochastic approximation (SA) methods for nonlinear optimization dates back to the pioneer work [26]. In the past decade, along with the development of complexity theory, profound research progress on SA methods for nonlinear optimization has been made, including but not limited to [2, 9, 11, 13, 14]. Quite recently, Milzarek *et al.* [18] propose a stochastic semismooth Newton method for nonsmooth nonconvex optimization by solving an equivalent nonsmooth fixed point-type equation and study global and local convergence properties of the proposed algorithm. However, to define proximal gradient steps and to set a growth condition for trial steps, the objective function of original optimization problem needs to be utilized, thus plays a crucial role in theoretical analysis. In [29], a stochastic Gauss-Newton method (SGN) was proposed for solving compositional optimization problems. This method can be used to solve stochastic nonlinear equations (1.1) by reformulating it into an optimization problem, namely, minimizing a given norm of $F(x)$. At each iteration, it solves an approximate prox-linear model which is constructed based on stochastic oracles. In contrast, SA methods directly designed for general nonlinear equations in the form of (1.1) are quite limited. In this paper, motivated by the success of classic Newton's method for deterministic nonlinear equations, we will propose a stochastic Newton method for (1.1) based on stochastic oracles and investigate its theoretical and numerical performances.

As is known, the step size has a great influence on both theoretical and numerical performances of an SA method. It is quite popular to set step sizes as either constants [18] or