Journal of Computational Mathematics
Vol.37, No.1, 2019, 76–94.
http://www.global-sci.org/jcm
doi:10.4208/jcm.1710-m2017-0021

DESCENT DIRECTION STOCHASTIC APPROXIMATION ALGORITHM WITH ADAPTIVE STEP SIZES*

Zorana Lužanin
Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia
Email: zorana@dmi.uns.ac.rs
Irena Stojkovska
Department of Mathematics, Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University, Skopje, Macedonia,
Email: irenatra@pmf.ukim.mk
Milena Kresoja
Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia
Email: milena.kresoja@dmi.uns.ac.rs

Abstract

A stochastic approximation (SA) algorithm with new adaptive step sizes for solving unconstrained minimization problems in noisy environment is proposed. New adaptive step size scheme uses ordered statistics of fixed number of previous noisy function values as a criterion for accepting good and rejecting bad steps. The scheme allows the algorithm to move in bigger steps and avoid steps proportional to $1/k$ when it is expected that larger steps will improve the performance. An algorithm with the new adaptive scheme is defined for a general descent direction. The almost sure convergence is established. The performance of new algorithm is tested on a set of standard test problems and compared with relevant algorithms. Numerical results support theoretical expectations and verify efficiency of the algorithm regardless of chosen search direction and noise level. Numerical results on problems arising in machine learning are also presented. Linear regression problem is considered using real data set. The results suggest that the proposed algorithm shows promise.

Mathematics subject classification: 90C15, 62L20.
Key words: Unconstrained optimization, Stochastic optimization, Stochastic approximation, Noisy function, Adaptive step size, Descent direction, Linear regression model.

1. Introduction

The main aim of the paper is to propose and analyse a new algorithm with adaptive step sizes for solving stochastic optimization problems. The problem under our consideration is an unconstrained minimization problem in noisy environment,

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

where $f:\mathbb{R}^n\to\mathbb{R}$ is a continuously differentiable, possibly nonconvex function bounded below on $\mathbb{R}^n$. We assume that only noisy observations of the objective function $f(x)$ and its gradient

* Received February 7, 2017 / Revised version received June 12, 2017 / Accepted October 23, 2017 / Published online August 14, 2018 /

$\nabla f(x) = g(x)$ are available for all $x \in \mathbb{R}^n$. Denote by ξ and ε random variable and random vector, respectively, defined on a probability space $(\Omega, \mathcal{F}, P)$. The noisy function and noisy gradient at each $x \in \mathbb{R}^n$, in this set-up, are given by

$$F(x) = f(x) + \xi \quad \text{and} \quad G(x) = g(x) + \varepsilon, \tag{1.2}$$

where ξ and ε represent the random noise terms. Also, we denote by $x^* \in \mathbb{R}^n$ a stationary point of $f(x)$ in (1.1), that is $g(x^*) = 0$. Throughout the paper we will use the following notation

$$\begin{aligned} F_k &= F(x_k) = f(x_k) + \xi_k = f_k + \xi_k \\ G_k &= G(x_k) = g(x_k) + \varepsilon_k = g_k + \varepsilon_k, \end{aligned} \tag{1.3}$$

where x_k is kth iteration. Index k used with ε and ξ allows us to consider the cases when the noise-generating process may change with k. We will refer the standard deviation of the noise term ε as *noise level*.

The most common method for solving problem (1.1) is *Stochastic Approximation* (SA) algorithm proposed by Robbins and Monro, [16]. It is introduced for finding roots of one-dimensional nonlinear scalar function and later extended to multidimensional systems by Blum, [2]. Iterative rule of SA algorithm is motivated by the gradient direction method and uses only noisy gradient observations. For a given initial iteration x_0, iterative rule is given by the formula

$$x_{k+1} = x_k - a_k G_k, \tag{1.4}$$

where $a_k > 0$ is a step size and G_k is the noisy gradient at x_k defined by (1.3). The sequence $\{a_k\}$ is called the *sequence of step size lengths* or *gain sequence*. The convergence of SA method is achievable in a stochastic sense under certain assumptions. Robbins and Monro established mean square (m.s.) convergence, [16], while almost sure (a.s.) convergence is established by Chen, [7] and Spall, [18]. They proved that method (1.4) converges to a solution of the system $g(x) = 0$.

The performance of SA method depends mostly on the choice of the step size sequence. Numerous modifications of SA algorithm based on the step size selection are proposed to improve the optimization process. Kesten, [9], proposed an accelerated SA algorithm, for one dimensional case, with the step sizes that depend on the frequency of sign changes of the differences between two successive iterations. The a.s. convergence of the accelerated SA algorithm is established. The method is extended for multidimensional problems and a.s. convergence is proved by Delyon and Juditsky, [8]. Idea of monitoring sign is further studied by Xu and Dai, [21]. An algorithm with adaptive step sizes is proposed by Yousefian et al., [22] where authors propose a scheme for minimizing strongly convex differentiable functions in noisy environment. The scheme generates a step size sequence that is a decreasing piecewise-constant function with a decrease that occurs when a suitable threshold error is met. SA algorithm with a line-search is proposed by Krejić et al., [10]. A line search along the negative gradient direction is applied while the iterates are far away from the solution and upon reaching some neighbourhood of the solution the method switches to SA rule. Approach in [10] is recently extended to general descent direction case by Krejić et al., [11]. This result allows application of faster, second-order methods while keeping the almost sure convergence. Algorithms that use second-order directions are frequently applied for solving large-scale problems in machine learning. SA algorithm with a quasi-Newton direction is successfully applied in [4–6]. A stochastic quasi-Newton method for solving nonconvex stochastic optimization problems is also proposed