Master's Thesis

Adaptive Gradient Method for Stochastic Optimization on Riemannian Manifolds

Guidance

Program-Specific Associate Professor Hiroyuki SATO
Professor Nobuo YAMASHITA

Shun TAKAYANAGI

Department of Applied Mathematics and Physics

Graduate School of Informatics

Kyoto University

February 2019

Abstract

In recent years, optimization theory on Riemannian manifolds and its applications have been attracting much attention. In this paper, we focus on stochastic optimization problems on Riemannian manifolds. For solving the problems, first order stochastic optimization algorithms such as the stochastic gradient descent (SGD) method on Riemannian manifolds have been proposed. The SGD is a kind of online-type method, and it uses a single observed data to stochastically approximate an unknown objective function. Thanks to the recent development of the Riemannian concepts such as a vector transport and retraction and systematical summarization of Riemannian optimization algorithms, the Riemannian SGD has been shown to be globally convergent provided that the step size is mildly decreasing. However, since the step size used in the existing algorithm is not well-scaled, the convergence is slow. Recently, several adaptive controls of the step size have been proposed for the stochastic optimization on usual Euclidean space. Since the adaptive controls exploit the values of each element of the past gradients, the step size is well-scaled in each coordinate of Euclidean space. However, since the adaptive controls are based on the global coordinate axes, with which a general Riemannian manifold is not endowed, it is not straightforward to generalize the adaptive control methods to Riemannian manifolds as the extension of the SGD.

Quite recently, G. Becigneul and O. E. Ganea generalized AdaGrad in 2018, one of the most famous adaptive methods, to a certain Riemannian manifold. However, since the manifold is restrictive, their method cannot be applied to a general Riemannian manifold. Therefore, in this paper, we generalize AdaGrad to complete Riemannian manifolds. This generalization deals with a broader class of Riemannian manifolds and our algorithm is more natural than the method by Becigneul and O. E. Ganea since we exploit the main idea of AdaGrad, which is to scale the search direction according to each local coordinate axis by repeatedly translating the orthonormal basis. we exploit the exponential mapping and parallel translation as a retraction and vector transport, respectively, which enable us to develop the theoretical analysis smoothly. In addition, we give a regret bound for the proposed method when the objective function is geodesically convex. The bound is better than the result in the previous researches. Furthermore, we conduct numerical experiments for solving the leading eigenvector problems and Riemannian centroid problems. The results show that the proposed method converges to an optimal solution faster than the SGDs.