

## Finite-Sum Optimization: Adaptivity to Smoothness and Loopless Variance Reduction

**Authors :** Bastien Batardière, Joon Kwon

**Abstract :** For finite-sum optimization, variance-reduced gradient methods (VR) compute at each iteration the gradient of a single function (or of a mini-batch), and yet achieve faster convergence than SGD thanks to a carefully crafted lower-variance stochastic gradient estimator that reuses past gradients. Another important line of research of the past decade in continuous optimization is the adaptive algorithms such as AdaGrad, that dynamically adjust the (possibly coordinate-wise) learning rate to past gradients and thereby adapt to the geometry of the objective function. Variants such as RMSprop and Adam demonstrate outstanding practical performance that have contributed to the success of deep learning. In this work, we present AdaLVR, which combines the AdaGrad algorithm with loopless variance-reduced gradient estimators such as SAGA or L-SVRG that benefits from a straightforward construction and a streamlined analysis. We assess that AdaLVR inherits both good convergence properties from VR methods and the adaptive nature of AdaGrad: in the case of  $L$ -smooth convex functions we establish a gradient complexity of  $O(n + (L + \sqrt{nL})/\epsilon)$  without prior knowledge of  $L$ . Numerical experiments demonstrate the superiority of AdaLVR over state-of-the-art methods. Moreover, we empirically show that the RMSprop and Adam algorithm combined with variance-reduced gradients estimators achieve even faster convergence.

**Keywords :** convex optimization, variance reduction, adaptive algorithms, loopless

**Conference Title :** ICOA 2024 : International Conference on Optimization Algorithms

**Conference Location :** Tokyo, Japan

**Conference Dates :** September 05-06, 2024