

Linear Convergence Rate in Convex Setup is Possible! First- and Zero-Order Algorithms under Generalized Smoothness

vendredi 1 août 2025 11:15 (30 minutes)

The gradient descent (GD) method – is a fundamental and likely the most popular optimization algorithm in machine learning (ML), with a history traced back to a paper in 1847 (Cauchy, 1847). In this paper, we provide an improved convergence analysis of gradient descent and its variants, assuming generalized smoothness (L0,L1). In particular, we show that GD has the following behavior of convergence in the convex setup: At first, the algorithm has linear convergence, and approaching the solution, has standard sublinear rate. Moreover, we show that this behavior of convergence is also common for its variants using different types of oracle: Normalized Gradient Descent as well as Clipped Gradient Descent (the case when the oracle has access to the full gradient); Random Coordinate Descent (when the oracle has access only to the gradient component); Random Coordinate Descent with Order Oracle (when the oracle has access only to the comparison value of the objective function). In addition, we also analyze the behavior of convergence rate of GD algorithm in a strongly convex setup.

Author: LOBANOV, Aleksandr (MIPT)

Orateur: LOBANOV, Aleksandr (MIPT)

Classification de Session: ML

Classification de thématique: Machine learning