Small Errors in Random Zeroth-Order Optimization are Imaginary

Daniel Kuhn
EPFL, Switzerland
daniel.kuhn@epfl.ch

Most zeroth-order optimization algorithms mimic a first-order algorithm but replace the gradient of the objective function with some noisy gradient estimator that can be computed from a small number of function evaluations. This estimator is constructed randomly, and its expectation matches the gradient of a smooth approximation of the objective function whose quality improves as the underlying smoothing parameter δ is reduced. Gradient estimators requiring a smaller number of function evaluations are preferable from a computational point of view. While estimators based on a single function evaluation can be obtained by a clever use of the divergence theorem from vector calculus, their variance explodes as δ tends to 0. Estimators based on multiple function evaluations, on the other hand, suffer from numerical cancellation when δ tends to 0. To combat both effects simultaneously, we extend the objective function to the complex domain and construct a gradient estimator that evaluates the objective at a complex point whose coordinates have small imaginary parts of the order δ. As this estimator requires only one function evaluation, it is immune to cancellation. In addition, its variance remains bounded as δ tends to 0. We prove that zeroth-order algorithms that use our estimator offer the same theoretical convergence guarantees as the state-of-the-art methods. Numerical experiments suggest, however, that they often converge faster in practice.

Joint work with Wouter Jongeneel (EPFL) and Man-Chung Yue (The University of Hong Kong).