Using Taylor-approximated Gradients to improve the Frank-Wolfe method for Empirical Risk Minimization

Robert M. Freund
joint work with Zikai Xiong

Workshop in honor of the 60th Birthday of Adrian Lewis
Seattle, 9–10 August, 2022

The Frank-Wolfe method has become increasingly useful in statistical and machine learning applications, due to the structure-inducing properties of the iterates, and especially in settings where linear minimization over the feasible set is more computationally efficient than projection. In the setting of Empirical Risk Minimization – one of the fundamental optimization problems in statistical and machine learning – the computational effectiveness of Frank-Wolfe methods typically grows linearly in the number of data observations n. This is in stark contrast to the case for typical stochastic projection methods. In order to reduce this dependence on n, we look to second-order smoothness of typical smooth loss functions (least squares loss and logistic loss, for example) and we propose amending the Frank-Wolfe method with Taylor series-approximated gradients, including variants for both deterministic and stochastic settings. Compared with current state-of-the-art methods in the regime where the optimality tolerance ε is sufficiently small, our methods are able to simultaneously reduce the dependence on large n while obtaining optimal convergence rates of Frank-Wolfe methods, in both the convex and non-convex settings. We also propose a novel adaptive step-size approach for which we have computational guarantees. Last of all, we present computational experiments which show that our methods exhibit very significant speed-ups over existing methods on real-world datasets for both convex and non-convex binary classification problems.