Abstract: There has been a recent surge of interest for finding local minima of nonconvex functions, mostly due to the outbreak of such instances in data science. In this setting, one aims at developing algorithms with suitable worst-case complexity properties. However, those guarantees may differ in terms of cost measure (number of iterations, evaluations or arithmetic calculations) as well as in strength (deterministic or high-probability results), thereby making it difficult to compare various strategies. In addition, endowing an algorithm with a complexity analysis is sometimes done in a theoretical fashion, and this can be detrimental to the practical use of such a method.

In this talk, we consider a classical approach in large-scale optimization, where the (linear) conjugate gradient algorithm is incorporated in a Newton-type method. We first present a simple instance of this scheme, for which complexity bounds can be derived: those are optimal within a large class of second-order methods. We then propose new variants that compare favorably to recently proposed algorithms based on accelerated gradient in terms of computational cost. To this end, we revisit the convergence theory of the conjugate gradient algorithm when applied to a nonconvex quadratic function. We also leverage randomized linear algebra techniques to allow for second-order complexity results at a suitable cost. By incorporating these tools within our Newton-type framework, we are able to match the guarantees of recently proposed algorithms based on accelerated gradient. Finally, we describe some features of a good implementation of our strategies, and illustrate their performance on several types of nonconvex problems.