Numerical optimization methods play a crucial role in large-scale supervised learning. In the first part of the talk, I will review the most popular methods used today in industry and discuss their limitations --- in particular the difficulty of parallelizing them. I will then present a new algorithm designed to yield good generalization error (not just training error) and illustrate its performance on machine learning test problems.

**BIO:** My research interests are in optimization and its application in machine learning and in disciplines involving differential equations. I specialize in nonlinear optimization, both convex and non-convex; deterministic and stochastic. There is a need for solving ever larger optimization problems, and throughout the years, I have developed algorithms that scale well with the number of variables, make judicious use of second-order information, and parallelize well. The motivation for my current algorithmic and theoretical research stems from applications in image and speech recognition, recommendation systems, and search engines.