

***SRKCD: a stabilized Runge-Kutta method for large-scale optimization problems***

**Måns Williamson** @ (Lund University), Tony Stillfjord

R 3.28 Wed Z2 11:20-11:30

In large-scale machine learning problems, one typically faces the problem of finding the minimum of a loss function which depends on many parameters of large dimensionality. Because of this, standard optimization algorithms such as the gradient descent are often too costly to be used. The classic, popular solution to this is to use an algorithm known as the stochastic gradient descent. This can be described as a version of the traditional gradient descent where each update depends on a random parameter. However, by reformulating the problem as a gradient flow, we see that gradient descent is equivalent to the explicit Euler time-stepping scheme, which has a very small stability region. For many problems, this prevents us from taking large steps and reaching the minimum quickly. This is an issue also for the stochastic version.

In this talk, we consider instead the Runge–Kutta–Chebyshev algorithms, which are explicit time integration methods designed such that their stability region are maximized. It has recently been shown that these methods perform very well in the deterministic optimization setting, i.e. for full gradient descent, with supporting theory for convex quadratic problems. Here, we extend this to the stochastic case. We prove convergence in expectation to a unique minimum for strongly convex, nonlinear functions. A similar approach also leads to convergence for non-convex functions, but in a weaker sense. Finally, we consider a numerical example arising from a supervised learning application which demonstrates the benefits of the scheme.