TIME INTEGRATION FOR OPTIMIZATION PROBLEMS ARISING IN SUPERVISED LEARNING

Tony Stillfjord[1*], Monika Eisenmann[1], Måns Williamson[1]

[1]Centre for Mathematical Sciences, Lund University, Sweden
* tony.stillfjord@math.lth.se

A key problem in supervised learning applications is to solve the arising optimization problem. In the AI community, this is referred to as the training phase. Doing this efficiently is critical if one wants to, e.g., rapidly investigate different neural network architectures. In this talk, I will present some initial results on this topic from the viewpoint of time integration.

Our main idea is to reformulate the optimization problem as a gradient flow. Then the optimum is a stationary point which we reach as time goes to infinity. In this context, the stochastic gradient descent method (on which most modern training algorithms are based) corresponds to an inexact explicit Euler scheme. It is well known that choosing the initial learning rate for this method is critical to its success, and from our new viewpoint we can observe that it is in fact a step size limitation due to stability problems.

If stability is an issue, then a natural solution would be to instead apply an implicit scheme, or at least a method with a large stability region. I will present results on a stochastic version of the implicit Euler scheme for certain problem classes such as support vector machines, as well as on the so-called tamed Euler scheme for more general problem classes. For both, we prove convergence with sub-linear (optimal) rates in infinite-dimensional settings. I will also show the results of a few numerical experiments that support our theoretical results and illustrate why these methods are beneficial.