CRUNCH Seminars at Brown, Division of Applied Mathematics

Friday – May 18, 2018

Deep Relaxation: partial differential equations for optimizing deep neural networks (Paper Review)

Mamikon Gulian

In this paper, the authors establish a connection between non-convex optimization methods for training deep neural networks and nonlinear partial differential equations (PDEs). Relaxation techniques arising in statistical physics which have already been used successfully in this context are reinterpreted as solutions of a viscous Hamilton-Jacobi PDE. Using a stochastic control interpretation allows them prove that the modified algorithm performs better in expectation that stochastic gradient descent. Well-known PDE regularity results allow us to analyze the geometry of the relaxed energy landscape, confirming empirical evidence. The PDE is derived from a stochastic homogenization problem, which arises in the implementation of the algorithm. The algorithms scale well in practice and can effectively tackle the high dimensionality of modern neural networks.