

## **CRUNCH Seminars at Brown, Division of Applied Mathematics**

**Friday – May 18, 2018**

### **Doing the impossible: Why neural networks can be trained at all (Paper Review)**

Anna Lischke

As deep neural networks grow in size, from thousands to millions to billions of weights, the performance of those networks becomes limited by our ability to accurately train them. A common naive question arises: if we have a system with billions of degrees of freedom, don't we also need billions of samples to train it? Of course, the success of deep learning indicates that reliable models can be learned with reasonable amounts of data. Similar questions arise in protein folding, spin glasses and biological neural networks. With effectively infinite potential folding/spin/wiring configurations, how does the system find the precise arrangement that leads to useful and robust results? Simple sampling of the possible configurations until an optimal one is reached is not a viable option even if one waited for the age of the universe. On the contrary, there appears to be a mechanism in the above phenomena that forces them to achieve configurations that live on a low-dimensional manifold, avoiding the curse of dimensionality. In the current work we use the concept of mutual information between successive layers of a deep neural network to elucidate this mechanism and suggest possible ways of exploiting it to accelerate training. We show that adding structure to the neural network that enforces higher mutual information between layers speeds training and leads to more accurate results. High mutual information between layers implies that the effective number of free parameters is exponentially smaller than the raw number of tunable weights.