June/10/2014

# IDENTIFYING AND ATTACKING THE SADDLE POINT PROBLEM IN HIGH-DIM NON-CONVEX OPTIMIZATION

They State as we all Know that it's Difficult to optimize nonconvex, functions in high dimentions. It is often thought that the issue is the prevalance of many local minima that have much higher cost that the global minima. Here they claim that a more profound difficulty originates from the fact that their are many saddle points and not local minima. They show evidance to support this then an better optimization algorithm.

They say that often our geometric intuition derived from the low dimentional physical word is inadiquate when thinking in higher dimentions.

They cite and explain a few reasonings for the prevolunce of Saddle point in higher dimentions.

One being: When in higher Dimentions the prop that a critical point has all the directions going up around it is low and get's lower as the dimentions go up. So Local minima become less likley as Saddel points Become more likley.

Now after they established the prevolunce of Saddel Points. we examine How diffrent optimization algo's work. So they Note many optimization Algos but since the paper is 10 years old they don't Note all optimizers use now.

SGD: - Steps in the right Direction near Saddel Point
- but very slow since Small step Size in directions with Small eigenvalues

They then propose a generalized trusted region + a new Saddle-free-Newtonian method that minimizes the first term of the taylors expansion + also has a constraint based on the absolute Hessian.

the SFN moves in direction with low curvature faster + slower in those with higher curvature to escape saddle points.

This works well But falls apart in higher Dimentions. However it was helpful in a Recurrent NN which are know to be difficult to train.

They did this by training with SGD then switching to SFN when Learing Stalled.

*Note not in paper*
But often doing Better on train set does not directly mean Better generalization results.