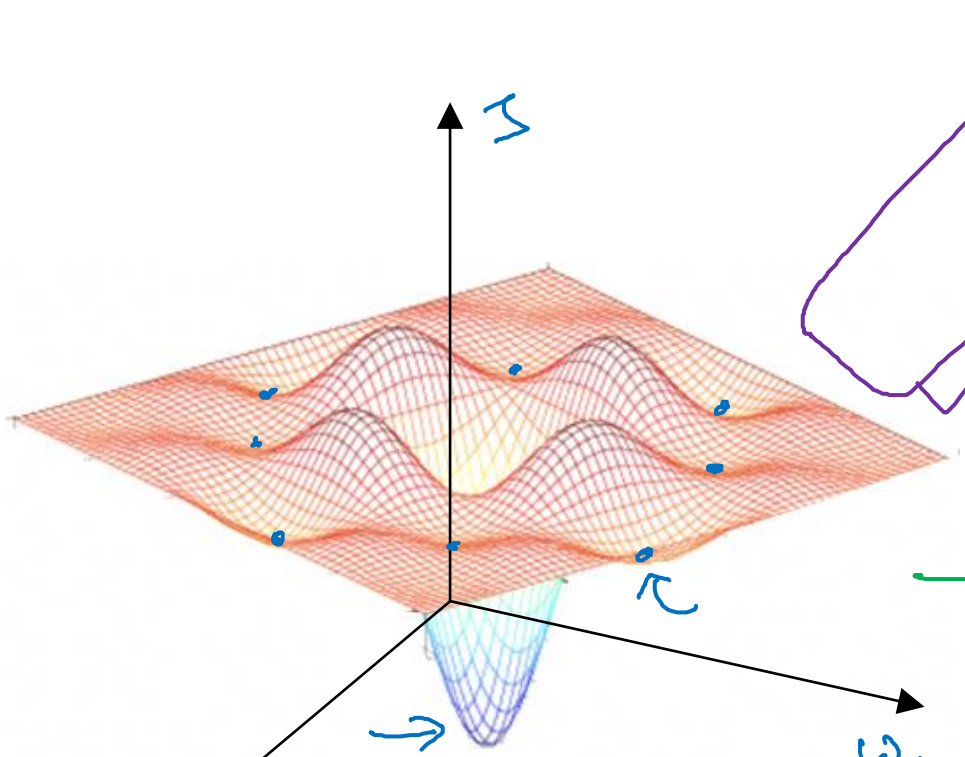
deeplearning.ai

Optimization Algorithms

The problem of local optima

Local optima in neural networks



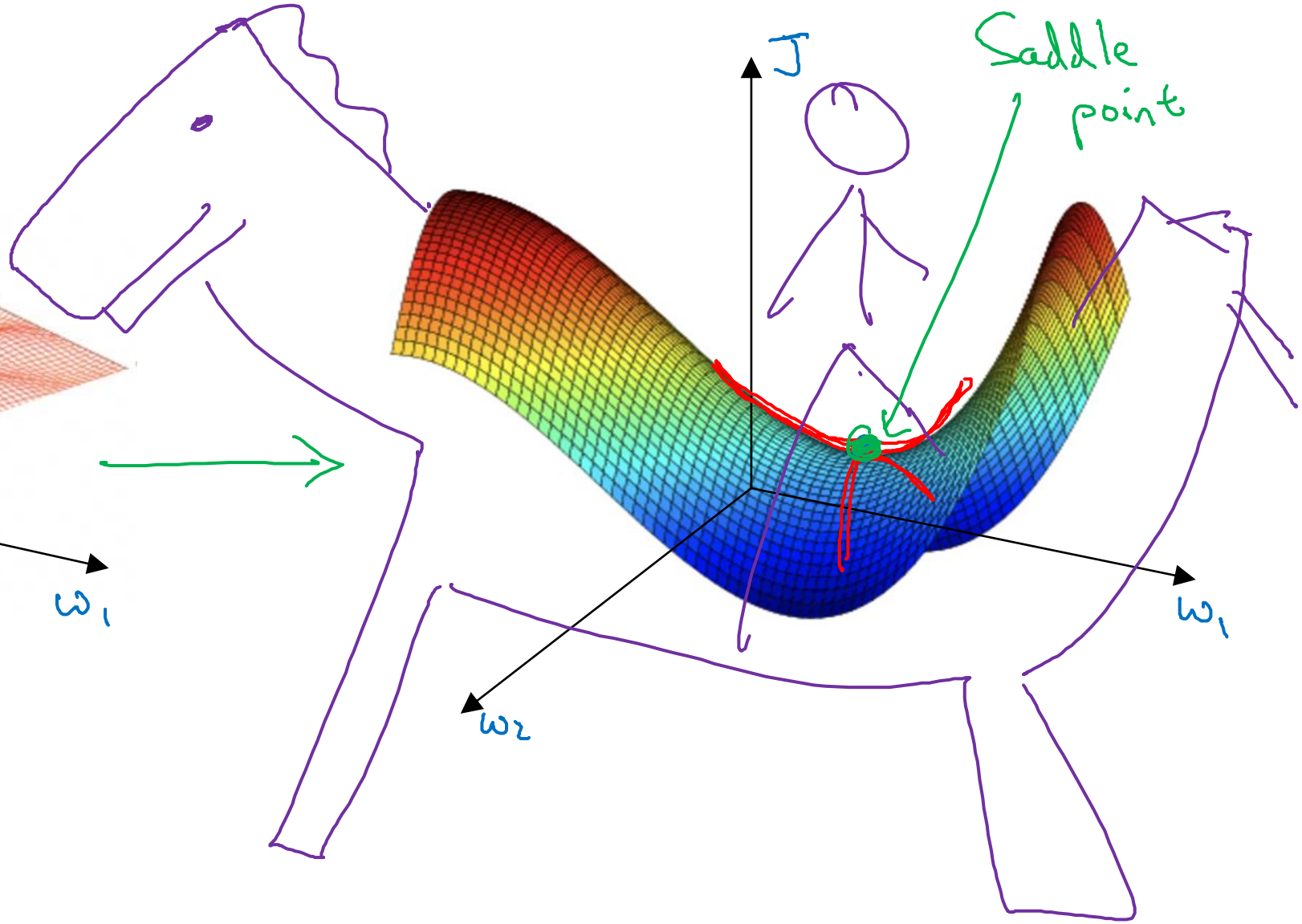
saddle point

convex light function

concave light function

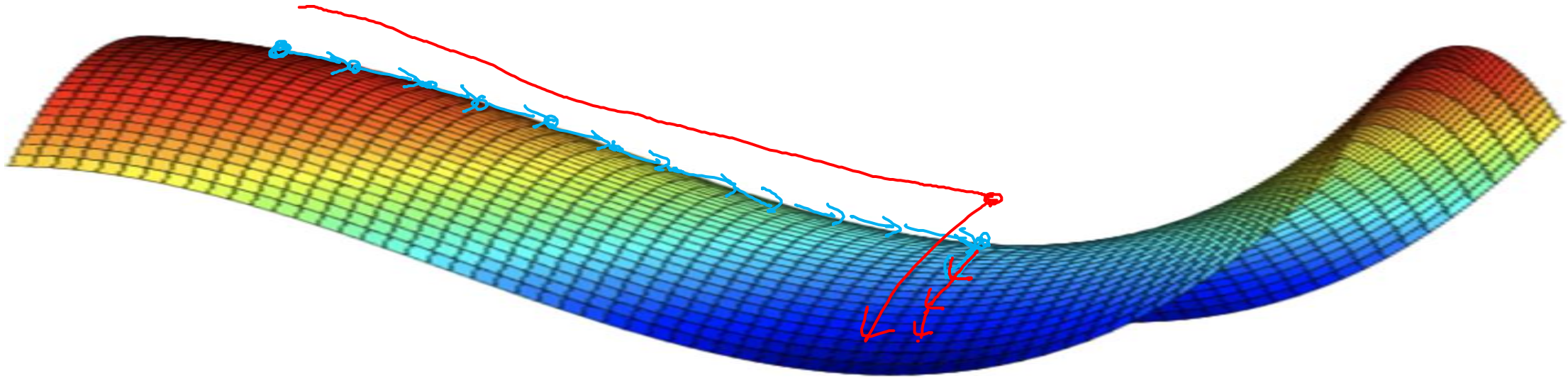
20000

2^{-20000}



Problem of plateaus

plateaus really slow down learning
A plateau is a region where the derivative is close to zero for a long time.



so long as you're training a reasonably large neural network, so a lot of parameters and the cost function J is define over a relatively high dimensional space.

- Unlikely to get stuck in a bad local optima
- Plateaus can make learning slow

And this is where algorithms like momentum or RMSprop or Adam can really help your learning algorithm as well. And these are scenarios where more sophisticated observation algorithms such as Adam can actually speed up the rate at which you could move down the plateau and then get off the plateau.

Andrew Ng