deeplearning.ai

Optimization Algorithms

The problem of local optima

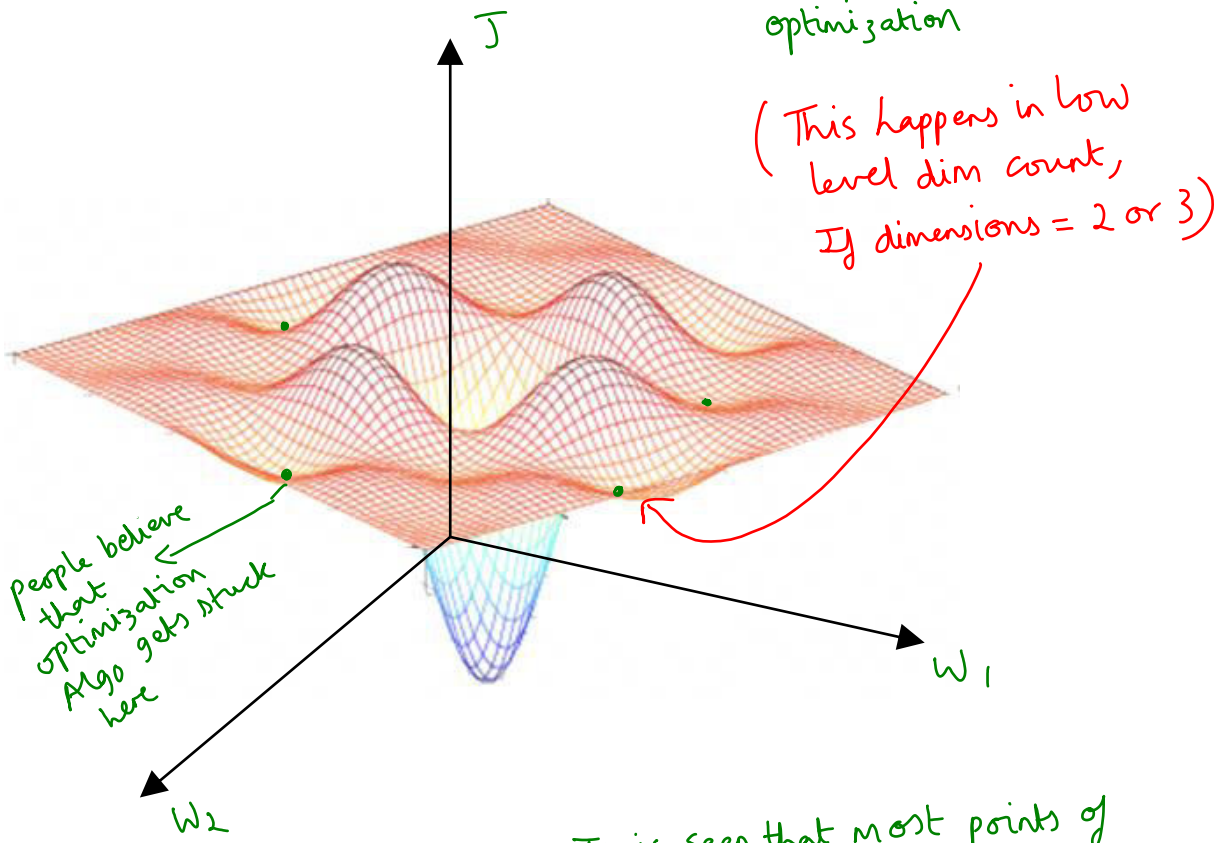
(People worry about the optimization Algo getting stuck in a bad local optima)

Local optima in neural networks

Saddle point = Horse saddle

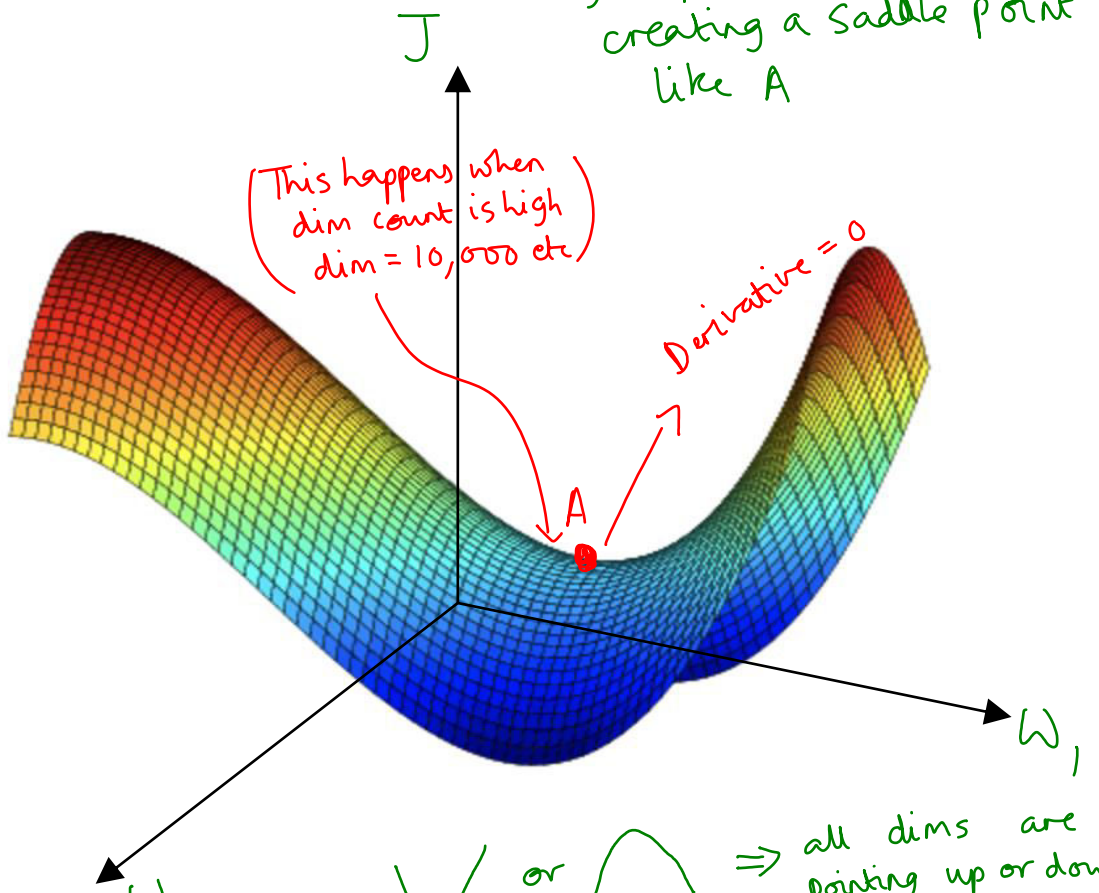
In most cases, some dims go up, some others go down creating a saddle point like A

Old view - wrong - doesn't happen in deep learning optimization



(This happens in low level dim count, If dimensions = 2 or 3)

People believe that optimization Algo gets stuck here



(This happens when dim count is high dim = 10,000 etc)

Derivative = 0

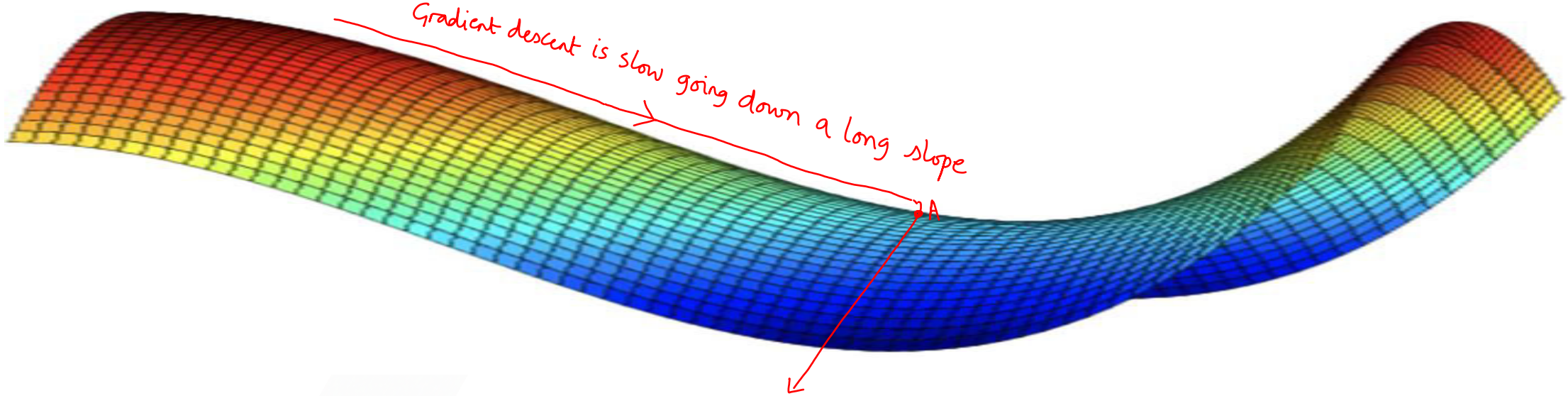
It is seen that most points of Zero gradients are not points on the left, rather they are "saddle points" - points on the right

* — why? In 20,000 dim space (NN are in high dim) when gradient is 0, the graph would have to look like convex or concave ⇒ all dims are pointing up or down ie, probability of that happening = tossing a coin 20,000 times w/ all tails $= 2^{-20,000} = 0$

Andrew Ng

Problem of plateaus

– The problem that does occur in deep learning (high dim) optimization is the problem of plateaus



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

– But RMS prop & momentum & Adam do a good job of optimizing for these cases

As long as J is defined over a high dim space