On Connected Sublevel Sets in Deep Learning

Quynh Nguyen

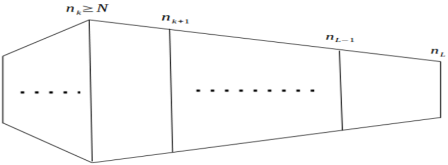
Department of Mathematics and Computer Science
Saarland University, Germany

- Objective: Understanding better the optimization landscape of deep neural networks

$$\boxed{\{\theta \in \Omega \mid \Phi(\theta) \leq \alpha\}}$$

See also Venturi-Bandeira-Bruna 2018, Safran-Shamir 2016, Nguyen-Mukkamala-Hein 2019

Our Main Result



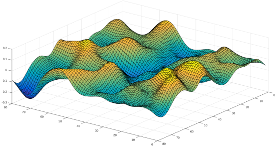
n_k : #hidden units at layer k N : #training samples L : #layers

Theorem (Informal)

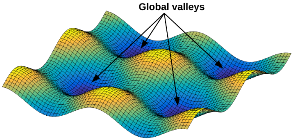
Consider training a **deep neural net** *with* **piecewise linear activation function** *and any convex loss (e.g. cross-entropy loss, square loss, non-smooth Hinge-loss). Then all the following hold:*

- *If the network has some layer k with $n_k \geq N$ and $n_{k+1} > \ldots > n_L$, then there exists a* **continuous descent path** *from any point to a global min.*
- *If $n_1 \geq 2N$ and $n_2 > \ldots > n_L$ then every sublevel set of the loss is connected and unbounded. As a result, there is always a* **continuous descent path** *from any point to a global minimum, and furthermore,* **all the global minima are connected** *within a unique & unbounded valley.*

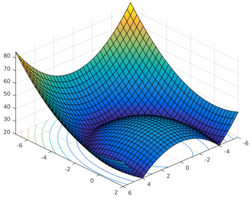
Summary



non-convex



$\exists k : n_k \geq N$



$n_1 \geq 2N$

POSTER #92