Global convergence of gradient descent

Global convergence of gradient descent

Theorem (Du et al. ’18, Allen-Zhu et al. ’18, Zou et al ’19) If the width of each layer is poly(n) where n is the number of data. Using random initialization with a particular scaling, gradient descent finds an approximate global minimum in polynomial time.

Gradient Flow: a Kernel Point of View

Gradient Flow: a Kernel Point of View

Gradient Flow: a Kernel Point of View

Gradient Flow: a Kernel Point of View

Gradient Flow: a Kernel Point of View

Gradient Flow: a Kernel Point of View