# Nonconvex Optimization for Deep Learning

Antonio Orvieto

1. Linear Regression GD on quadratic models: I, II
2. (a) Neural Network Basics Local Minima and Saddle Points Noisy Gradients: I, II
   (b) Linear Algebra Recap
3. (a) Initialization of Deep MLPs: I, II
4. (a) Nonconvex Gradient Descent Descent Lemma and Rates: I, II
   (b) Noisy Linear Models Deep Chains
5. (a) Nonconvex SGD Upper and Lower Bounds: I, II
   (b) Gradient Flow PL Condition
6. (a) Overparametrization Implicit Bias of SGD: I, II
   (b) Robbins-Monro Conditions
7. (a) Neural Tangent Kernel Lazy Training: I, II
   (b) Matrix Sensing
8. (a) Maximal Update Parametrization: I, II
   (b) Stochastic Matrices Exam-like questions
9. (a) Optimization Challenges in CNNs Batch Normalization and Skip Connections: I, II
   (b) Interpolated SGD Rates
10. (a) Optimization Challenges in Attention Layer Normalization and Rank Collapse: I, II
    (b) Residual Connection Mechanics
11. (a) Adaptive Methods Theory 1: I, II
    (b) Layer Normalization Mechanics
12. (a) Adaptive Methods Theory 2
    (b) Polyak Stepsize
13. Exam tips
14. Materials

September 4, 2025