

Learning rate

© 2019 Philipp Krähenbühl and Chao-Yuan Wu

Optimization algorithms

- Hyper parameters
 - Learning rate
 - Momentum
 - Batch size

What learning rate it use?

- Rule of thumb: Largest LR that trains
- Train for a few epochs and measure validation accuracy

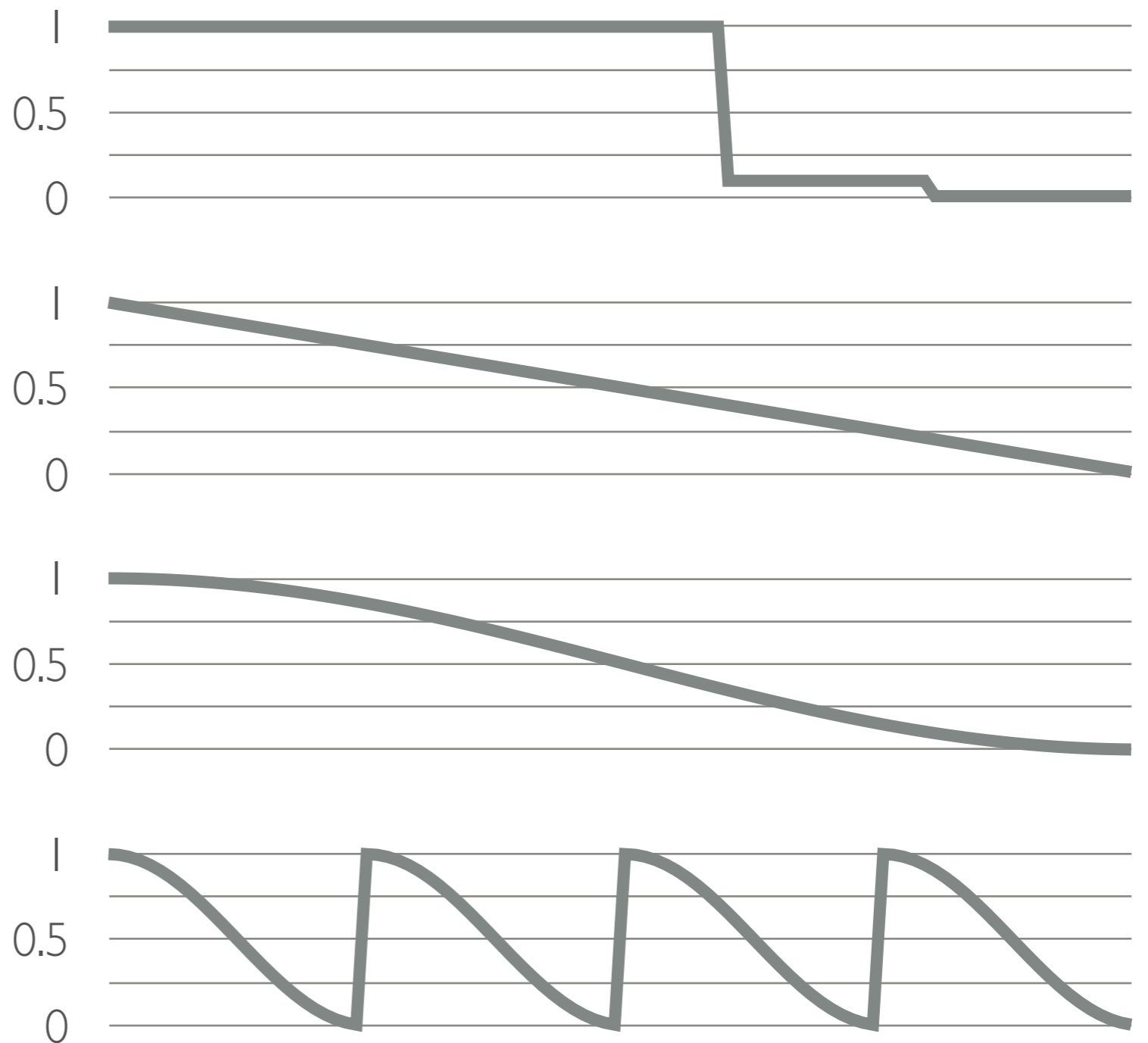
Learning rate vs batch size

- **Linear Scaling Rule:**
When the minibatch size is multiplied by k , multiply the learning rate by k .

Priya Goyal et al., "Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour", arXiv 2017

Learning rate schedules

- Step schedule
- Linear schedule
- Cosine schedule
- Cyclical schedules



Summary

- Use close to the largest LR that trains
- Step schedule