

### 0.0.1 Question 3d: Analyzing Learning Rates

In 1-2 sentences, describe what you notice about the path that  $\theta$  takes with a static learning rate vs. a decaying learning rate and for batch vs. stochastic gradient descent. Based on these comparisons, why do we often prefer SGD over batch GD in practice? In your answer, refer to the plots above.

Theta takes a much more variant and sharp path with static learning rate than with decaying learning rate. It also seems that decaying learning rate can more quickly decrease loss. We prefer SGD over batch GD because SGD is easier to read.

