Behavior of different optimizers for stochastic gradient descent. ([Source](#))

(Optional) Derivation

If you'd like to learn how to derive the equation that we use to approximate the gradient, please read the text below. Specifically, you'll learn how to derive

$$\nabla_\theta U(\theta) \approx \hat{g} = \frac{1}{m} \sum_{i=1}^{m} \sum_{t=0}^{H} \nabla_\theta \log \pi_\theta(a_t^{(i)} | s_t^{(i)}) R(\tau^{(i)})$$

This derivation is **optional** and can be safely skipped.

Likelihood Ratio Policy Gradient

We'll begin by exploring how to calculate the gradient $\nabla_\theta U(\theta)$. The calculation proceeds as follows: