

# 1 Problem 4: Experiments

## 1.1 Half Cheetah

Fig.1 shows that soft actor-critic (sac) can achieve very high return in half cheetah. Using the reparametrization trick can increase the return even more than the REINFORCE algorithm. This is because reparameterization tricks allows for better sampling which results in more accurate gradient estimation.

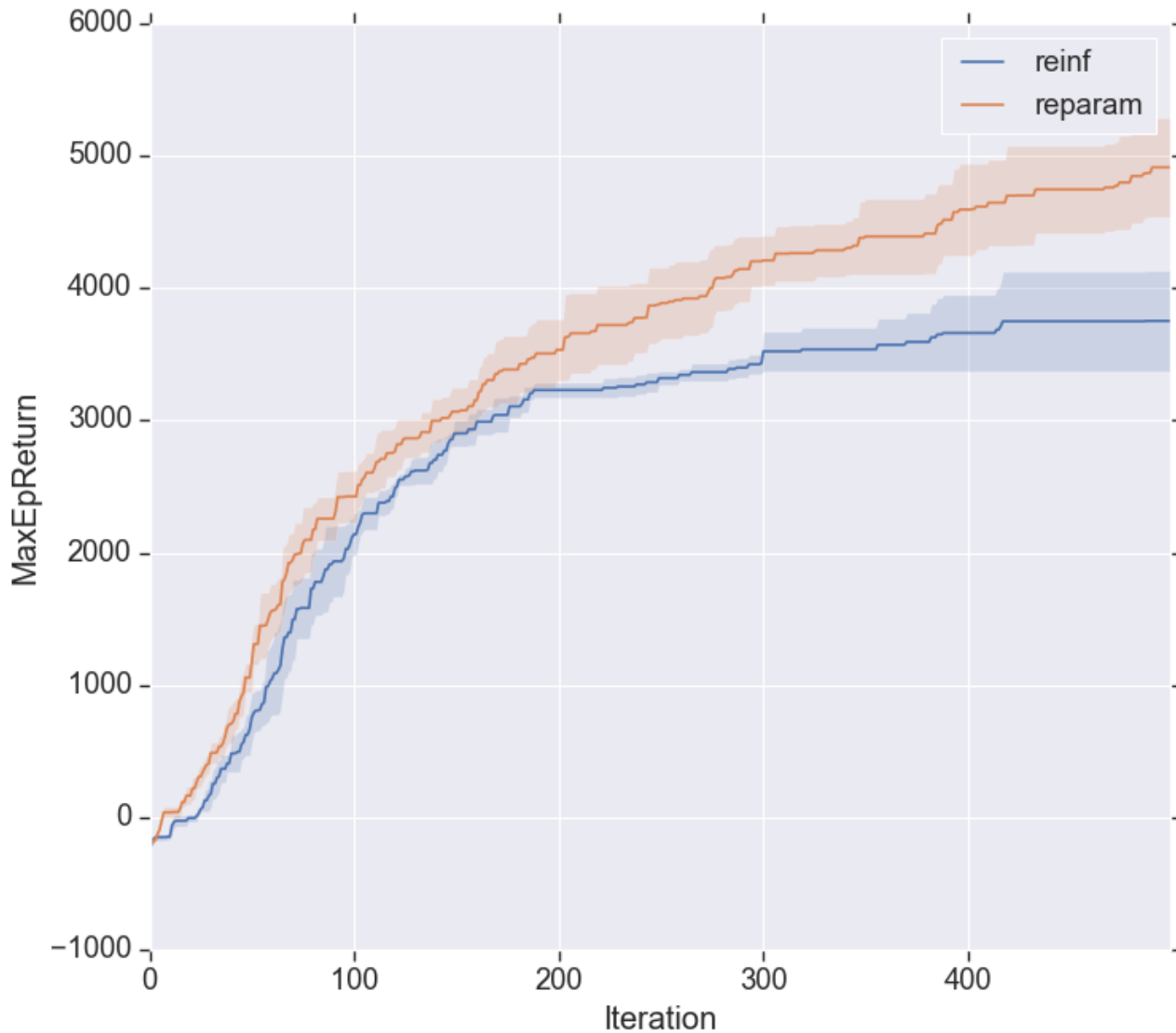


Figure 1: The performance of sac on half cheetah.

Blue: the maximum expected return of REINFORCE algorithm,

Orange: the maximum expected return using reparametrization trick.

## 1.2 Ant-v2

Fig.2 shows that soft actor-critic (sac) can achieve very high return in Ant-v2 as well. Using the two Q functions trick can increase the return a lot more. This is because double Q learning decorrelates sample points allowing for better Q function and Value function estimations.

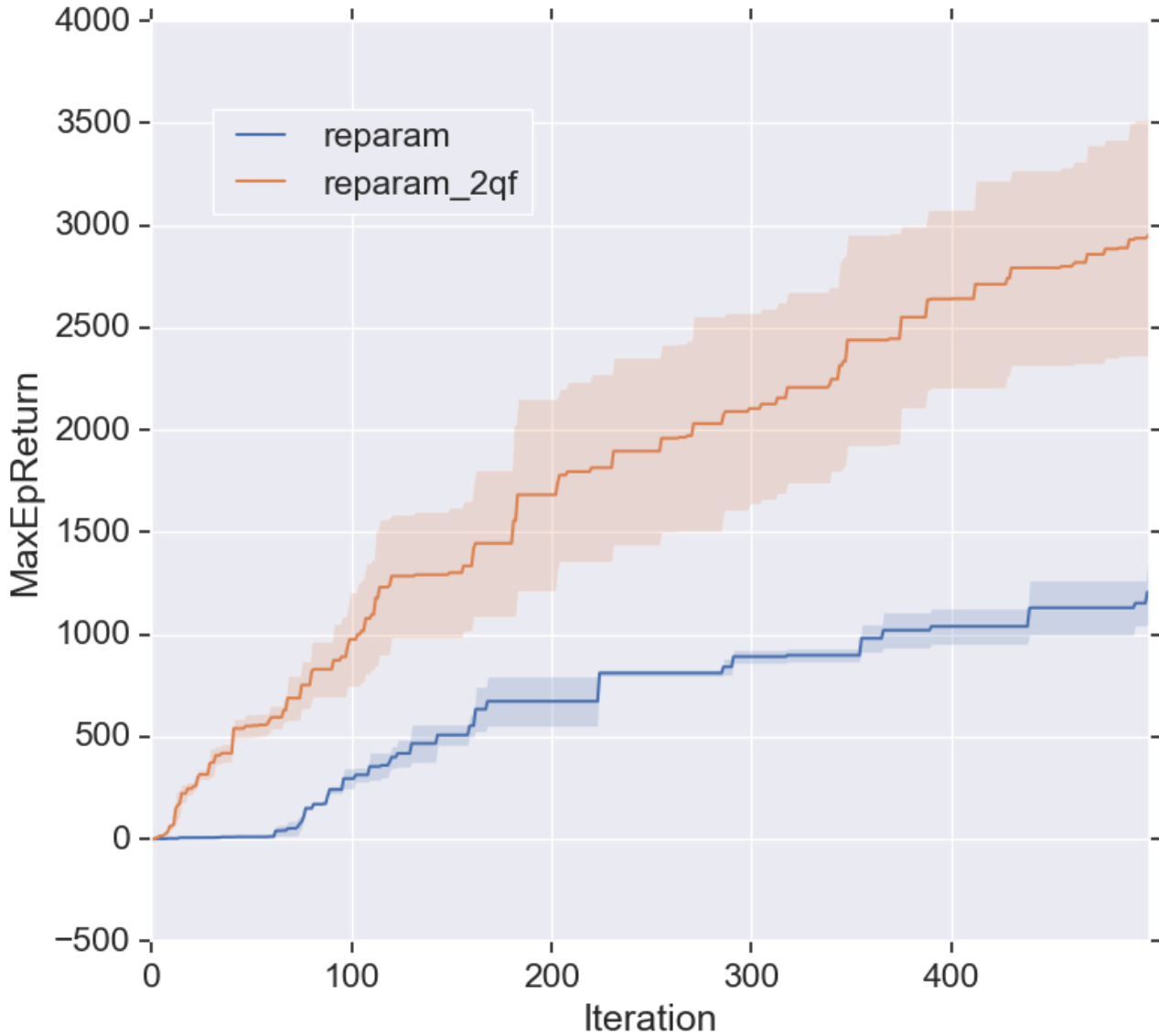


Figure 2: The performance of sac on Ant-v2.

**Blue:** the maximum expected return with a single Q function using reparametrization trick,  
**Orange:** the maximum expected return with two Q functions using reparametrization trick.