1 Model-based Policy Planning

1.1 In Action Space

1. The policy is used to propose the initial action sequences.
2. Instead of optimizing for the action sequence directly, CEM optimizes a sequence of noise to be added to the initial action sequences.

1.2 In Parameter Space

1. This approach is inspired by the observation that deep neural network is less likely to get stuck in sub-optimal local minima.

1.3 Policy Distillation Scheme

1. The agents iterate between interacting with the environments, and distilling the knowledge from planning trajectory into a policy network.
2. They consider different methods for distillation, including a L2 loss, GAN approach and averaging the optimized noise.

2 Experiments

1. A few interesting observations they have are:
2. The reward function surface is easier to optimize with when optimizing in the parameter space.
3. When optimizing in the parameter space, increasing the population size in CEM increases performance, while this is not true for optimization in action space.
4. Optimization in parameter space also leads to clear multi-modality in the solution obtained. This is less true for optimization in action space.

3 Questions

1. How should I interpret the dotted line in the plots? For example, in figure 1.
2. In what way can their noise injection in the parameter space be interpreted as stochastic policy network with reparameterization trick?
3. The performance on Ant has not reached SAC's performance at convergence. Can they match SAC's performance at convergence?
4. Where is to code to perform PCA?