RL: Policy Search

Gradient-free Optimization

Marius Lindauer

Leibniz Universität Hannover

Winter Term 2021

Policy optimization

- Policy based reinforcement learning is an optimization problem over θ
- $\rightsquigarrow$ Find policy parameters θ^* that maximize $V(s_0, \theta^*)$
- We can use gradient-free approaches (a.k.a. black-box optimization)
  - Hill climbing
  - Simplex / amoeba / Nelder Mead
  - Genetic algorithms
  - Cross-Entropy method
  - Covariance Matrix Adaptation (CMA)

Policy optimization

- Policy based reinforcement learning is an optimization problem over θ
- $\rightsquigarrow$ Find policy parameters θ^* that maximize $V(s_0, \theta^*)$
- We can use gradient-free approaches (a.k.a. black-box optimization)
  - Hill climbing
  - Simplex / amoeba / Nelder Mead
  - Genetic algorithms
  - Cross-Entropy method
  - Covariance Matrix Adaptation (CMA)
- gradient-free optimizers are (often) designed for
  - many function evaluations $\rightarrow$ possible in RL
  - parallel computation $\rightarrow$ possible in RL
  - a few to hundreds of dimensions $\rightarrow$ RL?

Policy optimization

- Policy based reinforcement learning is an optimization problem over θ
- $\rightsquigarrow$ Find policy parameters θ^* that maximize $V(s_0, \theta^*)$
- We can use gradient-free approaches (a.k.a. black-box optimization)
  - Hill climbing
  - Simplex / amoeba / Nelder Mead
  - Genetic algorithms
  - Cross-Entropy method
  - Covariance Matrix Adaptation (CMA)
- gradient-free optimizers are (often) designed for
  - many function evaluations $\rightarrow$ possible in RL
  - parallel computation $\rightarrow$ possible in RL
  - a few to hundreds of dimensions $\rightarrow$ RL?
- if we encode the policy π_θ as a DNN, we might have millions of dimensions (i.e., parameters in θ)