Meta-Gradient Reinforcement Learning [Xu et al. 2018], [Flennerhag et al. 2023]

Challenge:

- RL algorithms are sensible to return estimation
- How to compute return, i.e. how to set $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$?

Idea of Meta-Gradient RL:

- Learn a set of parameters η (e.g. $[\gamma, \lambda]$) online
- Based on *online cross-validation* to estimate performance [Sutton et al. 1992]
- Apply SGD to update η

Meta-Gradient Reinforcement Learning [Xu et al. 2018]

Algorithm: META-GRADIENT REINFORCEMENT LEARNING

Initialize policy parameters θ
Initialize hyperparameters (meta-parameters) η
for *each episode* **do**
  Collect rollout using current policy π_θ
  Update policy $\theta = \theta - \mu * \nabla_\theta J(\theta)$
  Compute meta-gradient ∇_η
  Update meta-prameters $\eta = \eta - \alpha * \nabla_\eta R$