Policy Continuation with Hindsight Inverse Dynamics

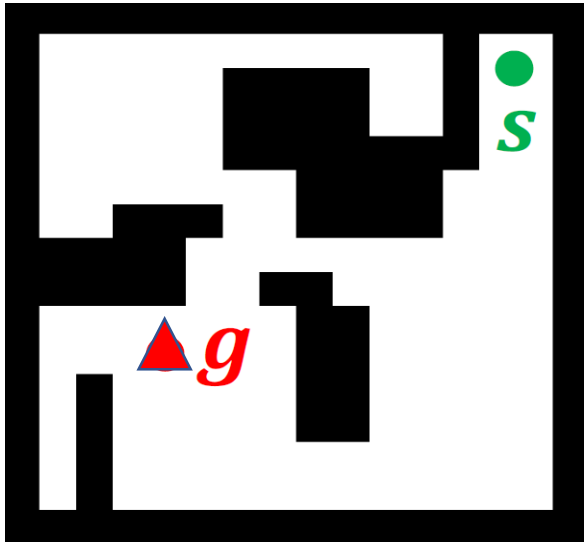
Hao Sun[1], Zhizhong Li[1], Xiaotong Liu[2], Dahua Lin[1], Bolei Zhou[1]

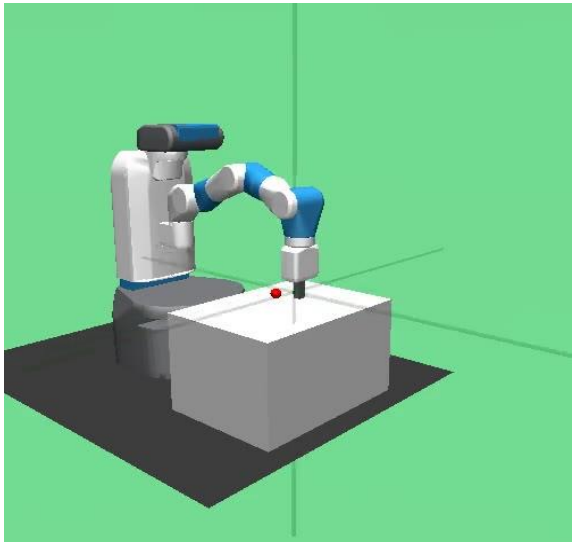
[1] The Chinese University of Hong Kong

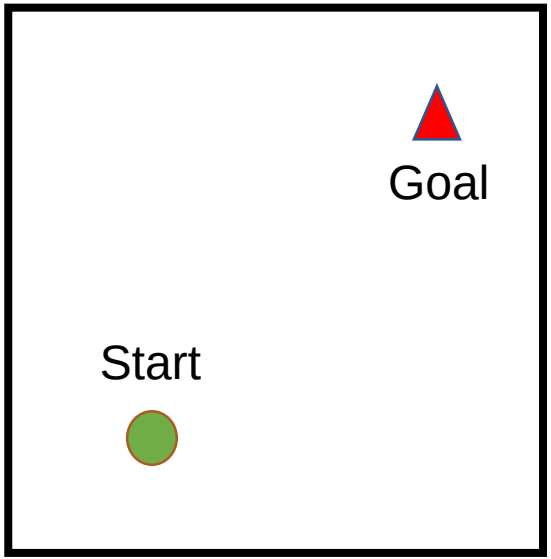
[2] Peking University

sh018@ie.cuhk.edu.hk

Goal-Oriented Reward Sparse Tasks

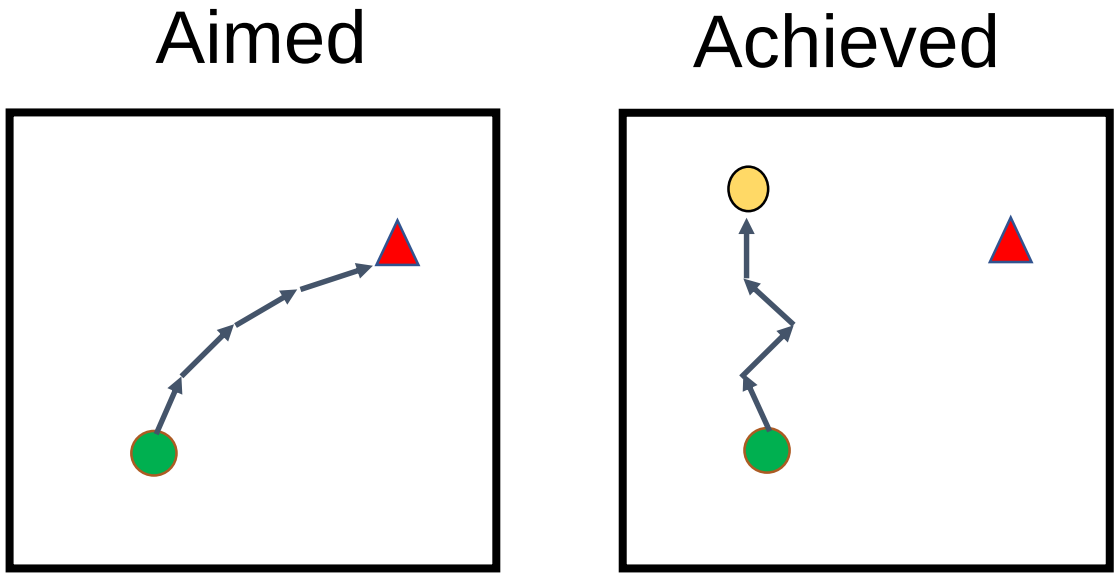






Inspirations from Human Learning

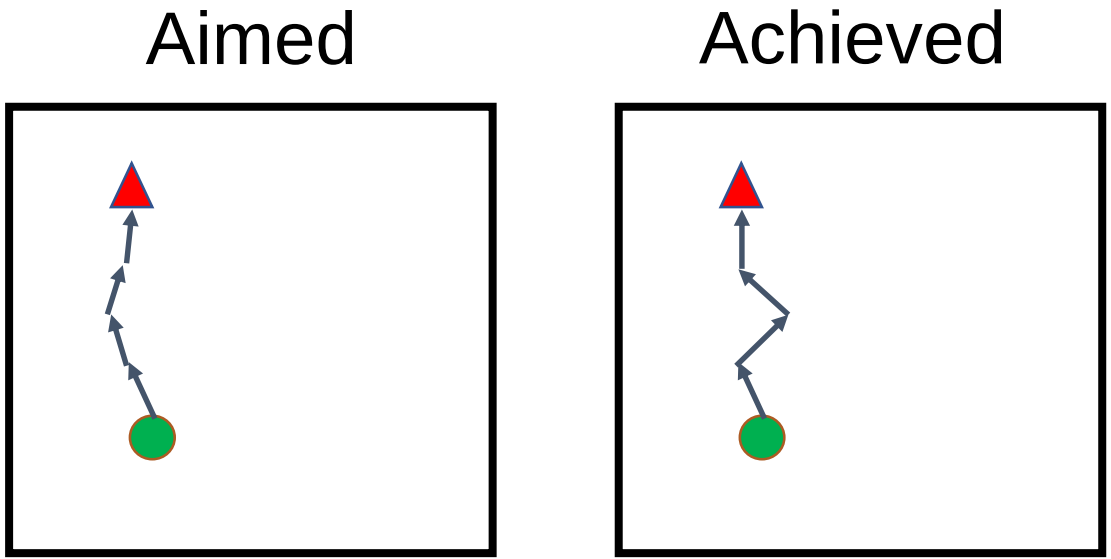
1. Learning from failures
[Hindsight Experience Replay, M Andrychowicz et al. 2017]



Inspirations from Human Learning

1. Learning from failures

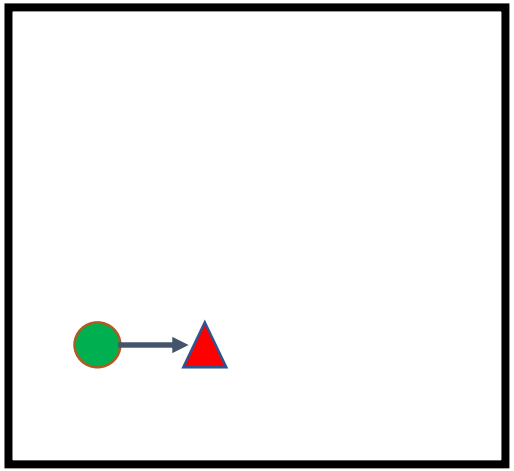
[Hindsight Experience Replay, M Andrychowicz et al. 2017]



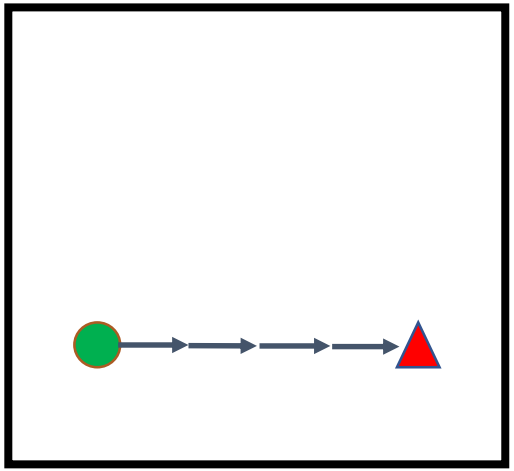
Inspirations from Human Learning

1. Learning from failures
2. Extrapolating Success

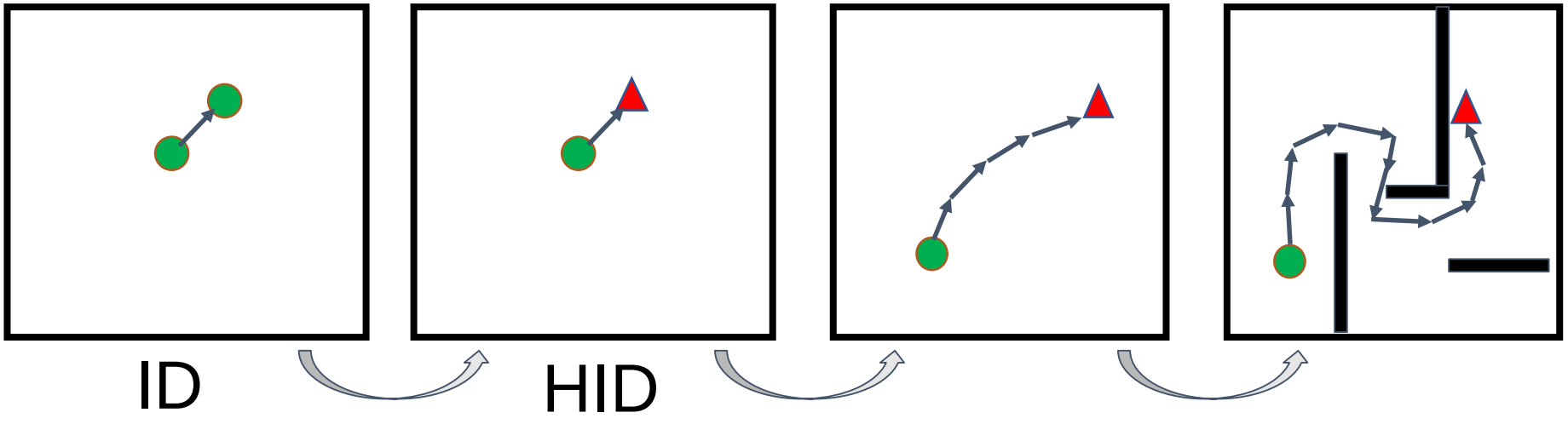
Learned



Extrapolate



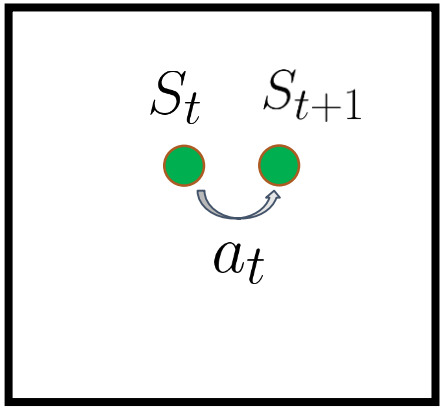
Our Proposed Method



1.Hindsight 2. Extrapolate 3. Policy Continuation

Equipe Inverse Dynamics with Hindsight

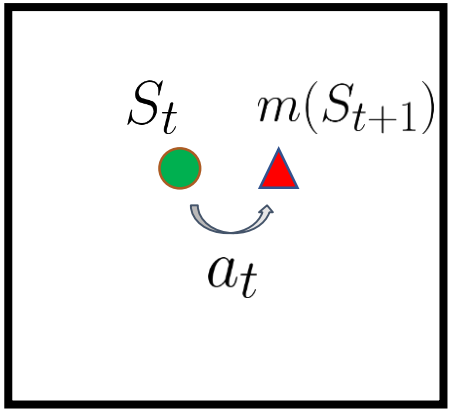
Inverse Dynamics:



State

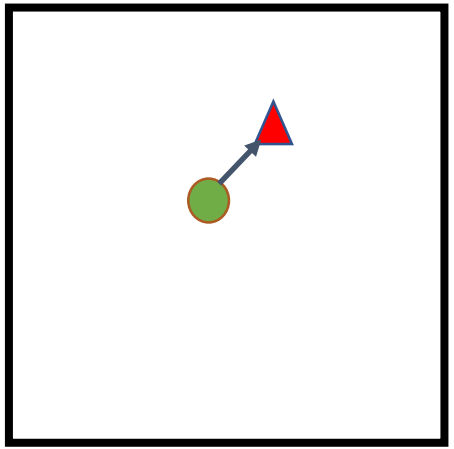
Goal

Hindsight Inverse Dynamics:

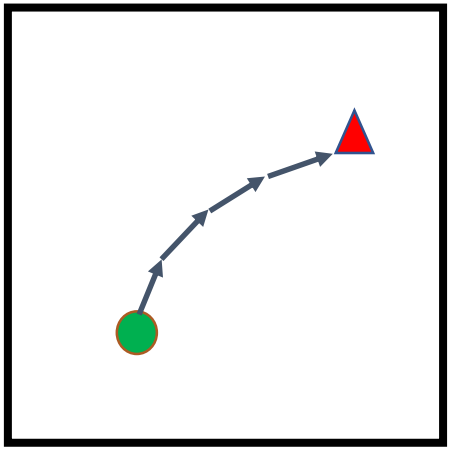


$$g = m(S_{goal})$$

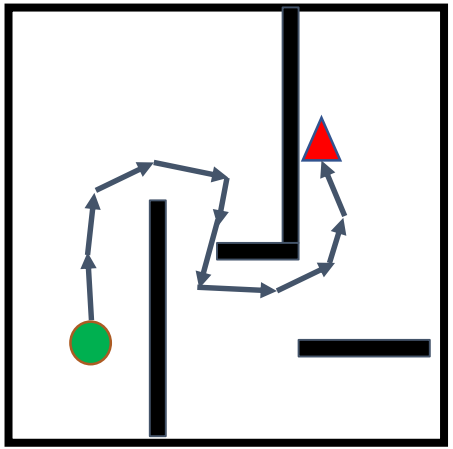
1-step HID Is Not Enough



1-step HID



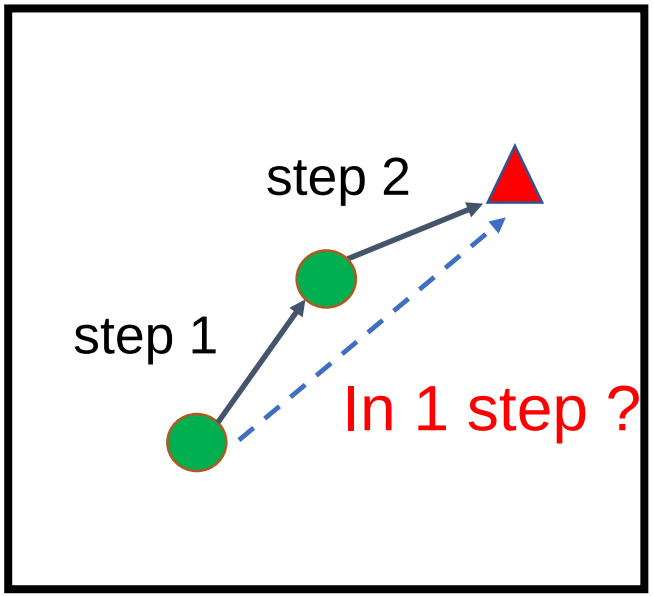
Linear Case



Non-linear Case

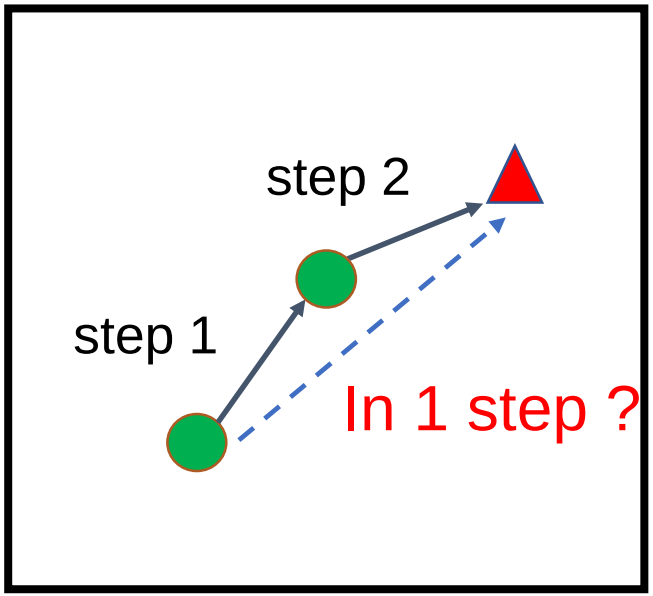
Multi-step Optimality?

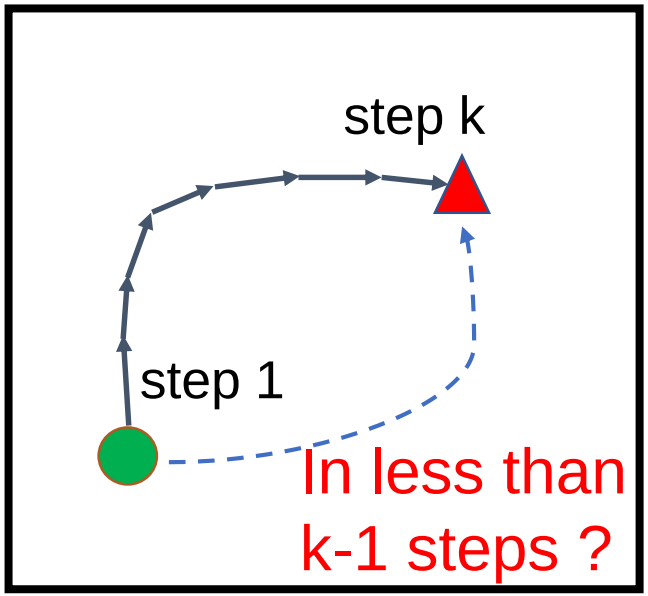
Policy Continuation: Test the optimality recursively

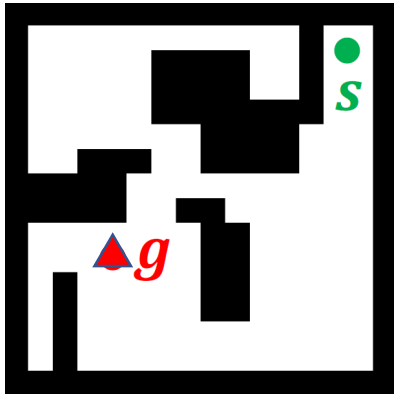


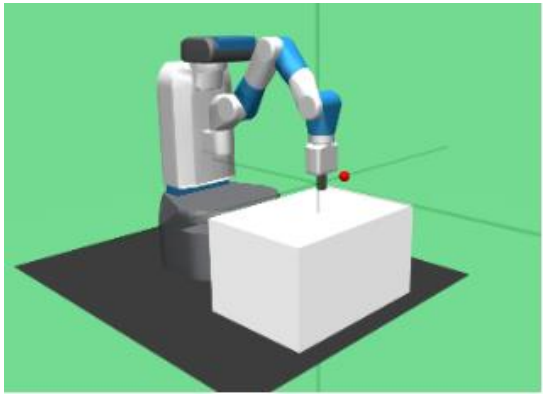
Multi-step Optimality?

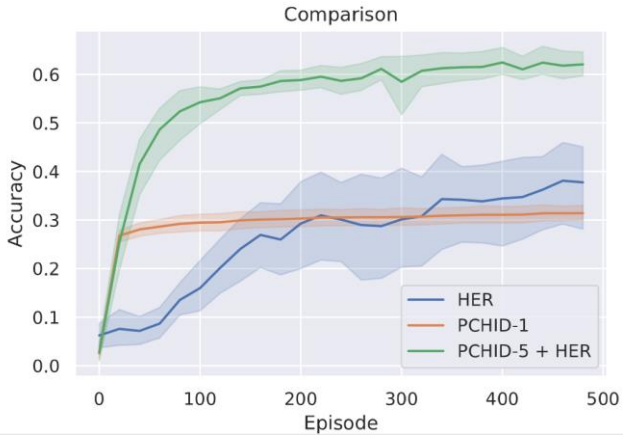
Policy Continuation: Test the optimality recursively

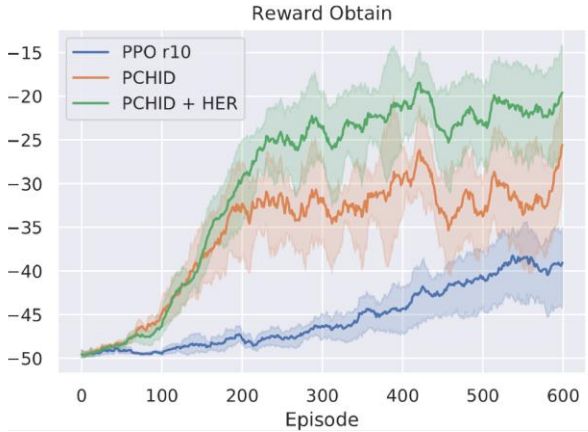












East Exhibition Hall B + C #194