Policy Continuation with Hindsight Inverse Dynamics

Hao Sun¹, Zhizhong Li¹, Xiaotong Liu², Dahua Lin¹, Bolei Zhou¹

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]

Aimed

Achieved
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:

\[ S_t \quad S_{t+1} \]

\[ a_t \]

- State
- Goal

Hindsight Inverse Dynamics:

\[ S_t \quad m(S_{t+1}) \]

\[ a_t \]

\[ g = m(S_{goal}) \]
1-step HID Is Not Enough

1-step HID

Linear Case

Non-linear Case
Multi-step Optimality?

Policy Continuation: Test the optimality recursively

In 1 step?
Multi-step Optimality?

Policy Continuation: Test the optimality recursively

In 1 step?

In less than \( k-1 \) steps?