LIDAR: Learning from Imperfect Demonstrations with Advantage Rectification

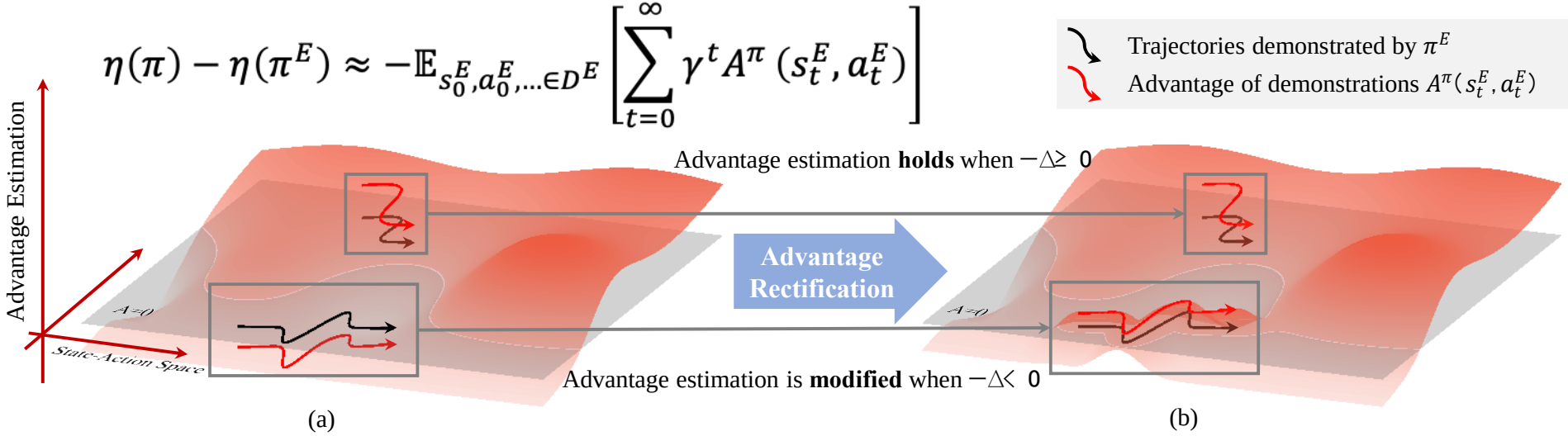
Xiaoqin ZHANG, Huimin MA, Xiong LUO, Jian YUAN

Frontiers of Computer Science, DOI: 10.1007/s11704-021-0147-9

Problems & Ideas

- Problems of learning from demonstrations
  - Imperfectness of the demonstrations
  - Effectiveness of learning and overfitting to flaws
- Ideas: learning from demonstrations using advantage rectification
  - Policy gap can be represented by advantage functions
  - Learning with a minimal assumption that the demonstrating policies have better performances than the current policy

$$\eta(\pi) - \eta(\pi^E) \approx -\mathbb{E}_{s_0^E, a_0^E, \ldots \in D^E}\left[\sum_{t=0}^{\infty} \gamma^t A^{\pi}\left(s_t^E, a_t^E\right)\right]$$



Policy gap representation (above) and the overview of Advantage Rectification (below)

Main experimental results

- **Experimental results of LIDAR on DDPG, TD3 and SAC**

DDPG:

TD3:

SAC: