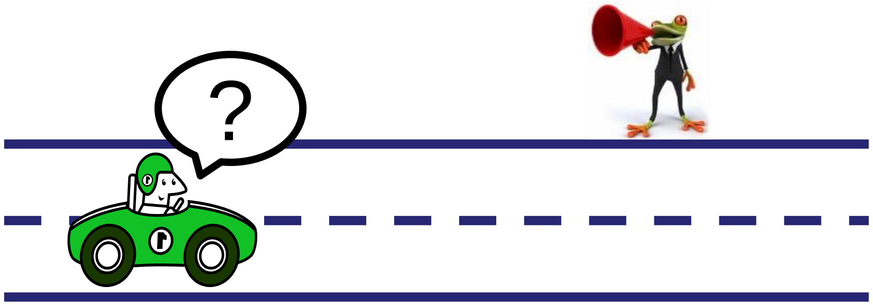
T-19

An Online Policy Gradient Algorithm for Continuous State and Action Markov Decision Processes with Bandit Feedback

Yao Ma (Tokyo Tech) and Masashi Sugiyama (UTokyo)

❑Motivating Example:

- Long Road Trip.
- The road condition may change frequently.
- Obtain the best driving strategy.



❑We proposed online policy gradient algorithm for continuous state and action online MDPs with sublinear regret.