Lipschitz Continuity Analysis for RA Value Function and Policy Conditioned on Parameters

Yichao Zhong

December 6, 2024

1 Problem Formulation

1.1 Dynamics Settings

The dynamics is defined by state $s \in \mathcal{S} \subset \mathbb{R}^{|s|}$ and action $a \in \mathcal{A} \subset \mathbb{R}^{|a|}$ and environmental physical parameters as $e \in \mathcal{E} \subset \mathbb{R}^{|e|}$: $s_{t+1} = s_t + f(s_t, a_t, e)$. For simplicity[1], in this paper, we denote e as the physical parameters, i.e., the mass of payload, the friction coefficient, the CoM shift, etc., which is assumed static within a trajectory in training sessions. The observations are from proprioceptive and exteroceptive sensors, denoted as $o = h(s)$ where h acts as the sensor mapping.

1.2 Goal Settings

Given local position and goals $\mathcal{T} \in \Gamma$, we learn a goal-conditioned reaching policy $\pi : \mathcal{O} \times \Gamma \to \mathcal{A}$ to maximize the expected return: $J(\pi) = \mathbb{E}_{\pi,\mathcal{T}}\left[\sum_{t=0}^{\infty} \gamma_{RL}^t r(s_t, a_t, \mathcal{T})\right]$, where $r(\cdot)$ is the reward at time t and γ_{RL} is the discount factor.

1.3 Safety Settings

First, we denote the system trajectory starting from state s while using control inputs from the policy π under environmental parameter e as $\xi_s^{\pi,e}(\cdot) : \mathbb{R} \to \mathcal{S}$. As in (1), we define several basic sets: The target set $\mathcal{T} \in \mathcal{S}$ which represents the area of the goal, the constraint set

[1]Here for simplicity, we assume the environment doesn't change and e as static, but in inference, e can be variant to time.

$\mathcal{K} \in \mathcal{S}$ which refers to the traversable areas for robots. And the failure set $\mathcal{F} = \mathcal{K}^C$, which is the complement of the constraint set and represents hazardous areas like obstacles.

Based on the basic sets, we can define the following sets in the context of the reachability theory. The safe set is defined as the set of states from which the robot can start and has a positive probability of rolling out a trajectory without failure, expressed as:

$$\omega^{\pi,e}(\mathcal{F}) := \{s \in \mathcal{S} \mid \forall \tau \geq 0, \xi_s^{\pi,e}(\tau) \notin \mathcal{F}\}$$

. The backward reachable set is the collection of states from which the robot has a positive probability of reaching the target:

$$\mathcal{R}^{\pi,e}(\mathcal{T}) := \{s \in \mathcal{S} \mid \exists \tau \geq 0, \xi_s^{\pi,e}(\tau) \in \mathcal{T}\}$$

. And the reach-avoid set combines the safe set and the backward reachable set:

$$\mathcal{RA}^{\pi,e}(\mathcal{T}, \mathcal{F}) := \{s \in \mathcal{S} \mid \exists \tau \geq 0, \xi_s^{\pi,e}(\tau) \in \mathcal{T} \wedge \forall \tau \geq 0, \xi_s^{\pi,e}(\tau) \notin \mathcal{F}\}$$

, which represents states from which the robot can reach the target while avoiding failure.

1.4 Reach-Avoid Value and Time-Discounted Reach-Avoid Bellman Equation (DRABE)

Identical to the vanilla reach-avoid analysis (1), we define two Lipschitz-continuous functions $l(\cdot), \zeta(\cdot) : \mathcal{S} \to \mathcal{R}$ which satisfy

$$\begin{cases} l(s) \leq 0 \iff s \in \mathcal{T} \\ \zeta(s) > 0 \iff s \in \mathcal{F} \end{cases}$$

to illustrate if it has reached without failure. Note that this function is only dependent on the state s and is environment-agnostic. Then we define the reach-avoid value function which satisfies $V_{RA}^{\pi}(s, e) \leq 0 \iff s \in \mathcal{RA}^{\pi,e}(\mathcal{T}, \mathcal{F})$:

$$V_{RA}^{\pi}(s, e) = \min_{\tau \in \{0,1,\dots\}} \max \{l(\xi_s^{\pi,e}(\tau), \max_{\kappa \in \{0,1,\dots,\tau\}} \zeta(\xi_s^{\pi,e}(\kappa)))\} \tag{1}$$

And as introduced in (2), Discounted Reach-Avoid Bellman Equation (DRABE) could make the value function iteration a contraction mapping, which guarantees the convergence of the value iteration through DRABE:

$$B_\gamma[V_{RA_\gamma}^{\pi}](s_t, e) = (1-\gamma) \max \{l(s_t), \zeta(s_t)\} + \gamma \max\{\min\{V_{RA_\gamma}^{\pi}(s_{t+1}, e), l(s_t)\}, \zeta(s_t)\} \tag{2}$$

(2) also mathematically proves the DRABE operator $B_\gamma[\cdot]$ is a contraction mapping, and having V_γ^{π} conditioned on a static physical parameter e doesn't alter the convergence.

2 Lipschitz Continuity of V_γ^π

Since (3) proves that $V_\gamma(s)$ is Lipschitz-continuous to s, we extend the proof to prove that $V_\gamma^\pi(s,e)$ is Lipschitz-continuous to both s and e. For s, introducing a static e trivially doesn't alter the Lipschitz continuity of V_γ^π to s.

So in this section we try to prove the Lipschitz continuity of the value function $V_\gamma^\pi(s,e)$ to environment factor e in Theorem 2.1.

Theorem 2.1. *The Learned Value Function $V_\gamma^\pi(s,e)$ Possesses Lipschitz Continuity w.r.t. Environmental Factors e under the following conditions:*

- *The functions $l(s)$ and $\zeta(s)$ are defined as L_l- and L_ζ-Lipschitz continuous functions of the state s.*
- *Given a specific policy π, the transition dynamics defined as $f_\pi(s,e) := f(s,\pi(s,e),e)$ are L_{f_π}-Lipschitz continuous w.r.t. the tuple (s,e).*
  L_{f_π} is defined as, for any states $s_1, s_2 \in \mathcal{S}$ and environmental factors $e_1, e_2 \in \mathcal{E}$

  $$\|f(s_1,\pi(s_1,e_1),e_1) - f(s_2,\pi(s_2,e_2),e_2)\| \le L_{f_\pi}(\|e_1-e_2\| + \|s_1-s_2\|),$$

  where L_{f_π} is conditioned upon the policy π.
- *And necessarily, $\gamma(1+L_{f_\pi}) < 1$.*
  Then the Lipschitz constant for V_γ^π is bounded by

  $$L_V \le C\max\{L_l, L_\zeta\}$$

  where C is a constant less than 1, which we will deduct in the proof.

Proof. Here we note V as for V_γ^π because there's only one value function. By definition, we got

$$V(s,e) := V_\gamma^\pi(s,e) = \min_{\tau\in\{0,1,\dots\}} \max\{\gamma^\tau l(\xi_s^{\pi,e}(\tau), \max_{\kappa\in\{0,1,\dots,\tau\}} \gamma^\kappa \zeta(\xi_s^{\pi,e}(\kappa)))\}$$

And define $P(s,e,t)$ as payoff at timestep t:

$$P(s,e,t) := \max\{\gamma^t l(\xi_s^{\pi,e}(t), \max_{\kappa\in\{0,1,\dots,t\}} \gamma^\kappa \zeta(\xi_s^{\pi,e}(\kappa)))\}$$

For all $e_1, e_2 \in \mathcal{E}$ and $s \in \mathcal{S}$, and $\theta > 0$ we have:

$$\begin{cases} \forall t \in \mathcal{R}, P(s,e_1,t) > V(s,e_1) - \theta \\ \exists \bar{t} \in \mathcal{R}, P(s,e_2,\bar{t}) < V(s,e_2) + \theta \end{cases} \tag{3}$$

Combining the two inequations:

$$\begin{aligned} &V(s,e_1) - V(s,e_2) - 2\theta < P(s,e_1,\bar{t}) - P(s,e_2,\bar{t}) \\ &\leq \max\{\gamma^{\bar{t}} L_l \|\xi_s^{\pi,e_1}(\bar{t}) - \xi_s^{\pi,e_2}(\bar{t})\|, \max_{\kappa\in\{0,1,\ldots,\bar{t}\}} \gamma^{\kappa} L_\zeta \|\xi_s^{\pi,e_2}(\kappa) - \xi_s^{\pi,e_2}(\kappa)\|\} \end{aligned}$$

We use $\Delta\xi(t) := \xi_s^{\pi,e_1}(t) - \xi_s^{\pi,e_2}(t)$, and by definition of f 's Lipschitz continuity, we have

$$\begin{aligned} \Delta\xi(\bar{t}) &\leq (1+L_f)\|\Delta\xi(\bar{t}-1)\| + L_f(\|p_1 - p_2\|) \\ &\leq \cdots = ((1+L_{f_\pi})^{\bar{t}} - 1)\|e_1 - e_2\| \end{aligned}$$

Because this holds for some $\bar{t}$, so it must be less than the maximum for all $\bar{t}$. Thus we got

$$V(s,e_1) - V(s,e_2) \leq 2\theta + \max\{L_l, L_\zeta\} \max_{\bar{t}} \{ \max_{t\in\{0,1,\ldots,\bar{t}\}} \gamma^t((1+L_{f_\pi})^t - 1)\}\|e_1 - e_2\|$$

As θ is an arbitrary variable, we can set it to infinitesimal. To guarantee the Lipschitz continuity of V to e, it must be assured that $\gamma(1+L_{f_\pi}) \leq 1$ which necessarily holds that the Lipschitz constant is finitely bounded. Then we got the Lipschitz constant for the Value Funtion V to environmental factor e:

$$L_V = \max\{L_l, L_\zeta\} \max_{t=0,1,\ldots,T} \gamma^t((1+L_{f_\pi})^t - 1) \ ,$$

where T denotes the maximum time steps for a system trajectory. Assuming $T \to \infty$, By calculating the maximum point of t in the right part, we got the upper bound of L_V:

$$L_V = \max\{L_\zeta, L_l\} \cdot L_{f_\pi} \gamma^{t^*} \frac{\log(1+L_{f_\pi})}{-\log(\gamma(1+L_{f_\pi}))} \ , \tag{4}$$

where

$$t^* := \frac{\log(\frac{\log(\gamma)}{\log(\gamma(1+L_{f_\pi}))})}{\log(1+L_{f_\pi})} \ .$$

So we've got $V(s,e)$ is L_V-Lipschitz-continuous to e. □

Following the proof, we can observe that L_V is also bounded by $\max\{L_l, L_\zeta\}$ because the exponential term $\gamma^t((1+L_{f_\pi})^t - 1)$ should be less than 1.

The assumption $\gamma(1+L_{f_\pi}) < 1$ also gives a constraint that the dynamics with respects to the policy π shouldn't be too sensitive to environment factor e, i.e. π should be a robust policy.

References

[1] Bansal, S., Chen, M., Herbert, S., and Tomlin, C. J. Hamilton-jacobi reachability: A brief overview and recent advances, 2017.

[2] Hsu*, K.-C., Rubies-Royo*, V., Tomlin, C., and Fisac, J. Safety and liveness guarantees through reach-avoid reinforcement learning. In *Robotics: Science and Systems XVII* (July 2021), RSS2021, Robotics: Science and Systems Foundation.

[3] Li, J., Lee, D., Sojoudi, S., and Tomlin, C. J. Infinite-horizon reach-avoid zero-sum games via deep reinforcement learning, 2024.