Lipschitz Continuity in Model-based Reinforcement Learning

Kavosh Asadi*, Dipendra Misra*, Michael L. Littman

* denotes equal contribution
Model-based RL

Model learning:

\[ T(s' \mid s, a) \approx \hat{T}(s' \mid s, a) \]
\[ R(s, a) \approx \hat{R}(s, a) \]

Planning:

...
Compounding Error

[Talvitie 2014, Venkatraman et al. 2015]

- happens when models are imperfect, which is almost always true
  - estimation error or partial observability
  - agnostic setting

credit to Matt Cooper for the video
github.com/dyelax
Main Takeaway

Lipschitz continuity plays a key role in compounding errors and more generally in the theory of model-based RL.

Given two metric spaces $(M_1, d_1)$ and $(M_2, d_2)$, a function $f : M_1 \mapsto M_2$ is Lipschitz if the Lipschitz constant defined below is finite:

$$K_{d_1, d_2}(f) := \sup_{s_1 \in M_1, s_2 \in M_1} \frac{d_2(f(s_1), f(s_2))}{d_1(s_1, s_2)}$$
Wasserstein Metric

in stochastic domains, we need to quantify difference between two distributions

\[ W(\mu_1, \mu_2) := \inf_{j \in \Lambda} \int \int j(s_1, s_2) d(s_1, s_2) ds_2 ds_1 \]

[Villani, 2008]
Three Theorems

- multi-step prediction error
- value function estimation error
- Lipschitz continuity of value function
Multi-step Prediction Error

assume a $\Delta$ accurate model:

$$\forall s \ \forall a \quad W(\hat{T}(\cdot \mid s, a), T(\cdot \mid s, a)) \leq \Delta$$

given a $\Delta$ accurate model with a Lipschitz constant $K(\hat{T})$ and a true model with Lipschitz constant $K(T)$ and a state distribution $\mu(s)$:

$$\delta(n) := W(\hat{T}^n(\cdot \mid \mu), T^n(\cdot \mid \mu)) \leq \Delta \sum_{i=0}^{n-1} (k)^i$$

$\delta$ : error  
$n$ : prediction horizon  
$k$ : $\min (K(T), K(\hat{T}))$
how inaccurate can the value function be?

\[ \forall s \quad |V_T(s) - V_{\hat{T}}(s)| \leq \frac{\gamma K(R)\Delta}{(1 - \gamma)(1 - \gamma k)} \]

\[ k : \min (K(T), K(\hat{T})) \]

\[ K(R) : \text{Lipschitz constant of reward} \]
Lipschitz Continuity of Value Function

- Generalized VI [Littman and Szepesvári, 96]:
  - repeat until convergence:
    \[
    Q(s, a) \leftarrow R(s, a) + \gamma \int T(s' \mid s, a) f(Q(s', \cdot)) \, ds'
    \]
  - value function is Lipschitz in every iteration (including the fixed point)
    \[
    K(Q) \leq \frac{K(R)}{1 - \gamma K(T)}
    \]
  - one implication: value-aware model learning [Farahmand et al, 2017] is equivalent to Wasserstein (will appear in PGMRL workshop later in the conference)
Controlling Lipschitz Constant with Neural Nets

for each layer, ensure the weights are in a desired norm ball:

Lipschitz constant of entire net is bounded by multiplication of Lipschitz constant of layers
Is Controlling the Lipschitz Constant of Transition Models Useful?

- Cartpole (left) and Pendulum (right)

- learn a model offline using random samples

- perform policy gradient using the model

- test the policy in the environment

- improved reward (higher is better) by an intermediate Lipschitz value

more experiments (including on stochastic domains) in the paper
Contributions:

- key role of Lipschitz constant in model-based RL:
  - compounding error
  - value function estimation error
  - Lipschitz continuity of value function

- learning stochastic models using EM (skipped, details in the paper)

- quantifying Lipschitz constant of neural nets (skipped, details in the paper)

- model regularization by controlling the Lipschitz constant

- usefulness of Wasserstein for model-based RL (skipped, details in the paper)

Questions?
References:


* Villani, "Optimal Transport, Old and New", 2014

* Talvitie, "Model Regularization for Stable Sample Rollouts", 2014

* Venkatraman, Hebert, and Bagnell, "Improving Multi-Step Prediction of Learned Time Series Models", 2015