
Modeling, Estimation and Control Conference 2021 (Austin, TX, October 2021)

清華大學 Tsinghua University

Carnegie Mellon University

# Joint Synthesis of Safe Control Policies and Safety Certificates using Constrained Reinforcement Learning

## MECC workshop on safe control and learning under uncertainty

Haitong Ma[1], Changliu Liu[2], Shengbo Eben Li[1], Sifa Zheng[1], Jianyu Chen[1]

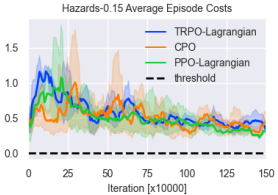
[1]Tsinghua University, [2]Carnegie Mellon University

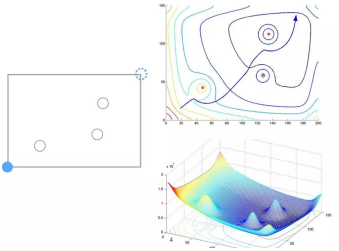
October 24, 2021

# Outline

# Why do we need safety certificates?

- Learn solidly safe policy in RL:



$$J_c(\pi) = \mathbb{E}_{\tau\sim\pi}\left\{ \sum_{t=0}^{\infty/T} \gamma_c^t \mathbf{1}(\text{danger}) \right\} < 0$$

CMDP constraints: posterior cost function, not for zero-violation

Energy-based safety certificate: prior modeling of safety

*A. Ray, J. Achiam, and D. Amodei, "Benchmarking safe exploration in deep reinforcement learning,"*

# Energy-Function-Based Safety Certificates

- $\phi$ energy function / safety index, high for unsafe states.
- Safety defined by a sub-level set: $\mathcal{S}_s = \{s|\phi_0(s) \leq 0\}$
- from $\phi_0$ to $\phi$: high order derivative for inevitably unsafe states

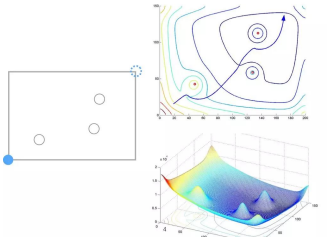
$$\phi = \phi_0 + k_1\phi_0^{'} + \cdots + k_n\phi_0^{(n)} \quad (1)$$

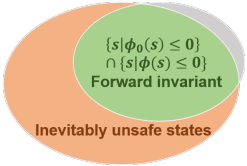
- Control safe set with energy function / safety index $\phi$

$$\mathcal{U}_S(s) := \{a \in \mathcal{A} \mid \phi(s') < \max\{\phi(s) - \eta, 0\}\} \quad (2)$$

- Valid safety certificates:

$$\mathcal{U}_S^D(s) \neq \emptyset, \forall s \in \mathcal{S} \quad (3)$$







*C. Liu and M. Tomizuka, "Control in a safe set: Addressing safety inhuman-robot interactions,"*

*T. Wei and C. Liu, "Safe control algorithms using energy functions: Auni ed framework, benchmark, and new directions,"*

# Why do we need joint synthesis?

**Challenging in complex environments with unknown models**

Current studies assume either one exists to learn the other

- *Learning valid safety certificates*: Synthesis safety certificate with/guided by known controllers (model-based, learning-based)[1, 6, 7, 5, 4];
- *Safe control for system with unknown dynamics*: Learn safe control with valid safety certificates [9, 2, 8, 3]


What if neither controller and valid certificate exists?

# Problem Formulation: Learning Safety Certificates

- Synthesis target: validness / feasibility

$$\mathcal{U}_S^D(s) := \{a \in \mathcal{A} \mid \phi(s') < \max\{\phi(s) - \eta, 0\}\} \neq \emptyset, \forall s \in \mathcal{S} \tag{4}$$

- Safety index synthesis optimization
  minimizing inevitable energy increase / violation of inequality

$$\min_{\xi} J(\phi) = \min_{\xi} \inf_{\pi} \mathbb{E}_s \Big\{ \left[\phi(s') - \max\{\phi(s) - \eta, 0\}\right]^+ \Big\} \tag{5}$$

- tunable variables $\xi = [\sigma, n, k]$

$$\phi(s) = \sigma + d_{\min}^n - d^n - k\dot{d} \tag{6}$$

*W. Zhao, T. He, and C. Liu, "Model-free safe control for zero-violation reinforcement learning,"*

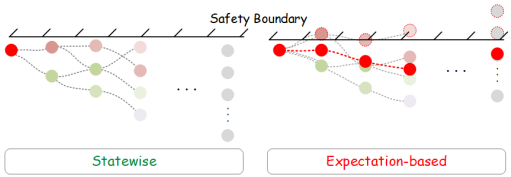
# Problem Formulation: Learning Safe Policies with Constrained RL

## Difficulty

Infinite **state-dependent constraints** on continuous state space?

RL with constrol safe set constraints

$$\max_{\pi} J(\pi) = \mathbb{E}_{\tau\sim\pi}\Big\{\sum_{t=0}^{\infty}\gamma^t r_t\Big\} = \mathbb{E}_s\big\{v^\pi(s)\big\}$$
$$\text{s.t. } \phi\left(s'\right) < \max\left\{\phi\left(s\right)-\eta, 0\right\}, \forall s \in \mathcal{S}$$




## A computable Lagrange function with neural multipliers

$$\mathcal{L}(\pi,\lambda) = \mathbb{E}_s\big\{-v^\pi(s) + \lambda(s)\big(\phi\left(s'\right) - \max\left\{\phi\left(s\right)-\eta, 0\right\}\big)\big\} \tag{7}$$

*H. Ma, Y. Guan, S. E. Li, X. Zhang, S. Zheng, and J. Chen, "Feasible actor-critic: Constrained reinforcement learning for ensuring statewisesafety,"*

# Problem Formulation

- Constrained RL optimization

$$\max_{\lambda} \min_{\pi} \mathcal{L}(\pi, \lambda) = \max_{\lambda} \min_{\pi} \mathbb{E}_s \big\{ -v^{\pi}(s) + \lambda(s)\big(\phi(s') - \max\{\phi(s) - \eta_D, 0\}\big) \big\} \tag{8}$$

- Safety index synthesis optimization
  minimizing inevitable energy increase / violation of inequality

$$\min_{\phi} J(\phi) = \min_{\phi} \inf_{\pi} \mathbb{E}_s \Big\{ \big[\phi(s') - \max\{\phi(s) - \eta, 0\}\big]^+ \Big\} \tag{9}$$

Difficulty: Two separate optimization?

# Joint Adversarial Optimization

- Recall the KKT condition for $(\pi^*, \lambda^*)$ for given state-dependent constraints
  If inequality constraints hold at $s$:

$$\lambda^*(s)(\phi(s') - \max\{\phi(s) - \eta_D, 0\}\big|_{\pi*}) = 0 \tag{10}$$

  else $\lambda(s) \to +\infty$.

$$\mathcal{L}(\pi^*, \lambda^*, \phi) = \mathbb{E}_s\{ \underbrace{-v^{\pi^*}(s)}_{\text{irrelevant term } \Delta} + \underbrace{\lambda^*(s)(\phi(s') - \max\{\phi(s) - \eta, 0\})|_{\pi=\pi^*}}_{=0 \text{ if inequality holds}} \} \tag{11}$$

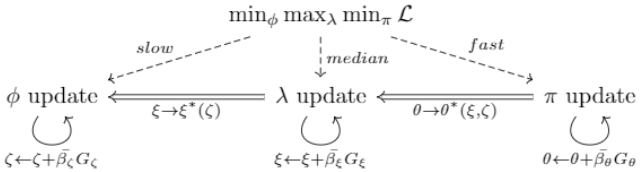
- Clip the infinite multiplier by $\lambda_{\max}$

$$\mathcal{L}(\pi^*, \lambda^*, \phi) = \lambda_{\max} J(\phi) + \Delta \to \arg\min J(\phi) = \arg\min \mathcal{L}'(\pi^*, \lambda^*, \phi)$$

# Joint Adversarial Optimization

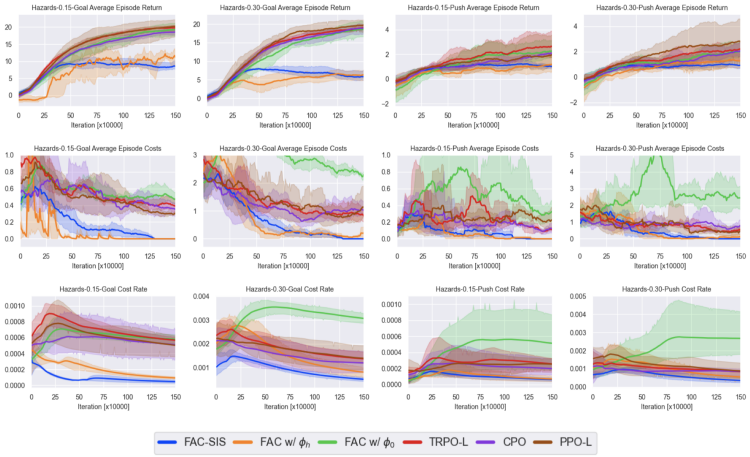
- Multi-scale adversarial training with proof, converging to optima of $\phi^*$ & $\pi^*$.

$$\min_{\phi} \max_{\lambda} \min_{\pi} \mathcal{L}(\pi, \lambda, \phi) \tag{12}$$



# Results - Safe Control

- FAC-SIS (proposed) achieves zer-violation in all environments ($\phi_h$ is a handcrafted safety certificate)



# Results - Feasibility Verification of Safety Certificate

- All transactions satisfy the safe action constraints, nearly all sampled states are feasible.

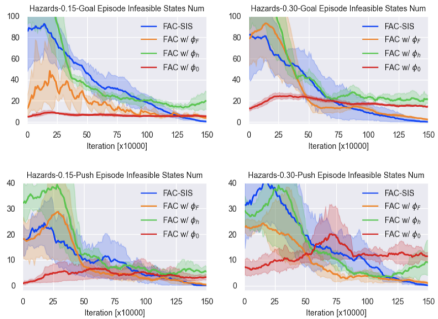



Figure 3: Average episodic number of violations of safe action constraint (2). A valid safety index and its corresponding safe control policy should have zero violation performance.

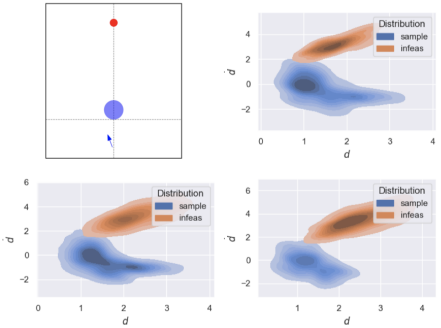



Figure 4: The custom environment and distributions of sampling state and infeasible states under three different initialization setups. The overlap of two distributions are small, which indicates that there exists feasible control for almost all sampled states.

# Conclusions

- We proposed a constrained RL algorithm that simultaneously learns the safe policies and synthesizes the safety certificates.
- We unified the loss function design of SIS and learning safe control policy (i.e., the RL loss), so we can prove the convergence of the proposed joint synthesis algorithm in a multi-timescale manner.

# Reference


Ya-Chien Chang, Nima Roohi, and Sicun Gao.
Neural lyapunov control.
*arXiv preprint arXiv:2005.00611*, 2020.

Richard Cheng, Gábor Orosz, Richard M Murray, and Joel W Burdick.
End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks.
In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3387–3395, 2019.

Haitong Ma, Jianyu Chen, Shengbo Eben Li, Ziyu Lin, and Sifa Zheng.
Model-based constrained reinforcement learning using generalized control barrier function.
*arXiv preprint arXiv:2103.01556*, 2021.

Yue Meng, Zengyi Qin, and Chuchu Fan.
Reactive and safe road user simulations using neural barrier certificates.
*arXiv preprint arXiv:2109.06689*, 2021.

Zengyi Qin, Kaiqing Zhang, Yuxiao Chen, Jingkai Chen, and Chuchu Fan.
Learning safe multi-agent control with decentralized neural barrier certificates.
*arXiv preprint arXiv:2101.05436*, 2021.

Matteo Saveriano and Dongheui Lee.
Learning barrier functions for constrained motion planning with dynamical systems.
In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 112–119. IEEE, 2019.

Mohit Srinivasan, Amogh Dabholkar, Samuel Coogan, and Patricio A Vela.
Synthesis of control barrier functions using a supervised machine learning approach.
In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7139–7145. IEEE, 2020.

Andrew Taylor, Andrew Singletary, Yisong Yue, and Aaron Ames.
Learning for safety-critical control with control barrier functions.
In *Learning for Dynamics and Control*, pages 708–717. PMLR, 2020.

Li Wang, Aaron D Ames, and Magnus Egerstedt.
Safety barrier certificates for collisions-free multirobot systems.
*IEEE Transactions on Robotics*, 33(3):661–674, 2017.

Thanks!