Feasible Reachable Policy Iteration

(An efficient policy space pruning algorithm of safe reinforcement learning)

Shentao Qin[1], Yujie Yang[1], Yao Mu[2], Jie Li[1], Wenjun Zou[1], Jingliang Duan[3], Shengbo Eben Li[3]

[1]Tsinghua University, [2]The University of Hong Kong, [3]University of Science and Technology Beijing

1. Introduction

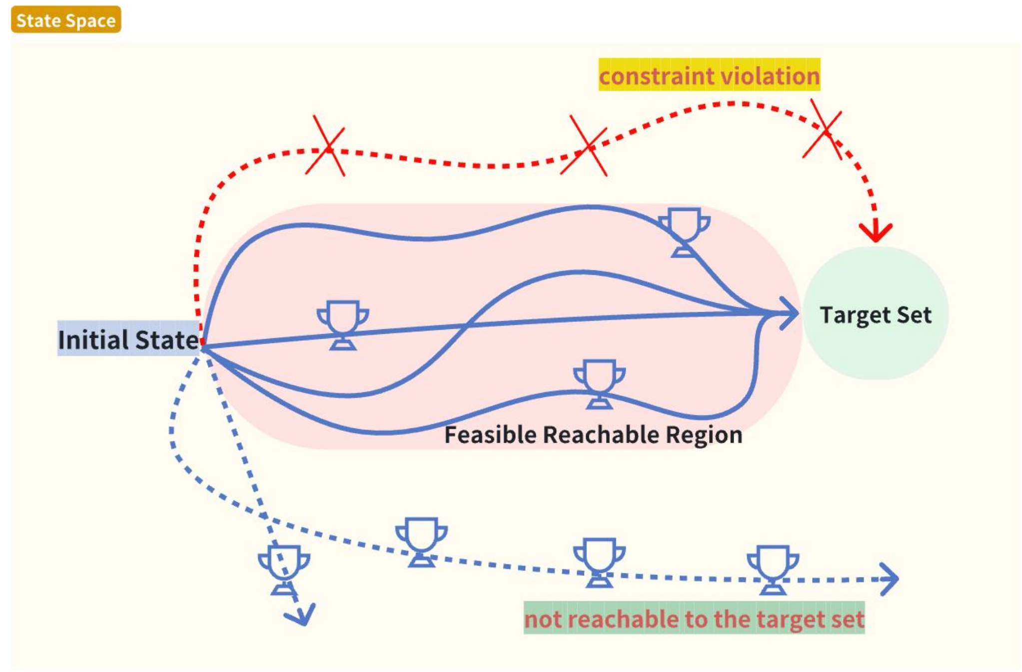
The goal-reaching tasks with safety constraints are common control problems in real world, such as intelligent driving and robot manipulation. The difficulty of this kind of problem comes from the **exploration termination caused by safety constraints** and the **sparse rewards caused by goals**. The existing safe RL avoids unsafe exploration by restricting the search space to a feasible region, the essence of which is **the pruning of the search space**. However, there are still many ineffective explorations in the feasible region because of the ignorance of the goals.

Contributions :

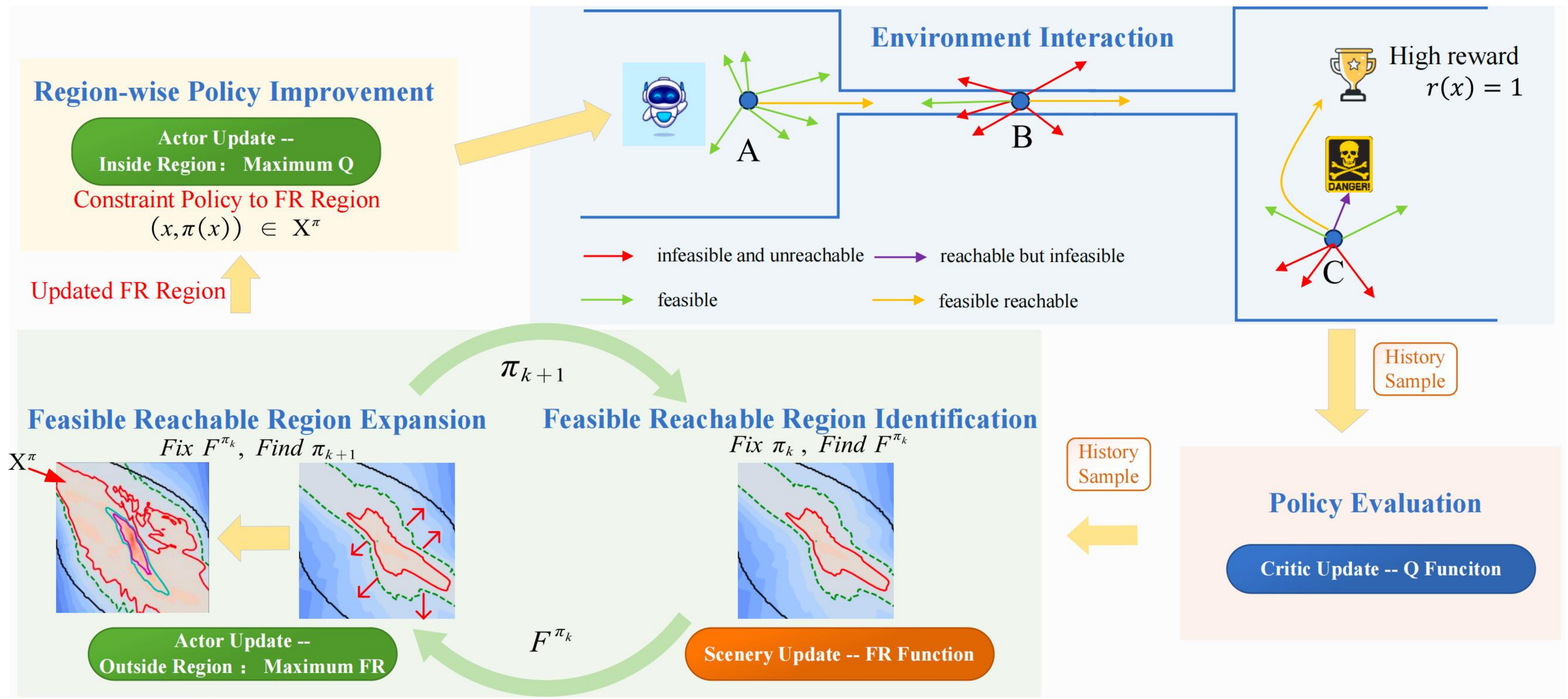
- We propose a novel feasible reachable function (FR function), which describes whether there is a policy to safely reach the target set. Our method takes both feasibility related to safety constraints and reachability related to goals into account, identifying the FR region to limit exploration.
- We propose a safe RL algorithm called feasible reachable policy iteration (FRPI), which uses the FR function to restrict policy improvement in the FR region to avoid inefficient exploration that is neither feasible nor reachable.
- The experiments show that FRPI achieves the best performance both in safety and return.

2. Objective

The objective of this paper is to find tbe max FR region where the policy can safely reach the target set. So we can restricte the search space to the region where make the policy improvement efficient.



3. What is the FRPI?



4. How FRPI Prune the State Space ?

- **Region Identification: Feasible Reachable Function**

$$F^\pi(x_0) = g(x_0) + c(x_0) + \sum_{m=1}^{T} \prod_{n=0}^{m-1} (1 + c(x_n))(1 - g(x_n))\gamma^n (g(x_m) + c(x_m))$$

we define $g(x) = \mathbf{1}_{x_{\text{goal}}}(x)$, indicating whether the target set is reached.

we define $c(x) = -\mathbf{1}_{\bar{x}_{\text{cstr}}}(x)$, indicating whether a state constraint is violated.

- **Region-wise Policy Improvement:**

Inside the FR Region:

$$\pi_{k+1}(x) = \arg\max_u r(x,u) + \gamma V^{\pi_k}(x') \quad \text{s.t. } F^{\pi_k}(x') > 0$$

Outside the FR Region:

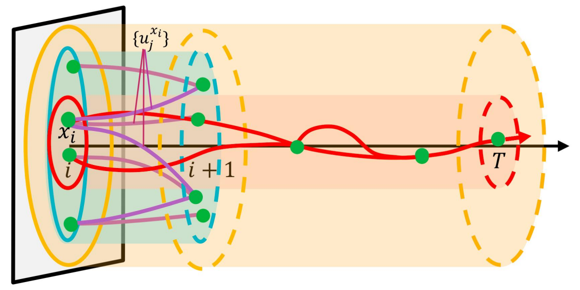
$$\pi_{k+1}(x) = \arg\max_u F^{\pi_k}(x')$$

- **Risk Bellman Equation**

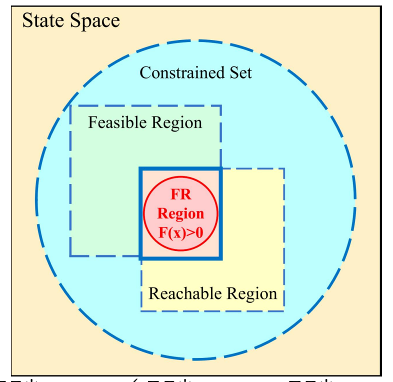
$$F(x) = c(x) + g(x) + (1 + c(x))(1 - g(x))\gamma \max_u F(x')$$

- **Feasible Reachable Bellman Equation**

$$V(x) = \max_{u \in \mathrm{U}^*(x)} r(x,u) + \gamma V(x')$$



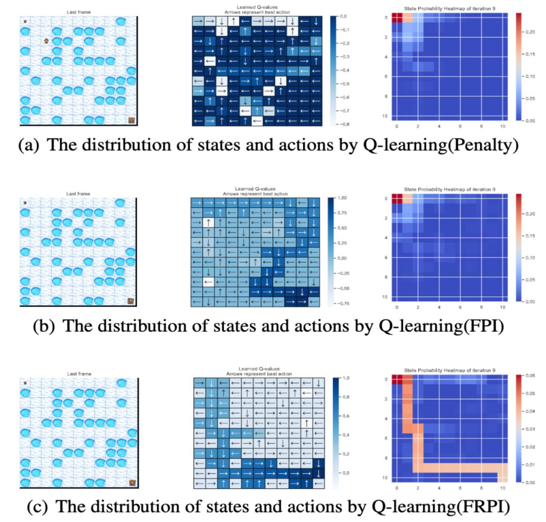
The green zone represents the feasible policy, while the red zone represents the FR policy.



$X^*_{\text{FR}} \subseteq (X^*_{\text{feas}} \cap X^*_{\text{reach}})$

5. Experiment Results

- **The Policy Pruning Verification**



(a) The distribution of states and actions by Q-learning(Penalty)

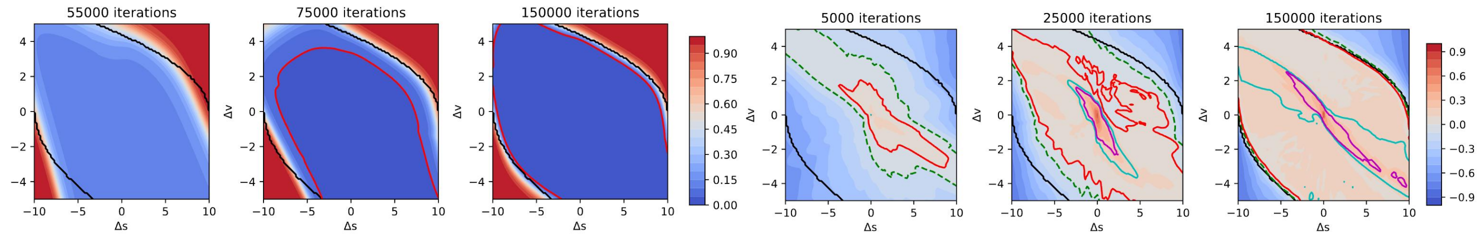
(b) The distribution of states and actions by Q-learning(FPI)

(c) The distribution of states and actions by Q-learning(FRPI)

Try to answer three questions:

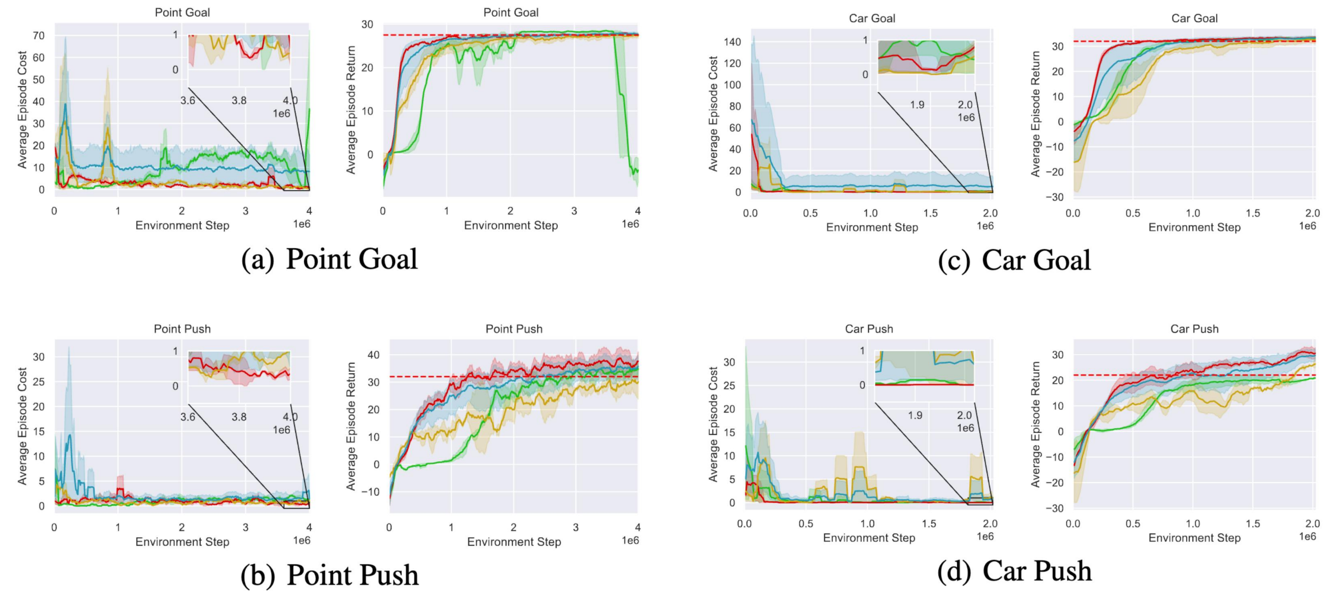
1. Can FR function enable an efficient policy space pruning to achieve faster convergence than other algorithms?
2. Can FRPI-SAC speed up feasible region expansion, and simultaneously identify a smaller FR region?
3. Do FRPI-SAC achieve a comparable performance faster than other algorithms without sacrificing safety?

- **The Feasible Region Expansion**



The region expansion between FPI(left) and FRPI(right)

- **The experiment on Safety Gym**



(a) Point Goal (c) Car Goal

(b) Point Push (d) Car Push

Ablation study & More Information

- Training on an NVIDIA GPU 0.1*3090 using JAX (allocates 2720 MB of GPU memory).
- WebSite : https://jackqin007.github.io/FRPI/
- Contact : qst23@mails.tsinghua.edu.cn

Table 1. The Inference Time on Safety Gym (ms)

Algorithm	FRPI-SAC	FPI-SAC	RAC	SAC-Lag	SAC
Inference Time (ms)	1.576	1.573	1.589	1.742	0.983

Tsinghua University