

Safe Execution of RL Policies

via Acceleration-based CBF-QP Constraint Enforcement
for Real-World Robotic Deployments

Runtime safety enforcement · No policy retraining · No performance degradation

Bastien Muraccioli^{*1}, Alice Cariou^{*1},
Pierre-Alexandre Leziart¹, Mathieu Celerier¹, Arnaud Demont¹, Gentiane Venture^{1,2}, Mehdi Benallegue¹

^{*} Indicates equal contribution, (1) CNRS-AIST Joint Robotics Laboratory, (2) The University of Tokyo



Paper · Code · Video
(open source)

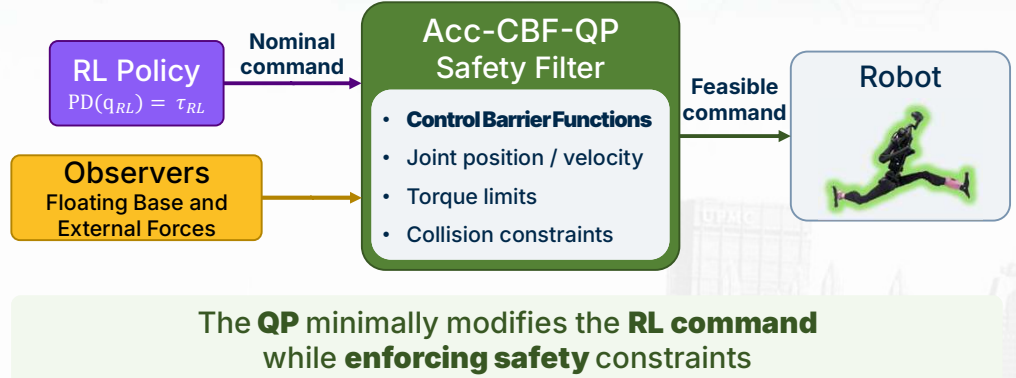
1 The Problem

Reinforcement Learning (RL) enables powerful robotic skills, but lacks **safety guarantees at deployment**

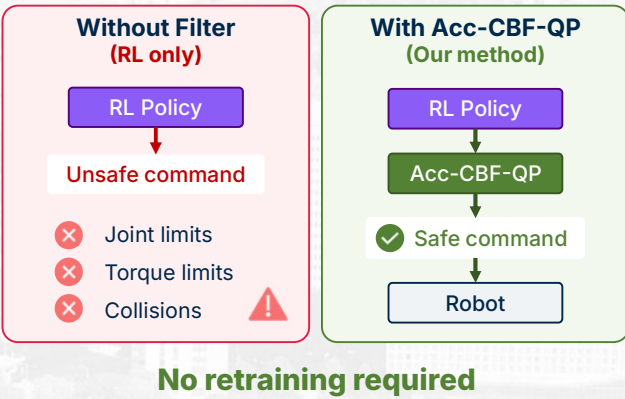
- ⚠️ Constraint violations
- ⚠️ Hardware damage
- ⚠️ Human & environmental

We need a method to **enforce safety constraints at runtime, without degrading performance**

2 The Solution: Acc-CBF-QP



3 Key Ideas



4 Method: RL + QP Tasks

$$\min_{\ddot{q}} \|A\ddot{q} - b\|^2 \quad \text{s.t. } h(q, \dot{q}, \ddot{q}) \geq 0$$

The QP enforces **safety constraints** while regulating deviation from the RL command

TorqueTask
Minimize torque deviation

- More robust to model error
- Lower hardware torque error

FDTask
Minimize accel. deviation

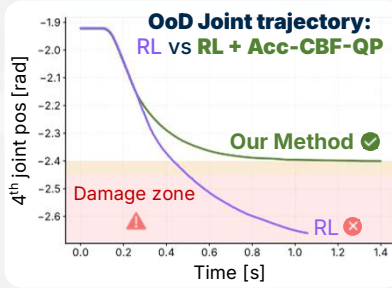
- Better accel. tracking
- Higher model sensitivity

Aspect	TorqueTask	FDTask
Objective	Torque dev.	Accel. dev.
Hardware torque error	Lower	Higher
Model sensitivity	Lower	Higher
Main advantage	Robustness	Accel. consistency

5 Results

Kinova Gen 3 - Manipulator

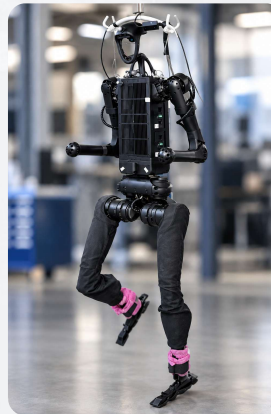
Inverse Kinematics policy without Safe RL during training



Acc-CBF-QP
✓ Safety constraints maintained under OoD states
RL alone
✗ Joint limits violated under OoD states → robot crash

Unitree H1 - Humanoid

Walking policy with Safe RL (CaT) during training



Nominal Behavior on Hardware

Safe RL alone
Our method
10.04 → **0.80**
violations/s → violations/s
-92%

200 Trials Simulation – Survival Time

Aggressive Vel. ⌚ +20% Survival Time Improvement	Uneven Terrain ⌚ No Degradation Survival Time
🎯 0 Constraint Termination	🎯 ~27x Less Constraint Termination

7 Positioning vs Prior Work

Training-time (Safe RL)

- Modify policy
- Retraining required to change constraints
- Statistical guarantees

Runtime (ours)

- Policy-agnostic filter
- No retraining required
- Theoretical guarantees, robust to unseen states

8 References

- B. Muraccioli et al., "Demonstrating a Control Framework for Physical Human-Robot Interaction Toward Industrial Applications," RSS, 2025.
- E. Chane-Sane et al., "CAT: Constraints as Terminations for Legged Locomotion Reinforcement Learning," IROS, 2024.

