Paper

Safety-Aware Preference-Based Learning for Safety-Critical Control

Video

Caltech

Ryan K. Cosner, Maegan Tucker, Andrew J. Taylor, Kejun Li, Tamás G. Molnár, Wyatt Ubellacker, Anil Alan, Gábor Orosz, Yisong Yue, Aaron D. Ames

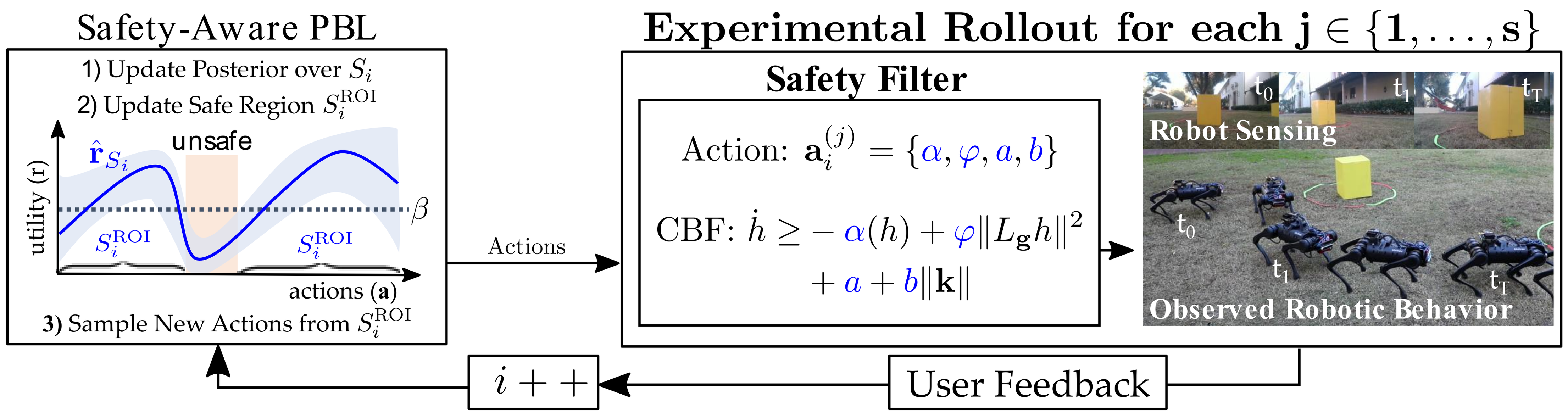
{rkcosner, mtucker, ajtaylor, kli5, tmolnar, wubellac, yyue, ames}@caltech.edu
{anilalan, orosz}@michigan.edu

UNIVERSITY OF MICHIGAN

Abstract

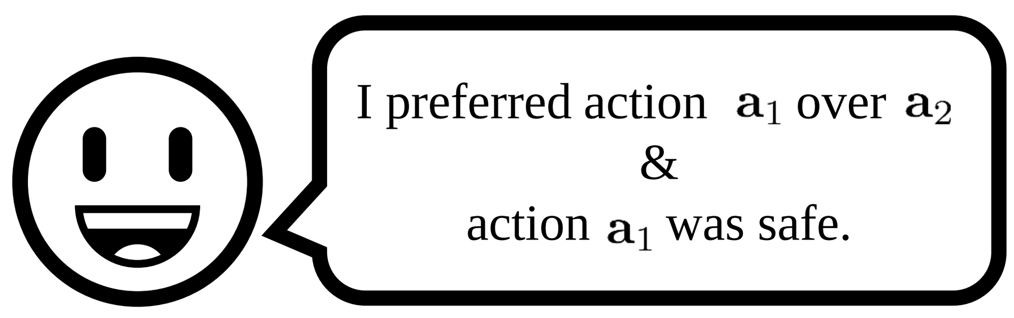
We present a design paradigm that tunes controller parameters to achieve a user's preferred safety-performance trade-off without requiring domain expertise. This is done in a *Safety-Aware* manner using Preference-Based Learning and Control Barrier Functions.

Overview



Safety-Aware Preference-Based Learning

User Feedback



Preference Likelihood Model:

$$\mathcal{P}(\mathbf{a}_1 \succ \mathbf{a}_2 \mid r(\mathbf{a}_1), r(\mathbf{a}_2)) = g_p\left(\frac{r(\mathbf{a}_1) - r(\mathbf{a}_2)}{c_p}\right)$$

Safety Label Likelihood Model:

$$\mathcal{P}(\mathbf{a}_1 \text{ is safe} \mid r(\mathbf{a}_1)) = g_o\left(\frac{\beta - r(\mathbf{a}_1)}{c_p}\right)$$

Gaussian Process Modeling

Given a dataset D of preferences and safety ordinal labels, the underlying reward is models as:

$$\mathcal{P}(\mathbf{r} \mid D) \propto \mathcal{P}(D_p \mid \mathbf{r})\mathcal{P}(D_o \mid \mathbf{r})\mathcal{P}(\mathbf{r})$$

The Maximum A Posteriori (MAP) is:

$$\hat{\mathbf{r}} \triangleq \operatorname{argmax}_{\mathbf{r}} \quad \mathcal{P}(\mathbf{r} \mid D)$$

We model the posterior as the multivariate Gaussian:

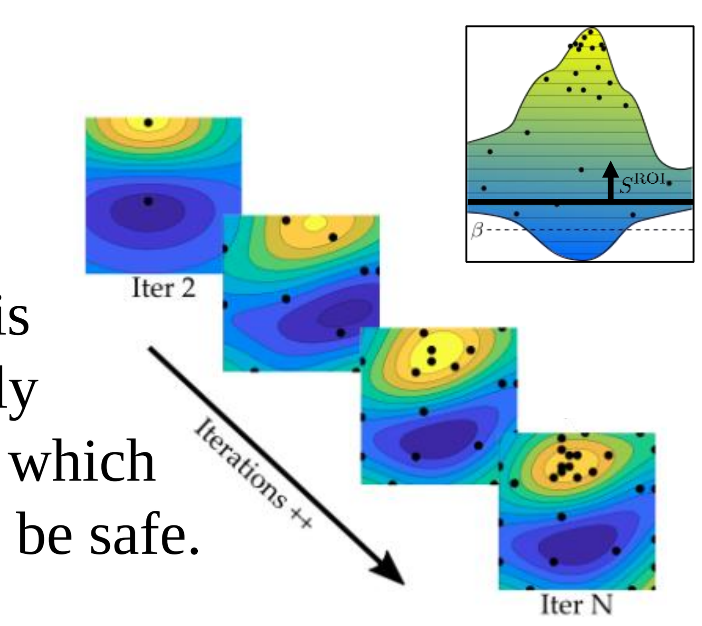
$$\mathcal{P}(\mathbf{r} \mid D) \approx \mathcal{N}(\hat{\mathbf{r}}, \Sigma)$$

Selecting New Actions

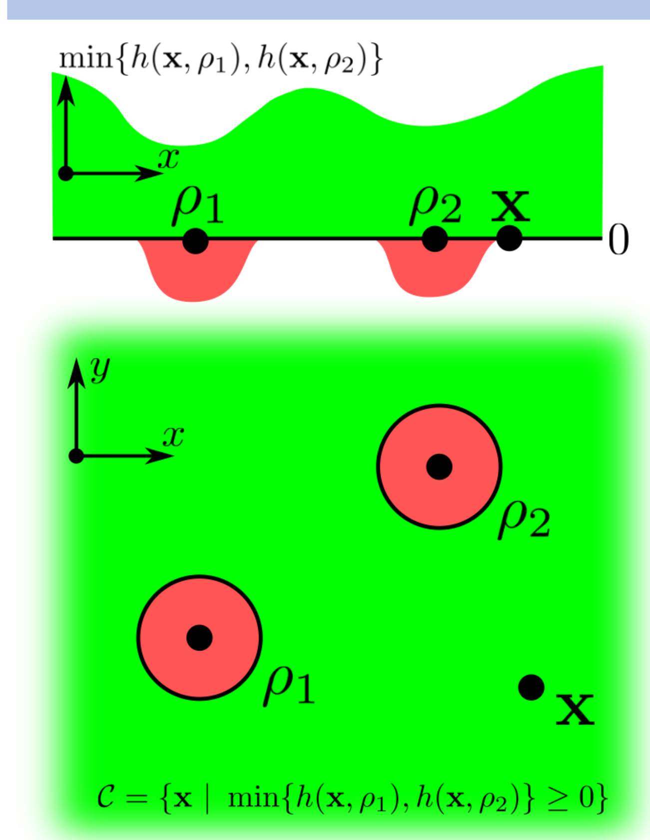
Safety-Aware LineCoSpar

$$S^{\mathrm{ROI}} = \{\mathbf{a} \in S \mid \hat{\mathbf{r}}_S(\mathbf{a}) + \lambda\sigma_S(\mathbf{a}) > \beta\}$$

LineCoSpar[1] is modified to only sample actions which are expected to be safe.



Robust Safety-Critical Control



Robustly-Safe Controllers often have bad Performance because they account for Worst-Case Scenarios.

Tunable Robustified Controller:

$$\mathbf{k}(\mathbf{x}) = \underset{\mathbf{v}\in\mathbb{R}^m}{\arg\min}\|\mathbf{v} - \mathbf{k}_{\mathrm{nom}}(\mathbf{x})\|^2$$

$$\text{s.t.} \;\; L_{\mathbf{f}}h(\mathbf{x}, \hat{\boldsymbol{\rho}}_i) + L_{\mathbf{g}}h(\mathbf{x}, \hat{\boldsymbol{\rho}}_i)\mathbf{v} - \varphi\|L_{\mathbf{g}}h(\mathbf{x}, \hat{\boldsymbol{\rho}}_i)\| - a - b\|\mathbf{v}\| \geq -\alpha h(\mathbf{x}, \hat{\boldsymbol{\rho}}_i), \quad \forall\, i \in \{1, 2\}$$

Tunable Parameters Provide Robustness to:

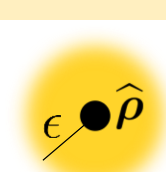
Model Error [2]

$$\dot{\mathbf{x}} = \hat{\mathbf{f}}(\mathbf{x}) + \hat{\mathbf{g}}(\mathbf{x})\mathbf{v} + \underbrace{\hat{\mathbf{g}}(\mathbf{x})\mathbf{d}(t)}_{\text{Matched Disturbance}}$$

Measurement Error [3]

True $\boldsymbol{\rho}$ vs. Measured $\hat{\boldsymbol{\rho}}$

$\boldsymbol{\rho} \in B_\epsilon(\hat{\boldsymbol{\rho}})$



Reduced-Order Modeling [4]

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\boldsymbol{\kappa}(\boldsymbol{\xi})$$

$$\dot{\boldsymbol{\xi}} = \mathbf{f}_{\boldsymbol{\xi}}(\boldsymbol{\xi}) + \mathbf{g}_{\boldsymbol{\xi}}(\boldsymbol{\xi})\mathbf{u}$$

Experimental Results

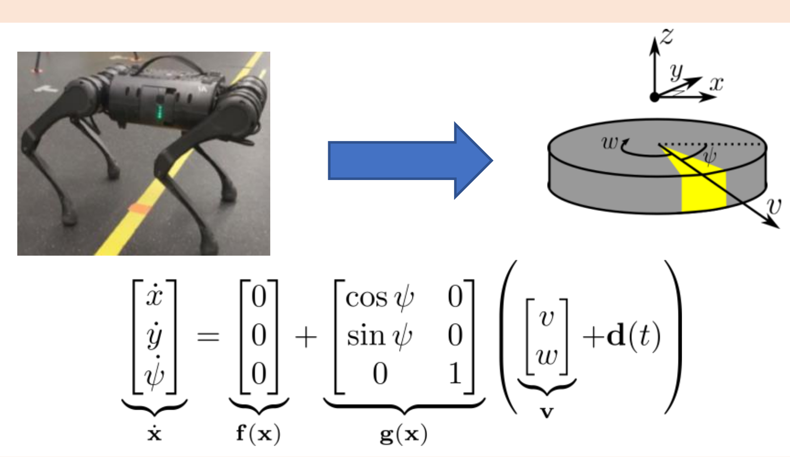
Experiment

For each iteration, the Unitree A1 robot was commanded using a reduced-order model to move across the room without prior knowledge of the obstacle locations. Obstacles were segmented visually and then tracked using SLAM. CBFs were generated for each perceived obstacle and used to filter the nominal input.

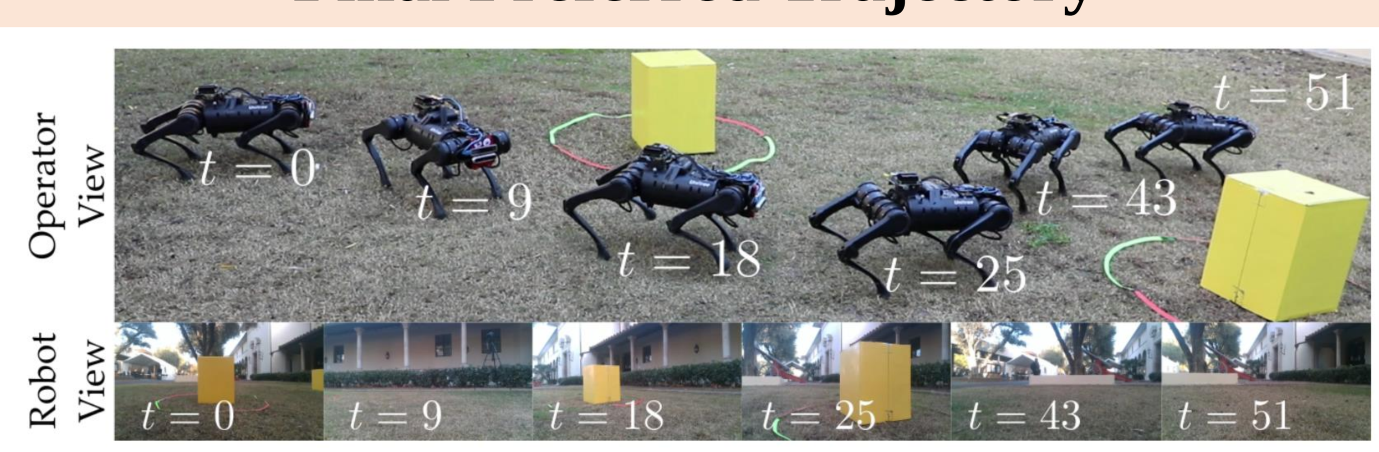
Safety-Aware PBL was conducted for the 4 tunable parameters $(\varphi,\ a,\ b,\ \alpha)$ with:

1. 30 iterations in simulation,
2. 7 iterations on hardware in lab,
3. 3 iterations on hardware outdoors.

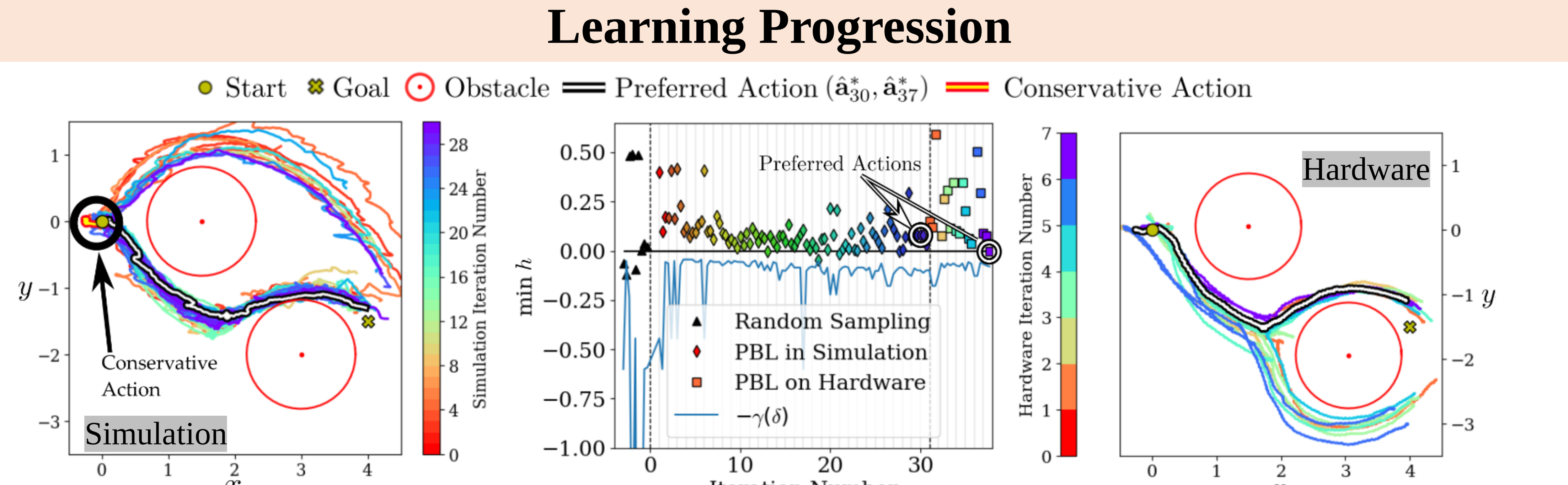
Reduced-Order Model



Final Preferred Trajectory



Learning Progression



References

1. Tucker, Maegan, et al. "Preference-based learning for exoskeleton gait optimization." *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020.
2. Kolathaya, Shishir, and Aaron D. Ames. "Input-to-state safety with control barrier functions." *IEEE control systems letters* 3.1 (2018): 108-113.
3. Dean, Sarah, et al. "Guaranteeing Safety of Learned Perception Modules via Measurement-Robust Control Barrier Functions." *Conference on Robot Learning*. PMLR, 2021.
4. Molnar, Tamas G., et al. "Model-free safety-critical control for robotic systems." *IEEE Robotics and Automation Letters* 7.2 (2021): 944-951.

Acknowledgements

DOW
WANDERCRAFT
bp
AeroVironment
NSF