Preference-Based Learning for Dynamic Bipedal Locomotion

Maegan Tucker, Noel Csomay-Shanklin, and Aaron D. Ames

California Institute of Technology

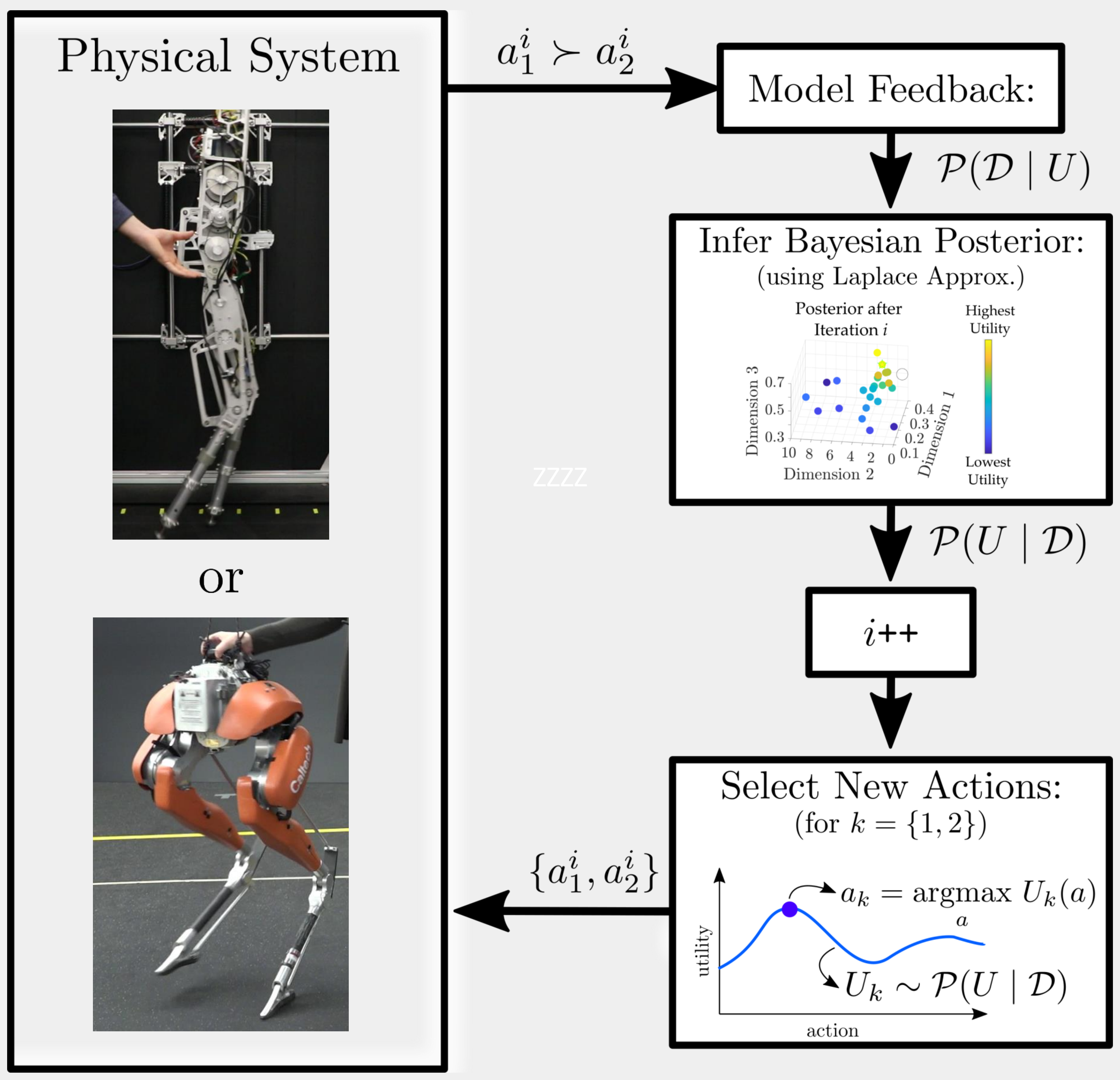
{mtucker,noelcs,ames}@caltech.edu

Abstract

- Use Preference-Based Learning to identify a^* with min. regret
- Experimentally demonstrate PBL towards identifying
  1) HZD constraints on AMBER-3M with unmodeled spring feet
  2) ID-CLF-QP$^+$ controller gains on Cassie.

Preference-Based Learning Algorithm (LineCoSparV2)

The LineCoSpar algorithm is aimed at identifying and sampling the optimal action, $a^* := \text{argmax}_{a \in \mathbb{R}^d} \; U(a)$ for some function $U : \mathbb{R}^d \to \mathbb{R}$, in as few iterations as possible.



Preference-based learning is beneficial for non-intuitive problems that aren't captured easily by a reward.

Limitations

- Action space bounds must be predefined
- Set of potential new actions is limited to a discrete set of actions
- Future work includes modifications to the learning framework to shift the action space based on the user's preferences

Conclusions

- The proposed preference-based learning framework is effective towards systematically exploring large parameter spaces using only a humans natural ability to judge "good" walking and experimentally resulted in improved locomotion for both platforms.

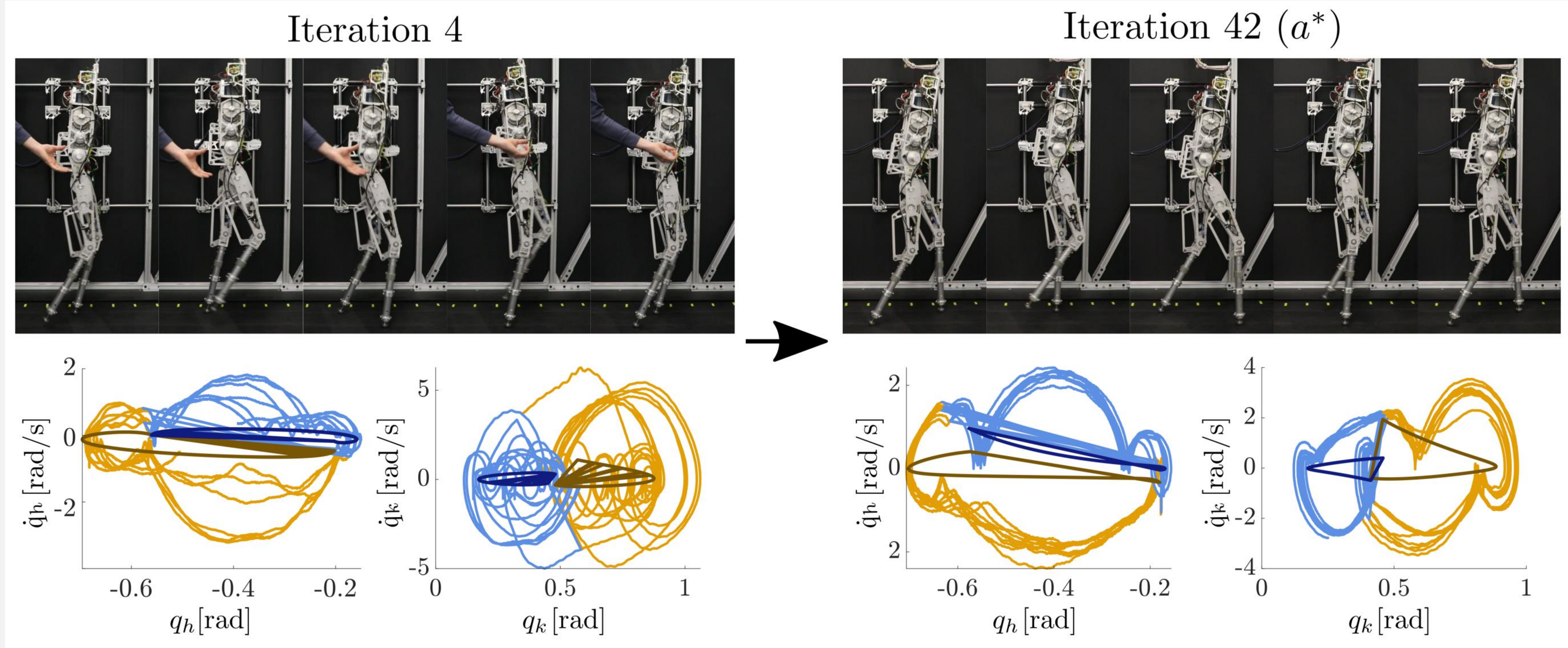
Learning Essential Constraints on AMBER-3M with Spring Feet

Experimental Setup:

HZD Optimization:

$$
\begin{aligned}
\{\alpha^*, X^*\} = \underset{\alpha, X}{\text{argmin}} \; & \Phi(X) \\
\text{s.t.} \quad & \dot{x} = f_{cl}(x) && \text{(Closed-loop Dynamics)} \\
& \Delta(\mathcal{S} \cap \mathcal{Z}_\alpha) \subset \mathcal{Z}_\alpha && \text{(HZD Condition)} \\
& X_{\min} \preceq X \preceq X_{\max} && \text{(Decision Variables)} \\
& c_{\min} \preceq c(X) \preceq c_{\max} && \text{(Physical Constraints)} \\
& a_{\min} \preceq p(X) \preceq a_{\max} && \text{(Essential Constraints)}
\end{aligned}
$$

Action Space: $a := [a_1, \ldots, a_5]$ s.t.:	
	Bounds
Avg. Vel. (m/s)	$a_1 : [0.3, 0.6]$
Clearance Tau $(\cdot)$	$a_2 : [0.4, 0.7]$
Min. Clearance (m)	$a_3 : [0.05, 0.19]$
Impact Vel. (m/s)	$a_4 : [-0.8, -0.2]$
Step Length (m)	$a_5 : [0.2, 0.4]$



Video: `https://youtu.be/rLJ-m65F6C4`

Learning ID-CLF-QP$^+$ Controller Gains on Cassie

Experimental Setup:

Rapidly Exponentially Stabilizing CLF (RES-CLF):

$$V(\eta) = \eta^\top \underbrace{I_\epsilon P I_\epsilon}_{P_\epsilon} \eta, \qquad I_\epsilon = \begin{bmatrix} \frac{1}{\epsilon} I & 0 \\ 0 & I \end{bmatrix},$$

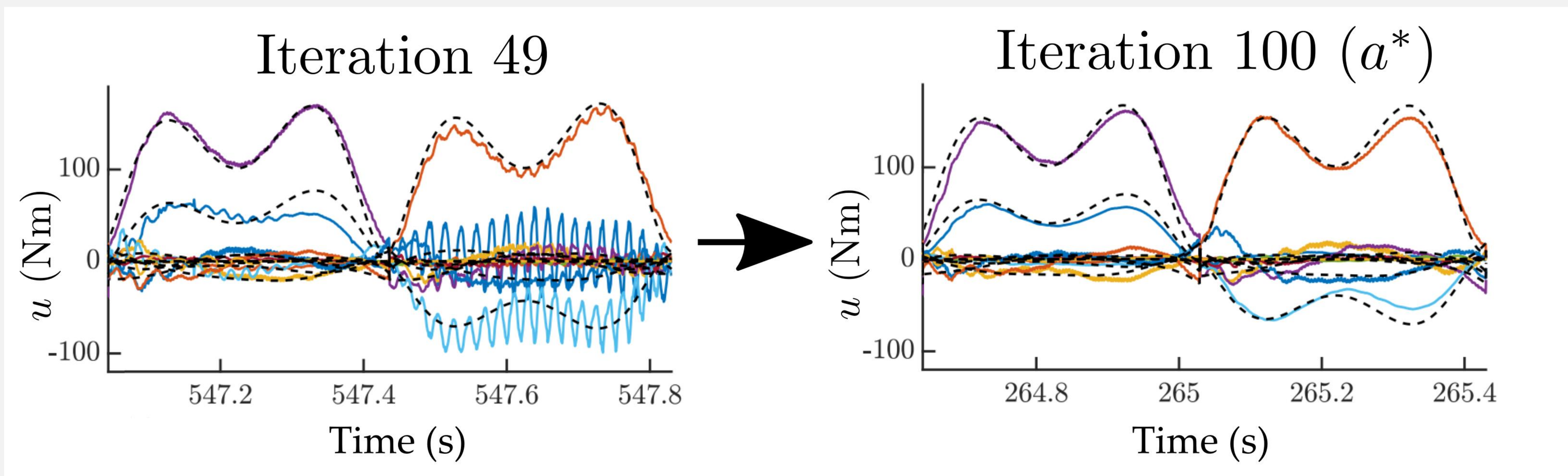
Continuous time algebraic Ricatti equation (CARE):

$$F^\top P + PF + PGR^{-1}G^\top P + Q = 0,$$

ID-CLF-QP$^+$:
(with $\mathcal{X} := [\ddot{q}^\top, u^\top, \lambda^\top]^\top \in \mathbb{R}^{39}$ on Cassie)

$$
\begin{aligned}
\mathcal{X}^* = \underset{\mathcal{X} \in \mathbb{X}_{ext}}{\text{argmin}} \quad & \|A(x)\mathcal{X} - b(x)\|^2 + \dot{V}(q, \dot{q}, \ddot{q}) \\
\text{s.t.} \quad & D(q)\ddot{q} + H(q, \dot{q}) = Bu + J(q)^\top \lambda \\
& u_{min} \preceq u \preceq u_{max} \\
& \lambda \in \mathcal{AC}(\mathcal{X})
\end{aligned}
$$

Action Space Definition: $a := [a_1, \ldots, a_{12}]$ such that:		
	Pos. Bounds	Vel. Bounds
Q Pelvis Roll	a_1:[2000, 12000]	a_7:[5, 200]
Q Pelvis Pitch	a_2:[2000, 12000]	a_8:[5, 200]
Q Stance Leg Length	a_3:[4000, 15000]	a_9:[50, 500]
Q Swing Leg Length	a_4:[4000, 20000]	a_{10}:[50, 500]
Q Swing Leg Angle	a_5:[1000, 10000]	a_{11}:[10, 200]
Q Swing Leg Roll	a_6:[1000, 8000]	a_{12}:[5, 150]



Video: `https://youtu.be/wrdNKK5JqJk`