# Integral Performance Approximation for Continuous-Time Reinforcement Learning Control

ICLR 2025

Brent A. Wallace, Jennie Si

School of Electrical, Computing and Energy Engineering
Arizona State University
April 24-26, 2025

# Motivation/Contributions



**Motivation:**

- Discrete-time (DT) RL theoretical and applications successes
  - Excellent convergence, optimality, closed-loop stability guarantees
  - Successfully applied to many real-world systems
- Prevalent CT-RL Adaptive Dynamic Programming (ADP) [1, 2, 3, 4]:
  - Substantial theoretical results: convergence, optimality, closed-loop stability
  - Few results on controller synthesis, thus no demonstrated applications [5]
- New CT-RL fitted value iteration (FVI) methods in deep RL (DRL) [6, 7]:
  - Great empirical promises: Results exceed ADP designs
  - But lack ADP's theoretical guarantees
  - And inherit high data/computational complexity from DRL in DT

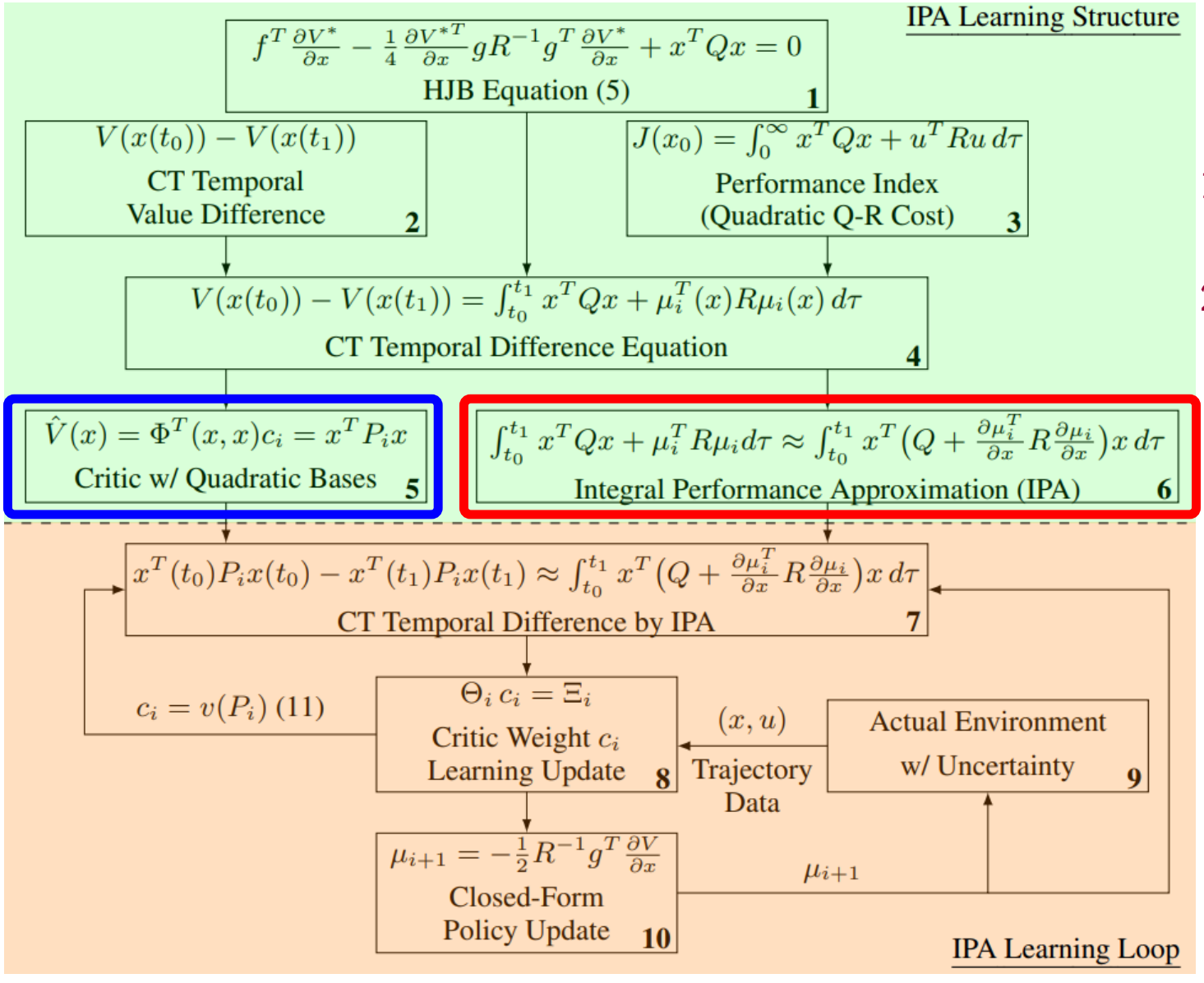
**Contributions:** A new SOTA CT-RL method (IPA) in the following context:

1. Leveraging affine nonlinear model, quadratic cost, and Kleinman control structures for great data efficiency and robust control performance
2. Theoretical guarantees of convergence, optimality, closed-loop stability
3. Comprehensive evaluations, including challenging hypersonic vehicle (HSV)

# How IPA Works: The Key Ideas



## IPA Algorithm Block Diagram




Use of:

1. Integral Performance Approximation (IPA)
2. Quadratic bases

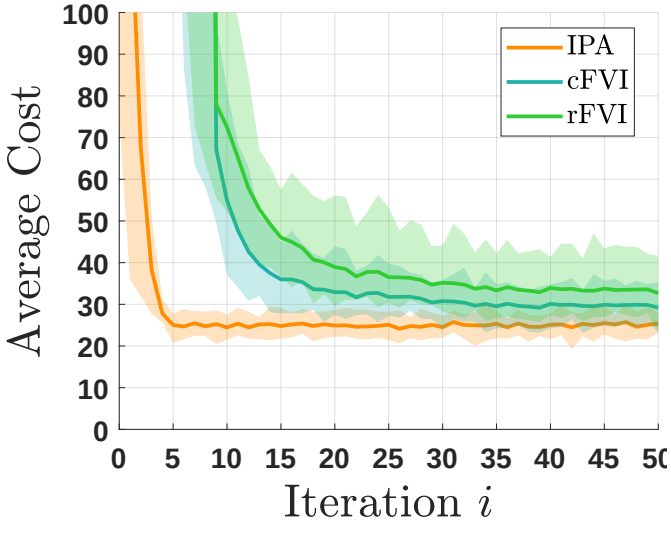
enable IPA to take advantage of Kleinman control structures for:

- High data efficiency
- Improved excitation/exploration
- Well-behaved system/learning responses

# IPA vs. SOTA Fitted Value Iteration (FVI)

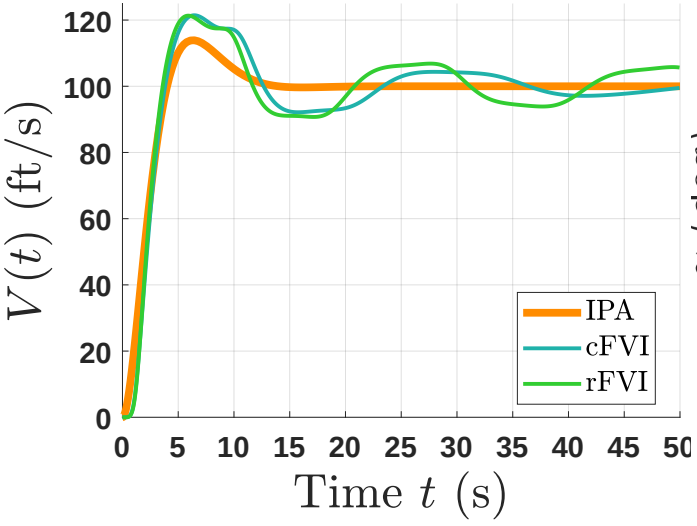


Learning Curves




Fastest Learning Convergence + Lowest Variance

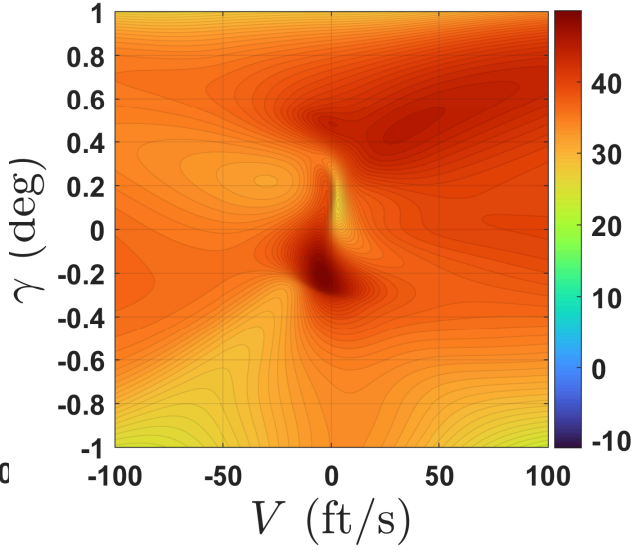
Step Response




Favorable Closed-Loop Time-Domain Performance (Overshoot, Settling Time)

Cost vs. Initial Condition $x_0$

rFVI – IPA ($> 0$: IPA Better)




Lowest Cost Performance + Best Generalization to Environment Uncertainty

IPA Enables Efficient Learning with Well-Behaved Responses on Significant HSV Application

# IPA Time/Data/Computational Complexity



Algorithm Time/Data/Computational Complexity on Hypersonic Vehicle

| Parameter | IPA | FVIs | Ratio FVIs/IPA |
|---|---|---|---|
| # Trajectory Data Samples | 30 | 173,000,000 | 5,800,000 |
| # Data Episodes | 1 | 5,250,000 | 5,250,000 |
| Average Training Time (s) | 2.75 | 4,600 | 1,673 |
| # Algorithm Iterations $i$ | 10 | 50 | 5 |

- **Table:** SOTA FVIs exhibit significantly greater time/data/computational complexity than IPA
  - Training deep networks $\Rightarrow$ greater complexity
- **Note:** Despite great FVI successes illustrated here, current FVI data complexity not practical for training in mission-critical flight control applications (e.g., HSV)

IPA's use of Integral Performance Approximation + Quadratic Bases + Kleinman Structures Significantly Lowers Complexity

# Key Takeaways from CT-RL



- IPA vs. ADP:
  - Both have comprehensive theoretical guarantees
  - IPA achieves synthesis and substantial learning generalization
- IPA vs. Deep RL FVIs:
  - IPA cost, approximation, and closed-loop performance meets and often exceeds FVIs
  - Reductions in data complexity of 6 orders of magnitude

# References




[1] D. Vrabie and F. L. Lewis, "Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems," *Neur. Net.*, vol. 22, no. 3, pp. 237–246, 2009.

[2] K. G. Vamvoudakis and F. L. Lewis, "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, 2010.

[3] Y. Jiang and Z.-P. Jiang, "Robust adaptive dynamic programming and feedback stabilization of nonlinear systems," *IEEE TNNLS*, vol. 25, no. 5, pp. 882–893, Jan. 2014.

[4] T. Bian and Z.-P. Jiang, "Reinforcement learning and adaptive optimal control for continuous-time nonlinear systems: A value iteration approach," *IEEE TNNLS*, vol. 33, no. 7, pp. 2781–2790, Jul. 2022.

[5] B. A. Wallace and J. Si, "Continuous-time reinforcement learning control: A review of theoretical results, insights on performance, and needs for new designs," *IEEE TNNLS*, Feb. 2023.

[6] M. Lutter, S. Mannor, J. Peters, D. Fox, and A. Garg, "Value iteration in continuous actions, states and time," in *Proc. 38th Int. Conf. Mach. Learn. (ICML)*, vol. 139, Jul. 2021, pp. 7224–7234.

[7] M. Lutter *et al.*, "Continuous-time fitted value iteration for robust policies," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 45, no. 5, pp. 5534–5548, May 2023.