Collection of Paper hyperlinks for Thesis Actor Critic MPC

July 22, 2025

Contents

Chapter 1

Differential AC MPC

1.1 Synthesis of Model Predictive Control and Reinforcement Learning: Survey and Classification

Summary:
This paper examines the shared foundations and key differences between Model Predictive Control (MPC) and Reinforcement Learning (RL). It highlights their complementary strengths and growing interest in combining the two..

Link to the paper: `https://arxiv.org/pdf/2502.02133`

1.2 Differentiable MPC for End-to-end Planning and Control

Summary:
Model Predictive Control (MPC) as a differentiable policy class for reinforcement learning in continuous state and action spaces.

Link to the paper: `https://arxiv.org/pdf/1810.13400`

1.3 OptNet:Differentiable Optimization as a Layer in Neural Networks

Summary:
Basis where the differential MPC is built on, OptNet embeds QP optimization as neural-network layers, enabling exact differentiation

Link to the paper: `https://arxiv.org/pdf/1703.00443`

1.4 Actor-Critic Model Predictive Control

Summary:
Development of framework called Actor-Critic Model Predictive Control. The key idea is to embed a differentiable MPC within an actor-critic RL framework.

Link to the paper: `https://arxiv.org/pdf/2306.09852v4`

1.5 Actor-Critic Model Predictive Control: Differentiable Optimization meets Reinforcement Learning

Summary:

follow-up paper expands and optimaze the Actor-Critic Model Predictive Control with Model-Predictive Value Expansion

Table 1.1: Comparative Analysis of Constraint-Handling Methodologies in Reinforcement Learning

Approach	Seminal Paper & Constraint Integration	Advantages	Limitations
Constrained RL (CPO)	**Source:** Achiam et al. (2017), arXiv:1705.10528 **Integration:** Lagrangian in policy gradient. **Differentiable:** Yes.	Integrated learning; model-free; iterative safety guarantees.	No per-step guarantee; conservative policies; potential infeasibility.
Safety Layer (Safety Filter)	**Source:** Dalal et al. (2018), arXiv:1801.08757 **Integration:** External module corrects actions. **Differentiable:** Yes.	Zero violations (w/ good model); very fast; algorithm-agnostic.	Model-dependent performance; frequent intervention hurts performance; single constraint assumption.
Barrier Functions (CBF) + QP	**Source:** Cheng et al. (2019), arXiv:1812.09528 **Integration:** QP filter enforces CBF condition. **Differentiable:** No (not end-to-end).	Rigorous per-step guarantee; stable; minimal intervention.	Needs accurate model; high online cost; hard to design barrier function.
Differentiable MPC	**Source:** Amos et al. (2018), arXiv:1810.13400 **Integration:** MPC solver as policy layer. **Differentiable:** Yes (end-to-end).	Hard input constraints; end-to-end training; sample efficient.	High computational overhead; complex setup; state constraints are difficult.
Recovery Policy (Recovery RL)	**Source:** Thananjeyan et al. (2020), arXiv:2010.15920 **Integration:** Switches between policies. **Differentiable:** Components are; switching is not.	Decouples objectives; better exploration; strong practical safety.	High system complexity; relies on recovery policy/critic; suboptimal switching.
Koopman-based Differentiable MPC	**Source:** Yang et al. (2023), arXiv:2307.03184 **Integration:** Differentiable MPC on learned linear model. **Differentiable:** Yes (end-to-end).	Handles nonlinearity with linear tools; hard input constraints; end-to-end.	Relies on embedding quality; high computational cost; assumes learnable linear map.

Chapter 2

Koopman Theory and Implementation

2.1 MODERN KOOPMAN THEORY FOR DYNAMICAL SYSTEMS

Summary: Theoretical Basis: Koopman spectral theory has emerged as a leading framework representing a nonlinear dynamics through an infinite-dimensional linear operator acting on the space of all possible measurement functions of the system.

Link to the paper: `https://arxiv.org/pdf/2102.12086`

2.2 Koopman Constrained Policy Optimization: A Koopman operator theoretic method for differentiable optimal control in robotics

Summary: Key paper for embedding hard constrains in differentiable MPC
Link to the paper:`https://openreview.net/pdf?id=3W7vPqWCeM`

2.3 End-to-End Reinforcement Learning of Koopman Models for Economic Nonlinear Model Predictive Control

Summary: Koopman end to end differentiable for economic nonlinear MPC with hard constrain
Link to the paper : `https://arxiv.org/pdf/2308.01674`

2.4 Koopman-Assisted Reinforcement Learning

Summary: Soft actor critic RL approach using koopman operator

Link to the paper : `https://arxiv.org/pdf/2403.02290v1`

Chapter 3

Advantage of Transformers as Critic

3.1 Learning Humanoid Locomotion over Challenging Terrain

Summary: The self-attention mechanism in Transformers enables effective credit assignment over time, allowing the model to capture key environmental features and dynamics. Additionally, Transformers can model diverse behaviors, which is essential for generalizing to and performing well in unseen scenarios.

Link to the paper: `https://arxiv.org/pdf/2410.03654`

3.2 Chunking the Critic: A Transformer-based Soft Actor-Critic with N-Step Returns

Summary: leverages the Transformer's ability to process sequential information, facilitating more robust value estimation. Empirical results show that this method not only achieves efficient, stable training but also excels in sparse reward/multi-phase environments-traditionally a challenge for step-based methods.

Link to the paper: `https://arxiv.org/pdf/2503.03660`

3.3 Transformers for Trajectory Optimization with Application to Spacecraft Rendezvous

Summary: Transformers learn near-optimal policies from previously collected data, and warm-start a sequential optimizer for the solution of non-convex optimal control problems, thus guaranteeing hard constraint satisfaction.

Link to the paper: `https://arxiv.org/pdf/2310.13831v3`
https://arxiv.org/pdf/2310.13831v3

3.4 Decision Transformer: Reinforcement Learning via Sequence Modeling

Summary: Decision Transformer, an architecture that frames Reinforcement Learning (RL) as a conditional sequence modeling problem. Instead of using value functions or policy gradients, it uses a causal-masked Transformer to generate optimal actions by conditioning the model on the desired return, past states, and past actions. The approach matches or outperforms state-of-the-art model-free offline RL methods on various benchmarks.

Link to the paper : `https://arxiv.org/pdf/2106.01345`

3.5 Reinforcement Learning as One Big Sequence Modeling Problem

Summary: This paper proposes to view RL as a generic sequence modeling problem, using a "Trajectory Transformer" to model distributions over sequences of states, actions, and rewards. This approach simplifies RL by replacing separate components with a single Transformer model and uses beam search as the planning algorithm. The paper demonstrates the method's flexibility in imitation learning, goal-conditioned RL, and offline RL.

Link to the paper : `https://papers.neurips.cc/099fe6b0b444c23836c4a5d07346082b-Paper.pdf`

3.6 Stabilizing Transformers for Reinforcement Learning

Summary: This paper addresses the optimization challenges of applying standard Transformer architectures to reinforcement learning. The authors propose architectural modifications resulting in the Gated Transformer-XL (GTrXL), which significantly improves stability and learning speed. GTrXL is shown to outperform LSTM networks on demanding memory-based tasks and achieves state-of-the-art results on the DMLab-30 benchmark suite.

Link to the paper : `https://arxiv.org/pdf/1910.06764`

3.7 Decision Mamba: Reinforcement Learning via Sequence Modeling with Selective State Spaces

Summary: This paper studies the integration of the Mamba framework, known for efficient sequence modeling, into the Decision Transformer architecture to improve performance in sequential decision-making tasks. The study evaluates this new architecture, "Decision Mamba," by comparing it with the traditional Decision Transformer in various decision-making environments.

Link to the paper : `https://arxiv.org/pdf/2403.19925`