DEEP LEARNING FOR SEQUENTIAL DECISION-MAKING PROBLEMS IN WIRELESS SYSTEMS

By

SPILIOS EVMORFOS

A dissertation proposal submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Athina P. Petropulu

and approved by

New Brunswick, New Jersey

January, 2025

ABSTRACT OF THE DISSERTATION PROPOSAL

Deep Learning for Sequential Decision-Making Problems in Wireless Systems

by Spilios Evmorfos

Dissertation Director:

Athina P. Petropulu

The recent advancements in deep learning, coupled with its integration into sequential decision-making frameworks such as dynamic programming, have transformed the approach to solving complex optimization problems in dynamic environments. In wireless systems, spanning both communication and sensing/localization, the need for intelligent and adaptive paradigms has grown increasingly critical. These systems operate in highly dynamic settings characterized by mobility, fluctuating channels, and varying performance demands, which render traditional myopic approaches inadequate. Deep learning enables the modeling of intricate dependencies, and its fusion with sequential decision-making frameworks allows for the approximation of optimal decision policies directly from data and system interactions. This combination facilitates adaptive responses to evolving conditions, making it essential for addressing the challenges and meeting the performance requirements of next-generation wireless networks.

This dissertation develops deep learning-based sequential decision-making approaches for various settings and challenges in wireless sensing and communications. Specifically, it addresses the general problem of antenna/sensor selection for thin array design in

wireless systems, a recurring issue tackled in prior work through supervised learning, convex optimization, or greedy methods. Here, the problem is reframed as a sequential decision-making task modeled as a deterministic Markov Decision Process. The Generative Flow Networks paradigm is adapted to learn an action-sampling policy, ensuring the probability of reaching each terminal state aligns with its reward. This approach outperforms greedy methods, convex optimization, and supervised learning across standard benchmarks.

The second focus is mobile relay motion control. A deep reinforcement learning approach is proposed to optimize relay movement over time under spatiotemporally correlated channels, maximizing the cumulative SINR at the destination. The channel variability introduces high-frequency components into the optimal value function, addressed by integrating Fourier features into the neural network for improved value function estimation.

The third setting involves designing Intelligent Reflective Surface (IRS) phase shift values for MISO communication systems under correlated channels. A deep reinforcement learning actor-critic method is developed, leveraging sufficient conditions on the critic's Neural Tangent Kernel to facilitate convergence under deep Q updates.

Acknowledgements

As I approach the end of my PhD journey, I find myself in a place I could hardly have imagined when I first set out. Nearly four and a half years ago, in the midst of a global pandemic, I left Greece—my hometown, my family, my parents' home—for the United States to pursue this goal. It was a time of great uncertainty, and leaving felt like an incredibly difficult choice. Looking back now, I am profoundly grateful for having taken that step. Despite the challenges and the times I missed my family and home, the experiences, growth, and connections I gained throughout this journey have been invaluable.

My PhD studies have given me the opportunity to meet and collaborate with remarkable individuals, each contributing to my growth as a researcher, professional, and human being. Now, as this chapter comes to an end and I prepare for the next steps in my career and life, I would like to take this moment to extend my deepest gratitude to the people who inspired, supported, and encouraged me along the way.

It is hard to express the depth of my gratitude to Professor Athina Petropulu, who has been my advisor throughout these years. Her unwavering support extended in every possible way—she was always there with guidance, encouragement, and technical insight whenever I needed them. Professor Petropulu also entrusted me with the freedom to explore new directions in my field, a trust that was crucial for my personal and professional growth. Above all, as great leaders do, she led by example; her commitment to excellence and dedication to her work served as a constant source of inspiration, encouraging me to always strive further. Her mentoring has shaped my mentality in so many ways. Professor, I am forever indebted to you—thank you so much.

I want to thank Professors Shirin Jalali, Aggelos Bletsas and Dionysios Kalogerias for agreeing to serve as members of my PhD committee and for all their valuable comments and suggestions.

I owe an enormous debt of gratitude to my friend—and, truly, my brother—George Vlachodimitropoulos. Without him, this dissertation would not have been possible. George and I were inseparable long before we decided to pursue our PhDs together in the United States. We made the transition together, navigating our first year side by side, which was invaluable to me. Amidst the uncertainty and global challenges, I had complete trust in us making this leap together. George eventually returned to Greece to pursue a different path, and I have deeply missed him over the last three and a half years. Still, his support from afar has never wavered. His encouragement and our conversations during my visits back to Greece have been invaluable to me. I don't think I will ever be able to thank him enough for all he has given me.

I would also like to extend my heartfelt thanks to Tasos Dimas, whom I had the pleasure of meeting when I began my PhD journey. As I was just starting, Tasos was completing his own PhD, and we quickly became friends. He is incredibly smart, kind, and a truly wonderful individual who treated me like a younger brother from the beginning. His guidance and advice have been invaluable, and I am grateful to have had his support and friendship along the way.

I am deeply grateful to Rutgers University for providing an exceptional environment for learning and growth. I would like to extend special thanks to Katie Yu, Christopher Reid and John Scafidi for their outstanding dedication and willingness to assist me whenever I encountered challenges. I am also particularly grateful to Professor Waheed Bajwa and Professor Yingying Chen, who served as graduate directors during my time here. I sought their guidance many times, and their support has been invaluable throughout my journey. Thank you for your unwavering assistance and kindness.

I am grateful to have met many remarkable individuals at Rutgers who helped build a supportive and motivating community. Together, we created an environment where we encouraged, supported, and inspired one another throughout our academic journeys. I would like to take this opportunity to thank, in no particular order: Anastasis Stathopoulos, George Chatzialexiou, Isidoros Marougkas, Jerry Chatzoudis, and Konstantinos Nikolakakis.

I am deeply grateful to my friends (or should I say brothers) back in Greece, with whom I have shared a lifetime of memories. From childhood through our mid twenties, we spent every day together, studying in the same classrooms and gathering in the same parks. I hold each of them close to my heart, and I have missed them immensely throughout these years in the States. Although our lives have taken different directions, our bond remains unwavering. Their constant support and encouragement, along with the conversations and laughter we share whenever I visit, have been invaluable in motivating me to keep pushing forward. I want to thank, in no particular order: Spyros Pougkakiotis, Alexandros Koukis, Nikolas Damianakos, John Vlaikidis, Efthimis Psaraftis, Aggelos Tzouchas, George Iniotakis, George Tofalos, Thanos Nikologiannis, Dimitris Apostolopoulos and Dimitris Konstantinidis.

My deepest gratitude belongs to my family—my parents, Panos and Kiki, my sister, Evangelia and my aunt Aleka. They have always shown me unwavering love and support, standing by me through every success and setback. In their quiet but profoundly reassuring way, they have been my foundation, present in both the highs and lows of this journey. They have allowed me the freedom to make my own choices, trusting in my path, yet they have always been there when I needed them most. I am forever grateful for their constant and unconditional presence in my life.

Lastly, I dedicate this dissertation, along with all the hard work it represents, to the most important person in my life—my love, my partner, and my soon-to-be wife, Evgenia. She is my other half, my constant source of strength and joy.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1 Introduction

The remarkable success of deep learning [72], a subset of machine learning, has fundamentally transformed the landscape of artificial intelligence [68]. Rooted in neural network architectures that mimic the structure of the human brain, deep learning enables the modeling of complex, high-dimensional relationships directly from data. The history of deep learning traces back to the development of early neural networks in the mid-20th century, such as the perceptron introduced by Rosenblatt in 1958 [113]. However, its practical potential was initially hindered by limitations in computational power, data availability, and algorithmic efficiency. The resurgence of deep learning in the 2010s can be attributed to breakthroughs such as backpropagation optimization, the availability of large datasets, and the advent of high-performance computing resources, particularly GPUs.

Deep learning's architecture is characterized by multiple layers of interconnected neurons, where each layer extracts increasingly abstract features from the input data. This hierarchical feature learning allows deep learning models to excel in tasks like image recognition [101], speech processing [65], and natural language understanding [95]. Architectures such as convolutional neural networks [71], recurrent neural networks [45], and transformer-based models [138] have driven state-of-the-art performance across numerous domains. The adaptability, scalability, and capacity to generalize from raw data have positioned deep learning at the forefront of technological advancements, fueling innovations in fields as diverse as healthcare [92], robotics [105], and financial modeling [50].

The rise of deep learning has led to its widespread adoption in wireless communications [27] and sensing [89], revolutionizing traditional settings with data-driven approaches. For instance, [125] investigates the time-frequency response of fast-fading communication channels and proposes a deep learning-based super-resolution technique for imputing missing channel values. In the domain of massive MIMO systems, [21] introduces a two-stage supervised deep learning framework for channel estimation, significantly improving accuracy and efficiency. Similarly, [17] presents a supervised deep learning method for radar detection, leveraging training exclusively on radar calibration data augmented through tailored techniques. Furthermore, [158] explores channel prediction and tracking in IRS-assisted UAV networks using a bidirectional LSTM neural network trained in a supervised manner. Lastly, [156] proposes a convolutional neural network that processes the range-Doppler ambiguity function to perform radar target detection effectively.

The 6th generation (6G) of wireless networks [39] is expected to power advanced applications such as connected vehicles [116], smart manufacturing [20], and smart cities [117], all operating in highly dynamic and uncertain environments. These environments are characterized by mobility, correlated channels, and the need to balance tradeoffs between sensing and communication, often integrating both functionalities on the same platform [164]. The success of these demanding applications hinges on their ability to adapt to rapidly changing conditions. Consequently, optimization and decision-making strategies must account for these variations over extended operational horizons. Traditional myopic methods, whether analytical or deep learning-based, fall short in addressing these requirements. Instead, sequential decision-making approaches, which can adapt dynamically and consider long-term implications, are essential for enabling the next generation of wireless systems.

The integration of deep learning with sequential decision-making frameworks has opened new frontiers in tackling complex optimization problems in dynamic and uncertain environments. At the core of this integration lies the paradigm of deep reinforcement learning (DRL) [8], which combines the representational power of deep learning with the decision-making framework of reinforcement learning (RL). RL, in its essence, addresses problems where an agent interacts with an environment, observes states, takes actions, and receives feedback in the form of rewards. The goal is to learn an optimal policy that maximizes cumulative rewards over time. Traditional RL methods often falter in environments with high-dimensional state and action spaces due to the inefficiency of classical function approximators. Deep learning addresses this limitation by employing neural networks to approximate the value function, policy, or both, enabling RL algorithms to scale effectively to such challenging domains.

DRL frameworks, such as deep Q learning [94] and actor-critic methods [37], allow for real-time learning and adaptation, making them particularly suitable for applications in dynamic settings like wireless systems. These methods can capture intricate dependencies in the state-action space while maintaining the ability to generalize across unseen scenarios. For example, deep Q learning approximates the state-action value function (Q-function) using a deep neural network, enabling the agent to learn effective policies in high-dimensional environments. Actor-critic frameworks, on the other hand, separate policy and value function estimation, facilitating more stable training and continuous action spaces.

Moreover, DRL excels in addressing problems with long-term objectives, where decisions at a given time step have cascading effects on future states. This capability is crucial for wireless systems, where optimizing parameters such as resource allocation, phase shifts, or relay movements requires foresight to maximize overall system performance. By integrating deep learning into RL, DRL methods not only overcome computational challenges but also provide a structured approach to adapt to the temporal and spatial dynamics of the environment. These features make DRL a cornerstone for enabling intelligent and adaptive decision-making in next-generation systems, including wireless communications and sensing.

Beyond DRL, several other sequential decision-making frameworks have emerged in deep learning, each tailored to address unique challenges and applications. One notable example is Generative Flow Networks (GFlowNets) [11], which provide a generative perspective on sequential decision-making by learning a stochastic policy that samples sequences such that the probability of reaching a terminal state is proportional to a given reward. GFlowNets are particularly well-suited for problems requiring diverse and high-quality solutions, such as combinatorial optimization and structured search spaces. Another framework is imitation learning [55], where the goal is to learn policies by mimicking expert demonstrations. Variants like behavior cloning [36] and inverse RL [5] extend sequential decision-making capabilities by leveraging supervised learning and inferring reward structures, respectively. Additionally, probabilistic graphical models [119] integrated with deep learning, such as deep probabilistic programming frameworks, offer structured ways to represent dependencies in sequential tasks, enabling robust reasoning under uncertainty.

This dissertation proposes deep learning-based sequential decision-making frameworks tailored for key challenges in wireless systems, addressing uncertainty, adaptability, and the need for reasoning over extended operational horizons.

1.1 Contributions

The first scenario explored in this dissertation addresses the fundamental challenge of antenna and sensor selection for thin array design. This problem arises across various contexts in wireless communications, sensing, and integrated sensing and communication systems, where both functions are performed on a shared hardware platform and must be jointly optimized. While the performance of array processing improves with an increasing number of deployed elements, this comes at the cost of higher energy consumption and financial expense. A practical solution is to activate only a subset of the deployed elements during operation, presenting the challenge of selecting the optimal subset from a combinatorially large number of possibilities. Furthermore, activating subsets can transform a uniform array into a sparse array, yielding benefits in multiple applications.

Traditionally, antenna and sensor selection has been approached as a discriminative, one-step task, with solutions based on convex optimization, greedy selection, or supervised machine learning methods. This dissertation reframes the problem as a sequential decision-making task, modeling the selection process as a deterministic Markov

Decision Process (MDP) with a single root. Terminal states represent subsets of elements with a specified number of active components, and their rewards correspond to the optimization objective evaluated for each subset. To address this, the GFlowNet paradigm is employed to parameterize an action-sampling policy, ensuring that the probability of reaching a terminal state is proportional to the reward associated with the corresponding subset.

This work has been published in:

- Evmorfos, Spilios, Zhaoyi Xu, and Athina Petropulu. "Gflownets for Sensor Selection." 2023 IEEE 33rd International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2023.

- Evmorfos, Spilios, Zhaoyi Xu, and Athina Petropulu. "Sensor selection via GFlowNets: A deep generative modeling framework to navigate combinatorial complexity." arXiv preprint arXiv:2407.19736 (2024).

- Evmorfos, Spilios, and Athina P. Petropulu. "Generative AI for Sparse Antenna Array Design in ISAC Systems." 2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2024.

- Evmorfos, Spilios and Athina P.Petropulu. "GFlowNet-Based Antenna Selection for ISAC Systems under the Presence of Eavesdroppers"2024 IEEE Asilomar Conference on Signals, Systems and Computers

The second scenario focuses on joint beamforming and relay motion control in mobile relay beamforming networks operating within spatiotemporally varying channel environments. A time-slotted approach is adopted, wherein, during each slot, the relays perform optimal beamforming and determine their optimal positions for the subsequent slot. The problem of relay motion control is formulated within a sequential decision-making framework.

A DRL approach is employed to guide relay motion, aiming to maximize the cumulative Signal-to-Interference-plus-Noise Ratio (SINR) at the destination. Initially, a model-based RL method is presented, where the SINR is estimated predictively, and relay motion is determined based on partial knowledge of the channel model and measurements at the relays' current positions. Subsequently, a model-free deep Q-learning approach is proposed, which does not depend on channel models.

For the deep Q learning method, two modified Multilayer Perceptron Neural Networks (MLPs) are introduced to approximate the value function. The first modification involves applying a Fourier feature mapping to the state before passing it through the MLP. The second modification leverages an alternative neural network architecture that uses sinusoidal activations between layers. Both modifications are shown to enhance the ability of the MLP to learn the high-frequency components of the value function, significantly improving convergence speed and SINR performance.

The work is published in:

- Evmorfos, Spilios, Konstantinos I. Diamantaras, and Athina P. Petropulu. "Reinforcement learning for motion policies in mobile relaying networks." IEEE Transactions on Signal Processing 70 (2022): 850-861.

- Evmorfos, Spilios, Konstantinos Diamantaras, and Athina Petropulu. "Deep q learning with fourier feature mapping for mobile relay beamforming networks." 2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2021.

- Evmorfos, Spilios, and Athina P. Petropulu. "Deep actor-critic for continuous 3D motion control in mobile relay beamforming networks." ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

- Evmorfos, Spilios, Konstantinos Diamantaras, and Athina Petropulu. "Double Deep Q Learning with Gradient Biasing for Mobile Relay Beamforming Networks." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

- Evmorfos, Spilios, Dionysios Kalogerias, and Athina Petropulu. "Adaptive discrete motion control for mobile relay networks." Frontiers in Signal Processing 2

(2022): 867388.

The third scenario examines the design of an IRS to support a Multiple-Input-Single-Output (MISO) communication system operating in a mobile, spatiotemporally correlated channel environment. The design objective is formulated to maximize the expected sum of Signal-to-Noise Ratio (SNR) at the receiver over an infinite time horizon, giving rise to a MDP.

An actor-critic algorithm is proposed for continuous control, which accounts for both channel correlations and destination motion by incorporating the history of destination positions and IRS phases into the state of the RL algorithm. To address the variability of the underlying value function caused by channel fluctuations, the critic's input is preprocessed using a Fourier kernel. This preprocessing enhances stability in the process of neural value approximation.

Additionally, the inclusion of the destination SNR as a component of the MDP state, a common practice in previous works, is investigated. Empirical results demonstrate that, under spatiotemporally varying channels, incorporating the SNR in the state representation leads to divergence. Insight into this divergence is provided by analyzing the impact of SNR inclusion on the Neural Tangent Kernel (NTK) of the critic network. Based on this study, a framework is proposed for designing actor-critic methods for IRS optimization and other general problems, predicated on sufficient conditions of the critic's NTK for convergence under neural value learning.

The work is published in:

- Evmorfos, Spilios, Athina P. Petropulu, and H. Vincent Poor. "Actor-critic methods for IRS design in correlated channel environments: A closer look into the neural tangent kernel of the critic." IEEE Transactions on Signal Processing (2023).

1.2 Outline

In Chapter 2, the GFlowNet-based approach for sensor and antenna selection is detailed, presenting its application across various wireless communication and sensing scenarios. Chapter 3 introduces the DRL framework developed for mobile relay motion control,