Parameterized Reinforcement Learning for Optical System Optimization

Heribert Wankerl, Maike L. Stern
Ali Mahdavi, Christoph Eichler
Elmar W. Lang

Multi-layer optical system generation as parameterized Markov decision processes

In our work, we optimize a multi-layer stack of dielectric materials such that it features a desired reflectivity behavior [1]. The optimization is conducted with respect to the thickness and material of each layer as well as the total number of layers.
This task states a so-called inverse design problem. Therefore, the generation of a multi-layer system is considered as the subsequent stacking of layers, which are determined by their thicknesses and materials. This framework characterizes a parameterized Markov decision process [2] that can be solved by parameterized Reinforcement learning by including:

- **states,** which represent the current multi-layer system
- continuous (thickness) & discrete (material) **actions** of the layers
- a **reward signal** based on how close we achieved the desired behavior

Note that most state of the art methods only optimize for the layer thicknesses of a pre-defined stack of dielectric materials. Our approach overcomes this limitation and takes into account each layer‘s material and the total number of layers as well.

Fig 1: In this work, a multi-layer optical system is defined by the total number of layers as well as each layer’s material and thickness. Our approach optimizes all three parameters, requiring a method able to combine continuous and discrete actions.

Parameterized reinforcement learning

Parameterized reinforcement learning (MP-DQN, [3]) is used to generate a multi-layer system by consecutive parameterized actions. Hereby, the material and thickness of a layer is determined until the algorithm terminates the stacking of layers. In the latter case or if the pre-defined maximum number of layers is reached, the suggested design is rewarded based on the deviation between the observed and required reflectivity. This feedback is used to train the policy based on Q-learning. As a result, the parameterized actions yield multi-layer systems that maximize the observed reward.
The policy consists of two neural networks, as shown in figure 2. The first estimates layer thicknesses based on the current state. The second network estimates Q-values associated to the discrete material decisions based on estimated thicknesses and the current state. In total, this leads to a parameterized action that determines which material at which thickness to place next.

Investigations & Experiments

Constrained optimization

Based on its intended functionality, an optical multi-layer system features a particular reflectivity behavior, as shown in the figure 3, bottom (dotted orange line). Evidently, our approach not only outperforms the reference design (blue line) and fulfils the customer specification (dotted grey line), it also suggests a system design that is easier to manufacture figure 3, top. This is achieved by adding a Lagrange term to the reward computation, which punishes overly complex systems. Namely, very thick layers as well as systems with a high total number of layers.

Reward computation and behavior of Q-values

Due to its quadratic form, using linear transformations of the mean-squared error as a reward signal yields similar reward values for resembling reflectivity behaviors that are close to optimal. Following the Bellman equation, this would impede Q-value reliability and thus decision making. To maintain the discriminability of the rewards, an exponential transformation is proposed, as depicted in figure 4, left.
Aside from their importance for decision making, Q-values impart some optical information with respect to the associated materials. As illustrated (figure 4, right), we found the functional relation between refractive indexes of the materials and the associated Q-value estimates to be monotonic for a particular layer.

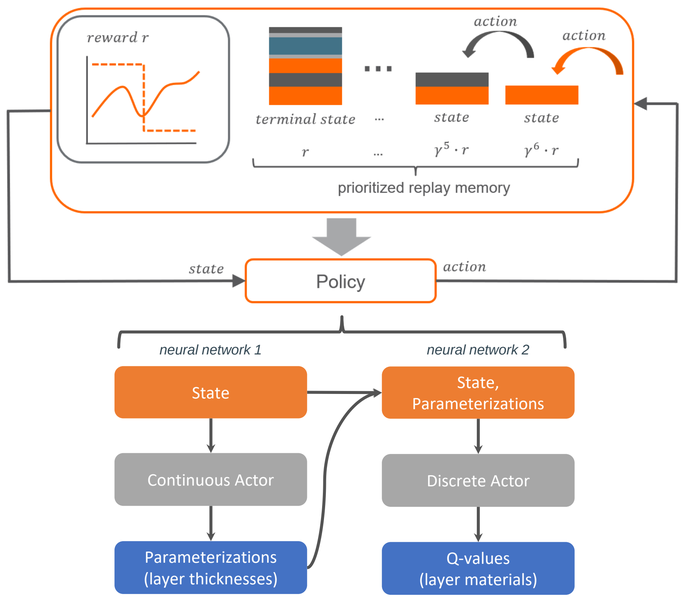

Fig. 2: Outline of the parameterized reinforcement learning algorithm employed in this work.

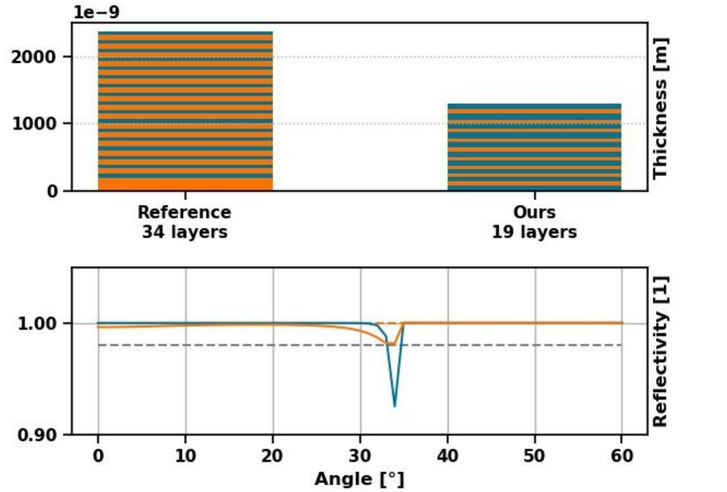

Fig 3: *Top:* Comparison between the multi-layer system suggested by optical experts (top, left) and the system suggested by our algorithm (top, right). Next to an overall thinner design our algorithm requires a lower total number of layers which eases manufacturing.
Bottom: Illustration of optimization regime over angle. Dotted orange: required reflectivity, gray dotted: specification, blue: reflectivity of the reference design, orange: reflectivity of our design.

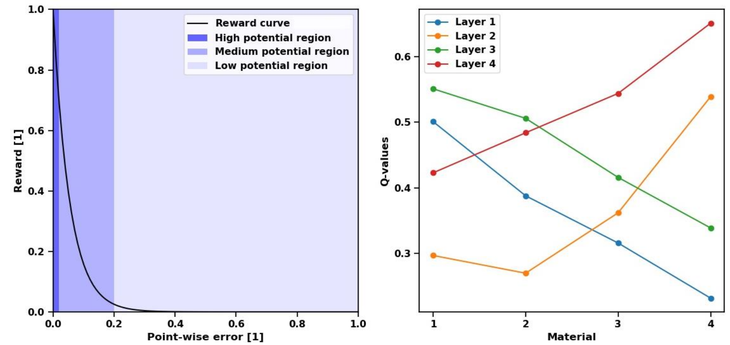

Fig 4: *Left:* Illustration of the reward computation, based on a point-wise error between target and observed reflectivity behavior. *Right:* Analysing the learned Q-values shows a rather systematic behavior, where the Q-value of an associated material monotonically decreases with increasing distance to the optimal material choice. The materials are sorted by refractive indexes in ascending order.

Conclusion

In this work, the generation of multi-layer optical systems is formulated and solved as a parameterized Markov decision process in the presence of both, continuous and discrete parameters.
The incorporation of a Lagrange term allows to conduct constraint optimization based on a specialized reward signal that keeps designs distinguishable especially in advanced optimization phases.

References

[1] H. A. MacLeod (2010). “Thin-Film Optical Filters”. CRC Press

[2] M. Hausknecht (2016). “Deep reinforcement learning in parameterized action space”. In: Proceedings of the International Conference on Learning Representations

[3] C. J. Bester (2019), “Multi-pass q-networks for deep reinforcement learning with parameterised action spaces”. arXiv preprint