MDPI

Review

Reinforcement Learning and Physics

José D. Martín-Guerrero [1,*] and Lucas Lamata [2]

1   IDAL, Electronic Engineering Department, ETSE-UV, University of Valencia,
     Avenida de la Universitat s/n Burjassot, 46100 Valencia, Spain
2   Departamento de Física Atómica, Molecular y Nuclear, Universidad de Sevilla, 41080 Sevilla, Spain;
     llamata@us.es
*   Correspondence: jose.d.martin@uv.es

Abstract: Machine learning techniques provide a remarkable tool for advancing scientific research, and this area has significantly grown in the past few years. In particular, reinforcement learning, an approach that maximizes a (long-term) reward by means of the actions taken by an agent in a given environment, can allow one for optimizing scientific discovery in a variety of fields such as physics, chemistry, and biology. Morover, physical systems, in particular quantum systems, may allow one for more efficient reinforcement learning protocols. In this review, we describe recent results in the field of reinforcement learning and physics. We include standard reinforcement learning techniques in the computer science community for enhancing physics research, as well as the more recent and emerging area of quantum reinforcement learning, inside quantum machine learning, for improving reinforcement learning computations.

Keywords: reinforcement learning; physics; artificial intelligence; machine learning; quantum technologies; quantum machine learning; quantum reinforcement learning

1. Introduction

The ubiquity of bigger and bigger data sets has made machine learning (ML) a common tool for knowledge discovery from those large data sets. The solid theoretical foundations of ML [1,2] have found their application in many different fields [3], physics being no exception [4,5]. In the particular case of quantum information processing [6], there are remarkable attempts that make use of quantum resources to enhance learning, particularly in terms of processing time, than can be reduced considerably, providing relevant speedups, sometimes quadratic or exponential [7].

The application of classical ML to different problems in physics has also become more and more common recently [8–10], even leading to the field of physics-based ML [11,12]. reinforcement learning (RL), the learning paradigm that this review focuses on, has been applied for the control of physical systems [13–16].

The next section will be devoted to go over the main contributions of classical RL to different problems in physics (Section 2.1), and quantum RL (Section 2.2), respectively. Section 3 summarizes this mini-review with some concluding remarks.

2. Reinforcement Learning and Physics

RL is a ML paradigm that optimizes decision making based on stages [17]. While it cannot be considered supervised learning because the desired outputs are not known in advance to train the model, it is not unsupervised or semisupervised learning either because there is not a training limitation due to a lack of labels [2,18].

The RL framework is shown in Figure 1; it consists of an agent that takes actions in a given environment; those actions have an associated immediate reward. The goal of the

learning process is to maximize the long-term reward, a function that depends on the sum of all the rewards that are collected over time:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{1}$$

where the immediate reward r_{t+k+1} is the value returned by the environment depending on the action taken by the agent at time $t + k + 1$ and $\gamma \in [0, 1]$ is the so-called discount-rate, that sets the relative importance of future rewards. The environment is explored by taking different actions, what leads to learning the action-value function or Q-function, that estimates the expected future reward R_t following the policy $\pi(s, a)$:
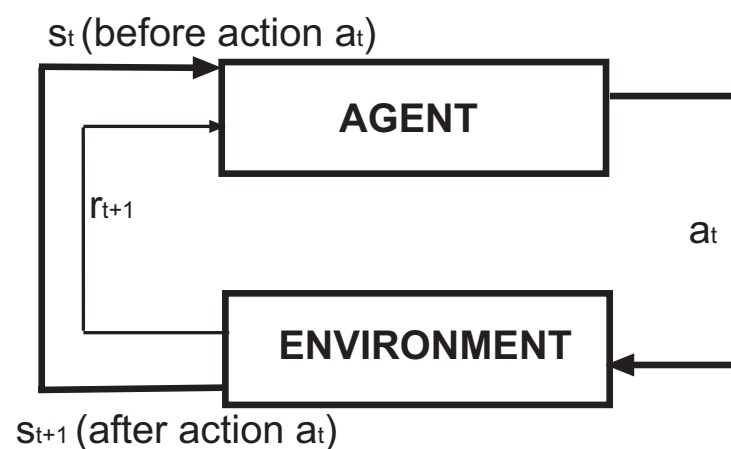
$$Q^\pi(s, a) = E_\pi[R_t | s_t = s, a_t = a] \tag{2}$$

Figure 1. Reinforcement learning framework: a_t is the action taken by the agent at time t, s_t is the state of the environment at time t, and r_{t+1} stands for the reward at time $t + 1$.

The optimal policy appears as the result of learning the Q-function:

$$\pi^*(s, a) = \arg\max_{a \in A} Q^*(s, a) \tag{3}$$

Equation (3) corresponds with a deterministic policy, in which given a state, only one action can be taken. Stochastical policies encompass more than one possible action according to probabilities encoded in the policy [17].

The goal of RL algorithms is to calculate $Q^\pi(s, a)$ in order to obtain the optimal policy using. Many different methods can be considered, basically grouped into three main approaches [17]:

- Dynamic programming (DP): it makes use of the Bellman equation [19] when a complete model of the environment is available.
- MonteCarlo (MC) methods [17]: They do not need a model of the environment but only sample sequences of states, actions and rewards. R_t is computed when an episode finishes, and $Q(s, a)$ accordingly updated.
- Temporal difference methods (TD) [20]: They do not require an environment model, either. The values of $Q_{t+1}(s, a)$ are updated using information from the environment (r_{t+1} and s_{t+1}) as well as estimations of $Q_t(s, a)$.

Sarsa and Q-learning are the most-widely used TD methods. Sarsa is an on-policy algorithm that modifies the starting policy towards the optimal one whereas Q-learning is off-policy and computes the optimal policy while the agent is interacting with the environment by means of another arbitrary policy, Equation (4):

$$Q_{t+1}(s, a) \leftarrow Q_t(s, a) + \alpha[r_{t+1} + max(Q_t(s', a')) - Q_t(s, a)] \tag{4}$$

where α is the rate of the update; s' and a' are the state and the action in time $t+1$, respectively. Q_t stands for the action-value function for a particular state before being visited at time t. Q_{t+1} is the updated value of that state once it has been visited.

RL is a natural approach for system control, that has been successfully applied to many different fields, from robotics [21] to marketing [22] and medicine [23], to name a few. In the case of physics, two main approaches arise, namely, the use of standard RL to control different parts of physical systems (either classical or quantum), and the so-called quantum RL, a quantum version of RL that shares objectives with other quantum ML approaches, i.e., to make use of quantum technologies to carry out ML calculations in a more efficient way. Both approaches are described in the next two subsections.

2.1. Standard Reinforcement Learning for Physics Research

The use of RL has spread across different physics applications in recent years, particularly in quantum physics. One of the first remarkable approaches proposed the application of RL to adaptive quantum metrology [15], in which RL-based control achieved a better control of of quantum processes than standard greedy approaches. In [24], RL demonstrates to be able to find the ground state and describe the unitary time evolution of complex interacting quantum systems. The ability of RL to optimize quantum-error-correction strategies, thus protecting qubits against noise is shown in [25]. In another original work [26], control based on RL shows a similar performance to optimal control methods in many-body quantum systems of interacting qubits.

RL has also been applied in the field of quantum computing for dynamic non-convex optimization in ultra-cold-atom experimentation [14] and measure control in order to facilitate the access to quantum states [13]. Other relevant works to deal with the issue of smart and efficient quantum measures make use of active learning [27,28].

The advent of Deep Learning (DL), that has allowed the resolution of data-driven problems that were unapproachable just a few years ago, has also produced an impact on RL by means of deep policies and DL-based function approximations, leading to the so-called Deep Reinforcement Learning (DRL) [29]. DRL has already been used for efficient measuring of quantum devices [30], for control optimization in quantum state preparation [31], for gate control [32] or for robust digital quantum control to break adiabatic quantum control [16].

Although this subsection has focused on RL to different quantum problems, its application to classical physics is also common. In [33], evolutionary RL was applied to estimate the likelihood of dynamical large deviations, thus showing the suitability of ML in path-extensive physics problems. An interesting review of RL approaches to solve fluid mechanics problems is provided in [34]. RL has also found its application in other fields of physics like optics, e.g., for an adaptive control of astronomy systems [35], or in thermodynamics, to optimize thermodynamic trajectories [36], thus learning previously unknown thermodynamic cycles. An interesting RL-based solution in dynamics is presented in [37], where an efficient sampling of rare trajectories is achieved; in particular, the idea is to make rare events typical so that dynamical behaviors that appear with very low probability in non-equilibrium systems can be accessed in a statistically significant way.

Therefore, we can conclude that standard RL is a common choice for optimization and control in different physics problems, particularly within the quantum realm. Recent RL approaches, like DRL, that enhances RL with DL have rapidly been applied to a number of physics control problems, with some relevant results, as shown in this subsection. Next, the quantum version of RL will be presented in Section 2.2.

2.2. Quantum Reinforcement Learning

Quantum machine learning [38] is an emerging field where the aim is either to employ quantum devices to carry out more efficient ML calculations, or to use ML algorithms to better control and design quantum systems. Inside quantum machine learning, quantum reinforcement learning (QRL) has been explored in the past few years [39–48]. Here the

motivation is to design "intelligent" quantum agents capable of interacting with their environment and adapting to it, by means of quantum resources such as entanglement and superposition.

In [39], a QRL algorithm based on Grover search was introduced. This kind of approach may achieve a polynomial speedup with respect to standard RL algorithms, by means of genuine quantum features such as superposition and entanglement.

The concept of quantum agent was proposed in [40], which analyzed the situation of a quantum agent with a quantum processing unit that interacts via a classical channel with a classical environment. Similarly to [39], the authors showed that a polynomial speedup in the processing of the information acquired from the environment, by means of the quantum processor, could be achieved.

An exponential speedup was predicted in [41] via a quantum oracular environment. This paper also analyzed a general framework for quantum machine learning, involving as well quantum supervised and unsupervised learning.

The possibility to have a speedup in quantum systems is given, in part, by the quantum mechanics properties of the Hilbert space, in which quantum superposition of different states in this space is possible, and quantum superpositions of composite states give rise to entanglement. This quantum parallelism is a crucial ingredient for achieving the quantum speedup in quantum technologies.

An implementation of QRL with superconducting circuits was proposed in [42], for basic protocols involving projective measurements and feedback inside the coherence time of the quantum system. This was extended in [43] to other quantum platforms that may not need to employ feedback, but just projective measurements and ancillary qubits.

Reference [44] introduced a QRL protocol where several copies of an environment state are available, and a quantum agent is able to learn this environment state via succesive measurements on the copies and feedback on its own state, following the outcome of the measurements. A convenient balance between exploration and exploitation was considered to optimize the outcome. This proposal was carried out in the quantum platforms of quantum photonics [45] as well as superconducting circuits [46].

An extension of the previous theoretical work of [44] was proposed in [47] for learning quantum operations instead of quantum states.

Moreover, a review of the field of quantum machine learning, and specifically of QRL, with the quantum platform of quantum photonics was given by [48].

In summary, QRL is an exciting and intriguing field that in some situations may provide a speedup with respect to standard RL algorithms, and in general terms may allow one for improved control and measurement of quantum systems. First steps in this direction have been produced, both in theory and experiments in a variety of quantum platforms, and now the follow up should be further focused on scalability, for aiming at larger quantum agents that may provide a faster speedup. In this respect, one should point out that a different speedup, in small quantum systems, has been obtained, e.g., in [45], in the context of a reduced amount of resources. This is what may be called the "reduced resource scenario", for which a speedup with QRL may be achieved, in this case with respect to standard quantum tomography. Further evidence that a quantum speedup with respect to classical computers may be achieved in this reduced resource scenario was given in an experimental implementation of a quantum memristor with a quantum photonics device [49]. In any case, achieving quantum agents of sizes above 50 qubits or so, will allow one for promising applications in quantum control and ML.

3. Conclusions

In summary, RL is one of the most prominent paradigms in standard ML, and its connection to physics is producing a plethora of interesting results and perspectives inside scientifical and technological discovery. Both in the realms of classical and quantum physics, it is expected that RL will provide an acceleration of the rate of breakthrough achievements. Thus, it will contribute to scientific productivity in a time, nowadays, in which it is more

expensive and harder as most of the "low hanging fruit" has already been taken since the first half of the 20th century.

References

1. Theodoridis, S. *Machine Learning: A Bayesian and Optimization Perspective*; Academic Press: Cambridge, MA, USA, 2020.
2. Shalev-Shwartz, S.; Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms*; Cambridge University Press: Cambridge, MA, USA, 2014.
3. Mathur, P. *Machine Learning Applications Using Python: Cases Studies from Healthcare, Retail and Finance*; APress: New York, NY, USA, 2014.
4. Martín-Guerrero, J.D.; Lisboa, P.J.G.; Vellido, A. Physics and machine learning: Emerging paradigms. In Proceedings of the ESANN 2016 European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 27–29 April 2016; pp. 319–326.
5. Schütt, K.T.; Chmiela, S.; von Lilienfeld, O.A.; Tkatchenko, A.; Tsuda, K.; Müller, K.-R. *Machine Learning meets Quantum Physics*; Lecture Notes in Physics; Springer Nature: Cham, Switzerland, 2020.
6. Nielsen, M.; Chuang, I.L.S. *Quantum Computation and Quantum Information: 10th Anniversary Edition*; Cambridge University Press: Cambridge, MA, USA, 2010.
7. Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* **2020**, *5549*, 195–202. [CrossRef]
8. Muñoz-Gil, G.; Garcia-March, M.A.; Manzo, C.; Martín-Guerrero, J.D.; Lewenstein, M. Single trajectory characterization via machine learning. *New J. Phys.* **2020**, *22*, 013010. [CrossRef]
9. Melko, R.G.; Carleo, G.; Carrasquilla, J.; Cirac, J.I. Restricted Boltzmann machines in quantum physics. *Nat. Phys.* **2019**, *15*, 887–892. [CrossRef]
10. Andreassen, A.; Feige, I.; Fry, C.; Schwartz, M.D. JUNIPR: A framework for unsupervised machine learning in particle physics. *Eur. Phys. J. C* **2019**, *2*, 102. [CrossRef]
11. Swischuk, R.;Mainini, L.; Peherstorfer, B.; Willcox, K. Projection-based model reduction: Formulations for physics-based machine learning. *Comput. Fluids* **2019**, *179*, 704–707. [CrossRef]
12. Ren, K.; Chew, Y.; Zhang, Y.F.; Fuh, J.Y.H.; Bi, G.J. Thermal field prediction for laser scanning paths in laser aided additive manufacturing by physics-based machine learning. *Comput. Meth. Appl. Mech. Eng.* **2020**, *362*, 112734. [CrossRef]
13. Tiersch, M.; Ganahl, E.J.; Briegel, H.J. Adaptive quantum computation in changing environments using projective simulation. *Sci. Rep.* **2012**, *5*, 12874. [CrossRef] [PubMed]
14. Wigley, P.B.; Everitt, P.J.; van den Hengel, A.; Bastian, J.W.; Sooriyabandara, M.A.; McDonald, G.C.;Hardman, K.S.; Quinlivan, C.D.; Manju, P.; Kuhn, C.C.N.; et al. Fast machine-learning online optimization of ultra-cold-atom experiments. *Sci. Rep.* **2012**, *6*, 25890. [CrossRef] [PubMed]
15. Palittapongarnpim, P.; Wittek, P.; Sanders, B.C. Controlling Adaptive Quantum Phase Estimation with Scalable Reinforcement Learning. In Proceedings of the 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN-16), Bruges, Belgium, 27–29 April 2016; pp. 327–332.
16. Ding, Y.; Ban, Y.; Martín-Guerrero, J.D.; Solano, E.; Casanova, J.; Chen, X. Breaking adiabatic quantum control with deep learning. *Phys. Rev. A* **2021**, *103*, L040401. [CrossRef]
17. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 1998.
18. Alpaydin, E. *Introduction to Machine Learning*, 4th ed.; MIT Press: Cambridge, MA, USA, 2020.
19. Bertsekas, D.P. *Dynamic Programming and Optimal Control: Volume II*; Athenas Scientific: Belmont, MA, USA, 2001.
20. Watkins, C.J.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef]
21. Zhang, T.; Mo, H. Reinforcement learning for robot research: A comprehensive review and open issues. *Int. J. Adv. Robot. Syst.* **2021**, *18*, 17298814211007305.
22. Gómez-Pérez, G.; Martín-Guerrero, J.D.; Soria-Olivas, E.; Balaguer-Ballester, E.; Palomares, A.; Casariego, N. Assigning discounts in a marketing campaign by using reinforcement learning and neural networks. *Expert Syst. Appl.* **2009**, *36*, 8022–8031. [CrossRef]
23. Martín-Guerrero, J.D.; Gomez, F.; Soria-Olivas, E.; Schmidhuber, J.; Climente-Martí, M.; Jiménez-Torres, N.V. A Reinforcement Learning approach for Individualizing Erythropoietin Dosages in Hemodialysis Patients. *Expert Syst. Appl.* **2009**, *36*, 9737–9742. [CrossRef]

24. Carleo, G.; Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **2017**, *355*, 602. [CrossRef] [PubMed]

25. Fösel, T.; Tighineanu, P.; Weiss, T.; Marquardt, F. Reinforcement Learning with Neural Networks for Quantum Feedback. *Phys. Rev. X* **2018**, *8*, 031084. [CrossRef]

26. Bukov, M.; Day, A.G.R.; Sels, D.; Weinberg, P.; Polkovnikov, A.; Mehta, P. Reinforcement Learning in Different Phases of Quantum Control. *Phys. Rev. X* **2018**, *8*, 031086. [CrossRef]

27. Melnikov, A.A.; Nautrup, H.P.; Krenn, M.; Dunjko, V.; Tiersch, M.; Zeilinger, A.; Briegel, H.J. Active learning machine learns to create new quantum experiments. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 1221–1226. [CrossRef] [PubMed]

28. Ding, Y.; Martín-Guerrero, J.D.; Sanz, M.; Magdalena-Benedicto, R.; Chen, X.; Solano, E. Retrieving Quantum Information with Active Learning. *Phys. Rev. Lett.* **2020**, *124*, 140504. [CrossRef] [PubMed]

29. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjel, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef]

30. Nguyen, V.; Orbell, S.B.; Lennon, D.T.; Moon, H.; Vigneau, F.; Camenzind, L.C.; Yu, L.; Zumbühl, D.M.; Briggs, G.A.D.; Osborne, M.A.; et al. Deep reinforcement learning for efficient measurement of quantum devices. *npj Quantum Inform.* **2021**, *7*, 100. [CrossRef]

31. Zhang, X.M.; Wei, Z.; Asad, R.; Yang, X.C.; Wang, X. When does reinforcement learning stand out in quantum control? A comparative study on state preparation. *npj Quantum Inform.* **2019**, *5*, 85. [CrossRef]

32. Zheng, A.; Zhou, D.L. Deep reinforcement learning for quantum gate control. *Europhys. Lett.* **2019**, *126*, 60002.

33. Whitelam, S.; Jacobson, D.; Tamblyn, I. Evolutionary reinforcement learning of dynamical large deviations. *J. Chem. Phys.* **2020**, *153*, 044113. [CrossRef] [PubMed]

34. Garnier, P.; Viquerat, J.; Rabault, J.; Larcher, A.; Kuhnle, A.; Hachem, E. A review on deep reinforcement learning for fluid mechanics. *Comput. Fluids* **2021**, *225*, 104973. [CrossRef]

35. Nousiainen, J.; Rajani, C.; Kasper, M.; Helin, T. Adaptive optics control using model-based reinforcement learning. *Opt. Express* **2021**, *29*, 15327–15344. [CrossRef]

36. Beeler, C.; Yahorau, U.; Coles, R.; Mills, K.; Whitelam, S.; Tamblyn, I. Optimizing thermodynamic trajectories using evolutionary and gradient-based reinforcement learning. *arXiv* **2019**, arXiv:1903.08453.

37. Rose, D.C.;Mair, J.F.; Garrahan, J.P. A reinforcement learning approach to rare trajectory sampling. *New J. Phys.* **2021**, *23*, 013013. [CrossRef]

38. Lamata, L. Quantum machine learning and quantum biomimetics: A perspective. *Mach. Learn. Sci. Technol.* **2020**, *1*, 033002. [CrossRef]

39. Dong, D.; Chen, C.; Li, H.; Tarn, T.-J. Quantum Reinforcement Learning. *IEEE Trans. Syst. Man Cybern. B* **2008**, *38*, 1207. [CrossRef]

40. Paparo, G.D.; Dunjko, V.; Makmal, A.; Martin-Delgado, M.A.; Briegel, H.J. Quantum speedup for active learning agents. *Phys. Rev. X* **2014**, *4*, 031002. [CrossRef]

41. Dunjko, V.; Taylor, J.M.; Briegel, H.J. Quantum-enhanced machine learning. *Phys. Rev. Lett.* **2016**, *117*, 130501. [CrossRef]

42. Lamata, L. Basic protocols in quantum reinforcement learning with superconducting circuits. *Sci. Rep.* **2017**, *7*, 1609. [CrossRef]

43. Cárdenas-López, F.A.; Lamata, L.; Retamal, J.C.; Solano, E. Multiqubit and multilevel quantum reinforcement learning with quantum technologies. *PLoS ONE* **2018**, *13*, e0200455. [CrossRef]

44. Albarrán-Arriagada, F.; Retamal, J.C.; Solano, E.; Lamata, L. Measurement-based adaptation protocol with quantum reinforcement learning. *Phys. Rev. A* **2018**, *98*, 042315. [CrossRef]

45. Yu, S.; Albarrán-Arriagada, F.; Retamal, J.C.; Wang, Y.T.; Liu, W.; Ke, Z.J.; Meng, Y.; Li, Z.P.; Tang, J.S.; Solano, E.; et al. Reconstruction of a Photonic Qubit State with Reinforcement Learning. *Adv. Quantum Technol.* **2019**, *2*, 1800074. [CrossRef]

46. Olivares-Sánchez, J.; Casanova, J.; Solano, E.; Lamata, L. Measurement-based adaptation protocol with quantum reinforcement learning in a Rigetti quantum computer. *Quantum Rep.* **2020**, *2*, 293. [CrossRef]

47. Albarrán-Arriagada, F.; Retamal, J.C.; Solano, E.; Lamata, L. Reinforcement learning for semi-autonomous approximate quantum eigensolver. *Mach. Learn. Sci. Technol.* **2020**, *1*, 015002. [CrossRef]

48. Lamata, L. Quantum Reinforcement Learning with Quantum Photonics. *Photonics* **2021**, *8*, 33. [CrossRef]

49. Spagnolo, M.; Morris, J.; Piacentini, S.; Antesberger, M.; Massa, F.; Ceccarelli, F.; Crespi, A.; Osellame, R.; Walther, P. Experimental quantum memristor. *arXiv* **2021**, arXiv:2105.04867.