Optimal Control and Reinforcement Learning for Robot: A Survey

Haodong Feng[1,2], Lei Yu[1,2], and Yuqing Chen[1(✉)]

[1] School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China
Yuqing.Chen@xjtlu.edu.cn

[2] Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, UK

Abstract. Along with the development of systems and their applications, conventional control approaches are limited by system complexity and functions. The development of reinforcement learning and optimal control has become an impetus of engineering, which has show large potentials on automation. Currently, the optimization applications on robot are facing challenges caused by model bias, high dimensional systems, and computational complexity. To solve these issues, several researches proposed available data-driven optimization approaches. This survey aims to review the achievements on optimal control and reinforcement learning approaches for robots. This is not a complete and exhaustive survey, but provides some latest and remarkable achievements for optimal control of robots. It introduces the background and facing problem statement at the beginning. The developments of the solutions to existed issues for robot control and some notable control methods in these areas are reviewed briefly. In addition, the survey discusses the future development prospects from four aspects as research directions to achieve improving the efficiency of control, the artificial assistant learning, the applications in extreme environment and related subjects. The interdisciplinary researches are essential for engineering fields based on optimal control methods according to the perspective; which would not only promote engineering equipment to be more intelligent, but extend applications of optimal control approaches.

Keywords: Data-driven optimization · Optimal control · Reinforcement learning · Model bias

1 Introduction

Optimal control is to make the performance index achieve optimal state, which can find an optimal channel to achieve predetermined destination [1]. The performance indexes are various based on real systems and missions. Optimal control

This work was supported by the Research Development Fund RDF-20-01-08 provided by Xi'an Jiaotong-Liverpool University.

H. Gao and X. Wang (Eds.): CollaborateCom 2021, LNICST 406, pp. 54–66, 2021.
https://doi.org/10.1007/978-3-030-92635-9_4

has two major principles on solving such issues which are the Pontryagin's minimum principle (PMP) and the dynamic programming (DP) [2]. PMP provides a requisite condition for optimality. Meanwhile, DP provides an ample condition for optimality via solving a differential equation, which is known as the Hamilton-Jacobi-Bellman (HJB) equation. They are the basis of optimal control theory and most algorithms were derived and designed based on PMP and DP approaches. In the following mentioned controllers, the ideas of PMP and DP were involved and applied. As one of the key components in optimal control algorithms, the cost functions or performance index functions are essential for optimization process. In the design of optimal controllers, some performance indexes are described in quadratic problems with destination state, process state, and action sections which are called linear quadratic regulator problems (LQR). However, along with the development of industry and technology, preceding conventional optimal control approaches are offline and they need complete model of system dynamics [3], it is hard to employ these kinds of optimal control solutions with dynamics uncertainties and changes as a result.

Currently, the widely utilized data-driven optimization algorithms of robot control are developed according to reinforcement learning (RL) which is known as a goal-oriented learning method. The agent with RL can learn an action control strategy to modify a long-term reward via interacting with outside environment [4]. At each iteration, an agent with RL evaluate feedback about the performance used state and action, to improve the performance in next step action. The purpose of strategies is to maximum the expected reward. Model-free reinforcement learning (MFRL) method has been employed to solve a wide list of robot tasks, such as playing graphical games and learning locomotion works. MFRL is generally applicable, require relatively fine tuning, and do not try to construct environment model [5]. However, such real-time model-free learning algorithm requires higher complexity of sampling and millions of samples are necessary if great performance must be achieved [6]. On the other hand, model-based reinforcement learning (MBRL) has much more sample efficiency and it uses a learned model via sample to assist learning [7]. To achieve great efficiency, MBRL has often utilized either function approximators or Bayesian models which can fit dynamics using few samples [8].

An advanced form of RL is called approximate DP which is also a solution methodology of optimal control. The purpose of MBRL is to find maximum reward or minimum cost in iterations as same as optimal control. MBRL connects the disparity between classical optimal control theory and adaptive control strategies. The objective is to learn the optimal strategy and cost function for an unknown dynamics. Unlike classical optimal control, MBRL finds the solution to the HJB equation online in real time. And MBRL algorithms are optimal, unlike traditional adaptive controllers that are not usually designed to be optimal in the sense of minimizing cost function. As a result, MBRL which can also be called optimal control has become an efficient and available approach on robot control [2].

There are three issues on robot control. Firstly, the purpose of optimal control, which is to achieve the optimal performance, is the major limitation of it, many researches focus on the precision and efficiency of control [6,8–11]. The second is capacity of computation which limits the algorithm application. The real time MFRL requires a large amount of computation and strong processor, so that it is hard to be employed on real-time robots [12–14]. Finally, the preceding model bias while model learning is also a problem, which is emphasized by many researchers who attempt to reduce the model bias by policy optimization [3,15–18].

The main objective of this paper is to investigate the advantages and disadvantages of optimal control and RL approaches for robots based on our understanding of present and recent advanced automatic techniques in the references. This paper can be applied to guide the design of efficient controllers from data-driven learning aspect for the automatic robots. The rest of this paper is organized as follows: Sect. 2 presents the general statement of optimal control problems' formations. Section 3 reviews the previous published solutions facing for robot optimal control design, regarding the improvement of precision and system complexity, the reduction of model bias, and the decrease of computation. Section 4 discusses the future researches directions to improve the performance of the robot optimal control applications, and the interdisciplinary development among optimal control and other engineering areas, different from the previous published related reviews. In Sect. 5, the main conclusions and contributions are described.

2 Optimal Control Problem Statement

The optimal control problem has fixed formation, this section will discuss cost function defining certain optional control problems. The optimal algorithms aim to solve such constrained optimization problem with cost function [1]

$$\min_{u(\cdot)} I[u(\cdot)] = \int_0^T L(x(t), u(t), t)dt \tag{1}$$

$$\text{Subject to : } \dot{x}(t) = \mathrm{f}(x(t), u(t), t) \text{ and } x(0) = x_0.$$

where $t \in \mathrm{T} = [0, T]$ represents the time interval, $T \in \mathrm{R}_+$ denotes the finally terminal time, $x(t) = [x_1(t), \ldots, x_n(t)]^T \in \mathrm{R^n}$ denotes the state and $u(t) = [u_1(t), \ldots, u_m(t)]^T \in \mathrm{R^m}$ denotes the control input at time $t, u(\cdot)$ denotes the control function, $I[u(\cdot)]$ denotes the total cost, $L(x(t), u(t), t)$ represents the running cost, $\mathrm{f}(x(t), u(t), t)$ denotes the dynamics of the system while x_0 denotes the initial state.

From a predefined initial state $x(\cdot)$, optimal policy can search action $u(\cdot)$ with bounded constrains to make cost index minimum. The Lagrange function is often employed as quadratic form in LQR problem [9]

$$L(x(t), u(t), t) = \frac{1}{2}\left(x^T Q x + u^T R u\right) \tag{2}$$

where $x \in \mathrm{R}^n, u \in \mathrm{R}^m, u \in \mathrm{R}^m, Q \in \mathrm{R}^{n \times n}, R \in \mathrm{R}^{m \times m}$.

If the terminal is predefined, the cost function often has a terminal cost weight ϕ

$$\min_{u(\cdot)} I[u(\cdot)] = \min_{u(\cdot)} \phi(x(T), T) + \int_0^T L(x(t), u(t), t)dt \tag{3}$$

ϕ is terminal cost weight. It would be predefined bigger than Q and R, if the destination is expected to achieve rapidly. The aim of optimal control algorithms is to minimum the cost using optimal control sequences, although they pay attention to different positions.

3 Solutions of Optimal Control for Robot

Several successful stories presented available solutions for problems caused by model bias, complex system, and large computation, which are developments of optimal control.

3.1 Overview the Related Approaches

The efficiency is of core important in optimal problem and several success stories are performed [19,20]. Supervised learning, the precise of which is better with preset labels, sometimes is combined with MBRL to improve efficiency and evaluate outside environment for DP [6]. A crucial choice for any practical utilization is advanced policy search algorithms which can be used to find optimal policy for complex system. Along with the system efficiency, the complexity of dynamic system is always a fundamental task in the field of robotics and engineering, and a variety of methods to solve it have been presented [6,21]. For the simple linear dynamic system, it is easier to find optimal solution, the dynamics formation is $\dot{x}(t) = Ax(t) + Bu(t)$ [22]. However, the nonlinear system is facing large tough issues as the dynamics function is uncertain. [15] uses linear formation to represent the unknown nonlinear dynamic system, but the formation of linear function has bias when it is utilized to express nonlinear dynamics. For many physical applications, the degree of difficulty is increasing along with the degree of freedom in nonlinear system.

Aforementioned applications in nonlinear systems with unknown dynamics, the first step of algorithms is often to learn unknown dynamics model called model learning. There are a large number of stories involved model leaning for optimal control and MBRL using multi-layer neural networks and Gauss mixed model (GMM) [6,9,10,23], however precise model of nonlinear high-dimensional dynamics cannot be learned directly. As learned model of dynamics leads to harmful error which influences policy improvement largely [3], the method of model learning can be optimized so that policy evaluation and policy update can be hold based on precise dynamics model architecture. Another method is to modify policy improvement process. Algorithms can interact with environment each iteration and continuously optimize policy with measured data to

reduce model bias. In several success stories, a large amount of measured samples of real system are employed to learn precise model [5,24,25] which improves computation of processor and adds burdens on operation. Improving sample efficiency has become an essential task in practice.

Some algorithms needing large computation capacity are available in simulation environment with high performance processors. However, they may meet low efficiency and poor performance in practice on robot system. Conventional model predictive control (MPC) [26] needs to predict predefined steps actions at each iteration [16] and compared with MPC several algorithms reduce computational complexity to reduce optimization time and improve the efficiency. The core problems of optimal control and MBRL on robot has been reviewed from three aspects, which hinder the developments and applications of robot in practice. Some excellent algorithms will be discussed in the following sections about the solutions of aforementioned optimal control issues.

3.2 Improve Precision and System Complexity

The degree of freedom (DoF) is often an essential factor which influences the performance of robot. It is hard to model and control a high-dimensional robot whose system is much more complex. [9] proposed an optimal control algorithm with learned local model, which can be employed to complete dexterous manipulation using a tendon-drive 24 DoF robot hand. The nonlinear system with high DoF faces challenging issue with less sample data and its dexterous manipulation needs higher dimension states and actions compared with single works. At the beginning of algorithm, time-varying linear-Gaussian model is learned and fitted from little size sample applying linear regression approach by a Gaussian mixture model, as the control law is locally linear at each time step although the controllers are time-varying. Time-varying linear-Gaussian dynamics are formed as Eq. (4)

$$p\left(x_{t+1} \mid x_t, u_t\right) = \mathrm{N}\left(f_{xt} x_t + f_{ut} u_t + f_{ct},\ \mathrm{F}_t\right) \tag{4}$$

KL-constrained optimization with LQR method and line search method are used to update and optimize controller at each iteration. This solution can complete complex work in high-dimension system and the precision of algorithm is great as same as doing the dexterous manipulation by a human being.

[6] used model-based deep reinforcement learning method to represent MFRL to reduce the value test sample. It utilizes a medium-sized deep neural network which was trained by gradient decent, to learn the system model. Then MPC method is employed to complete the task. To overcome the bias of model and improve the efficiency of MFRL, it combines these two learning strategies. The MBRL is used to initial the model free learner so that MFRL (policy gradient) algorithm just need fewer trials to achieve the destination. The algorithm was verified that it has larger rewards and higher precision than that using model-based approach for four different agents.

Approach in [10] is quiet similar as [6] which used the same model learning method (deep neural network). However, it merged computer vision into the RL. At the begin of the algorithm, two channels are input, one is the state and action channel to learn system model, another is the image which is input to the convolutional neural network (CNN) so that the algorithm can judge various surfaces outside. This approach can make the legged millirobots perform better in different surfaces. The image-conditioned approach can achieve less cost and high precision in different surfaces compared with other MBRL and optimal control methods.

Based on the learning system model, a neural-optimal control structure was proposed by [27] for the unknown discrete nonlinear system with discount factor in the cost function. For nonlinear discrete systems with additive disturbances, a robust model predictive control (MPC) method by self triggering was proposed in [28]. The method has an adaptive predictive time-domain scheme and can stabilize the system while ensuring suboptimal convergence. A novel error bound was designed by work of [29] based on the work in [30], it showed that a model's ability to self-correct is more tightly related to MBRL performance. The model of MBRL can guarantee the robust to model class limitations. A fault-tolerant control method of actuator based on tube-based MPC and set-based fault detection was proposed in [31]. It implemented an active fault isolation method after fault detection with the constraint-handling ability of MPC. The aforementioned researches tried and proposed several available approaches to improve the accurate of the state-of-the-art algorithms from various aspects and directions.

3.3 Overcome Model Bias

As solutions in which are used to improve precision of controllers, several algorithms were proposed to overcome model bias of the nonlinear systems. There are two directions on solving this problem. By using an advanced and efficient model learning approach, more exact model can be trained and learned via deep neural network or linear-Gaussian approximation. Another approach is to employ precision policy iteration algorithms which can optimize strategies according to interaction with environment at each iteration.

In work of [32], a new internal model control (IMC) structure based on fuzzy model was proposed to provide efficient and robust control achievement. A command filtered robust controller was proposed in [33]. In the presence of parameter uncertainty and interference, the given reference signal is tracked by adjusting the aircraft attitude angle. The method uses command filtering inversion to compensate the dynamic error and filtering error of the actuator. The improved stable linear filter is used to deal with the measurement error. [34] presented an approach to find area control error signal based on the frequency biased estimation, so that the performance of load-frequency control system can be improved. Aiming at the speed control problem of constrained nonlinear electric vehicles, a new nonlinear optimal model predictive controller design method was proposed in [35]. In terms of system dynamics, the proposed method can be regarded as a general case of the obtained results.

In order to realize iterative control, an improved embedded reinforcement learning algorithm was proposed by research [36]. The algorithm used off-policy learning to make the dynamics completely unknown, so as to reduce the influence of unknown disturbances and add disturbance compensation controller. For nonlinear continuous systems with state and input constraints, an event model-based predictive control approach was proposed by [37] to solve the uncertain problems such as random time delay caused by communication medium instability.

The effective learning of robot (humanoid) physical model was studied in work [38]. Aiming at the learning problem of body model, an active learning method was proposed. As the learning of serial robot kinematic model, the recursive least square method (RLS) is used to complete the learning process online, which is better than the commonly used gradient method.

Although the model learning methods have been developed by some success stories, the models of complex and high-dimension nonlinear systems are still hard to learn or need tremendous effort to complete. Several researches paid attention on using optimal control approaches to achieve expected objective. [3,16] were representatives of optimal feedback control algorithms. [16] presented an iterative online optimal feedback approach to solve optimal control problem for the continuous linear or nonlinear systems. A model can be employed to compute the variables for cQP using offline calculation. And the measured data are collected to compute optimal action set online. This method can overcome model bias caused by inexact model and have high efficiency to find optimal control policy. In addition, it can also be employed on discrete system after improvement. If model should be learned from the system discretely using precision model learning method, it would reduce the model bias and be more rapidly find the optimal control solution.

3.4 Reduce Computation

[16] tried to reduce computational complexity during iteration process, which is lower than the nonlinear MPC controller. Computation is an essential factor when implement algorithms in robot equipment caused by limitations of processor. MFRL approaches need much computation as high performance computer, which face limitations on real-time robot.

[12] proposed an adaptive moment estimation (Adam) approach, which is an effective random optimization method requires little memory and first-order gradients. It is computationally efficient and has little memory requirements. This algorithm combines two methods AdaGrad [39], which is very suitable for sparse gradient and RMS Prop, and has good on-line and non-stationary effects.

Conventional MPC method can be optimized to improve efficiency and reduce computational complexity. [13] showed an adaptive MPC algorithm for the linear systems with parameter uncertainty constraints. The proposed approach uses RLS based estimator to learn the unknown systems. And it extended the robust MPC controller in [40] to allow online model adaptation while ensuring closed-loop stability and recursive feasibility. The computational complexity

was concerned by research and an adaptive method is given with the reduced computational complexity as well.

[41] presented an arbitrary time control algorithm with limited processor resources, which calculates the components of the control input vector in order to maximize the available processing resources at each time step. To reduce the computational complexity, in the optimal feedback control system, two heuristic algorithms based on greedy search and multiplier alternating direction were proposed by work of [42]. [43] provided a problem solution from initial state to final state for stochastic linear systems controlling linear noise, and minimized the mean square deviation of the target state. The ability of offline computation in this algorithm can reduce computation during process. The aforementioned approaches are the solutions to reduce computational quantity and complexity.

4 Future Prospects and Discussion

Although the approaches in optimal control and MBRL are facing the aforementioned issues, the development is still prosperous as the prospect is bright, and their applications are very essential and wide. Various methods are proposed to overcome these numerous problems and adapt to different application scenarios. Some surveys about reinforcement learning based robotics were reviewed by works of [44–49]. According to these surveys, this paper discusses four directions which can be foreseen in the development of optimal control at present.

To overcome the bias caused by inexact model, aforementioned algorithms were presented in Sect. 3. However, more efficient and optimal policies are looked forward to explore by researches. This is the trend on the development of optimal control and MBRL algorithms. Meanwhile, less sample size and high sample efficiency are pursued as it can reduce the computational complexity and improve the efficiency of controller. If less sample are necessary during model learning, the controller would be easier to be employed in different environments widely. Therefore, more frequent interaction between the agent and environment will be necessary during optimization process to learn and know the parameters better. Optimal control policy can be learned via small sample size by robot as same as that via large sample size. On the other hand, model bias can be reduced using efficient model learning methods and policy update algorithms.

Let robot learn to complete works is hard to apply in real-time practice as numerous limitations in technique. However, if the state-of-the-art algorithms would be used in industrial practice, the efficiency of production would be highly improved and cost would be reduced a lot. Thus, combining actuality with the state-of-the-art control methods has demand prompt solution. It is unnecessary to make intelligent robot complete the whole complex missions in recent years. The help from human being is available when optimal control algorithms are employed in complex nonlinear industrial systems and artificial assistant learning method can be developed in several applications. Just like teachers at school when humans are young, humans act as teachers of robot at the beginning of learning certain works. Apprenticeship learning [50] is one of the example.

Humans can also do partial works and make robot learn another easy and tedious part of works. By this way, human and robot would complete works together and the efficiency of production would be improved.

Although the researches on optimal control theory applied in common environment have been held frequently, the applications of optimal control algorithms in complex, extreme, and special environment are also needed to be explored such as applications in marine technology, deep-sea equipment, and astronautics fields. The environments are more nonlinear complex and special compared with that in the ground. More parameters and factors are necessary to consider when algorithms of controllers are designed. Meanwhile, robot arm and other automatic devices are urgent needs during the scientific investigation in deep-sea and space. For example, Remotely Operated Vehicle (ROV), which is widely used in deep-sea investigation, needs cable to connect through water that constrains the range of ROV activity and increases cost of the equipment. Intelligent Autonomous Underwater Vehicle (AUV) can represent ROV in several suitable works, with high efficient MBRL algorithms it can learn complicated environment in deep-sea [51,52]. However, novel approaches of optimal control are necessary to develop so that extreme external environment can be considered. With optimal control optimization on robot, these types robots would work more efficient, which is helpful for conduct of research in extreme environment.

Optimal control algorithm is not an independent application, whose development has been influenced by computer science. Although novel optimal control methods are often used on robot system to show their availability, they can be utilized by other fields with data-driven optimization as well, like similar control methods. These applications involve smart grid [15], sustainable architecture [53], energy generation [54], and multidisciplinary optimization [55]. Meanwhile, optimal control is widely used in fault prediction and diagnosis [31,56–58] for industrial process control. Researches in interdisciplinary subjects can prompt the development in engineering and extend applications of optimal control algorithms. For instance, optimal control approaches can help power system keep stable voltage and frequency states when faults happening on any nodes via state evaluation and decision. Applications in related fields are one of the directions of optimal control development for existing algorithms and novel algorithms.

The aforementioned four research directions are about bias reduction, approach for applications on industrial robots, applications on complex environmental robots, and applications on other related fields, which are based on different aspects. The direction for bias reduction is a mission on general algorithms themselves, which can promote the generation and improvement of advanced optimal control algorithms. The second direction is an approach to apply the state-of-the-art algorithms on industrial practice by the way of cooperation. The third and forth directions are the applications in different environment and subjects. Although they seem to have little to do with the optimal control algorithm, the application is also an essential part of the development of optimal control. The applications in aforementioned fields are not sample transplant, they need to be tuned according to control objectives and different cases. Trying different

applications can also help adjust the performance of optimal controllers, which has a positive impact for the development of optimal control. Thus, this survey regards them as the future prospective directions.

5 Summary and Conclusions

This survey describes the development and applications of the previous optimal control algorithms on robot technology, and the comparison and analysis of different kinds of control approaches using for the automatic robots. In addition, three main problems of today's algorithms are discussed and several successful stories are reviewed to solve these kinds of problems. The survey discusses the future prospective directions of optimal control and MBRL which include overcoming limitations, artificial assistant learning, applications in complex environment, and interdisciplinary development in other data-driven optimization fields. With the development of advance optimization approaches, more problems on optimal control and RL will be addressed, which will be applied more widely in real-time industrial practice as well. Through this survey, the following conclusions can be drawn.

1. Researches will pay attention on the reduction of model bias and the improvement of the optimization efficiency. More precise modeling methods and more efficient control algorithms with little computational complexity will be developed for high-dimensional robot systems.
2. The optimal control approaches will be used in the other areas' control widely, and their applications will make the systems achieve the optimal solution and have better performance.
3. The applications on complex and extreme environmental robots will help to carry out researches in extreme environment regarding the deep-ocean and space for natural science exploration.
4. Artificial assisted learning applications for the industrial robots are available ways to apply the advanced optimal control algorithms in practice.

Acknowledgment. This work was supported by the Research Development Fund RDF-20-01-08 provided by Xi'an Jiaotong-Liverpool University.

References

1. Chen, Y., Roveda, L., Braun, D.J.: Efficiently computable constrained optimal feedback controllers. IEEE Rob. Autom. Lett. **4**(1), 121–128 (2019)
2. Kober, J., Bagnell, J.A., Peter, J.: Reinforcement learning in robotics: a survey. Int. J. Rob. Res. **32**(11), 1238–1274 (2013)
3. Chen, Y., Braun, D.J.: Hardware-in-the-loop iterative optimal feedback control without model-based future prediction. IEEE Trans. Rob. **35**(6), 1419–1434 (2019)
4. Rastogi, D., Koryakovskiy, I., Kober, J.: Sample-efficient reinforcement learning via difference models. In: 3rd Machine Learning in Planning and Control of Robot Motion Workshop at ICRA (2018)

5. Silver, D., et al.: Mastering the game of Go with deep neural networks and tree search. Nature **529**(7587), 484–489 (2016)
6. Nagabandi, A., Kahn, G., Fearing, R.S., Levine, S.: Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 7559–7566 (2018)
7. Kupcsik, A., Deisenroth, M.P., Peters, J., Loh, A.P., Vadakkepat, P., Neumann, G.: Model-based contextual policy search for data-efficient generalization of robot skills. Artif. Intell. **247**, 415–439 (2017)
8. Deisenroth, M., Rasmussen, C.E.: PILCO: a model-based and data-efficient approach to policy search. In: 28th International Conference on Machine Learning (ICML), pp. 465–472 (2011)
9. Kumar, V., Todorov, E., Levine, S.: Optimal control with learned local models: application to dexterous manipulation. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 378–383 (2016)
10. Nagabandi, A., et al.: Learning image-conditioned dynamics models for control of underactuated legged millirobots. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4606–4613 (2018)
11. Schulman, J., Levine, S., Abbeel, P., Jordan, M., Moritz, P.: Trust region policy optimization. In: 31th International Conference on Machine Learning (ICML), pp. 1889–1897 (2015)
12. Kingma, D.P., Ba, J.: ADAM: a method for stochastic optimization. In: 2015 International Conference for Learning Representations (ICLR) (2015)
13. Zhang, K., Shi, Y.: Adaptive model predictive control for a class of constrained linear systems with parametric uncertainties. Automatica **117**, 108974 (2020)
14. Rottmann, A., Burgard, W.: Adaptive autonomous control using online value iteration with gaussian processes. In: 2009 IEEE International Conference on Robotics and Automation (ICRA), pp. 2106–2111 (2009)
15. Vrabie, D., Pastravanu, O., Abu-Khalaf, M., Lewis, F.L.: Adaptive optimal control for continuous-time linear systems based on policy iteration. Automatica **45**(2), 477–484 (2009)
16. Chen, Y., Braun, D.J.: Iterative online optimal feedback control. IEEE Trans. Autom. Control **66**(2), 566–580 (2021)
17. Losey, D.P., McDonald, C.G., O'Malley, M.K.: A bio-inspired algorithm for identifying unknown kinematics from a discrete set of candidate models by using collision detection. In: 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob), pp. 418–423 (2016)
18. Saputra, A.A., Wi Tay, N.N., Toda, Y., Botzheim, J., Kubota, N.: Bézier curve model for efficient bio-inspired locomotion of low cost four legged robot. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4443–4448 (2016)
19. Morton, J., Witherden, F.D., Jameson, A., Kochenderfer, M.J.: Deep dynamical modeling and control of unsteady fluid flows. In: 2018 Conference on Neural Information Processing Systems (NIPS) (2018)
20. Corneil, D., Gerstner, W., Brea, J.: Efficient model-based deep reinforcement learning with variational state tabulation. In: 35th International Conference on Machine Learning (ICML), pp. 1049–1058 (2018)
21. Lioutikov, R., Paraschos, A., Peters, J., Neumann, G.: Sample-based informationl-theoretic stochastic optimal control. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 3896–3902 (2014)

22. Yaghmaie, F.A., Braun, D.J.: Reinforcement learning for a class of continuous-time input constrained optimal control problems. Automatica **99**, 221–227 (2019)
23. Levine, S., Wagener, N., Abbeel, P.: Learning contact-rich manipulation skills with guided policy search. In: 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 156–163 (2015)
24. Goedhart, M., Van Kampen, E.J., Armanini, S.F., de Visser, C.C., Chu, Q.P.: Machine learning for flapping wing flight control. In: 2018 AIAA Information Systems-AIAA Infotech @ Aerospace (2018)
25. Jordan, M.I., Rumelhart, D.E.: Forward models: supervised learning with a distal teacher. Cogn. Sci. **16**(3), 307–354 (1992)
26. Åkesson, B.M., Toivonen, H.T.: A neural network model predictive controller. J. Process Control **16**(9), 937–946 (2006)
27. Liu, D., Wang, D., Zhao, D., Wei, Q., Jin, N.: Neural-network-based optimal control for a class of unknown discrete-time nonlinear systems using globalized dual heuristic programming. IEEE Trans. Autom. Sci. Eng. **9**(3), 628–634 (2012)
28. Sun, Z., Dai, L., Liu, K., Dimarogonas, D.V., Xia, Y.: Robust self-triggered MPC with adaptive prediction horizon for perturbed nonlinear systems. IEEE Trans. Autom. Control **64**(11), 4780–4787 (2019)
29. Talvitie, E.: Self-correcting models for model-based reinforcement learning. In: 31 Conference on Artificial Intelligence (AAAI), pp. 1–12 (2017)
30. Talvitie, E.: Model regularization for stable sample rollouts. In: the 30th Conference on Uncertainty in Artificial Intelligence, pp. 780–789 (2014)
31. Xu, F., Ocampomartinez, C., Olaru, S., Niculescu, S.I.: Robust MPC for actuator-fault tolerance using set-based passive fault detection and active fault isolation. Int. J. Appl. Math. Comput. Sci. **27**(1), 43–61 (2017)
32. Kumbasar, T., Eksin, I., Guzelkaya, M., Yesil, E.: Adaptive fuzzy internal model control design with bias term compensator. In: 2011 IEEE International Conference on Mechatronics, pp. 312–317 (2011)
33. Li, X., Cao, L., Hu, X., Zhang, S.: Command filtered model-free robust control for aircrafts with actuator dynamics. IEEE Access. **7**, 139475–139487 (2019)
34. Daneshfar, F., Mansoori, F., Bevrani, H.: Multi-agent reinforcement learning design of load-frequency control with frequency bias estimation. In: The 2nd International Conference on Control, Instrumentation and Automation (ICCIA), pp. 310–314 (2011)
35. Vafamand, N., Arefi, M.M., Khooban, M.H., Dragicevic, T., Blaabjerg, F.: Nonlinear model predictive speed control of electric vehicles represented by linear parameter varying models with bias terms. IEEE J. Emerg. Sel. Topics Power Electron. **7**(3), 2081–2089 (2019)
36. Song, R., Lewis, F.L., Wei, Q., Zhang, H.: Off-policy actor-critic structure for optimal control of unknown systems with disturbances. IEEE Trans. Cybern. **46**(5), 1041–1050 (2016)
37. Varutti, P., Findeisen, R.: Event-based NMPC for networked control systems over UDP-like communication channels. In: 2011 American Control Conference, pp. 3166–3171 (2011)
38. Martinez-Cantin, R., Lopes, M., Montesano, L.: Body schema acquisition through active learning. In: 2010 IEEE International Conference on Robotics and Automation (ICRA), pp. 1860–1866 (2010)
39. Duchi, J., Hazan, E., Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. **12**(7), 2121–2159 (2011)

40. Fleming, J., Kouvaritakis, B., Cannon, M.: Robust tube MPC for linear systems with multiplicative uncertainty. IEEE Trans. Autom. Control **60**(4), 1087–1092 (2015)
41. Gupta, V., Luo, F.: On a control algorithm for time-varying processor availability. IEEE Trans. Autom. Control **58**(3), 743–748 (2013)
42. Demirel, B., Ghadimi, E., Quevedo, D.E., Johansson, M.: Optimal control of linear systems with limited control actions: threshold-based event-triggered control. IEEE Trans. Control Netw. Syst. **5**(3), 1275–1286 (2017)
43. Jenson, E.L., Chen, X., Scheeres, D.J.: Optimal control of sampled linear systems with control-linear noise. IEEE Control Syst. Lett. **4**(3), 650–655 (2020)
44. Nguyen, H., La, H.: Review of deep reinforcement learning for robot manipulation. In: 2019 3nd IEEE International Conference on Robotic Computing (IRC), pp. 590–595 (2019)
45. Khan, S.G., Herrmann, G., Lewis, F.L., Pipe, T., Melhuish, C.: Reinforcement learning and optimal adaptive control: an overview and implementation examples. Ann. Rev. Control. **36**(1), 42–59 (2012)
46. Polydoros, A.S., Nalpantidis, L.: Survey of model-based reinforcement learning: applications on robotics. J. Intell. Rob. Syst. **86**(2), 153–173 (2017)
47. Bhagat, S., Banerjee, H., Ho Tse, Z.T., Ren, H.: Deep reinforcement learning for soft, flexible robots: brief review with impending challenges. Robotics **8**(1), 4 (2019)
48. Khan, M.A.M., et al.: A systematic review on reinforcement learning-based robotics within the last decade. IEEE Access **8**, 176598–176623 (2020)
49. Zhao, W., Queralta, J.P., Westerlund, T.: Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In: 2020 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 737–744 (2020)
50. Abbeel, P., Coates, A., Ng, A.Y.: Autonomous helicopter aerobatics through apprenticeship learning. Int. J. Rob. Res. **29**(13), 1608–1639 (2010)
51. Cui, R., Yang, C., Li, Y., Sharma, S.: Adaptive neural network control of AUVs with control input nonlinearities using reinforcement learning. IEEE Trans. Syst. Man Cybern. Syst. **47**(6), 1019–1029 (2017)
52. Refsnes, J.E., Sorensen, A.J., Pettersen, K.Y.: Model-based output feedback control of slender-body underactuated AUVs: theory and experiments. IEEE Trans. Control Syst. Technol. **16**(5), 930–946 (2008)
53. Eller, L., Siafara, L. C., Sauter, T.: Adaptive control for building energy management using reinforcement learning. In: 2018 IEEE International Conference on Industrial Technology (ICIT), pp. 1562–1567 (2018)
54. Avila, L., De Paula, M., Carlucho, I., Sanchez Reinoso, C.: MPPT for PV systems using deep reinforcement learning algorithms. IEEE Lat. Am. Trans. **17**(12), 2020–2027 (2019)
55. Nguyen, T., Mukhopadhyay, S.: Multidisciplinary optimization in decentralized reinforcement learning. In: 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 779–784 (2017)
56. Dan, H., et al.: Error-voltage-based open-switch fault diagnosis strategy for matrix converters with model predictive control method. IEEE Trans. Ind. Appl. **53**(5), 4603–4612 (2017)
57. Yu, B., Zhang, Y., Qu, Y.: MPC-based FTC with FDD against actuator faults of UAVs. In: 15th International Conference on Control, Automation and Systems (ICCAS), pp. 225–230 (2015)
58. Kim, K., Raimondo, D.M., Braatz, R.D.: Optimum input design for fault detection and diagnosis: model-based prediction and statistical distance measures. Control Conference. In: 2013 European Control Conference (ECC), pp. 1940–1945 (2013)