1

Introduction

1.1 Human-Robot Interaction Control

If we know the robot dynamics, we can use them to design model-based controllers (See Figure 1.1). The famous linear controllers are: Proportional-Derivative (PD) [1], linear quadratic regulator (LQR), and Proportional-Integral-Derivative (PID) [2]. They use linear system theory, so the robot dynamics are required to be linearized at some point of operation. The LQR [3–5] control has been used as a basis for the design of reinforcement learning approaches [6].

The classic controllers use complete or partial knowledge of the robot's dynamics. In these cases (without considering disturbances), it is possible to design controllers that guarantee perfect tracking performance. By using the compensation or the pre-compensation techniques, the robot dynamics is canceled and establishes a simpler desired dynamics [7–9]. The control schemes with model compensation or pre-compensation in joint space can be seen in Figure 1.2. Here q_d is the desired reference, q is the robot's joint position, $e = q_d - q$ is the joint error, u_p is the compensator or pre-compensator of the dynamics, u_c is the control coming from the controller, and $\tau = u_p + u_c$ is the control torque. A typical model-compensation control is the proportional-derivative (PD) controller with gravity compensation, which helps to decrease the steady-state error caused by the gravity terms of the robot dynamics.

When we do not have exact knowledge of the dynamics, it is not possible to design the previous controllers. Therefore, we need to use model-free controllers. Some famous controllers are: PID control [10, 11], sliding mode control [2, 12], and neural control [13]. These controllers are tuned according to specific plant under certain conditions (disturbances, friction, parameters). When new conditions arise, the controllers do not display the same behavior, even reaching instability. Model-free controllers perform well for different tasks and are relatively

Human-Robot Interaction Control Using Reinforcement Learning, First Edition. Wen Yu and Adolfo Perrusquía.

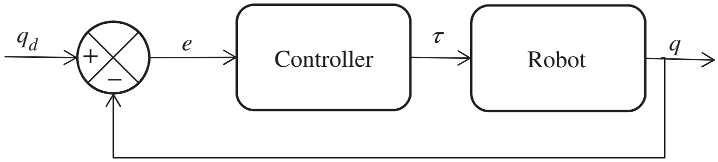

Figure 1.1 Classic robot control

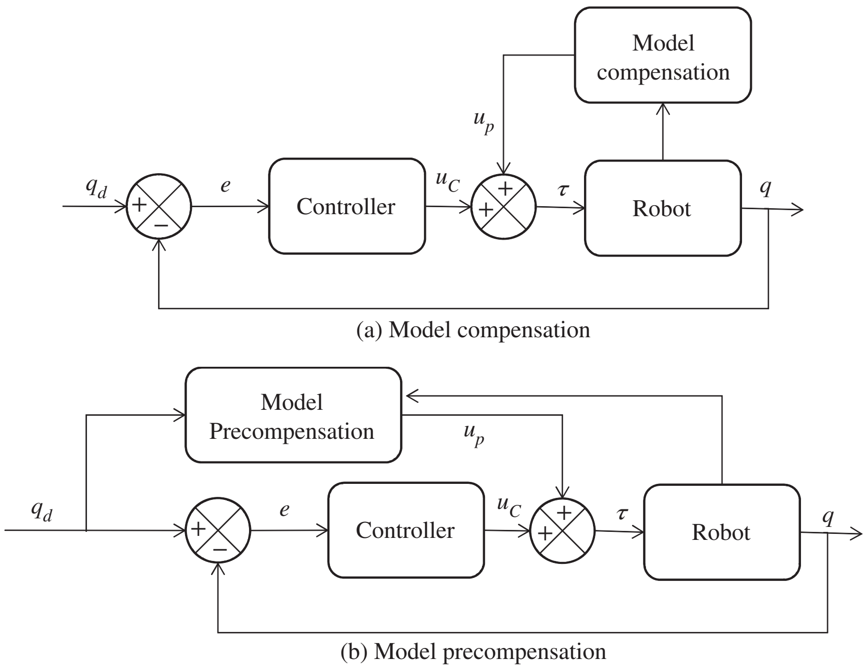

Figure 1.2 Model compensation control

easy to tune; however, they cannot guarantee an optimal performance and require re-tuning the control gains when the robot parameter are changed or a disturbance is applied.

All the above controllers are designed for position control and do not consider interaction with the environment. There is a great diversity of works related to the interaction, such as stiffness control, force control, hybrid control, and impedance control [14]. The force control regulates the interaction force using P (stiffness control), PD, and PID force controllers [15]. The position control can also use force control to perform position and velocity tracking [16, 17] (see Figure 1.3). Here f_d is the desired force, f_e is the contact force, $e_f = f_d - f_e$ is the force error, x_r is the output of the force controller, and $e_r = x_r - x_d$ is the position error in task space.

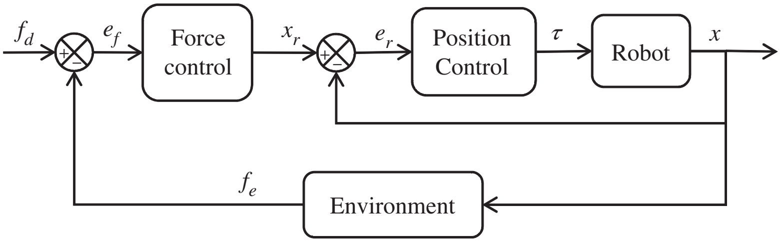

Figure 1.3 Position/force control

The force/position control uses the force for the compensation [17]. It can also use full dynamics to linearize the closed-loop system for perfect tracking [18].

Impedance control [7] addresses the problem of how to move the robot end-effector when it is in contact with the external environment. It uses a desired dynamic model, also known as mechanical impedance, to design the control. The simplest impedance control is the stiffness control, where the stiffness of the robot and the environment have a proportional interaction [19].

Traditional impedance control linearizes the system by assuming that the robot model is known exactly [20–22]. These algorithms need the strong assumption that the exact robot dynamics are known [23]. The robustness of the control lies in the compensation of the model.

Most impedance controllers assume that the desired inertia of the impedance model is equal to the robot inertia. Thus, we only have the stiffness and damping terms, which is equivalent to a PD control law [8, 21, 24]. One way to solve the inaccuracy of dynamic model compensation is through the use of adaptive algorithms, neural networks, or other intelligent methods [9, 25–31]

There are several implementations of impedance control. In [32], the impedance control uses human characteristics to obtain the inertia, damping, and stiffness components of the desired impedance. For the position control a PID control is used, which favors the omission of the model compensation. Another way to avoid the use of the model or to proceed without its full knowledge is to take advantage of system characteristics, that is, the high gear-ratio velocity reduction that causes the non-linear elements to become very small and the system to become decoupled [33].

In mechanical systems, particularly in the haptic field, the admittance is the dynamic mapping from force to motion. The input force “admits” certain amount of movement [11]. The position control based on impedance or admittance needs the inverse impedance model to obtain the reference position [34–38]. This type of scheme is more complete because there is a double control loop where the interaction with the environment can be used more directly.

The applications of impedance/admittance control are quite wide; for example, exoskeletons are used by a human operator. In order to maintain human safety, low mechanical impedance is required, while tracking control requires high impedance to reject the disturbances. So there are different solutions such as frequency molding and the reduction of mechanical impedance using the poles and zeros of the system [39, 40].

Model-based impedance/admittance control is sensitive to modeling error. There exist several modifications to the classical impedance/admittance controllers, such as the position-based impedance control, which improves robustness in the presence of modeling error using an internal position control loop [21].

1.2 Reinforcement Learning for Control

Figure 1.4 shows the control scheme with reinforcement learning. The main difference with the model-free controller in Figure 1.1 is that the reinforcement learning updates its value in each step using the tracking error and control torque.

The reinforcement learning schemes are first designed for discrete-time systems with discrete input space [6, 41]. Among the most famous methods are Monte Carlo [42], Q-learning [43], Sarsa [44], and critic algorithms [45].

If the input space is large or continuous, the classical reinforcement learning algorithms cannot be directly implemented due to the computational cost, and in most cases the algorithm would not converge to a solution [41, 46]. This problem is known as the curse of dimensionality of machine learning. For robot control, the curse of dimensionality increases because there are various degrees of freedom (DOFs), and each DOF needs its own input space [47, 48]. Another factor that makes the dimension problem more acute is the disturbances, because new states and controls must be considered.

To solve the curse of dimensionality, the model-based techniques can be applied to the reinforcement learning [49–51]. These learning methods are very popular; some algorithms are called “policy search” [52–59]. However, these methods require model knowledge to decrease the dimension of the input space.

There are a wide variety of model-free algorithms similar to the discrete-time algorithms. The main idea of these algorithms is to design adequate reward and approximators, which reduces the computational cost in presence of a large or continuous input space.

The simplest approximator to decrease the input space is the handcraft methods [60–65]. They speed up the learning time by looking for regions where the reward is minimized/maximized. [66, 67] use learning methods from input data, similarly to discrete-time learning algorithms, but the learning time increases. Other techniques are based on previously established actions in a sequential and related

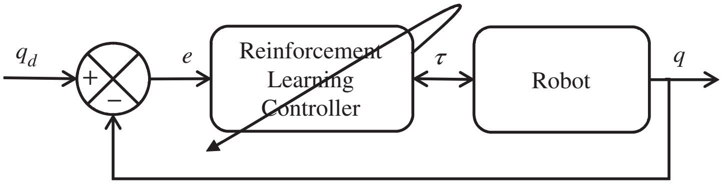

Figure 1.4 Reinforcement learning for control

way; that is, the actions that must be taken at each time instant are defined to do a simple task by themselves [68–72]. The main problems of these methods require an expert knowledge to obtain the best regions and to set the predefined actions.

A linear combination of approximators learn from input data without expert intervention. The most widely used approximators in robot control are inspired by human morphology [73, 74], neural networks [75–77], local models [74, 78], and Gaussian regression processes [79–82]. The success of these approximators is due to the adequate choice of their parameters and hyper-parameters.

A poor reward design can involve a long learning time, convergence to wrong solutions, or the algorithms never converging to any solution. On the other hand, the proper design of a reward helps the algorithm to find the best solutions in each moment of time in a faster way. This problem is known as the "curse of the reward design" [83].

When model-free methods are used, the reward should be designed in such a way that it adapts to changes in the system and possible errors, which is extremely useful in robust control problems where the controller is required to be able to compensate the disturbances or limit disturbances to obtain an optimal performance.

1.3 Structure of the Book

The book consists of two principal parts:

- The first part relates to the design of *human-robot interaction control* in different environments (Chapters 2, 3, 4 and 5).
- The second part deals with *reinforcement learning for robot interaction control* (Chapters 6, 7, 8, 9 and 10).

Part 1

Chapter 2: We address some important concepts for robot interaction control in a mechanical and electrical sense. The concepts of impedance and admittance play important roles for the design of robot interaction control and the environment modeling. The typical environment models and some of the most famous identification techniques for parameters estimation of the environment models are introduced.

Chapter 3: We discuss our first robot-interaction schemes using impedance and admittance controllers. The classical controllers are based on the design of feedback-linearization control laws. The closed-loop dynamics is reduced to a desired dynamics based on the proposed impedance model, which is designed as a second-order linear system. Precision and robustness problems are explained in detail for classical impedance and admittance control. The applicability of these controllers is illustrated by simulations in two different environments.

Chapter 4: We study some model-free controllers that do not need complete knowledge of the robot dynamics. The model-free controllers are designed for an admittance control scheme. The interaction is controlled by the admittance model, while the position controller uses the adaptive control, PID control, or sliding mode control. Stabilities of these controllers are given via Lyapunov stability theory. The applicability of these algorithms is proven via simulations and experiments using different environments and robots.

Chapter 5: We give new robot interaction control scheme known as human-in-the-loop control. Here the environment is the human operator. The human has no contact with the robot. This method uses the input forces/torques of the human operator and maps them into position/orientations of the end-effector via the admittance model. Since the human is in the control loop, she does not know if the applied force/torque yields to singular positions, so it is dangerous for real applications. Therefore, the admittance controllers of the previous chapters are modified to avoid the inverse kinematics, and the Jacobian matrix is modified by using the Euler angles. Experiments illustrate the effectiveness of the approach in both joint and task spaces.

Part 2

Chapter 6: The previous chapters use the desired impedance/admittance model to achieve the desired robot-environment interaction. In most cases, these interactions do not have the optimal performance, they have relative high contact forces or high position errors because they require the environment and robot dynamics. This chapter deals with the reinforcement learning approach for the position/force control in discrete time. The reinforcement learning techniques can achieve a sub-optimal robot-environment interaction.

The optimal impedance model is realized by two different approaches: dynamic programming using a linear quadratic regulator and reinforcement learning. The first one is the model-based control law, the and the second one is model-free. To accelerate the convergence of the reinforcement learning method, the eligibility traces and the temporal difference methods are used. Convergence of the reinforcement learning algorithms is discussed. Stability of the position and force control is analyzed using Lyapunov-like analysis. Simulations and experiments verify the approach in different environments.

Chapter 7: This chapter deals with the large and continuous-time counterpart of the reinforcement learning methods discussed in Chapter 6. Since we are considering big input spaces, the classical reinforcement learning methods cannot handle the problems of the optimal solutions and may not converge. It is required to use approximators to reduce the computational effort and to obtain the reliable optimal or near optimal solutions. This chapter deals with the dimensionality problem in both discrete and continuous time.

We use a parametric approximator based on the normalized radial basis function. The centers of each radial basis function are obtained through the K-means clustering algorithm and random clusters. The convergence of the discrete- and continuous-time versions of the reinforcement learning approximation is analyzed using the contraction property and Lyapunov-like analysis. A hybrid reinforcement learning controller is proposed to take advantage of both discrete- and continuous-time versions. Simulations and experiments are carried out to validate the performance of the algorithms in a position/force control task in different environments.

Chapter 8: We design robust controllers based on a modified reinforcement learning under the worst-case uncertainty, which are robust and present optimal or near optimal solutions. The reinforcement learning methods are designed in both discrete and continuous time. Both methods use a reward designed as an optimization problem under constraints.

In discrete time, the reinforcement learning algorithms are modified using the k-nearest neighbors and the double estimator technique in order to avoid overestimation of the action values. Two algorithms are developed: a large state and discrete action case and a large state-action case. The convergence of the algorithms is analyzed using the contraction property. In continuous time, we use the same algorithms of Chapter 7 under the modified reward. The effectiveness of the robust controllers is proven via simulations and experiments.

Chapter 9: For some kind of robots, such as redundant robots, it is impossible to compute the inverse kinematics or use the Jacobian matrix due to the singularities. This chapter uses the multi-agent reinforcement learning approach to deal with this issue by using only knowledge of the robot forward kinematics. The solutions of the inverse and velocity kinematics of robots and redundant robots are discussed.

To assure controllability and avoid a singularity or multiple solutions, we use the multi-agent reinforcement learning and a proposed double value function method. The kinematic approach is used to avoid the curse of dimensionality. We use small joint displacements as control input until the desired reference is achieved. We discuss the convergence of the algorithm. Simulations and experiments prove the approach with satisfactory results compared with he standard actor-critic methods.

Chapter 10: The reinforcement learning methods that we used in previous chapters learn an optimal control policy from scratch, which translates into large learning time. In order to give previous knowledge to the controller, this chapter gives an $\mathcal{H}_2$ neural control using reinforcement learning in discrete time and continuous time. The controller uses the knowledge of the learned dynamics in order to compute the optimal controller. Convergence of the proposed neural control is analyzed using the contraction property and Lyapunov-like analysis. Simulations are carried out to verify the optimization and robustness of the controller.

Appendices

Appendix A: We discuss some basic concepts and properties of the kinematic and dynamic models of robot manipulators. The dynamics is expressed in both joint space and task space. We also give the kinematic and dynamic models of the robots and systems used in this book by means of the Denavit-Hartenberg convention and the Euler-Lagrange formulation.

Appendix B: We give the basic theory of reinforcement learning and some of the most famous algorithms for controller design. The convergence of the reinforcement learning methods is discussed.

References

1 R. Kelly and V. Santibáñez, *Control de Movimiento de Robots Manipuladores*. Ribera del Loira, España: Pearson Prentice Hall, 2003.

2 A. Perrusquía, W. Yu, A. Soria, and R. Lozano, "Stable admittance control without inverse kinematics," *20th IFAC World Congress (IFAC2017)*, 2017.

3 L. Márton, A. Scottedward Hodel, B. Lantos, and Y. Hung, "Underactuated robot control: Comparing LQR, subspace stabilization, and combined error metric approaches," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 10, pp. 3724–3730, 2008.

4 R. Johansson and M. Spong, "Quadratic optimization of impedance control," *Proc. IEEE International Conference in Robot Autom*, no. 1, pp. 616–621, 1994.

5 M. Matinfar and K. Hashtrudi-Zaad, "Optimization based robot compliance control: geometric and linear quadratic approaches," *Int. J. Robot. Res.*, vol. 24, no. 8, pp. 645–656, 2005.

6 F. Lewis, D. Vrable, and K. Vamvoudakis, "Reinforcement learning and feedback control using natural decision methods to desgin optimal adaptive controllers," *IEEE Control Systems Magazine*, 2012.

7 N. Hogan, "Impedance Control: An Approach to Manipulation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 107, pp. 1–24, March 1985. Transactions of the ASME.

8 F. Ficuciello, L. Villani, and B. Siciliano, "Variable Impedance Control of Redundant Manipulators for Intuitive Human-Robot Physical Interaction," *IEEE Transactions on Robotics*, vol. 31, pp. 850–863, August 2015.

9 A. Khan, D. Yun, M. Ali, J. Han, K. Shin, and C. Han, "Adaptive Impedance Control for Upper Limb Assist Exoskeleton," *IEEE International Conference on Robotics and Automation*, pp. 4359–4366, May 26–30 2015. Seattle, Washington.

10 J. G. Romero, A. Donaire, R. Ortega, and P. Borja, "Global stabilisation of underactuated mechanical systems via PID passivity-based control.," *IFAC-PapersOnLine*, 2017.

11 W. Yu, J. Rosen, and X. Li, "PID Admittance Control for an Upper Limb Exoskeleton," *American Control Conference*, pp. 1124–1129, June 29–July 01 2011. O'Farrel Street, San Francisco, California.

12 R. Xu and U. Özgüner, "Sliding mode control of a class of underactuated systems.," *Automatica*, vol. 44, 2014.

13 Y. Zuo, Y. Wang, X. Liu, S. Yang, L. Huang, X. Wu, and Z. Wang, "Neural network robust H_∞ tracking control strategy for robot manipulators," *Applied Mathematical Modelling*, vol. 34, no. 7, pp. 1823–1838, 2010.

14 S. Chiaverini, B. Siciliano, and L. Villani, "A Survey of Robot Interaction Control Schemes with Experimental Comparison," *IEEE/ASME Transactions on Mechatronics*, vol. 4, pp. 273–285, September 1999.

15 R. Volpe and P. Khosla, "A Theoretical and Experimental Investigation of Explicit Force Control Strategies for Manipulators," *IEEE Transactions on Automatic Control*, vol. 38, pp. 1634–1650, November 1993.

16 M. Heidingsfeld, R. Feuer, K. Karlovic, T. Maier, and O. Sawodny, "A Force-controlled Human-assitive Robot for Laparoscopic Surgery," *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3435–3439, October 5–8 2014. San Diego, CA, USA.

17 S. Chiaverini, B. Siciliano, and L. Villani, "Force/Position Regulation of Compliant Robot Manipulators," *IEEE Transactions on Automatic Control*, vol. 39, no. 3, pp. 647–652, 1994.

18 O. Khatib, "A Unified Approach for Motion and Force Control of Robot Manipulators: The Operational Space Formulation," *IEEE Journal of Robotics and Automation*, vol. RA-3, pp. 43–53, February 1987.

19 M. Tufail and C. de Silva, "Impedance Control Schemes for Bilateral Teleoperation," *International Conference on Computer Science and Education*, pp. 44–49, August 2014. Vancouver, Canada.

20 F. Caccavele, B. Siciliano, and L. Villani, "The Tricept Robot: Dynamics and Impedance Control," *IEEE/ASME Transactions on Mechatronics*, vol. 8, pp. 263–268, June 2003.

21 S. Kang, M. Jin, and P. Chang, "A Solution to the Accuracy/Robustness Dilemma in Impedance Control," *IEEE/ASME Transactions on Mechatronics*, vol. 14, pp. 282–194, June 2009.

22 R. Bonitz and T. Hsia, "Internal Force-Based Impedance Control for Cooperating Manipulators," *IEEE Transactions on Robotics and Automation*, vol. 12, pp. 78–89, February 1996.

23 R. J. Anderson and M. W. Spong, "Hybrid Impedance Control of Robotic Manipulators," *IEEE Journal of Robotics and Automation*, vol. 4, pp. 549–556, October 1988.

24 T. Tsuji and M. Kaneko, "Noncontact Impedance Control for Redundant Manipulators," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 29, pp. 184–193, March 1999. Part A: Systems and Humans.

25 S. Singh and D. Popa, "An Analysis of Some Fundamental Problems in Adaptive Control of Force and Impedance Behavior: Theory and Experiments," *IEEE Transactions on Robotics and Automation*, vol. 11, pp. 912–921, December 1995.

26 W. Lu and Q. Meng, "Impedance Control with Adaptation for Robotic Manipulators," *IEEE Transactions on Robotics and Automation*, vol. 7, pp. 408–415, June 1991.

27 A. Abdossalami and S. Sirouspour, "Adaptive Control of Haptic Interaction with Impedance and Admittance Type Virtual Environments," *IEEE*, pp. 145–152, March 13–14 2008. Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems.

28 M. Chih and A. Huang, "Adaptive Impedance Control of Robot Manipulators based on Function Approximation Technique," in *Robotica* (Cambridge University Press, ed.), vol. 22, pp. 395–403, 2004.

29 R. Kelly, R. Carelli, M. Amestegui, and R. Ortega, "On Adaptive Impedance Control of Robot Manipulators," *IEEE Robotics and Automation*, vol. 1, pp. 572–577, May 14–19 1989. Scottsdale, AZ.

30 V. Mut, O. Nasisi, R. Carelli, and B. Kuchen, "Tracking Adaptive Impedance Robot Control with Visual Feedback," *IEEE International Conference on Robotics and Automation*, pp. 2002–2007, May 1998. Leuven, Belgium.

31 H. Mohammadi and H. Richter, "Robust Tracking/Impedance Control: Application to Prosthetics," *American Control Conference*, pp. 2673–2678, July 1–3 2015. Chicago, IL, USA.

32 T. Tsuji and Y. Tanaka, "Tracking Control Properties of Human-Robotic Systems Based on Impedance Control," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 35, pp. 523–535, July 2005.

33 G. Ferreti, G. Magnani, and P. Rocco, "Impedance Control for Elastic Joints Industrial Manipulators," *IEEE Transactions on Robotics and Automation*, vol. 20, pp. 488–498, June 2004.

34 W. Yu, R. Carmona Rodriguez, and X. Li, "Neural PID Admittance Control of a Robot," *American Control Conference*, pp. 4963–4968, June 17–19 2013. Washington, DC, USA.

35 A. Irawan, M. Moktadir, and Y. Tan, "PD-FLC with Admittance Control for Hexapod Robot's Leg Positioning on Seabed," *IEEE American Control Conference*, May 31–June 3, 2015. Kota Kinabalu.

36 K. Tee, R. Yan, and H. Li, "Adaptive Admittance Control of a Robot Manipulator Under Task Space Constraint," *IEEE International Conference on Robotics and Automation*, pp. 5181–5186, May 3–8 2010. Anchorage Convention District, Alaska.

37 W. Yu and J. Rosen, "A Novel Linear PID Controller for an Upper Limb Exoskeleton," *49th IEEE Conference on Decision and Control*, pp. 3548–3553, December 2010. Atlanta, Hotel, GA, USA.

38 M. Dohring and W. Newman, "The Passivity of Natural Admittance Control Implementations," *IEEE International Conference on Robotics and Automation*, pp. 371–376, September 2003. Taipei, Taiwan.

39 S. Oh, H. Woo, and K. Kong, "Frequency-Shaped Impedance Control for Safe Human-Robot Interaction in Reference Tracking Application," *IEEE/ASME Transactions on Mechatronics*, vol. 19, pp. 1907–1916, December 2014.

40 H. Woo and K. Kong, "Controller Design for Mechanical Impedance Reduction," *IEEE/ASME Transactions on Mechatronics*, vol. 20, pp. 845–854, April 2015.

41 L. Buşoniu, R. Babûska, B. De Schutter, and D. Ernst, *Reinforcement learning and dynamic programming using function approximators*. CRC Press, Automation and Control Engineering Series, 2010.

42 R. Sutton and B. A, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.

43 H. van Hasselt, "Double Q-learning,"*In Advances in Neural Information Processing Systems (NIPS)*, pp. 2613–2621, 2010.

44 M. Ganger, E. Duryea, and W. Hu, "Double sarsa and double expected sarsa with shallow and deep learning," *Journal of Data Analysis and Information Processing*, 2016.

45 C. Wang, Y. Li, S. Sam Ge, and T. Heng Lee, "Optimal critic learning for robot control in time-varying environments," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 10, 2015.

46 W. B. Powell, "AI, OR and control theory: A rosetta stone for stochastic optimization," tech. rep., Princeton University, 2012.

47 S. Hart and R. Grupen, “Learning generalizable control programs,” *IEEE Transactions on Autonomous Mental Development*, vol. 3, no. 3, pp. 216–231, 2011.

48 A. G. Barto and S. Mahadevan, “Recent advances in hierarchical reinforcement learning,” *Discrete Event Dynamic Systems*, vol. 13, no. 4, pp. 341–379, 2003.

49 C. G. Atkeson, A. Moore, and S. Stefan, “Locally weighted learning for control,” *AI Review*, vol. 11, pp. 75–113, 1997.

50 P. Abbeel, A. Coates, M. Quigley, and A. Y. Ng, “An application of reinforcement learning to aerobatic helicopter flight,” *In Advances in Neural Information Processing Systems (NIPS)*, 2007.

51 M. P. Deisenroth and C. E. Rasmusen, “PILCO: A model-based and data-efficient approach to policy search,” *In 28th International Conference on Machine Learning (ICML)*, 2011.

52 V. Gullapalli, J. Franklin, and H. Benbrahim, “Acquiring robot skills via reinforcement learning,” *IEEE Control Systems Magazine*, vol. 14, no. 1, pp. 13–24, 1994.

53 J. Kober and J. Peter, “Policy search for motor primitives in robotics,” *In Advances in Neural Information Processing Systems (NIPS)*, 2009.

54 N. Kohl and P. Stone, “Policy gradient reinforcement learning for fast quadrupedal locomotion,” *In IEEE International Conference on Robotics and Automation (ICRA)*, 2004.

55 J. A. Bagnell and J. C. Schneider, “Autonomous helicopter control using reinforcement learning policy search methods,” *In IEEE International Conference on Robotics and Automation (ICRA)*, 2001.

56 H. Miyamoto, S. Schaal, F. Gandolfo, H. Gomi, Y. Koike, R. Osu, E. Nakano, Y. Wada, and M. Kawato, “A Kendama learning robot bases on bi-directional theory,” *Neural Networks*, vol. 9, no. 8, pp. 1281–1302, 1996.

57 J. Peters and S. Schaal, “Learning to control in operational space,” *International Journal of Robotics Research*, vol. 27, no. 2, pp. 197–212, 2007.

58 R. Tedrake, T. W. Zhang, and H. S. Seung, “Learning to walk in 20 minutes,” *In Yale Workshop on Adaptive and Learning Systems*, 2005.

59 M. P. Deisenroth, C. E. Rasmussen, and D. Fox, “Learning to control low-cost manipulator using data-efficient reinforcement learning.,” *In Robotics: Science and Systems (R:SS)*, 2011.

60 H. Benbrahim, J. Doleac, J. Franklin, and O. Selfridge, “Real-time learning: A ball on a beam,” *International Joint Conference on Neural Networks (IJCNN)*, 1992.

61 B. Nemec, M. Tamošiūnaitė, F. Wörgötter, and A. Ude, “Task adaption thorough exploration and action sequencing,” *IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS)*, 2009.

62 M. Tokic, W. Ertel, and J. Fessler, "The crawler, a class room demonstrator for reinforcement learning," *International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, 2009.

63 H. Kimura, T. Yamashita, and S. Kobayashi, "Reinforcement learning of walking behaviour for a four-legged robot," *IEEE Conference on Decision and Control (CDC)*, 2001.

64 R. A. Willgross and J. Igbal, "Reinforcement learning of behaviors in mobile robots using noisy infrared sensing," *Australian Conference on Robotics and Automation*, 1999.

65 L. Paletta, G. Fritz, F. Kintzler, J. Irran, and G. Dorffner, "Perception and developmental learning of affordances in autonomous robots," *Hertzberg J., Beetz M. and Englert R., editors, KI 2007: Advances in Artificial Intelligence*, vol. 4667, pp. 235–250, 2007. Lecture Notes in Computer Science, Springer.

66 T. Yasuda and K. Ohkura, "A reinforcement learning technique with an adaptive action generator for a multi-robot system," *International Conference on Simulation of Adaptive Behavior (SAB)*, 2008.

67 J. H. Piater, S. Jodogne, R. Detry, D. Kraft, N. Krüger, O. Kroemer, and J. Peters, "Learning visual representations for perception-action systems," *International Journal of Robotics Research*, vol. 30, no. 3, pp. 294–307, 2011.

68 P. Fidelman and P. Stone, "Learning ball acquisition on a physical robot," *International Symposium on Robotics and Automation (ISRA)*, 2004.

69 G. D. Konidaris, S. Kuidersma, R. Grupen, and A. G. Barto, "Autonomous skill acquisition on a mobile manipulator," *AAAI Conference on Artificial Intelligence (AAAI)*, 2011.

70 G. D. Konidars, S. Kuindersma, R. Grupen, and A. G. Barto, "Robot learning from demonstration by constructing skill trees," *International Journal of Robotics Research*, vol. 31, no. 3, pp. 360–375, 2012.

71 R. Platt, R. A. Gruphen, and A. H. Fagg, "Improving grasp skills using schema structured learning," *International Conference on Development and Learning*, 2006.

72 V. Soni and S. P. Singh, "Reinforcement learning of hierarchical skills on the Sony AIBO robot," *International Conference on Development and Learning (ICDL)*, 2006.

73 C. An, C. G. Atkeson, and J. M. Hollerbach, "Model-based control of a robot manipulator," *MIT Press:, Cambridge*, MA, USA, 1988.

74 S. Schaal, "Learning from demonstration," *In Advances in Neural Information Processing Systems (NIPS)*, 1996.

75 Y. Duan, B. Cui, and H. Yang, "Robot navigation based on fuzzy RL algorithm," *International Symposium on Neural Networks (ISNN)*, 2008.

76 C. Gaskett, L. Fletcher, and A. Zelinsky, “Reinforcement learning for a vision based mobile robot,” *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2000.

77 R. Hafner and M. Riedmiller, “Reinforcement learning on a omnidirectional mobile robot,” *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2003.

78 I. Grondman, M. Vaandrager, L. Buşoniu, R. Babûska, and E. Schuitema, “Efficient model learning methods for actor-critic control,” *IEEE Transactions on Systems, man, and cybernetics. Part B: Cybernetics*, vol. 42, no. 3, 2012a.

79 O. Kroemeri, R. Detry, J. Piater, and J. Peters, “Active learning using mean shift optimization for robot grasping,” *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2009.

80 O. Kroemer, R. Detry, J. Piater, and J. Peter, “Combining active learning and reactive control for robot grasping,” *Robotics and Autonomous Systems*, vol. 58, no. 9, pp. 1105–1116, 2010.

81 A. Rottmann, C. Plagemann, P. Hilgers, and W. Burgard, “Autonomous blimp control using model-free reinforcement learning in a continuous state and action spac,” *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2007.

82 K. Gräve, J. Stückler, and S. Behnke, “Learning motion skills from expert demonstrations and own experience using Gaussian process regression,” *Joint International Symposium on Robotics (ISR) and German Conference on Robotics (ROBOTIK)*, 2010.

83 B. Kiumarsi, K. G. Vamvoudakis, H. Modares, and F. L. Lewis, “Optimal and autonomous control using reinforcement learning: A survey,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, 2018.