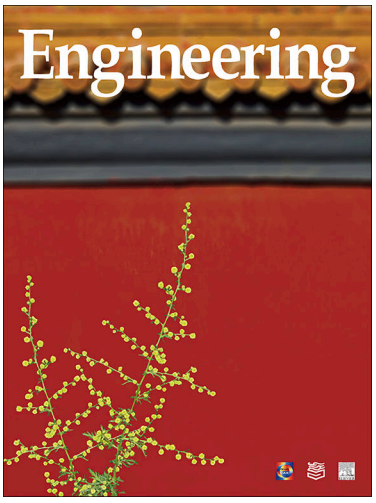
Data-Efficient and Robust Reinforcement Learning for Moving Devices

Citation for the original published paper (version of record):
Lennartson, B. (2026). Data-Efficient and Robust Reinforcement Learning for Moving Devices. Engineering, In Press. http://dx.doi.org/10.1016/j.eng.2026.02.005

Journal Pre-proofs



Views & Comments

Data-Efficient and Robust Reinforcement Learning for Moving Devices

Bengt Lennartson

PII: S2095-8099(26)00067-6
DOI: https://doi.org/10.1016/j.eng.2026.02.005
Reference: ENG 2244

To appear in: *Engineering*

Received Date: 23 June 2025
Accepted Date: 4 February 2026

Please cite this article as: B. Lennartson, Data-Efficient and Robust Reinforcement Learning for Moving Devices, *Engineering* (2026), doi: https://doi.org/10.1016/j.eng.2026.02.005

Views & Comments

Data-Efficient and Robust Reinforcement Learning for Moving Devices

Bengt Lennartson

Division of Systems and Control, Department of Electrical Engineering, Chalmers University of Technology, Gothenburg SE-412 96, Sweden

* Corresponding author.

E-mail address: bengt.lennartson@chalmers.se (B. Lennartson).

Introduction: The difference between model-free and model-based reinforcement learning (RL) is discussed in this paper. Focus is on data-efficiency and robustness against neglected high-frequency dynamics. The conclusion is that the less common model-based RL strategy has clear advantages. For moving devices in intelligent manufacturing systems, a more specific model-based RL method is therefore proposed. The method involves estimation of a nonlinear state-space model, where minor physical knowledge can be easily introduced. Based on this nonlinear model and a feedback/feed-forward controller design, a simple temporal optimization procedure is outlined. The path is then assumed to be given, while velocity and acceleration can be modified such that energy is saved. In robot applications, this temporal optimization strategy has shown to be able to save up to 25% energy and 50% peak power, still keeping the original desired makespan.

Moving devices in intelligent manufacturing: Moving devices are often important elements in intelligent manufacturing systems, where increased intelligence is achieved by adding more autonomy and adaption in such devices. Common examples of intelligent moving devices are automated guided vehicles (AGVs) and autonomous robots. Feedback control is a key element to achieve this adaption and autonomous system behavior of intelligent moving devices.

Control design requires dynamic knowledge: Knowledge about the dynamic system behavior is crucial in the design of feedback control systems. This knowledge is often achieved by formulating dynamic system models, based on physical laws and differential equations. Alternatively, this knowledge can be achieved by collecting data from physical experiments. Such experiments can be step-responses where the control signal makes a jump, and the system output response is measured. A quick response means a fast system dynamics, while a slow response means that it takes longer time before the output signal is adapted to the new control signal level. The response time is often measured by time constants, rise time, or settling time. An interesting alternative is time series analysis, where the system is excited by some type of random input signals, often in combination with closed loop control. The dynamic system behavior is then estimated by a parametric model that is determined by least squares regression [1]. Based on this model, a controller is computed and adapted such that the behavior of the closed loop system is improved, often based on an online optimization procedure.

Model-free RL: RL is a popular example of such an adaptive control strategy. Either a system model is then estimated, and based on this model a controller is designed. Alternatively, a controller is directly determined such that an estimated criterion is optimized. The first version is called model-based RL, while the second one is called model-free RL [2,3]. In the model-free version, a state feedback controller is determined, assuming that the states of the system are measured. Typical examples of states in moving devices are velocity and position, or angles and angular velocities for rotating systems such as robots. Based on an estimated criterion, the state feedback controller can be determined by experimental data without knowing a dynamic model for the system to be controlled. Assuming a linear quadratic criterion and a linear dynamic behavior, this model-free linear quadratic RL (LQRL) strategy is suggested among others in Ref. [4].

Robustness against neglected high frequency dynamics: Unfortunately, this learning strategy has recently been shown to be very sensitive to unmodeled dynamics. For instance, consider a mechatronic system with state feedback from angular velocity and rotating angle, but neglect an additional resonance due to a weak driving axis or the inductive time constant in the direct current (DC)-motor. The closed loop stability of the estimated LQRL controller is then significantly deteriorated even for a very small time constant or a high frequency resonance. The reason for this is that the criterion estimation is much more sensitive to this neglected dynamics than an ordinary feedback control design, based on a given system model with the same neglected resonance or time constant, see further details in Ref. [5]. Fortunately, low pass filtering of the control and nominal state signals is able to significantly reduce the sensitivity to the neglected dynamics.

Model-based RL: An interesting alternative is to evaluate the robustness to the neglected dynamics in a corresponding model-based RL strategy. An ordinary least squares estimation of a state-space model is then performed based on the nominal states and the control input signal [6]. This strategy also works well without filtering when a system model is estimated and an additional high frequency resonance or a short time constant is neglected in the estimated model. The model-based LQRL version also has some other benefits, since it is more open, flexible, and modular. First the estimated model can be evaluated, to see if it is reasonable or not. The evaluation can be performed by simple online experiments, such as step responses, but also based on physical knowledge of expected dynamic behavior. Then, a controller is designed, including evaluation of stability margins and performance in an offline study, before the designed controller is evaluated on the real system. The control structure and design are a not limited to state feedback control. Even a simple but robust proportional-integral-derivative (PID) controller can be chosen.

Data-efficient model-based RL: The most important difference between the model-free and the model-based RL strategy is the fact that the model-free version includes online iterative optimization. Instead of estimating one accurate model, as in model-based RL, the online optimization strategy in model-free RL requires a new estimate of the criterion in every optimization iteration. This means that the number of required data typically increases by a factor 10−100, replacing the model estimation in model-based RL with the criterion estimation in model-free RL. Based on these remarks, our conclusion is that model-based RL has clear advantages compared to model-free RL when data-efficiency and robustness against neglected high-frequency dynamics are considered.

Modularized RL: A model-based RL strategy has recently been proposed in Ref. [7], where the parameters in a nonlinear state-space are estimated. Nonlinearities are included in a black-box style based on polynomials in state and input variables, with the possibility to also include basic knowledge, such as relating force and acceleration by Newton's law. Given this nonlinear model, a combined feedforward and feedback controller is designed, where the goal is to determine an optimal input reference signal. For a moving device with focus on energy minimization, this reference signal is velocity or angular velocity. The optimal reference signal is then determined to minimize energy at the same time as constraints on velocity, acceleration and force or torque are considered.

Temporal optimization: A specific optimization strategy is proposed where discrete time sampling instances are adjusted for a given path such that the energy consumption is minimized. This sampling point adjustment, called temporal optimization, was proposed and applied to energy minimization in Ref. [8], where impressive energy and peak power reduction were obtained, still keeping the original makespan. In Ref. [7], this optimization has been combined with model estimation and feedback/forward design, and it has been compared with well-known model-free deep RL (DRL) strategies. The proposed model-based approach in Ref. [7] is shown to save more energy, while the number of evaluated time steps is reduced by a factor of 100 or more, compared to model-free DRL.

Summary: One main conclusion in this paper is that model-based RL is much more data-efficient, but also more robust against neglected high-frequency dynamics. A modularized model-based RL strategy is therefore proposed where a nonlinear state-space model is estimated. In this model some minor physical knowledge can be easily introduced. Combining feedback and feedforward control with temporal optimization based on the

estimated model, it is shown that energy and peak power for moving devices can be significantly reduced, utilizing much less data compared to standard model-free RL.

References

[1] Ljung L. System identification: theory for the user. 2nd ed. Upper Saddle River: Prentice Hall; 1998.

[2] Sutton RS, Barto AG. Reinforcement learning: an introduction. 2nd ed. Cambridge: MIT press; 2018.

[3] Bertsekas DP. Reinforcement learning and optimal control. Nashua: Athena Scientific; 2019.

[4] Lewis FL, Vrabie D, Vamvoudakis KG. Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers. IEEE Control Syst Mag 2012;32(6):76–105.

[5] Svedlund L, Lennartson B. Robust linear quadratic reinforcement learning by filtering. In: Proceedings of the 2025 IEEE 21st International Conference on Automation Science and Engineering (CASE); 2025 Aug 17–21; Los Angeles, CA, USA. Piscataway: IEEE; 2025. p. 2586–93.

[6] MathWorks. System identification toolbox [Internet]. Natick: The MathWorks, Inc.; 2025 [cited 23 June 2025]. Available from: https://mathworks.com/help/ident/index.html.

[7] Svedlund L, Cronrath C, Fredriksson J, Lennartson B. Model-based data-efficient and robust reinforcement learning. arXiv: arXiv:2602.00630

[8] Riazi S, Wigström O, Bengtsson K, Lennartson B. Energy and peak power optimization of time-bounded robot trajectories. IEEE Trans Autom Sci Eng 2017;14(2):646–57.

Declaration of Interest Statement

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for this journal and was not involved in the editorial review or the decision to publish this article.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: