

Article

A Comparison of the Use of Pontryagin's Maximum Principle and Reinforcement Learning Techniques for the Optimal Charging of Lithium-Ion Batteries

Andreas Rauh ^{*}, Marit Lahme  and Oussama Benzinane 

Group: Distributed Control in Interconnected Systems, School II—Department of Computing Science, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany

* Correspondence: andreas.rauh@uni-oldenburg.de

Abstract: Battery systems are one of the most important components for the development of flexible energy storage for future applications. These comprise energy storage in both the mobility sector and stationary applications. To ensure the safe operation of multiple battery cells connected in series and parallel in a battery pack, it is essential to implement state of charge (SOC) equalization strategies. Generally, two fundamentally different approaches can be distinguished. On the one hand, these are passive approaches for SOC equalization that are based on including additional Ohmic resistors in a battery back over which equalization currents flow as long as the correspondingly connected cells have different voltages. Despite the simple implementation of such equalization circuits, they have a major drawback, namely wasting stored energy to perform the SOC equalization. This waste of energy goes along with Ohmic heat production, which leads to the necessity of additional cooling for batteries with large power densities. On the other hand, active SOC equalization approaches have been investigated, which allow for an independent charging of the individual cells. Especially, this latter approach has big potential to be more energy efficient. In addition, the potential for a reduction of Ohmic heat production may contribute to extending the lifetime of battery cells. To perform the individual charging of battery cells in an energetically optimal manner, this paper provides a comparison of closed-form optimization approaches on the basis of Pontryagin's maximum principle and approaches for reinforcement learning. Especially, their accuracy and applicability for the implementation of optimal online cell charging strategies are investigated.

Keywords: Lithium-ion batteries; optimal control; Pontryagin's maximum principle; reinforcement learning



Citation: Rauh, A.; Lahme, M.; Benzinane, O. A Comparison of the Use of Pontryagin's Maximum Principle and Reinforcement Learning Techniques for the Optimal Charging of Lithium-Ion Batteries. *Clean Technol.* **2022**, *4*, 1269–1289. <https://doi.org/10.3390/cleantechnol4040078>

Academic Editor: Damien Guilbert

Received: 27 September 2022

Accepted: 1 December 2022

Published: 7 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Lithium-ion batteries have become one of the most important enablers for the implementation of fully electric and hybrid traction systems in all domains of transportation [1–4]. These include individual transport in terms of electric bicycles [5], motorcycles [6], or cars [7], but also battery storage for all-electric ships [8], as well as electric and hybrid tramways and locomotives [9]. Besides the sector of transportation, Lithium-ion batteries are widely employed as energy storage for systems in consumer electronics, as well as battery storage power stations [10], in which they are employed as medium-time-scale buffers. Among others, the latter serve as systems that cover peak-power demands and allow for storing an excess amount of renewable energy, but can also be employed as equipment covering ancillary services in the domains of operating reserves, as well as frequency control in electric power grids and, thus, serving as a means to reduce the risk of power outage. Depending on the exact area of application, the capacities of these storage systems range from a few watt-hours up to the gigawatt-hour range.

Regardless of those different storage capacities, a common problem that all of these systems have to cope with is the equalization of the SOC of individual battery cells that are

electrically connected in series and in parallel [11]. This SOC equalization is the prerequisite for the prevention of low discharge of individual cells or over-charging, which both may lead to irreversible damage of the cells in terms of capacity loss, accelerated degradation, and in the worst possible case, hazardous situations such as thermal runaway, which may lead to fires, exposing the technical equipment and the environment to high risks, causing extreme property and financial damage, and especially threatening the lives of humans [12].

Preventing low discharge and over-charging are typical functionalities of battery management systems. They are based on passive or active approaches for SOC equalization [11,13–19]. In the passive case, bleeder resistors can be included in battery packs (which can be activated and deactivated by means of semi-conductor switches) to equalize the voltage of individual cells that are compared with each other. The use of such resistors has the advantage of simple SOC equalization circuits. However, currents through these bleeder resistors lead to Ohmic heat production, which needs to be countered by cooling systems. In addition, wasting stored energy is inevitable by this kind of passive approach. In contrast, active approaches aim directly at charging (respectively, discharging) individual cells by decentralized sources. This active approach allows for implementing energy-optimal charging strategies, which are investigated in this paper from the point of view of minimizing the overall Ohmic heat production in terms of an energy equivalent when charging an individual cell from a given initial to a desired final SOC.

To account for the SOC-dependent nonlinearities of the dynamics of battery cells, this task is solved in this paper by means of two substantially different formulations. The first option is the derivation of optimal, finite-duration charging profiles with the help of Pontryagin's maximum principle as an indirect optimization approach [20–23]. As an alternative, a direct control optimization is presented in terms of a reinforcement learning formulation [24,25]. The first alternative is further extended toward a predictive control technique [26,27]. It is additionally compared with a linear-quadratic regulator design [20,28]. For that purpose, it is assumed that the nonlinear equivalent circuit model forming the basis for the solution of the maximum principle is approximated by a linear time-invariant state-space representation for which an infinite time-horizon control is designed.

This paper is structured as follows. Section 2 gives a summary of an equivalent circuit representation of Lithium-ion batteries that has been identified experimentally in previous work [29]. The methodological foundations of the alternative optimization approaches, namely Pontryagin's maximum principle and the reinforcement learning approach, are presented in Section 3 with a focus on energetically optimal battery charging. Representative simulation results, on the basis of the aforementioned experimentally validated battery model, are presented in Section 4, before conclusions and an outlook on future work are given in Section 5.

2. Equivalent Circuit Modeling and Criteria for the Energetically Optimal Charging of Lithium-Ion Batteries

In this section, the modeling assumptions for the dynamics of Lithium-ion batteries are described with a focus on the quantification of Ohmic losses occurring during instationary operating conditions.

2.1. Equivalent Circuit Modeling

As described in a large variety of scientific articles (cf. [30,31]), equivalent circuit models with lumped, SOC-dependent resistances and capacitances are applicable to model the dynamics of charging and discharging Lithium-ion batteries. An illustration of the structure of the corresponding electric network is given in Figure 1.

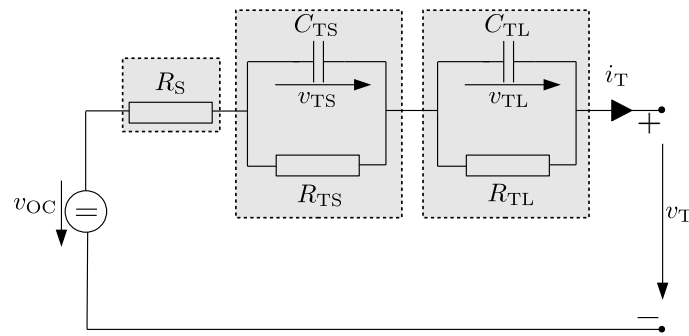


Figure 1. Equivalent circuit model of a Lithium-ion battery in which a series connection of two RC sub-networks is used to represent lag phenomena in the charging and discharging phases.

Basically, these models can be transferred into a corresponding integer-order state-space representation, where the SOC $\sigma(t)$ and the voltages across a series connection of RC sub-networks serve as the state variables. In addition to the storage of charge carriers, represented by the SOC, the voltages across the RC sub-models are used to represent the dynamic phenomena that are caused by electro-chemical polarization effects and concentration losses with short and large time constants (denoted by the indices TS and TL, respectively).

Under these considerations, the state vector of the equivalent circuit model is given by

$$\mathbf{x}(t) = [\sigma(t) \quad v_{TS}(t) \quad v_{TL}(t)]^T \quad (1)$$

with the associated quasi-linear, continuous-time state equations

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}(\sigma(t)) \cdot \mathbf{x}(t) + \mathbf{b}(\sigma(t)) \cdot u(t) \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{-1}{C_{TS}(t) \cdot R_{TS}(t)} & 0 \\ 0 & 0 & \frac{-1}{C_{TL}(t) \cdot R_{TL}(t)} \end{bmatrix} \cdot \mathbf{x}(t) + \begin{bmatrix} \frac{-1}{C_{Bat}} \\ \frac{1}{C_{TS}(t)} \\ \frac{1}{C_{TL}(t)} \end{bmatrix} \cdot u(t), \end{aligned}$$

in which the terminal current is the system input $u(t) := i_T(t)$.

The first row of the state Equation (2) represents the integrating behavior

$$\dot{\sigma}(t) = -\frac{i_T(t)}{C_{Bat}} \quad (2)$$

between the charging/discharging current $i_T(t)$ and the normalized SOC $\sigma(t) \in [0; 1]$, where C_{Bat} is the nominal battery capacitance. Throughout this paper, $\sigma = 1$ corresponds to the fully charged battery, while $\sigma = 0$ is the operating condition of the completely discharged one.

The two further differential equations in the system model express the voltages $v_{TS}(t)$ and $v_{TL}(t)$ with the time constants [30,31]

$$\tau_l = C_l(t) \cdot R_l(t), \quad l \in \{TS, TL\}, \quad (3)$$

which are temporally varying due to the SOC dependencies of the parameters

$$R_l(t) = R_{la} \cdot e^{R_{lb} \cdot \sigma(t)} + R_{lc} \quad (4)$$

and

$$C_l(t) = C_{la} \cdot e^{C_{lb} \cdot \sigma(t)} + C_{lc}. \quad (5)$$

Further SOC dependencies are relevant for the Ohmic series resistance

$$R_S(t) = R_{Sa} \cdot e^{R_{Sb} \cdot \sigma(t)} + R_{Sc}, \quad (6)$$

as well as for the approximation

$$v_{OC}(\sigma(t)) = v_0 \cdot e^{v_1 \cdot \sigma(t)} + \sum_{i=0}^3 v_{i+2} \cdot \sigma^i(t) \quad (7)$$

of the battery's open-circuit voltage.

The latter, as well as the voltage drops $v_l(t)$, $l \in \{TS, TL\}$, and the voltage drop over the series resistance $R_S(t)$ can be combined with the help of Kirchhoff's voltage law

$$v_T(t) = v_{OC}(t) - v_{TS}(t) - v_{TL}(t) - i_T(t) \cdot R_S(t) \quad (8)$$

to find a relation between the terminal voltage $v_T(t)$ and the terminal current $i_T(t)$.

All parameters of the system model above can be identified experimentally, either by means of impedance spectroscopy or by using dedicated identification experiments such as those summarized in [29]. The parameter values listed in [29] are used in this paper for the numerical evaluation of the presented optimization approaches.

Remark 1. *If the integer-order models mentioned before are not sufficiently accurate for describing long-term memory effects, they can be replaced with fractional-order differential equation models as presented, for example, in [32–34].*

2.2. Quantification of Ohmic Losses

Ohmic losses in the battery, leading effectively to a change of its temperature, are caused by the electric currents through all Ohmic resistances in the equivalent circuit in Figure 1. Summing up the corresponding electric powers leads to the expression

$$P_{el}(t) = R_S(t) \cdot i_T^2(t) + \frac{v_{TS}^2(t)}{R_{TS}(t)} + \frac{v_{TL}^2(t)}{R_{TL}(t)}, \quad (9)$$

which should be kept as small as possible when charging the battery from an initial SOC to a desired final one in a given time span. Besides a minimization of this power (in an integral sense over the complete charging duration), it is desired to limit the maximum absolute values of the charging currents. This latter aspect is accounted for in the following section by adding a penalty term to the cost function, which consists of the square of the control signal.

3. Optimal Control Synthesis

The optimal control design can be classified into indirect approaches (based on Pontryagin's maximum principle [23]), which compute the optimal system inputs in terms of an auxiliary boundary value problem for the set of canonical equations, as well as into direct ones, which directly determine the system inputs (e.g., by using reinforcement learning [25]). This section summarizes the methodological foundations of these alternatives and further provides the link between an optimal open-loop control, as well as closed-loop feedback control techniques.

3.1. Indirect Optimization Using Pontryagin's Maximum Principle

For the case of using Pontryagin's maximum principle, we distinguish two different options for specifying the terminal system states. These are fixed terminal conditions for all state vector components in Section 3.1.1 and partially free ones in Section 3.1.2.

3.1.1. Fixed Terminal State

Assume a continuous-time system model

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), u(t), t) \quad (10)$$

with the state vector $\mathbf{x}(t) \in \mathbb{R}^n$ and the scalar control signal $u(t)$. Assume further that the initial state is given by $\mathbf{x}_0 = \mathbf{x}(t_0)$.

In the case of a predefined terminal state $\mathbf{x}_f = \mathbf{x}(t_f)$ at the time instant $t = t_f > t_0$, the cost function to be minimized by the optimal control law

$$u^*(t) = u^*(\mathbf{x}(t), \mathbf{p}(t), t) \quad (11)$$

is given by

$$J(u) = \int_{t_0}^{t_f} f_0(\mathbf{x}(\tau), u(\tau), \tau) d\tau. \quad (12)$$

In (11), the vector $\mathbf{p}(t) \in \mathbb{R}^n$ denotes the co-state vector introduced in Pontryagin's maximum principle according to [20–23].

In addition, the integrand

$$f_0(\mathbf{x}(t), u(t), t) = \alpha \cdot i_T^2(t) + P_{el}(t) \quad (13)$$

in Equation (12) corresponds to the electric power according to (9) with an additive penalty, where $\alpha > 0$, that allows for limiting the control amplitudes. Note that time arguments are omitted in the following for the sake of compactness of the notation as long as the corresponding mathematical expressions are non-ambiguous. To find the candidate for the optimal control sequence u^* , the Hamiltonian

$$H(\mathbf{x}, u, \mathbf{p}, t) = -f_0(\mathbf{x}, u, t) + \mathbf{p}^T \cdot \mathbf{f}(\mathbf{x}, u, t) \quad (14)$$

needs to be maximized according to

$$\max_{u \in \mathbb{R}} (H(\mathbf{x}, u, \mathbf{p}, t)), \quad (15)$$

where we assume an unbounded control $u = u(t) \in \mathbb{R}$. Due to the unbounded control input and the differentiability of (16), the necessary optimality condition

$$\left. \frac{\partial H(\mathbf{x}, u, \mathbf{p}, t)}{\partial u} \right|_{u=u^*} = -2(R_S(t) + \alpha)i_T(t) + \mathbf{p}^T(t) \cdot \mathbf{b}(\sigma(t)) = 0 \quad (16)$$

leads to the unique solution candidate

$$u^* = \frac{1}{2(R_S(t) + \alpha)} \cdot \mathbf{p}^T(t) \cdot \mathbf{b}(\sigma(t)), \quad (17)$$

representing the global maximum of the Hamiltonian due to the strict negativity of its second derivative

$$\left. \frac{\partial^2 H(\mathbf{x}, u, \mathbf{p}, t)}{\partial u^2} \right|_{u=u^*} = -2(R_S(t) + \alpha) < 0. \quad (18)$$

To obtain a control law u^* that minimizes Ohmic losses and penalizes the maximum control effort, the parameter

$$\alpha > 0 \quad (19)$$

is selected by a trial-and-error approach in this paper.

With the help of this optimized control, the set of canonical equations

$$\begin{aligned}\dot{\mathbf{x}} &= \frac{\partial H(\mathbf{x}, u^*, \mathbf{p}, t)}{\partial \mathbf{p}} = \mathbf{f}(\mathbf{x}, u^*, t) \\ \dot{\mathbf{p}} &= -\frac{\partial H(\mathbf{x}, u^*, \mathbf{p}, t)}{\partial \mathbf{x}} = \frac{\partial f_0}{\partial \mathbf{x}} - \left(\frac{\partial \mathbf{f}(\mathbf{x}, u^*, t)}{\partial \mathbf{x}} \right)^T \cdot \mathbf{p}\end{aligned}\quad (20)$$

is obtained, which can be simplified according to

$$\begin{aligned}\dot{\mathbf{x}} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{-1}{C_{TS}(t) \cdot R_{TS}(t)} & 0 \\ 0 & 0 & \frac{-1}{C_{TL}(t) \cdot R_{TL}(t)} \end{bmatrix} \cdot \mathbf{x} + \begin{bmatrix} \frac{-1}{C_{Bat}} \\ \frac{1}{C_{TS}(t)} \\ \frac{1}{C_{TL}(t)} \end{bmatrix} \cdot \frac{1}{2(R_S(t) + \alpha)} \cdot \mathbf{p}^T \cdot \mathbf{b}(\sigma(t)) \\ \dot{\mathbf{p}} &= \begin{bmatrix} \frac{\partial P_{el}(t)}{\partial \sigma} \\ \frac{2v_{TS}(t)}{R_{TS}(t)} \\ \frac{2v_{TL}(t)}{R_{TL}(t)} \end{bmatrix} - \left(\frac{\partial \mathbf{f}(\mathbf{x}, u^*, t)}{\partial \mathbf{x}} \right)^T \cdot \mathbf{p}.\end{aligned}\quad (21)$$

For this set of canonical equations, a boundary value problem needs to be solved with the $2n$ boundary conditions $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\mathbf{x}(t_f) = \mathbf{x}_f$ in the case of fixed terminal states and perfectly known initial conditions. As shown in the following section, exactly prescribing all terminal state vector components \mathbf{x}_f , especially the voltages across the RC sub-networks, leads to higher Ohmic losses than the case of partially free boundary conditions in the following section.

The typical choice of boundary conditions in this case is those of equilibria with different values for the SOC; hence,

$$\mathbf{x}_0 = \begin{bmatrix} \sigma_0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_f = \begin{bmatrix} \sigma_f \\ 0 \\ 0 \end{bmatrix}.\quad (22)$$

3.1.2. Partially Free Terminal State

To avoid the exact prescription of an equilibrium state for the RC sub-network voltages, the second and third vector component of \mathbf{x}_f have to be left undefined. The correspondingly missing boundary conditions are then obtained after extending the cost function (12) by an additive term penalizing the final state according to

$$J_h(u) = h_0(\mathbf{x}(t_f), t_f) + \int_{t_0}^{t_f} f_0(\mathbf{x}(\tau), u(\tau), \tau) d\tau\quad (23)$$

with the typical choice

$$h_0(\mathbf{x}(t_f), t_f) = \beta \cdot \left(v_{TS}^2(t_f) + v_{TL}^2(t_f) \right).\quad (24)$$

The terminal cost term in (24) leads to the transversality conditions [20–23]

$$\begin{aligned}p_2(t_f) &= -\frac{\partial h_0(\mathbf{x}, t_f)}{\partial v_{TS}} \bigg|_{\mathbf{x}=\mathbf{x}(t_f)} = -2\beta v_{TS}(t_f) \\ p_3(t_f) &= -\frac{\partial h_0(\mathbf{x}, t_f)}{\partial v_{TL}} \bigg|_{\mathbf{x}=\mathbf{x}(t_f)} = -2\beta v_{TL}(t_f)\end{aligned}\quad (25)$$

for two of the co-state variables to be considered in addition to the exactly defined terminal SOC

$$x_1(t_f) = \sigma_f\quad (26)$$

when solving the boundary value problem for the canonical Equation (21).

3.2. Conversion into a Model Predictive Control Task

Both open-loop optimal control formulations from the previous subsection can be converted into appropriate model predictive control procedures. For that purpose, it is necessary to perform the optimization of the input sequence for the current state variables of the equivalent circuit, namely the SOC and the voltages across the RC sub-networks, after the previous optimization result has already been applied for a certain time interval. To perform the re-optimization, the equivalent circuit's state variables are typically estimated by a suitable observer or filtering approach to provide the state estimates $\hat{x}(t_k)$ at the time instants t_k , $k \in \{0, 1, \dots, K_c\}$. Widely applicable approaches, based on Kalman filters, neural networks, online estimation error minimization, or learning-type estimators can be found in [29,31,35–40]. Note that these estimation techniques are still an active field of research.

Figure 2 visualizes the computation of the optimal open-loop control sequence for the complete time horizon $[t_0 ; t_f]$ with either of the formulations of Pontryagin's maximum principle from the previous subsection. Due to the fact that this control sequence is then applied to the dynamic system without any further correction at runtime, it is not guaranteed that the desired terminal state (obtained perfectly by the prediction model employed for the offline optimization in the previous subsection) actually corresponds to the true terminal state. The reasons for this behavior are mismatches in the initial conditions, as well as imperfectly known system parameters and structural deviations between the assumed system dynamics and the true battery behavior.

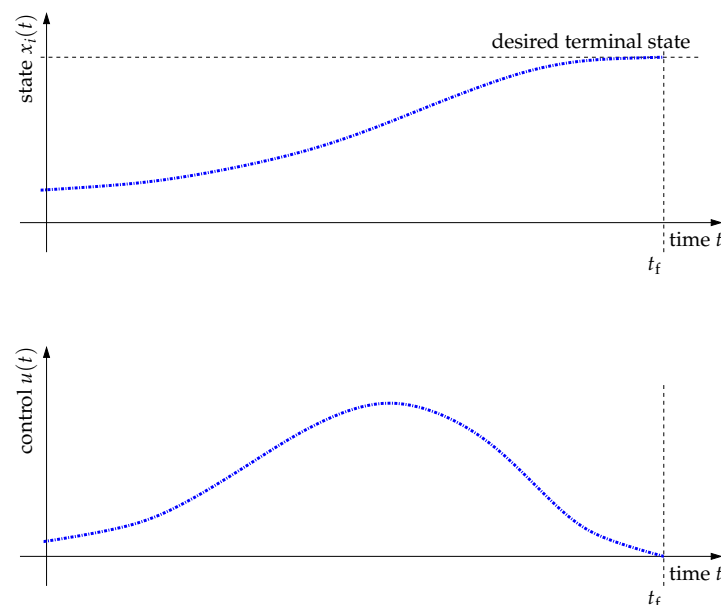


Figure 2. Computation of the optimal feedforward control $u(t) = u^*(r)$ for the complete optimization horizon $[t_0 ; t_f]$ with the predicted optimal state trajectory $x_i(t)$, $i \in \{1, \dots, n\}$.

Therefore, a predictive control setting can be implemented by following the strategies summarized in Figure 3. There, the open-loop control strategy is re-computed at each point of time t_k in an online manner by using the same optimization approach as in the previous subsection. The corresponding input signal $u_k(t)$ is then applied to the system over the time interval $[t_k ; t_{k+1}]$ before it is re-computed again with the new state estimates $\hat{x}_i(t_{k+1})$. Note that, in the implementation summarized in this figure, the prediction horizon becomes shorter after each control update. For noisy state estimates, used for the re-initialization of the optimization, this may have the drawback of large control variations towards the final time instant t_f . This can be countered by increasing the penalty actor α when increasing k ,

by avoiding an exact specification of the final SOC, but rather penalizing it in the terminal costs (24), by switching to a constant voltage charging for higher values of the SOC or by switching to the state-feedback implementation presented in the following subsection when the time t approaches the final time instant t_f .

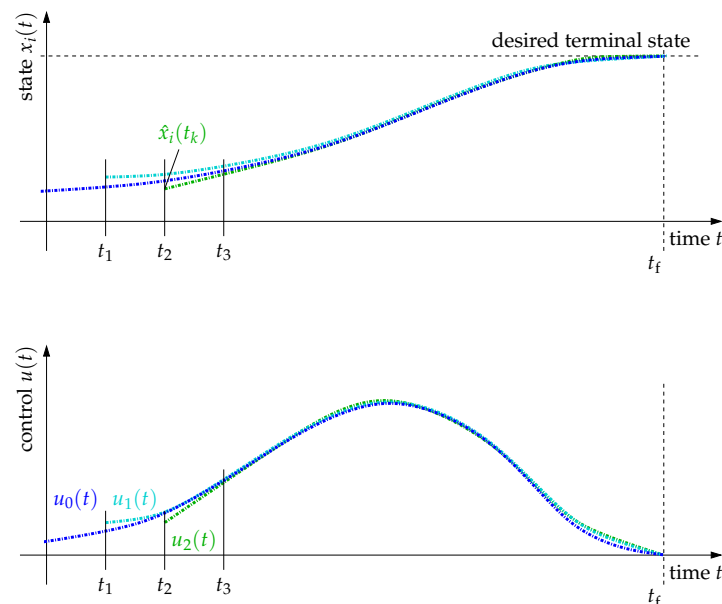


Figure 3. Re-computation of the optimal control sequences $u_k(t)$ for the reduced horizons $[t_k; t_f]$, $k \in \{0, \dots, K_c\}$, with the associated predicted state trajectories $x_{i,k}(t)$, $i \in \{1, \dots, n\}$, initialized with the state estimates $\hat{x}_i(t_k)$.

3.3. Relations to Feedback Control Based on a Linear-Quadratic Regulator Design

The previous finite-horizon optimization problems can be turned into a closed-loop control approach by a minimization of the cost function

$$\bar{J} = \frac{1}{2} \int_0^{\infty} \left((\mathbf{x}(\tau) - \mathbf{x}_f)^T \cdot \mathbf{Q} \cdot (\mathbf{x}(\tau) - \mathbf{x}_f) + R u^2(\tau) \right) d\tau. \quad (27)$$

Here, the vector \mathbf{x}_f corresponds to the terminal state constraint introduced in Section 3.1.1. The optimal controller is then given by the expression

$$u(t) = \mathbf{k}^T (\mathbf{x}_f - \mathbf{x}(t)), \quad (28)$$

if the gain vector

$$\mathbf{k}^T = R^{-1} \mathbf{b}^T \mathbf{P} \quad (29)$$

is computed in terms of the positive definite, state-independent solution of the algebraic Riccati equation

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} + \mathbf{Q} - \mathbf{P} \mathbf{b} R^{-1} \mathbf{b}^T \mathbf{P} = \mathbf{0}, \quad (30)$$

where the weighting matrix \mathbf{Q} and the scalar weight R are defined according to

$$\mathbf{Q} = \begin{bmatrix} \gamma & 0 & 0 \\ 0 & \frac{1}{\bar{R}_{TS}} & 0 \\ 0 & 0 & \frac{1}{\bar{R}_{TL}} \end{bmatrix} \quad (31)$$

with the parameter

$$\gamma > 0 \quad (32)$$

as well as

$$R = \alpha + \bar{R}_S. \quad (33)$$

Here, the parameter γ is introduced in addition to the parameterization of the fundamental cost function in (12) to ensure that the state trajectory approaches the desired terminal state \mathbf{x}_f .

Remark 2. State-dependent solutions of the algebraic Riccati equation, as introduced in [41–43], are not further studied in this paper because they can be interpreted as a refinement of the coarse state-independent solution of this subsection, which approaches the solution of Section 3.1.1 when setting the terminal time instant t_f to a large value.

Remark 3. A further combination of this state feedback controller with the optimization results of Pontryagin's maximum principle is the use of the result u^* of the open-loop optimization as a feedforward control signal, and to add a further feedback path that compensates the deviations between the offline optimized state trajectories and the currently estimated ones, see also Section 4.3.

3.4. Direct Optimization by Reinforcement Learning

Reinforcement learning control approaches [24,25] can be used as an alternative to the solution techniques of the previous subsections to directly determine optimal control strategies. The general reinforcement learning environment, adapted to the case of optimal charging for Lithium-ion batteries, is summarized in Figure 4.

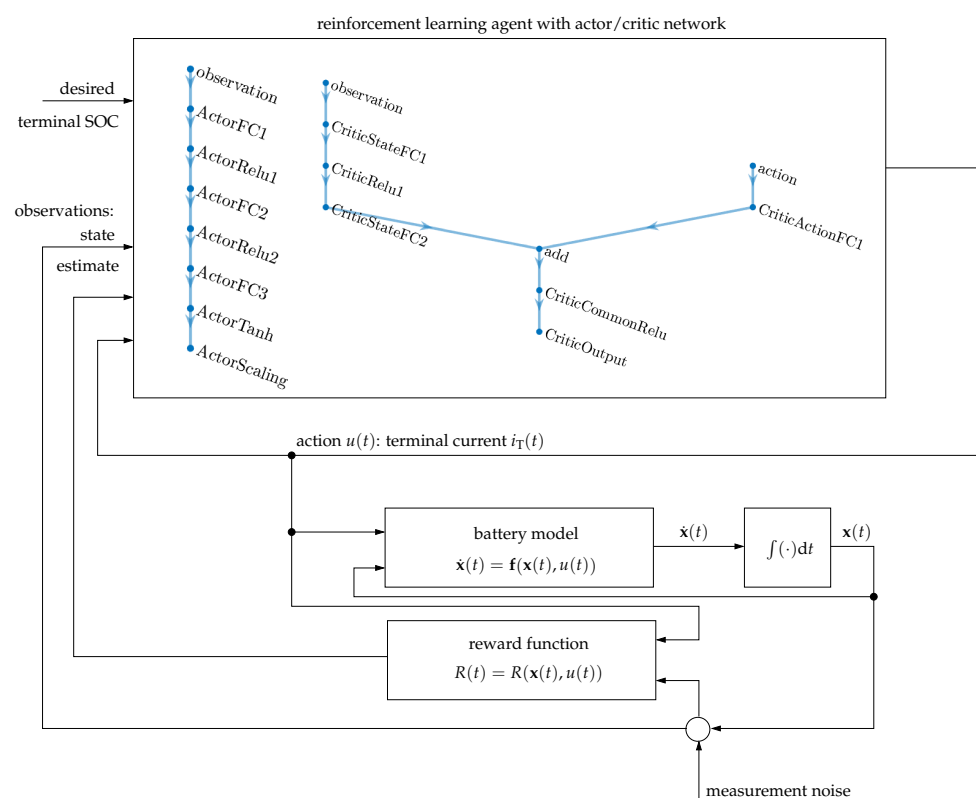


Figure 4. Structure of the reinforcement learning environment.

There, the reward function serves as a substitute for the performance criterion $J(u)$, where the quite commonly used notation is a maximization of the reward, as opposed to the minimization discussed before. Therefore, the basic constituent of the reward function is the multiplication of the electric power P_{el} (respectively, the integrand f_0 according to (13))

with a negative sign. As discussed in Section 4, it is extended by a penalization of the deviations of the actual SOC from the desired terminal state.

Due to the fact that the reinforcement learning approach avoids the solution of boundary value problems for a specific set of (estimated) initial states, it can be robustified directly by adding suitably chosen measurement noise (representing also the effect of state reconstruction errors in an aleatory form) in both the computation of the reward function and in the observation vector. In the application at hand, this observation vector corresponds to the full state vector consisting of the SOC and the voltages across the RC sub-networks of the equivalent circuit model in Figure 1.

The reinforcement learning itself, implemented as a deep deterministic policy gradient (DDPG) agent in MATLAB within this paper, includes two neural networks:

- The *actor network* determines the control signal (action) in terms of the observations, where the individual layers in the network according to Figure 4 are fully connected layers denoted by *FC* and layers with *ReLU* and *tanh* activation functions;
- The *critic network* contains both observations and actions as inputs to determine the reward of the current policy, where the two corresponding input paths are superposed additively, as shown again in Figure 4.

3.5. Summary of the Properties of the Control Approaches

The fundamental properties of all control approaches presented in this section are summarized in Table 1. There, the characterization of the capability to minimize the Ohmic losses during the optimal charging, denoted as optimization efficiency, is based on the results presented in Section 4.

Table 1. Summary of the algorithmic properties of the considered control approaches.

Approach	Offline Effort	Online Effort	Robustness against Noise	Generalizability with Respect to Initial SOC	Optimization Efficiency
Maximum principle	medium	low	independent (pure offline solution)	low (recomputation required)	good
Predictive control	—	medium–high	depending on cost function parameters	excellent	excellent
Linear-quadratic feedback control	low	low	depending on cost function parameters	medium–good	low
Reinforcement learning	high	low–medium	high for sufficiently rich training data	good–excellent	excellent

4. Simulation and Optimization Results

In this section, numerical simulation results are presented to illustrate the performance of the optimization approaches summarized in Section 3. The presentation of these results focuses on both the capability to transfer the Lithium-ion battery from a given SOC σ_0 to a desired terminal value σ_f in a fixed time span and to analyze the robustness against sensor (respectively, state reconstruction) noise when adapting the control signal at runtime.

For all the following simulations, we assumed that the battery model according to Figure 1 is parameterized with the values listed in [29]. Moreover, the following initial and terminal conditions hold:

$$\sigma_0 = 0.5 \quad \text{and} \quad \sigma_f = 0.9. \quad (34)$$

4.1. Indirect Optimization Using Pontryagin's Maximum Principle

The numerical evaluation in this subsection is based on precomputing a charging profile for the Lithium-ion battery that transfers the initial SOC (σ_0) to the final one (σ_f) in 3600 s. This phase is followed by a relaxation phase of again 3600 s with the terminal current $i_T \equiv 0$ A, which guarantees that the desired final SOC is kept, but that the voltages across both RC sub-networks approach zero, with a corresponding quantification of the additional Ohmic losses.

The corresponding results are summarized for two different values of the parameter α included in the cost function (13) in Tables 2 and 3. It can be seen clearly that the partially free terminal states (penalized by the terminal cost according to (24)) lead to much smaller Ohmic losses than the case of setting all voltages exactly to zero at the end of the charging phase. For that reason, only the case of partially free terminal conditions is investigated further in the following graphical illustrations in Figures 5 and 6.

Table 2. Ohmic losses (energy) during charging and relaxation phases for fixed terminal states.

α	Losses during Charging ($t \in [0 ; 3600]$ s)	Total Losses ($t \in [0 ; 7200]$ s)
1	843.11 Ws	843.16 Ws
0.01	841.26 Ws	841.32 Ws

Table 3. Ohmic losses (energy) during charging and relaxation phases for partially fixed terminal states with $\beta = 50$.

α	Losses during Charging ($t \in [0 ; 3600]$ s)	Total Losses ($t \in [0 ; 7200]$ s)
1	690.14 Ws	702.62 Ws
0.01	686.13 Ws	702.52 Ws

The left column of these two figures corresponds to the case with the larger penalization of the charging current (larger value of α) as compared to the right column. It can be seen that this larger value of α leads to a solution that is close to a constant current charging policy, which is additionally included in Figure 5a,b. For this constant current profile, the Ohmic losses in the pure charging phase result in 690.97 Ws, while the overall losses including also the relaxation phase are 703.05 Ws. It can be seen that these values are larger than the optimized charging profile, which is characterized (except for the very beginning and end) by currents that are smaller than the temporally constant alternative. Moreover, the remaining sub-graphs of Figures 5 and 6 indicate the relaxation processes in the RC sub-networks after terminating the charging at $t = 3600$ s. These results indicate also that too small values for α may lead to large current peaks at the end of the charging process, which need to be prevented by suitably selecting this parameter.

4.2. Model Predictive Control

As already discussed in Section 3.2, a pure offline optimization of the charging profile suffers from the disadvantage that it cannot adapt to uncertainty in the initial conditions, as well as to deviations of the actual battery parameters from the ones assumed for solving the optimization problem. Therefore, this subsection shows the results of a model predictive implementation of the optimal control law, where the duration $t_{k+1} - t_k$ between two points at which the control strategy is recomputed was set to 120 s. At these points, knowledge about the current state vector $\hat{\mathbf{x}}(t_k)$ is assumed to be known, which deviates, however, in the simulation by an additive zero-mean Gaussian noise from the true state values $\mathbf{x}(t_k)$ with a standard deviation of 0.01 for the SOC and 0.001 for both voltages.

The simulation results in Figure 7, again computed for $\alpha = 0.01$ with the extended cost function that includes the terminal state penalization (24), contain in yellow color the one-standard-deviation bounds for the terminal current and the SOC, respectively, in

addition to the corresponding average values. To obtain these one-standard-deviation bounds, the complete simulation was repeated 100 times.

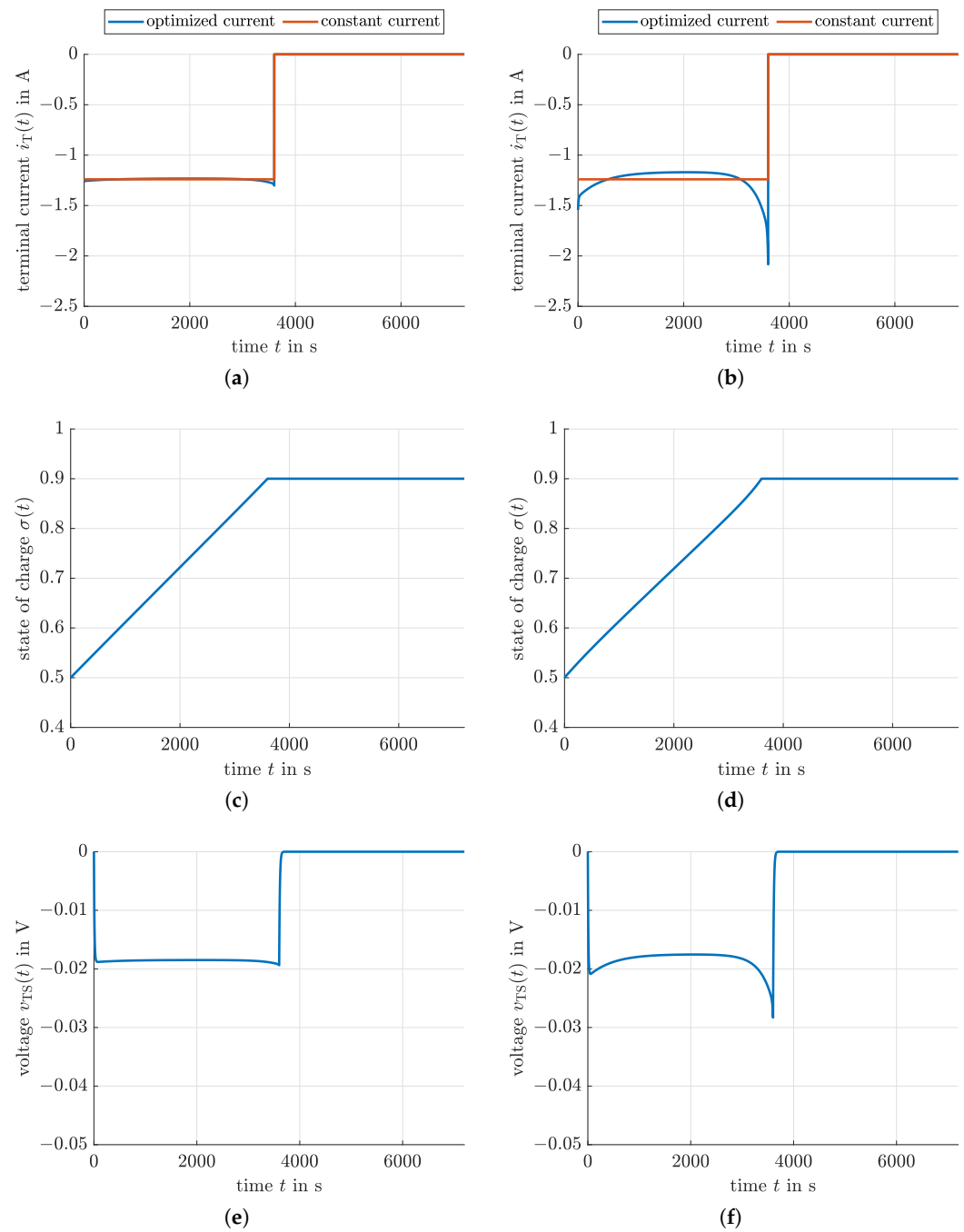


Figure 5. Offline optimization of the charging profile according to Section 3.1.2. (a) Terminal current $i_T(t)$ for $\alpha = 1$; (b) Terminal current $i_T(t)$ for $\alpha = 0.01$; (c) SOC $\sigma(t)$ for $\alpha = 1$; (d) SOC $\sigma(t)$ for $\alpha = 0.01$; (e) Voltage $v_{TS}(t)$ for $\alpha = 1$; (f) Voltage $v_{TS}(t)$ for $\alpha = 0.01$.

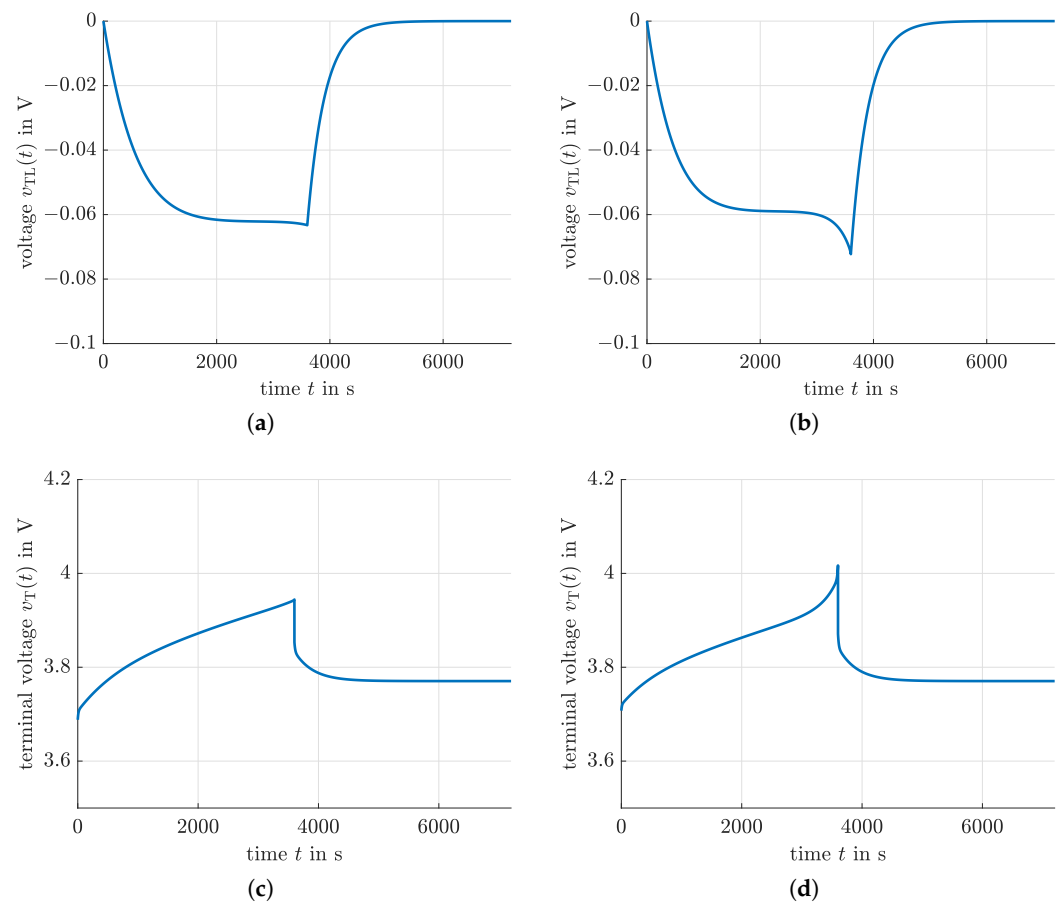


Figure 6. Offline optimization of the charging profile according to Section 3.1.2 (continued). (a) Voltage $v_{TL}(t)$ for $\alpha = 1$; (b) Voltage $v_{TL}(t)$ for $\alpha = 0.01$; (c) Terminal voltage $v_T(t)$ for $\alpha = 1$; (d) Terminal voltage $v_T(t)$ for $\alpha = 0.01$.

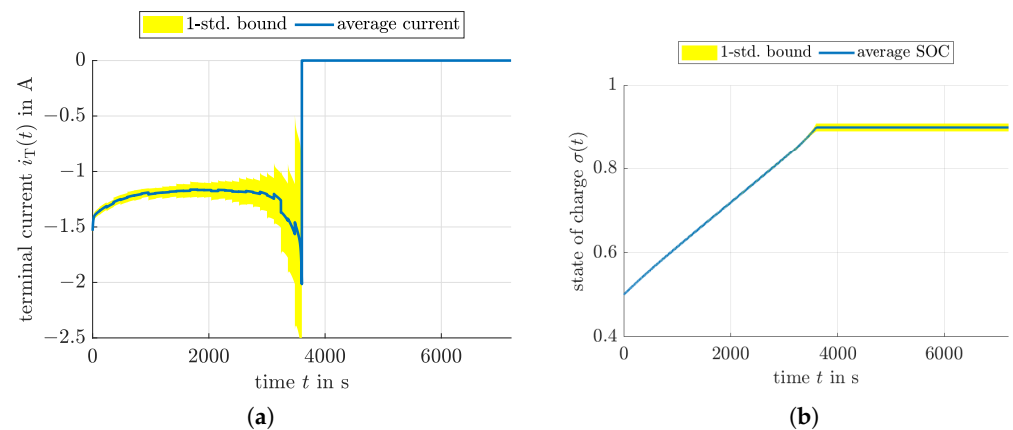


Figure 7. Model predictive control under the influence of state reconstruction noise. (a) Terminal current $i_T(t)$ for $\alpha = 0.01$; (b) SOC $\sigma(t)$ for $\alpha = 0.01$.

Obviously, as also shown in Figure 8, the noise leads to the fact that the Ohmic losses in both the pure charging phase, as well as in the longer time span including the charging and relaxation phases become themselves uncertain; cf. Figure 8a,b. Moreover, also the desired terminal SOC is not reached perfectly at the final time instant $t_f = 3600$ s according to Figure 8c. This latter aspect is re-discussed in the investigation of the reinforcement

learning control approach, where it is shown that a remaining correction of the SOC for $t > 3600$ s in a few minutes leads to only small additional Ohmic losses.

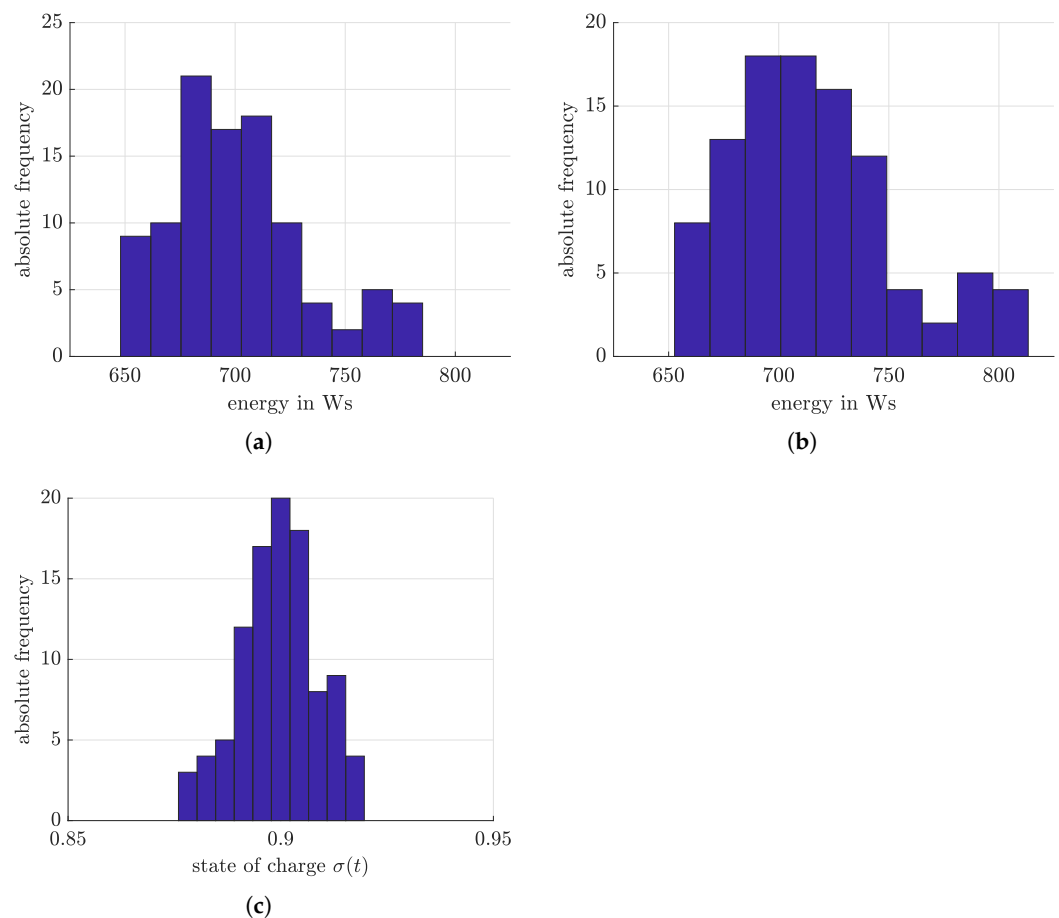


Figure 8. Sensitivity of the Ohmic losses and the terminal SOC for the model predictive control. (a) Variability of the Ohmic losses occurring on the time interval $t \in [0; 3600]$ s; (b) Variability of the Ohmic losses occurring on the time interval $t \in [0; 7200]$ s; (c) Variability of the SOC at the end of the charging phase.

4.3. Feedback Control Based on a Linear-Quadratic Regulator Design

Besides the use of a model predictive control approach, the design of a linear-quadratic state feedback controller is a further option to adapt the charging current to the state estimates of the battery equivalent circuit model at each point in time t_k . Using the same noise model as in the previous subsection and also a control update at integer multiples of 120 s, the feedback controller derived in Section 3.3 with the parameters $\alpha = 1$, $\gamma = 100$, as well as $\bar{R}_{TS} = R_{TS}(0.7)$ and $\bar{R}_{TL} = R_{TL}(0.7)$ leads to the charging dynamics summarized in Figure 9.

As in the previous subsection, this simulation was repeated 100 times to quantify one-standard-deviation tolerance bounds for the terminal current i_T and for the SOC σ . Those tolerance bounds are shown in Figure 9 as the yellow bounds along the complete trajectory, as well as in the form of histograms in Figure 10 for the Ohmic losses and the SOC at the points of time $t = 3600$ s and $t = 7200$ s.

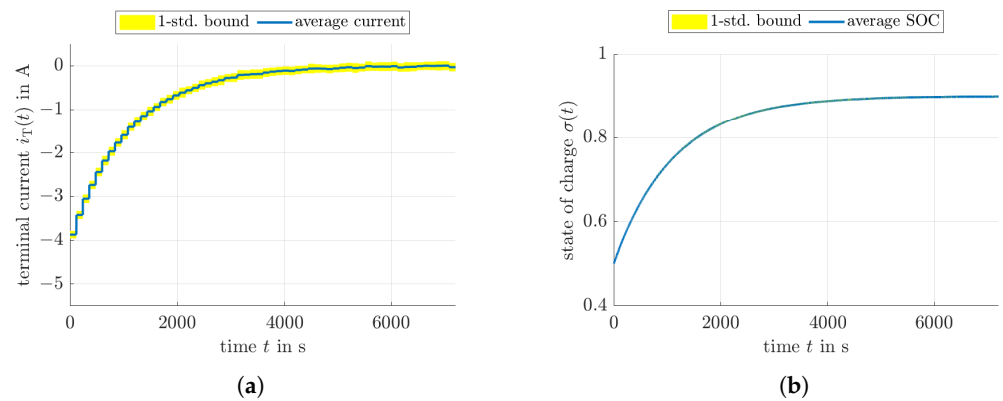


Figure 9. Feedback control under the influence of state reconstruction noise. (a) Terminal current $i_T(t)$; (b) SOC $\sigma(t)$.

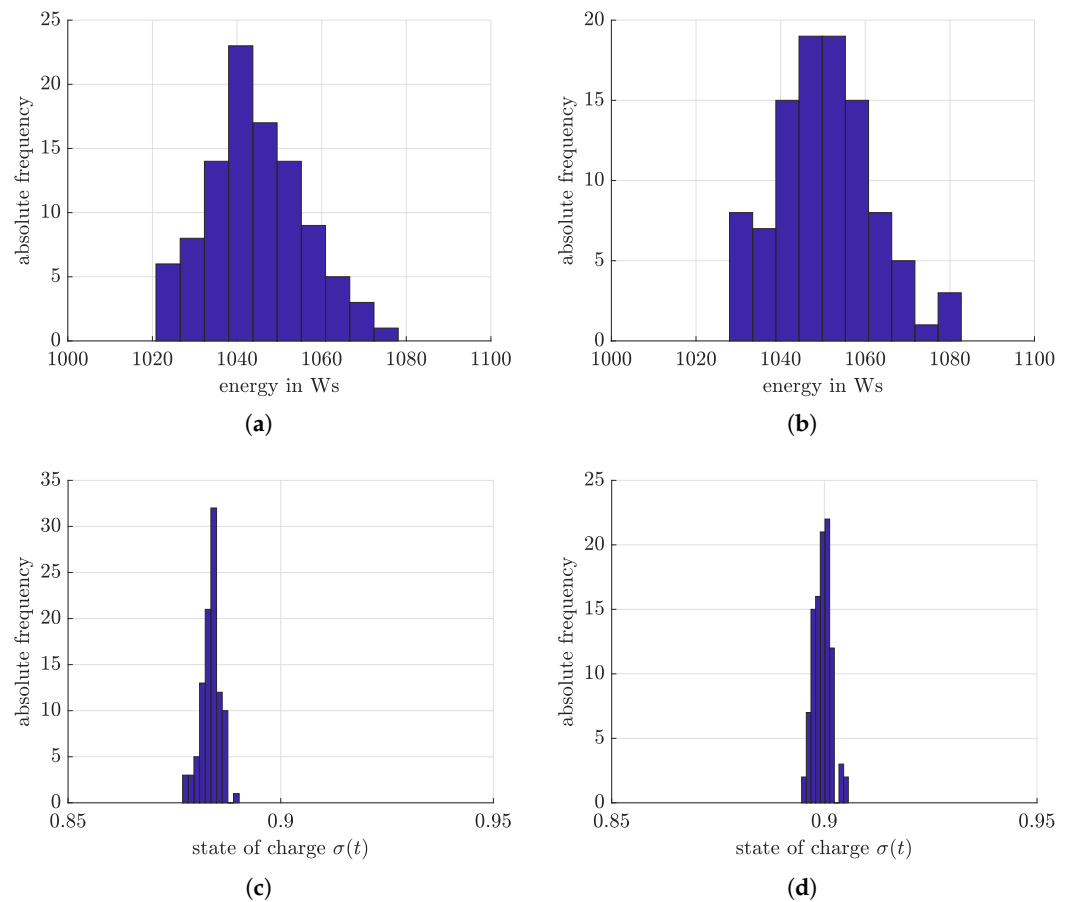


Figure 10. Sensitivity of the Ohmic losses and the terminal SOC for the feedback control approach. (a) Variability of the Ohmic losses occurring on the time interval $t \in [0 ; 3600]$ s; (b) Variability of the Ohmic losses occurring on the time interval $t \in [0 ; 7200]$ s; (c) Variability of the SOC at $t = 3600$ s; (d) Variability of the SOC at $t = 7200$ s.

It can be seen clearly that the feedback controller (based on the integral quadratic cost function) leads to significantly larger Ohmic losses than the previous two approaches. However, the advantage of this approach, becoming visible in Figure 10d, is the reduction of the remaining uncertainty of the SOC. For that reason, a practical application of this approach would be given by determining an optimal reference trajectory according to the offline evaluation of Pontryagin's maximum principle (Section 3.1.2) and using the feedback

controller to compensate the remaining deviations from the optimal (open-loop) charging profile. This approach will have similar Ohmic losses as the predictive controller presented in the previous subsection, however with the advantage of less online computational effort. Its disadvantage, however, will be the fact that the reference trajectory is restricted to a single initial point with a fixed charging duration, unless these are not pre-computed in a look-up-table-like manner for various initial SOC values and various process durations. This disadvantage can be avoided easily by the reinforcement learning approach investigated in detail in the following subsection.

4.4. Optimized Charging Based on Reinforcement Learning

To implement a fundamental reinforcement learning control approach that is comparable with the previous subsections, assume a given initial SOC and the desired final SOC σ_0 and σ_f , respectively. In addition, the charging duration is set to the fixed length of 3600 s.

Now, a reinforcement learning environment is set up as shown in Figure 4 according to the details listed in Appendix A, where all relevant information concerning the numbers of neurons and the types of activation functions in each layer (both for the critic and actor networks), as well as the parameters of the learning agent are summarized in terms of MATLAB code excerpts.

Note that the reinforcement learning control cannot guarantee a perfect SOC of σ_f at $t = 3600$ s. Therefore, this time span is followed by a 120 s period with constant charging current to exactly achieve the desired SOC before the relaxation phase takes place until $t = 7200$ s.

To make sure that the learning is close to the application of Pontryagin's maximum principle discussed above, the running costs (included in the reward function) are chosen as in Equation (13) with $\alpha = 1$, to which the additive time-dependent penalty term

$$5 \cdot 10^3 \cdot (\sigma(t) - \sigma_f)^2 \cdot \exp\left(\frac{t - 3600}{50}\right) + 10^4 \cdot \delta(t) \quad (35)$$

was added to obtain charging profiles approaching the desired SOC and, by means of

$$\delta(t) = \begin{cases} (\sigma(t) - 0.05)^2 & \text{for } \sigma(t) < 0.05 \\ (\sigma(t) - 0.95)^2 & \text{for } \sigma(t) > 0.95 \\ 0 & \text{else,} \end{cases} \quad (36)$$

preventing the optimizer from using control strategies violating the physical constraints on the SOC. Note that the scaling parameters in (35) were selected in such a way that undesired and physically meaningless solutions are penalized at least by one order of magnitude worse than the optimal control policies according to Sections 3.1.2 and 3.2.

Using these settings, the results summarized in Figure 11 were obtained. In detail, Figure 11a,b indicate the case that the noise in Figure 4 is deactivated both during training and application of the trained agent.

Figure 11c,d contain the noise-free training and the evaluation of the learning agent with the same noise standard deviations that were also used in Section 4.2; finally, Figure 11e,f show the cases that the noise was active during the training and the application of the trained agents.

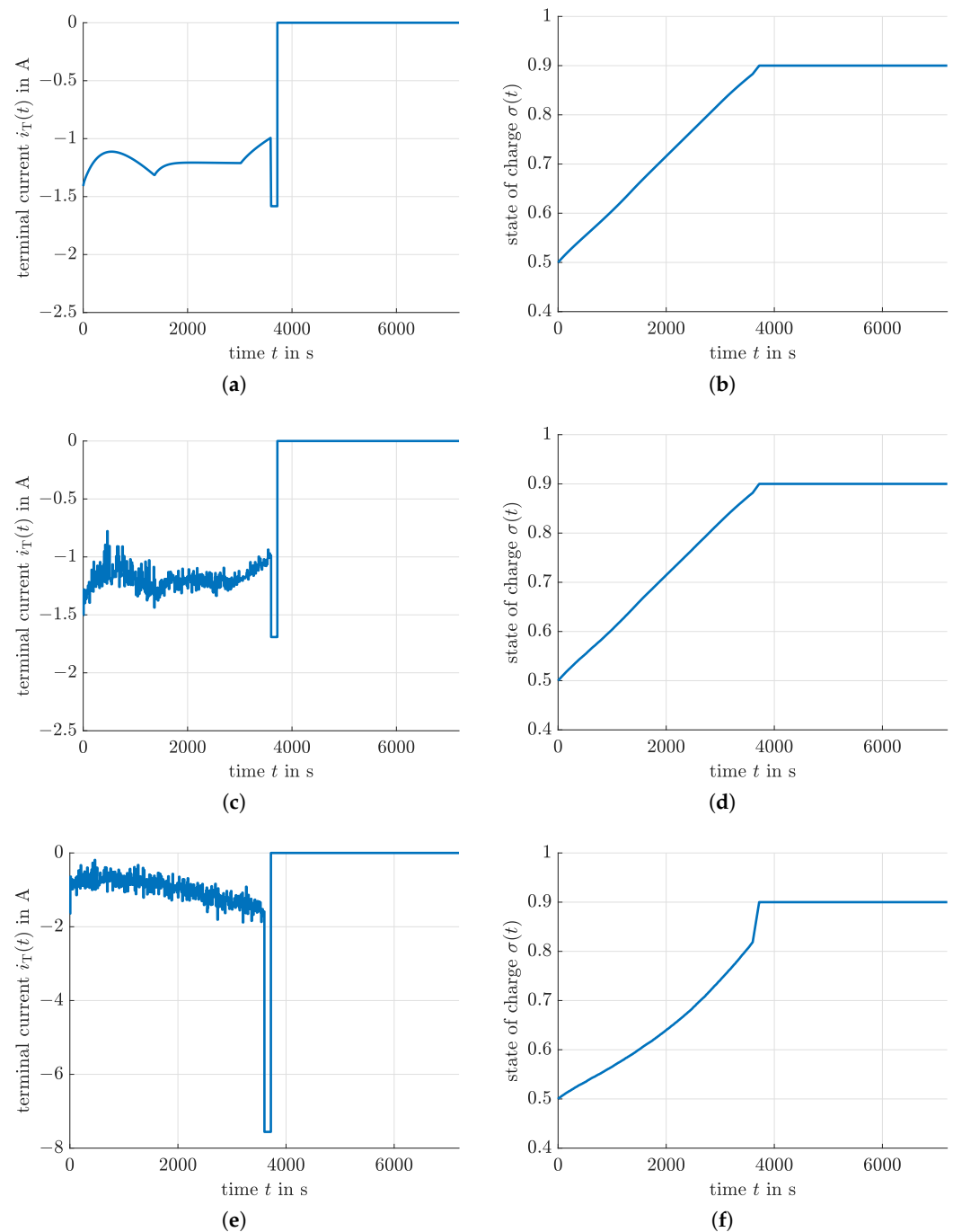


Figure 11. Evaluation of the control optimization by means of reinforcement learning. (a) Terminal current $i_T(t)$ (optimization and evaluation without noise); (b) SOC $\sigma(t)$ (optimization and evaluation without noise); (c) Terminal current $i_T(t)$ (optimization without noise; evaluation with noise); (d) SOC $\sigma(t)$ (optimization without noise; evaluation with noise); (e) Terminal current $i_T(t)$ (optimization and evaluation with noise); (f) SOC $\sigma(t)$ (optimization and evaluation with noise).

The corresponding Ohmic losses for these three different cases are summarized in Table 4. It becomes obvious that these values are close to the solutions obtained by means of the offline control optimization (cf. Table 3) with the help of Pontryagin's maximum principle and also the predictive control approach according to Figure 8. The big advantage of this solution is its small online computational effort (no boundary value problems have to be solved at runtime) and its simple generalizability to different charging durations and different initial conditions. To perform this generalization, it is only necessary to provide

suitable observations to the reinforcement learning agent during the offline training. Future work concerning this task will especially deal with incorporating information on measurement uncertainty (as well as state reconstruction uncertainty) into the training, not only by the Gaussian noise processes used in Figure 11e,f, but also by set-valued counterparts. In addition, it will be investigated how the reward functions can be generalized reasonably to minimize the sensitivity of the resulting control approximation against such uncertainty, for example by directly including the noise-induced control signal variances as a further penalty term in the reward function.

Table 4. Ohmic losses (energy) during the charging and relaxation phases for the different variants of the reinforcement learning control procedure.

Case According to	Losses during Charging ($t \in [0 ; 3600]$ s)	Total Losses ($t \in [0 ; 7200]$ s)
Figure 11a,b	637.62 Ws	683.41 Ws
Figure 11c,d	636.43 Ws	685.61 Ws
Figure 11e,f	467.36 Ws	1163.7 Ws

Remark 4. It is possible that the total Ohmic losses over the complete time span $t \in [0 ; 7200]$ s for the reinforcement learning control approach are smaller than those for the offline-optimized solution using Pontryagin’s maximum principle. This is foremost caused by the fact that the terminal SOC is not predefined and that the additional 120 s time interval available for achieving the desired terminal SOC leads to a reduction of the maximum current, entering the expression for the Ohmic losses in a quadratic form.

5. Conclusions and Outlook on Future Work

In this paper, a thorough comparison between the use of Pontryagin’s maximum principle and reinforcement learning control approaches for the optimization of the charging strategies of Lithium-ion batteries was presented. This comparison was based on an experimentally validated equivalent circuit model. It was shown that the reinforcement learning control approach can achieve robust, close-to-optimal solutions despite the presence of state reconstruction noise if the utilized reward function for the learning agents is extended in a problem-specific manner.

Future work will especially aim at the consideration of more general uncertainty models included in the training phase, as well as the derivation of reinforcement-learning-based SOC equalization strategies when multiple Lithium-ion battery cells are electrically connected in series and in parallel.

Author Contributions: Conceptualization, A.R., M.L. and O.B.; data curation, A.R., M.L. and O.B.; formal analysis, A.R.; investigation, A.R.; methodology, A.R.; software, A.R.; validation, A.R., M.L. and O.B.; visualization, A.R.; writing—original draft, A.R.; writing—review and editing, A.R., M.L. and O.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All data are included in the article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Parameterization of the Reinforcement Learning Approach

Appendix A.1. Parameterization of the Critic Network

The following code excerpt gives details on the layer configuration of the critic network, composed of fully connected layers and layers with both the *ReLU* and *tanh* activation functions.

```
statePath = [
    featureInputLayer(numObs, 'Normalization', 'none', 'Name', 'observation')
    fullyConnectedLayer(200, 'Name', 'CriticStateFC1')
```

```

reluLayer('Name', 'CriticRelu1')
fullyConnectedLayer(150, 'Name', 'CriticStateFC2');
actionPath = [
featureInputLayer(1, 'Normalization', 'none', 'Name', 'action')
fullyConnectedLayer(150, 'Name', 'CriticActionFC1', 'BiasLearnRateFactor', 0)];
commonPath = [
additionLayer(2, 'Name', 'add')
reluLayer('Name', 'CriticCommonRelu')
fullyConnectedLayer(1, 'Name', 'CriticOutput')];

```

Appendix A.2. Parameterization of the Actor Network

The actor network, parameterized according to the following code excerpt, is parameterized in such a way that the maximum absolute value of the charging current i_T is limited to 10 A.

```

actorNetwork = [
featureInputLayer(numObs, 'Normalization', 'none', 'Name', 'observation')
fullyConnectedLayer(200, 'Name', 'ActorFC1')
reluLayer('Name', 'ActorRelu1')
fullyConnectedLayer(150, 'Name', 'ActorFC2')
reluLayer('Name', 'ActorRelu2')
fullyConnectedLayer(1, 'Name', 'ActorFC3')
tanhLayer('Name', 'ActorTanh')
scalingLayer('Name', 'ActorScaling', 'Scale', max(actInfo.UpperLimit));
% upper limit = 10

```

Appendix A.3. Parameterization of the Learning Agent

All parameters listed below form the parameterization of the learning agent, where especially the sampling time of 10 s is of practical importance as it determines the temporal difference between two subsequent points in time at which the charging current for the Lithium-ion battery is updated.

```

agentOpts = rlDDPGAgentOptions(...
'SampleTime', 10, ...
'TargetSmoothFactor', 1e-3, ...
'ExperienceBufferLength', 1e5, ...
'NumStepsToLookAhead', 1, ...
'DiscountFactor', 0.99, ...
'MiniBatchSize', 128);
agentOpts.NoiseOptions.Variance = 1e-1;
agentOpts.NoiseOptions.VarianceDecayRate = 1e-5;

```

References

1. Xiong, R.; He, H.; Guo, H.; Ding, Y. Modeling for Lithium-Ion Battery used in Electric Vehicles. *Procedia Eng.* **2011**, *15*, 2869–2874. [\[CrossRef\]](#)
2. Kennedy, B.; Patterson, D.; Camilleri, S. Use of Lithium-Ion Batteries in Electric Vehicles. *J. Power Sources* **2000**, *90*, 156–162. [\[CrossRef\]](#)
3. Han, U.; Kang, H.; Song, J.; Oh, J.; Lee, H. Development of Dynamic Battery Thermal Model Integrated with Driving Cycles for EV Applications. *Energy Convers. Manag.* **2021**, *250*, 114882. [\[CrossRef\]](#)
4. Pesaran, A.A. Battery Thermal Models for Hybrid Vehicle Simulations. *J. Power Sources* **2002**, *110*, 377–382. [\[CrossRef\]](#)
5. Liu, Y.C.; Chang, S.B. Design and Implementation of a Smart Lithium-Ion Battery Capacity Estimation System for E-Bike. *World Electr. Veh. J.* **2011**, *4*, 370–378. [\[CrossRef\]](#)
6. Nguyen, V.T.; Pyung, H.; Huynh, T. Computational Analysis on Hybrid Electric Motorcycle with Front Wheel Electric Motor Using Lithium Ion Battery. In Proceedings of the 2017 International Conference on System Science and Engineering (ICSSE), Ho Chi Minh City, Vietnam, 21–23 July 2017; pp. 355–359.
7. Chen, W.; Liang, J.; Yang, Z.; Li, G. A Review of Lithium-Ion Battery for Electric Vehicle Applications and Beyond. *Energy Procedia* **2019**, *158*, 4363–4368. [\[CrossRef\]](#)

8. Misyris, G.S.; Marinopoulos, A.; Doukas, D.I.; Tegnér, T.; Labridis, D.P. On Battery State Estimation Algorithms for Electric Ship Applications. *Electr. Power Syst. Res.* **2017**, *151*, 115–124. [[CrossRef](#)]
9. Haji Akhoundzadeh, M.; Panchal, S.; Samadani, E.; Raahemifar, K.; Fowler, M.; Fraser, R. Investigation and Simulation of Electric Train Utilizing Hydrogen Fuel Cell and Lithium-Ion Battery. *Sustain. Energy Technol. Assess.* **2021**, *46*, 101234. [[CrossRef](#)]
10. Feehall, T.; Forsyth, A.; Todd, R.; Foster, M.; Gladwin, D.; Stone, D.; Strickland, D. Battery Energy Storage Systems for the Electricity Grid: UK Research Facilities. In Proceedings of the 8th IET International Conference on Power Electronics, Machines and Drives (PEMD 2016), Glasgow, UK, 19–21 April 2016; pp. 1–6.
11. Cao, J.; Schofield, N.; Emadi, A. Battery Balancing Methods: A Comprehensive Review. In Proceedings of the IEEE Vehicle Power and Propulsion Conference, Harbin, China, 3–5 September 2008; pp. 1–6.
12. Zhang, X.; Li, L.; Zhang, W. Review of Research about Thermal Runaway and Management of Li-ion Battery Energy Storage Systems. In Proceedings of the 9th IEEE International Power Electronics and Motion Control Conference (IPEMC2020-ECCE Asia), Nanjing, China, 29 November–2 December 2020; pp. 3216–3220.
13. Wang, Q. State of Charge Equalization Control Strategy of Modular Multilevel Converter with Battery Energy Storage System. In Proceedings of the 5th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, 12–14 September 2020; pp. 316–320.
14. Speltino, C.; Stefanopoulou, A.; Fiengo, G. Cell Equalization in Battery Stacks Through State Of Charge Estimation Polling. In Proceedings of the 2010 American Control Conference, Baltimore, MD, USA, 30 June–2 July 2010; pp. 5050–5055.
15. Raeber, M.; Heinzelmann, A.; Abdeslam, D.O. Analysis of an Active Charge Balancing Method Based on a Single Nonisolated DC/DC Converter. *IEEE Trans. Ind. Electron.* **2021**, *68*, 2257–2265. [[CrossRef](#)]
16. Quraan, M.; Abu-Khaizaran, M.; Sa’ed, J.; Hashlamoun, W.; Tricoli, P. Design and Control of Battery Charger for Electric Vehicles Using Modular Multilevel Converters. *IET Power Electron.* **2021**, *14*, 140–157. [[CrossRef](#)]
17. Kim, T.; Qiao, W.; Qu, L. Series-Connected Self-Reconfigurable Multicell Battery. In Proceedings of the Twenty-Sixth Annual IEEE Applied Power Electronics Conference and Exposition (APEC), Fort Worth, TX, USA, 6–11 March 2011; pp. 1382–1387.
18. Ci, S.; Zhang, J.; Sharif, H.; Alahmad, M. A Novel Design of Adaptive Reconfigurable Multicell Battery for Power-Aware Embedded Networked Sensing Systems. In Proceedings of the IEEE GLOBECOM 2007—IEEE Global Telecommunications Conference, Washington, DC, USA, 26–30 November 2007; pp. 1043–1047.
19. Visairo, H.; Kumar, P. A Reconfigurable Battery Pack for Improving Power Conversion Efficiency in Portable Devices. In Proceedings of the 7th International Caribbean Conference on Devices, Circuits and Systems, Cancun, Mexico, 28–30 April 2008; pp. 1–6.
20. Stengel, R. *Optimal Control and Estimation*; Dover Publications, Inc.: Mineola, NY, USA, 1994.
21. Leitmann, G. *An Introduction to Optimal Control*; McGraw-Hill: New York, NY, USA, 1966.
22. Athans, M.; Falb, P.L. *Optimal Control: An Introduction to the Theory and Its Applications*; McGraw-Hill: New York, NY, USA, 1966.
23. Pontryagin, L.S.; Boltjanskij, V.G.; Gamkrelidze, R.V.; Misčenko, E.F. *The Mathematical Theory of Optimal Processes*; Interscience Publishers: New York, NY, USA, 1962.
24. Buşoniu, L.; de Bruin, T.; Tolić, D.; Kober, J.; Palunko, I. Reinforcement Learning for Control: Performance, Stability, and Deep Approximators. *Annu. Rev. Control* **2018**, *46*, 8–28. [[CrossRef](#)]
25. Gosavi, A. Reinforcement Learning: A Tutorial Survey and Recent Advances. *Inform. J. Comput.* **2009**, *21*, 178–192. [[CrossRef](#)]
26. Borrelli, F.; Bemporad, A.; Morari, M. *Predictive Control for Linear and Hybrid Systems*; Cambridge University Press: Cambridge, UK, 2017.
27. Maciejowski, J. *Predictive Control with Constraints*; Prentice Hall: Essex, UK, 2002.
28. Sontag, E. *Mathematical Control Theory—Deterministic Finite Dimensional Systems*; Springer: New York, NY, USA, 1998.
29. Reuter, J.; Mank, E.; Aschemann, H.; Rauh, A. Battery State Observation and Condition Monitoring Using Online Minimization. In Proceedings of the 21st International Conference on Methods and Models in Automation and Robotics, Miedzyzdroje, Poland, 29 August–1 September 2016.
30. Erdinc, O.; Vural, B.; Uzunoglu, M. A Dynamic Lithium-Ion Battery Model Considering the Effects of Temperature and Capacity Fading. In Proceedings of the International Conference on Clean Electrical Power, Capri, Italy, 9–11 June 2009; pp. 383–386.
31. Rauh, A.; Butt, S.; Aschemann, H. Nonlinear State Observers and Extended Kalman Filters for Battery Systems. *Intl. J. Appl. Math. Comput. Sci. AMCS* **2013**, *23*, 539–556. [[CrossRef](#)]
32. Hildebrandt, E.; Kersten, J.; Rauh, A.; Aschemann, H. Robust Interval Observer Design for Fractional-Order Models with Applications to State Estimation of Batteries. In Proceedings of the 21st IFAC World Congress, Berlin, Germany, 11–17 July 2020.
33. Wang, B.; Liu, Z.; Li, S.; Moura, S.; Peng, H. State-of-Charge Estimation for Lithium-Ion Batteries Based on a Nonlinear Fractional Model. *IEEE Trans. Control Syst. Technol.* **2017**, *25*, 3–11. [[CrossRef](#)]
34. Zou, C.; Zhang, L.; Hu, X.; Wang, Z.; Wik, T.; Pecht, M. A Review of Fractional-Order Techniques Applied to Lithium-Ion Batteries, Lead-Acid Batteries, and Supercapacitors. *J. Power Sources* **2018**, *390*, 286–296. [[CrossRef](#)]
35. Plett, G. Extended Kalman Filtering for Battery Management Systems of LiPB-Based HEV Battery Packs—Part 1. Background. *J. Power Sources* **2004**, *134*, 252–261. [[CrossRef](#)]
36. Plett, G. Extended Kalman Filtering for Battery Management Systems of LiPB-Based HEV Battery Packs—Part 2. Modeling and Identification. *J. Power Sources* **2004**, *134*, 262–276. [[CrossRef](#)]

37. Plett, G. Extended Kalman Filtering for Battery Management Systems of LiPB-Based HEV Battery Packs—Part 3. State and Parameter Estimation. *J. Power Sources* **2004**, *134*, 277–292. [[CrossRef](#)]
38. Bo, C.; Zhifeng, B.; Binggang, C. State of Charge Estimation Based on Evolutionary Neural Network. *J. Energy Convers. Manag.* **2008**, *49*, 2788–2794. [[CrossRef](#)]
39. Rauh, A.; Chevet, T.; Dinh, T.N.; Marzat, J.; Raïssi, T. Robust Iterative Learning Observers Based on a Combination of Stochastic Estimation Schemes and Ellipsoidal Calculus. In Proceedings of the 25th International Conference on Information Fusion (FUSION), Linköping, Sweden, 4–7 July 2022; pp. 1–8.
40. Lahme, M.; Rauh, A. Combination of Stochastic State Estimation with Online Identification of the Open-Circuit Voltage of Lithium-Ion Batteries. In Proceedings of the 1st IFAC Workshop on Control of Complex Systems (COSY 2022), Bologna, Italy, 24–25 November 2022.
41. Friedland, B. Quasi-Optimum Control and the SDRE Method. In Proceedings of the 17th IFAC Symposium on Automatic Control in Aerospace, Toulouse, France, 25–29 June 2007; pp. 762–767.
42. Mracek, C.P.; Cloutier, J.R. Control Designs for the Nonlinear Benchmark Problem via the State-Dependent Riccati Equation Method. *Int. J. Robust Nonlinear Control* **1998**, *8*, 401–433. [[CrossRef](#)]
43. Çimen, T. State-Dependent Riccati Equation (SDRE) Control: A Survey. In Proceedings of the 17th IFAC World Congress, Seoul, South Korea, 6–11 July 2008; pp. 3761–3775.