2021 European Control Conference (ECC)
June 29 - July 2, 2021. Rotterdam, Netherlands

Battery State of Health Estimation via Reinforcement Learning

D. Natella, *Member IEEE*, F. Vasca, *Senior Member IEEE*

***Abstract*— The state of health of a battery characterizes its performance in terms of loss of capacity compared to the beginning of its life. This paper proposes a reinforcement learning algorithm for identifying the capacity of lithium-ion batteries. The training phase of the algorithm is based on data derived from constant current and constant voltage charging operations. The technique exploits a state observer based on a dynamic model of the battery and on the capacity estimation obtained with the reinforcement learning technique. The reward is defined as the error between the estimated and measured battery voltage. The effectiveness of the proposed solution is validated by considering different C-rates battery charging.**

I. Introduction

The continuously growing use of batteries as flexible power sources requires the development of specific management systems aimed at controlling of charging and discharging operations in order to ensure safe, reliable and efficient operations under demanding operating conditions. Due to its agile implementation and low computational load, a common solution adopted in lithium-ion battery management systems consists of using a dynamic equivalent circuit model (ECM) [1]. In this model the open circuit voltage depends on the state of charge (SOC). In particular, the battery capacity is a parameter which relates the evaluation of the SOC with the estimation of the state of health (SOH) [2]–[4]. The battery capacity determines the SOH which is used for monitoring the battery functionalities and provides information on possible maintenance and substitution interventions [5].

Among others, the battery capacity is a model parameter which depends on the type of accumulator considered. At the beginning of the battery life its capacity can be obtained through experimental tests and measurements of the ampere-hour charged and discharged during a complete cycle. To get aging information the voltage curves can be analyzed by using (offline) differential voltage techniques [6].

The capacity of a battery changes during its life. Therefore, online identification techniques are required in order to get a more precise estimation of the SOH [7]. For instance, in automotive vehicles is fundamental to have an onboard estimation of this parameter [8]. Online estimation algorithms used to quantify the degradation of lithium-ion batteries techniques can be roughly classified as model-based and data-driven. Adaptive algorithms are commonly based on electrochemical and ECM models [9], [10], while data-driven techniques can be defined as a black box model which interacts with the real battery and makes advanced classifications by using intelligent optimization algorithms [11]. An estimation of the battery capacity can be obtained by considering the constant-current (CC)/constant-voltage (CV) charging operations. In particular, in [12] it is shown that a possible reduction of the battery capacity can be identified by comparing the current time evolution during different CV phases.

The estimation technique of SOH proposed in this paper combines the advantages of data-driven approaches with those coming from the use of an ECM to describe the battery behaviour. To the best of our knowledge this paper is the first one that proposes reinforcement learning (RL) techniques for the battery SOH estimation. As a model-free technique, RL represents a valuable solution in engineering applications where learning processes from an uncertain environment play a key role [13], [14]. In control theory this approach has been widely used. For instance, RL has been proposed for the design of controllers (agents) which must take actions applied to an unknown environment by maximizing a cumulative reward [15].

In the proposed RL estimation procedure, the training phase is based on a learning approach from the interaction with the battery model and real measurements of current and voltage during CC-CV operating conditions. The reward of the algorithm is defined as the integral of the error between the battery voltage and the one obtained from a SOC observer. The proposed approach is validated through different cycles of charge and C-rates.

The reminder of this paper is organized as follows. In Section II the battery model is presented. The proposed RL approach is described in Section III and its implementation for the battery capacity estimation in Section IV. The effectiveness of the proposed solution is verified in Section V through numerical simulations based on real data. In Section VI some conclusions are summarized.

II. Battery model

In this section the ECM used for the design of the proposed RL algorithm is described.

A. Dynamic subsystem

A typical approach for representing the dynamic electrical behavior of a battery consists of using an ECM. This type of model is widely applied for the design of battery management systems and vehicular energy management systems [16], [17].

Department of Engineering, University of Sannio, 82100 Benevento, Italy. Email: {dnatella, vasca}@unisannio.it.

A class of ECM can be represented in the following state-space form

$$R_1 C_1 \, \dot{v}_1 = -R_1 i_b - v_1 \tag{1a}$$

$$Q \, S\dot{O}C = -i_b \tag{1b}$$

$$e_b = OCV(SOC) - R_0 i_b + v_1 \tag{1c}$$

where $\dot{v}_1$ and $S\dot{O}C$ indicate the continuous-time derivatives of the internal battery voltage v_1 and the state of charge SOC, respectively.

The input of the model is the battery current i_b which is assumed with a positive sign when the battery is operating as a generator by supplying some loads.

The first state variable of the model is the internal equivalent voltage v_1, see (1a). The resistor-capacitor branch which consists of R_1 and C_1 is used to describe the diffusion phenomena in the battery, see [8], [17], [18]. Some authors consider ECMs with more resistor-capacitor branches in order to obtain a better fitting of the nonlinear behaviour of the battery voltage. The model considered in this paper includes a single branch, so as shown in Fig. 1 and in the corresponding equations (1). This widely used simplified model has been experimentally confirmed to be effective [10] and can be considered to be sufficiently accurate for the analysis proposed herein.

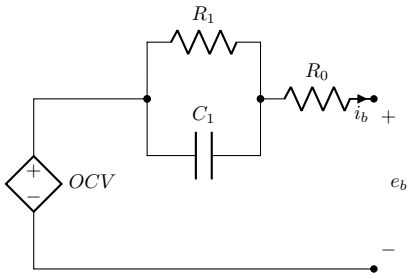

Fig. 1. Equivalent circuit model of the battery.

The second state variable of the model (1) is the state of charge SOC. In the corresponding equation (1b) the self-discharge phenomenon has been neglected. The parameter Q is the battery capacity which plays a key role in the SOH determination, so as it will be shown in next section.

The output of the model is the battery voltage e_b which can be measured at the external terminals of the battery. The open circuit voltage which is denoted by OCV depends on the state of charge SOC through a suitable nonlinear map which will be described in next subsection. The series resistance R_0 is used to represent the resistive behavior at the cell's terminal voltage. This resistance allows one to represent the power dissipated by the cell as heat. Some authors have added a noise signal on the right hand side of (1c) in order to represent the model uncertainty, see [19].

B. The open circuit voltage characteristic

The characteristic $OCV(SOC)$ plays a key role for the determination of the battery SOH. Several representations for this map have been proposed in the literature.

A quite common approach for defining the function $OCV(SOC)$ comes from the so called Partnership for a New Generation of Vehicle (PNGV) battery model, see [10], [20]–[22]. Indeed, the PNGV model can be obtained from (1) by considering a representation of the characteristic $OCV(SOC)$ expressed as the sum of two terms, i.e.

$$OCV(SOC) = v_\ell(SOC) + v_{n\ell}(SOC) \tag{2}$$

where v_ℓ depends linearly on SOC and $v_{n\ell}(SOC)$ is a suitable nonlinear map.

In the PNGV model the voltage v_ℓ is obtained by integrating the following differential equation

$$C_b \dot{v}_\ell = -i_b \tag{3}$$

where the capacitor C_b weights the linear dependence of the open circuit voltage on the integral of the battery current. By comparing (3) and (1b) and by assuming that C_b and Q are constant, it follows that

$$v_\ell(SOC) = \alpha_0 + \alpha_1 SOC \tag{4}$$

where α_0 is a suitable constant and $\alpha_1 = Q/C_b$. In other words, if the equation (1b) is adopted in the dynamic model, the dynamic equation (3) is not required and v_ℓ in (2) represents the linearization of the characteristic $OCV(SOC)$ around some specific operating point.

The term $v_{n\ell}(SOC)$ represents the nonlinear contribution to the $OCV(SOC)$ function. A possible expression for $v_{n\ell}(SOC)$ consists of a sigmoid-like function which is used to capture the typical three voltage plateaus of the characteristic and the corresponding two transitions, see [23]–[25] for details.

The complete (nonlinear) third order dynamic model of the battery can be obtained by considering as state variables v_1, SOC and v_ℓ, the dynamic equations (1a)–(1b) and (3), the output equation (1c) and the function $OCV(SOC)$ defined by (2) with (4) and a suitably defined nonlinear function $v_{n\ell}(SOC)$.

C. State of health

The battery capacity Q is the model parameter used to define the SOH which can be expressed as

$$SOH = \frac{Q}{Q^*} \tag{5}$$

where Q^* is the battery capacity measured at the beginning of battery life. The value of Q^* is provided by the battery manufacturer and is obtained by implementing several controlled charging and discharging processes. In particular, each test is carried out by imposing specific current and/or voltage profiles such that, by starting from pre-defined initial values of SOC, some specific values of SOC are obtained at the end of the test [8].

During the battery operation, the desired combinations of electrical variables which allow to reproduce the offline process adopted for the evaluation of Q^* may never occur. Therefore suitable online techniques must be used for the

estimation of Q and the corresponding SOH during the battery life.

III. Reinforcement learning algorithm

In this section some contents on RL techniques are recalled in order to present the proposed algorithm. The interested reader should refer to [15] for more technicalities and formal details.

Let us define the quadruple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$, where $\mathcal{S}$ is the set of the states of the environment, $\mathcal{A}(s)$ is the set of actions which the agent can take given the state $s \in \mathcal{S}$, $\mathcal{R}$ is the set of possible rewards, $\mathcal{P}$ is the probability of transitioning to state $s' \in \mathcal{S}$ given that the environment is in the state $s \in \mathcal{S}$ and the agent takes the action $a \in \mathcal{A}(s)$. Furthermore, the control policy $\pi(s,a)$ is defined as the probability to take the action $a \in \mathcal{A}(s)$ given a state s, i.e. $\pi : \mathcal{S} \times \mathcal{U} \to [0,1]$.

At the discrete time-step $t \in \mathbb{N}_0$ the agent receives from the environment the state $s_t \in \mathcal{S}$. The value function $J_t^\pi(s) : \mathcal{S} \to \mathbb{R}$ under a policy π can be formulated as the expected return in a future time interval, say $T \in \mathbb{N}$, when starting from $s_t = s$ and following π:

$$J_t^\pi(s) = \mathbb{E}^\pi \left[\sum_{i=t}^{t+T} \gamma^{i-t} r_i \big| s_t = s \right] \tag{6}$$

where $r_t \in \mathcal{R}$ is the reward obtained from the environment at the discrete time-step t and $\gamma \in [0,1]$ is a parameter which weights how much the future events lose their values according to how far away in time they are.

Consider now the case that the policy π must be chosen within a given set of policies. Assume that at the discrete time-step t the state of the environment is $s_t = s$. A policy π_1 is said to be better than a policy π_2 if $J_t^{\pi_1}(s) > J_t^{\pi_2}(s)$ where the value functions are defined as in (6). Then, the optimal policy at t from the state $s_t = s$ is the policy corresponding to the largest value function:

$$J_t^*(s) = \max_\pi \; J_t^\pi(s). \tag{7}$$

By using the Bellman optimality principle one can find that the maximum value function satisfies the following backward recursive equation

$$J_t^*(s) = \max_\pi \sum_{a \in \mathcal{A}(s)} \pi(s,a) \sum_{s' \in \mathcal{S}} p_{ss'}^a [\rho_{ss'}^a + \gamma J_{t+1}^*(s')], \tag{8}$$

where $s = s_t$, $s' = s_{t+1}$, $a = a_t$, $p_{ss'}^a$ is the probability of the transition from the state s to s' given the action a, and $\rho_{ss'}^a = \mathbb{E}[r_t | s_t = s, a_t = a, s_{t+1} = s']$ is the expected value of the reward r_t starting from the state s_t, applying the action a_t and going to the state s_{t+1}. The optimal policy at the time-step t is given by $\pi^*(s,a) = arg_\pi \, J_t^*(s)$, where $J_t^*(s)$ is defined by the recursive equation (8). By assuming that there exists a deterministic optimal policy, one can think at such policy as a deterministic map of the action to be taken from each state. Then, the recursive equation (8) can be equivalently written as

$$J_t^*(s) = \max_{a \in \mathcal{A}(s)} \sum_{s' \in \mathcal{S}} p_{ss'}^a [\rho_{ss'}^a + \gamma J_{t+1}^*(s')]. \tag{9}$$

One can now introduce the Q-learning approach. Let us assume that at the time-step t the environment provides the state $s_t = s$ and the agent takes an arbitrary action $a \in \mathcal{A}(s)$. The quality function $\Theta_t^*(s,a) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, also called action-value function, is defined as the expected return obtained when the optimal policy is followed thereafter:

$$\Theta_t^*(s,a) = \sum_{s' \in \mathcal{S}} p_{ss'}^a [\rho_{ss'}^a + \gamma \, J_{t+1}^*(s')]. \tag{10}$$

By using (9) in (10) one can write

$$\Theta_t^*(s,a) = \sum_{s' \in \mathcal{S}} p_{ss'}^a [\rho_{ss'}^a + \gamma \max_{a' \in \mathcal{A}(s')} \Theta_{t+1}^*(s',a')], \tag{11}$$

which associates at each pair (s,a) the corresponding expectation of the short-term reward. On an infinite time horizon and under stationary policies, from (11) one can write

$$\Theta^*(s,a) = \sum_{s' \in \mathcal{S}} p_{ss'}^a [\rho_{ss'}^a + \gamma \max_{a' \in \mathcal{A}(s')} \Theta^*(s',a')]. \tag{12}$$

A possible approach for obtaining a solution of (12) is the temporal difference learning method. This technique is the one which will be adopted for the proposed battery capacity estimation. Let us assume that the action-value function is evaluated along a specific trajectory of the whole system. As a consequence, the probability $p_{ss'}^a$ for going from s to s' can be assumed to be unitary and the expected value of the reward $\rho_{ss'}^a$ can be assumed to be equal to the actual reward r_t obtained from the environment. Under these hypotheses an estimation of the action-value function can be obtained through the following iteration

$$\begin{aligned} \widetilde{\Theta}_{t+1}(s,a) = & \widetilde{\Theta}_t(s,a) + \alpha [r(s,a) \\ & + \gamma \max_{\widetilde{a} \in \mathcal{A}(s')} \widetilde{\Theta}_t(s',\widetilde{a}) - \widetilde{\Theta}_t(s,a)], \end{aligned} \tag{13}$$

where $\alpha \in (0,1]$ is the learning rate which determines to what extent newly acquired information overrides old information. An interpretation of (13) through the previous analysis can be obtained by assuming that the iterative equation (13) has a steady-state solution, say $\widetilde{\Theta}^*(s,a)$. In particular, by using $\widetilde{\Theta}^*(s,a) = \widetilde{\Theta}_t(s,a)$ for all t in (13) one can write

$$\widetilde{\Theta}^*(s,a) = r(s,a) + \gamma \max_{a' \in \mathcal{A}(s')} \widetilde{\Theta}^*(s',a'), \tag{14}$$

which has a straightforward interpretation by looking at (12) under the considered scenario.

The iterative equation (13) is the basis for the proposed estimation technique which is synthesized in Algorithm 1.

IV. SOH estimation

In this section the proposed RL strategy for the battery SOH estimation is presented.

A. Estimator block scheme

Figure 2 represents a block scheme of the training phase for the proposed SOH estimator based on the RL technique. During the validation phase the estimator provides the value of the battery SOH based on the trained map.

Algorithm 1: Q-learning pseudocode

Parameter: α, γ, Δ_r, Δ_{ep}, θ, $\mathcal{A}$, $\mathcal{S}$;
Input : r_t, s_t;
Output : a_t;
Initialize : $\widetilde{\Theta}_{\Delta_{ep}}(s,a) = 0,\ \forall (s,a) \in \mathcal{S} \times \mathcal{A}$;
$p_\varepsilon = 1$; $r_{av} = 0$;
begin
  while $r_{av} \leq \Delta_r$ **do**
    $t \leftarrow 0$; $r_{av} \leftarrow 0$;
    $\widetilde{\Theta}_t(s,a) \leftarrow \widetilde{\Theta}_{\Delta_{ep}}(s,a),\ \forall (s,a) \in \mathcal{S} \times \mathcal{A}$;
    read(s_t);
    $\hat{a} \leftarrow rand\{a\},\ a \in \mathcal{A}(s_t)$;
    $a_t \leftarrow \hat{a}$; apply(a_t);
    $t \leftarrow t + 1$;
    while $t \leq \Delta_{ep}$ **do**
      read(r_t, s_t);
      $\widetilde{\Theta}_t(s_{t-1}, a_{t-1}) \leftarrow$
       $(1-\alpha)\widetilde{\Theta}_{t-1}(s_{t-1}, a_{t-1})$
       $+\alpha(r_t + \gamma \max_{a \in \mathcal{A}(s_t)} \widetilde{\Theta}_{t-1}(s_t, a))$;
      $\widetilde{\Theta}_t(s,a) \leftarrow \widetilde{\Theta}_{t-1}(s,a)$,
       $\forall (s,a) \neq (s_{t-1}, a_{t-1})$;
      $r_{av} \leftarrow ((t-1)\, r_{av} + r_t)/t$;
      $a_{max} \leftarrow argmax_{a \in \mathcal{A}(s_t)} \widetilde{\Theta}_t(s_t, a)$;
      if $rand\{[0,1]\} > p_\epsilon$ **then**
        $\hat{a} \leftarrow a_{max}$;
      end
      else
        $\hat{a} \leftarrow rand\{a\},\ a \in \mathcal{A} \backslash \{a_{max}\}$;
      end
      $a_t \leftarrow \hat{a}$; apply(a_t);
      $t \leftarrow t + 1$;
    end
    $p_\varepsilon \leftarrow \theta p_\varepsilon$;
  end
end

At the top of the scheme the block *Battery* represents the real battery with the current i_b as an input and voltage e_b as an output, which are both measured at the battery terminals. The block *SOC observer*, by using the model (1), provides an estimation of the output voltage, say $\hat{e}_b$. The integral of the error over the entire t-th CC-CV operation determines the reward r_t for the t-th episode of Algorithm 1. The block *Agent* implements Algorithm 1 by computing a state which is obtained from the battery current, so as described below.

B. State for reinforcement learning

The RL SOH estimator described by Algorithm 1 requires the definition of the state s_t and reward r_t, where the discrete time-step t is the subscript used to indicate a single episode. An episode is defined as a battery charging consisting of a complete CC-CV operation. The action at the end of each episode is an estimation of the battery capacity, say $a_t = \hat{Q}_t$.

In order to define the state of each episode we exploit the

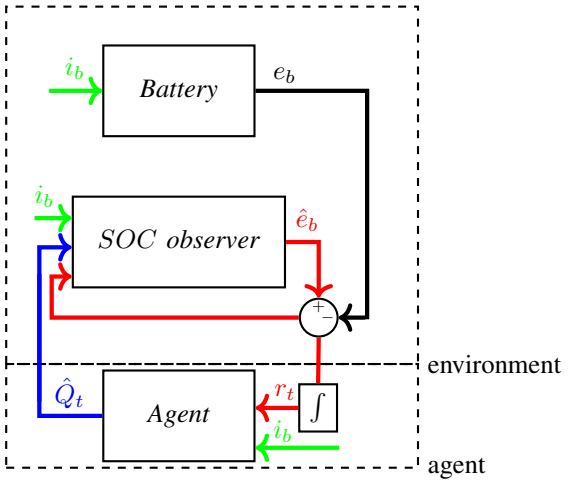

Fig. 2. Environment-agent scheme during the training phase for the battery capacity estimation.

integration of the differential equation (1b). Let us define the constant sampling period $\Delta \in \mathbb{R}$ and, for each episode, the following three sampling time instants: the time instant $k_{0t}\Delta$ with $k_{0t} \in \mathbb{N}$ at which the CC phase of the t-th episode starts; the time instant $k_{1t}\Delta$ with $k_{1t} \in \mathbb{N}$ when the CC phase ends and the CV phase starts; the time instant $k_{2t}\Delta$ with $k_{2t} \in \mathbb{N}$ when the CV phase ends. At the t-th episode from (1b) one can write

$$-\int_{k_{0t}\Delta}^{\bar{k}\Delta} i_b(\xi)\mathrm{d}\xi = \big(SOC(k_{0t}) - SOC(\bar{k})\big)\, Q \qquad (15)$$

where $\bar{k} \in \{k_{0t}, k_{0t}+1, \ldots, k_{2t}\} \subset \mathbb{N}$ is a generic time-step during the charging operation. By evaluating the integral in (15) with the backward Euler technique one obtains

$$-\sum_{k=k_{0t}}^{\bar{k}-1} i_b(k) = \frac{SOC(k_{0t}) - SOC(\bar{k})}{\Delta} Q. \qquad (16)$$

Say $\bar{I}_t$ the constant current of the CC phase of the t-th episode. From (16) at the end of the CC phase one obtains

$$-\bar{I}_t(k_{1t} - k_{0t} - 1) = \frac{SOC(k_{0t}) - SOC(k_{1t})}{\Delta} Q \qquad (17)$$

where k_{1t} is obtained as the smallest time instant of the t-th episode when the current is not constant, i.e.

$$k_{1t} = \min\{k \geq k_{0t} : i_b(k) \neq i_b(k-1)\}. \qquad (18)$$

In practice one can easily substitute the inequality condition in (18) with some threshold condition. The first component of the state s_t is given by the left hand side of (17).

The second component of the state is obtained by considering the CV phase only. In particular the discrete time instant k_{2t} is defined as the time instant when the absolute value of the current reaches a sufficiently small value, say $i_{\min}$. By rewriting (16) during the entire CV phase one obtains

$$-\sum_{k=k_{1t}}^{k_{2t}-1} i_b(k) = \frac{SOC(k_{1t}) - SOC(k_{2t})}{\Delta} Q \qquad (19)$$

where k_{2t} is obtained as the smallest time instant of the t-th episode when the current reaches $i_{\min}$, i.e. $k_{2t} = \min\{k \geq k_{1t} : |i_b(k)| \leq i_{\min}\}$. Clearly, at the beginning of each CV phase the procedure for the evaluation of k_{2t} must be restarted.

C. Reward for reinforcement learning

The typical battery measurements available are the voltage e_b and the current i_b. The battery current provides the state for Algorithm 1. The integral of the battery voltage measured during the charging operation is considered as the reward for the proposed RL approach.

The voltage estimation $\hat{e}_b$, which is the output of the block *SOC observer* in Fig. 2, is obtained from (1c) with the simplifying assumption $v_1 = 0$ which leads to

$$\hat{e}_b(k) = OCV(S\hat{O}C(k)) - R_0 i_b(k) \tag{20}$$

where $S\hat{O}C(k)$ is the state of charge to be estimated at step k. In order to obtain such estimation one can discretize (1b) by using the backward Euler technique which provides

$$S\hat{O}C(k) = S\hat{O}C(k-1) - \frac{\Delta}{\hat{Q}_{t-1}} i_b(k) + \gamma_o(e_b(k) - \hat{e}_b(k)) \tag{21}$$

where Δ is the sampling period used for the dynamic model of the battery, $\hat{Q}_{t-1}$ is the battery capacity which is determined at the $(t-1)$-th episode of activation of the RL algorithm with $t \in \mathbb{N}$, γ_o is the observer gain.

The reward for the RL algorithm can be written as

$$r_t = \sum_{k=k_{0t}}^{k_{2t}} -|e_b(k) - \hat{e}_b(k)| \tag{22}$$

where $\hat{e}_b(k)$ is given by (20) with (21).

V. Simulation Results

The effectiveness of the proposed reinforcement learning technique for the battery capacity estimation is verified by considering real data taken from [26]. The parameters of the battery model are: $R_0 = 11.2\,\Omega$, $R_1 = 2.5\,\Omega$, $C_1 = 1.03\,\mathrm{F}$, $Q^* = 7.13\,\mathrm{A\,h}$.

The parameters of the training phase implemented through Algorithm 1 are: $\alpha = 0.4$, $\gamma = 0.9$, $\Delta_{\mathrm{r}} = -0.05$, $\Delta_{\mathrm{ep}} = 100$, $\theta = 0.9$. Each episode corresponds to a real CC-CV charging cycle and several repetitions of the same cycle are considered for the training phase. The battery current is used as an input for the battery model (1) from which the battery voltage e_b is determined. In order to emulate the capacity fading phenomenon, during the training phase at every Δ_{ep} episodes the value of Q is decreased. The total number of episodes considered for the training phase is $5\Delta_{\mathrm{ep}}$ where for the different groups of Δ_{ep} episodes the value of Q has been chosen equal to Q^*, $0.75Q^*$, $0.55Q^*$, $0.25Q^*$ and $0.15Q^*$, respectively. The training phase was implemented in the Matlab/Simulink environment with a laptop having 16 GB of RAM and a quad-core CPU at 4.6 GHz; the total duration of the training phase was approximately 2 hours.

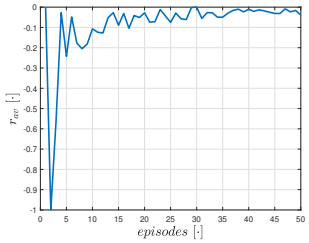

Fig. 3. Cumulative reward for the RL algorithm applied to a single CC-CV charging.

In Fig. 3 it is shown the cumulative reward r_{av} defined in Algorithm 1 with r_t given by (22), evaluated during the first section of the training phase corresponding to $Q = Q^*$. The error of the RL algorithm is less than 2.5% (the voltage is around $4.2\,\mathrm{V}$) for $t \geq 20$ episodes which is an acceptable time by considering that the variation of a battery capacity is a phenomenon which shows its effects at a greater time scale. The reward has a similar evolution also for the other sections of the training phase.

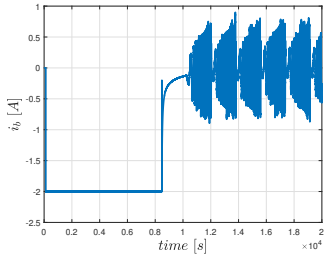

Fig. 4. Battery current i_b profile for the validation scenario.

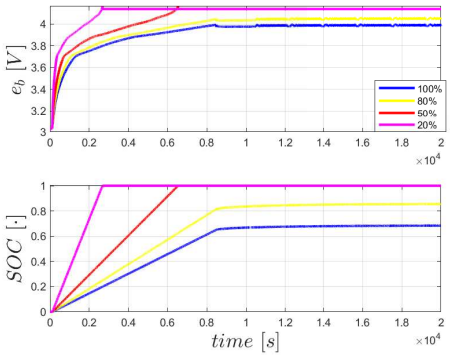

Fig. 5. Validation tests on a given i_b profile for different SOH values: battery voltage e_b (top), battery SOC (bottom) vs. time

The validation phase is tested by considering the realistic current profile shown in Fig. 4. Figure 5 shows the time evolution of e_b and SOC obtained by integrating the battery model with different values of Q. The values of the estimated variables obtained from the observer (21) are shown in Fig. 6. The initial condition of the observer is $S\hat{O}C(0) = 0.4$ for all numerical tests. After 100 s, the voltage error is less than 1% and the absolute value of the state of charge error is less

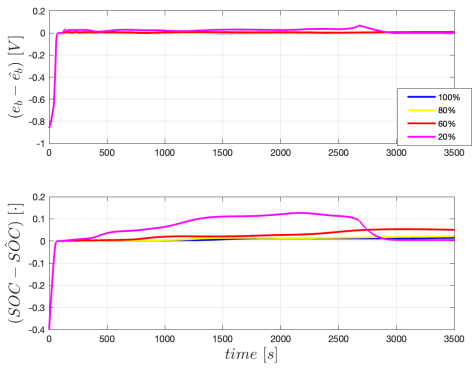

Fig. 6. Errors $e_b - \hat{e}_b$ (top) and $SOC - \hat{SOC}$ (bottom).

than 0.1. This error starts increasing when the battery voltage e_b is constant and the voltage error is negligible. However, these operating conditions can be easily detected and one could correspondingly stop the SOC estimation.

VI. Conclusion

State of health of batteries, defined as the ratio of the battery capacity with respect to its nominal value, is a crucial variable for the battery maintenance and energy management strategies. By assuming a uniform use of the battery, its capacity variation can be considered to be a long time-scale process which can be influenced by unexpected operating conditions too. Data coming from the battery history in the variety of these scenarios could be a useful source of information for getting an SOH estimation. The proposed technique combines data-driven and model-based approaches. A reinforcement learning algorithm provides the battery capacity estimation. The training phase of the algorithm is activated during the battery charging. In online implementation, the trained map determines the SOH estimation which is exploited by a model-based SOC observer. Numerical results obtained by emulating the battery aging have shown the effectiveness of the proposed technique in providing a good estimation of the battery capacity.

Further steps are required for moving towards an actual implementation of the proposed solution. In this line, directions for future work are the evaluation of the computational burden required for an on-board realization of the proposed strategy, the influence of internal resistance variations on the SOH estimation accuracy and the validation of the solution for different charging and discharging phases, and temperatures.

References

[1] X. Hu, S. Li, and H. Peng, "A comparative study of equivalent circuit models for Li-ion batteries," *J. Power Sources*, vol. 198, pp. 359–367, 2012.

[2] R. Xiong, J. Cao, Q. Yu, H. He, and F. Sun, "Critical review on the battery state of charge estimation methods for electric vehicles," *IEEE Access*, vol. 6, pp. 1832–1843, 2017.

[3] H. Chaoui, N. Golbon, I. Hmouz, R. Souissi, and S. Tahar, "Lyapunov-based adaptive state of charge and state of health estimation for lithium-ion batteries," *IEEE Trans. on Industrial Electronics*, vol. 62, no. 3, pp. 1610–1618, 2014.

[4] J. Du, Z. Liu, Y. Wang, and C. Wen, "An adaptive sliding mode observer for lithium-ion battery state of charge and state of health estimation in electric vehicles," *Control Engineering Practice*, vol. 54, pp. 81–90, 2016.

[5] S. Ebbesen, P. Elbert, and L. Guzzella, "Battery state-of-health perceptive energy management for hybrid electric vehicles," *IEEE Trans. on Vehicular Technology*, vol. 61, no. 7, pp. 2893–2900, 2012.

[6] L. Zheng, J. Zhu, D. D.-C. Lu, G. Wang, and T. He, "Incremental capacity analysis and differential voltage analysis based state of charge and capacity estimation for lithium-ion batteries," *Energy*, vol. 150, pp. 759–769, 2018.

[7] L. Tang, G. Rizzoni, and S. Onori, "Energy management strategy for HEVs including battery life optimization," *IEEE Trans. on Transportation Electr.*, vol. 1, no. 3, pp. 211–222, 2015.

[8] A. Farmann, W. Waag, A. Marongiu, and D. U. Sauer, "Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles," *J. Power Sources*, vol. 281, pp. 114–130, 2015.

[9] R. Klein, N. A. Chaturvedi, J. Christensen, J. Ahmed, R. Findeisen, and A. Kojic, "Electrochemical model based observer design for a lithium-ion battery," *IEEE Trans. on Control Systems Technology*, vol. 21, no. 2, pp. 289–301, 2012.

[10] H. He, R. Xiong, and J. Fan, "Evaluation of lithium-ion battery equivalent circuit models for state of charge estimation by an experimental approach," *Energies*, vol. 4, no. 4, pp. 582–598, 2011.

[11] X. Hu, S. E. Li, and Y. Yang, "Advanced machine learning approach for lithium-ion battery state estimation in electric vehicles," *IEEE Trans. on Transportation Electr.*, vol. 2, no. 2, pp. 140–149, 2015.

[12] A. Eddahech, O. Briat, and J.-M. Vinassa, "Determination of lithium-ion battery state-of-health based on constant-voltage charge phase," *J. Power Sources*, vol. 258, pp. 218–227, 2014.

[13] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. Boston, USA: MIT press, 2018.

[14] D. Natella and F. Vasca, "A Q-learning Approach for SoftECU Design in Hybrid Electric Vehicles," in *Proc. of IEEE International Conference on System Theory, Control and Computing*, Sinaia, Romania, 8–10 Oct. 2020, pp. 763–768.

[15] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control," *IEEE Control Systems Magazine*, vol. 32, no. 6, pp. 76–105, 2012.

[16] Y. He, W. Liu, and B. J. Koch, "Battery algorithm verification and development using hardware-in-the-loop testing," *J. Power Sources*, vol. 195, no. 9, pp. 2969–2974, 2010.

[17] D. Andre, C. Appel, T. Soczka-Guth, and D. U. Sauer, "Advanced mathematical methods of SOC and SOH estimation for lithium-ion batteries," *J. Power Sources*, vol. 224, pp. 20–27, 2013.

[18] S. Li, M. Stapelbroek, and J. Pfluger, "Model-In-The-Loop Testing of SOC and SOH Estimation Algorithms in Battery Management Systems," *SAE Technical Paper*, no. 2017-26-0094, 2017.

[19] I.-S. Kim, "A technique for estimating the state of health of lithium batteries through a dual-sliding-mode observer," *IEEE Trans. on Power Electronics*, vol. 25, no. 4, pp. 1013–1022, 2009.

[20] P. Nelson, I. Bloom, K. Amine, and G. Henriksen, "Design modeling of lithium-ion battery performance," *J. Power Sources*, vol. 110, no. 2, pp. 437–444, 2002.

[21] X. Liu, W. Li, and A. Zhou, "PNGV equivalent circuit model and SOC estimation algorithm for lithium battery pack adopted in AGV vehicle," *IEEE Access*, vol. 6, pp. 23 639–23 647, 2018.

[22] D. Haifeng, W. Xuezhe, and S. Zechang, "A new SOH prediction concept for the power lithium-ion battery used on HEVs," in *Proc. of Vehicle Power and Propulsion Conference*, Dearborn, MI, USA, 7–10 Sept. 2009, pp. 1649–1653.

[23] C. Weng, J. Sun, and H. Peng, "A unified open-circuit-voltage model of lithium-ion batteries for state-of-charge estimation and state-of-health monitoring," *J. Power Sources*, vol. 258, pp. 228–237, 2014.

[24] S.-C. Huang, K.-H. Tseng, J.-W. Liang, C.-L. Chang, and M. G. Pecht, "An online SOC and SOH estimation model for lithium-ion batteries," *Energies*, vol. 10, no. 4, p. 512, 2017.

[25] K. Khan, B. Zhou, and A. Rezaei, "Real time application of battery state of charge and state of health estimation," *SAE Technical Paper*, no. 2017-01-1199, 2017.

[26] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 2. Modeling and identification," *J. Power Sources*, vol. 134, no. 2, pp. 262–276, 2004.