State of Charge Estimation Using the Extended Kalman Filter for Battery Management Systems Based on the ARX Battery Model

Shifei Yuan *, Hongjie Wu and Chengliang Yin

National Engineering Laboratory for Automotive Electronic Control Technology, Shanghai Jiao Tong University, 200240 Shanghai, China; E-Mails: wuhongjie@sjtu.edu.cn (H.W.); clyin1965@sjtu.edu.cn (C.Y.)

* Author to whom correspondence should be addressed; E-Mail: kuaidian@sjtu.edu.cn; Tel./Fax: +86-21-3420-6805.

Received: 6 November 2012; in revised form: 11 December 2012 / Accepted: 24 December 2012 / Published: 17 January 2013

Abstract: State of charge (SOC) is a critical factor to guarantee that a battery system is operating in a safe and reliable manner. Many uncertainties and noises, such as fluctuating current, sensor measurement accuracy and bias, temperature effects, calibration errors or even sensor failure, etc. pose a challenge to the accurate estimation of SOC in real applications. This paper adds two contributions to the existing literature. First, the auto regressive exogenous (ARX) model is proposed here to simulate the battery nonlinear dynamics. Due to its discrete form and ease of implementation, this straightforward approach could be more suitable for real applications. Second, its order selection principle and parameter identification method is illustrated in detail in this paper. The hybrid pulse power characterization (HPPC) cycles are implemented on the 60AH LiFePO$_4$ battery module for the model identification and validation. Based on the proposed ARX model, SOC estimation is pursued using the extended Kalman filter. Evaluation of the adaptability of the battery models and robustness of the SOC estimation algorithm are also verified. The results indicate that the SOC estimation method using the Kalman filter based on the ARX model shows great performance. It increases the model output voltage accuracy, thereby having the potential to be used in real applications, such as EVs and HEVs.

Keywords: state of charge; ARX battery model; HPPC test; lithium-ion battery; extended Kalman filter
1. Introduction

As energy storage components, batteries play a significant role in the development and widespread market penetration of electric vehicles (EVs). For potential EV and HEV customers, safety and reliability of the high efficiency battery are the top concerns. Due to volatility, flammability and entropy changes, batteries have a potential to ignite or even explode when overcharged. These serious safety problems have caused several fatal accidents and also caused a great deal of public concern and skepticism about EVs and HEVs. Solutions to all these problems lie not only in battery technologies but also the management systems for the batteries. Battery management system (BMS), which integrate the battery system with the rest of the power train, play a vital role in improving battery performance, safety and reliability. Meanwhile, as one of the critical tasks of the BMS, accurate estimation of state of charge (SOC) has attracted considerable attention and effort.

Many different approaches to estimate SOC can be found in the literature. Accurate estimation of SOC is a challenge, mainly because of the uncertainties involved, such as temperature, varying power requests, aging effects and so on. Reviewing the extensive literature of SOC estimation, one can find that they all fall into two categories, those based on open-loop models and those based on closed-loop models.

Pattipati et al. [1] have summarized the methods most commonly used for SOC estimation, including Ah counting approach, open-circuit voltage (OCV) measurement, electrochemical models, electrochemical impedance spectroscopy, equivalent circuit models, neural network and fuzzy logic methods. All aforementioned models have their own advantages and suitable implementation situations. The Ah counting approach, as shown in Figure 1, is the basic approach followed directly by the definition of SOC and also the simplest. This method is accurate only if the current transducer is fairly precise and sufficient SOC re-calibration points are available. The major drawback is that it is sensitive to current measurement accuracy and needs regular re-calibration. As time goes on, voltage errors will accumulate and probably diverge, which may provide harmful information and have unfavorable effects on the battery system and control strategy.

![Figure 1. Open-loop battery model.](image)

To improve the output accuracy of the battery model, the closed-loop model, as shown in Figure 2, was developed and implemented in several applications [2–8]. Plett [9–11] first applied the Kalman filter theory to estimate the SOC by modeling the battery system to include the unknown quantities in its state description. From Figure 2, we can see that the Kalman filter algorithm is a closed-loop form...
based upon a feedback mechanism. It adjusts the SOC value dynamically according to the voltage error between the measured voltage value and the estimated voltage value from the battery model. Then the adjusted SOC value and the current feed back to the battery model, to generate a new estimated battery voltage. After much iteration, the output voltage of the model will achieve a dynamical equilibrium, approximately equal to the measured terminal voltage. An additional benefit of the Kalman filter is that it automatically provides dynamic error bounds on the estimation as well. This method provides the safety information specifics of the battery pack, reducing the chance of overcharge or overdischarge. The drawbacks of this method lay in the computational complexity and more stringent requirements for the accuracy of the battery model.

**Figure 2.** Closed-loop battery model.

A battery is a very complex electrochemical system with physical/chemical processes and some extra side reactions, such as aging and self-discharge effects. Therefore, it is difficult to build the relationships between the load voltage and varied current based on its internal chemical reaction process and electrode dynamics. In most instances, we just build a data-driven model, which means that the battery model is established according to the measured data.

The time series modeling approach is a mathematical model, which is built only based on the observed input and output sequence, without knowing the system’s detailed characteristics and internal principle. For its simplicity and validity, this modeling method has been widely used in stochastic optimal control and adaptive control applications. The time series models can be classified into two types: the stationary time series and non-stationary time series, respectively. In the stationary time series, the output sequence can be taken as the output of a zero-mean white noise sequence through the linear filter, such as Equation (1). Here the dynamic voltage of the battery \( u_d \), as shown in Figure 3, can be seen as a zero-mean sequence once the input signal has zero-mean white noise properties. For this reason, we bring in the concept of time series, taking the current and dynamic voltage \( u_d \) as two separate time series sequences. The relation of the current and dynamic voltage could be modeled as a time-series model.
In the time series family, there are many models to be chosen, such as Box-Jenkins, ARX, ARMAX, and ARMA, etc. A summary of the most used time series models is provided in Table 1.

Table 1. Summary of the time series models.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Equation Form</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box-Jenkins</td>
<td>[ y(k) = \frac{B(z^{-1})}{A(z^{-1})} u(k) + \frac{D(z^{-1})}{C(z^{-1})} e(k) ]</td>
<td>It’s a universal form and its transfer function involves with the input and error signals separately. The drawback maybe the computation complexity</td>
</tr>
<tr>
<td>ARMAX</td>
<td>[ y(k) = \frac{B(z^{-1})}{A(z^{-1})} u(k) + \frac{D(z^{-1})}{A(z^{-1})} e(k) ]</td>
<td>This is an alternation of B-J equation with the ( C(z^{-1}) ) changed by ( A(z^{-1}) ) for simplicity. It builds the connection between the output and error.</td>
</tr>
<tr>
<td>ARMA</td>
<td>[ y(k) = \frac{D(z^{-1})}{A(z^{-1})} e(k) ]</td>
<td>It is used to characterize the noise process, without the presence of the input “or exogenous” variable.</td>
</tr>
<tr>
<td>ARX</td>
<td>[ y(k) = \frac{B(z^{-1})}{A(z^{-1})} u(k) + \frac{1}{A(z^{-1})} e(k) ]</td>
<td>This model reveals that the additive noise is not white noise but colored noise that is strongly influenced by the nature of the ( A(z^{-1}) ), just a special case of ARMAX.</td>
</tr>
</tbody>
</table>

From the above comparison of time series models, we can infer that the Box-Jenkins model is one universal form, which involves the input and error signals separately, suitable for various conditions. The weakness of this model lies in the intensive computational requirements. The ARMAX model is also excessively concerned about the stochastic error signals. The ARMA model is just a noise process model, with no consideration of the input or “exogenous” variable effects. As the simplest, the ARX model cares about the additive noise influenced by the \( A(z^{-1}) \), which is the autoregressive operator of battery dynamic voltage. Therefore, as a specific case of ARMAX, the ARX model for the battery cell was adopted in this paper, as shown in Figure 3.

Based on an ARX model, the SOC estimation method using the extended Kalman filter is studied. Experiments are performed on a 60 Ah LiFePO4 battery module. The hybrid pulse power characterization (HPPC) schedule is used to identify the proposed model, as well to verify the model and the SOC estimation method using Kalman filters.

This paper is organized as follows: Section 2 gives an introduction of the experimental setup and test sequences. The ARX model and its order selection, parameter identification algorithm and results are discussed in Section 3. The Kalman filter SOC estimation method based on the ARX model is described in Section 4. Evaluation of the adaptability of the battery models and robustness of the SOC estimation algorithm are also verified. Section 5 gives a brief summary of this paper and provides some future outlook.
2. Experimental Setup

2.1. Battery Testing System

The test bench setup is shown in Figure 4. It consists of a battery testing system (MT-30), a thermal chamber for environment control, a computer for the human-machine interface and experimental data storage and the battery module or pack. The battery testing system is responsible for loading the battery module based on the customer’s program with a highest voltage of 120 V and maximum charging/discharging current of ±500 A. The host computer is used to control the MT-30 and thermal chamber, and record many quantities, such as load current, terminal voltage, accumulative Ah and Wh. The measurement inaccuracy of the current transducer inside the MT-30 system is within 0.25%. The MT-30 is connected to the battery module, which is placed inside the thermal chamber to maintain the temperature. The temperature operation range of the thermal chamber is between −50 °C and 120 °C. As for the cell voltage measurement, the Smart-Guard system with a precision of 0.25% is adopted in our test.
2.2. Battery Test Design

The lithium-ion battery cells selected for our test are from the Sinopoly Battery Ltd. company, and the key specifications of the LiFePO₄ cells are shown in Table 2. During this test, four battery cells are connected in series.

<table>
<thead>
<tr>
<th>Type</th>
<th>Nominal capacity (Ah)</th>
<th>Nominal voltage (V)</th>
<th>Upper cut-off voltage (V)</th>
<th>Lower cut-off voltage (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiFePO₄</td>
<td>60</td>
<td>3.20</td>
<td>3.65</td>
<td>2.50</td>
</tr>
</tbody>
</table>

**Figure 5.** A single HPPC test profile.

**Figure 6.** Battery terminal voltage profile in HPPC test at 50% SOC.

In order to identify the model parameters, hybrid pulse power characterization (HPPC) [12–14] tests were consecutively conducted with the lithium-ion battery module at each 10% SOC decrement across the range from 100% to 0%, while the temperature was maintained within 25 ± 0.5 °C. One HPPC test profile at SOC = 50% is shown in Figure 5. Figure 6 displays the terminal voltage profile of...
one battery cell at SOC = 50% during the HPPC test. It is assumed that the current is positive (+) when the battery charges and negative (-) when the battery discharges.

3. Battery Model

The lithium-ion battery is a dynamic system, which varies according to the different operating conditions, such as current fluctuation, variable temperatures, and altering SOC ranges. As shown in Figure 7, the equivalent circuit model [15] can be considered as a reduced order model based on the electrochemistry principle. In this model, the internal resistances include ohmic resistance $R_0$, and electrochemical polarization reaction resistance $R_1$. The polarization capacitance $C_1$ is used to describe the transient response during the charging and discharging transient state. The model output voltage comprises two parts: one is the open-circuit voltage $u_{oc}$; the other is the dynamical effect voltage $u_d$. A model with a combination of these two parts is generally used to overcome their individual shortcomings. The open-circuit voltage method is especially important for a dynamic battery model, which describes the battery cell under different conditions and is also necessary for the SOC estimation.

For its simplicity and effectiveness, this model is widely used in real applications. However, some questions about this model still remain to be considered: (1) The first is the model order selection principle. As the model's order increases, the model precision will be improved considerably at the cost of model complexity. There should be a trade-off between the model computational complexity and generalization ability; (2) another question is about the discretization and implementation. Based on the above considerations, a much more straightforward approach will be an optimal solution to these problems. The auto regressive exogenous (ARX) model can provide an appropriate description of battery cell dynamics process in terms of two polynomials. One is the auto-regression term, which is used to simulate the cell dynamic voltage, correlating with its value during the previous time interval; the other one is the moving average term, which is used to predict the cell’s future behavior based on the effect of the input signal. As one universal battery modeling method, the ARX can tune its orders more flexibly according to user’s specific weight. At this point, that represents a notable difference from the equivalent circuit model. Furthermore, we can evidently see the difference
of each model’s precision and complexity in the process of modeling. In brief, the ARX battery model is adopted here for its simplicity and efficiency, and its general diagram is shown in Figure 8.

**Figure 8.** The detail diagram of the ARX battery model.

As seen in Figure 8, the battery model consists of two main parts: one is the open-circuit voltage (OCV), denoted as $U_{oc}$, which is modeled by a look-up table; another is the dynamic effect voltage, denoted as $U_d$, which is modeled by the ARX form. The explicit content of each part will be illustrated separately in the following subsections.

### 3.1. OCV Model

The SOC is an indication of the amount of energy available in the battery compared to its maximum capacity, and it is very similar to the fuel gauge in its concept. The SOC can be mathematically expressed as an integration of the current over time as shown in Equation (2) [5]:

$$SOC(t)=SOC(t_0)+\int_{t_0}^{t} \frac{\eta \cdot i(\tau)}{C_n} d\tau \quad (2)$$

where $SOC(t_0)$ is the initial SOC, $\eta$ is the (dis)charge efficiency, $t$ is the time, and $C_n$ is the nominal battery capacity, which is defined as the amount of charge the battery can hold under normal conditions.

Open circuit voltage (OCV) [16,17] tests are performed by consecutive discharge of the battery under various SOC and temperature conditions. In this paper, the test procedure is as follows: the battery is first discharged at a constant current of 0.3C ampere from the fully charged state to 95% of the nominal capacity at 25 °C in the thermal chamber. Afterwards, it is left in the open-circuit condition for 3 hours to reach the equilibrium potential. The procedure is repeatedly performed to obtain the equilibrium potentials while the SOC changes from 100% to 0%, and afterward a similar test procedure by charging the battery at a constant current of 0.3C ampere from the fully discharged state with the step of 5% of the nominal capacity is performed. The final results of the charge and discharge process are shown in Figure 9(a).

It can be found the hysteresis phenomenon occurs: the equilibrium potential is higher during the charging process than that during the discharging process. In this paper, the hysteresis is ignored for its
minor difference, and to simplify the model the open-circuit voltage $v_{oc}$ is defined as the average of the equilibrium potentials of charging and discharging.

Similarly, the same test procedures are repeated under different temperature conditions, such as $-20, -10, 5, 15, 35$ and $45$ °C and the test results are shown in Figure 9(b). Through the contrast, the reconstructed OCV curve at $25$ °C can be taken as a balanced choice without regard to the temperature effect. Finally, the average OCV can be modeled explicitly by a regression Equation (3):

$$u_{oc}(s_k) = -0.80132 \cdot s_k^4 + 2.1575 \cdot s_k^3 - 2.0167 \cdot s_k^2 + 0.86971 \cdot s_k + 3.1561, \quad s_k \in [0,1]$$

(3)

**Figure 9.** Open-circuit voltage *versus* state of charge (SOC). (a) OCV varies with charging/discharging; (b) OCV varies with temperature.

### 3.2. ARX Model

The ARX form is especially popular for discrete systems when there are just one input and one output. An $n$th order linear time invariant ARX [18–20] model is given by:

$$y[k] + a_1y[k-1] + \ldots + a_{n_y}y[k-n_y] = b_0u[k] + b_1u[k-1] + \ldots + b_{n_u}u[k-n_u] + e[k]$$

where $y \in \mathbb{R}^2$ is the output sequence, $u \in \mathbb{R}^1$ is the input sequence, and $a_i$ and $b_j$ are the system coefficients to be identified. The popularity of this form comes from the fact that it can be used to approximate any single-input, single-output (SISO) linear and time invariant system dynamics and is very easy to identify. The latter aspect will be discussed in detail in the following section.

To be explicit, we will take the first order equivalent circuit model and its discrete form as an example. As shown in Figure 7, it is mainly composed of three parts, including open-circuit voltage $u_{oc}$ internal resistance and equivalent capacitance. As for the resistance inside, it contains the ohmic resistance $R_0$ and the polarization resistance $R_1$. The equivalent capacitance $C_1$ is used to describe the transient response during the charging and discharging process. $u_{oc}$ is the voltage across $R_0$ and RC network, which is used to describe the dynamic effects of the battery. Based on this model, the
mathematical relationship between the equivalent circuit model and ARX model will be illustrated clearly. In the equivalent circuit model as shown in Figure 7, the transfer function of $R_0$ and RC network can be shown as follows:

$$G(s) = \frac{Y(s)}{U(s)} = \frac{R_l}{1 + (R_l \cdot C_l)s} = \frac{(R_0 \cdot R_l \cdot C_l)s + (R_0 + R_l)}{(R_l \cdot C_l)s + 1}$$  \hspace{1cm} (5)$$

From transfer function (5) and using the bilinear transform \( s \rightarrow \frac{2(1-z^{-1})}{T(1+z^{-1})} \), we can get the discrete transfer function of battery system with sample time $T$:

$$G(z^{-1}) = \frac{Y(z^{-1})}{U(z^{-1})} = \frac{b_0 + b_1z^{-1}}{1 + a_1z^{-1}}$$  \hspace{1cm} (6)$$

where:

$$a_1 = \frac{T - 2R_lC_l}{T + 2R_lC_l}$$  \hspace{1cm} (6a)$$

$$b_0 = \frac{R_0T + R_lT + 2R_0R_lC_l}{T + 2R_lC_l}$$  \hspace{1cm} (6b)$$

$$b_1 = \frac{R_0T + R_lT - 2R_0R_lC_l}{T + 2R_lC_l}$$  \hspace{1cm} (6c)$$

According to Equations (6) and 6(a–c), the time-domain relationship between different samples of I/O is as follows:

$$y[k] = -a_1y[k - 1] + b_0u[k] + b_1u[k - 1]$$  \hspace{1cm} (7)$$

The above equation is the specific form for the equivalent circuit model with first order, and the ARX model, as a more compact way, can be written in the following equation:

ARX model: \( A(q)y(t) = B(q)u(t) + e(t) \)  \hspace{1cm} (8)$$

Specifically:

$$A(q) = 1 + a_1q^{-1} + \ldots + a_nq^{-n_a}$$  \hspace{1cm} (8a)$$

$$B(q) = b_0 + b_1q^{-1} + \ldots + b_nq^{-n_b}$$  \hspace{1cm} (8b)$$

where $n_a$, $n_b$ are the model orders, which need to be determined firstly by the order selection principle. The order of auto-regression term $n_a$ and the moving average term $n_b$ aims to simulate the cell nonlinear dynamics; \([a_1, \ldots, a_n, b_0, b_1, \ldots, b_n]\) are the model coefficients, which need to be identified. $e(t)$ stands for the noisy disturbance, which was caused by the current sensor sampling error and signal conversion loss and so on. The related system identification approach will be introduced in detail by the next section.

### 3.3. Parameter Identification Algorithm

Generally speaking, parameter identification includes two essential parts: one is for the model order selection; another one is for the model parameter identification. As for the model order selection
principle, numerous approaches exist in this field, such as the Minimum description length principle (MDL), the Bayesian information criterion (BIC), the Akaike information criterion (AIC), etc. Some studies [12,21–23] suggest that AIC has theoretical and practical/performance advantages over the others. Furthermore, the rate at which AIC converges to the optimum is the best in a certain sense. Therefore, the AIC is a common and most adopted method. About the model parameter identification, there also exist different schemes, such as the errors-in-variables (EIV) estimation [24–26], maximum likelihood techniques [27,28] or total least-squares [29,30] and other least-squares methods. Here the extended least-squares method [31–33] has been employed for the model parameter identification. To be more specific, two aspects will be shown subsequently: the model order selection principle in Section 3.3.1 and the parameter identification algorithm in Section 3.3.2.

3.3.1. Order Selection

The battery system is a complex object, which involves chemical reactions and thermal/heat generation. In a sense, it’s hard to choose the optimal model structure. The input-output data may provide an effective way to obtain an accepted model structure. As we know, by increasing the model order, the model accuracy will be improved significantly. However, the model with a higher order will encounter the tendency of over-parameterization and unfavorable prediction ability, while the model complexity will increase greatly.

To address this problem, a common measure is Akaike’s information criterion (AIC), which provides a trade-off between the model complexity and the goodness of fit to the experimental data. The AIC is given by the following equation:

$$AIC = \log V + \frac{2d}{N}$$  \hspace{1cm} (9)

where $V$ is the loss function, $d$ is the number of estimated parameters in the model, and $N$ is the number of values in the estimation dataset. The loss function $V$ is defined by the following equation:

$$V = \det \left( \frac{1}{N} \sum_{t=1}^{N} e(kt, \theta)(e(kt, \theta))^T \right)$$  \hspace{1cm} (10)

where, $\theta$ represents the estimated parameters. For $d \ll N$:

$$AIC = \log \left( V \left( 1 + \frac{2d}{N} \right) \right)$$  \hspace{1cm} (11)

The model with the lowest AIC should be preferred. It provides an optimal principle for the model order selection.

Based on the above mentioned order selection principle [Equation (11)], a good choice for the model structure would be ARX (2,2), as shown in Figure 10. As the number of parameters increases, the output variance decreases insignificantly.
3.3.2. Parameter Identification

After the determination of the model order, the most important next step is to identify the model parameters. Here we will give a brief introduction to the parameter identification method. Let the inputs \( u(0), u(1), \ldots, u(N) \) and the corresponding outputs \( y(0), y(1), \ldots, y(N) \) be recorded with \( N > \max(n_a, n_b) \). If we introduce the vector:

\[
\phi(k, \vartheta)^T = [-y(k-1), -y(k-2), \ldots, -y(k-n_a), u(k-1), u(k-2), \ldots, u(k-n_b)]
\]

where \( \vartheta = [a_1, \ldots, a_{n_a}, b_1, \ldots, b_{n_b}]^T \), then the model output can be rewritten as:

\[
\hat{y}(k, \vartheta) = \phi(k, \vartheta)^T \vartheta
\]

which is sometimes called a pseudo-linear regression because of the non-linear effect of \( \vartheta \) on the model output. The iterative method can be used for the parameter identification:

\[
\Phi^{(i)} = \begin{bmatrix}
y(n_a-1) & \cdots & y(0) & u(n_a-1) & \cdots & u(n_a-n_b) \\
y(n_a) & \cdots & y(1) & u(n_a) & \cdots & u(n_a-n_b+1) \\
\vdots & & \vdots & & \vdots & \\
y(N-1) & \cdots & y(N-n_a) & u(N-1) & \cdots & u(N-n_b) 
\end{bmatrix}
\]

Usually, for \( i = 0 \), an ordinary least-squares solution is used. Subsequently, this solution provides prediction errors, which are used in the next step. This procedure is repeated a number of times until the estimates converge or the maximum number of iterations is reached.

**Algorithm: Extended Least-Squares** method for ARX model parameters identification:
1. Specify an ARX model structure in terms of \( n_a \) and \( n_b \), by the order selection principle;
2. Define the vector \( y := [y(n_a), \ldots, y(N)]^T \) and the matrix \( \Phi^{(i)} \) with \( i = 0 \), as in Equation (14);
3. Calculate the unknown \( (n_a + n_b) \)-dimensional parameter vector \( \vartheta^{(0)} \), based on the least squares estimation \( \hat{\vartheta} = (\Phi^T \Phi)^{-1} \Phi^T y \);
4. Calculate the prediction errors $\varepsilon(k, \mathcal{G}^0)$, where $\mathcal{G} = [a_1, \ldots, a_n, b_1, \ldots, b_m]^T$ and $\varepsilon(k, \mathcal{G}) = y(k) - \hat{y}(k, \mathcal{G})$;

5. Given the prediction errors from an ordinary least-squares estimation of the ARX-parameters, and then back to step 3, until reaching the final terminal criterion.

3.4. ARX Modeling Results

According to the parameter identification method mentioned above and the HPPC test cycle, we can identify the parameters of a specific ARX model. The previous research work [34] has clarified that the optimal model for lithium-ion battery is the first-order RC model with one-state hysteresis. Therefore, to contrast the different performance of each ARX model, five ARX models, the ARX(1,1), ARX(1,2), ARX(2,1), ARX(2,2) and ARX(3,3) are listed here. By way of the system identification toolbox of MATLAB, final simulation results are shown in Table 3.

<table>
<thead>
<tr>
<th>model structure $(n_a, n_b)$</th>
<th>ARX model equations in detail (identified by the HPPC test data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(1,1)$</td>
<td>$(1 + a_1 q^{-1})u_d(t) = b_1 q^{-1} \cdot i(t) + e(t)$</td>
</tr>
<tr>
<td>$n_a = 1, n_b = 1$</td>
<td>with $a_1 = -0.8971$; $b_1 = 2.679e^{-4}$</td>
</tr>
<tr>
<td>$(1,2)$</td>
<td>$(1 + a_1 q^{-1})u_d(t) = (b_1 q^{-1} + b_2 q^{-2}) \cdot i(t) + e(t)$</td>
</tr>
<tr>
<td>$n_a = 1, n_b = 2$</td>
<td>with $a_1 = -0.9268$; $b_1 = 5.399e^{-4}$; $b_2 = -3.431e^{-4}$</td>
</tr>
<tr>
<td>$(2,1)$</td>
<td>$(1 + a_1 q^{-1} + a_2 q^{-2})u_d(t) = b_1 q^{-1} \cdot i(t) + e(t)$</td>
</tr>
<tr>
<td>$n_a = 2, n_b = 1$</td>
<td>with $a_1 = -0.9985$; $a_2 = 0.09182$; $b_1 = 2.4841e^{-4}$</td>
</tr>
<tr>
<td>$(2,2)$</td>
<td>$(1 + a_1 q^{-1} + a_2 q^{-2})u_d(t) = (b_1 q^{-1} + b_2 q^{-2}) \cdot i(t) + e(t)$</td>
</tr>
<tr>
<td>$n_a = 2, n_b = 2$</td>
<td>with $a_1 = -1.421$; $a_2 = 0.448$; $b_1 = 5.391e^{-4}$; $b_2 = -4.662e^{-4}$</td>
</tr>
<tr>
<td>$(3,3)$</td>
<td>$(1 + a_1 q^{-1} + a_2 q^{-2} + a_3 q^{-3})u_d(t) = (b_1 q^{-1} + b_2 q^{-2} + b_3 q^{-3}) \cdot i(t) + e(t)$</td>
</tr>
<tr>
<td>$n_a = 3, n_b = 3$</td>
<td>with $a_1 = -1.229$; $a_2 = -0.07347$; $a_3 = 0.3184$; $b_1 = 5.398e^{-4}$; $b_2 = -4.085e^{-4}$; $b_3 = -9.66e^{-5}$</td>
</tr>
</tbody>
</table>

Figure 11 presents the terminal voltage curve of the test data and the five-model simulation results. The comparison curves of the voltage error between the experimental data and the simulation data are drawn as shown in Figure 12. To facilitate observation, the partially enlarged figure of the voltage error curves at SOC = 90% is presented in Figure 13.
Figure 11. The terminal voltage profiles of the five ARX models simulation and experiment.

Figure 12. The terminal voltage error curves between the simulation and experiment.

Figure 13. The terminal voltage error curves between the simulation and experiment (enlarged at SOC = 90%).
Table 4. The statistic analysis list of the absolute values of terminal voltage errors.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Maximum (V)</th>
<th>Mean (V)</th>
<th>Variance (V²)</th>
<th>Max. Error Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARX(1,1)</td>
<td>0.0504</td>
<td>0.0038</td>
<td>5.0609e−005</td>
<td>1.4175</td>
</tr>
<tr>
<td>ARX(1,2)</td>
<td>0.0463</td>
<td>0.0020</td>
<td>1.5309e−005</td>
<td>1.3041</td>
</tr>
<tr>
<td>ARX(2,1)</td>
<td>0.0510</td>
<td>0.0027</td>
<td>2.8467e−005</td>
<td>1.4337</td>
</tr>
<tr>
<td>ARX(2,2)</td>
<td>0.0395</td>
<td>0.0013</td>
<td>6.1995e−006</td>
<td>1.1178</td>
</tr>
<tr>
<td>ARX(3,3)</td>
<td>0.0285</td>
<td>0.0007</td>
<td>2.0243e−006</td>
<td>0.8019</td>
</tr>
</tbody>
</table>

A statistical analysis on the absolute values of the terminal voltage errors is conducted, and the results are as shown in Table 4. It shows that as the model order increases, the model error will decrease significantly at the expense of computational complexity. Here both the ARX(1,2) and ARX(2,2) have better dynamic simulation performance, which reveals that these two models are preferable for the lithium-ion batteries.

3.5. ARX Model Stability Evaluation

In this paper, the ARX model is defined as the Equation (8), and the parameter identification results are listed in Table 3. Another critical problem about the ARX model remains to be confirmed: that is the model stability analysis.

Auto Regression (AR) Part:

As in Equation 8(a), we define the characteristic polynomial as:

\[ A(q) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n} \]

Thereby, the characteristic equation is defined as:

\[ z^n + a_1 z^{n-1} + a_2 z^{n-2} + \ldots + a_n = 0 \]  \hspace{1cm} (15)

To keep the ARX model output stable, the roots of the above equation should lie inside the unit circle.

Figure 14. The root distribution of the AR part of five ARX models.
Moving Average (MA) Part:

As in Equation 8(b), similarly we can find the characteristic polynomial of MA part as:

\[ B(q) = b_0 + b_1 q^{-1} + \ldots + b_{n_B} q^{-n_B} \]

Thereby, the characteristic equation of MA is defined as:

\[ b_0 z^{n_B} + b_1 z^{n_B - 1} + b_2 z^{n_B - 2} + \ldots + b_{n_B} = 0 \]

To keep the ARX model stable, the roots of the above equation should lie inside the unit circle.

**Figure 15.** The root distribution of the MA part of five ARX models.

We conduct the stability analysis of the five proposed ARX models based on the stability Equations (15,16), and the roots distribution of these models are presented in Figures 14 and 15. The graphic results show clearly that the eigenvalues of the characteristic equations of the five models are all located inside the unit circle. Therefore, all five ARX models are stable.

**Summary of Battery Cell Model:**

Simple model: note that \( OCV(s_k) \) is open-circuit-voltage:

\[
 s_{k+1} = s_k + \left( \frac{\eta_i \Delta t}{C} \right) i_k \\
 y_{k} = OCV(s_k) + ARX(i_k) = u_{oc}(s_k) + u_d(i_k)
\]

Here, if we take the ARX(2,2) structure as an example, then the discrete state space model will be as:
Define the state matrix $X$ as:

$$X = \begin{bmatrix} s(k) \\ u_d(k) \\ u_d(k-1) \end{bmatrix}$$

Then, the matrix $A$, $B$, $C$ and $D$ can be written as follows:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -a_1 & -a_2 \\ 0 & 1 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{\eta \Delta t}{C} & 0 & 0 \\ 0 & b_1 & b_2 \\ 0 & 0 & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} \frac{u_{soc}}{s(k)} & 1 & 0 \end{bmatrix}$$

$$D = 0$$

$$Y_k = [y(k)]$$

where $s(k)$ is SOC, $i(k)$ is the current, $u_d(k)$ is the dynamic effect voltage, $y(k)$ is the predicted cell voltage, $\eta_i$ is the (dis)charge efficiency (here assumed to be 1), $\Delta t$ is the time sampling interval, $C$ is the nominal capacity.

4. Kalman Filter for SOC Estimation

The Kalman filter is a mature and advanced method to filter the measurement noise to get an optimal estimation of a dynamic system’s state. If the system is nonlinear, a linearization process at each time step will be necessary to approximate the non-linear system. The extended Kalman filter (EKF) [35–40] will play a great role in these systems. In our research, the extended-Kalman-filter is adopted to correct the proposed battery model, and the general diagram can be seen in Figure 16. Based on the error between the model estimated voltage and the measured voltage, the EKF adjusts the SOC to change the model output voltage to minimize the voltage error. After a certain number of iterations, this error will converge to zero, while the SOC will converge to its optimal value. The whole operating principle of the model and EKF will be stated in detail here.
4.1. SOC Estimation Based on EKF and ARX Model

The extended Kalman filter (EKF) is a method for system state estimation in real time. In our application, to estimate the SOC during charging and discharging, the EKF can be constructed as shown in the following steps.

The state space representation (15) can be shortly expressed as:

\[
\begin{align*}
    x_{k+1} &= f(x_k, i_k) + w_k \\
    y_k &= h(x_k, i_k) + v_k \\
    w_k &\sim \left(0, Q_k = \text{diag}(Q_{uk}, Q_{uk}, Q_{uk})\right), \quad v_k \sim \left(0, R_k\right)
\end{align*}
\]

where \( x_k \) is the \([s(k) \quad u_d(k) \quad u_d(k-1)]\), \( k \) is the time index, \( h(\cdot) \) is the output equation of the battery model, \( w_k \) is a discrete time process white noise with a covariance matrix \( Q_k = \text{diag}(Q_{uk}, Q_{uk}, Q_{uk}) \), and similarly \( v_k \) is a discrete time observation white noise with covariance matrix \( R_k \). As for the \( Q_k \), the \( Q_{uk}, Q_{uk} \) are the covariance of SOC and dynamic voltage \( u_d \) respectively.

Compute the following partial derivative matrices:

\[
\begin{align*}
    \dot{A}_k &= \frac{\partial f(x_k, i_k)}{\partial x_k} \bigg|_{x_k = \hat{x}_k} \\
    \dot{C}_k &= \frac{\partial h(x_k, i_k)}{\partial x_k} \bigg|_{x_k = \hat{x}_k}
\end{align*}
\]

The initialization can be given by:

\[
\begin{align*}
    &\text{for } k = 0, \text{ set} \\
    &\hat{x}_0^p = E[x_0] = x_0 \\
    &P_0^p = E[(x_0 - \hat{x}_0^p)(x_0 - \hat{x}_0^p)^T] = P_0
\end{align*}
\]

where \( P_0^p \) is the prediction error covariance matrix.
For \(k = 1,2,\ldots\), the following steps are performed:

Step 1: perform the time update of the state estimate and estimation error covariance:

- State estimation time update:
  \[
  \hat{x}_k = f(\hat{x}_{k-1}^+, i_{k-1})
  \]
  \( (22) \)

- Error covariance time update:
  \[
  P_k^- = \hat{A}_{k-1}^+ P_{k-1}^+ \hat{A}_{k-1}^T + Q_k
  \]
  \( (23) \)

Step 2: compute the Kalman gain matrix:

- Kalman gain matrix:
  \[
  L_k = P_k^- C_k [C_k P_k^- C_k^T + R_k]^{-1}
  \]
  \( (24) \)

Step 3: measurement update:

- State estimation measurement update:
  \[
  \hat{x}_k^- = \hat{x}_k^+ + L_k [y_k - g(\hat{x}_k^+,i_k)]
  \]
  \( (25) \)

- Error covariance measurement update:
  \[
  P_k^- = (I - L_k C_k) P_k^-
  \]
  \( (26) \)

The general diagram of the EKF can be illustrated by Figure 17. After a certain number of iterations, the estimated voltage will converge to the measured value while the estimated SOC will be near to the actual value.

**Figure 17.** General diagram of the extended Kalman filter.

### 4.2. Results of EKF and ARX Model

#### 4.2.1. Evaluation on the Adaptability of the Battery Models for SOC Estimation

In this part, five different identified ARX battery models are employed for the SOC calculation. We set the initial value as:

- For the process noise, \(Q_k = diag(1e-6, 1e-3, 1e-3)\), for the measurement noise, \(R_k = [1e-1]\) (herein, \(Q_k\) is the covariance matrix of process noise, \(R_k\) is the covariance matrix of the measurement noise). The final ARX-based SOC estimation results with EKF are shown in Figure 18(a). The dynamic voltages predicted by ARX models are presented in Figure 18(b). Figure 18(c) is a zoom in of the chart of Figure 18(b), in order to indicate the detailed dynamic voltage estimation by different ARX models.
Figure 18. SOC estimation profiles based on HPPC cycles with $Q_{sh} = [1e-6], R_k = [1e-1]$.

The SOC errors between the estimated value and Ah-counting value are shown in Figure 19. A statistic analysis on the absolute SOC estimation errors is conducted and the results including the terminal SOC error are listed in Table 5.

Figure 19. SOC estimation errors based on HPPC cycles with $Q_{sh} = [1e-6], R_k = [1e-1]$.

From Figure 19, two obvious conclusions could be drawn: (1) through the error comparison, we can see that as the model accuracy increases, the SOC estimation error will be decreased notably. In our research, the maximum SOC error decreased from 3.12% to 1.39%, as the model changed from ARX(1,1) to ARX(3,3); (2) There is a tendency of SOC error variation. It continues to increase during the SOC range of 100%–40%, and then turns around to decrease during the SOC range of 30%–0%.
This could be attributed to the property of OCV-SOC curves. As we know, the open circuit voltage is flat and does not vary greatly when the SOC is from 40% to 100%. The feedback mechanism of the Kalman filter is based on the OCV-SOC curve. For this sake, to keep our estimated voltage follow the measurement voltage closely, the adjustment of the SOC range will be enlarged. However, the situation changes along the OCV-SOC curve during 0%–30%. The open circuit voltage changes dramatically versus the SOC. Therefore, to make the model output match the measurement voltage, much smaller adjustments of the SOC will be performed.

4.2.2. Evaluation on the Voltage Estimation Accuracy Influenced by the Kalman Filter

Due to the feedback mechanism, one advantage of the Kalman filter is that it can give a more precise estimation of terminal voltage. To illustrated this explicitly, a comparison is conducted on the EKF effect. The battery model of ARX(2,2), as a reference, is chosen here to compare with the ARX(2,2) + EKF. The terminal voltages of these two models are shown in Figure 20, and the voltage error comparisons between the ARX(2,2) and the ARX(2,2) + EKF are shown in Figure 21. Similarly, a statistic analysis on the absolute voltage estimation error is conducted, and some results are listed in Table 5.

Table 5. The statistic list of the absolute SOC estimation error and terminal SOC error.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Maximum</th>
<th>Mean</th>
<th>Variance</th>
<th>Terminal</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,1,1)</td>
<td>0.0312</td>
<td>0.0103</td>
<td>1.2746e−004</td>
<td>−0.0007</td>
</tr>
<tr>
<td>(1,2,1)</td>
<td>0.0261</td>
<td>0.0085</td>
<td>0.8781e−004</td>
<td>−0.0005</td>
</tr>
<tr>
<td>(2,1,1)</td>
<td>0.0288</td>
<td>0.0096</td>
<td>1.1017e−004</td>
<td>−0.0006</td>
</tr>
<tr>
<td>(2,2,1)</td>
<td>0.0176</td>
<td>0.0056</td>
<td>0.3895e−004</td>
<td>−0.0003</td>
</tr>
<tr>
<td>(3,3,1)</td>
<td>0.0139</td>
<td>0.0047</td>
<td>2.4982e−005</td>
<td>−0.0002</td>
</tr>
</tbody>
</table>

Figure 20. ARX (2,2) model voltage vs. ARX(2,2) & EKF model voltage.
Figure 21. Model voltage errors between two different models.

Table 6. The statistic analysis list of the absolute values of terminal voltage errors.

<table>
<thead>
<tr>
<th>Model Type (2,2,1)</th>
<th>Maximum (V)</th>
<th>Mean (V)</th>
<th>Variance (V²)</th>
<th>Max. Error Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>without EKF</td>
<td>0.0397</td>
<td>0.0014</td>
<td>6.1231e−006</td>
<td>1.242</td>
</tr>
<tr>
<td>with EKF</td>
<td>0.0165</td>
<td>0.0007</td>
<td>1.3492e−006</td>
<td>0.513</td>
</tr>
</tbody>
</table>

According to Figure 21 and Table 6, it can be seen that with the use of EKF, the maximum terminal voltage error decreases notably from 0.0397 V to 0.0165V, meanwhile the mean error value reduces from 0.0014 V to 0.0007 V. This could be of great value and potential for real applications.

4.2.3. Evaluation on the SOC Estimation Accuracy Influenced by Its Initial Value

Another factor affecting the SOC estimation is the initial SOC bias. In order to investigate whether the ARX model and EKF algorithm can effectively avoid the original SOC offset, a further simulation and analysis is conducted.

Figure 22. SOC curves for ARX(2,2) + EKF estimation with different initial SOC values ($Q_{st} = [1e−6]$, $R_s = [1e−1]$) (zoom in view).
Five different SOC initial values—1.00, 0.85, 0.70, 0.55, 0.30—are preset and the corresponding SOC estimations are performed based on the HPPC test data. The results are shown in Figure 22 for the first 15,000 s and in Figure 23 for the full view.

According to Figure 22, we can see that the EKF estimations with different initial SOC values converge at different rates. About 5,000 seconds afterward, the SOC will converge to the same level, near to the true value, regardless of their initial SOC values. Figure 23 is a full view of the whole convergence process. From this figure, we can find that even when the current is zero the SOC can adjust itself to the true value. The reason is that when the floating current is zero, the battery terminal voltage does not remain stationary, as a result of its dynamic effects. Therefore, the model and EKF algorithm itself will adjust the SOC to keep track of the real battery terminal voltage.

Figure 24. SOC curves for ARX(2,2) + EKF estimation with different initial SOC values ($Q_{sk} = [1e-6], R_i = [1e-1]$) (full view).
To speed up the convergence rate, we change the covariance matrix of noise from $Q_k = [1e-6]$, $R_k = [1e-1]$ to $Q_k = [1e-4]$, $R_k = [5e-1]$. By increasing the covariance matrix of process noise $Q_k$ from $1e-6$ to $1e-4$ and covariance matrix of measurement noise $R_k$ from $1e-1$ to $5e-1$, a much more rapid convergence rate will be obtained, as shown in Figure 24.

We can infer that the time for its convergence is shortened dramatically and limited within 2,000 seconds at the cost of great variations. Herein another question is posed about how to determine the covariance matrix of noise $Q_k$, $R_k$. Because of all kinds of noise, such as the sensor calibration bias, temperature effect, and environment electro-magnetic interference and so on, the noise of process and measurement is variable. A better solution to this problem may refer to the adaptive techniques [4,6,8,41] and markov chain ideas.

4.3. Discussion

To improve the convergence rate, we can change the sample time to a smaller time interval, such as 0.1 second. A smaller sampling time interval means more iterations and more feedback, which will do good to the SOC convergence. To verify the flexibility of the proposed model and algorithm, more validation sequences, such as the disturbance signal, sensor failure or temperature effects signals could be adopted. In this paper, we propose a more universal form of battery modeling and SOC estimation and this approach could also be applied to other kinds of batteries and components.

5. Conclusions

This paper proposed a new approach for battery modeling, the auto regressive exogenous (ARX) model. Its order selection principle and parameter identification algorithm are given in theoretical and practical form. Then, based on the concept of Extended Kalman Filtering (EKF), the techniques for estimating the SOC of a lithium-ion battery cell are illustrated in detail. The final results verify that the ARX model and EKF algorithm shows excellent performance and high robustness in SOC estimation for the entire range from 0% to 100%. Additionally, it can increase the model output voltage accuracy greatly. The future work will be focused on the hardware-in-loop simulation and validation for the proposed model and EKF algorithm, finally this technique should be suitable for real-time applications for electric vehicles.

Acknowledgements

This research work is supported by a grant from the National High Technology Research and Development Program of China (863 Program) (2011AA11A229); (SRFDP) Specialized Research Fund for the Doctoral Program of Higher Education (20090073120051) and CERC-CVC: U.S.-China Clean Energy Research Center Clean Vehicles Consortium (2010DFA72760-305). The Sinopoly Battery Ltd., a sponsor of the battery cells for experimental test, is also gratefully acknowledged.
References


15. He, H.; Xiong, R.; Fan, J. Evaluation of lithium-ion battery equivalent circuit models for state of charge estimation by an experimental approach. *Energies* 2011, 4, 582–598.


© 2013 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 license (http://creativecommons.org/licenses/by/3.0/).