On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression

Caihao Weng a,*, Yujia Cui b, Jing Sun a, Huei Peng c

a Department of Naval Architecture and Marine Engineering, University of Michigan, Ann Arbor, MI 48109, USA
b Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA
c Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA

Article info
Article history:
Received 24 September 2012
Received in revised form
23 January 2013
Accepted 5 February 2013
Available online 11 February 2013

Keywords:
Electric vehicles
Lithium-ion batteries
State-of-health
Incremental capacity analysis
Support vector regression

Abstract
Battery state of health (SOH) monitoring has become a crucial challenge in hybrid electric vehicles (HEVs) and all electric vehicles (EVs) research, as SOH significantly affects the overall vehicle performance and life cycle. In this paper, we focus on the identification of Li-ion battery capacity fading, as the loss of capacity and therefore the driving range is a primary concern for EV and plug-in HEV (PHEV). While most studies on battery capacity fading are based on laboratory measurement such as open circuit voltage (OCV) curve, few publications have focused on capacity loss monitoring during on-board operations. We propose a battery SOH monitoring scheme based on partially charging data. Through analysis of battery aging cycle data, a robust signature associated with battery aging is identified through incremental capacity analysis (ICA). Several algorithms to extract this signature are developed and evaluated for on-board SOH monitoring. The use of support vector regression (SVR) is shown to provide the most consistent identification results with moderate computational load. For battery cells tested, we show that the SVR model built upon the data from one single cell is able to predict the capacity fading of 7 other cells within 1% error bound.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The research and development of electric vehicles (EVs) progressed at unprecedented pace in recent years, driven primarily by their energy efficiency and environmental benefits [1]. Locally, EVs do not emit any pollutants or consume any gasoline, and in combination with electricity from renewable energy, they could achieve low emission and fuel consumption on a well-to-wheel basis [2]. However, the performance of EVs is limited due to the challenges in the development of reliable, low cost and long life cycle battery systems.

While researchers continue to develop next-generation batteries with higher energy and power density [3], there are many difficulties to be solved for the battery management. Two important functions are the state-of-charge (SOC) estimation and state-of-health (SOH) determination, and both have been studied extensively in the literature [4]. SOC is commonly defined as “the percentage of the maximum possible charge that is present inside a rechargeable battery” and SOH is “a ‘measure’ that reflects the general condition of a battery and its ability to deliver the specified performance in comparison with a fresh battery” [5]. Typically, the quantitative definition of SOH is based either on the battery capacity or the internal resistance, depending on specific applications.
Many methods for on-line SOC estimation have been studied including coulomb counting, open circuit voltage—SOC (OCV—SOC) mapping and model based approach with extended Kalman filter (EKF) [6–8]. In contrast, the development of on-line SOH monitoring technique is more challenging because of the complicated electrochemical mechanism involved in battery aging. Whereas it is possible to assess the resistance growth issue by both off-line tests such as electrochemical impedance spectroscopy (EIS) [9,10] and on-line identification algorithms such as the use of least squares methods [11–13], the detection of capacity fading still largely relies on laboratory measurements and off-line analysis [14–16].

One conventional and most common method in determining battery capacity fading is based on the OCV—SOC curve [15]. However, it requires fully charging or discharging the battery at low rate (e.g., 1/25 C) or measuring the open circuit voltage after a long relaxation period (more than 2 h) at SOC levels that span the entire range. Both methods require time-consuming tests and thus are not applicable for on-board implementation with in-situ operation data. An alternative approach to estimate capacity loss is the so-called incremental capacity analysis (ICA) [17]. By differentiating the battery charged capacity (Q) versus the terminal voltage (V), ICA transforms voltage plateaus on the charging/discharging voltage (V–Q) curve, which are associated to the staging of graphite anode, into clearly identifiable dQ/dV peaks on the incremental capacity (IC) curve [14,17]. The concept of ICA originates from the study of the lithium intercalation process and the corresponding staging phenomenon [14,18–20]. ICA has the advantage to detect a gradual change in cell behavior during a life-cycle test, with greater sensitivity than those based on conventional charge/discharge curves and yield key information on the cell behavior associated with its electrochemical properties [16,21]. Although ICA was shown to be an effective tool for analyzing battery capacity fading, most studies have focused on understanding the electrochemical aging mechanism and no study has been reported based on the on-board application of ICA. Meanwhile, since all the peaks on an IC curve lie within the voltage plateau region of the V–Q curve, which is relatively flat and more sensitive to measurement noise, calculating dQ/dV directly from the data set is difficult. Hence, effective and robust algorithms of obtaining the IC curve need to be developed.

This paper aims at developing algorithms that can apply ICA for battery SOH monitoring with partially charging data. Several numerical procedures were developed and evaluated for extracting the IC peaks and associating them with capacity fading. In particular, our support vector regression (SVR) approach was found to be:

(a) Insensitive to measurement noises
(b) Robust to data range and size
(c) Effective in extracting a signature that shows strong dependence to battery age

Our study is based on a battery life cycle test data set collected over a period of 18 months. We showed that the SVR model built upon the data from one single cell is able to predict the capacity fading of 7 other cells with less than 1% error. Even though the aging data are collected for A123 LiFePO4 cells and the quantitative relation identified in this work have not been demonstrated to work on other types of cells, we believe that the procedure and algorithms reported in this paper are general and should be applicable to many of Li-ion battery systems.

The remainder of this paper is organized as follows: Section 2 introduces experiment setup and results obtained by applying ICA off-line. In Section 3, ICA results based on partial data with conventional methods are presented. Section 4 introduces the linear programming based SVR (LP-SVR) and elucidates the implementation of LP-SVR in obtaining the IC curve. The conclusions are given in Section 5.

2. Experiment setup and off-line ICA results based on OCV

2.1. Battery testing systems and schedule

The test data used for this study are acquired through the battery test bench set up in Ref. [22], which includes an Arbin BT2000 tester, a thermal chamber for environment control, a computer for user-machine interface and data storage, a switch board for cable connection, and battery cells. The data acquisition system has a sampling frequency of 10 Hz.

The batteries used for this test are lithium ion phosphate (LiFePO4) APR18650M1A cells from A123 with the specifications listed in Table 1. Eight cells are tested with the same load profile for comparison and validation. As shown in Fig. 1, the experiment procedure starts with a series of characterization tests (which consist of a static capacity test using current rate 1/2 C, a hybrid pulse test, a DC resistance test, a dynamic stress test (DST) and a Federal Urban Driving Schedule (FUDS) test) conducted at three different temperatures (in the order of 10 °C, 35 °C and 22 °C). After these tests, 100 aging cycles are conducted at 22 °C. In each aging cycle, the cells are charged and discharged at a constant rate until the cutoff voltage is detected [22]. For this study, the data sets from the 1/2 C rate static capacity test are mainly used.

More detailed discussion of the battery testing systems and schedules can be found in Ref. [22].

2.2. Off-line OCV identification and ICA results

As indicated in Ref. [17], the ideal ICA results should be obtained with a close-to-equilibrium OCV curve. Hence, we first start the ICA study by identifying OCV curves from the experiment data set. Note

![Fig. 1. Battery test schedule from Ref. [22].](image-url)
that there is no OCV data collected at extremely low charging rate (i.e., 1/25 C); we therefore perform identification on the static capacity test data (which has charging/discharging rate at ½ C) using a parametric model developed in Refs. [23–25].

The model of the OCV curve is expressed as

\[
OCV(z) = K_0 - \frac{K_1}{z} - K_2 z + K_3 \ln(z) + K_4 \ln(1-z) \tag{1}
\]

where \(K_0\)–\(K_4\) are the model parameters and \(z\) is the normalized SOC, and the discrete battery model used for charging/discharging process is formulated as

\[
z_{k+1} = z_k - \left( \frac{\eta \Delta t}{C} \right) i_k, \quad y_k = OCV(z_k) - R_i k
\]

\[
\begin{align*}
y_k &= OCV(z_k) - R_i k \\
&= K_0 - \frac{K_1}{z_k} - K_2 z_k + K_3 \ln(z_k) + K_4 \ln(1-z_k) - R_i k
\end{align*} \tag{2}
\]

where \(\eta\) is the charging/discharging efficiency, \(\Delta t\) is the time step, \(C\) is the battery capacity, \(i_k\) is the input current, \(y_k\) is the model output (terminal voltage) and \(R_i\) is the parameter that represents battery internal resistance [22].

Because the battery model is linear in parameters, we could thus formulate the estimation problem as the following

\[
y_k = \theta_V^T \phi_{y_k}, \quad \theta_V = [K_0, K_1, K_2, K_3, K_4, R_i]^T, \quad \phi_{y_k} = \left[ 1, -\frac{1}{z_k} - z_k, \ln(z_k), \ln(1 - z_k), i_k \right]^T.
\] \tag{3}

The parameters \(\theta_V\) can be solved by the standard least squares method.
\[
\theta_Y = (\Phi_Y^T \Phi_Y)^{-1} \Phi_Y^T Y,
\]
where \(Y = [y_1, y_2, \ldots, y_n]^T\),
\(\Phi_Y = [\phi_{y_1}, \phi_{y_2}, \ldots, \phi_{y_n}]^T\).

An example of the OCV identification results and the corresponding IC curve is shown in Fig. 2. Note that the OCV is plotted against the battery charged capacity \(Q\), instead of the normalized SOC. With the parametric OCV model identified, the IC curve is obtained by taking analytic derivative of the mathematic expression (1). One single peak can be observed from the IC curve. It should be noted, according to Ref. [17], there are three identifiable peaks on the IC curve. However, the number of peaks that can be extracted numerically from the experimental data depends on the OCV model and the numerical approach. Using the parametric model (1) and analytic derivative calculation, on one peak is identified. Nevertheless, the purpose of using this parametric OCV model is to reveal and validate that the aging signature can indeed be extracted from battery data through ICA. In later sections, however, we will discuss how ICA can be applied directly to battery charging data without the parametric OCV model, and show results where all three peaks are identified.

Applying the identification procedure to different data sets collected at different aging cycles, the results shown in Fig. 3 represent the change in OCV and IC curve for the first 2300 cycles of one battery under testing. Compared with the other cells, the battery #7 in Fig. 4 shows the most consistent aging behavior and is therefore used as the reference battery cell in our study. Unless otherwise notified, all single cell results in this paper are based on the battery #7. The numerical values in the plot legends represent the aging cycle number. Monotonic trends in the peaks, as the battery ages, can be clearly identified on both OCV and IC curves. However, notice that the IC curve provides greater sensitivity than the OCV curve.

The same trend can be observed from all eight cells at all temperatures. Fig. 4 shows the normalized IC peak values for all battery cells at the displayed aging cycles. With the exception of 4 outliers out of 48 data sets, the eight cells give consistent decreasing IC peaks, which confirms the validity of using IC peaks as the signature for battery capacity fading.
At last, the battery capacity is plotted versus the IC peak in Fig. 5 and a linear correlation can be observed between the two values. Up to now, the monotonic characteristic of IC peaks as batteries degrade has been verified through the off-line identification, and a correlation between capacity and IC peak is established. Therefore ICA is shown to be a useful tool for battery SOH assessment. For on-line SOH monitoring, however, OCV curves change as batteries age, and the updated OCV curves are not available for conducting ICA. Moreover, most real-life charging data does not span the entire SOC range, thus the off-line identification method discussed in this section cannot be applied. On the other hand, since the peak of IC curve appears around the nominal voltage of 3.3 V, we believe that this signature can be extracted from normal EV charging data (which is limited in its SOC range) with appropriate algorithms. Toward this end, we propose and analyze several SOH monitoring frameworks in the following sections.

3. ICA results using conventional data processing and curve fitting methods

3.1. Numerical derivative with smoothing

As mentioned before, the updated OCV curve is not available during real-life operation, the ICA has to therefore rely on the vehicle charging data, which gives the charging voltage (V–Q) curve.

Applying numerical derivative directly to the data is the most intuitive approach, given the definition of IC. Because the measurement noise has a rather big influence on the flat portion of the charging voltage curve, judicious data processing is required before numerical derivative can be performed. In our study, the raw data curve is fitted with a 3rd order polynomial curve piecewisely with a moving window. The derivative of the middle point of each window is recorded. The resulting derivative curve is then smoothed by averaging. The results are shown in Fig. 6. There are three noticeable peaks in the plot, which agrees with the results shown in Ref. [17]. The peaks are associated with the staging process in the negative electrode as discussed in Refs. [14,18,21]. The peak at higher voltage gives a clearly decreasing trend as battery ages (shown in Fig. 7) and is thereby selected for further study.

Although the numerical derivative results do yield a clear monotonic change on the IC peaks, the results are very noisy and not suitable for further quantitative analysis. In addition, the data processing procedure is time-consuming and thus not computationally efficient. Nonetheless, the numerical results identify the range where the relevant peaks appear and therefore a local polynomial curve can be fitted. The polynomial fitting results are discussed in the next subsection.

3.2. Polynomial curve fitting

The charging data around the main voltage plateaus (e.g., ranging from 0.45 Ah to 0.85 Ah for the battery of nominal capacity...
1.1 Ah) is selected for the analysis. Because the selected data only contains two IC peaks at the higher voltage range as shown in the previous subsection and Fig. 6, a 5th order polynomial (which can represent exactly two peaks) is chosen to be fitted with the charging data and then differentiated for ICA. Two ICA peaks can be observed on the IC curve. Same as before, the value of the peak at higher voltage is recorded for analysis. An illustration of this method is shown in Fig. 8 and a comparison of the curves involving aging effect is plotted in Fig. 9.

Polynomial curve fitting gives much smoother IC curves compared to the ones by numerical derivative and efficient algorithms (such as the least squares method) are readily available for use. However, our analysis reveals an unacceptably high sensitivity of the results to the selected data range. As shown in Fig. 10, the shape of the IC curve may change significantly when slightly different data range is used for the fitting. A more robust method for ICA is thus needed.

The correlation between faded battery capacity and IC peaks obtained from the above two methods are displayed in Fig. 11. Both methods lead to 2nd order correlations between the faded capacity value and the corresponding IC peak value. Although quantitatively the two methods give different results, the qualitative relationships are the same.

4. ICA results using SVR

4.1. ICA results obtained by LP-SVR

The support vector algorithm is a nonlinear generalization of the Generalized Portrait algorithm developed by Vapnik et al. in the sixties [26]. SVR adopts the original machine learning algorithm and applies it for non-parametric function estimation. Conventional SVR is formulated as a convex quadratic programming (QP) problem and has been successfully applied in identifying nonlinear dynamic systems [27]. However, the implementation of QP-SVR is computationally expensive and sufficient model sparsity cannot be guaranteed. LP-SVR that employs $\ell_1$ norm as regularizer was then proposed to improve the model sparsity and computational efficiency [28,29].

SVR is chosen for this study because of its excellent approximation and generalization capability, and its demonstrated potential in the realm of nonlinear system identification [27,28,30].

Since we are only interested in obtaining the IC ($dQ/dV$) curve, we decide to use SVR to fit the reverse of charging curve ($Q-v$). The kernel based $Q-V$ model is described by

$$ f(x_n) = \sum_{i=1}^{N} \beta_i k(x_i, x_n), $$

where $x_n$ and $y_n$ are model input and output, $x_i$ is the base data point used for the kernel function, $q_n$ is the battery charged
capacity, \(v_n\) and \(v_l\) represent the measured voltage, \(N\) is the number of data points in the data set, \(\beta_i\) is the model parameters and \(k(\cdot, \cdot)\) is the selected kernel. In this study, the Gaussian radial basis function (rbf) kernel is used and is expressed as

\[
k(x, x') = \exp\left(-\frac{|x - x'|^2}{2\sigma^2}\right)
\]

where \(\sigma\) is the adjustable parameter for the kernel function.

The parameters \(\beta_i\)s in SVR are computed by an optimization algorithm.

In this work, we used LP as the optimization engine to derive the SVR model. The LP-SVR formulates the optimization problem as follows,

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \| \beta \|_1 + w \sum_{n=1}^{N} \xi_n, \\
\text{subject to} & \quad y_n - \sum_{i=1}^{N} \beta_i k(x_i, x_n) \leq \varepsilon + \xi_n, \\
& \quad \sum_{i=1}^{N} \beta_i k(x_i, x_n) - y_n \leq \varepsilon + \xi_n, \\
& \quad \xi_n \geq 0
\end{align*}
\]

where \(\xi_n\)s are the slack variables, \(w\) is the weighting factor, \(\varepsilon\) is the precision parameter, \(\| \cdot \|_1\) denotes the \(\ell_1\) norm in coefficient space and \(\beta\) is defined as

\[
\beta = [\beta_1 \ \beta_2 \ \ldots \ \beta_N].
\]

The optimal result usually gives zero value for most of the \(\beta_i\)s and the \(x_i\)s corresponding to non-zero \(\beta_i\)s are called support vectors (SVs).

An example of the IC curve obtained through LP-SVR is shown in Fig. 12. Note that the SVR algorithm gives a robust and smooth result even though a shorter range of data (0.6 Ah–0.85 Ah), which only contains the one peak that we are interested in, is used here. The algorithm performed well in terms of both model sparsity and data approximation. The insensitivity to data range is further demonstrated in Figs. 13 and 14, where the same results are obtained when different data range is used (Fig. 13), and when only a sub-segment of data is used (Fig. 14). The IC peak values with respect to different data range are compared in Table 2. One can see that the maximum difference between the tabulated IC peak values is about 1%.

### 4.2. ICA results obtained by SV based parameter identification (SVPI)

To further improve the computational efficiency, and make it feasible for on-board implementation, we investigated the performance of the IC peak identification algorithm when the SVs are fixed from cycle to cycle (i.e., only apply LP-SVR to the initial data set to find SVs and favorable kernel functions, and then use the SVs and kernels as a parametric model for other aged battery data sets).

The LP-SVR problem is then reduced to an SVPI problem,

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \| \beta \|_1 + w \sum_{n=1}^{N} \xi_n, \\
\text{subject to} & \quad y_n - \sum_{i=1}^{N} \beta_i k(s v_i, x_n) \leq \varepsilon + \xi_n, \\
& \quad \sum_{i=1}^{N} \beta_i k(s v_i, x_n) - y_n \leq \varepsilon + \xi_n, \\
& \quad \xi_n \geq 0
\end{align*}
\]

where \(\beta_i\)s and \(s v_i\)s are called SVs and favorable kernel functions, and then use the SVs and kernels as a parametric model for other aged battery data sets.)

\[
f(x_n) = \sum_{i=1}^{N} \beta_i k(s v_i, x_n) = \theta_{SV}^T \phi_{SV},
\]

\[
\theta_{SV} = [\beta_1, \beta_2, \ldots, \beta_N]^T, \quad \phi_{SV} = [k(s v_1, x_n), k(s v_2, x_n), \ldots, k(s v_N, x_n)]^T
\]
where \( s_{svs} \) are the SVs identified previously by LP-SVR formulated in (7), \( N_{sv} \) is the total number of SVs and the parameters \( \beta \) can be solved by parameter identification method such as least squares. Typically, \( N_{sv} \ll N \). In this study, we have \( N_{sv} = 9 \) and \( N = 200 \).

Comparisons between the IC curves and corresponding peak values obtained by LP-SVR and by SVPI are displayed in Fig. 15. One can see that both methods give qualitatively the same results but the IC peak values from the two methods start to deviate from each other after 1900 cycles, which implies that the SVPI with a fixed set of SVs works well over a wide range of data and it is only necessary to update the SVs after a long period operation. Moreover, as plotted in Figs. 15 and 16, if the SVs are re-derived with the 1900 cycle data, we can obtain qualitatively (Fig. 15) and quantitatively (Fig. 16) the same results by both methods at all aging cycles.

We can establish a 2nd order correlation between battery capacity and IC peak through the SVR results as well. The correlation curves identified by LP-SVR and SVPI are both plotted in Fig. 17. By updating the SVs at 1900 cycles, the two identified correlations are almost identical (the two curves overlap each other as shown in Fig. 17).

The correlation shown in Fig. 16, which is obtained for battery #7, will be used as the capacity fading prediction model. To validate the model, SVPI is also performed (with updated SVs) on the data sets of all other 7 battery cells at the chosen aging cycles. The signatures of decreasing IC peaks can be extracted clearly from all cells with partially charging data (about 60%–85% SOC range). The normalized IC peaks and the capacities of all batteries are plotted in Fig. 18 with the correlation curve identified using data of battery #7. The maximum difference in the capacity is about 1% except the 5 outliers. Therefore, by normalizing both the capacities and IC peak values, the identified quantitative correlation can be used for effective on-board capacity estimation and SOH monitoring.

In terms of computational efficiency, because the major burden comes from the process of solving the LP problem, SVPI using least squares method for identification saves a significant amount of time compared to LP-SVR and is very promising for on-line applications. In the case of battery #7, the total processing times of LP-SVR and SVPI are 187 s and 64 s respectively.\(^1\) SVPI reduces 66% computational time compared to LP-SVR.

\(^1\) Recorded on a computer with a 2.53 GHz Intel Core 2 Duo CPU and 4.0 GB RAM.

Fig. 19. ICA based on-board SOH monitoring framework.
4.3. Development of SOH monitoring framework using ICA results

With the established quantitative correlation between the battery capacities and the IC peak values, an on-board SOH (mainly on battery capacity fading) monitoring framework is developed (Fig. 19). Compared to conventional method which monitors battery capacity loss based on fully charging data, our ICA based SOH monitoring framework utilizes partially charging data (about 60%–85% SOC range), that are frequently available during real-life operations, to estimate the faded battery capacity on-board. Fully charging data is only needed for calibration after a long operation period. By using SVPI, the advantages of our framework in robustness and computational efficiency have both been demonstrated.

5. Conclusions

In this paper, we consider the battery SOH monitoring problem with a specific emphasis on using partially charging data for on-board implementation. ICA is used to correlate capacity fading with the IC curve peaks, thereby rendering the latter as the robust signature for SOH monitoring. Several algorithms are developed to extract this signature from normal vehicle charging data. Using SVR, an SOH monitoring framework is developed to provide a definite and quantitative correlation between IC peaks and faded battery capacity. The capacity loss can thereby be estimated/predicted through normal charging data during real-life operation. For data collected on 8 LiFePO4 cells, the model developed with the SVR approach using one cell data is able to predict the capacity fading of other cells with less than 1% absolute error except a few outliers. As the IC peak degradation shown in the charging data may be more related to the loss of active material at the anode, further studies based on discharging data (during driving) are needed to verify the applicability as well as the robustness of the ICA methodology to other aging mechanisms. Future research will also investigate the sensitivity of the method to different charging patterns and operating conditions (such as temperatures), and develop adaptive control strategies that can incorporate the SVR SOH monitoring for battery energy management.

Acknowledgments

This material is based upon work supported by the Department of Energy under Award Number DE-P10000012.

The authors appreciate the discussion with and inputs from Dr. Zhao Lu from Tuskegee University on LP-SVR.

References