Capacity Estimation of Serial Lithium-ion Battery Pack Using Dynamic Time Warping Algorithm

YANG LIU[1], **CAIPING ZHANG**[1], **(Senior Member, IEEE),**
JIUCHUN JIANG[1,2], **(Senior Member, IEEE), YAN JIANG**[1],
LINJING ZHANG[1], **AND WEIGE ZHANG**[1]

[1]National Active Distribution Network Technology Research Center, Beijing Jiaotong University, Beijing 100044, China
[2]Shenzhen Precise Testing Technology Company, Ltd., Shenzhen 518108, China

Corresponding author: Caiping Zhang (zhangcaiping@bjtu.edu.cn)

ABSTRACT The existence of the consistency degradation of the battery pack hinders the accurate estimation of pack capacity and cell capacity in the battery pack. The paper focuses on the capacity estimation of cells in the serial battery pack. The shape invariance of the charging voltage curve is discussed and used as the theoretical foundation of cell capacity difference identification. The matching relationship between two voltage curves is obtained based on the dynamic time warping algorithm. Then the capacity difference identification algorithm to calculate the capacity difference between the two cells is proposed. Based on the algorithm, a three-step capacity estimation method is established. The proposed method can only use the previous charging curve of one cell in the pack and the current charging data of the battery pack to rapidly estimate the capacity of each cell in the battery pack. A 16 serial LiFePO$_4$ battery pack is employed to verify the method. The result shows the estimation error of cell capacities is less than 3% rated capacity. With this method, the cell capacities in the pack can be rapidly and accurately estimated, providing a foundation for the consistency analysis and equalization of the battery pack.

INDEX TERMS Charging voltage curve, dynamic time warping algorithm, capacity estimation, lithium-ion battery pack.

I. INTRODUCTION

Lithium-ion batteries with high power performance and high energy density, are widely used in fields such as electric vehicles and energy storage [1]–[4]. For security consideration, the capacity of the commercial lithium-ion cell is limited. So, the mode of the series-parallel group is adopted to meet various battery system requirements of energy and power levels. For example, in all the electric vehicles, batteries are used as a group with cells in parallel and series [5], [6]. The initial performance of the battery is affected by the battery manufacturing process, and it was easy to diverge in this process, which makes the internal structure or property of the battery unable to be entirely consistent, leading to inconsistent performance parameters such as different battery capacity and internal resistance in the same batch. The inconsistency

The associate editor coordinating the review of this manuscript and approving it for publication was Gaetano Zizzo.

of parameters such as cell capacity, internal resistance, and state of charge (SOC) will be increased due to differences in ambient temperature, heat dissipation conditions and self-discharge degree of cells in the battery pack [7]–[9]. The inconsistency among batteries will affect the available capacity and available energy of the battery pack [10]. Furthermore, the working SOC range of battery will differ among them due to the battery parameters inconsistency, and then will result in a discrepancy of battery degradation rate during usage. The capacity accuracy estimation of each cell in the pack is hence a considerable challenge.

The initial pack capacity can be measured through the capacity test, but the pack capacity will change as cells aging. The capacity of the aged pack is determined by the battery pack consistency, including the consistency of capacity, internal resistance and SOC [11]. But the capacity degradation affects the pack capacity directly. Researches on single-cell capacity estimation have been the most concerned topics for

a long time. There is plenty of literature proposing many kinds of methods for single-cell capacity prediction. The reference [12] has summarized some estimation methods for the state of health (SOH). Typically, a system model of the battery for capacity, open-circuit voltage (OCV) and the SOC is established, then the extended Kalman filter (EKF) is used to simultaneously estimate SOC and capacity [2], [13]. Some other researchers build the cycling model of the battery about capacity degradation by testing lots of batteries at multi-temperature conditions and different charging rates [2], [15]. But the fade rates of cell capacities differ among different cells even under the same test conditions, so it will be very inaccurate to estimate pack capacity by using the same methods for estimating the cell capacity [16]. And the further estimation for each cell capacity in the pack is harder. Reference [16] proposed a transformation method on the cell charging voltage curve and used the genetic algorithm (GA) to estimate the cell capacity in the pack. And reference [17] proposed a geometrical approach method to estimate cell capacity. So far, lots of achievements from these researchers have already contributed to the accuracy of capacity estimation. However, the rapidity and the simplicity of the estimation algorithm are still an issue waiting for resolution.

In this paper, a rapid and accurate method of cell capacity estimation in the pack is proposed. By using the data of the battery pack in one charging process, the capacity of each cell in the pack can be estimated rapidly. In Section II, the shape invariance of the battery voltage curve and the mechanism of dynamic time warping (DTW) algorithm are discussed. In Section III, a three-step method for estimating all the cells' capacity in the pack is introduced. And the method is validated by some experimental data in Section IV. Section V summarizes the findings and primary work of this paper.

II. PRINCIPLES
A. SHAPE INVARIANCE OF BATTERY VOLTAGE CURVE
The loss of lithium inventory (LLI) and the loss of active material (LAM) are considered to be the primary reason causing battery capacity reduction, which is a significant phenomenon during the degradation of Lithium-ion batteries. LAM can be categorized into four types, depending on the type of electrode and the degree of lithiation: loss of lithiated active material of positive electrode (LAM_{liPE}), loss of delithiated active material of positive electrode (LAM_{dePE}), loss of delithiated active material of negative electrode (LAM_{deNE}), and loss of lithiated active material of negative electrode (LAM_{liPE}). Reference [18] has systematically discussed LLI, four types of LAM, and degradation behavior of the half-cell electrode voltage curve by a simulation battery model of lithium iron phosphate (LFP). The shape change of the half-cell electrode voltage curve during LLI or LAM can be summarized as the curve shifts or shrinks in a horizontal direction (SOC axis) in brief.

The voltage curve of the full cell is composed of voltage curves of two half-cell electrodes, the positive and the negative. For the LFP battery, the material of its positive electrode

is LFP. It is demonstrated that LFP has only one phase change process during charging and discharging. So, the half-cell voltage curve of the LFP positive electrode shows one voltage plateau. The negative material is usually graphite, which has five-phase change processes when lithium-ions intercalation or deintercalation. The LLI and LAM affect half-cell voltage curves and further affect the full cell voltage curve. The behavior of LLI and LAM acting on voltage curves can be summarized as follows [18], [19]:

  a. When LAM_{liPE} or LAM_{dePE} occur, the half-cell voltage curve of the positive electrode will shrink towards low SOC direction or high SOC direction respectively, but the negative curve will not change. If the loss of positive material is severe, the part of the negative curve in use will be narrowed, so the shape of the full cell voltage curve will be extremely different from the previous curve.

  b. LAM_{deNE} can make a half-cell voltage curve of the negative electrode shrink towards low SOC direction. The shape of the full cell voltage curve will change, only if the load ratio (LR) between the negative and positive electrode is less than about 1.1.

  c. LAM_{liNE} has almost no effect on the shape of the full cell voltage curve. The half-cell voltage curve of the negative electrode shrinks towards the high SOC direction, but all the voltage platforms on the negative electrode will still be used.

  d. LLI will shift the negative voltage curve towards high SOC direction. Like LAM_{liNE}, the last voltage platform on the negative electrode is long enough, and it can still be used after LAM_{liNE} or LLI. So, the LLI is considered to have little effect on the shape of the full cell voltage curve only if the LLI is too large.

It is worth mentioning that the LFP positive material has excellent chemical stability. So, the capacity loss of LFP battery primarily comes from LLI and loss of negative active material, and LR between the negative and positive electrode keeps larger than 1.1 over the entire life of LFP battery [19]. Hence no condition can change the shape of the full-cell voltage curve. It is illustrated in Figure 1 that the voltage curve of the battery shrinks right along the time axis as the number of cycles increases.

As a battery pack, the cells in it need to be filtered before they are grouped. Cells in a pack are considered to have a similar voltage curve. For the battery pack in series, the same operation current and the uniform temperature distribution in the pack may lead to a similar aging path of cells. So, the shape invariance of the battery voltage curve can be applied to the cells in the battery pack, and the shape invariance means that the full-cell voltage curve only stretches or shrinks but has no shifting or angle changing.

B. DYNAMIC TIME WARPING ALGORITHM
Based on the shape invariance of the battery voltage curve, a method for estimating the capacity difference between two cells by matching similar parts of the curves is proposed.
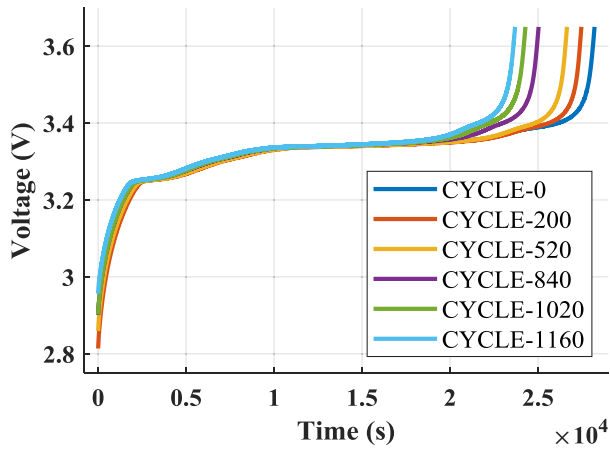
FIGURE 1. The voltage curves of a battery under different cycle times.

FIGURE 2. An example of the warping path map.

In the paper, the DTW algorithm is applied to find the matching segment of the voltage curves. The reason for DTW algorithm selection is that the algorithm can expediently export the "warping path", by using which can partition the voltage curve into matched segments and mismatched segments.

In general, the DTW algorithm is applied with some restrictions [20]:

a. Every index from the first sequence must be matched with one or more indices from the other sequence and vice versa.

b. The first index from the first sequence must be matched with the first index from the other sequence, but it does not have to be its only match. And the last index from the first sequence must be matched with the last index from the other sequence.

c. The mapping of the indices from the first sequence to indices from the other sequence must be monotonically increasing, and vice versa.

The general steps of the DTW algorithm are:

a. Assuming that there are two sequences X and Y.

$$X = (x_1, x_2, \ldots, x_n) \tag{1}$$
$$Y = (y_1, y_2, \ldots, y_m) \tag{2}$$

b. Find a warping path W:

$$W = w_1, w_2, \ldots, w_k, \ldots, w_K$$
$$\max(n, m) \le K \le m + n - 1 \tag{3}$$

where $w_k = (i, j)$ represents the distance between point i in X and point j in Y. The Euclidean distance is usually adopted. For the situation that the sequences in this paper are one-dimension time-series, the Euclidean distance is the absolute value of the difference of two-point.

And it needs to obey these rules as described above, which can be formulized as:

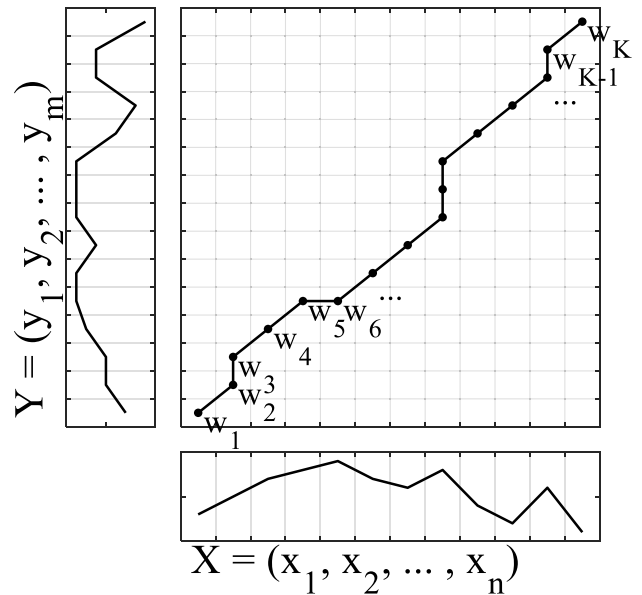$$w_1 = (1, 1), w_K = (n, m) \tag{4}$$
$$w_k = (a, b), w_{k-1} = (a', b')$$

$$0 \le a - a' \le 1, 0 \le b - b' \le 1 \tag{5}$$

c. Minimize the warping path, and that is the DTW distance:

$$DTW(X, Y) = \min \frac{\sqrt{\sum_{k=1}^{K} w_k}}{K} \tag{6}$$

Figure 2 is an example of a DTW path, which represents the matching relationship of every point in two sequences. In Figure 2, the point w_1 locates at $(1, 1)$ that means x_1 matching y_1, so the coordinates of the point w_2 and w_3 mean x_2 matching both y_2 and y_3, which can be interpreted as that the sampling point x_2 and its neighborhood in the real continuous signal X can match the segment from y_2 to y_3 in the real signal Y, and which is where the "warping" comes from.

It is illustrated in Figure 2, and there are three types of moving orientations of the matching points under the algorithm constraints:

a. Along the sequence X (the point in sequence Y unchanged). Such as the path from w_5 to w_6 in Figure 2.

b. Along the sequence Y (the point in sequence X unchanged). Such as the path from w_2 to w_3 in Figure 2.

c. Along the diagonal line (the points in sequence X and Y both changed). Such as the path from w_1 to w_2 in Figure 2.

In some general DTW algorithms, a specific value will be chosen as the common length of sequences when the lengths of two sequences are different. If the lengths are inequality, some preprocessing will apply in the initial sequences before the DTW process. For example, when n in (1) is larger than m in (2), the value n can be selected as the common length, then the sequence (2) will be stretched to the common length n.
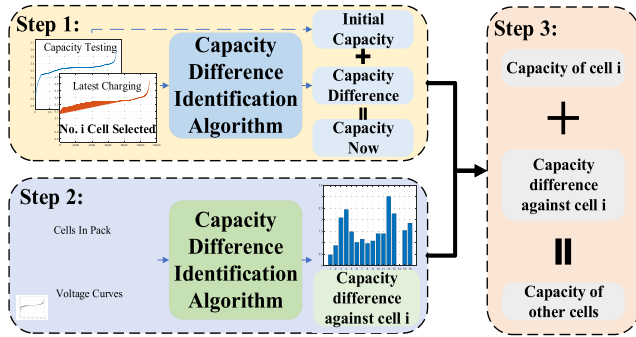
FIGURE 3. The process of cell capacity estimation in a battery pack.

FIGURE 4. A division example of the voltage curve according to scale coefficient.

So, the warping path map will be a square matrix, and the warping path of two same sequences is a diagonal line.

III. IMPLEMENTATIONS

Based on the shape invariance assumption of the voltage curve, a method of identifying the difference in capacities between two cells is proposed. Then the capacities of cells in a battery pack are further estimated by applying this method. The whole process can be divided into three steps (shown in Figure 3):

Step 1: Estimate the capacity of a cell (Number i in the battery pack) by using the capacity difference identification method to analyze voltage sequence of the cell in latest charging process and voltage sequence of the same cell in the capacity testing before grouped as a pack (the initial charging voltage curve of the cell is obtained in this capacity testing).

Step 2: Identify the capacity differences of cells in the pack also by using the capacity difference identification method.

Step 3: Calculate the capacities of other cells in the pack by using the capacity value of the cell i estimated in **Step 1** and the capacity differences between cell i and other cells in **Step** 2. Then compute the capacity of the whole battery pack according to the cell' capacities and the information of capacity offset from **Step 2**.

Sixteen LFP cells were picked to form a battery pack, and the rated capacity of these cells is about 60 ampere-hour (Ah). The capacity testing data of a cell (assuming it is the No. i cell in the pack) before grouped was stored. The capacity testing data contains the initial capacity of the cell, the charging voltage sequence at a certain current, and the charged-in capacity sequence during the whole process. Then these cells were grouped as a serial pack in the laboratory after they were numbered. The cycling test is performed on the battery pack, and the performance test is executed during the cycling. In this test schedule, the performance of cells is tested after the 200th, 520th, 840th, 1020th, 1160th cycle. The pack will be disassembled before the performance test. After the performance test, each cell will be reassembled, and the cell
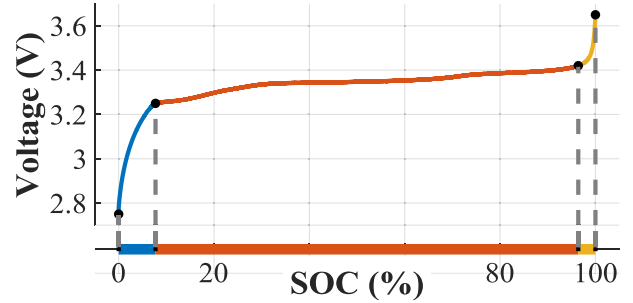
SOC is recovered to the pre-disassembly state. The present capacity of each cell will be recorded in the performance test.

The capacity of each cell tested in the performance test after the 1160th cycle is used as the present capacity of each cell. The cycle data before the performance test is used to identify the capacity difference between No. i cell and other cells. And the cycle data of cell i before the performance test at the 1160th cycle is used to estimate the present capacity of cell i.

The method and the steps will be introduced in more detail below.

A. THE METHOD OF IDENTIFYING THE CAPACITY DIFFERENCE

1) DIVISION OF THE VOLTAGE CURVE ACCORDING TO SHAPE SCALABILITY

The charging voltage curve of a battery can be separated into three parts: two parts located at low SOC and high SOC position with high polarization voltage (the blue and yellow line in Figure 4); one part located at the voltage platform representing the phase transition of chemical reaction (the red line in Figure 4). The upper curve in Figure 4 is the voltage curve, and the lower curve in Figure 4 is the projection line of the voltage curve on the SOC-axis. The three parts of the voltage curve are considered to have different scale coefficients when comparing with other voltage curves. For the LFP battery, the primary usable capacity mainly locates at the voltage platform. The capacity of the part at high SOC and low SOC region is so small and can be ignored sometimes. So, the middle part has a capacity scale coefficient k, which is unequal or unclosed to 1. But the factors of two parts at high SOC and low SOC region seem to be equal or closely approximate to 1. So, the part on the voltage platform is supposed to obey the curve scale rules, but the other parts do not.

To find the demarcation points of the curve and divide the three parts are essential to work. As for the LFP battery, the voltage curve has a relatively fixed shape. And it is easy to find the demarcation points by a simple algorithm:

a. ake a straight line l through the start point and the endpoint of the voltage curve:
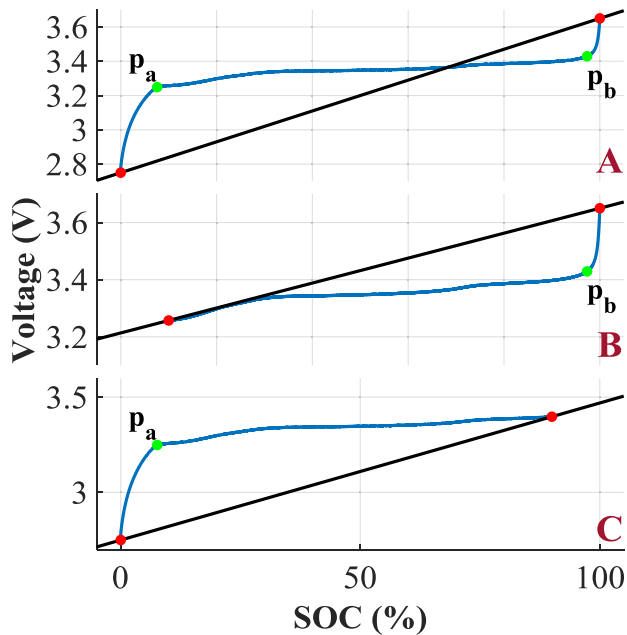
$$l: ax + by + c = 0 \tag{7}$$

FIGURE 5. Three cases of voltage curve (two of them are incomplete) and the corresponding straight lines for calculating the demarcation points.

FIGURE 6. The warping path map of sequence X and reference sequence Y.

where the a, b, c are the undetermined coefficients of the straight line. All the unknown factors can be determined by substituting the coordinates of the start and endpoint into this equation.

b. Calculate the distance d from all the points on the voltage curve to the line l :

$$\delta(p_i, l) = \begin{cases} +1, ax_i + by_i + c \leq 0 \\ -1, ax_i + by_i + c > 0 \end{cases} \quad (8)$$

$$d(p_i, l) = \frac{|ax_i + by_i + c|}{\sqrt{a^2 + b^2}} \cdot \delta(p_i, l) \quad (9)$$

where the $\delta(p_i, l)$ is a sign function which is positive when the point on the voltage curve is above the straight line and negative when the point is below the line, the $p_i(x_i, y_i)$ is the point on the voltage curve, and the $d(p_i, l)$ represents the distance from the point to the line.

c. Then find the point p_a whose distance to the straight line is maximum among the positive distance points (as point p_a in Figure 5 A, and find the point p_b whose distance is minimum among the negative distance points (as point p_b in Figure 5 A).p_a and p_b are regarded as the demarcation points.

d. Sometimes, the voltage curve is incomplete (as case B and C shown in Figure 5). Two limit threshold points are set as voltage 3.25V and 3.4V to avoid finding the wrong p_a and p_b. When the voltage of p_a is more than 3.25V, the p_a changes to the closest point to 3.25V or the start point on the curve. And similarly, when the voltage of p_b is less than 3.4V, the p_b changes to the closest point to 3.4V or the endpoint on the curve.
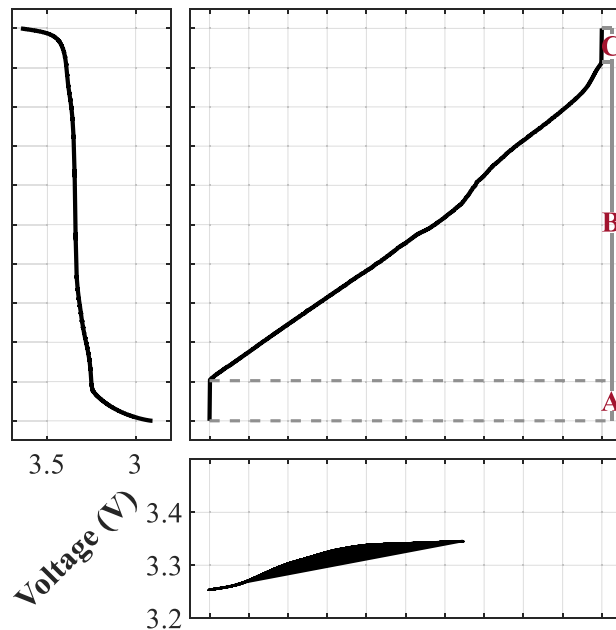
2) DIVISION OF THE VOLTAGE CURVE ACCORDING TO CURVE MATCHING RELATIONSHIP

The voltage curve is divided according to shape scalability in order to find a scalable portion of the voltage curve that varies with the charging current. And in this part, the division of the voltage curve according to the curve matching relationship is to get similar parts between two voltage curves.

It is mentioned above that the DTW algorithm is utilized for evaluating the similarity of two voltage curves and partitioning the matched and mismatched parts.

Assuming that there is a sequence X to be analyzed and the sequence Y is for reference. Considering the actual situation during the capacity estimating process, where the reference curve in **Step 1** is from capacity testing, and the reference curve in **Step 2** is synthesized by two curves in the battery pack. The SOC range of the sequence X is therefore contained in the SOC range of the sequence Y. There are two phenomena about the warping path of the two sequences:

a. As the part A in Figure 6, the initial point of sequence X matches plenty of points at the front of sequence Y, which means the x coordinate of warping path at part A is fixed;

b. As the part C in Figure 6, the endpoint of sequence X matches plenty of points at the end of sequence Y, which means the x coordinate of warping path at part C is also fixed;

Figure 7 is the matching illustration between sequence X and reference sequence Y. In Figure 7, the top line is the projection of sequence X on the SOC axis, the bottom line is the projection of sequence Y on the SOC axis, and the middle graph is the voltage curves of X and Y. As illustrated
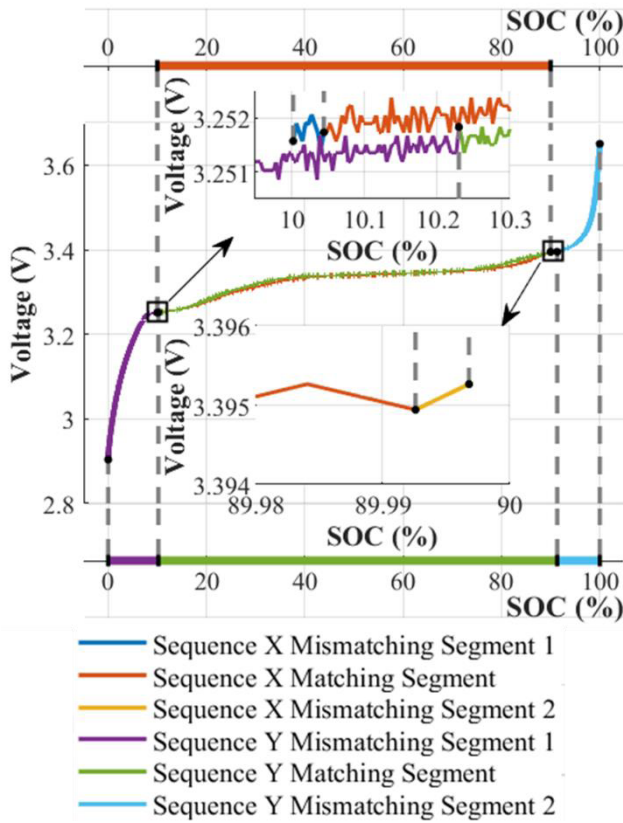
FIGURE 7. The matching illustration between sequence X and reference sequence Y.
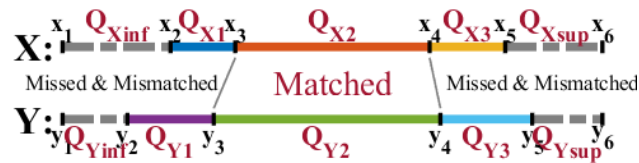

FIGURE 8. The abstract illustration of two voltage curve matching relationships.

in Figure 7, the segment of sequence X or Y corresponding to the part A and part C of the warping path map is the two mismatched segments, and the segment corresponding to the part B is the matched segment.

3) INTRODUCTION TO CAPACITY DIFFERENCE IDENTIFICATION METHOD

The method for identifying the capacity difference between two cells is to remain the matched part of sequence X but replace the missed and mismatched parts of sequence X with the corresponding parts of sequence Y.

Figure 8 is the abstraction of two curve matching relationship according to reality. The x_i and y_i ($i = 1, 2, \ldots, 6$) in the illustration are the split point of every part. The Q_* represents the charged-in capacity of each part. The grey dotted lines at both ends of the sequence X and sequence Y the in Figure 8 are the missing part of the voltage curve considering the real situation. The $Q_{Xinf}, Q_{Xsup}, Q_{Yinf}, Q_{Ysup}$,
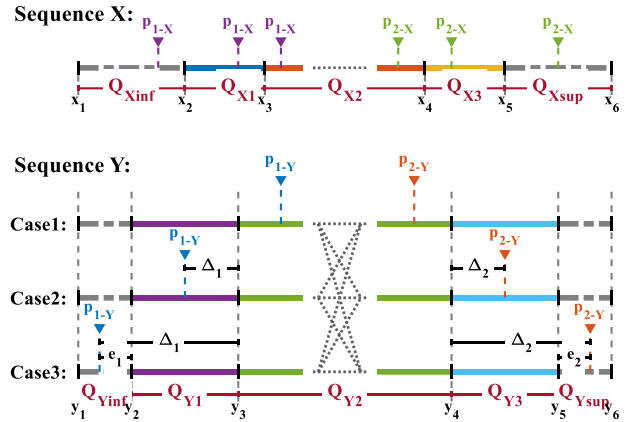

FIGURE 9. The abstract illustration of the voltage curve showing two kinds of division simultaneously.

which represent the unknown value of the charged-in capacity of these missing part. As shown in Figure 8, The capacities of sequence X and Y are:

$$Q_X = Q_{Xinf} + Q_{X1} + Q_{X2} + Q_{X3} + Q_{Xsup} \qquad (10)$$
$$Q_Y = Q_{Yinf} + Q_{Y1} + Q_{Y2} + Q_{Y3} + Q_{Ysup} \qquad (11)$$

Due to the existence of the unknowns in (10) and (11), the difference of capacity cannot be found out by subtracting the two formulas. But by replacing the unknown part and mismatched part of sequence X with the corresponding parts of sequence Y, the unknowns might be eliminated when subtracting the two equations. And the new Q_X' is:

$$Q_X' \approx \underbrace{\left(Q_{Yinf} + Q_{Y1}\right)'}_{Transformed\ part1} + Q_{X2} + \underbrace{\left(Q_{Y3} + Q_{Ysup}\right)'}_{Transformed\ part2} \qquad (12)$$

where the $\left(Q_{Yinf} + Q_{Y1}\right)'$ is the transformed part of sequence Y to replace the corresponding part of sequence X, so is the $\left(Q_{Y3} + Q_{Ysup}\right)'$.

So, the capacity difference between the two cells are:

$$\begin{aligned}
diff\ Q &\approx Q_X' - Q_Y \\
&= \left(Q_{Yinf} + Q_{Y1}\right)' \\
&\quad - \left(Q_{Yinf} + Q_{Y1}\right) + Q_{X2} \\
&\quad - Q_{Y2} + \left(Q_{Y3} + Q_{Ysup}\right)' \\
&\quad - \left(Q_{Y3} + Q_{Ysup}\right) \qquad (13)
\end{aligned}$$

When carrying out the replacement process, the shape scalability of the curve mentioned in Chapter III (A. 1)) needs to be considered. The shape scalability of the voltage curve makes the charged-in capacity of every part have a different transformation coefficient. As mentioned above, the scale coefficient of the voltage curve is equal to 1 at both ends of the curve, but the coefficient might be k which is not equal to 1 at the voltage platform. So, the scaling relationship also applies to the corresponding charged-in capacity.

There are three cases that the start points of the voltage platform located in the sequence, and so is the endpoint. The cases are shown in Figure 9. The point

p_1 (p_{1-X}, p_{1-Y}) and p_2 (p_{2-X}, p_{2-Y}) are the demarcation points based on shape scaling relationship, and the points x_i and y_i $(i = 1, 2, 3, 4, 5, 6)$ are the demarcation points based on matching relationship. The locations of p_1 and p_2 have three cases respectively, which means there are nine combinations of two points' locations.

For simplicity, there are only three combinations of sequences Y shown. Q_* are the charged-in capacity value of every part. The Δ_1 is defined as the charged-in capacity value of the part between y_3 and p_1, and e_i is defined as the charged-in capacity value of the part between y_2 and p_1, the Δ_2 and e_2 can be defined in the same way as they are shown in the figure. As illustrated in Figure 9, the capacity scaling coefficient of the part on the left of the p_1 is 1, the coefficient of the part on the right of the p_2 is also 1, and only the coefficient of the part between p_1 and p_2 may not equal 1. The scaling coefficient in the part of voltage platform is called the capacity scaling coefficient k, which can be calculated as follows:

$$k = \frac{Q_{X-matched\&scalable}}{Q_{Y-matched\&scalable}} \tag{14}$$

$$Q_{X-matched\&scalable} = Q_{charged-in}(m_1, m_2)$$

$$m_1 = \begin{cases} p_{1-X}, & p_{1-X} \in [x_3, x_4] \\ x_3, & p_{1-X} \in [x_1, x_3] \end{cases}$$

$$m_2 = \begin{cases} p_{2-X}, & p_{2-X} \in [x_3, x_4] \\ x_4, & p_{2-X} \in [x_4, x_6] \end{cases} \tag{15}$$

$$Q_{Y-matched\&scalable} = Q_{charged-in}(m_1, m_2)$$

$$m_1 = \begin{cases} p_{1-Y}, & p_{1-Y} \in [y_3, y_4] \\ y_3, & p_{1-Y} \in [y_1, y_3] \end{cases}$$

$$m_2 = \begin{cases} p_{2-Y}, & p_{2-Y} \in [y_3, y_4] \\ y_4, & p_{2-Y} \in [y_4, y_6] \end{cases} \tag{16}$$

where the $Q_{charged-in}(m_1, m_2)$ means the charged-in capacity between the point m_1 and m_2. Taking "transformed part 1" as an example, only the part on the right of the p_{1-X} needs to multiply with the capacity scaling coefficient k. The transformed capacity is expressed as:

$$(Q_{Yinf} + Q_{Y1})'$$
$$= \begin{cases} Q_{Yinf} - e_1 + k\Delta_1, & p_{1-Y} \in [y_1, y_2] \\ Q_{Yinf} + Q_{Y1} - \Delta_1 + k\Delta_1, & p_{1-Y} \in (y_2, y_3) \\ Q_{Yinf} + Q_{Y1}, & p_{1-Y} \in [y_3, y_4] \end{cases} \tag{17}$$

And the reduction formula of (17) is:

$$(Q_{Yinf} + Q_{Y1})' = (Q_{Yinf} - e_1)$$
$$+ (Q_{Y1} - \Delta_1 + e_1) + k\Delta_1$$
$$= Q_{Yinf} + Q_{Y1}$$
$$+ (k - 1)\Delta_1$$
$$\begin{cases} \Delta_1 > 0, & p_{1-Y} \in [y_1, y_3) \\ \Delta_1 = 0, & p_{1-Y} \in [y_3, y_4] \end{cases}$$
$$\begin{cases} e_1 > 0, & p_{1-Y} \in [y_1, y_2) \\ e_1 = 0, & p_{1-Y} \in [y_2, y_4] \end{cases} \tag{18}$$

Similarly, the capacity of "transformed part 2" is:

$$(Q_{Y3} + Q_{Ysup})'$$
$$= k\Delta_2 + (Q_{Y3} - \Delta_2 + e_2)$$
$$+ (Q_{Ysup} - e_2)$$
$$= (k - 1)\Delta_2 + Q_{Y3}$$
$$+ Q_{Ysup}$$
$$\begin{cases} \Delta_2 0, & p_{2-Y} \in [y_3, y_4] \\ \Delta_2 > 0, & p_{2-Y} \in (y_4, y_6] \end{cases}$$
$$\begin{cases} e_2 = 0, & p_{2-Y} \in [y_3, y_5] \\ e_2 > 0, & p_{2-Y} \in (y_5, y_6] \end{cases} \tag{19}$$

Substituting (18) and (19) into (13) and simplifying the equation:

$$diffQ \approx (k - 1)(\Delta_1 + \Delta_2) + (Q_{X2} - Q_{Y2}) \tag{20}$$

∗ The conditions of Δ_1 and Δ_2 are omitted here

Because of the absence of the parts at both ends of the sequence, e_1 and e_2 is always equal to 0, which should not equal 0 when calculating. It is illustrated in Figure 9 that $\Delta_1 = Q_{Y1} + e_1, p_{1-Y} \in [y_1, y_2)$ and $\Delta_2 = Q_{Y3} + e_2, p_{2-Y} \in (y_5, y_6]$. And the errors causing by the absence of the values e_1 and e_2 are:

$$error_1 = (k - 1)e_1, \quad p_{1-Y} \in [y_1, y_2)$$
$$error_2 = (k - 1)e_2, \quad p_{2-Y} \in (y_5, y_6] \tag{21}$$

It is revealed that the more incomplete the reference curve, the higher the recognition error.

B. ESTIMATING THE CAPACITY OF ALL CELLS IN BATTERY PACK

1) ESTIMATING THE CAPACITY OF ONE CELL IN THE BATTERY PACK

The data from the initial capacity test of cell i is set as the reference sequence Y, and the data from the latest cycle test of cell i is set as sequence X. By using the capacity difference identification method, the capacity difference between cell i in the initial state and cell i now can be calculated as:

$$Q_{i,now} = Q_{i,initial} + dQ \tag{22}$$

where the $Q_{i,now}$ is the estimating result of cell i, the $Q_{i,initial}$ is the initial capacity of cell i, and the dQ is the capacity difference of present capacity of cell i against the initial capacity of cell i.

2) IDENTIFYING THE CAPACITY DIFFERENCE OF CELLS IN THE PACK

Similarly, a reference sequence is demanded when identifying the capacity difference between cells. Different from the process of estimating the capacity of one cell, the reference sequence in the process of identifying the capacity of cells is synthesized but not picked from the battery pack. The curves, which have the minimum voltage at the start of the
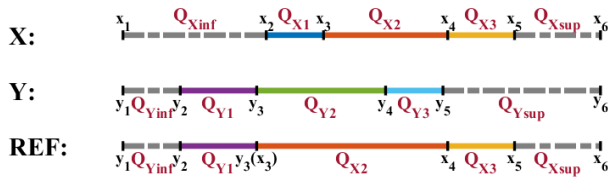
FIGURE 10. The illustration of the synthesizing process for reference sequence.

FIGURE 11. The illustration of the calculating process of the whole battery pack capacity.

charging process and the maximum voltage value at the end of the charging process, are selected to synthesize the reference sequence.

The synthesizing method is also similar to the capacity difference identification method introduced in Part A but not identical. It is assumed that the curve with the minimum voltage at the end of the charging process is sequence X and the curve with the maximum voltage at the start of the charging process is sequence Y. The segment between y_1 and y_3 (abbreviated to $[y_1, y_3]$) and the segment between x_3 and x_6 (abbreviated to $[x_3, x_6]$) in Figure 10 are remained for synthesizing the reference sequence. The voltage curves of the two parts are combined and the voltage of the segment $[y_1, y_3]$ times a factor k_x to make the voltage at y_3 equal to the voltage at x_3. The k_x is the ratio of voltage at x_3 to voltage at y_3. And the capacity of the synthesized sequence obeys the capacity transformation rules mentioned in Part A. After generating the reference sequence, all the capacity differences of the cells against the reference sequence can be calculated by equation (20).

After the above two steps, the capacity of all cells in the pack can be estimated.

3) ESTIMATING CAPACITY OF THE WHOLE BATTERY PACK

The capacity of the whole battery pack is defined by the maximum available capacity of the battery pack [21]. It can be explained as the common capacity of cells in the pack (such as the Q_{pack} in the Figure 11). In the process of capacity difference identification of the battery pack, the matching points of each cell voltage curve against the reference curve can be obtained, and the matching points are considered as the offset of each cell against the reference curve. So the capacity offset of each cell can be known. Assuming that all the cells are charged to the cut-off voltage before they are grouped, so the capacity of the whole battery pack can be calculated by this formula:

$$Q_{pack} = \min \left(Q_i + Q_{offset,i} \right) - \max \left(Q_{offset,i} \right) \quad (23)$$

where the Q_{pack} refers to the capacity of the whole battery pack, the Q_i is the estimated capacity of each cell in the pack, the $Q_{offset,i}$ refers to the capacity offset of each cell against the reference cell. All the variables are illustrated in Figure 11.

IV. VALIDATIONS AND RESULTS

Considering that the capacity differences may be tiny, which may cause a huge relative error when it applied to be
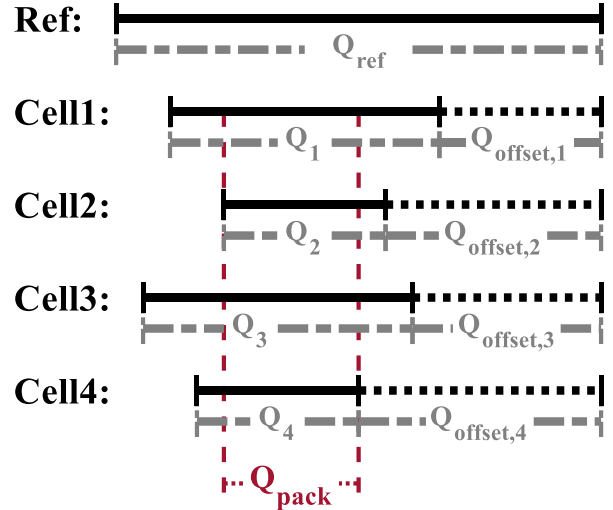
a denominator. For unifying the measurement of errors, the error ratio of rated capacity (ERRC) is proposed to evaluate the effects of identification, and the "error" in this paper refers to ERRC for convenience. The calculation formula is as follows:

$$ERRC = \frac{|Q_{real} - Q_{calc}|}{Q_{rated}} \times 100\%$$
$$\text{Or}$$
$$ERRC = \frac{|dQ_{real} - dQ_{calc}|}{Q_{rated}} \times 100\% \quad (24)$$

where Q_{real} is the cell capacity in actual, Q_{calc} is the cell capacity calculated, dQ_{real} is the capacity difference in actual, dQ_{calc} is the capacity difference calculated, and Q_{rated} is the rated capacity of cells in the battery pack. The ERRC is a relative value.

A. THE VALIDATION OF ESTIMATING ONE CELL CAPACITY

All the capacity of cells in the pack tested after specific cycle are shown in Table 1. In the 16 series LFP battery pack, the #1, #2 and #3 cells are selected to validate the method. The whole voltage curve of the initial capacity test is set as the reference sequence Y, and the sequence X is the voltage curve of performance test after the special number of cycles. Before executing the algorithm, the sequence X is intercepted to about 10% ∼ 95% SOC range. The results are illustrated in Figure 12, where the errors are all within 0.5% (about 0.3 Ah while the rated capacity is 60 Ah). The algorithm applying on different data of the same cell is precisely accurate.

B. THE VALIDATION OF IDENTIFYING CAPACITY DIFFERENCES IN PACK

The data of the pack at the 1160th cycle is used for validating the results of capacity difference identification. The real capacity of each cell is from the performance test after the 1160th cycling test. The results are illustrated in Figure 13.
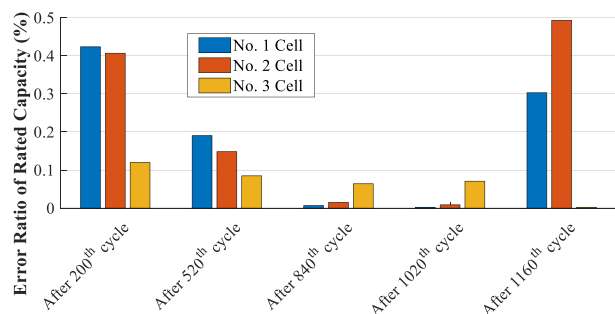
FIGURE 12. The results of the capacity estimation.

TABLE 1. The capacity of each cell after the specific aging cycle.

Cycle	0	200	520	840	1020	1160
#1	47.70	46.41	44.71	42.02	41.04	39.76
#2	48.20	46.42	44.21	41.13	40.06	38.33
#3	49.54	48.26	46.68	44.26	43.24	42.28
#4	49.48	48.30	46.73	44.27	43.35	42.37
#5	49.04	47.95	46.52	44.00	42.86	42.01
#6	48.73	47.44	45.93	43.19	42.21	40.97
#7	48.02	46.30	44.02	41.15	39.96	38.11
#8	47.90	46.78	45.07	41.97	40.93	39.42
#9	48.69	47.81	46.33	43.91	43.19	42.09
#10	49.12	48.10	46.68	44.16	43.38	42.29
#11	50.19	48.99	47.56	45.37	44.38	43.51
#12	49.51	48.62	47.16	45.21	44.22	43.38
#13	48.88	47.93	46.43	44.14	43.25	42.38
#14	48.24	47.13	45.66	42.77	41.70	40.34
#15	49.10	47.36	45.49	42.66	41.38	39.84
#16	48.20	47.42	45.91	43.35	42.43	41.45

*Cycle 0 means the initial capacity of the cells in the battery pack

The capacity benchmark in Figure 13 (a) is the capacity of the synthesized cells, which does not exist in actual. So, the cell with minimum capacity in the pack is selected as the benchmark when calculating the error of the capacity difference. From Figure 13 (b), the ERRC of the identified capacity difference is all less than 1.5% (less than 1 Ah while the rated capacity is 60 Ah).

The identification algorithm of capacity difference is the most important content in this paper. The error of capacity difference identifying is the primary source of the estimating error. The data from another 16 series battery pack is applied to validate the identification effect further. The 16 cells are also LFP batteries, whose rated capacity is 80 Ah. And the
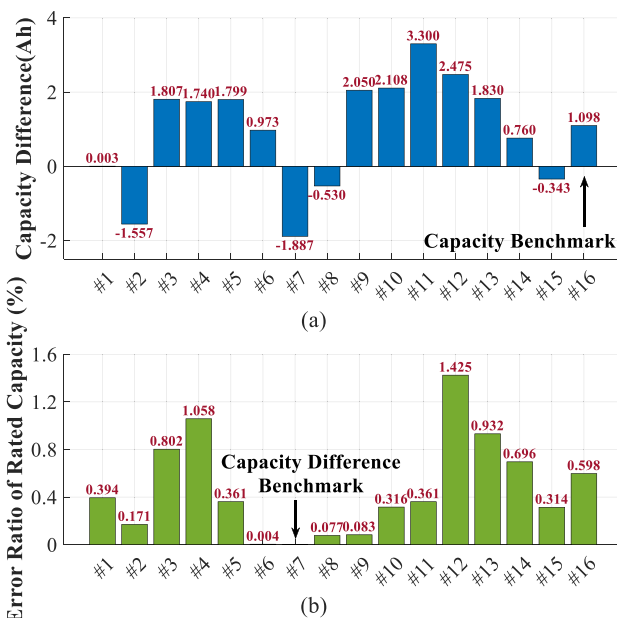

FIGURE 13. The identification result of capacity difference. (a) The capacity difference against reference curve; (b) The ERRC of the cells' capacity difference against the cell with minimum capacity.

capacity range of the pack is less than 6 Ah. The SOC of these cells are well controlled to simulate many kinds of conditions. Two groups of experiments are conducted: (1) all the cells are charged to charge cut-off voltage 3.65 V before grouped; (2) all the cells are discharged to discharge cut-off voltage 2.75 V before grouped. Control the SOC of the smallest capacity cell in the pack. The controlled depth of discharge (DOD) are $0 \sim 90\%$, $0 \sim 95\%$, $0 \sim 100\%$, $5\% \sim 100\%$, $10\% \sim 100\%$, $15\% \sim 100\%$, $20\% \sim 100\%$. The boxplots of the identification results are illustrated in Figure 14. The boxplot is a statistical representation of the data. Each boxplot has five horizontal lines, which from the bottom up are respectively minimum value, quarter quantile, mean value, three-quarter quantile, and maximum value. In the two SOC alignment situations and seven different DOD, the ERRC can be less than 3%. And when the DOD is $20\% \sim 100\%$ which can most easily happen in practices among these DOD situations, the maximum error can still be less than 2%. As the SOC at the end of charging decreases, the average level of error increases significantly. It is revealed that this method is sensitive to the high SOC part of the voltage curve. When the SOC at the start of charging increased, the average level of error increases slightly. Comparing to the condition of $0\sim100\%$ DOD, the degree of dispersion of the error distribution increases under other DOD conditions.

C. THE VALIDATION OF CAPACITY ESTIMATION IN PACK

The three steps process mentioned above is conducted to estimate the capacity of each cell in the pack. Take the 1# cell for instance, the charging data of 1# cell in the initial capacity test
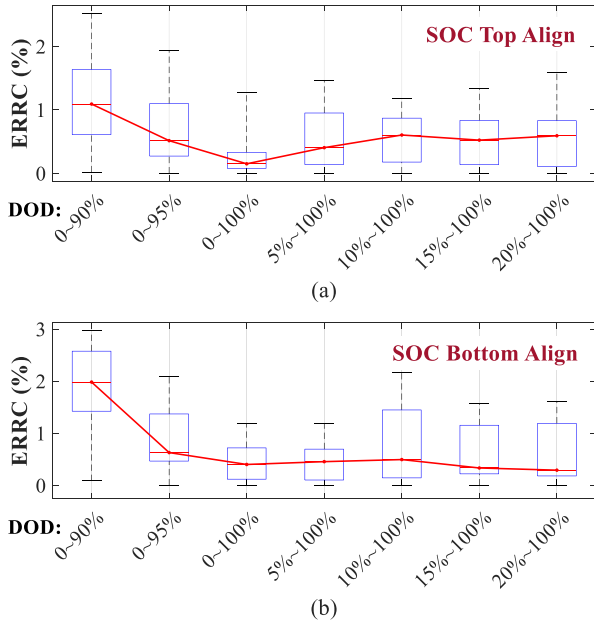
FIGURE 14. The boxplots of the identification results with different DOD. (a) The results of all the cells charging to 3.65 V (SOC Top Align); (b) The results of all the cells discharging to 2.75 V (SOC Bottom Align).

TABLE 2. The ERRC of the capacity of the whole capacity.

Cycle	200	520	840	1020	1160	MAE	RMSE
ERRC (%)	0.30	0.68	2.01	2.82	1.97	1.59	1.80

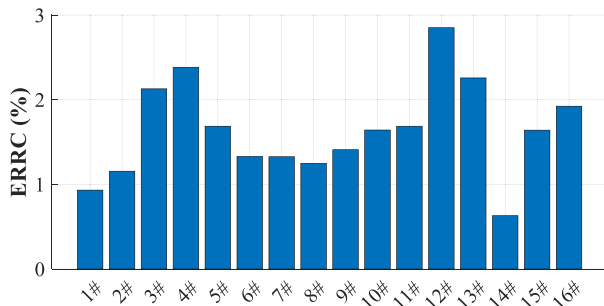*MAE: Mean Absolute Error
*RMSE: Root Mean Square Error

FIGURE 15. The ERRC of the cell capacity estimating.

is used as the reference sequence (sequence Y) in step 1, and the charging data of 1# cell in the 1160th cycling is used as the experimental sequence (sequence X). Then the estimated value of the 1# cell capacity in the 1160th cycle can be figured out. In step 2, the charging data of each cell in the 1160th cycling is applied. The capacity difference against 1# cell can be computed in step 2. In the last step 3, add the capacity of 1# cell estimated in step 1 and the capacity difference calculated in step2 together, the estimating capacity of each cell can be obtained. The ERRC of the estimation result is illustrated in Figure 15, and the error is less than 3% (about 1.8 Ah while the rated capacity is 60 Ah).

The estimated capacity of the whole battery pack is shown in Table 2. The errors increase slightly with the number of cycles but remain below 3%.

D. THE ADVANTAGE OF THE METHOD AGAINST OTHER METHODS

Many capacity estimation methods (such as the SVM-based method, the Bayesian-based method, the Neural Networks based method, etc.) all require a ''Data Training'' process. In other words, plenty of pre-experiments are needed before the capacity modeling. Changes in environmental conditions, such as the temperature and charging rate, can affect the accuracy of the model. But the method proposed in this paper only needs one charging voltage curve and the initial capacity of one cell in the pack, which greatly simplifies the modeling process. And according to the estimation results, the estimation accuracy of the method proposed in this paper is as good as the SVM-based method.

V. CONCLUSION

The capacity estimating for cells in the pack is one of the important and necessary technologies in lithium-ion battery management. By combining the mechanism of the battery aging and the consistency of the battery pack, this paper proposes a capacity estimation algorithm based on capacity difference identification. The major works are summarized as below:

(1) The mechanism of battery degradation is discussed. The loss of lithium inventory and the loss of active materials will not affect the full-cell voltage curve in the full life cycle of the cell, even the retired battery can still maintain the shape invariance of full-cell voltage curve. The shape invariance of the full-cell voltage curve is also known as the invariance of the OCV, which is the most primary principle in this paper.

(2) Based on the shape invariance of the battery voltage curve, the method of capacity difference identification between two similar voltage curves (sequence X & sequence Y) five-phase. By applying the DTW algorithm, the voltage curve is divided into matched segments and mismatched segments. By curve transformation, some parts of sequence X can be replaced by sequence Y, so the unknown quantities of sequence X can be substituted by the quantities of sequence Y. Then these unknown quantities can be eliminated to some extent when the subtracting capacity formula of each sequence. The replacement process obeys complex rules, which are based on the shape scalability and the matching relationship of the two similar curves. Following these rules, the capacity difference between the two curves can be calculated.

(3) A three-step method for estimating the capacity of each cell in the pack is proposed. Record the whole charging curve and initial capacity value of one cell before grouping as a pack. Then the method can be

applied on the pack. Firstly, calculate the capacity difference between the initial and now of the cell, and get the estimated capacity of the cell. Secondly, calculate the capacity difference between other cells and the selected cell in the pack. Finally, adding the capacity difference of cells to the capacity value of the reference cell, the estimating capacity of each cell can be obtained.

(4) The capacity difference identification algorithm is validated by changing the DOD of the pack, and the algorithm can remain a high accuracy (2%) in 20% ∼ 100% SOC. The three-step method is validated by some experimental data of the battery pack. The validation of every step of the method is carried out. The algorithm has high accuracy in each step. The final error of capacity estimation is less than 3% rated capacity, the absolute error is about 1.8 Ah and the rated capacity is 60Ah in this paper.

REFERENCES

[1] K. Liu, K. Li, Q. Peng, and C. Zhang, "A brief review on key technologies in the battery management system of electric vehicles," *Frontiers Mech. Eng.*, vol. 14, no. 1, pp. 47–64, 2019.

[2] S. Lee, J. Kim, J. Lee, and B. H. Cho, "State-of-charge and capacity estimation of lithium-ion battery using a new open-circuit voltage versus state-of-charge," *J. Power Sources*, vol. 185, no. 2, pp. 1367–1373, Dec. 2008.

[3] B. Scrosati and J. Garche, "Lithium batteries: Status, prospects and future," *J. Power Sources*, vol. 195, no. 9, pp. 2419–2430, May 2010.

[4] L. Lu, X. Han, J. Li, J. Hua, and M. Ouyang, "A review on the key issues for lithium-ion battery management in electric vehicles," *J. Power Sources*, vol. 226, pp. 272–288, Mar. 2013.

[5] L. Y. Wang, C. Wang, G. Yin, F. Lin, M. P. Polis, C. Zhang, and J. Jiang, "Balanced control strategies for interconnected heterogeneous battery systems," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 189–199, Jan. 2016.

[6] G. J. Offer, V. Yufit, D. A. Howey, B. Wu, and N. P. Brandon, "Module design and fault diagnosis in electric vehicle batteries," *J. Power Sources*, vol. 206, pp. 383–392, May 2012.

[7] J. Cao, N. Schofield, and A. Emadi, "Battery balancing methods: A comprehensive review," in *Proc. IEEE Vehicle Power Propuls. Conf.*, Sep. 2008, pp. 1–6.

[8] J. Kim, "Discrete wavelet transform-based feature extraction of experimental voltage signal for Li-ion cell consistency," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1150–1161, Mar. 2015.

[9] M. Dubarry, N. Vuillaume, and B. Y. Liaw, "Origins and accommodation of cell variations in Li-ion battery pack modeling," *Int. J. Energy Res.*, vol. 34, no. 2, pp. 216–231, 2010.

[10] T. B. Reddy, *Linden's Handbook of Batteries*, vol. 4. New York, NY, USA: McGraw-Hill, 2011.

[11] C. Zhang, Y. Jiang, J. Jiang, G. Cheng, W. Diao, and W. Zhang, "Study on battery pack consistency evolutions and equilibrium diagnosis for serial-connected lithium-ion batteries," *Appl. Energy*, vol. 207, pp. 510–519, Dec. 2017.

[12] Y. Li, K. Liu, A. M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, H. E. Hoster, and J. Van Mierlo, "Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review," *Renew. Sustain. Energy Rev.*, vol. 113, Oct. 2019, Art. no. 109254.

[13] C. Hu, B. D. Youn, and J. Chung, "A multiscale framework with extended Kalman filter for lithium-ion battery SOC and capacity estimation," *Appl. Energy*, vol. 92, pp. 694–704, Apr. 2012.

[14] J. Wang, P. Liu, J. Hicks-Garner, E. Sherman, S. Soukiazian, M. Verbrugge, P. Finamore, H. Tataria, and J. Musser, "Cycle-life model for graphite-LiFePO$_4$ cells," *J. Power Sources*, vol. 196, no. 8, pp. 3942–3948, 2011.

[15] M. Ecker, J. B. Gerschler, J. Vogel, S. Käbitz, F. Hust, P. Dechent, and D. U. Sauer, "Development of a lifetime prediction model for lithium-ion batteries based on extended accelerated aging test data," *J. Power Sources*, vol. 215, pp. 248–257, Oct. 2012.

[16] Y. Zheng, L. Lu, X. Han, J. Li, and M. Ouyang, "LiFePO$_4$ battery pack capacity estimation for electric vehicles based on charging cell voltage curve transformation," *J. Power Sources*, vol. 226, pp. 33–41, Mar. 2013.

[17] C. Lu, L. Tao, and H. Fan, "Li-ion battery capacity estimation: A geometrical approach," *J. Power Sources*, vol. 261, pp. 141–147, Sep. 2014.

[18] M. Dubarry, C. Truchot, and B. Y. Liaw, "Synthesize battery degradation modes via a diagnostic and prognostic model," *J. Power Sources*, vol. 219, pp. 204–216, Dec. 2012.

[19] M. Dubarry, C. Truchot, and B. Y. Liaw, "Cell degradation in commercial LiFePO$_4$ cells with high-power and high-energy designs," *J. Power Sources*, vol. 258, pp. 408–419, Jul. 2014.

[20] E. J. Keogh and M. J. Pazzani, "Derivative dynamic time warping," in *Proc. SIAM Int. Conf. Data Mining*. Philadelphia, PA, USA: SIAM, Apr. 2001, pp. 1–11.

[21] J.-C. Jiang, F. Wen, J.-P. Wen, and C.-P. Zhang, "Battery management system used in electric vehicles," *Dianli Dianzi Jishu/Power Electron.*, vol. 45, no. 12, pp. 2–10, 2011.
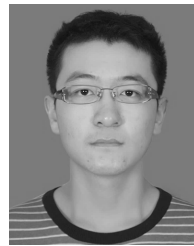
YANG LIU was born in Shandong, China. He received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2016. He is currently pursuing the Ph.D. degree with Beijing Jiaotong University, Beijing.

He is currently with the National Active Distribution Network Technology Research Center, Beijing Jiaotong University. His research interests include Li-ion batteries, parameters estimation, and safety on battery.

CAIPING ZHANG was born in Henan, China. She received the B.S. degree in vehicular engineering from the Henan University of Science and Technology, Luoyang, China, in 2004, and the Ph.D. degree in vehicle engineering from the Beijing Institute of Technology, Beijing, China, in 2010.

From 2010 to 2012, she was a Postdoctoral Researcher with Beijing Jiaotong University, Beijing, where she is currently an Associate Professor. Her research interests include battery modeling, state estimation, optimal charging, battery second use technology, and battery energy-storage systems.

JIUCHUN JIANG was born in Jilin, China. He received the B.S. degree in electrical engineering and the Ph.D. degree in power system automation from Northern Jiaotong University, Beijing, China, in 1993 and 1999, respectively.

He is currently a Professor with the School of Electrical Engineering, Beijing Jiaotong University, Beijing. His research interests include battery application technology, electric car charging stations, and microgrid technology.

Dr. Jiang was a recipient of the National Science and Technology Progress 2nd Award for his work on electric vehicle (EV) bus systems, and the Beijing Science and Technology Progress 2nd Award for his work on EV charging systems.

YAN JIANG was born in Jilin, China. He received the B.S. degree in electrical engineering and automation from Beijing Jiaotong University, Beijing, China, in 2014, where he is currently pursuing the Ph.D. degree. He is currently with the National Active Distribution Network Technology Research Center, Beijing Jiaotong University. His research interests include battery state estimation, battery second use technology, and battery energy storage systems.

LINJING ZHANG was born in Hebei, China. She received the Ph.D. degree from the Beijing Institute of Technology, Beijing, China, in 2015. She is currently a Lecturer with Beijing Jiaotong University. Her main research focuses on decay mechanisms, thermodynamics and kinetics characteristics, fast charging technique, and safety of lithium-ion batteries.

WEIGE ZHANG was born in Gansu, China. He received the M.S. and Ph.D. degrees in electrical engineering from Northern Jiaotong University, Beijing, China, in 1997 and 2013, respectively.

From 1993 to 1994, he was an Engineer with Hohhot Railway Bureau. He is currently an Associate Professor with the School of Electrical Engineering, Beijing Jiaotong University, Beijing. His research interests include battery pack application technology, power electronics, and intelligent distribution systems.

• • •