Robust Battery Lifetime Prediction with Noisy Measurements by Total-Least-Squares Regression

Ting Lu, *Member, IEEE*, Xiaoang Zhai, Sihui Chen, Yang Liu, Jiayu Wan, Guohua Liu, and Xin Li, *Fellow, IEEE*

***Abstract*—Machine learning techniques have recently gained significant attentions in renewable energy research, and they have been extensively applied to provide data-driven solutions to tackle a variety of grand challenges. In this paper, we focus on the problem of battery lifetime prediction. Our objective is to accurately estimate the remaining lifetime of a battery given the measurement data collected by physical testing. Importantly, we propose to adopt total-least-squares (TLS) regression to accommodate the noisy measurements by minimizing the total errors associated with both input features and predicted outcomes. In addition, a novel numerical solver is developed to determine all unknown model coefficients both efficiently and robustly. Our numerical experiments based on three public datasets for commercial Lithium-Ion batteries demonstrate that TLS can effectively reduce the modeling error by up to 8.8%, compared against ordinary-least-squares (OLS) regression.**

***Index Terms*—Battery lifetime prediction, total-least-squares regression, measurement error, energy storage**

I. Introduction

RECHARGEABLE battery technology is gaining rapidly increasing attentions in the modern world, as sustainability is becoming a critical issue of global interests. As a popular class of rechargeable battery, Lithium-Ion battery has been widely deployed due to its advantages such as high energy density and low cost, which enable reliable and economic energy storage in multiple application domains including portable devices, electric vehicles, and power grids [1]-[2]. The quality of a Lithium-Ion battery can be quantitatively assessed by its lifetime, and hence, battery lifetime prediction has been extensively studied in the literature, providing a strategic guideline to the manufacturing, maintenance, and usage of batteries [1], [3]-[5].

A variety of machine learning techniques have been recently adopted to predict battery lifetime with high accuracy and low cost [1], [6]. These approaches aim to train a regression model that takes the battery measurements as its inputs to predict the corresponding lifetime value. For example, a significant progress has been made in [1] by identifying a set of important features from the early-cycle battery discharge curves, and these features are strongly correlated with the battery lifetime. This seminal work enables battery lifetime prediction at an early stage and is of practical utility in real-world applications. Furthermore, various regression modeling algorithms, such as elastic net [1], Gaussian process regression [7], support vector regression [6], [8], XGBoost [8] and deep neural networks [9]-[10], have been proposed to predict battery lifetime in practice.

As a common practice in regression modeling, most conventional approaches attempt to minimize the prediction error for battery lifetime when a regression model is trained. However, the input features (e.g., voltage, current, resistance, temperature, etc.) are often noisy because they are measured through probe techniques with measurement errors [5]. In addition, when the measured physical metrics are used to derive new features for battery lifetime prediction, the measurement errors on these physical metrics may be further propagated to the derived features [11]-[12]. Consider the discharge capacity curve as an example. While it is computed by integrating the current over the testing period, it suffers from the cumulative error posed by both time-domain sampling and current measurement. These measurement errors associated with input features have not been appropriately taken into account by the conventional regression modeling methods.

In order to tackle the aforementioned challenge, we propose to adopt total-least-squares (TLS) regression to predict battery lifetime in this paper. Unlike the conventional approaches, TLS minimizes the total errors associated with both input features and predicted outcomes [13]-[14]. In other words, it is able to learn an accurate regression model with consideration of noisy features. Due to this advantage, TLS has been broadly used in many practical applications with an errors-in-variables (EIV) nature, such as image reconstruction and financial data forecast [14]. Our numerical experiments based on three public datasets [1] for commercial Lithium-Ion batteries demonstrate that TLS can reduce the modeling error by up to 8.8%, compared against ordinary-least-squares (OLS) regression [15].

The remainder of this paper is organized as follows. In Section II, we derive the mathematical formulation of TLS for

T. Lu, X. Zhai and G. Liu are with Donghua University, Shanghai, China (e-mails: luting@dhu.edu.cn, xiaoang_zhai@mail.dhu.edu.cn, ghliu@dhu.edu.cn).

S. Chen and J. Wan are with Southern University of Science and Technology, Shenzhen, Guangdong, China (e-mails: 12232304@mail.sustech.edu.cn, wanjy@sustech.edu.cn).

Y. Liu and X. Li are with Duke Kunshan University, Kunshan, Jiangsu, China (e-mails: {yang.liu2, xinli.ece}@dukekunshan.edu.cn).

Y. Liu, J. Wan and X. Li are the corresponding authors.

battery lifetime prediction. The numerical solver for TLS is discussed in Section III. Experimental results are presented to demonstrate the efficacy of TLS in Section IV. Finally, we conclude in Section V.

II. Problem Formulation

Considering a set of features $\mathbf{x} = [x_1\ x_2\ \cdots\ x_N]^T$, where N denotes the total number of features and the superscript T represents the transpose of a vector or matrix, our objective is to learn a mapping from $\mathbf{x}$ to the battery lifetime y [16]-[17]:

$$y = \sum_{m=1}^{M} w_m \cdot g_m(\mathbf{x}) - \varepsilon_y , \tag{1}$$

where $g_m(\mathbf{x})$ is the m-th basis function, w_m denotes the m-th model coefficient, ε_i stands for the modeling error following a zero-mean Gaussian distribution $N(0, \sigma_y^2)$, and M represents the total number of basis functions.

Given a set of samples $\{(\mathbf{x}_k, y_k);\ k = 1, \cdots, K\}$, the model coefficients $\{w_m;\ m = 1, \cdots, M\}$ can be determined by minimizing the total squared error [15]:

$$\min_{\mathbf{w}} \quad \left\| \mathbf{y} - \mathbf{G} \cdot \mathbf{w} \right\|_2^2 , \tag{2}$$

where $\|\bullet\|_2$ denotes the L_2 norm of a vector and

$$\mathbf{G} = \begin{bmatrix} g_1(\mathbf{x}_1) & g_2(\mathbf{x}_1) & \cdots & g_M(\mathbf{x}_1) \\ g_1(\mathbf{x}_2) & g_2(\mathbf{x}_2) & \cdots & g_M(\mathbf{x}_2) \\ \vdots & \vdots & \vdots & \vdots \\ g_1(\mathbf{x}_K) & g_2(\mathbf{x}_K) & \cdots & g_M(\mathbf{x}_K) \end{bmatrix} \tag{3}$$

$$\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_M \end{bmatrix}^T \tag{4}$$

$$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_K \end{bmatrix}^T . \tag{5}$$

The aforementioned approach is referred to as OLS regression in the literature and it aims to find the maximum-likelihood solution $\mathbf{w}$ for the unknown model coefficients [15].

In practice, each basis function $g_m(\mathbf{x})$ may be noisy, as the feature vector $\mathbf{x}$ is measured with measurement errors. In this case, Eq. (1) should be re-written as:

$$y + \varepsilon_y = \sum_{m=1}^{M} w_m \cdot \left[g_m(\mathbf{x}) + \varepsilon_{g,m} \right], \tag{6}$$

where $\varepsilon_{g,m}$ represents the measurement error associated with the m-th basis function $g_m(\mathbf{x})$. In this paper, we assume that $\varepsilon_{g,m}$ follows a zero-mean Gaussian distribution $N(0, \sigma_{g,m}^2)$.

To derive the maximum-likelihood solution $\mathbf{w}$ for the unknown model coefficients in (6), we make two further assumptions. First, both ε_y and $\{\varepsilon_{g,m};\ m = 1, \cdots, M\}$ can be normalized to standard Gaussian distribution $N(0, 1)$ by appropriately scaling y and $\{g_m(\mathbf{x});\ m = 1, \cdots, M\}$ respectively. Second, ε_y and $\{\varepsilon_{g,m};\ m = 1, \cdots, M\}$ are statistically independent.

With these two assumptions, it is straightforward to show that the likelihood of observing a sample $(\mathbf{x}_k, y_k)$ is equal to:

$$\begin{aligned} p(\mathbf{x}_k, y_k) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varepsilon_{y,k}^2}{2} \right) \times \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varepsilon_{g,m,k}^2}{2} \right) \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^{M+1}} \exp\left[-\frac{\varepsilon_{y,k}^2}{2} - \sum_{m=1}^{M} \frac{\varepsilon_{g,m,k}^2}{2} \right] \end{aligned} , \tag{7}$$

where $\varepsilon_{y,k}$ and $\varepsilon_{g,m,k}$ denote the k-th samples for ε_y and $\varepsilon_{g,m}$ respectively. Furthermore, by assuming that all samples in the dataset $\{(\mathbf{x}_k, y_k);\ k = 1, \cdots, K\}$ are statistically independent, the likelihood for observing these K samples is equal to:

$$\begin{aligned} \prod_{k=1}^{K} p(\mathbf{x}_k, y_k) &\propto \prod_{k=1}^{K} \exp\left[-\frac{\varepsilon_{y,k}^2}{2} - \sum_{m=1}^{M} \frac{\varepsilon_{g,m,k}^2}{2} \right] \\ &= \exp\left[-\sum_{k=1}^{K} \frac{\varepsilon_{y,k}^2}{2} - \sum_{k=1}^{K} \sum_{m=1}^{M} \frac{\varepsilon_{g,m,k}^2}{2} \right] \end{aligned} . \tag{8}$$

Hence, the maximum-likelihood solution $\mathbf{w}$ can be found by solving the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\varepsilon}, \mathbf{E}, \mathbf{w}} \quad & \|\boldsymbol{\varepsilon}\|_2^2 + \|\mathbf{E}\|_F^2 \\ \text{s.t.} \quad & \mathbf{y} + \boldsymbol{\varepsilon} = (\mathbf{G} + \mathbf{E}) \cdot \mathbf{w} \end{aligned} , \tag{9}$$

where $\|\bullet\|_F$ denotes the Frobenius norm of a matrix and

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{y,1} & \varepsilon_{y,2} & \cdots & \varepsilon_{y,K} \end{bmatrix}^T \tag{10}$$

$$\mathbf{E} = \begin{bmatrix} \varepsilon_{g,1,1} & \varepsilon_{g,2,1} & \cdots & \varepsilon_{g,M,1} \\ \varepsilon_{g,1,2} & \varepsilon_{g,2,2} & \cdots & \varepsilon_{g,M,2} \\ \vdots & \vdots & \vdots & \vdots \\ \varepsilon_{g,1,K} & \varepsilon_{g,2,K} & \cdots & \varepsilon_{g,M,K} \end{bmatrix} . \tag{11}$$

Note that minimizing the cost function in (9) is equivalent to maximizing the likelihood in (8). Such an approach is referred to as TLS regression in the literature [13].

III. Numerical Solver

The optimization problem in (9) is non-convex due to its quadratic equality constraints [18]. Hence, it is difficult to find the optimal solution both efficiently (i.e., with low computational cost) and robustly (i.e., with guaranteed global convergence). In this section, we transform the original non-convex problem to an equivalent version that can be easily solved.

First, we note that Eq. (9) is reduced to the following convex optimization problem if the vector $\mathbf{w}$ is known:

$$\begin{aligned} \min_{\boldsymbol{\varepsilon}, \mathbf{E}} \quad & \|\boldsymbol{\varepsilon}\|_2^2 + \|\mathbf{E}\|_F^2 \\ \text{s.t.} \quad & \mathbf{y} + \boldsymbol{\varepsilon} = (\mathbf{G} + \mathbf{E}) \cdot \mathbf{w} \end{aligned} , \tag{12}$$

where $\boldsymbol{\varepsilon}$ and $\mathbf{E}$ are the problem unknowns. Eq. (12) minimizes a convex quadratic cost function subject to a set of linear equality constraints. Its optimal solution can be analytically derived by using the KKT condition [18]. For this purpose, we calculate the Lagrangian of (12):

$$L(\boldsymbol{\varepsilon}, \mathbf{E}, \boldsymbol{\lambda}) = \|\boldsymbol{\varepsilon}\|_2^2 + \|\mathbf{E}\|_F^2 + \boldsymbol{\lambda}^T \left[(\mathbf{G} + \mathbf{E}) \cdot \mathbf{w} - \mathbf{y} - \boldsymbol{\varepsilon} \right], \tag{13}$$

where $\boldsymbol{\lambda}$ is a vector containing K Lagrange multipliers. By taking the partial derivatives of $L(\boldsymbol{\varepsilon}, \mathbf{E}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\varepsilon}$ and $\mathbf{E}$ and setting them to zero, we have:

$$2 \cdot \boldsymbol{\varepsilon} - \boldsymbol{\lambda} = \mathbf{0} \tag{14}$$

$$2 \cdot \mathbf{E} + \boldsymbol{\lambda} \cdot \mathbf{w}^T = \mathbf{0} . \tag{15}$$

The optimal solutions of $\boldsymbol{\varepsilon}$ and $\mathbf{E}$ must satisfy (14)-(15) and the linear equality constraints:

$$\mathbf{y} + \boldsymbol{\varepsilon} = (\mathbf{G} + \mathbf{E}) \cdot \mathbf{w} . \tag{16}$$

Solving the linear system of (14)-(16) yields:

$$\boldsymbol{\varepsilon} = \frac{\mathbf{G}\cdot\mathbf{w}-\mathbf{y}}{\|\mathbf{w}\|_2^2+1} \tag{17}$$

$$\mathbf{E} = -\frac{(\mathbf{G}\cdot\mathbf{w}-\mathbf{y})\cdot\mathbf{w}^T}{\|\mathbf{w}\|_2^2+1}. \tag{18}$$

Substituting (17)-(18) into (9), the cost function of TLS regression can be expressed as:

$$\|\boldsymbol{\varepsilon}\|_2^2 + \|\mathbf{E}\|_F^2 = \frac{\|\mathbf{G}\cdot\mathbf{w}-\mathbf{y}\|_2^2}{\|\mathbf{w}\|_2^2+1}. \tag{19}$$

Note that the cost function in (19) has been minimized over $\boldsymbol{\varepsilon}$ and $\mathbf{E}$, and it is now a non-convex function of $\mathbf{w}$.

To efficiently minimize the cost function in (19), we construct the following optimization problem [18]:

$$\begin{aligned}\min_{\mathbf{w},t} \quad & \frac{\|\mathbf{G}\cdot\mathbf{w}-t\cdot\mathbf{y}\|_2^2}{\|\mathbf{w}\|_2^2+t^2} \\ \text{s.t.} \quad & t=1\end{aligned}. \tag{20}$$

It is equivalent to:

$$\begin{aligned}\min_{\boldsymbol{\theta}} \quad & \frac{\boldsymbol{\theta}^T\cdot\mathbf{B}\cdot\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|_2^2} \\ \text{s.t.} \quad & \theta_{M+1}=1\end{aligned}, \tag{21}$$

where

$$\mathbf{B} = \begin{bmatrix} \mathbf{G}^T\mathbf{G} & -\mathbf{G}^T\mathbf{y} \\ -\mathbf{y}^T\mathbf{G} & \mathbf{y}^T\mathbf{y} \end{bmatrix} \tag{22}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \mathbf{w} \\ t \end{bmatrix}, \tag{23}$$

and θ_{M+1} is the (M+1)-th element of $\boldsymbol{\theta}$. Studying (21), we notice that its cost function is independent of the magnitude of $\boldsymbol{\theta}$. Hence, solving the following optimization problem without the constraint $\theta_{M+1} = 1$ yields the same optimal cost function value:

$$\begin{aligned}\min_{\boldsymbol{\theta}} \quad & \frac{\boldsymbol{\theta}^T\cdot\mathbf{B}\cdot\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|_2^2} \\ \text{s.t.} \quad & \boldsymbol{\theta}\neq\mathbf{0}\end{aligned}. \tag{24}$$

Eq. (24) represents the well-known problem of Rayleigh quotient minimization [18]. Its optimal solution $\boldsymbol{\theta}^*$ is the eigenvector corresponding to the least eigenvalue of $\mathbf{B}$. Once $\boldsymbol{\theta}^*$ is known, we can scale $\boldsymbol{\theta}^*$ by its (M+1)-th element θ^*_{M+1} to determine the optimal solution $\boldsymbol{\theta}$ for (21):

$$\boldsymbol{\theta} = \frac{\boldsymbol{\theta}^*}{\theta^*_{M+1}}. \tag{25}$$

Thus, the optimal solution $\mathbf{w}$ for unknown model coefficients can be obtained by taking the first M elements of $\boldsymbol{\theta}$:

$$\mathbf{w} = \begin{bmatrix}\theta_1 & \theta_2 & \cdots & \theta_M\end{bmatrix}^T. \tag{26}$$

The major steps for TLS regression are summarized in Algorithm 1.

Algorithm 1: Total-Least-Squares (TLS) Regression

1. Start from a set of basis functions $\{g_m(\mathbf{x}); m=1, \cdots, M\}$ and a set of samples $\{(\mathbf{x}_k, y_k); k=1, \cdots, K\}$.
2. Calculate the matrix $\mathbf{G}$ in (3) and the vector $\mathbf{y}$ in (5).
3. Construct the matrix $\mathbf{B}$ based on (22).
4. Compute the eigenvector $\boldsymbol{\theta}^*$ of $\mathbf{B}$ corresponding to its least eigenvalue.
5. Calculate the vector $\boldsymbol{\theta}$ in (25) by normalizing $\boldsymbol{\theta}^*$.
6. Compute the vector $\mathbf{w}$ in (26), containing the optimal model coefficients.

IV. Experimental Results

In this section, three public datasets based on commercial Lithium-Ion batteries [1] are used to demonstrate the efficacy of our proposed TLS method for battery lifetime prediction. These three datasets are referred to as "Dataset 1", "Dataset 2" and "Dataset 3", and they are composed of 41, 43 and 40 samples, respectively.

While 20 features in total have been extracted for these datasets [1], we further manually select a subset of 5 important features, denoted as $\mathbf{x} = [x_1\ x_2 \cdots x_5]^T$, based on domain expertise. In this example, because a small number of samples are available in each dataset, we only use 5 features for regression modeling in order to avoid overfitting. The physical meanings of these selected features are summarized in Table I. To reduce the nonlinearity for our modeling task, we take the logarithm for both the battery lifetime y and the first feature x_1 [1]. With these nonlinear transformations, we adopt the following linear model template to predict log(y):

$$\log(y) \approx w_1\log(x_1) + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6. \tag{27}$$

Table I
Description of Selected Features

Feature Name	Description
x_1	Variance of the difference in the discharge capacity curves as a function of voltage between the 10th and 100th cycles
x_2	Slope of the capacity fade curve fitted by a linear function
x_3	Discharge capacity of the 2nd cycle
x_4	Average charging time from the 2nd cycle to the 6th cycle
x_5	The difference of internal resistances between the 2nd and 100th cycles

For testing and comparison purposes, two regression modeling methods have been implemented: (i) TLS and (ii) OLS. To assess the modeling accuracy, we partition each dataset into (i) a training dataset to determine the unknown model coefficients, and (ii) a testing dataset to evaluate the modeling accuracy based on root mean squared error (RMSE) [19]. We randomly select 30% samples to form the testing dataset and the remaining 70% samples form the training dataset. In addition, we vary the training dataset size from 70% samples to 20% samples in order to study the robustness of TLS and OLS when only an extremely small number of training samples are available. To improve numerical stability, we normalize the predicted outcome log(y) and all features {log(x_1), x_2, x_3, x_4, x_5} so that they have zero mean and unit variance over the training dataset. The aforementioned experiments are repeated for 200 times, where the training and testing datasets are independently and randomly generated at each run. The median of 200 RMSE values is reported for each method so that the error metric is not strongly biased due to random fluctuations.

Fig. 1 shows the RMSEs of TLS and OLS for three datasets, where 70% samples are used to form the training dataset in each case. By taking into account the measurement errors for $\mathbf{x}$, TLS consistently achieves higher modeling accuracy than OLS over all three datasets. Most notably, the modeling error is reduced by 8.8% for “Dataset 3”. It, in turn, demonstrates that TLS can effectively improve the modeling accuracy by minimizing the total squared error for both $\mathbf{x}$ and y when the input features are noisy.

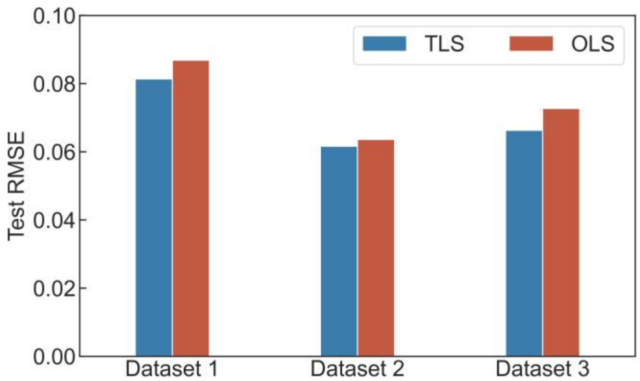

Fig. 1. The RMSEs of TLS and OLS are shown for three datasets, where 70% samples are used to form the training dataset in each case.

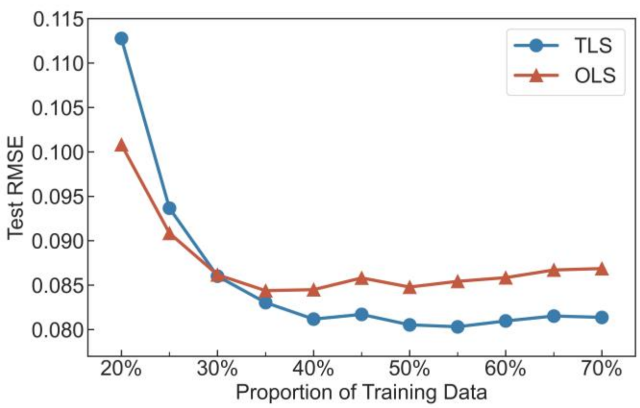

Fig. 2. The RMSEs of TLS and OLS are reduced, as the training dataset size varies from 20% samples to 70% samples for “Dataset 1”.

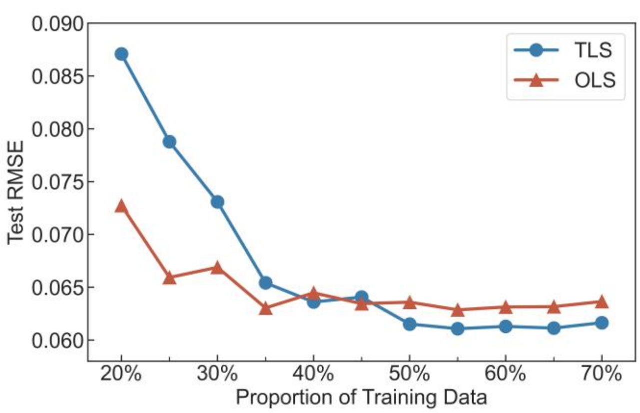

Fig. 3. The RMSEs of TLS and OLS are reduced, as the training dataset size varies from 20% samples to 70% samples for “Dataset 2”.

Fig. 2, Fig. 3 and Fig. 4 further show the RMSEs of TLS and OLS as functions of the training dataset size for three cases, respectively. Studying these figures reveals three important observations. First, the RMSEs of both TLS and OLS are reduced, as more training samples are available and, hence, more information is incorporated for model training. Second, when only a small number of training samples are used, OLS outperforms TLS. Note that TLS needs to solve $\boldsymbol{\varepsilon}$, $\mathbf{E}$ and $\mathbf{w}$ in (9), while OLS only needs to solve $\mathbf{w}$ in (2). As TLS involves more problem unknowns, it is less robust to over-fitting and requires more training samples to reliably determine all unknowns. Third, when the training dataset is sufficiently large, TLS is more accurate than OLS, which is consistent with our observation in Fig. 1.

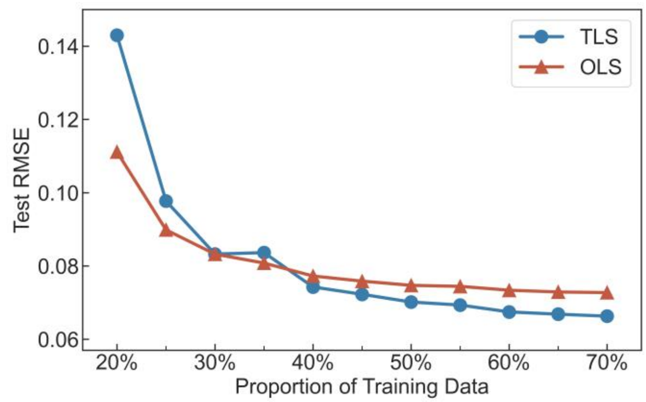

Fig. 4. The RMSEs of TLS and OLS are reduced, as the training dataset size varies from 20% samples to 70% samples for “Dataset 3”.

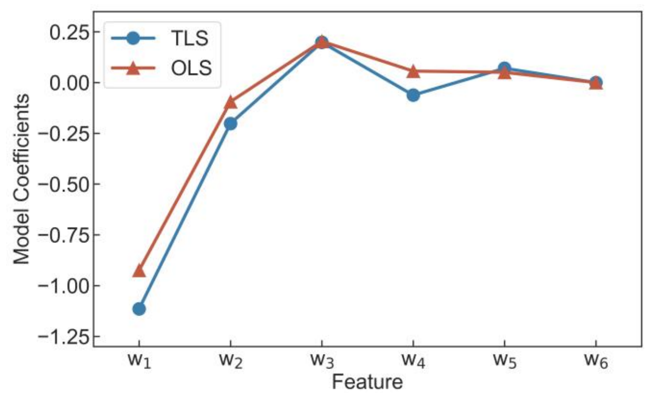

Fig. 5. The model coefficients of TLS and OLS are shown for “Dataset 1”, where 70% samples are used to form the training dataset.

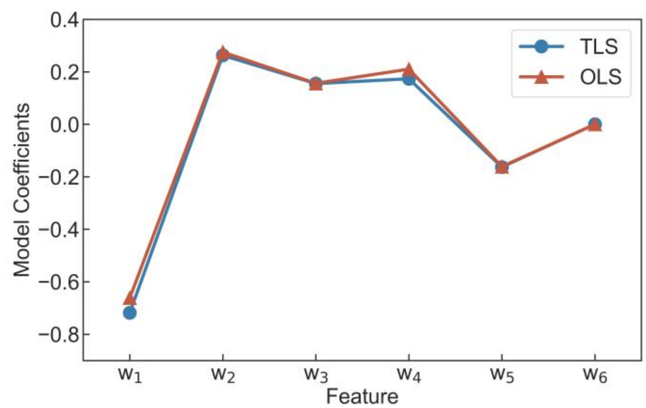

Fig. 6. The model coefficients of TLS and OLS are shown for “Dataset 2”, where 70% samples are used to form the training dataset.

Finally, we visualize the model coefficients of TLS and OLS for three datasets, as shown in Fig. 5, Fig. 6 and Fig. 7 respectively, where 70% samples are used to form the training datasets. Note that the model coefficient w_6 is equal to 0, because the predicted outcome log(y) has been normalized to have zero mean. The model coefficients solved by TLS and OLS are similar, but not identical, due to the measurement errors associated with input features. In this example, as the TLS errors are less than the OLS errors, the model coefficients solved by TLS are expected to be more accurate.

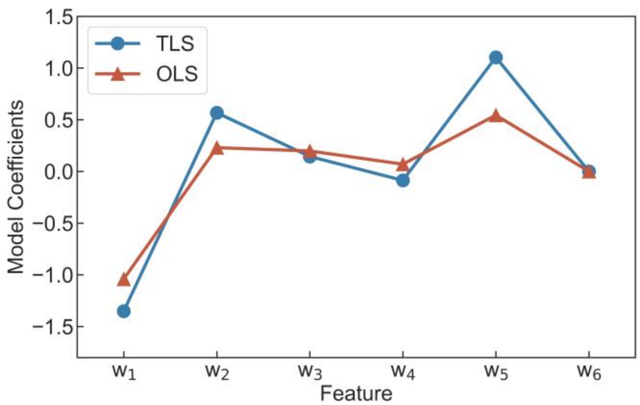

Fig. 7. The model coefficients of TLS and OLS are shown for “Dataset 3”, where 70% samples are used to form the training dataset.

V. Conclusions

In this paper, we propose a novel TLS method to predict battery lifetime based on regression modeling. TLS can effectively take into account the measurement noises associated with both input features and battery lifetime. While the TLS problem is cast as a non-convex optimization, a numerical solver is developed to determine the unknown model coefficients both efficiently (i.e., with low computational cost) and robustly (i.e., with guaranteed global convergence). Numerical experiments based on three public datasets for commercial Lithium-Ion batteries show that TLS can effectively reduce the modeling error by up to 8.8%, compared against the traditional OLS method. It, in turn, demonstrates the efficacy of TLS as a promising approach for battery lifetime prediction in practice.

References

[1] K. A. Severson *et al.*, “Data-driven prediction of battery cycle life before capacity degradation,” *Nature Energy*, vol. 4, no. 5, pp. 383–391, 2019.

[2] B. Scrosati and J. Garche, “Lithium batteries: Status, prospects and future,” *Journal of Power Sources*, vol. 195, no. 9, pp. 2419-2430, 2010.

[3] S. Chetoui and S. Reda, “Workload- and user-aware battery lifetime management for mobile SoCs,” *Proceedings of Design, Automation & Test in Europe Conference & Exhibition*, pp. 1679-1684, 2021.

[4] M. Rossi, A. Toppano and D. Brunelli, “Real-time optimization of the battery banks lifetime in hybrid residential electrical systems,” *Proceedings of Design, Automation & Test in Europe Conference & Exhibition*, pp. 1-6, 2014.

[5] J. Wei, G. Dong and Z. Chen, “Remaining useful life prediction and state of health diagnosis for Lithium-Ion batteries using particle filter and support vector regression,” *IEEE Trans. on Industrial Electronics*, vol. 65, no. 7, pp. 5634-5643, 2018.

[6] M. A. Patil *et al.*, “A novel multistage support vector machine based approach for Li ion battery remaining useful life estimation,” *Applied Energy*, vol. 159, pp. 285-297, 2015.

[7] Y. Zhang *et al.*, “Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning,” *Nature Communications*, vol. 11, no. 1, p. 1706, 2020.

[8] J. Zhu, *et al.*, “Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation,” *Nature Communications*, vol. 13, no. 1, p. 2261, 2022.

[9] D. Chen *et al.*, “A novel deep learning-based life prediction method for lithium-ion batteries with strong generalization capability under multiple cycle profiles,” *Applied Energy*, vol. 327, p. 120114, 2022.

[10] C.-W. Hsu *et al.*, “Deep neural network battery life and voltage prediction by using data of one cycle only,” *Applied Energy*, vol. 306, p. 118134, 2022.

[11] J. Taylor *et al.*, “An insight into the errors and uncertainty of the lithium-ion battery characterization experiments,” *Journal of Energy Storage*, vol. 24, p. 100761, 2019.

[12] A. Moradpour *et al.*, “Measurement uncertainty in battery electrochemical impedance spectroscopy,” *IEEE Trans. on Instrumentation and Measurement*, vol. 71, pp. 1-9, 2022.

[13] A. Beck, A. Ben-Tal and M. Teboulle. “Finding a global optimal solution for a quadratically constrained fractional quadratic problem with applications to the regularized total least squares,” *SIAM Journal on Matrix Analysis and Applications*, vol. 28, no. 2, pp. 425-445, 2006.

[14] H. Zhu, G. Leus and G. B. Giannakis, “Sparsity-cognizant total least-squares for perturbed compressive sampling,” *IEEE Trans. on Signal Processing*, vol. 59, no. 5, pp. 2002-2016, 2011.

[15] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[16] X. Li, “Finding deterministic solution from underdetermined equation: large-scale performance variability modeling of analog/RF circuits,” *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 29, no. 11, pp. 1661-1668, 2010.

[17] F. Wang *et al*., “Bayesian model fusion: large-scale performance modeling of analog and mixed-signal circuits by reusing early-stage data,” *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 35, no. 8, pp. 1255-1268, 2016.

[18] A. Beck, *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*. Society for Industrial and Applied Mathematics, 2014.

[19] W. Zhang *et al.*, “Automatic clustering of wafer spatial signatures”, *Proceedings of Design Automation Conference*, pp. 1-6, 2013.