Computers and Chemical Engineering 198 (2025) 109099

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/cace

Short communication

Interpretation of high-dimensional regression coefficients by comparison with linearized compressing features

Joachim Schaeffer [a,b], Jinwook Rhyu [b], Robin Droop [c], Rolf Findeisen [a], Richard D. Braatz [b,*]

[a] Control and Cyber-Physical Systems Laboratory, Technical University of Darmstadt, Germany
[b] Massachusetts Institute of Technology, Cambridge, MA, USA
[c] Technical University of Munich, Germany

ARTICLE INFO

Dataset link: https://github.com/JoachimSchaeffer/HDRegAnalytics, https://data.matr.io/1/

Keywords:
Interpretable machine learning
Linear regression
Linearization
High dimensions
Functional data
Lithium-ion batteries

ABSTRACT

Linear regression is often deemed inherently interpretable; however, challenges arise for high-dimensional data. We focus on further understanding how linear regression approximates nonlinear responses from high-dimensional functional data. We develop a linearization method to derive feature coefficients, which we compare with the closest regression coefficients of the path of regression solutions. We showcase the methods on battery data case studies where a single nonlinear compressing feature, $g : \mathbb{R}^p \to \mathbb{R}$, is used to construct a synthetic response, $\mathbf{y} \in \mathbb{R}$. This unifying view of linear regression and compressing features for high-dimensional functional data helps to understand (1) how regression coefficients are shaped in the highly regularized domain, (2) how regression coefficients relate to linearized feature coefficients, and (3) how the shape of regression coefficients changes as a function of regularization to approximate nonlinear responses by exploiting local structures.

1. Introduction

Interpretable machine learning methods aim to turn data into insights and knowledge for decision-making (Verdonck et al., 2024; Zheng and Amanda, 2018; Domingos, 2012). Linear regression provides simple models that allow interpretation of how inputs $\mathbf{X} \in \mathbb{R}^{n \times p}$ affect responses $\mathbf{y} \in \mathbb{R}^n$ when the input dimension p is smaller than the sample size n (Hastie et al., 2009; James et al., 2021). However, for many problems in domains such as chemical engineering, chemistry, and biology, the input dimension is much larger than the sample size, $p \gg n$ (Schaeffer and Braatz, 2022). Then, multicollinearity is inherently present because any column can be reconstructed by a linear combination of other columns, giving rise to the nullspace (Strang, 2016; Schaeffer et al., 2024b). Regression methods commonly used in high dimensions, such as RR and PLS, learn coefficients that are orthogonal to the nullspace by construction (Schaeffer et al., 2024b), affecting interpretability. For data with a functional structure, the fused lasso can improve interpretability (Schaeffer et al., 2024b) but might affect model accuracy. Another research area is concerned with functional data analytics (Ramsay and Silverman, 2005) and improving interpretability there (James et al., 2009).

We are motivated by the problem of prediction of lithium-ion battery cycle life (Schaeffer et al., 2024a; Severson et al., 2019) based on data collected from early-stage cycling. Voltage and current were continuously measured with high resolution ($p = 1000$), but the total number of batteries tested was small ($n = 124$), leading to high-dimensional data with functional structure. The variance, a single nonlinear compressing feature, displayed high predictive power for cycle life (Severson et al., 2019).[1] A different approach is to use linear regression directly on high-dimensional data, leading to comparable prediction performance (Schaeffer et al., 2024b,a). We are interested in understanding the connection between high-dimensional linear regression and regression using a single nonlinear feature. In particular, we are interested in understanding how regression coefficients, $\boldsymbol{\beta} \in \mathbb{R}^p$, learn a mapping to a response, $\mathbf{y}$, that is a nonlinear response of the input data, $\mathbf{X}$. To investigate this question, we construct synthetic nonlinear responses $\mathbf{y}$ by using the compressing feature g that takes lithium-ion battery data (Severson et al., 2019) as input. We then derive linearized feature coefficients, $\boldsymbol{\beta}_{\text{T1}}$, by applying the first-order Taylor expansion to g. Subsequently, we analyze the path of regression solutions and compare the feature coefficients with the regression coefficients closest to the feature coefficients. This linearization perspective

* Corresponding author.
E-mail address: braatz@mit.edu (R.D. Braatz).

[1] The term *compressing feature* refers to transformations of data to produce a scalar, $g : \mathbb{R}^p \to \mathbb{R}$.

https://doi.org/10.1016/j.compchemeng.2025.109099
Received 17 November 2024; Received in revised form 9 February 2025; Accepted 11 March 2025
Available online 21 March 2025

helps to improve the understanding of the highly regularized regression domain in high dimensions. The key contribution of this short note is the connection between linearized features and high-dimensional linear regression.

2. Feature linearization

Linear regression models can be written as $(\mathbf{y}-\bar{y}\mathbf{1}) = (\mathbf{X}-\mathbf{1}\bar{\mathbf{x}}^\top)\boldsymbol{\beta}+\epsilon$ for a data set, and as

$$(y_i-\bar{y}) = (\mathbf{x}_i-\bar{\mathbf{x}})^\top\boldsymbol{\beta}+\epsilon_i, \tag{1}$$

for a single sample i, where $\mathbf{y}\in\mathbb{R}^n$ is a vector of observed responses, $\bar{y}\in\mathbb{R}$ is the mean of the responses, $\boldsymbol{\epsilon}\in\mathbb{R}^n$ is the model error, $\mathbf{1}\in\mathbb{R}^n$ is a vector of ones, $\mathbf{x}_i\in\mathbb{R}^p$ contains the data of sample i, and $\bar{\mathbf{x}}\in\mathbb{R}^p$ contains the column means of the data $\mathbf{X}\in\mathbb{R}^{n\times p}$. Furthermore, $\boldsymbol{\beta}\in\mathbb{R}^p$ are the regression coefficients, implying that they are estimated by regression. Here, we are concerned with high-dimensional data with n samples and p measurements made over a continuous domain, and therefore $p\gg n$. Such discrete measurements of an assumed smooth underlying process are also called *functional data* (Ramsay and Silverman, 2005). Additionally, (1) shows that regression coefficients are gradients with respect to the input (Hastie et al., 2009).

We want to compare nonlinear compressing features, $g:\mathbb{R}^p\to\mathbb{R}$, with regression coefficients. In the following, we linearize g and construct *feature coefficients*. We make three assumptions: Assumption 1: There exists a nonlinear compressing feature, g, that maps the data and the response, i.e., $g(\mathbf{x}_i)=y_i+\epsilon_i^*$, where ϵ_i^* is the error of the ground truth model (irreducible error). Assumption 2: For simplicity, we set $\epsilon_i^*=0\ \forall i\in\{1,\dots,n\}$, but relax this assumption later. Assumption 3: g is differentiable.

The first-order Taylor series approximation for functions with multiple inputs (Duistermaat, 2010; Hörmander, 2003), applied to g and evaluated at $\bar{\mathbf{x}}$, is

$$g_{\mathrm{T1}}(\mathbf{x}_i;\bar{\mathbf{x}}) = g(\bar{\mathbf{x}})+(\mathbf{x}_i-\bar{\mathbf{x}})^\top\nabla g(\bar{\mathbf{x}}) = z_i. \tag{2}$$

where $\nabla g(\bar{\mathbf{x}}) := \partial g/\partial\mathbf{x}|_{\mathbf{x}=\bar{\mathbf{x}}}$ is the gradient of g evaluated at $\bar{\mathbf{x}}$, and z_i is the linearized feature corresponding to sample i. Rearranging (2) yields

$$z_i-g(\bar{\mathbf{x}}) = (\mathbf{x}_i-\bar{\mathbf{x}})^\top\nabla g(\bar{\mathbf{x}}). \tag{3}$$

The linearized feature z_i is an estimate for y_i. Inserting $z_i+\epsilon_{i,\mathrm{T1}}^g=y_i$, where $\epsilon_{i,\mathrm{T1}}^g$ is the linearization error of g corresponding to sample i, into (3) yields

$$y_i-g(\bar{\mathbf{x}}) = (\mathbf{x}_i-\bar{\mathbf{x}})^\top\nabla g(\bar{\mathbf{x}})+\epsilon_{i,\mathrm{T1}}^g. \tag{4}$$

Using Assumption 1 and the superscript g to indicate that the coefficients correspond to response that is subtracted by $g(\bar{\mathbf{x}})\mathbf{1}$, we can reformulate the regression problem (1) to

$$(\mathbf{y}-g(\bar{\mathbf{x}})\mathbf{1}) = (\mathbf{X}-\mathbf{1}\bar{\mathbf{x}}^\top)\boldsymbol{\beta}^g+\epsilon^g. \tag{5}$$

Comparing (4) and (5) shows that both equations have the same structure. However, the regression coefficients obtained by (5) differ from (1) because

$$g(\bar{\mathbf{x}}) =: \bar{z} \tag{6}$$

$$\bar{z}\approx\bar{y} = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{x}_i) \tag{7}$$

might not be a good approximation. To correct for this, we suggest to estimate a scalar factor m via an additional regression step

$$(\mathbf{y}-\bar{y}\mathbf{1}) = m(\mathbf{z}-\bar{z}\mathbf{1})+\epsilon_{\mathrm{T1}}, \tag{8}$$

which also allows to relax Assumption 2. Here, $\mathbf{z}$ is the vector of all z_i and ϵ_{T1} is the vector of linearized feature model errors. The scalar m is obtained by Ordinary Least Squares (OLS) and ensures that the feature coefficients estimate the true response. Combining (3), (6), and (8) gives

$$(\mathbf{y}-\bar{y}\mathbf{1}) = m(\mathbf{X}-\mathbf{1}\bar{\mathbf{x}}^\top)\nabla g(\bar{\mathbf{x}})+\boldsymbol{\epsilon}_{\mathrm{T1}}. \tag{9}$$

We define the vector $\boldsymbol{\beta}_{\mathrm{T1}} := m\nabla g(\bar{\mathbf{x}})$ and refer to it as *feature coefficients*. The feature coefficients can be compared with the path of regression solutions, as described in the next section.

Remark. For simplicity, we work with a fixed training set in the following case studies and do not model the regression coefficients probabilistically.

Regression solution path:. Various linear models are suitable for regressing high-dimensional data. The PLS model is commonly used in the chemometrics community, while RR, lasso, and Elastic Net (EN) are more popular in the machine learning community. We do not consider the lasso and EN because their sparse regression coefficients inevitably differ from the structure of smooth derivatives that yield $\boldsymbol{\beta}_{\mathrm{T1}}$. Analyzing the lasso or EN would require further investigation and relaxation of assumptions. The PLS model has a single discrete regularization parameter, the number of components. RR has a single continuous regularization parameter, λ, making the path of regression solutions continuous. We want to explore the regression solution path and compare the regression coefficients *closest* to the feature coefficients with the feature coefficients. A description and introduction to high-dimensional linear regression can be found in Schaeffer and Braatz (2022).

While various norms can be used for calculating the distance between the regression and feature coefficients, we suggest estimating the regularization parameters based on the ℓ^2-norm,

$$\min_{\lambda}\|\boldsymbol{\beta}(\lambda)-\boldsymbol{\beta}_{\mathrm{T1}}\|_2^2, \tag{10}$$

due to its convenient properties. Inserting the closed-form solution of RR in (10) yields

$$\min_{\lambda}\left\|(\mathbf{X}^\top X+\lambda\mathbf{I})^{-1}\mathbf{X}^\top y-\boldsymbol{\beta}_{\mathrm{T1}}\right\|_2^2, \tag{11}$$

which can be solved numerically. The regression coefficients of PLS and RR are orthogonal to the nullspace of the data $\mathbf{X}$ by construction (Schaeffer et al., 2024b), However, due to the nullspace, there can be situations where $\|\boldsymbol{\beta}-\boldsymbol{\beta}_{\mathrm{T1}}\|_2^2$ is *large* but the corresponding predictions are *similar* (see Appendix A.5 and Schaeffer et al. (2024a) for further discussion).

3. Case studies

The feature linearization method in this short note is motivated by cycle life prediction for lithium-ion batteries from early-stage cycling experiments where it was observed that a single compressing feature, the variance, had a good prediction performance, similar to the performance of regression in high dimensions (compare models in Schaeffer et al. (2024a) and Severson et al. (2019)). Therefore, we use the lithium-ion battery data set (Severson et al., 2019) for the case studies in this work. The data is the difference in discharge capacity of cycles 100 and 10 as a function of voltage (Fig. 1). Here, we use the training data set ($\mathbf{X}^{\mathrm{train}}\in\mathbb{R}^{40\times1000}$ after removing one outlier), and test data set ($\mathbf{X}^{\mathrm{test}}\in\mathbb{R}^{43\times1000}$) with the split according to Severson et al. (2019). To test our methodology, we only use the training set in the following. We provide predictions results on the test set in Appendix A.1. Furthermore, we do not z-score the data because the variance feature was applied to the non-z-scored data in Severson et al. (2019). Furthermore, the noise in the data is heteroscedastic, and the signal-to-noise ratio drops significantly in the region 3.2–3.5 V for physical reasons (see Schaeffer et al. (2024b) for more details on z-scoring as a design choice). To study feature linearization, we construct synthetic responses $\mathbf{y}$ by using two different compressing features.

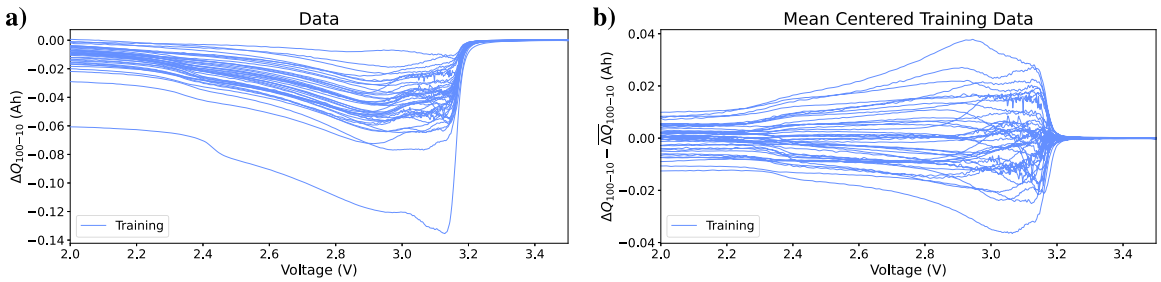

Fig. 1. Lithium-ion data from discharge cycles (Severson et al., 2019). (a) Training data are plotted as curves. (b) Mean-centered training data curves with one outlier removed.

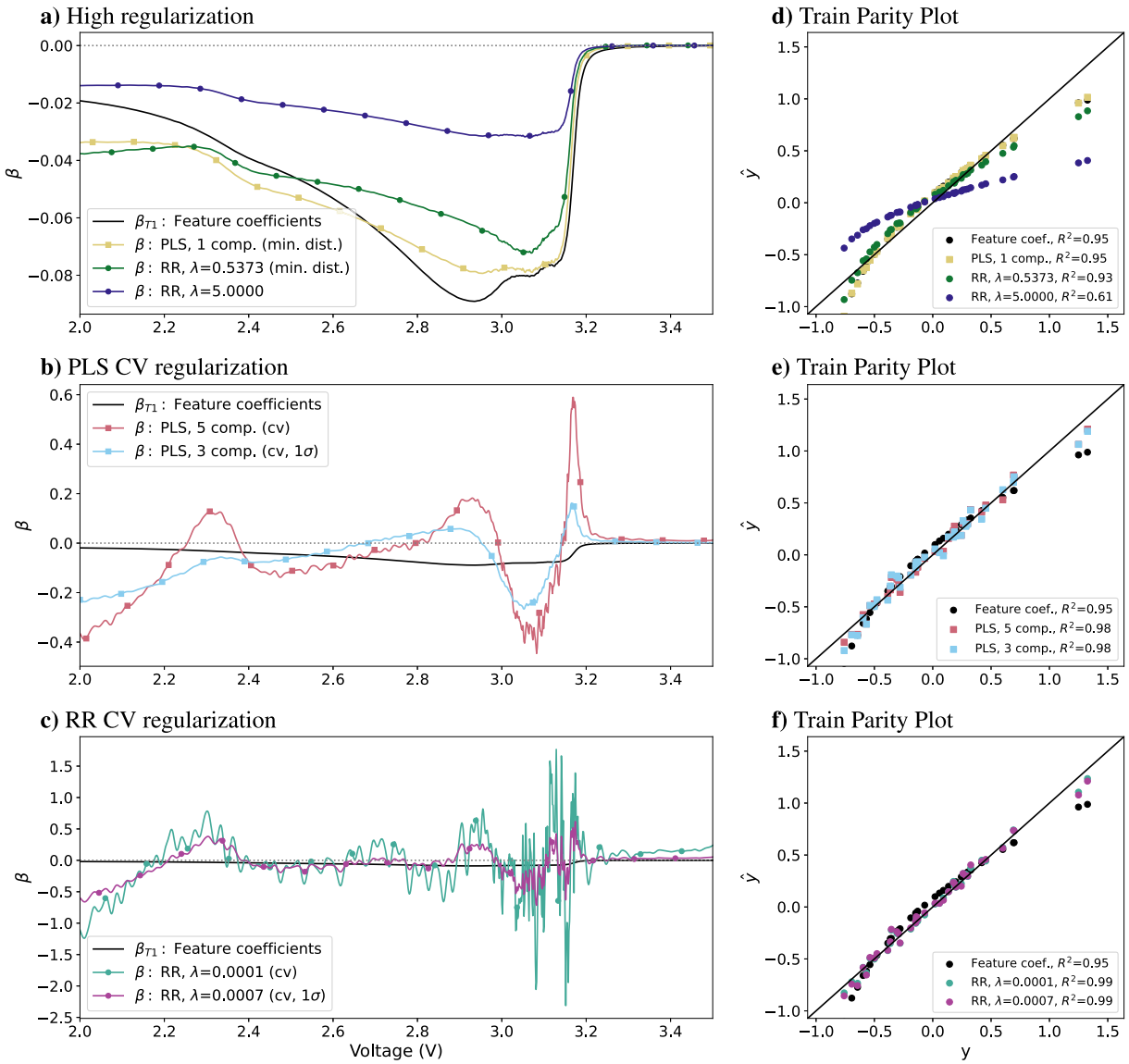

Fig. 2. First case study with the sum-of-squares response. (a) PLS and RR with high regularization and sum-of-squares feature coefficients. (b) PLS regression coefficients obtained by cross-validation and sum-of-squares feature coefficients. (c) RR regression coefficients obtained by cross-validation and sum-of-squares feature coefficients. (d), (e), (f) Train data parity plots.

Case study I:. The first case study uses the sum-of-squares $g(\mathbf{x}_i) = \sum_{j=1}^{p} x_{i,j}^2 = y_i$ (Fig. 2). The feature coefficients of the sum-of-squares feature are the scaled mean of the data (compare Fig. 2a and Fig. A.1d; the derivation is shown in Appendix A.4). For large regularization, the regression coefficients of RR approach zero over the entire domain (Fig. 2a) while remaining a shape that is similar to the standard deviation of the data (Fig. A.1d). The RR and PLS regression coefficients with minimal ℓ^2-distance to the feature coefficients have a very similar shape as the feature coefficients. However, they do not match exactly due to the way how regularization shapes the regression coefficients. The corresponding predictions shown in Fig. 2d are very similar for the feature coefficients and RR and PLS regression coefficients with minimal ℓ^2-distance. The feature coefficients yield an R^2 value of 0.95, which indicates the sum-of-squares feature is well approximated by

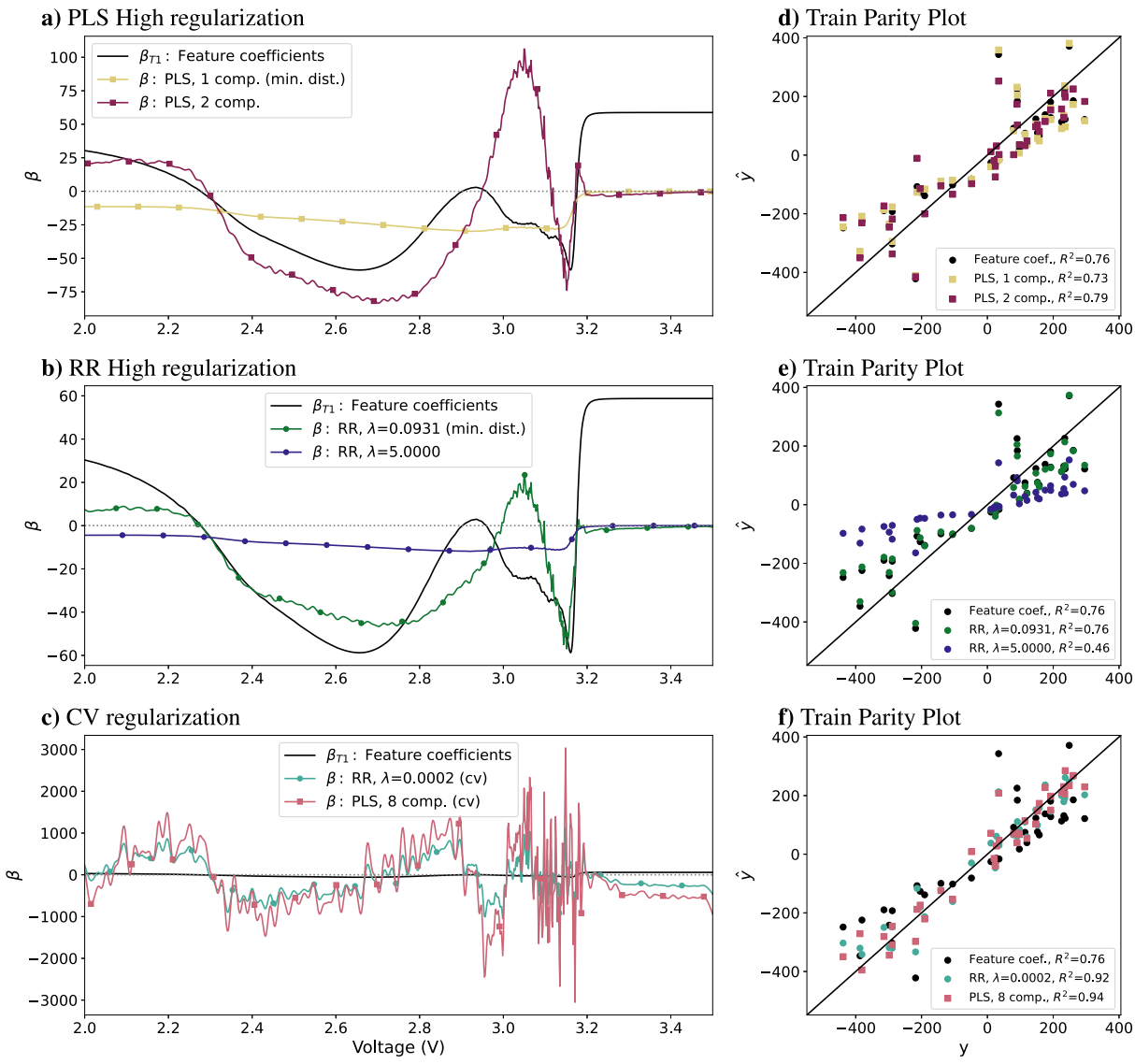

Fig. 3. Second case study with the sinusoidal response. (a) PLS high regularization and sinusoidal feature coefficients. (b) RR high regularization and sinusoidal feature coefficients. (c) RR and PLS regression coefficients obtained by cross-validation and sinusoidal feature coefficients. (d), (e), (f) Train data parity plots.

its linearization. Lowering the regularization further by choosing the regularization value by 10-fold cross-validation according to the one-standard-error rule increases the magnitude of regression coefficients. The cross-validation procedure is described in Appendix A.3. Further decreasing the regularization by not using the one-standard-error rule leads to an even larger magnitude of the regression coefficients that change faster across the voltage input domain (Fig. 2). The prediction results of the regression coefficients based on cross-validation are fairly similar with R^2 values close to one (Fig. 2ef). These results indicate that, for this data set, the sum-of-squares response can be well approximated by regression in high dimensions.

Case study II:. The second case study considers a sinusoidal feature, $g(\mathbf{x}_i) = \sum_{j=1}^{p} \sin\left(\frac{2\pi}{0.06} x_{i,j}\right) = y_i$. The sinusoidal feature coefficients have a very different shape than the variance of the data (Fig. 3a and Fig. A.1). Similar to the sum-of-squares feature, the shape of the highly regularized regression coefficients of RR and PLS are dominated by the directions with high variance (Figs. 3ab). The PLS model with one component has minimum ℓ^2-distance to the feature coefficients but is still dominated by the variance of the data and does not yet pick up the characteristic shape of the sinusoidal feature coefficients. The PLS model with two components shows a similar shape as the feature coefficients but overshoots them (Fig. 3a). The predictions of the feature coefficients and PLS coefficients with one and two components are very similar (Fig. 3d). For RR, the minimum distance regression coefficients are closer to the feature coefficients in terms of ℓ^2-norm than the PLS coefficients and also have the characteristic shape of the feature coefficients (Fig. 3b). The predictions of the feature coefficients and minimum distance RR coefficients are very similar (Fig. 3e). Choosing the regularization value by 10-fold cross-validation yields regression coefficients with a larger magnitude, exploiting local structures (Fig. 3c), phenomenologically similar to the sum-of-squares case study.

Approximation of nonlinearity:. Conceptually, a highly regularized linear model is expected to capture the "key" directions of the data, subject to the regularization constraint. When decreasing the regularization, the model tries to approximate the nonlinearity linearly in high dimensions. Although the linear model might not fully capture nonlinearity, the case studies in this work and Schaeffer and Braatz (2022) suggest that linear models can perform well on high-dimensional data when a limited amount of nonlinearity is present. For the sum-of-squares case study, the nonlinearity is well approximated by the linear models (Fig. 2f, training set, and Fig. A.2a, test set predictions). For

the sinusoidal response, the nonlinearity is approximated slightly less well by the linear models (Fig. 3f, training set, and Fig. A.2a test set predictions). However, in both cases, the regression models perform significantly better that the feature coefficients, demonstrating that linear regression in high dimensions can approximate a nonlinear response by exploiting local structures.

Cycle life prediction and limitations:. Case studies I and II linearize the true function that maps **X** and **y**. However, for any practical application, this mapping is not known and needs to be learned by regression. In the motivating application of battery cycle life prediction (Severson et al., 2019), a single nonlinear feature, the variance, performed well. However, the variance is not the true mapping. The case studies showed that the regression coefficients can look very similar to the linearized response if it is known, but do not match exactly due to the assumptions made in Section 2. Therefore, even if the regression coefficients closest to the unknown linearized true mapping would be available, it would still be an open challenge to use them for feature design.

4. Conclusion

Linear regression is often deemed inherently interpretable; however, challenges arise for high-dimensional data. The key contribution of this work is to further the understanding of how the shape of regression coefficients for high-dimensional functional data is affected by the degree of regularization. We developed a linearization method to compare regression coefficients with feature coefficients and demonstrated it on two case studies with synthetic responses. We show how the variance of the data dominates regression coefficients obtained with high regularization. The regression solution path contains regression coefficients similar to the linearized true mapping of input data and response (i.e., feature coefficients) for strong regularization. Furthermore, reducing the regularization to the regularization parameter estimated by cross-validation yields regression coefficients that change faster in the input data domain, exploiting the data's local structures to approximate the nonlinear response linearly in high dimensions. While we used synthetic responses for case studies, in reality, there is no ideal feature, and good features that might exist are difficult to discover. Systemically designing features (e.g., Rhyu et al. (2024)) that generalize well and are interpretable is an active area of research. We hope that, ultimately, the ideas in this work can support such efforts.

CRediT authorship contribution statement

Joachim Schaeffer: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jinwook Rhyu:** Writing – review & editing, Writing – original draft, Validation, Investigation. **Robin Droop:** Writing – review & editing, Investigation. **Rolf Findeisen:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Richard D. Braatz:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Dr. Eric Lenz for very helpful feedback.

We acknowledge base funding from the Technical University of Darmstadt, Germany. Furthermore, this work was refined during Joachim Schaeffer's time at the Massachusetts Institute of Technology, for which we acknowledge financial support by a fellowship within the IFI program of the German Academic Exchange Service (DAAD), funded by the Federal Ministry of Education and Research (BMBF), Germany.

Appendix A

A.1. Lithium-ion battery data set

The Lithium Iron Phosphate (LFP) battery data set used in this article contains cycling data of lithium-ion batteries and was initially published with (Severson et al., 2019). Each battery is charged with a fixed charging protocol. Each charging protocol was applied to multiple cells thus the number of unique charging protocols is smaller than the number of batteries. The discharge was constant at 4C (corresponding to a 15-minute discharge) and identical for all cells (Severson et al., 2019). The batteries responded to the stress of the experiment by showing different capacity fade curves. The objective in Severson et al. (2019) was to develop a model that learns signs of battery degradation from early discharge cycles and uses this information to predict the number of cycles until the battery capacity drops below 80% of its nominal capacity (i.e., the cycle life). Severson et al. (2019) found that significant information about cell degradation, and thus a source of characteristics, is the difference between discharge vectors that map the integral of current over time to the voltage of the cell for two distinct cycles. The data matrix used here contains the training (43 cells) and test set (40 cells) $\mathbf{X} \in \mathbb{R}^{83\times1000}$ and has a large number of columns due to the high measurement resolution of current over voltage. More technical details can be found in Severson et al. (2019). Furthermore, there also exists a secondary test set (41 cells) which was published in Severson et al. (2019) but is not used here for simplicity. Similarly shaped high-dimensional data often appear in chemical and biological systems due to measurements over a continuous domain.

A.2. Test set predictions

See Fig. A.1.

A.3. Cross-validation details

We use 10-fold random cross-validation to estimate regression coefficients denoted with *"cv"*. Furthermore, we report regression coefficients based on the one-standard-error rule (see Hastie et al. (2009), Filzmoser et al. (2009) for more details). The regression coefficients based on the one-standard-error rule are denoted with *"cv, 1σ"*. These are the coefficients obtained by using the regularization parameter (number of components for PLS) that yields a mean cross-validation prediction accuracy within one standard error of the prediction accuracies with the minimum mean cross-validation error. As such, the one-standard error rule picks more conservative regularization, which can help with generalization.

A.4. Derivation of feature coefficients for the sum-of-squares feature

For the sum-of-squares case study, the nonlinear response **y** is

$$y_i = \sum_{j=1}^{p} x_{i,j}^2. \tag{12}$$

The feature coefficients are

$$\frac{\partial f(\mathbf{x}_i)}{\partial x_k} = 2x_{i,k}$$
$$\nabla f(\bar{\mathbf{x}}) = 2\bar{\mathbf{x}}$$
$$\boldsymbol{\beta}_{\mathrm{T1}} = m\nabla f(\bar{\mathbf{x}}) = 2\ m\bar{\mathbf{x}}, \tag{13}$$

where $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ is the vector containing column averages of the data matrix. The feature coefficients for the sum-of-squares response are, therefore, the scaled column mean. Deriving the feature coefficients for a large number of features is tedious and error-prone but can be automated via automatic differentiation.

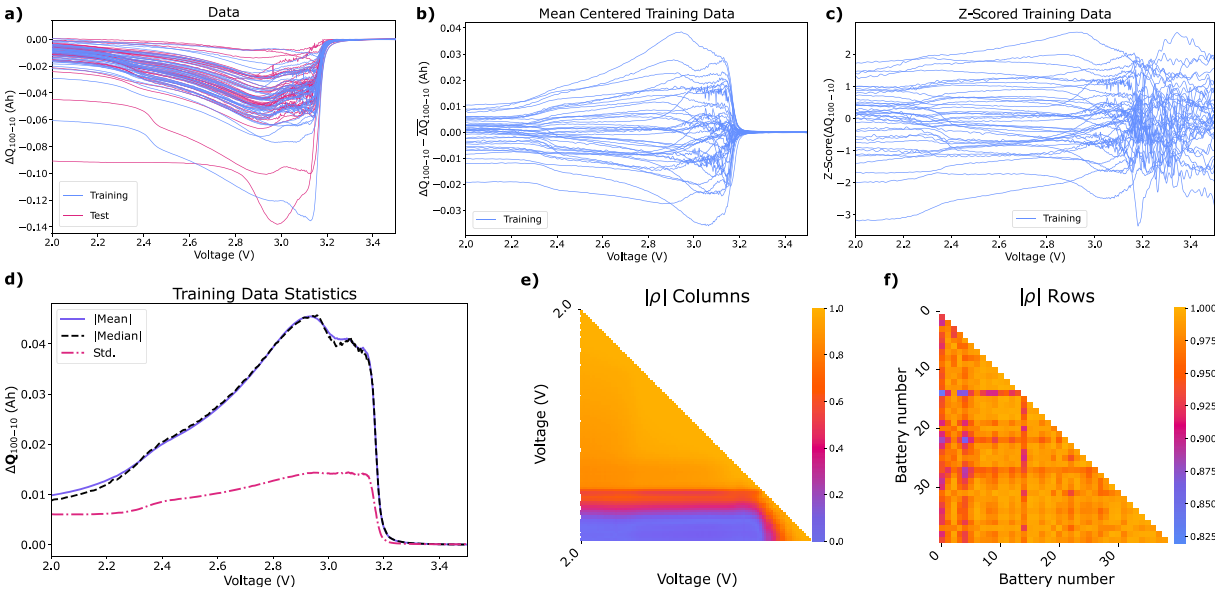

Fig. A.1. Lithium-ion battery data from discharge cycles (Severson et al., 2019). (a) Training, primary test, and secondary test data plotted as curves, (b) mean-centered training data curves with one outlier removed, (c) z-scored training data curves with one outlier removed, (d) training data statistics, (e) Pearson correlation coefficients of training data columns, (f) Pearson correlation coefficients of training data rows.

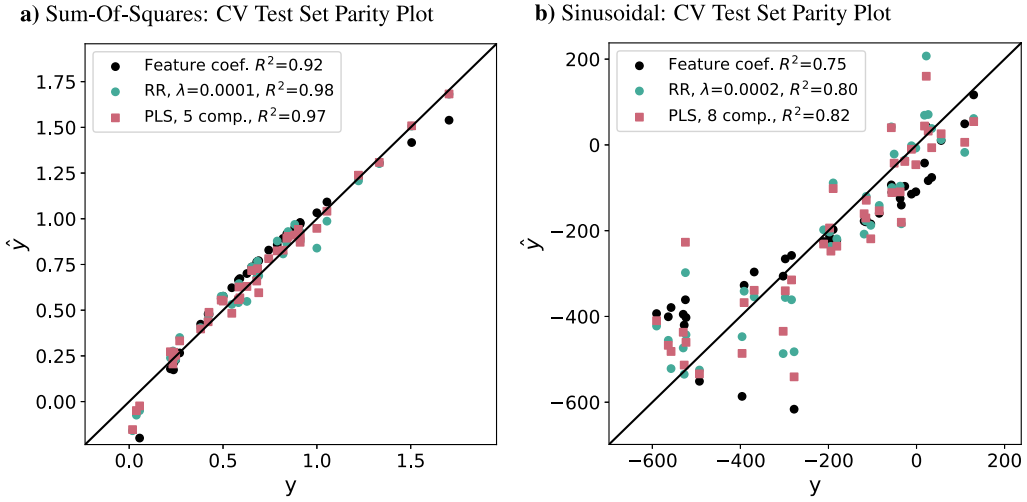

Fig. A.2. Minimum cross-validation test set prediction results for RR and PLS. (a) Case study 1 with sum-of-squares response, corresponding to the minimum cross-validation regression coefficients in Fig. 2bc. (b) Case study 2 with sinusoidal response, corresponding to the minimum cross-validation regression coefficients in Fig. 3c.

A.5. Nullspace note

In high dimensions, the nullspace allows very different-looking regression coefficients to yield similar predictions (Schaeffer et al., 2024b). Therefore, as an alternative to (11),

$$\min_{\lambda} \|\mathbf{X}\boldsymbol{\beta}(\lambda) - \mathbf{X}\boldsymbol{\beta}_{\mathrm{T1}}\|_2^2 \tag{14}$$

could be used to yield regression coefficients with predictions closest to the feature coefficient predictions. However, applying (14) might make it more difficult to compare the resulting coefficients and is therefore not investigated further here.

Data availability

The data that we used is publicly available. For convenience, we recommend the GitHub repository https://github.com/JoachimSchaeffer/HDRegAnalytics as a starting point. The repository contains the data matrix used in this article and further utility functions to help reproduce the results of this work. The full battery data set is available here: https://data.matr.io/1/.

References

Domingos, Pedro, 2012. A few useful things to know about machine learning. Commun. ACM 55 (10), 78–87.
Duistermaat, J.J., 2010. Distributions: Theory and Applications. Birkhäuser, Boston, Massachusetts.
Filzmoser, Peter, Liebmann, Bettina, Varmuza, Kurt, 2009. Repeated double cross validation. J. Chemom. 23 (4), 160–171.
Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome H., Friedman, Jerome H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.
Hörmander, Lars, 2003. The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis, second ed. Springer, Berlin–Heidelberg.
James, Gareth M., Wang, Jing, Zhu, Ji, 2009. Functional linear regression that's interpretable. Ann. Statist. 37 (5A), 2083–2108.
James, Gareth, Witten, Daniela, Hastie, Trevor, Tibshirani, Robert, 2021. An Introduction to Statistical Learning: with Applications in R. Springer, New York.

Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, second ed. Springer, New York, p. 38.

Rhyu, Jinwook, Schaeffer, Joachim, Li, Michael L., Cui, Xiao, Chueh, William C., Bazant, Martin Z., Braatz, Richard D., 2024. Systematic feature design for cycle life prediction of lithium-ion batteries during formation. arXiv preprint arXiv: 2410.07458.

Schaeffer, Joachim, Braatz, Richard D., 2022. Latent variable method demonstrator – Software for understanding multivariate data analytics algorithms. Comput. Chem. Eng. 167, 108014.

Schaeffer, Joachim, Galuppini, Giacomo, Rhyu, Jinwook, Asinger, Patrick A., Droop, Robin, Findeisen, Rolf, Braatz, Richard D., 2024a. Cycle life prediction for lithium-ion batteries: Machine learning and more. In: Proceedings of the American Control Conference. pp. 763–768.

Schaeffer, Joachim, Lenz, Eric, Chueh, William C., Bazant, Martin Z., Findeisen, Rolf, Braatz, Richard D., 2024b. Interpretation of high-dimensional linear regression: Effects of nullspace and regularization demonstrated on battery data. Comput. Chem. Eng. 180, 108471.

Severson, Kristen A., Attia, Peter M., Jin, Norman, Perkins, Nicholas, Jiang, Benben, Yang, Zi, Chen, Michael H., Aykol, Muratahan, Herring, Patrick K., Fraggedakis, Dimitrios, Bazant, Martin Z., Harris, Stephen J., Chueh, William C., Braatz, Richard D., 2019. Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy 4 (5), 383–391.

Strang, Gilbert, 2016. Introduction to Linear Algebra, Fifth ed. Cambridge Press, Wellesley, Massachusetts.

Verdonck, Tim, Baesens, Bart, Óskarsdóttir, María, vanden Broucke, Seppe, 2024. Special issue on feature engineering editorial. Mach. Learn. 113 (7), 3917–3928.

Zheng, Alice, Amanda, Casari, 2018. Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists. O'Reilly, Beijing.