Data-Driven Dimension Reduction for Industrial Load Modeling Using Inverse Optimization

Ruike Lyu, Hongye Guo*, Goran Strbac, Chongqing Kang, *Fellow, IEEE*

Abstract—The intricate mixed-integer constraints in industrial load models not only pose challenges for their direct integration into economic dispatch or market clearing processes but also render current analytical dimension-reduction methods ineffective. We propose a novel data-driven dimension-reduction approach for industrial load modeling, which uses the optimal energy usage data from industrial loads to train a dimension-reduced model that best fits the original constraints. Our approach, implemented by the adjustable load fleet model, outperformed analytical methods across three industrial load datasets.

Index Terms—industrial load, dimension reduction, inverse optimization, data-driven, load modeling

I. INTRODUCTION

To harness industrial users' flexibility through grid-load interactions while mitigating adverse effects, it is crucial to mathematically model the constraints of their production processes [1]. The complex physics of industrial processes poses a key challenge, requiring integer variables or nonlinear functions for representation [2]. Therefore, precise industrial load models encounter obstacles when integrated into economic dispatch or market clearing. To address this issue, we explore dimension reduction techniques to approximate the intricate mixed-integer constraints of industrial loads with simplified lower-dimensional constraints.

Due to the complexity of temporally coupled constraints, exact dimension reduction through projection becomes computationally intractable [3]. Therefore, the common approach is to obtain a computationally feasible result using inner or outer approximation methods [4]. Currently, most approximation-based dimension reduction approaches are analytical methods, which first specify a template geometry (e.g., boxes [5], ellipses [6], or zonotopes [7]) to approximate the original constraints, then determine the constraint parameters through mathematical derivation based on certain assumptions about the original constraints. While these analytical methods with template-based geometries enable efficient computation of Minkowski sums, they face two key limitations [8]: 1) the pre-specified geometric templates lack adaptability across various industrial loads with different constraint structures; 2) they require convexity in the original constraints, making them incompatible with industrial load models with integer variables.

In this paper, we propose a data-driven dimension reduction (D3R) framework that leverages optimal energy consumption
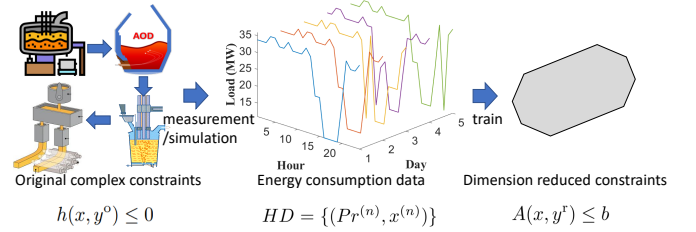
Fig. 1. Proposed data-driven dimension reduction framework for industrial load modeling, illustrated with a steel plant example. h: high-dimensional constraints; HD: historical dataset; A: low-dimensional constraints.

data of industrial loads derived from the original constraints to train a reduced-constraint model through inverse optimization (Fig. 1). D3R exhibits enhanced adaptability and is compatible with constraints involving integer variables.

II. PROBLEM DESCRIPTION

Without loss of generality, our objective is to replace the original constraints of a given industrial load (original constraints, OCs), denoted as $h(x, y^{\mathrm{o}}) \leq 0$, with a set of dimension-reduced linear constraints (reduced constraints, RCs) $A(x, y^{\mathrm{r}}) \leq b$. Here, $h(\cdot)$ is the original constraint function which typically involves complex mixed-integer constraints from Resource-Task Network (RTN) [2] or State-Task Network (STN) models [1], [9] of industrial processes, $x \in \mathbb{R}^T$ represents the hourly energy consumption (continuous), T is the number of time intervals, y^{o} represents other variables of the OCs (including both continuous variables for states/resources and integer variables for operating modes/task scheduling), y^{r} represents other variables of the RCs (continuous), distinguished between the superscripts o (original) and r (reduced), respectively; A and b are the parameter matrix of and the right-hand side of the RCs, respectively.

If we require strict equivalence between the RCs and OCs, then this involves projecting the variable space associated with the OCs onto the RCs, which means the following:

$$\forall x : \exists y^{\mathrm{r}}, A(x, y^{\mathrm{r}}) \leq b \Leftrightarrow \exists y^{\mathrm{o}}, h(x, y^{\mathrm{o}}) \leq 0. \qquad (1)$$

However, strict projection may be computationally infeasible in cases where $h(\cdot)$ has high dimensionality, y^{o} contains integer variables, or T is large. Meanwhile, considering the current market mechanisms for demand-side resources, strict projection may not be necessary. In electricity markets, demand-side resources typically provide services such as demand response, reserves and frequency regulation, where performance requirements are more flexible compared to generation resources. For instance, a 10% to 20% tolerance is generally

acceptable, which motivates us to pursue an approximation-based approach rather than strict equivalence or guaranteeing feasibility that may be overly conservative.

Considering these practical factors, we relax Equ. (1) and introduce a loss function $J(h(\cdot), A, b)$ to quantify the error of the RCs relative to the OCs. Therefore, our objective transforms into determining appropriate A and b values to minimize the approximation error $J(h(\cdot), A, b)$.

Specifically, we do not derive A and b analytically on the basis of the form and parameters of the constraint $h(\cdot)$. Instead, we observe the historical energy usage of industrial users and parameterize (train) A and b in a data-driven manner. In doing so, we assume access to historical data of industrial users $HD = \{(Pr^{(n)}, x^{(n)})\}$, where $Pr^{(n)}$ and $x^{(n)}$ represent the hourly electricity price and the user's hourly electricity consumption on day n, respectively, $n = 1, ..., N$. We also assume that $x^{(n)}$ are optimized on the basis of the OCs to minimize the energy cost $Pr^{(n)\intercal}x^{(n)}$. Note that HD can be realistic or simulated historical data. In practice, historical energy consumption data from industrial users may not originate from optimized decisions. For instance, factories might employ heuristic strategies rather than complex optimization models to schedule their production and energy usage. In such cases, the D3R framework remains applicable but requires an additional step: first, simulating optimal energy usage data under varying electricity prices using the original constraints (OCs), then using these simulated data to train the reduced constraints (RCs).

Under the above assumptions, the approximation error can be expressed as the distance between the optimal energy consumption under historical electricity prices given by the OCs and RCs, and the training problem is formulated as follows:

$$\min_{A,b}.J = \frac{1}{N} \sum_{n=1}^{N} ||x^{(n)} - x_n||^2 \quad (2a)$$

$$\text{s.t.} \quad x_n = \text{argmin}.\{Pr^{(n)\intercal}x_n \ : \ A(x_n, y^{\text{r}}) \leq b\}, \ \forall n \quad (2b)$$

where x_n is the variable for the optimization problem regarding day n based on the RCs to distinguish it from $x^{(n)}$ in the dataset. (2) is a data-driven inverse optimization problem [8], which will be further elaborated in Section III.

III. METHODOLOGY

In principle, D3R allows for the exploration of suitable parameter matrix sizes and element values within the space A/b. However, the high-dimensional search space and the high-dimensional nonlinear constraints of the optimality conditions in (2b) may lead to unaffordable computational costs. To address this challenge, our approach is twofold: first, we selected an appropriate form of the parameter matrix on the basis of the energy consumption characteristics of industrial users (Section III-A), and second, we devised a solution algorithm on the basis of zeroth-order stochastic gradient descent (ZOSGD) [10] to iteratively solve the inverse optimization problem, with a transformation of the optimality conditions to reduce the computational complexity at each iteration (Section III-B).

A. ALF-based Reduced Constraints

We chose an adjustable load fleet (ALF) composed of multiple adjustable loads (ALs) as the form of the RC to be trained. The motivation for choosing ALF is that it can directly model the actual composition and behavior of industrial loads without requiring baseline load definitions or assuming bidirectional energy exchange capabilities as in virtual battery models [4]. The ALF model captures three key physical characteristics of industrial loads through linear constraints:

1) Equipment Composition: Multiple pieces of equipment in industrial facilities are represented by parallel ALs, where the total energy consumption is the sum of individual device consumption.

2) Equipment Power Rating: Each device's power consumption is bounded by its rated capacity.

3) Production Goals: Daily energy consumption limits reflect production targets, as they are typically proportional to energy usage through product energy intensity factors.

The RCs in the form of an ALF are then:

$$x_t = \sum_{i=1}^{I} p_{t,i}\Delta t, \quad \underline{P}_i \leq p_{t,i} \leq \overline{P}_i \ : \ \underline{\mu}^{\text{P}}_{t,i}, \overline{\mu}^{\text{P}}_{t,i}, \ \forall t \quad (3a)$$

$$\underline{E}_i \leq \sum_{t=1}^{T} p_{t,i}\Delta t \leq \overline{E}_i \ : \ \underline{\mu}^{\text{E}}_i, \overline{\mu}^{\text{E}}_i \quad (3b)$$

where the subscript n is omitted, I is the number of ALs as a hyperparameter that balances model complexity and approximation accuracy. A larger I allows RCs to better approximate OCs but increases computational costs, while limited training data may constrain the choice of I to avoid overfitting. x_t is the net energy consumption at time t, which is the summation of $p_{t,i}\Delta t$, representing the average consumption power of AL i at time t multiplied by the length of time interval Δt. $\theta = \{\underline{P}_i, \overline{P}_i, \underline{E}_i, \overline{E}_i\}$ include the power and energy limits of AL i, respectively, which are the parameters to be fitted. The Lagrange multipliers $\underline{\mu}^{\text{P}}_{t,i}, \overline{\mu}^{\text{P}}_{t,i}, \underline{\mu}^{\text{E}}_i, \overline{\mu}^{\text{E}}_i$ are the dual variables associated with the power and energy limits, respectively. The above linear constraints can be easily transformed into the form of $A(x, y^{\text{r}}) \leq b$, where $x = [x_t]$ and $y^{\text{r}} = [p_{t,i}]$.

While this simplified model may overlook some temporal coupling between production processes and relax integer variables, it effectively balances computational complexity and approximation accuracy. Moreover, the D3R framework is not limited to the ALF model - other reduced constraint forms can be chosen based on specific needs while maintaining the framework's core methodology.

B. Training of the ALF model

Substituting the form of the ALF, the optimality conditions (Karush-Kuhn-Tucker conditions) represented by Equation (2b) for day n include primal feasibility (3), stationarity condition derived from the Lagrangian function (4a), dual feasibility (4b), and complementary slackness (4c, 4d):

$$Pr^{\text{E}}_t - \underline{\mu}^{\text{P}}_{t,i} + \overline{\mu}^{\text{P}}_{t,i} - \underline{\mu}^{\text{E}}_i + \overline{\mu}^{\text{E}}_i = 0, \ \forall t, \forall i \quad (4a)$$

$$(\underline{\mu}^{\text{P}}_{t,i}, \overline{\mu}^{\text{P}}_{t,i}, \underline{\mu}^{\text{E}}_i, \overline{\mu}^{\text{E}}_i) \geq 0, \ \forall t, \forall i \quad (4b)$$

$$\underline{\mu}^{\text{P}}_{t,i}(\underline{P}_i - p_{t,i}) = 0, \quad \overline{\mu}^{\text{P}}_{t,i}(p_{t,i} - \overline{P}_i) = 0, \quad \forall t, i \quad (4c)$$

$$\underline{\mu}^{\text{E}}_i(\underline{E}_i - \sum_{t=1}^{T} p_{t,i}) = 0, \quad \overline{\mu}^{\text{E}}_i(\sum_{t=1}^{T} p_{t,i} - \overline{E}_i) = 0, \quad \forall i \quad (4d)$$

The complementary slackness (4c, 4d) conditions involve bilinear constraints, which can be transformed via the Fortuny-Amat transformation [11] into a more easily solvable form. For example, the first term can be replaced by the following:

$$\underline{\mu}^{\mathrm{P}}_{t,i} \leq M(1 - z_{t,i}), \quad p_{t,i} - \underline{P}_i \leq M z_{t,i}, \quad \forall t, i \quad (5)$$

where $z_{t,i}$ is a binary variable and M is a large positive number. The transformed conditions are used in inverse optimization (2), forming a mixed-integer linear programming (MILP) problem that can be solved by commercial solvers.

Since the number of integer variables scales with dataset size N, direct solving becomes computationally intractable. We employ ZOSGD [10] for iterative solving, which is suitable because it: 1) avoids explicit gradient computation, 2) reduces computational burden through batch processing, and 3) maintains tractability via iterative optimization. In each iteration, certain days of data are randomly selected and parameters are updated using estimated gradients:

1: Initialize $j = 0$ and $\theta^{(0)} = \{\underline{P}_i, \overline{P}_i, \underline{E}_i, \overline{E}_i\} = [0]$.

2: Randomly select B days from the dataset and solve the batch problem: $\min_\theta . J = \frac{1}{B} \sum_{n=1}^{B} ||x^{(n)} - x_n||^2$ s.t.$(3) - (4)$ to obtain θ^*.

3: Update $\theta^{(j+1)} = (1 - \alpha)\theta^{(j)} + \alpha\theta^*$; if $\theta^{(j+1)}$ converges, break; otherwise, set $j = j + 1$ and go to step 2.

where α is the learning rate, which can be adaptively adjusted. The ZOSGD algorithm can be easily implemented by using commercial solvers, and the computational complexity is independent of the number of days in the dataset.

IV. NUMERICAL RESULTS

a) Dataset and Task Description: We tested the D3R framework on three datasets from a cement plant [9], a steel powder plant [1], and a steelmaking plant [2]. The first two were originally modeled via the STN, whereas the steelmaking plant was modeled via the RTN. Both are high-dimensional models with integer variables and were used to simulate the optimal energy consumption results for training the ALF.

We used historical hourly electricity market prices from PJM in July 2022 for the simulations. The data from July 1st to 21st contained the training set, whereas the data from July 22nd to 31st were used to test the accuracy of the reduced constraints. The detailed settings and codes can be found in [12]. We employed Gurobi (V11.0.0) with YALMIP in MATLAB to solve the optimization problems on a workstation equipped with an Intel Core i9-10900X CPU (3.7 GHz) and 128 GB of RAM.

For complex industrial load models, there are currently no mature methods for dimension reduction. Therefore, we compared the accuracy of D3R with that of:

1) the simple AL model (SAL), which assigns values to θ based on the maximum and minimum values of the corresponding quantities in the historical load curves. For instance, the maximum power value from historical data is directly taken as the $P^{\max}(\overline{P})$ for the SAL model.

2) the optimal virtual battery model (OVB) [4], an analytical inner-approximation method. For the STN model, we need to relax the decision variables to continuous variables as the OVB

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT METHODS

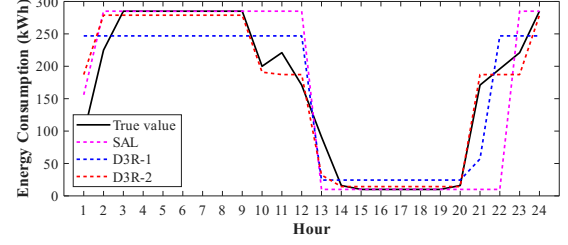| Method	Normalized RMSE (%)			Continuous (Integer) Variables		
	Cement	Steel powder	Steel-making	Cement	Steel powder	Steel-making
Original	-	-	-	196(288)	490(720)	0(10208)
OVB	25.4	36.7	N/A	48(0)	48(0)	N/A
SAL	10.2	23.0	8.3	24(0)	24(0)	24(0)
D3R-1	**8.9**	17.2	**3.6**	24(0)	24(0)	24(0)
D3R-2	9.5	**10.3**	4.2	48(0)	48(0)	48(0)

Fig. 2. Optimal energy consumption results provided by the original model and the reduced constraints under the same electricity prices.
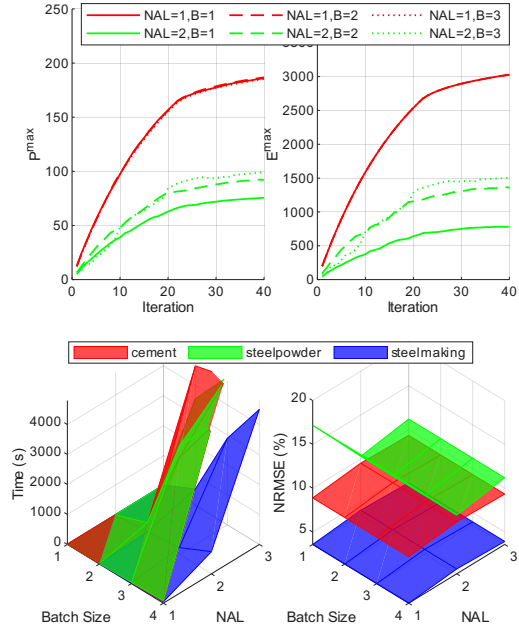

Fig. 3. NAL: number of ALs. B: batch size. Top: Convergence of the ALF parameters during training for the steel powder plant case, exemplified by AL No.1. Bottom: Comparison of computation time and optimal energy consumption errors with different NAL and B.

method only applies to linear constraints. Moreover, it cannot be utilized for steelmaking plants modeled by RTN due to the absence of linear relaxation methods.

b) Results and comparison: The proposed D3R achieved the best performance on all three datasets. This means that the dimensionality reduction constraints obtained through this method yielded the lowest normalized root mean square error (nRMSE) compared with the results based on the original constraints in the test set (Table I). Here, D3R-1(2) represents D3R via the ALF model with 1(2) ALs. The results from OVB based on the analytical inner approximation method were significantly inferior. Fig. 2 visually demonstrates the

effectiveness of D3R, where it reduces the industrial users' complex constraints and variables in the optimization problem with an error ranging from 4% to 10% (Table I). Fig. 3 illustrates the training process and computational efficiency of D3R. In the cement plant case (4 production stages), the OVB method converged after 767 seconds of iteration. However, in the steel powder plant case (10 production stages), the OVB did not converge in 24 hours; therefore, the values of the unconverted states were used.

V. CONCLUSION

Diverging from bottom-up analytical approaches, we propose D3R, a data-driven dimension reduction framework that trains a low-dimensional constraint model to fit the optimal solutions derived from the original constraints. D3R was tested on three industrial load datasets, and the results demonstrated that D3R reduced the high-dimensional constraints without losing much optimality, outperforming the analytical methods in terms of accuracy. D3R provides a new perspective for dimension reduction in industrial load modeling and other tasks with complex constraints to reduce, enhancing adaptability and compatibility with integer variables.

Future research could focus on: (1) developing more realistic forms of dimension-reduced constraints within the D3R framework to capture industrial load characteristics more accurately; (2) addressing how to decompose the overall energy consumption derived from reduced constraints into individual devices.

REFERENCES

[1] R. Lu, R. Bai, Y. Huang, Y. Li, J. Jiang, and Y. Ding, "Data-driven real-time price-based demand response for industrial facilities energy management," *Appl. Energy*, vol. 283, p. 116291, Feb. 2021.

[2] X. Zhang, G. Hug, and I. Harjunkoski, "Cost-Effective Scheduling of Steel Plants With Flexible EAFs," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 239–249, Jan. 2017.

[3] Z. Tan, Z. Yan, H. Zhong, and Q. Xia, "Non-Iterative Solution for Coordinated Optimal Dispatch Via Equivalent Projection—Part I: Theory," *IEEE Trans. Power Syst.*, pp. 1–9, 2023.

[4] Z. Tan, A. Yu, H. Zhong, X. Zhang, Q. Xia, and C. Kang, "Optimal virtual battery model for aggregating storage-like resources with network constraints," *CSEE J. Power Energy Syst.*, vol. 10, no. 4, pp. 1843–1847, 2024.

[5] X. Chen, E. Dall'Anese, C. Zhao, and N. Li, "Aggregate Power Flexibility in Unbalanced Distribution Systems," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 258–269, Jan. 2020.

[6] X. Chen and N. Li, "Leveraging Two-Stage Adaptive Robust Optimization for Power Flexibility Aggregation," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 3954–3965, Sep. 2021.

[7] F. L. Muller, J. Szabo, O. Sundstrom, and J. Lygeros, "Aggregation and Disaggregation of Energetic Flexibility From Distributed Energy Resources," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1205–1214, Mar. 2019.

[8] R. Lyu, H. Guo, and Q. Chen, "Approximating Energy-Regulation Feasible Region of Virtual Power Plants: A Data-driven Inverse Optimization Approach," in *2024 IEEE Power & Energy Society General Meeting*.

[9] H. Golmohamadi, R. Keypour, B. Bak-Jensen, J. R. Pillai, and M. H. Khooban, "Robust Self-Scheduling of Operational Processes for Industrial Demand Response Aggregators," *IEEE Trans. Ind. Electron.*, vol. 67, no. 2, 2020.

[10] S. Liu, P.-Y. Chen, B. Kailkhura, G. Zhang, A. Hero, and P. K. Varshney, "A Primer on Zeroth-Order Optimization in Signal Processing and Machine Learning," Jun. 2020, arXiv:2006.06224.

[11] J. Fortuny-Amat and B. McCarl, "A Representation and Economic Interpretation of a Two-Level Programming Problem," *J. Oper. Res. Soc*, vol. 32, no. 9, pp. 783–792, Sep. 1981.

[12] R. Lyu. [Online]. Available: https://github.com/Rick10119/Data-Driven-Dimension-Reduction