■ Supplementary Material

For the manuscript "Shuai Lu, Jiayi Ding, Wei Gu, et al., On the Solution Uniqueness of Data-Driven Modeling of Flexible Loads".

NUMERICAL TEST

I. SETTINGS OF TESTS

To validate the results, we perform simulations on a hypothetical PFL consisting of a fixed load, a time-decoupled adjustable load, and four batteries. The length of the period is set to 2 for visualization. The detailed parameters and codes are provided at https://github.com/Shuai-Lu/On-the-Solution-Uniqueness-of-Data-Driven-Modeling-of-Flexible-Loads.

The simulation is divided into two steps, including:

Step 1. We generate a random set of electricity price samples for $\lambda \in \mathbb{R}^{N\times T}$. Then, the corresponding aggregate power $P_* \in \mathbb{R}^{N\times T}$ for is calculated by the response model (1). This process generates the dataset $(\lambda^k, P_*^k), \forall k \in \mathrm{K}$, which is required for the identification of the PFL.

Step 2. We select the physical model Ω_{phy} parameterized by θ_{vb}^n as defined in (2), in which the number of the components, i.e., $\mathrm{N}_{vb}/\mathrm{N}_{td}/\mathrm{N}_{fix}$, are prescribed. Then, we substitute the dataset $(\lambda^k, P_*^k), \forall k \in \mathrm{K}$ and the selected physical model Ω_{phy} into the identification model (3), and choose 1-norm as the objective function. Finally, we solve the identification model (3) to obtain the model parameters $\theta_{vb}^n = \left(\underline{P}_{vb}^n, \overline{P}_{vb}^n, \underline{P}_{td}^n, \overline{P}_{td}^n, \underline{E}_{vb}^n, \overline{E}_{vb}^n, E_{vb,0}^n, \sigma^n\right)$. In the simulations, to reduce the computational burden, the value of the parameter σ^n is prescribed.

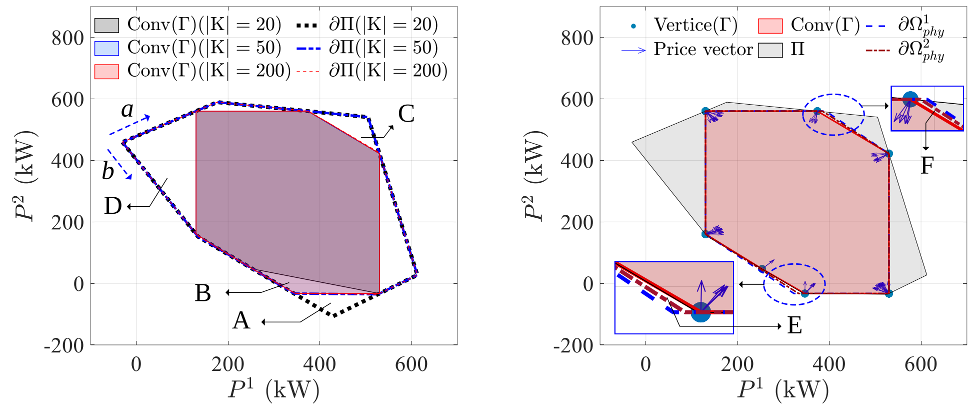
After getting the model parameters θ_{vb}^n, we can get the physical model Ω_{phy} by substituting the parameters. In the following analysis, the aggregated power feasible region of Ω_{phy} is compared with $\mathrm{Conv}(\Gamma)$ and Π to evaluate the accuracy of the identified results.

II. TEST WITH NOISE-FREE DATA

The numerical simulations in this section are based on noise-free data, i.e., we assume that the dataset $(\lambda^k, P_*^k), \forall k \in \mathrm{K}$ are noise-free. The identifications under different dataset sizes are tested.

First, we investigate the impact of the sample size on the sets $\mathrm{Conv}(\Gamma)$ and Π, as given in Fig. 1 (a). Consistent with theoretical results, the sets $\mathrm{Conv}(\Gamma)$ (or Π) expands (or shrinks) as the sample size increases. For example, observing the four regions labeled A, B, C, and D in Fig. 1 (a) and combining with Theorem 2, we can conclude that in the results under 20 samples, A, B, C, and D are all the undetermined regions since they are out of the set $\mathrm{Conv}(\Gamma)$ but bounded by the boundary $\partial\Pi$. When the sample size increases from 20 to 50, the boundary $\partial\Pi$ shrinks so that the region A is not within the boundary $\partial\Pi$, indicating that the region A can be identified as the infeasible region of the PFL. At the same time, as the set $\mathrm{Conv}(\Gamma)$ expands, the region B is contained within the set $\mathrm{Conv}(\Gamma)$, and thus it can be identified as the feasible region of the PFL. In the result under 200 samples, the boundary $\partial\Pi$ shrinks further and accordingly the region C is identified as an infeasible region. Besides, the $\Delta\Omega$ is still nonempty under 200 samples, indicating that the information in the dataset is insufficient. The theoretical results also show that as the sample size increases, the set $\Delta\Omega$ will gradually converge to 0, and finally the exact feasible region can be uniquely determined. Obviously, the simulation results are in agreement with the theoretical ones. Furthermore, the results indicate that when $\Delta\Omega \neq \emptyset$, there always exist some undetermined regions that we cannot ensure whether they are part of the feasible region of the PFL. This inspires us to design specific price vectors to detect if some undetermined region is feasible for the PFL. For example, we can choose any price $\lambda \in \{\lambda | \lambda^\intercal a \geq 0, \lambda^\intercal b \geq 0\}$ to probe the undetermined region D in Fig. 1 (a).

Second, we analyze the identification results using 50 samples based on the physical model Ω_{phy}, in which the numbers of storage are set to 1 and 2 for Ω_{phy}^1 and Ω_{phy}^2, respectively. The feasible region of the aggregated power in Ω_{phy} is given in Fig. 1 (b). Obviously, the physical model Ω_{phy} in (2) is incorrect since it includes the practically infeasible region E that is outside of the set Π. Interestingly, the undetermined region F is also identified, which is not covered by $\mathrm{Conv}(\Gamma)$. This indicates that neither the 50 samples nor the physical model Ω_{phy} includes the information that can determine whether the region F is part of the feasible region of PFL. Besides, it can be seen that Ω_{phy}^2 is closer to $\mathrm{Conv}(\Gamma)$ than Ω_{phy}^1 since both the infeasible region and the undetermined region shrink. This is because the increase in the number of energy storages increases the number of the model parameters, making the selected physical model more refined and flexible. Therefore, we can conclude that as the number of the storages in the Ω_{phy} increases, the identification result is more accurate, which is consistent with the theoretical results.



(a) (b)

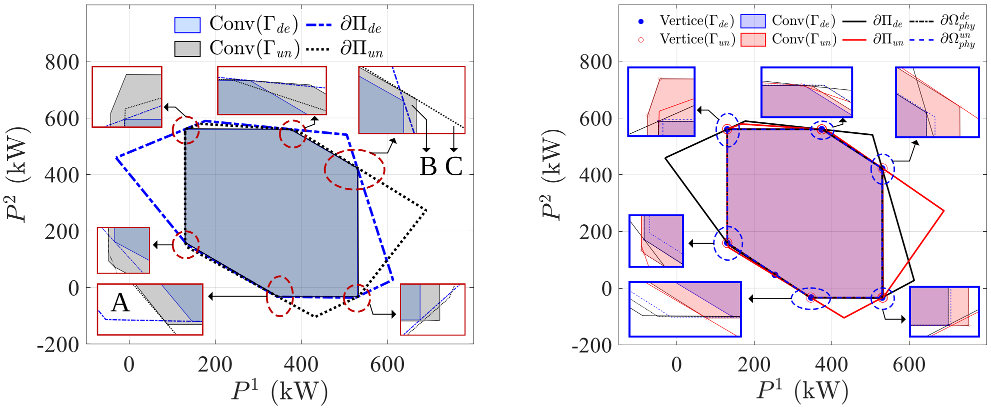
Fig. 1. (a) $\mathrm{Conv}(\Gamma)$ and Π under different sample sizes; (b) Identification results of Ω_{phy} under different number of storage (1 and 2 for Ω_{phy}^1 and Ω_{phy}^2, respectively). (A: Π shrinks as the sample sizes $|\mathrm{K}|$ increases from 20 to 50; B: $\mathrm{Conv}(\Gamma)$ expands as $|\mathrm{K}|$ increases from 20 to 50; C: Π shrinks as the $|\mathrm{K}|$ increases from 50 to 200; D: Undetermined region; E: Practically infeasible region; F: Undetermined region in Ω_{phy}.)

The above results have some practical significance for real-world applications. First, after collecting the data, we should test if they contain sufficient information about the operational characteristics of the PFL by analyzing the set $\Delta\Omega$. Second, after obtaining the identification result of the PFL, we should check if the physical model Ω_{phy} is appropriately selected by inspecting if the practically infeasible region is included in the identification result.

III. Test With Noisy Data

To provide a deeper understanding of the robustness and applicability of the model under noisy data, we perform the simulations using the datasets with noise. According to the requirements of China's national standard for the accuracy of electric meters, we assume that the error of the measured electric power is ±0.5% of the true value. We assume that the electricity price has a relative error of ±10%. Based on these settings, we use Gaussian white noise in the simulations to model the errors. Specifically, for the aggregate power P_*^k, we assume that the relative error ϵ_P obeys a Gaussian distribution with a mean of 0 and a standard deviation of σ_P, i.e., $\epsilon_P \sim N(0, \sigma_P^2)$. We adopt the 3σ criterion, i.e., the probability that the noise value is distributed in $(-3\sigma_P, +3\sigma_P)$ is 99.74%, and hence σ_P is set to (0.5/3)%. For the electricity price λ^k, we assume that the relative error ϵ_λ obeys a Gaussian distribution with a mean of 0 and a standard deviation of σ_λ, i.e., $\epsilon_\lambda \sim N(0, \sigma_\lambda^2)$, in which σ_λ is set to (10/3)% based on the 3σ criterion. Based on the above settings, the original noise-free dataset $(\lambda^k, P_*^k), \forall k \in \mathrm{K}$ turns to the new noisy dataset, denoted as $(\tilde{\lambda}^k, \tilde{P}_*^k), \forall k \in \mathrm{K}$. Note that since the aggregated power is calculated based on the response model (1), we first obtain $\tilde{\lambda}^k = (1+\epsilon_\lambda)\lambda^k$, and then calculate the aggregated power $P_*^k(\tilde{\lambda}^k)$ using the model (1) under the noisy electricity price $\tilde{\lambda}^k, \forall k \in \mathrm{K}$, and then obtain the noisy aggregated power $\tilde{P}_*^k$ by adding the noise to $P_*^k(\tilde{\lambda}^k)$, i.e., $\tilde{P}_*^k = (1+\epsilon_P)P_*^k(\tilde{\lambda}^k)$. We use 50 samples that are the same as those in the noise-free case in the following test.

First, we investigate the impact of the noise on the set $\mathrm{Conv}(\Gamma)$, as given in Fig. 2 (a). The results under noise-free and noisy data are subscripted with de and un, respectively. It can be observed that there are some minor differences between the sets $\mathrm{Conv}(\Gamma_{de})$ and $\mathrm{Conv}(\Gamma_{un})$. Interestingly, the set $\mathrm{Conv}(\Gamma_{un})$ completely covers the set $\mathrm{Conv}(\Gamma_{de})$. The reason is given as follows: (1) The response of the PFL, as shown in the model (1), depends on the direction of the electricity price $\tilde{\lambda}^k$, and the noise ϵ_λ in this case does not significantly change the direction of the electricity price λ^k; (2) Hence, the aggregated power $P_*^k(\tilde{\lambda}^k)$ calculated under the noisy electricity price $\tilde{\lambda}^k$ almost overlapping with $P_*^k(\lambda^k)$; (3) After adding the noise to $P_*^k(\tilde{\lambda}^k)$, the points on the boundary of $\mathrm{Conv}(\Gamma_{de})$ have 50% probability of falling outside the boundary $\partial\mathrm{Conv}(\Gamma_{de})$ since the errors are symmetric, making the set $\mathrm{Conv}(\Gamma_{un})$ with $\tilde{P}_*^k$ as the extreme points cover the set $\mathrm{Conv}(\Gamma_{de})$. This result reveals that in practical use, it is crucial to accurately predict the relative magnitude of electricity prices in different time periods. Besides, under noisy data, the region B is included in the set $\mathrm{Conv}(\Gamma_{un})$, which is easily mistaken as part of the feasible region of the PFL.



(a) (b)

Fig. 2. (a) $\mathrm{Conv}(\Gamma)$ and Π under noise-free data and noisy data respectively; (b) Identification results of Ω_{phy} under noise-free data and noisy data (noise-free data for Ω_{phy}^{de} and noisy data for Ω_{phy}^{un}). (A: Undetermined region under noise-free data but infeasible region under noisy data; B: Infeasible region under noise-free data but feasible region under noisy data; C: Infeasible region under noise-free data but undetermined region under noisy data.)

Second, we investigate the impact of the noise on sets Π. It is easy to notice that the shape of Π has changed considerably by the noise. The reason is that the direction of the boundary $\partial\Pi$, as shown in Fig. 2 (a), is determined by the direction of the electricity price, which is changed by the noise. Note that the change in the shape of Π is not essential, and what we are concerned about is the operational characteristics of the PFL. In the results under noise-free data, the region A is an undetermined region since it is out of the set $\mathrm{Conv}(\Gamma_{de})$ but bounded by the boundary $\partial\Pi_{de}$. Nevertheless, it is mistakenly identified as the infeasible region of the PFL under noisy data since it is not within the boundary $\partial\Pi_{un}$. Moreover, the region B and the region C are identified as infeasible regions since they are not within the boundary $\partial\Pi_{de}$. Besides, it also easy to mistake the region C as an undetermined region only based on the evidence that it is bounded by $\partial\Pi_{un}$. In summary, we can conclude that, under noisy data, if we use the results in Fig. 2 (a) to prejudge the operational characteristics of the PFL as what we do under the noise-free dataset, we will get some wrong results, including misidentifying the regions that are actually feasible as infeasible regions, and mistakenly identifying the actually infeasible areas as feasible or undetermined regions.

Third, we analyze the identification results. The feasible regions of the aggregated power in Ω_{phy}^{de} and Ω_{phy}^{un} are given in Fig. 2 (b). In addition to the similar conclusions that we have previously obtained, we also find that there is very little difference between Ω_{phy}^{de} and Ω_{phy}^{un} either in shape or in the size of the regions. This means that a certain range of noise on the dataset has little effect on the identification results, although both the set $\mathrm{Conv}(\Gamma)$ and the set Π are considerably affected by noise. That is, the physical model identified using the actual noisy data is very close to the one identified using the noise-free data. A potential reason is that the prior physical model, Ω_{phy}^{de} or Ω_{phy}^{un}, endow the PFL model with a specific structural characteristic, making it less sensitive to the noise in the dataset. This result indicates that in practical use, embedding the prior physical knowledge into the data-driven modeling of the PFL greatly improves the robustness against the noise.

IV. CONCLUSION

The above results validate the effectiveness of theoretical results. Some conclusions drawn from the simulation results can be summarized as follows.

(1) When $\Delta\Omega \neq \emptyset$, there always exist some undetermined regions, which we cannot use data alone to ensure whether they are part of the feasible region of the PFL.

(2) The identification results under noisy data could misidentify the regions that are actually feasible as infeasible regions, or mistakenly identify the actually infeasible areas as feasible or undetermined regions.

(3) If a priori physical model Ω_{phy} is appropriately selected, allowing a certain range of noise on the dataset has little effect on the identification results, although both the set $\mathrm{Conv}(\Gamma)$ and the set Π are affected by noise.

Based on the simulation results, we can also get some practical implications for the data-driven modeling of the PFL in the real world, as follows.

(1) After collecting the data, we should test if they contain sufficient information about the operational characteristics of the PFL by analyzing the set $\Delta\Omega$. The information of the undetermined region can be used to design the probing price to eliminate the information gap between $\mathrm{Conv}(\Gamma)$ and Π, and thus ensure the practical identifiability of the PFL.

(2) Selecting an appropriate physical model Ω_{phy} will help identify the PFL model under incomplete and noisy information, while an incorrect Ω_{phy} produces wrong results. After obtaining the identification result of the PFL, we should check if the physical model Ω_{phy} is appropriately selected by inspecting if the practically infeasible region is included in the identification result.

(3) It is crucial to accurately predict the relative magnitude of electricity prices in different time periods because the response of the PFL depends on the direction of the electricity price vector.