

## ■ Supplementary Material

For the manuscript “Shuai Lu, Jiayi Ding, Mingji Chen, et al., On the Solution Uniqueness of Data-Driven Modeling of Flexible Loads”.

This material, as a supplement to the manuscript “On the Solution Uniqueness of Data-Driven Modeling of Flexible Loads”, clarifies some fundamental problems and gives extended simulation results. Appendix A provides further explanations for *Remark 3. (4)*, showing how checking if  $\Pi \subseteq \text{Conv}(\Gamma_*)$  or not can be reformulated into (mixed-integer) linear programming problems. Appendix B provides the proof for *Theorem 3*. Appendix C introduces the solution method for the PFL identification model. Appendix D explains how to design price signals to probe the undetermined regions to handle the information insufficiency problem. Finally, Appendix E introduces the simulation settings and gives detailed simulation results.

### Appendix A

#### CHECKING IF $\Pi \subseteq \text{Conv}(\Gamma_*)$ OR NOT

In *Remark 3. (4)* in the manuscript, we discuss how to judge if  $\Delta\Omega = \emptyset$ , i.e., to check whether  $\Pi \subseteq \text{Conv}(\Gamma_*)$ . As we analyzed in *Remark 3*, whether  $\Delta\Omega = \emptyset$  holds depends on the information sufficiency of the dataset  $(\lambda^k, P_*^k), k \in K$ , which is the endogenous property of the dataset. We said in the original manuscript that checking if  $\Delta\Omega = \emptyset$  can be formulated into equivalent (mixed-integer) linear programming problems. Next, we will give more details about this problem.

First, let us recall the definition of  $\Pi$  and  $\Gamma_*$ :  $\Pi \triangleq \{P | \Lambda P \geq \Xi\}$ ,  $\Gamma_* \triangleq [P_*^1 \cdots P_*^K]$ , wherein  $\Lambda \triangleq [\lambda^1 \cdots \lambda^K]^\top$  and  $\Xi \triangleq [(\lambda^1)^\top P_*^1 \cdots (\lambda^K)^\top P_*^K]^\top$ . According to this definition, after we collect the dataset (i.e., the price-power pairs)  $(\lambda^k, P_*^k), k \in K$ , we can calculate the parameter matrix  $\Lambda$  and parameter vector  $\Xi$ . To check whether  $\Pi \subseteq \text{Conv}(\Gamma_*)$  holds is equivalent to checking if  $\forall P \in \Pi, P \in \text{Conv}(\Gamma_*)$  holds.

Note that the set  $\text{Conv}(\Gamma_*)$  can be reformulated into an equivalent linear form by introducing auxiliary variables  $\alpha = [\alpha^1 \cdots \alpha^K]^\top \in \mathbb{R}^K$  as:

$$\text{Conv}(\Gamma_*) = \{P | P = \Gamma_* \alpha, 1^\top \alpha = 1, 0 \leq \alpha \leq 1\}. \quad (\text{A.1})$$

Then, to check if  $\forall P \in \Pi, P \in \text{Conv}(\Gamma_*)$  holds is equivalent to checking if  $\forall P \in \Pi, \exists 0 \leq \alpha \leq 1, 1^\top \alpha = 1$  making  $P = \Gamma_* \alpha$ . For this problem, we can use two methods: vertex enumeration and robust optimization. The details are given as follows.

#### (1) The vertex enumeration method

In this method, we need to check if each vertex of the set  $\Pi$  is in the set  $\text{Conv}(\Gamma_*)$ . The corresponding mathematical model is

$$\begin{aligned} & \forall P^j \in \text{Vertex}(\Pi) = \{P^1, P^2, \dots, P^J\}: \\ & \text{Solve feasibility-checking problem:} \\ & \min_{P^j} \quad 0 \\ & \text{s.t.} \quad P^j = \Gamma_* \alpha \\ & \quad \quad 1^\top \alpha = 1 \\ & \quad \quad 0 \leq \alpha \leq 1 \end{aligned}, \quad (\text{A.2})$$

wherein  $\text{Vertex}(\Pi)$  denotes the vertex set of the set  $\Pi$ .

If for each  $P^j$ , the above optimization model is always feasible, we can conclude that  $\Pi \subseteq \text{Conv}(\Gamma_*)$ . For this method, we need to calculate the vertices of the set  $\Pi$ , and then check that each vertex is in the set  $\text{Conv}(\Gamma_*)$  by solving the feasibility checking problem (linear optimization problem) (A.2). If one of the feasibility-checking problems is infeasible, we can conclude that the corresponding vertex is not in the set  $\text{Conv}(\Gamma_*)$ , i.e.,  $\Pi \not\subseteq$

$\text{Conv}(\Gamma_*)$ .

Furthermore, we would like to clarify that this method is computationally expensive if the dimension of the set  $\Pi$  is very high. Hence, it is unsuitable for the problem where the dimension of the set  $\Pi$  is very high.

(2) *The robust optimization method*

Using the robust optimization method, the feasibility-checking problem can be formulated as a bilevel problem, as:

$$\max_{P \in \Pi} \min_{P_{\#} \in \text{Conv}(\Gamma_*)} \|P - P_{\#}\|_2^2. \quad (\text{A.3})$$

If the optimal objective of this model is 0, it means that  $\forall P \in \Pi, P_{\#} \in \text{Conv}(\Gamma_*)$  holds. This model can be further reformulated as:

$$\begin{aligned} \max_{P \in \Pi} \min_{\alpha} \quad & \|P - \Gamma_* \alpha\|_2^2 \\ \text{s.t.} \quad & \mathbf{1}^\top \alpha = 1: \quad \mu \\ & 0 \leq \alpha \leq 1: \quad v_1, v_2 \end{aligned} \quad (\text{A.4})$$

The above model can be equivalently transformed into a mixed-integer linear programming model using the KKT conditions and big-M method, as:

$$\begin{aligned} \max_{P \in \Pi, \alpha, \mu, v_1, v_2, \varepsilon_1, \varepsilon_2} \quad & 0 \\ \text{s.t.} \quad & \mathbf{1}^\top \alpha = 1 \\ & 0 \leq \alpha \leq 1 \\ & 2\Gamma_*^\top (\Gamma_* \alpha - P) + \mu \cdot \mathbf{1} + v_1 - v_2 = 0 \\ & 0 \leq \alpha \leq \varepsilon_1 M \\ & -(1 - \varepsilon_1)M \leq v_1 \leq 0 \\ & 0 \leq 1 - \alpha \leq \varepsilon_2 M \\ & -(1 - \varepsilon_2)M \leq v_2 \leq 0 \\ & \varepsilon_1, \varepsilon_2 \in \{0, 1\} \end{aligned} \quad (\text{A.5})$$

Based on the above, we show that checking if  $\Delta\Omega = \emptyset$  can be formulated into equivalent (mixed-integer) linear programming problems.

## Appendix B

### PROOF FOR THEOREM 3

**Theorem 3.** For the model (4), under *Assumption 1* and the  $l^2$ -norm objective function  $f(\theta)$ , supposing that  $\forall k \in K, e_P^k \sim \mathcal{N}(0_T, \Sigma_P)$ , wherein  $\Sigma_P \in \mathbb{R}^{T \times T}$  is the covariance matrix, we have:

- (a)  $\lim_{|K| \rightarrow +\infty} (\hat{\theta}_{noise} - \hat{\theta}_{nf}) = 0$ ;
- (b)  $\left( f(\hat{\theta}_{noise}) - f(\hat{\theta}_{nf}) \right)_K \xrightarrow{\text{Probability}} \text{tr}(\Sigma_P)$ ;

wherein  $\hat{\theta}_{noise}$  and  $\hat{\theta}_{nf}$  denote the solutions of model (4) under noisy  $\tilde{P}_*^k$  and noise-free  $P_*^k, \forall k \in K$ , respectively, and  $\text{tr}(\Sigma_P)$  denotes the trace of the matrix  $\Sigma_P$ .

**Proof.** Recall that the model (4) in the manuscript is as follows:

$$\begin{aligned} \hat{\theta} = \arg \min_{\theta} f(\theta) = \frac{1}{|K|} \sum_{k \in K} \|\tilde{P}_*^k - \hat{P}_*^k(\lambda^k, \Omega)\|, \quad & ((4) \text{ in the manuscript}) \\ \text{s.t.} \quad & \hat{P}_*^k(\lambda^k, \Omega) = \arg \min_{P \in \Omega(\theta)} (\lambda^k)^\top P, \quad \forall k \in K \end{aligned}$$

wherein the superscript  $k$  denotes the sample index,  $K$  denotes the sample set;  $\hat{\theta}$  is the optimal estimated value

of the model parameter  $\theta$ ;  $\hat{P}_*^k$  is the estimated aggregate power; and  $\tilde{P}_*^k \in \mathbb{R}^T$  is the measurement of the aggregated power  $P_*^k$ , which can be modeled as  $\tilde{P}_*^k = P_*^k + e_p^k$ , wherein  $e_p^k \in \mathbb{R}^T$  denotes the measurement errors of  $P_*^k$ .

We assume that  $\forall k \in K$ ,  $e_p^k \sim \mathcal{N}(\mu_p, \Sigma_p)$ . Under *Assumption 1*, if the  $l^2$ -norm is used for  $f(\theta)$ , the model (4) can be rewritten as:

$$\begin{aligned} \hat{\theta}_{noise} = \arg \min_{\theta} f(\theta) &= \frac{1}{|K|} \sum_{k \in K} \left\| \tilde{P}_*^k - \hat{P}_*^k(\lambda^k, \Omega) \right\|_2^2 \\ \text{s.t. } \hat{P}_*^k(\lambda^k, \Omega) &= \underset{P \in \Omega(\theta)}{\operatorname{argmin}} (\lambda^k)^\top P, \quad \forall k \in K \end{aligned} \quad (\text{B.1})$$

Now, let us calculate the expression of  $f(\theta)$ , as:

$$\begin{aligned} f(\theta) &= \frac{1}{|K|} \sum_{k \in K} \left\| \tilde{P}_*^k - \hat{P}_*^k \right\|_2^2 \\ &= \frac{1}{|K|} \sum_{k \in K} \left\| P_*^k + e_p^k - \hat{P}_*^k \right\|_2^2 \\ &= \frac{1}{|K|} \sum_{k \in K} \left\| P_*^k - \hat{P}_*^k \right\|_2^2 + \frac{1}{|K|} \sum_{k \in K} \|e_p^k\|_2^2 + \frac{1}{|K|} \sum_{k \in K} 2(P_*^k - \hat{P}_*^k)^\top e_p^k \\ &= \frac{1}{|K|} \sum_{k \in K} \left\| P_*^k - \hat{P}_*^k \right\|_2^2 + \frac{1}{|K|} \sum_{k \in K} 2(P_*^k - \hat{P}_*^k)^\top \mu_p + \frac{1}{|K|} \sum_{k \in K} (\mu_p)^\top \mu_p \\ &\quad + \frac{1}{|K|} \sum_{k \in K} 2(P_*^k - \hat{P}_*^k)^\top (e_p^k - \mu_p) + \frac{1}{|K|} \sum_{k \in K} \|e_p^k\|_2^2 - \frac{1}{|K|} \sum_{k \in K} (\mu_p)^\top \mu_p \\ &= \frac{1}{|K|} \sum_{k \in K} \left\| P_*^k + \mu_p - \hat{P}_*^k \right\|_2^2 \\ &\quad + \frac{2}{|K|} \sum_{k \in K} (P_*^k - \hat{P}_*^k)^\top (e_p^k - \mu_p) + \frac{1}{|K|} \sum_{k \in K} \|e_p^k\|_2^2 - (\mu_p)^\top \mu_p \end{aligned} \quad (\text{B.2})$$

First, we analyze the term  $\frac{2}{|K|} \sum_{k \in K} (P_*^k - \hat{P}_*^k)^\top (e_p^k - \mu_p)$ . Note that the term  $P_*^k - \hat{P}_*^k(\lambda^k, \Omega)$  is independent of the random error  $e_p^k$ , and  $P_*^k - \hat{P}_*^k(\lambda^k, \Omega), \forall k \in K$  can also be regarded as independent of each other. Since  $\forall k \in K, e_p^k \sim \mathcal{N}(\mu_p, \Sigma_p)$ , we have  $e_p^k - \mu_p \sim \mathcal{N}(0_T, \Sigma_p)$ . Then, we have  $(P_*^k - \hat{P}_*^k)^\top (e_p^k - \mu_p) \sim \mathcal{N}(0_T, (P_*^k - \hat{P}_*^k)^\top \Sigma_p (P_*^k - \hat{P}_*^k))$ . Thus, we have:

$$\begin{aligned} \frac{2}{|K|} \sum_{k \in K} (P_*^k - \hat{P}_*^k)^\top (e_p^k - \mu_p) &\sim \mathcal{N}(\mu_a, \sigma_a^2) \\ \left( \mu_a = 0, \sigma_a^2 = \frac{4}{|K|^2} \sum_{k \in K} (P_*^k - \hat{P}_*^k)^\top \Sigma_p (P_*^k - \hat{P}_*^k) \right) \end{aligned} \quad (\text{B.3})$$

Without loss of generality, the vector  $P_*^k - \hat{P}_*^k$  can be considered as bounded, i.e., there exists  $M_a \in \mathbb{R}$  satisfying  $|P_*^k - \hat{P}_*^k| \leq M_a \cdot 1_T, \forall k \in K$ . Hence, we have:

$$\sigma_a^2 = \frac{4}{|K|^2} \sum_{k \in K} (P_*^k - \hat{P}_*^k)^\top \Sigma_p (P_*^k - \hat{P}_*^k) \leq \frac{4}{|K|^2} \sum_{k \in K} M_a^2 (1_T)^\top \Sigma_p 1_T = \frac{4}{|K|} M_a^2 (1_T)^\top \Sigma_p 1_T. \quad (\text{B.4})$$

Taking the limit of  $\sigma_a^2$ , we have:

$$\lim_{|K| \rightarrow +\infty} \sigma_a^2 \leq \lim_{|K| \rightarrow +\infty} \frac{4}{|K|} M_a^2 (1_T)^\top \Sigma_p 1_T = 0. \quad (\text{B.5})$$

Hence, from Chebyshev's inequality, we have:

$$\lim_{|K| \rightarrow +\infty} P \left( \left| \frac{2}{|K|} \sum_{k \in K} (P_*^k - \hat{P}_*^k)^\top (e_p^k - \mu_p) \right| > \varepsilon \right) = 0, \quad \forall \varepsilon > 0. \quad (\text{B.6})$$

Second, we analyze the term  $\frac{1}{|\mathbf{K}|} \sum_{k \in \mathbf{K}} \|e_p^k\|_2^2$ . Since  $e_p^k \sim \mathcal{N}(\mu_p, \Sigma_p)$ , we have that  $\forall k \in \mathbf{K}$ ,  $\|e_p^k\|_2^2$  follows the identical independent distribution. Then, based on the Central Limit Theorem, we have:

$$\begin{aligned} \frac{1}{|\mathbf{K}|} \sum_{k \in \mathbf{K}} \|e_p^k\|_2^2 &\xrightarrow{\text{Distribution}} \mathcal{N}(\mu_b, \sigma_b^2) \\ \left( \mu_b = \text{tr}(\Sigma_p) + (\mu_p)^\top \mu_p, \sigma_b^2 = \frac{1}{|\mathbf{K}|} \text{var}(\|e_p^k\|_2^2) \right) \end{aligned} \quad (\text{B.7})$$

Note that

$$\begin{aligned} \text{var}(\|e_p^k\|_2^2) &= \mathbb{E}(\|e_p^k\|^4) - \left[ \mathbb{E}(\|e_p^k\|^2) \right]^2 \\ &\leq \mathbb{E}(\|e_p^k\|^4) \\ &\stackrel{\text{Cauchy inequality}}{\leq} \mathbb{E} \left( T \sum_{i \in \mathbf{T}} \left( (e_p^k)_i \right)^4 \right) \\ &\leq T \sum_{i \in \mathbf{T}} \left[ 3((\Sigma_p)_{ii})^4 + 6((\Sigma_p)_{ii})^2 ((\mu_p)_i)^2 + ((\mu_p)_i)^4 \right] \end{aligned} \quad (\text{B.8})$$

We introduce a constant  $M_b < +\infty$ , defined as:

$$M_b = T \sum_{i \in \mathbf{T}} \left[ 3((\Sigma_p)_{ii})^4 + 6((\Sigma_p)_{ii})^2 ((\mu_p)_i)^2 + ((\mu_p)_i)^4 \right] \quad (\text{B.9})$$

Then, we have:

$$\sigma_b^2 \leq \frac{1}{|\mathbf{K}|} M_b \quad \text{and} \quad \lim_{|\mathbf{K}| \rightarrow +\infty} \sigma_b^2 \leq \lim_{|\mathbf{K}| \rightarrow +\infty} \frac{1}{|\mathbf{K}|} M_b = 0 \quad (\text{B.10})$$

Hence, from Chebyshev's inequality, we have:

$$\lim_{|\mathbf{K}| \rightarrow +\infty} P \left( \left| \frac{1}{|\mathbf{K}|} \sum_{k \in \mathbf{K}} \|e_p^k\|_2^2 - \mu_b \right| > \varepsilon \right) = 0, \forall \varepsilon > 0. \quad (\text{B.11})$$

Now, we define a random variable  $\xi \in \mathbb{R}$ , as:

$$\xi = \frac{2}{|\mathbf{K}|} \sum_{k \in \mathbf{K}} (P_*^k - \hat{P}_*^k)^\top (e_p^k - \mu_p) + \frac{1}{|\mathbf{K}|} \sum_{k \in \mathbf{K}} \|e_p^k\|_2^2 - (\mu_p)^\top \mu_p - \text{tr}(\Sigma_p). \quad (\text{B.12})$$

Then, when  $\mu_p = 0_T$ , based on (B.2) and (B.12), the model (4) can be rewritten as:

$$\begin{aligned} \hat{\theta}_{noise} &= \arg \min_{\theta} f(\theta) = \frac{1}{|\mathbf{K}|} \sum_{k \in \mathbf{K}} \|P_*^k - \hat{P}_*^k(\lambda^k, \Omega)\|_2^2 + \text{tr}(\Sigma_p) + \xi \\ \text{s.t.} \quad \hat{P}_*^k(\lambda^k, \Omega) &= \arg \min_{P \in \Omega(\theta)} (\lambda^k)^\top P, \quad \forall k \in \mathbf{K} \end{aligned} \quad (\text{B.13})$$

Compared with the case wherein *Assumption 2* holds, there are two extra terms in the objective function, including the constant  $\text{tr}(\Sigma_p)$  and the random variable  $\xi$ .

By substituting (B.6) and (B.11) into (B.12), we have:

$$\xi \xrightarrow{\text{Distribution}} \mathcal{N}(\mu_\xi, \sigma_\xi^2), \quad (\text{B.14})$$

wherein the mean and variance can be calculated as:

$$\begin{cases} \mu_\xi = \mu_a + \mu_b - \text{tr}(\Sigma_p) - (\mu_p)^\top \mu_p = 0, \\ \sigma_\xi^2 = \sigma_a^2 + \sigma_b^2 = \frac{4}{|\mathbf{K}|^2} \sum_{k \in \mathbf{K}} (P_*^k - \hat{P}_*^k)^\top \Sigma_p (P_*^k - \hat{P}_*^k) + \frac{1}{|\mathbf{K}|} \text{var}(\|e_p^k\|_2^2) \end{cases} \quad (\text{B.15})$$

Based on this, we have:

$$\lim_{|K| \rightarrow +\infty} \mu_\xi = 0, \lim_{|K| \rightarrow +\infty} \sigma_\xi^2 = 0, \quad (\text{B.16})$$

Hence, we have:

$$\lim_{|K| \rightarrow +\infty} P(|\xi| > \varepsilon) = 0, \forall \varepsilon > 0, \quad (\text{B.17})$$

Therefore, as  $|K| \rightarrow +\infty$ , the random variable  $\xi$  will not affect the solution of the model (B.13), i.e.,

$$\lim_{|K| \rightarrow +\infty} \hat{\theta}_{noise} = \lim_{|K| \rightarrow +\infty} \hat{\theta}_{nf}. \quad (\text{B.18})$$

This proves *Theorem 3. (a)*.

Further, note that:

$$f(\hat{\theta}_{nf}) = \frac{1}{|K|} \sum_{k \in K} \left\| P_*^k - \hat{P}_*^k(\lambda^k, \Omega(\hat{\theta}_{nf})) \right\|_2^2, \quad (\text{B.19})$$

and

$$\begin{aligned} f(\hat{\theta}_{noise}) - f(\hat{\theta}_{nf}) &= \frac{1}{|K|} \sum_{k \in K} \left\| P_*^k - \hat{P}_*^k(\lambda^k, \Omega(\hat{\theta}_{noise})) \right\|_2^2 + \text{tr}(\Sigma_P) + \xi - f(\hat{\theta}_{nf}) \\ &= \frac{1}{|K|} \sum_{k \in K} \left\| P_*^k - \hat{P}_*^k(\lambda^k, \Omega(\hat{\theta}_{noise})) \right\|_2^2 + \text{tr}(\Sigma_P) + \xi - \frac{1}{|K|} \sum_{k \in K} \left\| P_*^k - \hat{P}_*^k(\lambda^k, \Omega(\hat{\theta}_{nf})) \right\|_2^2. \end{aligned} \quad (\text{B.20})$$

Based on (B.18), we have:

$$\begin{aligned} &\lim_{|K| \rightarrow +\infty} P\left(\left|f(\hat{\theta}_{noise}) - f(\hat{\theta}_{nf}) - \text{tr}(\Sigma_P)\right| > \varepsilon\right) \\ &= \lim_{|K| \rightarrow +\infty} P\left(\left|\frac{1}{|K|} \sum_{k \in K} \left\| P_*^k - \hat{P}_*^k(\lambda^k, \Omega(\hat{\theta}_{noise})) \right\|_2^2 + \text{tr}(\Sigma_P) + \xi - f(\hat{\theta}_{nf}) - \text{tr}(\Sigma_P)\right| > \varepsilon\right) \\ &= \lim_{|K| \rightarrow +\infty} P(|\xi| > \varepsilon) \end{aligned} \quad (\text{B.21})$$

Combining (B.17) and (B.21), we have:

$$\lim_{|K| \rightarrow +\infty} P\left(\left|f(\hat{\theta}_{noise}) - f(\hat{\theta}_{nf}) - \text{tr}(\Sigma_P)\right| > \varepsilon\right) = 0, \forall \varepsilon > 0. \quad (\text{B.22})$$

This indicates that:

$$\left(f(\hat{\theta}_{noise}) - f(\hat{\theta}_{nf})\right)_K \xrightarrow{\text{Probability}} \text{tr}(\Sigma_P). \quad (\text{B.23})$$

This proves *Theorem 3. (b)*. ■

From the above derivation, we can summarize the influence of noise on this identification model as follows:

(1) If the mean of measurement error  $\mu_P$  is not 0, it will result in differing levels of drift in our estimated parameters, i.e.,  $\hat{\theta}_{noise} \neq \hat{\theta}_{nf}$ . Fortunately, in practical engineering, the mean of the noise can be regarded as 0, i.e.,  $\mu_P = 0, \forall k$ . Therefore, the estimated parameters  $\hat{\theta}$  will not be affected when the sample size is large enough, i.e.,  $\hat{\theta}_{noise} = \hat{\theta}_{nf}$ .

(2) While the constant term  $\text{tr}(\Sigma_P)$  does not influence the optimal solution for  $\theta$ , it does affect the value of the objective function. Namely, the value of  $f(\hat{\theta}_{noise})$  will always be greater than 0, although the parameter  $\theta$  is accurately estimated. Therefore, in this case, we do not have an exact certificate denoting that  $\hat{\theta}_{noise}$  is the exact estimation, different from the noise-free case wherein  $f(\hat{\theta}_{nf}) = 0$  indicates that  $\hat{\theta}_{nf}$  is (one of) the optimal estimations. Still, a smaller  $f(\hat{\theta}_{noise})$  means a better estimation of the parameters  $\theta$ . In this case, the value of  $f(\hat{\theta}_{noise}) - \text{tr}(\Sigma_P)$  ( $\approx f(\hat{\theta}_{nf})$ ) provides a basis for determining if the physical model  $\Omega_{phy}(\theta)$  is correctly selected.

## Appendix C

### SOLUTION METHOD OF THE PFL IDENTIFICATION MODEL

Let us recall the PFL identification model (4) in the manuscript, as follows:

$$\begin{aligned} \hat{\theta} = \arg \min_{\theta} f(\theta) &= \frac{1}{|\mathbf{K}|} \sum_{k \in \mathbf{K}} \|\tilde{P}_*^k - \hat{P}_*^k(\lambda^k, \Omega)\| \\ \text{s.t. } \hat{P}_*^k(\lambda^k, \Omega) &= \underset{P \in \Omega(\theta)}{\operatorname{argmin}} (\lambda^k)^\top P, \quad \forall k \in \mathbf{K} \end{aligned}, \quad ((4) \text{ in the manuscript})$$

wherein the superscript  $k$  denotes the sample index,  $\mathbf{K}$  denotes the sample set;  $\hat{\theta}$  is the optimal estimated value of model parameter  $\theta$ ;  $\tilde{P}_*^k$  is the estimated aggregate power; and  $\tilde{P}_*^k \in \mathbb{R}^T$  is the measurement of the aggregated power  $P_*^k$ , which can be modeled as  $\tilde{P}_*^k = P_*^k + e_P^k$ , wherein  $e_P^k \in \mathbb{R}^T$  denotes the measurement errors of  $P_*^k$ .

For the PFL identification model (4), if  $\forall k \in \mathbf{K}$ ,  $e_P^k = 0$  and the storage-like model  $\Omega_{phy}$  is used, after the parameter  $\sigma^n$  is prescribed, the above model can be reformulated as follows:

$$\begin{aligned} \hat{\theta} = \arg \min_{\theta = \{E_{vb}^n, \bar{E}_{vb}^n, P_{vb}^n, \bar{P}_{vb}^n, P_{id}^n, \bar{P}_{id}^n\}} f(\theta) &= \frac{1}{|\mathbf{K}|} \sum_{k \in \mathbf{K}} \|P_*^k - \hat{P}_*^k(\lambda^k, \Omega)\| \\ \text{s.t. } \forall k \in \mathbf{K}: & \\ \hat{P}_*^k(\lambda^k, \Omega) &= \underset{P^k}{\operatorname{argmin}} (\lambda^k)^\top P^k \\ \text{s.t. } P^k &= \sum_{n \in \mathbf{N}_{vb}} P_{vb}^{n,k} + \sum_{n \in \mathbf{N}_{id}} P_{id}^{n,k} + \sum_{n \in \mathbf{N}_{fix}} P_{fix}^{n,k} : \omega^k \\ E_{vb}^n &\leq \Upsilon_1^n P_{vb}^{n,k} \Delta t + E_{vb,0}^n \Upsilon_2^n \leq \bar{E}_{vb}^n, : \nu_1^{n,k}, \nu_2^{n,k} \\ P_{vb}^n &\leq P_{vb}^{n,k} \leq \bar{P}_{vb}^n, \forall n \in \mathbf{N}_{vb} : \nu_3^{n,k}, \nu_4^{n,k} \\ P_{id}^n &\leq P_{id}^{n,k} \leq \bar{P}_{id}^n, \forall n \in \mathbf{N}_{id} : \nu_5^{n,k}, \nu_6^{n,k} \end{aligned} \quad (C.1)$$

The model (C.1) is a typical bilevel optimization problem, wherein the lower level is a linear optimization problem. Based on the duality theory, the lower-level model can be reformulated into KKT conditions, as:

$$\begin{aligned} \forall k \in \mathbf{K}: \hat{P}_*^k &= \sum_{n \in \mathbf{N}_{vb}} P_{vb}^{n,k} + \sum_{n \in \mathbf{N}_{id}} P_{id}^{n,k} + \sum_{n \in \mathbf{N}_{fix}} P_{fix}^{n,k} \\ E_{vb}^n &\leq \Upsilon_1^n P_{vb}^{n,k} \Delta t + E_{vb,0}^n \Upsilon_2^n \leq \bar{E}_{vb}^n, P_{vb}^n \leq P_{vb}^{n,k} \leq \bar{P}_{vb}^n, \forall n \in \mathbf{N}_{vb} \\ P_{id}^n &\leq P_{id}^{n,k} \leq \bar{P}_{id}^n, \forall n \in \mathbf{N}_{id} \\ \lambda^k + \omega^k &= 0 \\ -\omega^k + (\Upsilon_1^n)^\top (\nu_2^{n,k} - \nu_1^{n,k}) \Delta t - \nu_3^{n,k} + \nu_4^{n,k} &= 0, \forall n \in \mathbf{N}_{vb} \\ -\omega^k - \nu_5^{n,k} + \nu_6^{n,k} &= 0, \forall n \in \mathbf{N}_{id} \\ \nu_1^{n,k}, \nu_2^{n,k}, \nu_3^{n,k}, \nu_4^{n,k} &\geq 0, \forall n \in \mathbf{N}_{vb} \\ \nu_5^{n,k}, \nu_6^{n,k} &\geq 0, \forall n \in \mathbf{N}_{id} \\ \nu_1^{n,k} \odot (E_{vb}^n - \Upsilon_1^n P_{vb}^{n,k} \Delta t - E_{vb,0}^n \Upsilon_2^n) &= 0, \forall n \in \mathbf{N}_{vb} \\ \nu_2^{n,k} \odot (\Upsilon_1^n P_{vb}^{n,k} \Delta t + E_{vb,0}^n \Upsilon_2^n - \bar{E}_{vb}^n) &= 0, \forall n \in \mathbf{N}_{vb} \\ \nu_3^{n,k} \odot (P_{vb}^n - P_{vb}^{n,k}) &= 0, \forall n \in \mathbf{N}_{vb} \\ \nu_4^{n,k} \odot (P_{vb}^{n,k} - \bar{P}_{vb}^n) &= 0, \forall n \in \mathbf{N}_{vb} \\ \nu_5^{n,k} \odot (P_{id}^n - P_{id}^{n,k}) &= 0, \forall n \in \mathbf{N}_{id} \\ \nu_6^{n,k} \odot (P_{id}^{n,k} - \bar{P}_{id}^n) &= 0, \forall n \in \mathbf{N}_{id} \end{aligned} \quad (C.2)$$

Based on this, the model (C.1) using can be converted into a single-level optimization model, as:

$$\hat{\theta} = \arg \min_{\substack{\theta = \{E_{yb}^n, \bar{E}_{yb}^n, E_{yb}^n, \bar{E}_{yb}^n, E_{id}^n, \bar{E}_{id}^n\}, \\ \bar{P}_{yb}^k, P_{yb}^{n,k}, \bar{P}_{id}^k, P_{id}^{n,k}, \bar{P}_{fix}^k, P_{fix}^{n,k}, \\ \omega^k, V_1^{n,k}, V_2^{n,k}, V_3^{n,k}, V_4^{n,k}, V_5^{n,k}, V_6^{n,k}}} f(\theta) = \arg \min_{\substack{\theta = \{E_{yb}^n, \bar{E}_{yb}^n, E_{yb}^n, \bar{E}_{yb}^n, E_{id}^n, \bar{E}_{id}^n\}, \\ \bar{P}_{yb}^k, P_{yb}^{n,k}, \bar{P}_{id}^k, P_{id}^{n,k}, \bar{P}_{fix}^k, P_{fix}^{n,k}, \\ \omega^k, V_1^{n,k}, V_2^{n,k}, V_3^{n,k}, V_4^{n,k}, V_5^{n,k}, V_6^{n,k}}} \frac{1}{|K|} \sum_{k \in K} \|P_*^k - \hat{P}_*^k\| \quad (C.3)$$

s.t. (C.2)

By using the big-M method and introducing binary variables, the complementary constraints in (C.2) can be converted into mixed-integer linear programming, and thus, the model (C.3) can be finally converted into a mixed-integer linear optimization model for the  $l^1$ -norm objective function (or mixed-integer quadratic optimization model for the  $l^2$ -norm objective function).

Here, we clarify that the above method is not computationally efficient since many binary variables need to be introduced, especially when the sample size is large. Some more efficient solution methods have been investigated in [1, 2]. Since this work mainly focuses on whether the PFL model can be uniquely identified from the data, i.e., the solution uniqueness of the model (C.1), we do not go further into the solution method.

## Appendix D

### PRICE SIGNAL DESIGN FOR PROBING UNDETERMINED REGION

Denote  $P_{\#}^q$  as the  $q$ th vertex of the set  $\Pi$ ,  $q = 1, 2, \dots, Q_{\Pi}$ . We assume that  $P_{\#}^j$  is the vertex that is not in the set  $\text{Conv}(\Gamma)$ . We need to collect more data to determine if  $P_{\#}^j$  is in the feasible region  $\Omega$  of the PFL. Specifically, we need to design a price signal  $\lambda_{\#}$  to probe the response of the PFL to check if  $P_{\#}^j$  is a feasible point for the PFL. In the following, we discuss two cases.

(1) *The first case:  $P_{\#}^j$  is a feasible point of the set  $\Omega$*

In this case, for the price signal  $\lambda_{\#}$ , we should have:

$$\lambda_{\#} : P_{\#}^j = \arg \min_{P \in \Omega} (\lambda_{\#})^{\top} P. \quad (D.1)$$

Since we do not know the exact  $\Omega$  now, it is not easy to design the price signal  $\lambda_{\#}$ . However, since  $\Omega$  is a subset of  $\Pi$ , we can use  $\Pi$  to replace  $\Omega$  in the above model to get a conservative signal, i.e.,

$$\lambda_{\#} : P_{\#}^j = \arg \min_{P \in \Pi} (\lambda_{\#})^{\top} P. \quad (D.2)$$

Note that the  $\lambda_{\#}$  that makes the model (D.1) hold will naturally make the model (D.2) hold in this case. Therefore, we only need to analyze how to design  $\lambda_{\#}$  for the model (D.2). The model (D.2) indicates:

$$(\lambda_{\#})^{\top} P_{\#}^j < (\lambda_{\#})^{\top} P_{\#}^q \quad \forall q = 1, \dots, j-1, j+1, \dots, Q_{\Pi} \quad (D.3)$$

This is equivalent to

$$(\lambda_{\#})^{\top} (P_{\#}^j - P_{\#}^q) < 0 \quad \forall q = 1, \dots, j-1, j+1, \dots, Q_{\Pi} \quad (D.4)$$

Denoting  $\Delta P_{\#}^{j,q} = P_{\#}^q - P_{\#}^j$ , we have:

$$(\lambda_{\#})^{\top} (-\Delta P_{\#}^{j,q}) < 0 \quad \forall q = 1, \dots, j-1, j+1, \dots, Q_{\Pi} \quad (D.5)$$

Note that since  $\Pi$  is a polyhedron, and thus there exists  $0 \leq \alpha^{q,s} \leq 1, \sum_{s \in S_j} \alpha^{q,s} = 1$  satisfying

$$\Delta P_{\#}^{j,q} = \sum_{s \in S_j} \alpha^{q,s} \Delta P_{\#}^{j,s}, \quad \forall q = 1, 2, \dots, Q_{\Pi}, \quad (D.6)$$

wherein  $S_j$  is the set of the vertices connecting to the vertex  $j$ .

Substituting (D.6) into (D.5), we have:

$$(\lambda_{\#})^{\top} \sum_{s \in S_j} \alpha^{q,s} \Delta P_{\#}^{j,s} > 0 \quad \forall q = 1, \dots, j-1, j+1, \dots, Q_{\Pi}, \quad (D.7)$$

which is equivalent to the following:

$$(\lambda_{\#})^{\top} \sum_{s \in S_j} \alpha^s \Delta P_{\#}^{j,s} > 0 \quad \forall \alpha^s : 0 \leq \alpha^s \leq 1, \sum_{s \in S_j} \alpha^s = 1. \quad (\text{D.8})$$

The above inequalities can be reformulated into

$$(\lambda_{\#})^{\top} \Delta P_{\#}^{j,s} > 0 \quad \forall s \in S_j. \quad (\text{D.9})$$

Using the price signal satisfying the condition (D.9), if the vertex  $P_{\#}^j$  is a feasible point of the  $\Omega$ , we will get a pair  $(\lambda_{\#}, P_{\#}^j)$  that will be added into the dataset. Then,  $\text{Conv}(\Gamma)$  will also contain  $P_{\#}^j$ .

Note that in our numerical test, since the dimension of the set  $\Pi$  is 2, the formula (9) turns into:

$$(\lambda_{\#})^{\top} P_1 > 0, (\lambda_{\#})^{\top} P_2 > 0. \quad (\text{D.10})$$

(2) *The second case:  $P_{\#}^j$  is not a feasible point of the set*

In this case, using the price signal satisfying the condition (D.9), the response of the PFL is assumed to be  $\hat{P}_{\#}^j$ , which is not  $P_{\#}^j$ . This indicates

$$(\lambda_{\#})^{\top} \hat{P}_{\#}^j > (\lambda_{\#})^{\top} P_{\#}^j. \quad (\text{D.11})$$

With this new pair  $(\lambda_{\#}, \hat{P}_{\#}^j)$ , the set  $\Pi$  is updated as:

$$\Pi_{\#} = \left\{ P \left[ \lambda^1 \cdots \lambda^K \lambda_{\#} \right]^{\top} P \geq \left[ (\lambda^1)^{\top} P_{\#}^1 \cdots (\lambda^K)^{\top} P_{\#}^K (\lambda_{\#})^{\top} \hat{P}_{\#}^j \right] \right\}. \quad (\text{D.12})$$

The formula (D.11) indicates that  $[\lambda^1 \cdots \lambda^K \lambda_{\#}]^{\top} P_{\#}^j \geq [(\lambda^1)^{\top} P_{\#}^1 \cdots (\lambda^K)^{\top} P_{\#}^K (\lambda_{\#})^{\top} \hat{P}_{\#}^j]$  does not hold. Hence,  $P_{\#}^j$  is not in the set of  $\Pi_{\#}$ . This means that if  $P_{\#}^j$  is not a feasible point of the set  $\Omega$ , by adding the data pair consisting of the price signal  $\lambda_{\#}$  and the corresponding response  $\hat{P}_{\#}^j$ , the new set  $\Pi_{\#}$  will not contain  $P_{\#}^j$ .

In summary, the price signal  $\lambda_{\#}$  can be used to probe the response of the PFL to check if  $P_{\#}^j$  is a feasible point for the PFL and thus to reduce the set  $\Delta\Omega$ .

## Appendix E NUMERICAL TEST

### (1) 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 [3].

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  $(\lambda^k, P_{\#}^k), \forall k \in K$ , which is required for the identification of the PFL.

*Step 2.* We select the physical model  $\Omega_{phy}$  parameterized by  $\theta_{vb}^n$  as defined in (2), in which the number of the components, i.e.,  $N_{vb}/N_{td}/N_{fix}$ , are prescribed. Then, we substitute the dataset  $(\lambda^k, P_{\#}^k), \forall k \in K$  and the selected physical model  $\Omega_{phy}$  into the identification model (4). Finally, we solve the identification model (4) to obtain the model parameters  $\theta_{vb}^n = (P_{vb}^n, \bar{P}_{vb}^n, P_{td}^n, \bar{P}_{td}^n, E_{vb}^n, \bar{E}_{vb}^n, E_{vb,0}^n, \sigma^n)$ . In the simulations, to facilitate the model solving, the value of the parameter  $\sigma^n$  is prescribed.

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



## (2) Test with noise-free data

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

First, we investigate the impact of the sample size on the sets  $\text{Conv}(\Gamma)$  and  $\Pi$ , as given in Fig. E-1 (a). Consistent with theoretical results, the sets  $\text{Conv}(\Gamma)$  (or  $\Pi$ ) expands (or shrinks) as the sample size increases. For example, observing the four regions labeled A, B, C, and D in Fig. E-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  $\text{Conv}(\Gamma)$  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  $\text{Conv}(\Gamma)$  expands, the region B is contained within the set  $\text{Conv}(\Gamma)$ , 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  $\lambda \in \{\lambda | \lambda^\top P_1 \geq 0, \lambda^\top P_2 \geq 0\}$  to probe the undetermined region D in Fig. E-1 (a).

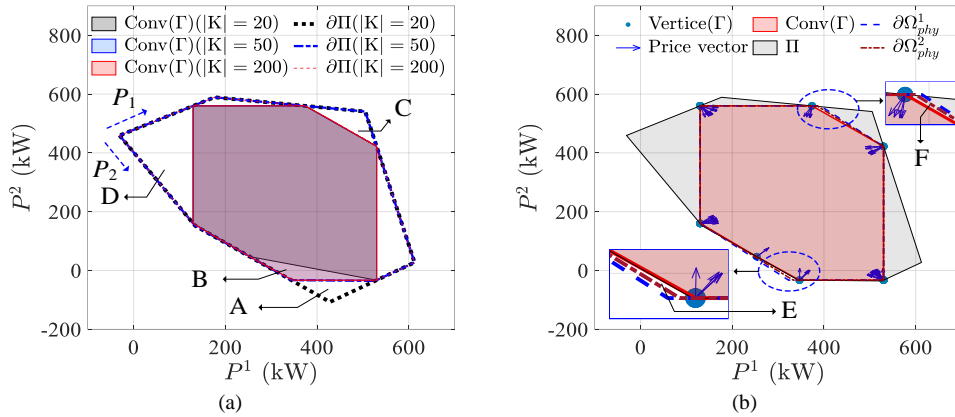


Fig. E-1. (a)  $\text{Conv}(\Gamma)$  and  $\Pi$  under different sample sizes; (b) Identification results of  $\Omega_{\text{phy}}$  under different number of storage (1 and 2 for  $\Omega_{\text{phy}}^1$  and  $\Omega_{\text{phy}}^2$ , respectively). (A:  $\Pi$  shrinks as the sample sizes  $|K|$ 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

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

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  $\Omega_{phy}$  is appropriately selected by inspecting if the practically infeasible region is included in the identification result.

### (3) 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  $\pm 0.5\%$  of the true value. We assume that the electricity price has a relative error of  $\pm 10\%$ . We use Gaussian white noise in the simulations to model the errors based on these settings. Specifically, for the aggregate power  $P_*^k$ , we assume that the relative error  $e_p$  obeys a Gaussian distribution with a mean of 0 and a standard deviation of  $\sigma_p$ , i.e.,  $e_p \sim \mathcal{N}(0, \sigma_p^2)$ . We adopt the  $3\sigma$  criterion, i.e., the probability that the noise value is distributed in  $(-3\sigma_p, +3\sigma_p)$  is 99.74%, and hence  $\sigma_p$  is set to  $(0.5/3)\%$ . For the electricity price  $\lambda^k$ , we assume that the relative error  $e_\lambda$  obeys a Gaussian distribution with a mean of 0 and a standard deviation of  $\sigma_\lambda$ , i.e.,  $e_\lambda \sim \mathcal{N}(0, \sigma_\lambda^2)$ , in which  $\sigma_\lambda$  is set to  $(10/3)\%$  based on the  $3\sigma$  criterion. Based on the above settings, the original noise-free dataset  $(\lambda^k, P_*^k), \forall k \in K$  turns to the new noisy dataset, denoted as  $(\tilde{\lambda}^k, \tilde{P}_*^k), \forall k \in K$ . Note that since the aggregated power is calculated based on the response model (1), we first obtain  $\tilde{\lambda}^k = (1 + e_\lambda)\lambda^k$ , and then calculate the aggregated power  $P_*^k(\tilde{\lambda}^k)$  using the model (1) under the noisy electricity price  $\tilde{\lambda}^k, \forall k \in K$ , and then obtain the noisy aggregated power  $\tilde{P}_*^k$  by adding the noise to  $P_*^k(\tilde{\lambda}^k)$ , i.e.,  $\tilde{P}_*^k = (1 + e_p)P_*^k(\tilde{\lambda}^k)$ . We use 50 samples, the same as those in the noise-free case in the following test.

First, we investigate the impact of the noise on the set  $\text{Conv}(\Gamma)$ , as given in Fig. E-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  $\text{Conv}(\Gamma_{de})$  and  $\text{Conv}(\Gamma_{un})$ . Interestingly, the set  $\text{Conv}(\Gamma_{un})$  completely covers the set  $\text{Conv}(\Gamma_{de})$ . 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  $e_\lambda$  in this case does not significantly change the direction of the electricity price  $\lambda^k$ ; (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  $\text{Conv}(\Gamma_{de})$  have 50% probability of falling outside the boundary  $\partial\text{Conv}(\Gamma_{de})$  since the errors are symmetric, making the set  $\text{Conv}(\Gamma_{un})$  with  $\tilde{P}_*^k$  as the extreme points cover the set  $\text{Conv}(\Gamma_{de})$ . 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  $\text{Conv}(\Gamma_{un})$ , which is easily mistaken as part of the feasible region of the PFL.

Second, we investigate the impact of the noise on sets  $\Pi$ . It is easy to notice that the shape of  $\Pi$  has changed considerably by the noise. The reason is that the direction of the boundary  $\partial\Pi$ , as shown in Fig. E-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  $\Pi$  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  $\text{Conv}(\Gamma_{de})$  but bounded by the boundary  $\partial\Pi_{de}$ . Nevertheless, it is identified as the infeasible region of the PFL under noisy data since it is not within the boundary  $\partial\Pi_{un}$ . Moreover, the regions B and C are identified by using noise-free data as infeasible regions since they are not within the boundary  $\partial\Pi_{de}$ . Besides, it is 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. E-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.

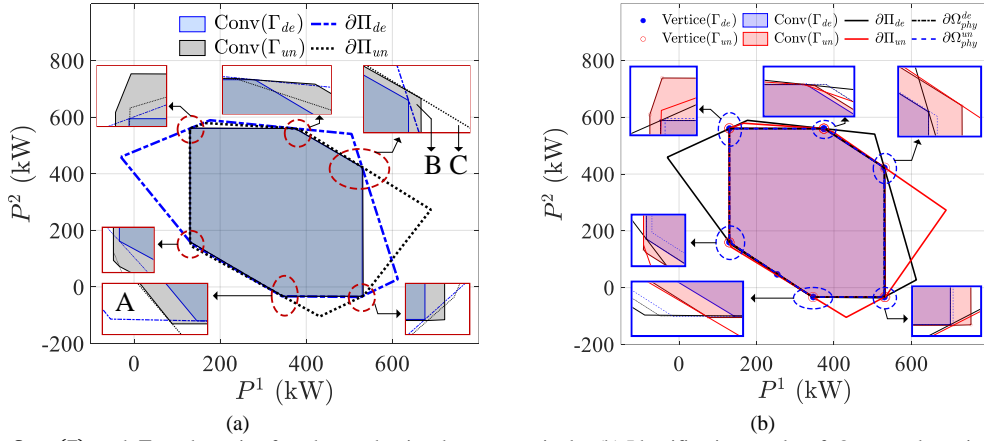


Fig. E-2. (a)  $\text{Conv}(\Gamma)$  and  $\Pi$  under noise-free data and noisy data, respectively; (b) Identification results of  $\Omega_{phy}$  under noise-free data and noisy data (noise-free data for  $\Omega_{phy}^{de}$  and noisy data for  $\Omega_{phy}^{un}$ ). (A: Undetermined region under 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.)

Third, we analyze the identification results. The feasible regions of the aggregated power in  $\Omega_{phy}^{de}$  and  $\Omega_{phy}^{un}$  are given in Fig. E-2 (b). In addition to the similar conclusions that we have previously obtained, we also find that there is very little difference between  $\Omega_{phy}^{de}$  and  $\Omega_{phy}^{un}$  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  $\text{Conv}(\Gamma)$  and the set  $\Pi$  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,  $\Omega_{phy}^{de}$  or  $\Omega_{phy}^{un}$ , 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.

#### (4) Conclusions

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  $\Omega_{phy}$  is appropriately selected, allowing a certain range of noise on the dataset has little effect on the identification results, although both the set  $\text{Conv}(\Gamma)$  and the set  $\Pi$  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  $\text{Conv}(\Gamma)$  and  $\Pi$ , and thus ensure the practical identifiability of the PFL.
- 2) Selecting an appropriate physical model  $\Omega_{phy}$  will help identify the PFL model under incomplete and noisy information, while an incorrect  $\Omega_{phy}$  produces wrong results. After obtaining the identification result of the PFL, we should check if the physical model  $\Omega_{phy}$  is appropriately selected by inspecting if any response power newly measured is out of the feasible range calculated by the identification model.
- 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.

#### REFERENCES

- [1] Y. Wen, Z. Hu, and L. Liu, "Aggregate temporally coupled power flexibility of DERs considering distribution system security constraints," *IEEE Transactions on Power Systems*, vol. 38, no. 4, pp. 1-13, 2022, 2022.
- [2] Z. Tan, Z. Yan, Q. Xia *et al.*, "Data-Driven Inverse Optimization for Modeling Intertemporally Responsive Loads," *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 4129-4132, 2023/09//, 2023.
- [3] S. Lu, J. Ding, and M. Chen. "Supplementary Material," 2024; <https://github.com/Shuai-Lu/ShuaiLu-On-the-Solution-Uniqueness-of-Data-Driven-Modeling-of-Flexible-Loads> or <https://arxiv.org/abs/2403.15029v3>.