On the Sample Complexity of Storage Control

Zhiqi Wang, Zeren Tan, Yang Yu

***Abstract*—Understanding the data value for energy-storage control is critical. The performance of the control policy is highly related to the quality of demand information. An accurate prediction about future demand can better the performance of energy storage control. Thus, the storage control asks for sufficient data-sample collection for qualified prediction. However, the lack of a theory to quantify the data sufficiency for the energy-storage control problem. Meanwhile, demand data samples include privacy information while the storage managers have to procure the data from data owners. Thus, it is necessary to determine the relationship between the data size and the storage-control performance. In addition, a growing number of studies have proposed many storage-control policies. However, we are unknown how to theoretically verify their data-use efficiency. Here, we develop the *sample complexity* theory of storage-control problem, which enables us to theoretically measure the data-use efficiency of the control strategy and assess the data value. We proposed the sample-based dynamic programming (SDP) algorithm that is both cost-minimization and data-use efficient. Based on the SDP and the sample complexity theory, we manifest the trade-off between data size, computational load, and storage-control performances. Finally, we used real-world data to conduct numerical experiments to validate the effectiveness of the proposed method.**

***Index Terms*—Storage Control, Sample Complexity, Dynamic Pricing, Threshold Policy.**

I. Introduction

Energy storage is valuable for managing the uncertainty and volatility in future power grid operations. While the decarbonization progress of the power grid is associated with the enlargement of the system uncertainty and volatility, the studies on energy storage have sufficiently discussed the storage control strategies for mitigating the uncertainty due to new electricity technologies, such as wind power plants [1], [2], [3], [4], rooftop solar panels [5], [6], [7], distribution systems [8], [9], [10], [11], demand response markets [12], electric vehicles [13], [14], and emission reduction [15]. Many studies also addressed the strategies of arbitrage by storage in the markets associated with the time-of-use (ToU) rate or dynamic price. The strategies of shifting demand from peak period to off-peak period have been comprehensively examined during the last decades.

Data plays a vital role during energy storage control. However, there lacks a systematic way of discussing the value of data. While the energy storage is operated in a multi-stage and uncertain environment, the storage control strategy has to consider the future forecasts, which provide distributional information about uncertain factors, such as the demand level and renewable energy generation [5], [7], [16]. The predictions highly rely on historical data. However, those pieces of data either raise privacy concerns or needs to be procured from the data dealers. Therefore, it is necessary to understand how much data is necessary to guarantee energy-storage control performance. Further, we notice that the literature clues have shown that the estimation model and control algorithm will influence the performance of energy storage control given the same distribution information. Therefore, it is critical to verify the data-use efficiency of the pair of estimation models and control algorithm effectively before assessing the value of data. However, there lacks a theory enabling the systematic methods to verify the data-use efficiency of the control algorithm and assess the value of data.

In this research, we develop the sample complexity theory for the energy storage control problem to verify the data-use efficiency of the control algorithm and assess the value of data. Currently, the sample complexity theory has been developed to analyze the relationship between the data size and the performance of machine learning [17], [18], [19]. To our best knowledge, there still lacks a sample complexity theory of multi-stage dynamic and stochastic control problems.

We use the demand data as an example to develop the sample complexity theory for the energy-storage control problem. Many storage control studies focus on using storage to manage demands. However, the demand data include the consumer's privacy. The more data collected, the more pieces of privacy exposes to the risk of leakage. Therefore, verifying the data size of demand samples and energy storage performance is critical.

To discuss the value of data, we develop a sample-based energy storage approach that efficiently uses the

(Corresponding author: Yang Yu and Zhiqi Wang.) This work was supported by the National Key R&D Program of China (Grant/Award Number: 2020AAA0105402), the Shanghai Qi Zhi Institute, and the Institute for Interdisciplinary Information Core Technology, Xi'an.

Zhiqi Wang is with the Institute for Theoretical Computer Science, Shanghai University of Finance and Economics (e-mail: wangzhiq19@mails.tsinghua.edu.cn)

Zeren Tan is with the Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, Beijing, 100084, P.R. China (e-mail: tanzr20@mails.tsinghua.edu.cn).

Yang Yu is with the Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, Beijing, 100084, P.R. China, and Shanghai Qi Zhi Institute (e-mail: yangyu1@tsinghua.edu.cn).

Method	Input	Guarantee
DP-based Algorithm	Exact Distribution	Optimal
Online Algorithm	None	Competitive Ratio
Sample-base DP (SDP)	Samples	Sample Complexity

TABLE I: The difference between SDP and previous algorithms

data. The current studies on energy-storage control either assume the demand distribution is fully known or consider the online control problem when the distribution information is absent. However, both of these two streams partially reflect real-world conditions. The storage operators in the real world neither fully know the distribution of the demand nor have the relevant information. Instead, the operators have historical records that are samples of the demand. Hence, it is necessary to consider the optimal control strategy according to the sample rather than the distribution. The samples only record partial information about the distribution. Thus, the information loss must cause the storage-control strategy according to samples to be worse than that in the scenario when the distribution is fully known. It is necessary to calibrate the regret due to the information loss. Simultaneously, it also deserves a discussion about whether the optimal charging strategy according to the sample is still threshold based.

To our best knowledge, this is the first study introducing the sample complexity theory into the energy-storage control. We first examine and analyze the optimal strategy of storage control according to the sample. Subsequently, we develop a theoretical approach to clarify the mechanism by which the information gap between samples and distribution causes the regret of the storage control. We also propose a method to clarify the calibration of the regret associated with the given data size. The results manifest the trade-off between the data demand and computational load during the storage control: the more samples, the less the computing load required for accurate storage control.

The remainder of this paper is organized as follows: Related studies are summarized in Section II. In Section III, the storage control problem is first presented and the double-threshold form of the optimal control policy, which acts as the base for sample complexity is proven. Subsequently, we state the specific process of the proposed sample-based dynamic programming (SDP) algorithm. Finally, we measure the discretization loss on the strategy set and demand distribution and introduce the sample complexity bound. Section V evaluates the performance of the proposed methods using real-world data. The conclusions of this study are presented in Section VI.

II. Related Work

A. Storage Control

Presently, numerous studies have been conducted aiming at minimizing user costs using storage systems [20], [21], [22], [23], [24], [25]. Many techniques have been employed in previous studies to solve the optimal policy, such as dynamic programming, reinforcement learning, optimization methods, greedy algorithms, and online algorithms.

Present studies on storage control can be divided into two groups according to their assumptions on demand information known by the storage operators. The first stream of the literature assumes that the storage operators know the distributions of the demands. The current studies have addressed the optimal strategies of storage control according to the time-varying price signal. For instance, Van de ven *et al.* proposed the optimal control policy where the demands follow known Markovian processes and proved the threshold structure of the policy [26]. Oudalov *et al.* presented an optimal operation strategy based on DP to maximize the customer's economic benefit [27]. Wang *et al.* proposed an efficient arbitrage policy and studied the optimal sizing problem based on dynamic programming [16]. Wu *et al.* used one-shot load decomposition to solve the optimal control policy in the setting of dynamic pricing and fixed demand [28].

The second stream of the literature assumes that the storage operators do not know the distribution of the demands and consider the robust control problem of storage charging. The objective function for such problems is usually of a max-min form, where max is taken over the action space and min is taken over all possible demands, which can be regarded as the action space of an adversary. For example, Qin *et al.* designed an online modified greedy algorithm for storage control when the demand is uncertain [29]. Chau *et al.* designed an online algorithm to minimize cost with a competitive ratio guarantee [30]. Some studies consider the setting of online convex optimization with switching cost [31], [32]. In this setting, the metric is often the regret with the optimal offline algorithm in the worst case.

Numerous studies have solved the optimal control policy based on reinforcement learning algorithms such as Q-learning and the Markov decision process. Bui *et al.* considered the load and renewable energy uncertainty and designed a double deep Q-learning algorithm to solve the optimal operation strategy in community battery energy storage systems [33]. Henri *et al.* introduced a reinforcement learning-based machine learning algorithm to schedule the operations of energy storage systems [34]. Hu *et al.* used a Markov decision process to model battery operations considering battery aging [35].

Other techniques have been considered as well. For example, Xu *et al.* proposed a linear time constant space control algorithm and associated the optimal control policy with the Lagrangian multiplier [36]. Hashmi *et al.* developed a quadratic time algorithm and observed that it is sufficient to consider prices in a sub-horizon to determine the optimal action under net-metering policies [37].

B. Sample Complexity

Sample complexity is significant in theoretical machine learning [17], [18], [19], [38]. It represents the number of samples required to train a nearly optimal algorithm with high probability. Sample complexity is a theory discussing the following questions in control: 1. How can a good control policy be solved according to samples? 2. What is the theoretical guarantee of the control policy? As shown, with positive probability, the data may be considerably different from those of the true distribution although the data are sampled from the distribution. The controller cannot solve an even nearly optimal control policy with such samples. Hence the theoretical guarantee we use is *sample complexity*; that is, how many samples are required to obtain a nearly optimal control policy with high probability.

Sample complexity has been widely studied in optimal transport theory [39], machine learning [18], [40], [17], [41], algorithmic game theory and mechanism design [42], [43], [44], [45]. However, to the best of our knowledge, we are the first to introduce the notion of sample complexity in storage-control research. In this study, we investigate the sample complexity bound to ensure a small error bound on the computed cost.

Guo *et al.* converted the control problem to probably approximately correct (PAC) learning and proved that the problem can have a sample complexity guarantee if it has two special forms [38]. Assuming that the input space of a control problem is a product distribution of n independent distributions, a control policy can be regarded as a *hypothesis* in PAC learning, which is a mapping between the input space and $[0, 1]$. The set of all possible policies can be regarded as the hypothesis set. The cost incurred by the policy can be interpreted as the reward of PAC learning. The true distribution is not known; however, we can learn an empirical distribution from the known samples. The difference in the reward of a hypothesis between the empirical distribution and true distribution is just a famous concept, *the generalization error*. In the PAC learning theory, if we sample a sufficiently large number of times, the generalization error of any hypothesis will be sufficiently small. Therefore, if we simply choose the optimal hypothesis on the empirical distribution, it is a nearly optimal hypothesis on the true distribution.

C. Our Contribution

In this study, we consider a general and complex storage model and derive the sample complexity bound to extract a nearly optimal control policy. The sample complexity can be regarded as a criterion to measure the value of data. Our contributions are summarized as follows:

- We theoretically defined and analyzed the sample complexity problem of energy-storage control. The sample complexity problem clarifies the trade-off frontier between the performance of a control policy and the size of data collected from the consumers. Thus, the data-use efficiency of every control policy can be assessed. To our best knowledge, we are the first study systematically developing the sample complexity theory in the area of optimal control in smart grid. In contrast, the current sample complexity discussions are mainly active in machine learning.
- We proposed a double-threshold control policy and demonstrated its optimality of utilizing the data according to our sample complexity theory. In addition, the double-threshold control policy is also economically optimal, which is able to figure out the benefit-maximization storage charging-discharging strategy. Therefore, our research demonstrates that there exists an economically energy-control strategy that also most efficiently utilizes the data.
- We modeled the trade-off relationship between the computational load, the size of data collected from consumers, the computational load, and the cost of energy-storage operation. Given the same energy-storage operation cost, the fewer consumer data collected induce a higher computational load of computing the optimal control strategy. The trade-off relationship further allows us to develop the model calibrating the marginal value of the collected data.

III. Model

Our storage control model is based on a previous study [26]. A power system in which the demand at the i^{th} time slot is d_i, which yields a distribution whose cumulative distribution function is D_i. Assume that the support set of D_i is $[0, \overline{D}_i]$, and let $\overline{D} = \max_i\{\overline{D}_i\}$. The demand d_i is not necessarily identical for different time slots; however, we assume that they are independent. At each time slot, the user can satisfy the demand via two options: either buying from the grid at the current price or discharging from the storage. Consumers make their decisions according to a ToU price. Let p_i be the ToU price in the i^{th} period; we assume that $0 \leq p_i \leq \overline{p}$.

Based on the storage-operation model presented in [26], we further consider additional constraints, including

- *Storage loss*: The storage is not completely efficient. The energy in the storage is lost over time.
- *Charge & discharge loss*: The charging and discharging process is also not completely efficient. It depends on the parameter of charging and discharging efficiency.

We, particularly, include these two constraints mainly because they cause the storage control to be sensitive to the accuracy of demand forecast. We also include the constraints in the model of [26], which include

- *Capacity constraint*: During the process, the storage level is upper bounded by the capacity.
- *Charge & discharge speed*: The user has a charging/discharging limit per slot.

We assume that the energy level at the beginning of the i^{th} time slot is C_i. During the i^{th} time slot, the energy directly purchased from the grid is denoted as $A_1(i)$, the energy injected into the storage as $A_2(i)$, and the energy discharged from the storage as $A_3(i)$. The capacity of the storage, and charging and discharging speeds are denoted as $\overline{C}$, $\overline{A}_c$, and $\overline{A}_d$, respectively. Assuming that the storage loss is γ, while the charging and discharging losses are μ_c and μ_d, respectively. Therefore, the constraints of the storage control in the i^{th} time slot are modeled as below:

$$A_j(i) \geq 0, \quad \forall j \in \{1,2,3\}, \tag{1}$$

$$A_2(i) \leq \overline{A}_c \quad \text{(Charge speed constraint)}, \tag{2}$$

$$A_3(i) \leq \overline{A}_d \quad \text{(Discharge speed constraint)}, \tag{3}$$

$$A_1(i) + \mu_d A_3(i) = d_i \quad \text{(Discharge loss)}, \tag{4}$$

$$C_i + \mu_c A_2(i) - A_3(i) \leq \overline{C} \quad \text{(Capacity constraint)}, \tag{5}$$

$$C_{i+1} = \gamma(C_i + \mu_c A_2(i) - A_3(i)) \quad \text{(Storage loss)}. \tag{6}$$

Although the demand in each period yields a latent distribution, the distribution is unknown to the storage operator. Instead, the operator has a set of historical demand data, which can be used to estimate the information of D_i. We aimed at designing a nearly optimal storage operation policy based on data and discuss the theory of how the size of the data set influences the performance of the nearly-optimal control strategy.

IV. Theory and Algorithm

The key question here is how much extra cost will incur due to the use of the demand samples compared with the control according to the distribution of the demand. The other question is whether all samples are necessary. When the size of the sample set is large, using all samples incurs a high computational load. Therefore, the sample-based control has to make a unique decision about how many samples should be used. We discuss this question by examining how the resolution of samples impacts the performance of the sample-based DP.

In this section, we first clarify the optimal storage control according to the distribution of demand. Then, we propose a sampled-based DP algorithm, which generates the optimal control strategy according to the samples of demand. Finally, we calibrate the difference between using the distribution and using the samples.

A. Optimal Control Policy: Double-Threshold Structure

The optimal storage control needs to make three decisions: the energy purchased from grid $A_1(i)$, the amount of charging $A_2(i)$, and the amount of discharging $A_3(i)$. To simplify the discussion, we demonstrate that we only need to focus on the strategy for setting C_{i+1}, which will fully decide the value of $A_1(i)$, $A_2(i)$, and $A_3(i)$. The proof of the theorem is presented in the appendix A.

Theorem IV.1. *If the storage level at the beginning of the next time slot, C_{i+1} is provided, the optimal solutions for $A_1(i)$, $A_2(i)$, and $A_3(i)$ are fixed.*

The above theorem reduced the number of controlled variables from three to one. In the rest part of this subsection, we mathematically formulate the sample-based storage-control problem of deciding C_{i+1} in order to enable the discussion about the value of data. Here, we use the sample-based DP problem to formulate the storage control problem. We select the DP problem because the DP formulation can clarify the optimal strategy in both the sample-based and the distribution-based cases. Thus, the distribution-based optimal strategy is able to play the role of benchmark for calibrating the value of data.

To develop the sample-based DP, we have to sequentially clarify the state transition function of battery cross-time slots, feasible action space, and the objective function of the storage control problem. Then, we can get the optimal control strategy.

State transition function: We first propose the associated state transition function of the energy storage control in every stage. The cost in each time slot arises from the purchase from the grid. If the storage is not charged or discharged, the battery level in the next time slot will be $\gamma \cdot C_i$, and the operator will suffer a cost of $d_i p_i$. When $\gamma^{-1}C_{i+1} < C_i$, The battery is discharged while $A_2(i) = 0$. In this case, the residual demand is satisfied by directly purchasing energy from the grid. Symmetrically, the battery discharge when $\gamma^{-1}C_{i+1} > C_i$ while $A_3(i) = 0$. According to the above analysis, the transaction of the cost function time slot i to $i+1$ is that:

$$\begin{aligned} g_i(C_{i+1}, C_i, d_i) =& [d_i + (\gamma^{-1}C_{i+1} - C_i)^+ \mu_c^{-1} \\ &+ (\gamma^{-1}C_{i+1} - C_i)^- \mu_d] p_i, \end{aligned} \tag{7}$$

where $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$.

Feasible action space: The constraints of the storage in equations 1–6 define the feasible set of C_{i+1}. We convert equations 1-6 to three types of constraints about C_{i+1}. The first type is the charging-discharging speed limit, which is $-\overline{A}_d \leq \gamma^{-1}C_{i+1} - C_i \leq \overline{A}_c\mu_c$.

The second type is the constrain about the lower-bound of energy stored in the battery because of the discharge speed constraint: $\gamma^{-1}C_{i+1} - C_i \geq -\mu_d^{-1}D_i$.

The above two types of constraints and the storage capacity constrain together determine the upper and lower bounds of the C_{i+1}, which is denoted by $\overline{U_i}(C_i)$ and $\underline{U_i}(C_i)$ respectively.

$$\begin{aligned}\overline{U_i}(C_i) &= \gamma \cdot \min\{\overline{C}, C_i + \overline{A}_c\}, \\ \underline{U_i}(C_i) &= \gamma \cdot \max\{0, C_i - \overline{A}_d, C_i - \mu_d^{-1}D_i\}.\end{aligned} \tag{8}$$

Objective function The storage control pursues minimizing the total energy cost during the whole operation period. We use $G_i(C_i)$ to denote the total cost after time i. Therefore, the objective function of the energy storage control is to minimize

$$G_i(C_i) = \mathbb{E}_{d_i \sim D_i}\left[G_{i+1}(C^*_{i+1}) + g_i(C^*_{i+1}, C_i, d_i)\right], \tag{9}$$

where C^*_{i+1} is the optimal solution for the cost.

Double-threshold structure of the optimal control strategy: According to the above objective function and feasible action space, the optimal control strategy of the battery is

$$C^*_{i+1} = \mathop{\arg\min}_{x \in [\underline{U_i}(C_i), \overline{U_i}(C_i)]} \{G_{i+1}(C_i) + g_i(x, C_i, d_i)\}. \tag{10}$$

We argue that the optimal control strategy of Ea. Equation 10 satisfies a double-threshold rule, which is summarized in the following theorem.

Theorem IV.2. *There exist two thresholds β_i^-, β_i^+, which determine the optimal control strategy C^*_{i+1}.*

$$C^*_{i+1} = \begin{cases} \min\{\beta_i^-, \overline{U_i}(C_i)\} & C_i \leq \frac{\beta_i^-}{\gamma}, \\ \max\{\beta_i^+, \underline{U_i}(C_i)\} & C_i \geq \frac{\beta_i^+}{\gamma}, \\ \gamma C_i & \frac{\beta_i^-}{\gamma} \leq C_i \leq \frac{\beta_i^+}{\gamma}. \end{cases} \tag{11}$$

The entire proof can be found in appendix B.

B. Sample-based dynamic programming (SDP)

Because of the double-threshold structure explained in Theorem IV.2, all the optimal control strategies of the total n periods can be characterized by $2n$ thresholds. In the following subsections, we utilize this observation to analyze the difference between the scenario of using the distribution and that of using the samples.

We first develop the sample-based DP algorithm of optimal storage control. We develop the sample-based DP by two steps. First, we estimate the empirical distribution from the samples. Then, we use the DP algorithm based on the empirical distribution to obtain the sample-based control strategy.

Algorithm 1 uses the empirical distribution function (EDF) to estimate the true distribution. The EDF is broadly adopted in the main-stream sample complexity literature. The popularity of EDF is due to two advantages. First, the EDF does not rely on any assumption about the family and parameters of the estimated true distribution. Therefore, once the data is sufficiently sampled, the EDF can always converge to the true distribution. Second, in many cases, the EDF is empirically efficient. The estimation divides the entire support set of the demand level into a sequence of bins, whose width is ϵ_2. Every sample is rounded up to the nearest multiple of ϵ_2.

Algorithm 1 Discretize Empirical Distribution

Require: Demand data D_{ij}
1: $N \leftarrow$ the number of samples
2: **for** $i = 1, \cdots, n$ **do**
3:     **for** $j = 1, \cdots, N$ **do**
4:         $P_i^e(\lceil \frac{D_{ij}}{\epsilon_2} \rceil \cdot \epsilon_2) \leftarrow P_i^e(\lceil \frac{D_{ij}}{\epsilon_2} \rceil \cdot \epsilon_2) + \frac{1}{N}$
5:     **end for**
6: **end for**

According to the empirical distribution, we can develop the associated DP algorithm to clarify the threshold rules for determining C_{i+1}. We refer to the DP algorithm using the empirical distribution as the sample-based DP algorithm (SDP), which is summarized in Algorithm 2. Note that the feasible domain of the decision variable C_{i+1} is continuous. To develop the SDP, the feasible domain of C_{i+1} needs to be discretized for all i. We refer the feasible domain of C_{i+1} as $[\underline{U_i}(C_i), \overline{U_i}(C_i)]$ and discretize the domain by the resolution of ϵ_1. The discretized feasible domain of C_{i+1} is defined as $\mathcal{A} = \{0, \epsilon_1, \cdots, \lfloor \frac{\overline{C}}{\epsilon_1} \rfloor \cdot \epsilon_1\} \cap [\underline{U_i}(C_i), \overline{U_i}(C_i)]$.

Note, the SDP still made the decision of charging and discharging according to the double-threshold rule defined in Equation 11. In contrast to the theoretical optimal strategy, the optimal decision of SDP C^*_{i+1} must be a multiple of ϵ_1. Line 2-5 computing the approximation of $G^{e'}_{i+1}(\cdot)$. The empirical distribution we use is discretized. Therefore, we use $\frac{\Delta(j)}{\epsilon_1}$ to approximate $G^{e'}_{i+1}(\cdot)$. To figure out the double thresholds, we notice that $g(C_{i+1}, C_i, d_i)$ is piecewise linear respect to C_{i+1} while the slope of both pieces are $\gamma^{-1}\mu_c^{-1}p_i$ and $\gamma^{-1}\mu_d p_i$ respectively. The two slopes are the thresholds.

According to Lemma B.2, $G^e_{i+1}(\cdot)$ is a convex function. Thus, its derivative is monotone decreasing. If we find the point such that $G^{e'}_{i+1}(\cdot) = \gamma^{-1}\mu_c^{-1}p_i$ or $\gamma^{-1}\mu_d p_i$ then it is not profitable to charge or discharge anymore at that point. Line 9-12 compute the optimal $C_(i+1)^*$ based on the optimal thresholds computed in line 6 and line 7. Because of the limitation of charge/discharge speed, if we take floor function when we discharge it may violate the constraint, so we take the ceil to discretize. For the same reason, we take the floor when charge. The line 13 updates $G^e_i(\cdot)$.

Algorithm 2 Sample-based Dynamic Programming

1: Initialize $\forall i, G^e_{n+1}(i) \leftarrow 0$
2: **for** $i = n, \cdots, 1$ **do**
3: **for** $j = 0, \epsilon_1, \cdots, (\lfloor \frac{\overline{C}}{\epsilon_1} \rfloor - 1) \cdot \epsilon_1$ **do**
4: $\Delta(j) \leftarrow (G^e_{i+1}(j+\epsilon_1) - G^e_{i+1}(j))$
5: **end for**
6: $\beta_i^- \leftarrow$ Smallest j such that $\Delta(j) \leq \gamma^{-1}\mu_c^{-1}p_i\epsilon_1$ (if there is no such j, then $\beta_i^- \leftarrow \lfloor \frac{\overline{C}}{\epsilon_1} \rfloor \cdot \epsilon_1$)
7: $\beta_i^+ \leftarrow$ Smallest j such that $\Delta(j) \leq \gamma^{-1}\mu_d p_i\epsilon_1$ (if there is no such j, then $\beta_i^+ \leftarrow \lfloor \frac{\overline{C}}{\epsilon_1} \rfloor \cdot \epsilon_1$)
8: **for** $j = 0, \epsilon_1, \cdots, \lfloor \frac{\overline{C}}{\epsilon_1} \rfloor \cdot \epsilon_1$ **do**
9: **if** Storage discharge **then**
10: $C^* \leftarrow \lceil \frac{C^*_{i+1}}{\epsilon_1} \rceil \cdot \epsilon_1$
11: **else** $C^* \leftarrow \lfloor \frac{C^*_{i+1}}{\epsilon_1} \rfloor \cdot \epsilon_1$
12: **end if**
13: $G^e_i(j) \leftarrow \sum_{d_i=0}^{\lceil \frac{\overline{D}}{\epsilon_2} \rceil \cdot \epsilon_2} P^e_i(d_i)(G^e_{i+1}(C^*) + g_i(C^*))$
14: **end for**
15: **end for**

Remark 1: The sample complexity of deep-neural network (DNN) method for storage control can also be analyzed. In the above energy control problem, we argue that reinforcement learning (RL) can also achieve the same sample complexity as DP. In fact, the RL is equivalent to the DP in the storage-control model formulated in Section III. Here, we explain how Q-learning, one type of RL, is equivalent to DP. During the training process, the Q-learning algorithm iteratively updated the Q-value in every state. In the storage control model formulated in Section III, we can first compute all possible costs of the n^{th} time slot $G_n(C_n)$. Then, we can run the Q-learning algorithm to determine the optimal battery level of the $(n-1)^{th}$ time slot. The optimal battery level makes the Q-values of the $(n-1)^{th}$ time slot equal to $G_n(C_n)$. Then, we can compute all possible costs of the $(n-1)^{th}$ time slot. Following the same process, we can use Q-learning to determine the optimal battery level of every time slot. Thus, the Q-learning is equivalent to the DP in our storage control problem.

C. Sample Complexity of SDP

The above discussion reveals two steps through which using data to control storage incurs a loss. First, there can exist a difference between the true distribution of the demand from the empirical distribution estimated from the demand data. Particularly, the difference between the two distributions is contingent on the size of ϵ_2. Consequently, it is the resolution of empirical distribution estimation that determines the gap between the optimal strategy and SDP strategy. Further, the gap is also contingent on the resolution of the control-space discretization ϵ_1.

In this subsection, we calibrate the total loss determined using ϵ_1 and ϵ_2 by performing a backward analysis. We first focus on the loss incurred by discretizing the action space and assume that the empirical distribution is given. We denote the total cost of the storage control problem as $H(A,F)$ associated with the control policy A and demand distribution F. Therefore, the loss caused by discretizing the action space is $H(\hat{S},\boldsymbol{D}) - H(S,\boldsymbol{D})$, where $\hat{S}$ is the strategy solved using the SDP when the empirical distribution is $\boldsymbol{D}$. S is the theoretical optimal control strategy associated with the empirical distribution is $\boldsymbol{D}$. Note that $\boldsymbol{D} = \times_{i=1}^n D_i$ because we assume that the hourly demands are independent of each other. We argue that the loss owing to the discretization on the action space is linearly related to ϵ_1.

Theorem IV.3. *Assuming the optimal policy for $\boldsymbol{D}$ is S and the optimal policy given in Eq. 11 on the discretized action space is $\hat{S}$, we have,*

$$H(\hat{S},\boldsymbol{D}) - H(S,\boldsymbol{D}) \leq \overline{p}n\epsilon_1. \tag{12}$$

Proof. We first develop a particular control policy that discretizes the action space by the resolution of ϵ_1 and denote it by S'. Subsequently, we demonstrate that S' satisfies

$$H(S',\boldsymbol{D}) - H(S,\boldsymbol{D}) \leq \overline{p}n\epsilon_1.$$

Because the SDP algorithm guarantees that $\hat{S}$ is the optimal strategy when the discretization resolution is ϵ_1. Therefore,

$$H(\hat{S},\boldsymbol{D}) - H(S,\boldsymbol{D}) \leq H(S',\boldsymbol{D}) - H(S,\boldsymbol{D});$$

thus, we have the conclusion.

We define S' as a proxy of S: when the strategy of S at time slot t decides to discharge the storage to the level of $C_{i+1,S}$, the strategy of S' is rounding $C_{i+1,S}$ up to the nearest multiple of ϵ_1. If the strategy of S at time slot t is to charge or maintain the energy stored in the battery to the level of $C_{i+1,S}$, the corresponding strategy of S' is rounding $C_{i+1,S}$ down to the nearest multiple of ϵ_1.

We use backward induction to prove that S' satisfies Equation 8. We first argue that the total energy cost $G'_i(C_i)$ at the end of the time slot i satisfies

$$\forall C_i, G'_i(C_i) - G_i(C_i) \leq \overline{p}(n-i+1)\epsilon_1.$$

In the last time slot n, the storage only needs to discharge. S' may buy at most ϵ_1 extra energy from the grid, and the loss is at most $\overline{p}\epsilon_1$. Thus, $G'_n(C_n)$ satisfies the condition Equation 8.

Assume that the conclusion holds for $i+1$, and that, for the i^{th} time slot, S chooses c. If the storage discharges, the cost for S' to buy extra energy is at most $\overline{p}\epsilon_1$. According to the induction hypothesis,

$$\begin{aligned} G'_{i+1}(\lceil \frac{c}{\epsilon_1} \rceil \cdot \epsilon_1) &\leq G_{i+1}(\lceil \frac{c}{\epsilon_1} \rceil \cdot \epsilon_1) + \overline{p}(n-i)\epsilon_1 \\ &\leq G_{i+1}(c) + \overline{p}(n-i)\epsilon_1 \end{aligned}$$

Otherwise, if the storage charges, the choice of S' may be smaller than c by at most ϵ_1. Therefore,

$$\begin{aligned} G'_{i+1}(\lfloor \frac{c}{\epsilon_1} \rfloor \cdot \epsilon_1) &\leq G_{i+1}(\lfloor \frac{c}{\epsilon_1} \rfloor \cdot \epsilon_1) + \overline{p}(n-i)\epsilon_1 \\ &\leq G_{i+1}(c) + \overline{p}(n-i+1)\epsilon_1 \end{aligned}$$

□

According to the above Theorem, we bounded the total loss due to the discretization of action space. Now, we present the main theorem of this study, which describes the relationship between the number of samples and the loss of using SDP to decide the storage control.

Theorem IV.4. *If there are $O(\frac{1}{\epsilon_2^3}\log\frac{1}{\delta})$ samples, with a probability of at least $1-\delta$, we have*

$$H(SDP, \boldsymbol{D}) - H(S, \boldsymbol{D}) = O(n\epsilon_1 + n\epsilon_2). \tag{13}$$

Because we have proved that the loss due to the discretization of action space is bounded in Theorem IV.3, we only need to bound the loss due to the error of estimating the distribution of demand from the samples. Therefore, the task is to prove that $H(S,D) - H(\hat{S}^d, D^d)$ is bounded for a sufficient number of samples. Here, $H(S,D)$ denotes the total energy cost when the battery is controlled according to the theoretical optimal strategy S based on the true demand-distribution D. $H(\hat{S}^d, D^d)$ denotes the energy cost when the battery is controlled using the SDP strategy $\hat{S}^d$ according to the empirical distribution D^d, which is estimated from the samples of demand.

We decompose the proof of this theorem into three parts. In the first part, we prove that using the optimal control strategy according to the empirical distribution and true distribution will yield sufficiently similar control results. We prove this argument in Lemma IV.1. In the second part, we argue that the results of control according to the optimal strategy and those according to the SDP strategy are sufficiently close when we have a sufficient amount of data to estimate the empirical distribution. We prove this argument in Lemma IV.2. In the third part, we prove that the optimal and SDP strategies are significantly similar when they both rely on the empirical distribution. We prove this argument in Lemma IV.2.

Lemma IV.1. *Assuming the optimal control policy for $\boldsymbol{D}^d$ is $\hat{S}^d$, we have:*

$$H(\hat{S}, \boldsymbol{D}) \geq H(\hat{S}^d, \boldsymbol{D}^d) - (\sum_{i=1}^{n} p_i)\epsilon_2. \tag{14}$$

Proof. We first analyze the optimal strategy to control the battery according to the empirical distribution estimated using the samples. When the demand in the ith time slot $D_i(x|x \in [(m-1)\epsilon_2, m\epsilon_2])$, $\hat{S}^d$ returns the optimal strategy, which is computed according to $\hat{S}$ when the demand is $m\epsilon_2$. Therefore, the $\hat{S}^d$-based strategy always purchases no less electricity from the grid than the $\hat{S}$-based strategy. However, the extra purchased demand is less than ϵ_2 and incurs an additional cost $p_i\epsilon_2$ in the ith time slot. For the total n time slots, $H(\hat{S}, \boldsymbol{D}) \geq H(\hat{S}^d, \boldsymbol{D}^d) - (\sum_{i=1}^{n} p_i)\epsilon_2$. □

Lemma IV.2. *If there are $O(\frac{1}{\epsilon_2^3}\log\frac{1}{\delta})$ samples, then, with a probability of at least $1-\delta$, we have*

$$H(\hat{S}^d, \boldsymbol{D}^d) \geq H(SDP, \boldsymbol{D}^d) - M\epsilon_2, \tag{15}$$

where $M = \sum_{j=1}^{n}(\mathbb{E}[D_i] + \overline{C}\mu_c^{-1} + \epsilon_2)p_i = O(n)$.

Proof. We can directly apply the conclusion of Theorem 1 of [38] to prove this lemma. According to Theorem 1 of [38], if the cost function is bounded in $[0,1]$, and the sample space is a product distribution of n distributions whose support size is not larger than k, then with probability $1-\delta$ we can have an ϵ-additive optimal strategy with $O(\frac{nk}{\epsilon^2}\log\frac{1}{\delta})$ samples. The support size k in our storage control problem is $\frac{\overline{C}}{\epsilon_2}$, and the only thing remaining to apply the theorem is to bound the cost in our problem.

We argue that $H(A, \boldsymbol{D}) \leq M$ in the energy storage control problem. In the i^{th} time slot, the maximum expected unit of energy we can consume when faced with D_i^d is $(\mathbb{E}[D_i^d] + \overline{C}\mu_c^{-1})$. Further, we found that $\mathbb{E}[D_i^d] \leq \mathbb{E}[D_i] + \epsilon_2$.

Therefore, by applying Theorem 1 of [38], we obtain Lemma IV.2. □

Lemma IV.3.

$$H(SDP, D^d) \geq H(SDP, D) - \overline{p}n\epsilon_1. \tag{16}$$

Proof. This is intuitive because $\boldsymbol{D}$ is stochastically dominated by $\boldsymbol{D}^d$; hence, using the same control policy will incur more cost. Equation 16 can be proven by induction. $GS_i(C_i)$ and $GS_i^d(C_i)$ can be denoted as the cost of

SDP after the i^{th} time slot when faced with $\boldsymbol{D}$ and $\boldsymbol{D}_d$, respectively. We prove that, for all i and C_i,

$$GS_i^d(C_i) \geq GS_i(C_i) - \overline{p}(n-i+1)\epsilon_1.$$

Particularly, $GS_1^d(0) \geq GS_1(0) - \overline{p}n\epsilon_1$ is our objective.

First, $\forall C_i$, $GS_n^d(C_i) \geq GS_n(C_i) - \overline{p}\epsilon_1$ because we only need to discharge in the last time slot. S'' only needs to purchase at most ϵ_1 units of extra energy from the grid. Assume that the conclusion is valid for $i+1$. In the i^{th} time slot, simultaneous charging and discharging is not profitable. If the storage charges, the extra demand of $\boldsymbol{D}^d$ can only be bought from the grid; thus, the conclusion is valid. However, if the storage discharges, S'' needs to also purchase at most ϵ_1 units of extra energy from the grid. Therefore, according to the monotonicity of GS and the induction hypothesis, the conclusion is valid. □

Combining equations 14-16 and Theorem IV.3,

$$H(\hat{S}, \boldsymbol{D}) \geq H(SDP, \boldsymbol{D}) - \overline{p}n\epsilon_1 - (M + \sum_{i=1}^{n} p_i)\epsilon_2,$$

which completes the proof of Theorem IV.4.

D. Value of data: tradeoff between data size, computing load, and performance of storage control

Theorem IV.4 also provides an insight into the trade-off between data values and computational load. The SDP strategy's performance is guaranteed only if the size of the sample set is larger than $O(\frac{1}{\epsilon_2^3}\log\frac{1}{\delta})$, which is contingent to the resolution of control-space discretization ϵ_2. If the discretization has a high resolution, a small set of samples can guarantee a low cost of storage control. However, the high resolution leads to a heavy computational load. According to an computational complexity analysis of algorithm 2 we have the following proposition:

Proposition IV.1. *The running time of algorithm 2 is* $O(\frac{nDC}{\epsilon_1\epsilon_2})$.

Thus, our theoretical discussion about the sample complexity manifests that more computational power can reduce the requirement for demand data, which include the privacy information of consumers. On the other hand, more data can reduce the requirement of computational power, which leads more energy use for computing as well as carbon emissions. The data of demands and computational power are substitute inputs for exploring storage-control strategy.

V. Numerical Experiments

A. Experiment setting

We evaluate the performance of SDP by the public data set, Pecan street [46] which provides real electricity

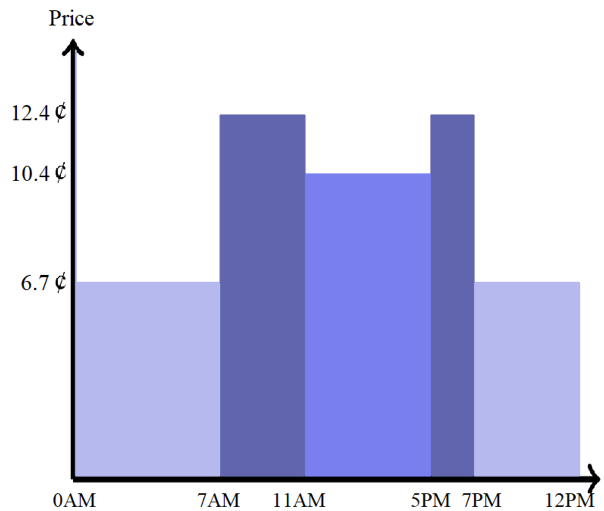


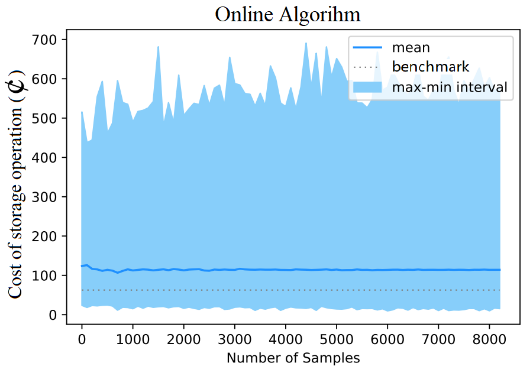
Fig. 1: A multi-peaked ToU pricing scheme

consumption data from 1000 households in the summer of 2016 which is decomposed to 1 minute granularity. We took 10^5 samples as the test set and 10^4 of them as the training set. For the price part, we consider a 4-tier multi-peaked ToU pricing from Ontario Energy Board [47], which is shown in Figure 1.

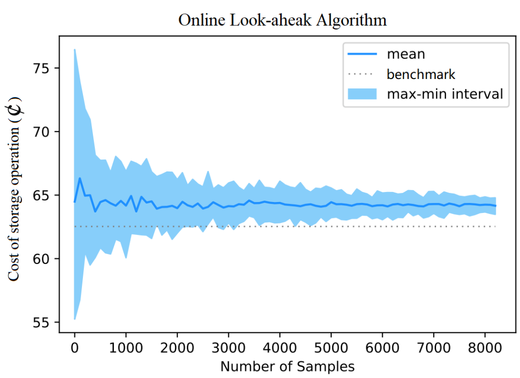
B. Necessity and loss of the SDP

To evaluate the performance of the SDP, we conducted numerical experiments using real-world data considering that the retail market uses a multi-peaked ToU pricing scheme. We first illustrate the necessity of using samples to design a control policy. In the experiment, whose results are summarized in Figure 2, we compare the performances of SDP and two online algorithms proposed in [30]. One online algorithm is threshold-based: the storage is discharged only if the current price is higher than a given threshold (Figure 2(a)). Although this algorithm has a competitive ratio guarantee, its performance is worse than that of SDP. The second is a look-ahead algorithm, which has the information on future demands within a time window. The performance of this algorithm is better than that of the threshold-based online algorithm because it knows the future information (Figure 2(b)). However, the look-ahead algorithm is still worse than the SDP.

The comparison shown in Figure 3 is designed to numerically demonstrate that the SDP is the optimal strategy from the data-use. We compared the SDP with the approaches for designing the control strategies according to the distribution estimated from the data obtained using two other methods. One method is a parametric method that estimates the distribution by the truncated-normal approximation. The other is a non-parametric method



(a) Threshold-Based Online Algorithm



(b) Look-Ahead Algorithm

Fig. 2: Sampling Results of Online Algorithms in [30]

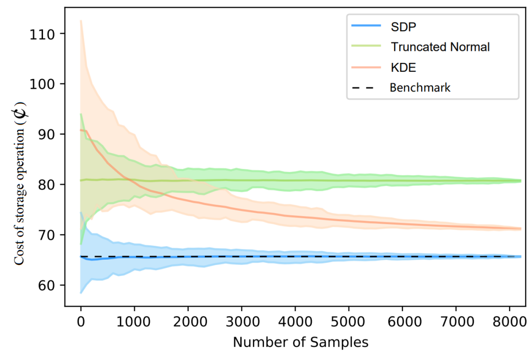


Fig. 3: Sampling Results of Different Methods

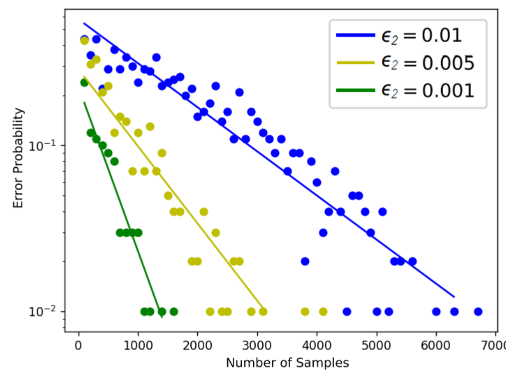


Fig. 4: Error Probability

that estimates the distribution using the kernel density estimation method (KDE). We can see that the SDP method performs better than the other two methods although all three methods develop the control strategy according to the estimated distribution learned from the samples. The result shown that the data-use efficiency is contingent to the pair of estimation approach and control strategy. The data will be efficiently use only if both the estimation approach and the control strategy are appropriately designed. For instance, the empirical distribution and the SDP together can guarantee the efficient use of data.

In fact, we observed that the results of the SDP strategy are significantly close to those of the optimal strategy even if the data size is small. Further, the variance of the SDP strategy due to the sample uncertainty is also limited even if the dataset is small. When there are only 1000 samples, the variance of the SDP strategy's performances is as narrow as that of the look-ahead algorithm when there are more than 3000 samples. Thus, among all the examined methods, the SDP performed the best from both the expected cost and sample-uncertain risk perspectives.

C. Value of data: trade off between data size, computational load, and the control performance

We also verify the values of data given different resolutions of estimating the empirical distribution. Considering three resolutions shown in Figure 4, It is clear that when the resolution is high, a small set of samples is sufficient to limit the gap between the SDP strategy and theoretically-optimal strategy. When $\epsilon_2 = 0.001$, we only need less than 2000 data points to obtain a zero error probability. In contrast, when $\epsilon_2 = 0.001$, we cannot obtain the zero error probability even if the dataset size is over 8000.

Beyond manifesting the trade-off between ϵ_2 and the size of data sample, Figure 4 also presents the elbow-shape of data's value of reducing the error. For all three scenarios, there always exists a tipping point after which more data has very limited value for reducing the SDP's extra cost. Therefore, our method also generates a method to decide whether the size of data collection is necessary.

VI. Conclusion and Discussion

In this study, we investigated the optimal control problem of battery storage systems under the ToU pricing scheme. We formulated the problem using dynamic programming and derived the optimal solution structure while considering the charging/discharging inefficiency and limit, and storage loss. We are the first to introduce the concept of sample complexity into storage control research. With this important tool, we proved the high probability error bound. Using numerical experiments, we demonstrated the effectiveness of our proposed dynamic programming method and error bound. Our research is only the first step to discuss the sample complexity of energy-storage control. There still exists a sequence of open questions, which deserve the future comprehensive studies. For instance, the future study shall examine the sample complexity of controlling various types of storage that have different technical features. It is critical to verify how much data is necessary for the control of every type of storage. From the privacy information protection perspective, it also needs to know whether the data-efficient control method varies by storage types. When the storages belong to different owners, the sample complexity analysis can clarify the size of data sharing that enables the best cooperation equilibrium.

The future studies also need to examine the situation when the assumptions of our research are absent. As the first study discussing the sample complexity of the storage control, our model is built upon a sequence of assumptions. For instance, we in this study assume the hourly demands are independent. In the future study, it is necessary to discuss the sample complexity when the hourly demands are correlated. If the hourly demands are correlated, one hour's demand can store information about the distributions of the demands in other hours. The covariance matrix of the hourly demands determines the optimal sampling approach, which can significantly affect the data efficiency of each control method. Thus, the discussion has to mainly focus on the relation between the covariant matrix, the sampling approach, and the sample complexity of the control method.

It is necessary in future study to discuss the sample complexity of applying the method of deep-neural network (DNN) for smart-grid control. For the complex control problems, such as multi-storage control, the DNN-based methods such as RL are able to guarantee a better outcome than conventional approaches such as DP. In some cases, alternative DNN-based methods can all guarantee the optimal control, e.g. the convolutional neural network (CNN) model vs the full-connected network model. It is necessary to discuss which DNN-based methods are more data efficient.

In summary, the sample complexity of control problem in smart grid deserves more attentions. Data and data-based methods play a core role in system operation and control in smart grid. For the same controlling problem, the alternative data-based controlling methods can all be available. The operator has to select one from them. Simultaneously, the data collection has raised the concerns about privacy leakage. The public debates and legislation practices have explored the approach of limiting data collection over the world. Thus, it is necessary to systematically examine the methods' data-use efficiencies and assess the data value. Analyzing the sample complexity of control method provide us the information about the method's data-use efficiency. The sample complexity study can also theoretically provide the bound of the data value. Thus, the sample complexity for smart control in power system deserves more discussions and analysis.

REFERENCES

[1] E. Bitar, R. Rajagopal, P. Khargonekar, and K. Poolla, "The role of co-located storage for wind power producers in conventional electricity markets," in *Proceedings of the 2011 American Control Conference*. IEEE, 2011, pp. 3886–3891.

[2] E. Bitar, P. Khargonekar, and K. Poolla, "On the marginal value of electricity storage," *Systems & Control Letters*, vol. 123, pp. 151–159, 2019.

[3] Y. Zhao, J. Qin, R. Rajagopal, A. Goldsmith, and H. V. Poor, "Risky power forward contracts for wind aggregation," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2013, pp. 54–61.

[4] H. Khazaei and Y. Zhao, "Indirect mechanism design for efficient and stable renewable energy aggregation," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1033–1042, 2018.

[5] Y. Zhao, "Solar energy sharing in net metered community microgrids: Can the social goals be achieved?" in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2018, pp. 1–6.

[6] R. Henriquez-Auba, P. Pauli, D. Kalathil, D. S. Callaway, and K. Poolla, "The sharing economy for residential solar generation," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 7322–7329.

[7] P. Chakraborty, E. Baeyens, P. P. Khargonekar, K. Poolla, and P. Varaiya, "Analysis of solar energy aggregation under various billing mechanisms," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4175–4187, 2018.

[8] J. A. Taylor, D. S. Callaway, and K. Poolla, "Inventory control of storage in distribution systems," in *2012 American Control Conference (ACC)*. IEEE, 2012, pp. 2147–2152.

[9] J. Qin and R. Rajagopal, "Dynamic programming solution to distributed storage operation and de-

sign," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–5.

[10] W. Tushar, B. Chai, C. Yuen, S. Huang, D. B. Smith, H. V. Poor, and Z. Yang, "Energy storage sharing in smart grid: A modified auction-based approach," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1462–1475, 2016.

[11] D. Kalathil, C. Wu, K. Poolla, and P. Varaiya, "The sharing economy for the electricity storage," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 556–567, 2017.

[12] J. A. Taylor, J. L. Mathieu, D. S. Callaway, and K. Poolla, "Price and capacity competition in zero-mean storage and demand response markets," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2012, pp. 1316–1323.

[13] C. Wu, H. Mohsenian-Rad, and J. Huang, "Vehicle-to-aggregator interaction game," *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 434–442, 2011.

[14] S. Han, S. Han, and K. Sezaki, "Development of an optimal vehicle-to-grid aggregator for frequency regulation," *IEEE Transactions on smart grid*, vol. 1, no. 1, pp. 65–72, 2010.

[15] J. Sun, Y. Zhang, Y. Yu, and C. Wu, "Storage control for carbon emission reduction: Opportunities and challenges," *arXiv preprint arXiv:1912.10464*, 2019.

[16] Z. Wang, K. Wang, Y. Yu, and C. Wu, "An algorithmic view on optimal storage sizing," in *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2020, pp. 1–5.

[17] S. M. Kakade *et al.*, "On the sample complexity of reinforcement learning," Ph.D. dissertation, University of London London, England, 2003.

[18] S. Hanneke, "The optimal sample complexity of pac learning," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1319–1333, 2016.

[19] N. Golowich, A. Rakhlin, and O. Shamir, "Size-independent sample complexity of neural networks," in *Conference On Learning Theory*. PMLR, 2018, pp. 297–299.

[20] P. Xie, Z. Cai, P. Liu, X. Li, Y. Zhang, and D. Xu, "Microgrid system energy storage capacity optimization considering multiple time scale uncertainty coupling," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5234–5245, 2018.

[21] C. Opathella, A. Elkasrawy, A. A. Mohamed, and B. Venkatesh, "A novel capacity market model with energy storage," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5283–5293, 2018.

[22] T. Li and M. Dong, "Residential energy storage management with bidirectional energy control," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3596–3611, 2018.

[23] Y. Wang, X. Lin, and M. Pedram, "Adaptive control for energy storage systems in households with photovoltaic modules," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 992–1001, 2014.

[24] P. Harsha and M. Dahleh, "Optimal management and sizing of energy storage under dynamic pricing for the efficient integration of renewable energy," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1164–1181, 2014.

[25] N. Padmanabhan, M. Ahmed, and K. Bhattacharya, "Battery energy storage systems in energy and reserve markets," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 215–226, 2019.

[26] P. M. van de Ven, N. Hegde, L. Massoulié, and T. Salonidis, "Optimal control of end-user energy storage," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 789–797, 2013.

[27] A. Oudalov, R. Cherkaoui, and A. Beguin, "Sizing and optimal operation of battery energy storage system for peak shaving application," in *2007 IEEE Lausanne Power Tech*. IEEE, 2007, pp. 621–625.

[28] J. Wu, Z. Wang, Y. Yu, and C. Wu, "Optimal storage control for dynamic pricing," in *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*, 2020, pp. 395–397.

[29] J. Qin, Y. Chow, J. Yang, and R. Rajagopal, "Online modified greedy algorithm for storage control under uncertainty," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1729–1743, 2015.

[30] C.-K. Chau, G. Zhang, and M. Chen, "Cost minimizing online algorithms for energy storage management with worst-case guarantee," *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2691–2702, 2016.

[31] M. Lin, A. Wierman, A. Roytman, A. Meyerson, and L. L. Andrew, "Online optimization with switching cost," *ACM SIGMETRICS Performance Evaluation Review*, vol. 40, no. 3, pp. 98–100, 2012.

[32] Y. Li, G. Qu, and N. Li, "Using predictions in online optimization with switching costs: A fast algorithm and a fundamental limit," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 3008–3013.

[33] V.-H. Bui, A. Hussain, and H.-M. Kim, "Double deep q-learning-based distributed operation of battery energy storage system considering uncertainties," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 457–469, 2019.

[34] G. Henri and N. Lu, "A supervised machine learning approach to control energy storage devices," *IEEE Transactions on Smart Grid*, vol. 10, no. 6,

pp. 5910–5919, 2019.
[35] Y. Hu and B. Defourny, "Optimal price-threshold control for battery operation with aging phenomenon: a quasiconvex optimization approach," *Annals of Operations Research*, pp. 1–28, 2017.
[36] B. Xu, M. Korpås, A. Botterud, and F. O'Sullivan, "A lagrangian policy for optimal energy storage control," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 224–230.
[37] M. U. Hashmi, A. Mukhopadhyay, A. Bušić, and J. Elias, "Storage optimal control under net metering policies," *arXiv preprint arXiv:2002.01524*, 2020.
[38] C. Guo, Z. Huang, Z. G. Tang, and X. Zhang, "Generalizing complex hypotheses on product distributions: Auctions, prophet inequalities, and pandora's problem," *arXiv preprint arXiv:1911.11936*, 2019.
[39] G. Mena and J. Niles-Weed, "Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem," in *Advances in Neural Information Processing Systems*, 2019, pp. 4541–4551.
[40] Y. Chen and M. Wang, "Stochastic primal-dual methods and sample complexity of reinforcement learning," *arXiv preprint arXiv:1612.02516*, 2016.
[41] S. Dasgupta, "Coarse sample complexity bounds for active learning," in *Advances in neural information processing systems*, 2006, pp. 235–242.
[42] C. Guo, Z. Huang, and X. Zhang, "Settling the sample complexity of single-parameter revenue maximization," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, 2019, pp. 662–673.
[43] M.-F. F. Balcan, T. Sandholm, and E. Vitercik, "Sample complexity of automated mechanism design," in *Advances in Neural Information Processing Systems*, 2016, pp. 2083–2091.
[44] M.-F. Balcan, T. Sandholm, and E. Vitercik, "A general theory of sample complexity for multi-item profit maximization," in *Proceedings of the 2018 ACM Conference on Economics and Computation*, 2018, pp. 173–174.
[45] V. Syrgkanis, "A sample complexity measure with applications to learning optimal auctions," in *Advances in Neural Information Processing Systems*, 2017, pp. 5352–5359.
[46] P. Street, "Dataport: the world's largest energy data resource," *Pecan Street Inc*, pp. 01–01, 2015.
[47] "Time-of-Use price of OEB." Accessed: Feb. 2022. [Online]. Available: https://www.oeb.ca/rates-and-your-bill/electricity-rates/historical-electricity-rates

APPENDIX A
PROOF OF THEOREM IV.1

The demand d_i is realized at time i and is satisfied by $A_1(i)$ and $A_3(i)$:

$$d_i = A_1(i) + \mu_d A_3(i).$$

The energy remaining at the next time slot is then,

$$C_{i+1} = \gamma(C_i + \mu_c A_2(i) - A_3(i)).$$

The cost incurred is $(A_1(i) + A_2(i))p_i$.

Owing to the charging and discharging inefficiency, simultaneous charging and discharging are not profitable. This conclusion is drawn from an intuitive observation that to satisfy the demand, one would reduce their total cost by directly purchasing energy from the grid rather than buying energy to charge their battery and then discharging to meet the demand. With this in mind, we can reformulate $A_i(i)$ in terms of C_{i+1} as follows:

$$\begin{aligned} A_2(i) &= \mu_c^{-1}(\gamma^{-1}C_{i+1} - C_i) \cdot \mathbb{1}_{\gamma^{-1}C_{i+1} \geq C_i}, \\ A_3(i) &= -(\gamma^{-1}C_{i+1} - C_i) \cdot \mathbb{1}_{\gamma^{-1}C_{i+1} < C_i}, \\ A_1(i) &= d_i - \mu_d A_3(i) \\ &= d_i + \mu_d(\gamma^{-1}C_{i+1} - C_i) \cdot \mathbb{1}_{\gamma^{-1}C_{i+1} < C_i}. \end{aligned}$$

Thus, it suffices to only determine C_{i+1} to determine $A_j(i)$.

APPENDIX B
PROOF OF THEOREM IV.2

Lemma B.1. *$\forall i$, given d_i we have $h_1(C_{i+1}, C_i) = g_i(C_{i+1}, C_i, d_i)$ is convex.*

Proof. We first verify that the feasible domain of $h_1(\cdot)$ is a convex set. In fact, $\forall$ feasible (x_1, y_1) and (x_2, y_2), if $x_1 \in [\underline{U_i}(y_1), \overline{U_i}(y_1)]$ and $x_2 \in [\underline{U_i}(y_2), \overline{U_i}(y_2)]$, then it is easy to verify that $\forall \lambda \in [0, 1]$ we have $\lambda x_1 + (1 - \lambda)x_2 \in [\underline{U_i}(\lambda y_1 + (1 - \lambda)y_2), \overline{U_i}(\lambda y_1 + (1 - \lambda)y_2)]$ by definition. For example, if $x_1 \leq \gamma \cdot (y_1 + \overline{A_c})$ and $x_2 \leq \gamma \cdot (y_2 + \overline{A_c})$, then $\lambda x_1 + (1 - \lambda)x_2 \leq \gamma \cdot (\lambda y_1 + (1 - \lambda)y_2 + \overline{A_c})$. Similarly, the other terms can be verified.

Furthermore, because $\mu_d < 1 < \mu_c^{-1}$,

$$\begin{aligned} &(\gamma^{-1}C_{i+1} - C_i)^+ \mu_c^{-1} + (\gamma^{-1}C_{i+1} - C_i)^- \mu_d \\ = &\max\{(\gamma^{-1}C_{i+1} - C_i)\mu_c^{-1}, (\gamma^{-1}C_{i+1} - C_i)\mu_d\}, \end{aligned}$$

which is maximum for linear functions. Therefore, h is convex. □

Lemma B.2. *$\forall i$, $G_i(\cdot)$ is convex.*

Proof.

$$G_i(C_i) = \mathbb{E}_{d_i \sim D_i}\big[\min_{C_{i+1}^*}\{G_{i+1}(C_{i+1}^*) + g(C_{i+1}^*, C_i, d_i)\}\big].$$

We prove this by induction. This statement is valid for the last time slot n because $G_n(\cdot) = 0$. Assuming that $G_{i+1}(\cdot)$ is convex, we need to prove that $G_i(\cdot)$ is convex.

It is sufficient to prove that, given d_i, $h_2(C_i) = \min_{C^*_{i+1}}\{G_{i+1}(C^*_{i+1})+g(C^*_{i+1}, C_i, d_i)\}$ is convex. $G_i(\cdot)$ is non-negative weighted sum of convex functions; therefore, $G_i(\cdot)$ is also convex.

According to the induction hypothesis and Lemma B.1, $G_{i+1}(x) + g(x, y, d_i)$ is convex with respect to (x, y) pair. Assuming that the corresponding optimal solutions for $C_{i,1}$ and $C_{i,2}$ are $C^*_{i+1,1}$ and $C^*_{i+1,2}$, respectively, $\forall\lambda \in [0, 1]$

$$\begin{aligned}
&h_2(\lambda C_{i,1} + (1-\lambda)C_{i,2}) \le G_{i+1}(\lambda C^*_{i+1,1} + (1-\lambda)C^*_{i+1,2})\\
&\quad + g(\lambda C^*_{i+1,1} + (1-\lambda)C^*_{i+1,2}, \lambda C_{i,1} + (1-\lambda)C_{i,2}, d_i)\\
\le&\lambda \cdot (G_{i+1}(C^*_{i+1,1}) + g(C^*_{i+1,1}, C_{i,1}, d_i))\\
&\quad + (1-\lambda) \cdot (G_{i+1}(C^*_{i+1,2}) + g(C^*_{i+1,2}, C_{i,2}, d_i))\\
=&\lambda \cdot h_2(C_{i,1}) + (1-\lambda) \cdot h(C_{i,2}),
\end{aligned}$$

which complete the proof. □

Using Lemma B.2, because $[\underline{U_i}, \overline{U_i}]$ is compact, a minimum element exists, and it is unique. Hence, C^*_{i+1} is well-defined. Further, it can be shown that $\forall i, G_i(\cdot)$ is non-increasing.

Lemma B.3. *$\forall i$, $G_i(\cdot)$ is non-increasing.*

Proof. This is proven by induction. First, $G_n(\cdot) = 0$ and is non-increasing. Assuming that $G_{i+1}(\cdot)$ is non-increasing, we prove the following stronger proposition: If $C_{i,1} < C_{i,2}$ and their optimal solutions are $C^*_{i+1,1}$ and $C^*_{i+1,2}$, respectively, then $\forall d_i$

$$\begin{aligned}
&G_{i+1}(C^*_{i+1,1}) + g(C^*_{i+1,1}, C_{i,1}, d_i)\\
\ge&G_{i+1}(C^*_{i+1,2}) + g(C^*_{i+1,2}, C_{i,2}, d_i).
\end{aligned}$$

$g(\cdot, \cdot, \cdot)$ is non-decreasing with respect to the first dimension and non-increasing with respect to the second dimension. If $C^*_{i+1,1} \in [\underline{U_i}(C_{i,2}), \overline{U_i}(C_{i,2})]$, then

$$\begin{aligned}
&G_{i+1}(C^*_{i+1,2}) + g(C^*_{i+1,2}, C_{i,2}, d_i)\\
\le&G_{i+1}(C^*_{i+1,1}) + g(C^*_{i+1,1}, C_{i,2}, d_i)\\
\le&G_{i+1}(C^*_{i+1,1}) + g(C^*_{i+1,1}, C_{i,1}, d_i),
\end{aligned}$$

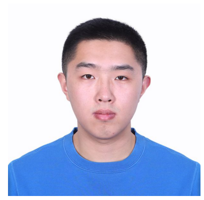
where the last inequality arises from the monotonicity of $g(\cdot, \cdot, \cdot)$.

However, when $C^*_{i+1,1} \notin [\underline{U_i}(C_{i,2}), \overline{U_i}(C_{i,2})]$, $C^*_{i+1,1} < \underline{U_i}(C_{i,2})$,

$$\begin{aligned}
&G_{i+1}(C^*_{i+1,2}) + g(C^*_{i+1,2}, C_{i,2}, d_i)\\
\le&G_{i+1}(\underline{U_i}(C_{i,2})) + g(\underline{U_i}(C_{i,2}), C_{i,2}, d_i)\\
\le&G_{i+1}(C^*_{i+1,1}) + g(C^*_{i+1,1}, C_{i,1}, d_i),
\end{aligned}$$

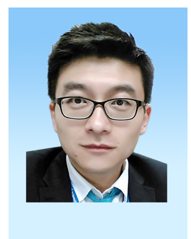
where the last inequality arises from the monotonicity of $g(\cdot, \cdot, \cdot)$ and the induction hypothesis. □

With these two lemmas in hand, a similar method can be used as in [26] to prove Theorem IV.2, which is omitted in this paper.



Zhiqi Wang (Student Member, IEEE) received the B.S. and M.S. degree from Institution for Interdisciplinary Information Sciences, Tsinghua University, in 2019 and 2022, respectively. He is pursuing the Ph.D. degree at the Institute for Theoretical Computer Science, Shanghai University of Finance and Economics. His research interests include algorithmic game theory and storage control in smart grids.

Zeren Tan (Student Member, IEEE) received the B.S. degree from Civil Engineering, Tsinghua University, in 2020. He is pursuing the Ph.D. degree in Institution for Interdisciplinary Information Sciences, Tsinghua University. His research interests lie in the robustness of machine learning models and interpretable AI.



Yang Yu (Member, IEEE) received the Ph.D. Degree from Stanford University and the B.S. Degree from Nanjing University. He is an Assistant Professor with Institution for Interdisciplinary Information Sciences, Tsinghua University. His research focuses on smart regulation of the future electricity market and market designs for energy transmission.