

RESEARCH

Open Access



A cooperative demand-response framework for day-ahead optimization in battery pools

Georgios C. Chasparis^{1*}, Mario Pichler¹, Johanna Spreitzhofer² and Tara Esterl²

From The 8th DACH+ Conference on Energy Informatics
Salzburg, Austria. 26–27 September, 2019

*Correspondence:

georgios.chasparis@scch.at

¹Software Competence Center
Hagenberg GmbH, Softwarepark 21,
4232 Hagenberg, Austria
Full list of author information is
available at the end of the article

Abstract

The constantly increasing electricity and energy demand in residential buildings, as well as the need for higher absorption rates of renewable sources of energy, demand for an increased flexibility at the end-users. This need is further reinforced by the rising numbers of residential Photovoltaic (PV) and battery-storage systems. In this case, flexibility can be viewed as the excess energy that can be charged to or discharged from a battery, in response to a group objective of several such battery-storage systems (aggregation).

One such group objective considered in this paper includes marketing flexibility (charging or discharging) to the Day-ahead (DA) spot market, which can provide both a) financial incentives to the owners of such systems, and b) an increase in the overall absorption rates of renewable energy. The responsible agent for marketing and offering such flexibility, herein aggregator, is directly controlling the participating batteries, in exchange to some financial compensation of the owners of these batteries. In this paper, we present an optimization framework that allows the aggregator to optimally exchange the available flexibility to the DA market. The proposed scheme is based upon a reinforcement-learning approach, according to which the aggregator learns through time an optimal policy for bidding flexibility to the DA market. By design, the proposed scheme is flexible enough to accommodate the possibility of erroneous forecasts (of weather, load or electricity price). Finally, we evaluate our approach on real-world data collected from currently installed battery-storage systems in Upper Austria.

Keywords: Demand-response, Aggregation, Reinforcement learning, Day-ahead spot market

Introduction

A recent trend in home automation is a constant increase in the number of battery-storage systems (Kairies et al. 2016). So far, such storage systems are mainly used to maximize the on-site absorption of the PV generation. Given the current need for increasing the percentage of renewable energy fed into the grid, the available charging/discharging flexibility potential can also be used to react to price variations in the Day-ahead (DA) or the Intra-day (ID) electricity spot-market. The responsible party (or

aggregator) could make such decisions over the specific use of the participating storage units with respect to the optimal participation in the DA and ID spot markets, the benefits of which are then transferred to the owners of the participating batteries.

In this paper, we focus on addressing the problem of optimal participation of a set of residential battery-storage systems in the DA market. The underlying assumption is that the aggregator (of a pool of batteries) may directly control the operation of the participating batteries if required, thus any charging/discharging flexibility potential can be extracted in real-time. The proposed scheme will be based upon an approximate-dynamic-programming (or reinforcement-learning) methodology. According to this scheme, an approximation function of the performance is being trained (using historical data) that can be used to generate optimal biddings/schedules for the DA market. By design, the proposed scheme is flexible enough to accommodate the possibility of erroneous forecasts as well as the need for re-optimizing in real-time upon receipt of corrected/updated forecasts.

It is worth noting that the proposed framework is part of the Flex+ project (Flex+). In this project, several component-pools consisting of a single technology (amongst heat pumps, boilers, battery-storage systems and e-cars) are enabled to participate in the short-term electricity markets, whilst also considering the interests of the consumers (e.g., comfort, eco-friendliness, greediness, etc.).

The remainder of this paper is organized as follows. In “[Related work and contributions](#)” section, we present related work and the main contributions of this paper. “[Participation costs and notation](#)” section presents the notion of participation costs in the flexibility extraction program. “[System dynamics](#)” section presents a detailed formulation of the overall system dynamics, and “[Optimal activation](#)” section addresses the problem of (instantaneous) optimal activation of a set of batteries, given a desirable amount of flexibility extraction (charging/discharging). Both the system dynamics and the instantaneous optimal activation are essential parts for formulating the problem of optimal flexibility bidding in the DA market (briefly, *DA-optimization*), which is presented in “[Day-ahead \(DA\) optimization](#)” section. In “[Approximate dynamic programming \(ADP\) for DA optimization](#)” section, we propose a reinforcement-learning scheme specifically tailored for addressing the DA-optimization, the efficiency of which is evaluated on real-world data. Finally, “[Conclusions and future work](#)” section presents concluding remarks.

Related work and contributions

With the constantly increasing renewable generation, users need to be flexible in adjusting their energy consumption, giving rise to *demand response* mechanisms. Demand response refers to the ability of each user to respond to certain requests reported by the network operator. This is usually performed either in the form of a *commitment* of the consumer to reduce load during peak hours (Ruiz et al. 2009; Chen et al. 2014) or by introducing *financial incentives* that affect prices during peak hours (Herter 2007; Triki and Violi 2009; Xu et al. 2016). For example, a *commitment-based approach* has been proposed by Chen et al. (2014), in which the operator distributes portions of its desired aggregated demand in the household users, using an average consensus algorithm. In particular, each one of the households receives a local demand objective which may only be fulfilled through the adjustment of its own flexible loads. On the other hand, an

incentive-based approach has been proposed by Xu et al. (2016), where each participating household communicates to the operator a bidding curve, that is a function that provides the load adjustment that each user is willing to perform at a given price. Then, the group operator computes the clearing prices, so that the overall cost of the participating households is minimized while achieving the desired demand adjustment. A similar approach to the one proposed in this paper, including also battery charging/discharging control, is considered in Nguyen et al. (2015). However, there is no specific centralized operator objective for either production/consumption. Instead, the objective in Nguyen et al. (2015) is to drive all instantaneous energy demand close to the average demand of the network.

Apart from these approaches (*commitment-based* or *incentive-based*), there is an alternative methodology which can be considered as a combination of the two and it is the one employed in this paper. According to such methodology, an aggregator directly extracts the required flexibility from the participating equipment when necessary. In return the aggregator offers to the owners of the participating equipment an agreed financial compensation. Such methodology is usually referred to as *demand-response aggregation* (Parvania et al. 2013). It has been employed in Parvania et al. (2013), where aggregators can activate load reduction in a set of consumers according to an agreed demand-response strategy for each consumer. Similar in spirit is also the work in references (Iria et al. 2017; Nan et al. 2018), where an aggregator directly controls a set of different types of loads in residential buildings to reduce total electricity consumption. As expected, a feature that distinguishes *demand-response aggregation* is the self-scheduling or activation optimization problem, that is the optimization of optimally utilizing the available flexibility (stemming from several households) over a future time horizon. Such feature (of multiple households) is not usually considered in the context of participation in a wholesale electricity market (see, e.g., (Gomez-Villalva and Ramos 2003; Philpott and Pettersen 2006)).

In the context of battery-storage systems, *demand-response aggregation* (as discussed in the previous paragraph) has not yet been addressed in an effective and computationally efficient way. In this context, the aggregator wishes to compute an optimal (day-ahead) schedule for extracting flexibility (charge, discharge or do nothing) for each one of the participating batteries. So far, such optimization problem has mostly been addressed for a *single* battery system, e.g., (Mohsenian-Rad 2016; He et al. 2016). Existing methodologies also include a detailed modeling of the battery as well as a detailed description of the cycle life costs of the battery due to the frequent charging/discharging (He et al. 2016). It may include computations of optimal bids for the day-ahead electricity market, as in Mohsenian-Rad (2016); He et al. (2016), or the intra-day/hour-ahead electricity market, as in Jiang and Powell (2015). In order to effectively address the uncertainty of the initial/final stage-of-charge of the battery, reference (Jiang and Powell 2015) also employs an approximate dynamic programming formulation.

In this paper, participation in the day-ahead wholesale electricity market is implemented by directly controlling the battery-storage systems, as in Mohsenian-Rad (2016); He et al. (2016); Jiang and Powell (2015). The specific contributions and novelty of this paper can be summarized as follows:

- (C1) In comparison to the work on demand-response aggregation, the aggregator derives specific optimal scheduling (or activation) strategies for the set of participating

- batteries over the duration of the next day. The specifications and cost functions of the participating batteries are directly included in the optimization (contrary to a generic load reduction strategy, e.g., (Parvania et al. 2013)).
- (C2) In comparison to the work on day-ahead optimal bidding of flexibility available in battery-storage systems, we consider aggregators that can directly control multiple battery-storage systems simultaneously. This extends prior work presented in Mohsenian-Rad (2016); He et al. (2016); Jiang and Powell (2015) which are restricted to a *single* battery.
- (C3) In order to address the complexity of deriving optimal bidding schedules for multiple battery-storage systems simultaneously, we introduce a novel optimization method that hierarchically decomposes the problem into a two-level optimization problem: a) *at the upper-level*, we compute the optimal (aggregate) flexibility (charging/discharging) that can be extracted from the set of available batteries for each 15-min interval over the duration of the next day, b) *at the lower-level*, we compute the optimal activation strategy of the participating batteries in order to generate the overall flexibility determined by the upper level (a) for each 15-min interval. Naturally, this decomposition significantly reduces the computational complexity of the problem, contrary to Iria et al. (2017); Nan et al. (2018) where these two computations are treated simultaneously.
- (C4) The computation of the flexibility bids/schedules is based upon an Approximate Dynamic Programming (ADP) methodology specifically tailored to this problem. An ADP approach can naturally incorporate uncertainties or erroneous forecasts, while it provides optimal strategies (instead of direct schedules). Thus, the schedules can directly be updated in case of updated forecasts. An ADP optimization approach also offers an attractive computational efficiency, as it will be discussed in detail in the forthcoming “Discussion” section.

Participation costs and notation

The participation costs are the costs experienced by the owners of the battery-storage systems due to the intervention of the aggregator through the DA flexibility schedules. We will consider a generic form of participation costs that can be used to model *discomfort* of the participants. Discomfort could be expressed with respect to the weight that the battery owner puts in certain preferences, e.g., *autarky* (i.e., maintaining a high state-of-charge at all times), *eco-friendliness* (i.e., priority on charging the battery only with PV generation), *greediness* (i.e., always selling available PV or battery energy), etc.

The optimization horizon (next day) is divided into time intervals of duration ΔT . We enumerate the resulting time intervals (T in total) using index $t = 1, \dots, T$. Let N be the total number of participating battery-storage systems and let $i \in \{1, \dots, N\}$ be a representative element of this set. The parameter $u_i(t) \in [-1, 1]$ will denote the *activation factor* of battery i at time t . If $u_i(t) \geq 0$, then $|u_i(t)|$ expresses the percentage of the *charging flexibility potential* that is activated. Analogously, if $u_i(t) \leq 0$, then $|u_i(t)|$ expresses the percentage of the *discharging flexibility potential* that is activated. The quantities $v_{c,i}(t) \geq 0, v_{d,i}(t) \leq 0$ will denote the *charging and discharging flexibility potential that is available in household i* , respectively. We will also denote $\mathbf{v}_c(t) = [v_{c,i}(t)]_i$ and $\mathbf{v}_d(t) = [v_{d,i}(t)]_i$ as the vectors of the available positive and negative flexibility potential, respectively. Finally, let $V_d(t) \doteq \mathbf{1}^T \mathbf{v}_d(t) \leq 0$ and

$V_c(t) \doteq \mathbf{1}^T \mathbf{v}_c(t) \geq 0$) be the total discharging and charging flexibility potential, respectively.

The user’s participation costs could be expressed through a set of bidding curves. In particular, we introduce the positive constants $\beta_{c,i}(t), \beta_{d,i}(t)$, that represent the cost that the provider pays to the participating household per unit of charging and discharging potential retrieved at time interval t , respectively. For example, if the user values *autarky* a lot, then we can select a large $\beta_{d,i}(t)$ when the state-of-charge of the battery is high, and small $\beta_{c,i}(t)$ when the state-of-charge is low. Then, when there is a decision for a charging activation of user i at time interval t , i.e., $u_i(t) > 0$, then $u_i(t)v_{c,i}(t)\beta_{c,i}(t) > 0$ represents the cost of activation. Analogously, in case of a discharging activation of user i at time interval t , i.e., $u_i(t) < 0$, then $u_i(t)v_{d,i}(t)\beta_{d,i}(t) > 0$ represents the cost of activation.

We will denote $E_i(t)$ to be the energy charged to or discharged from household i . We will also adopt the convention that the energy is positive if it is charged to the battery and negative otherwise. In other words, if $u_i(t) \geq 0$, then $E_i(t) = u_i(t)v_{c,i}(t) \geq 0$ (energy is charged to household i), and if $u_i(t) \leq 0$, then $E_i(t) = -u_i(t)v_{d,i}(t) \leq 0$ (energy is discharged from household i). In several cases, we will denote by $E(t)$ as the total energy charged/discharged from all participating households, i.e., $E(t) \doteq \sum_{i=1}^N E_i(t)$.

System dynamics

The $SOC_i(t)$ (*state-of-charge*) of the participating battery-storage systems, $i = 1, \dots, N$, together with the *exogenous* parameters of $P_{PV,i}(t)$ (*PV power-generation*) and $P_{load,i}(t)$ (*load consumption*), and the *activation factors* (or *control parameters*) $u_i(t)$, are sufficient to determine the evolution of the state-of-charge through time. We will often refer to this update mechanism as the *system dynamics*, which are also visualized in Fig. 1. We use the notation $\Delta P_i(t) \doteq P_{PV,i}(t) - P_{load,i}(t)$ to denote the excess PV generation. Furthermore, for any quantity x , the notation “ \hat{x} ” will denote either estimated or forecast quantity.

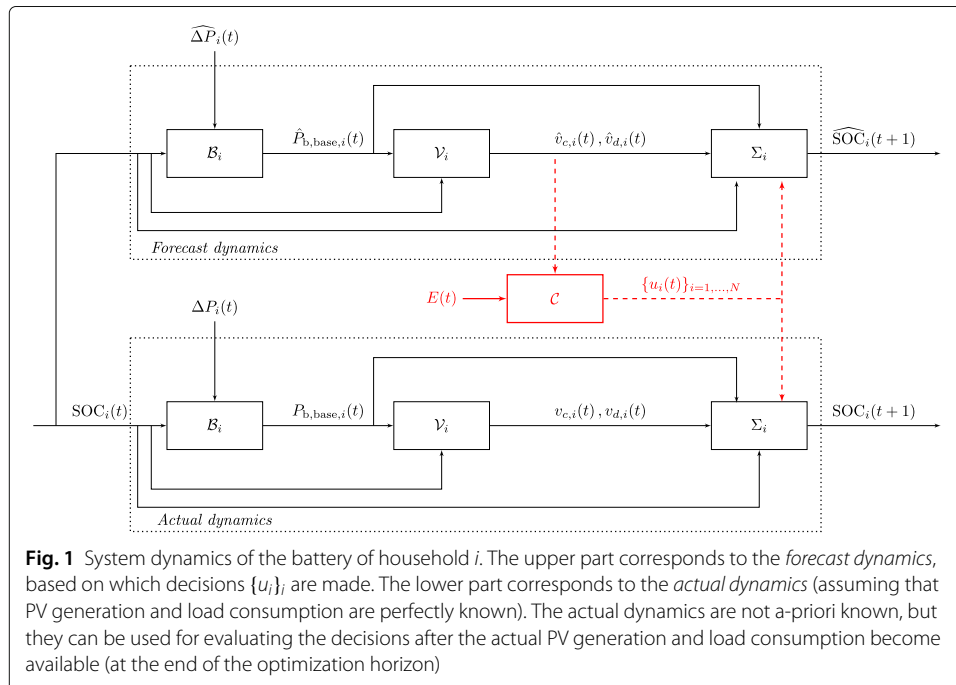


Fig. 1 System dynamics of the battery of household i . The upper part corresponds to the *forecast dynamics*, based on which decisions $\{u_i\}_i$ are made. The lower part corresponds to the *actual dynamics* (assuming that PV generation and load consumption are perfectly known). The actual dynamics are not a-priori known, but they can be used for evaluating the decisions after the actual PV generation and load consumption become available (at the end of the optimization horizon)

The system dynamics comprises three main procedures (operations), namely:

- *Baseline computation* which refers to the computation of the power from/to the battery under normal operation (i.e., when no additional charging or discharging potential is activated by the aggregator). This procedure is described by Algorithm 1.
- *Flexibility potential computation*, that computes the total amount of energy that can be charged or discharged to the participating households. This procedure is described by Algorithm 2.
- *State-of-Charge update*, that computes the new state-of-charge (at the beginning of the next time interval $t + 1$), when energy commitment level $u_i(t)$ has been assigned. This procedure is described by Algorithm 3.

Regarding the *baseline computation*, we assume here that the participating battery-storage systems follow the following simple strategy: a) if $\Delta P_i(t) \geq 0$ (i.e., PV power generation is larger than the load consumption), then any excess PV power generation is first used to charge the battery, and any additional power is fed into the grid, b) if $\Delta P_i(t) \leq 0$ (i.e., the PV power is less than the load consumption), then the excess load consumption is first covered by the battery and if not sufficient additional power is withdrawn from the grid. We may argue that this criterion values autarky and eco-friendliness, since at times of energy excess the priority is to charge the battery, while at times of energy shortage, the priority is to use the energy available in the battery. Alternative criteria may also be employed, e.g., an economic optimum, where the objective would have been the maximization of the monetary utility of the participant.

Algorithm 3 defines the system dynamics update (i.e., the update of the $\text{SOC}_i(t)$) and uses the results of Algorithms 1 and 2 for the baseline and flexibility computation. In summary, given the current state-of-charge $\text{SOC}_i(t)$ (at the beginning of time interval t) and the planned flexibility extraction $u_i(t)$ for that interval, updating the state-of-charge is performed as follows:

$$P_{b,\text{base},i}(t) = \mathcal{B}_i(\Delta P_i(t), \text{SOC}_i(t)) \quad (1a)$$

$$[v_{d,i}(t), v_{c,i}(t)] = \mathcal{V}_i(P_{b,\text{base},i}(t), \text{SOC}_i(t)) \quad (1b)$$

$$\text{SOC}_i(t + 1) = \Sigma_i(\text{SOC}_i(t), v_{d,i}(t), v_{c,i}(t), u_i(t)). \quad (1c)$$

Since we will be addressing a DA-optimization, we will particularly be interested in updating the total flexibility potential, which leads to a different ordering of the above sequence of equations. In particular, given $v_{d,i}(t)$, $v_{c,i}(t)$, $\text{SOC}_i(t)$ and $u_i(t)$, we have:

$$\text{SOC}_i(t + 1) = \Sigma_i(\text{SOC}_i(t), v_{d,i}(t), v_{c,i}(t), u_i(t)) \quad (2a)$$

$$P_{b,\text{base},i}(t + 1) = \mathcal{B}_i(\Delta P_i(t + 1), \text{SOC}_i(t + 1)) \quad (2b)$$

$$[v_{d,i}(t + 1), v_{c,i}(t + 1)] = \mathcal{V}_i(P_{b,\text{base},i}(t + 1), \text{SOC}_i(t + 1)). \quad (2c)$$

In this case, we will refer to the overall available potentials $\{V_d(t), V_c(t)\}$ as the **state variables** of the overall process and $\mathbf{u}(t) = [u_i(t)]_i$ as the **control variables**. The terms $\Delta P \doteq [\Delta P_i(t)]_i$, which capture the available excess of power available in all households, are considered exogenous parameters or **disturbances**. To summarize, we can write the state dynamics of Eq. 2 more compactly as:

$$[v_{d,i}(t + 1), v_{c,i}(t + 1)] = \Phi_i(v_{d,i}(t), v_{c,i}(t), u_i(t), \Delta P_i(t)), \quad i = 1, \dots, N, \quad (3)$$

Algorithm 1 Baseline power to battery i at time interval t

```

1: procedure  $\mathcal{B}_i(\Delta P_i(t), \text{SOC}_i(t))$ 
2:   Step 1a. maximum possible (average) charging power to battery
3:    $\Delta T_{c,i}^*(t) = (1 - \text{SOC}_i(t)) \cdot \frac{\chi_i}{c_{\max,i} \eta_{c,i}}$   $\triangleright$  We make the assumption here that the
   maximum charging power,  $c_{\max,i}$ , is evaluated at the output of the battery. Thus, the actual
   charging power (that directly affects the  $\text{SOC}_i(t)$ ) will be a bit smaller than  $c_{\max,i}$  due to the
   energy losses during charging (captured by the efficiency rate  $\eta_{c,i}$ ).  $\Delta T_{c,i}^*(t)$  is the minimum
   time needed for the battery to reach its full energy capacity, denoted by  $\chi_i$ , at the maximum
   charging power  $c_{\max,i}$ .
4:    $P_{b,c,\max,i}(t) = c_{\max,i} \cdot \min\{\Delta T_{c,i}^*, \Delta T\} / \Delta T$   $\triangleright$  This is the maximum (average over a  $\Delta T$ 
   time-interval) charging power that could be achieved within  $\Delta T$ .
5:   Step 1b. maximum possible (average) discharging power to battery
6:    $\Delta T_{d,i}^*(t) = \text{SOC}_i(t) \cdot \frac{\chi_i \eta_{d,i}}{d_{\max,i}}$   $\triangleright$  Analogously to the
   charging case, we make the assumption here that the maximum discharging power, denoted
   by  $d_{\max,i}$ , is evaluated at the output of the battery. Thus, the actual discharging power (that
   directly affects the  $\text{SOC}_i(t)$ ) will be a bit larger than  $d_{\max,i}$  due to the energy losses during
   discharging (captured by the efficiency rate  $\eta_{d,i}$ ).  $\Delta T_{d,i}^*$  is the minimum time needed for the
   battery to empty at the maximum discharging speed  $d_{\max,i}$ .
7:    $P_{b,d,\max,i}(t) = -d_{\max,i} \cdot \min\{\Delta T_{d,i}^*, \Delta T\} / \Delta T$   $\triangleright$  This is the maximum (average over a  $\Delta T$ 
   time-interval) discharging power that could be achieved within  $\Delta T$ .
8:   Step 2. baseline power to the battery
9:   if  $\Delta P_i(t) \geq 0$  then
10:     $P_{b,\text{base},i}(t) = \min\{P_{b,c,i}(t), \Delta P_i(t)\} \geq 0$   $\triangleright$  if there is an excess of power, a battery is
    always charged at maximum possible (average) power
11:   else
12:     $P_{b,\text{base},i}(t) = \max\{P_{b,d,i}(t), \Delta P_i(t)\} \leq 0$   $\triangleright$  if there is a shortage of power, the battery
    is always discharged at maximum possible (average) power
13:   return  $P_{b,\text{base},i}(t)$   $\triangleright$  baseline power to battery at interval  $t$ 

```

Algorithm 2 Flexibility potential of household i at time interval t

```

1: procedure  $\mathcal{V}_i(P_{b,\text{base},i}(t), \text{SOC}_i(t))$ 
2:   Maximum charging (flexibility) potential
3:    $v_{c,i}(t) = [P_{b,c,\max,i}(t) - P_{b,\text{base},i}]_+ \cdot \Delta T$ 
4:   Maximum discharging (flexibility) potential
5:    $v_{d,i}(t) = [P_{b,d,\max,i}(t) - P_{b,\text{base},i}]_- \cdot \Delta T$ 
6:   return  $v_{c,i}(t), v_{d,i}(t)$   $\triangleright$  flexibility potential at interval  $t$ 

```

from which we can derive an aggregate flexibility update recursion of the form:

$$[V_d(t+1), V_c(t+1)] = \overline{\Phi}(V_d(t), V_c(t), \mathbf{u}(t), \Delta P(t)), \quad (4)$$

which updates the aggregate flexibility potential.

Optimal activation

In this section, we address the optimization problem of the *optimal activation* of some given flexibility (positive or negative), $E(t)$, over a time interval $t \in \{1, \dots, T\}$. That is, we wish to compute the set of batteries that should be activated to provide total energy $E(t)$, as well as the amount of flexibility extracted from each activated battery. This optimization problem is solved centrally by an aggregator which tries to minimize the overall activation or participation cost (defined in “[Participation costs and notation](#)” section through the positive cost factors $\beta_{c,i}, \beta_{d,i}$). The solution of this optimization problem is the basis of the upcoming day-ahead market optimization, and it has been

Algorithm 3 State-of-Charge of battery i at time $t + 1$

```

1: procedure  $\Sigma_i(\text{SOC}_i(t), v_{d,i}(t), v_{c,i}(t), u_i(t))$ 
2:   Step 1. average power to battery under  $u_i(t)$ ,
3:   if  $u_i(t) > 0$  then
4:      $P_{b,i}(t) = P_{b,\text{base},i}(t) + u_i(t) \cdot v_{c,i}(t)/\Delta T$   $\triangleright$  Charging flexibility potential  $v_{c,i}(t)$  during
     time interval  $t$  has been computed by first applying Algorithm 1 to compute  $P_{b,\text{base},i}(t)$  and
     then applying Algorithm 2 for computing the available charging potential  $v_{c,i}(t)$ 
5:   else
6:      $P_{b,i}(t) = P_{b,\text{base},i}(t) - u_i(t) \cdot v_{d,i}(t)/\Delta T$   $\triangleright$  Discharging flexibility potential  $v_{d,i}(t)$ 
     during time interval  $t$  has been computed by first applying Algorithm 1 to compute  $P_{b,\text{base},i}$ 
     and then applying Algorithm 2 for computing the available discharging potential  $v_{d,i}(t)$ 
7:   Step 2. State-of-Charge at time  $t + 1$ 
8:   if  $P_{b,i}(t) > 0$  then
9:      $\text{SOC}_i(t + 1) = \text{SOC}_i(t) + \eta_{c,i} \frac{P_{b,i}(t)\Delta T}{\chi_i}$ 
10:  else
11:     $\text{SOC}_i(t + 1) = \text{SOC}_i(t) + \frac{P_{b,i}(t)\Delta T}{\eta_{d,i}\chi_i}$ 
12:  return  $\text{SOC}_i(t + 1)$   $\triangleright$  State-of-Charge at  $t + 1$ 

```

presented in detail in Chasparis et al. (2019). It is evident that any flexibility potential $E(t)$ extracted from the pool of batteries, at any given time interval t , should satisfy: $V_d(t) \leq E(t) \leq V_c(t)$.

Forecast flexibility potential

The optimal activation problem is driven by our forecast energy potential for the next day. It is evident that we do not have knowledge of the actual flexibility potential during the time interval t of the next day, since this depends on the actual PV generation $P_{PV,i}(t)$ and the actual load consumption $P_{\text{load},i}(t)$. For this reason, any optimization for computing the optimal activation for time interval t may only be based upon the *forecast* (or *predicted*) flexibility potential, $\hat{v}_{c,i}(t)$ and $\hat{v}_{d,i}(t)$. Analogously, we define the overall forecast charging and discharging flexibility potential as $\hat{V}_c(t)$ and $\hat{V}_d(t)$, respectively. Thus, the above activation constraint should be replaced by $\hat{V}_d(t) \leq E(t) \leq \hat{V}_c(t)$.

Activation optimization

For each battery $i = 1, \dots, N$, we wish to compute the *activation factor* $u_i(t) \in [-1, 1]$ for time interval t . This computation is provided by a function of the form $\mathcal{O}_{\text{act},t}(E(t), \hat{v}_d(t), \hat{v}_c(t))$, which accepts as inputs, the desired total activation $E(t)$ and the available forecast flexibility potential of the participating batteries. The output of this function will be the optimal activation factors $\{u_i^*(t)\}_i$ for each one of the participating batteries $i = 1, \dots, N$ at time interval t . In other words, the optimal activation optimization is summarized by $\mathcal{O}_{\text{act},t} : [\hat{V}_d(t), \hat{V}_c(t)] \times \mathbb{R}_- \times \mathbb{R}_+ \mapsto [-1, 1]^N$, such that:

$$\mathbf{u}^*(E(t), \hat{v}_d(t), \hat{v}_c(t)) \in \mathcal{O}_{\text{act},t}(E(t), \hat{v}_d(t), \hat{v}_c(t)), \quad (5)$$

where \mathbf{u}^* denotes the optimal activation. At the optimal activation of a positive flexibility ($E(t) > 0$), we should expect that $\mathbf{u}^* \geq 0$ (elementwise). In other words, given that a positive activation has a strictly positive activation cost, a positive flexibility can only be optimally extracted through positive activations. Analogously, when $E(t) < 0$, then $\mathbf{u}^* \leq 0$ (elementwise). To this end, we can decompose the problem of optimal activation as follows:

$$\mathbf{u}^*(E, \hat{\mathbf{v}}_d, \hat{\mathbf{v}}_c) \in \mathcal{O}_{\text{act},t}(E, \hat{\mathbf{v}}_d, \hat{\mathbf{v}}_c) = \begin{cases} \mathcal{O}_{\text{act},d,t}(E, \hat{\mathbf{v}}_d), & \text{if } E \leq 0, \\ \mathcal{O}_{\text{act},c,t}(E, \hat{\mathbf{v}}_c), & \text{if } E > 0. \end{cases} \quad (6)$$

In particular, let us first consider the case of a negative (or discharging) desirable energy activation, i.e., $E < 0$. In this case, the optimal activation is the solution to the following optimization problem:

$$\mathcal{O}_{\text{act},d,t}(E, \hat{\mathbf{v}}_d(t)) : \begin{cases} \arg \min & C_{d,t}(\mathbf{u}, \hat{\mathbf{v}}_d(t)) \doteq \sum_{i=1}^N u_i v_{d,i}(t) \beta_{d,i}(t) \\ \text{s.t.} & \sum_{i=1}^N u_i \hat{v}_{d,i}(t) = -E \\ \text{var.} & u_i \in [-1, 0], i = 1, 2, \dots, N. \end{cases} \quad (7)$$

Analogously, for the case of positive (or charging) desirable activation $E > 0$, the optimization takes on the following form:

$$\mathcal{O}_{\text{act},c,t}(E, \hat{\mathbf{v}}_c(t)) : \begin{cases} \arg \min & C_{c,t}(\mathbf{u}, \hat{\mathbf{v}}_c(t)) \doteq \sum_{i=1}^N u_i v_{c,i}(t) \beta_{c,i}(t) \\ \text{s.t.} & \sum_{i=1}^N u_i \hat{v}_{c,i}(t) = E \\ \text{var.} & u_i \in [0, 1], i = 1, 2, \dots, N. \end{cases} \quad (8)$$

Optimal activation algorithm

Reference (Chasparis et al. 2019) provides an algorithm for computing the optimal activation (cf., (Chasparis et al. 2019, Algorithm 1)), and takes the form of a merit-order optimization. We will present here only the case of discharging activation, i.e., when $E(t) < 0$, since the charging activation will be similar.

We first order the set of participating batteries in ascending order, with respect to the cost coefficient $\beta_{d,i}(t)$. In other words, we order the batteries as follows: $\beta_{d,1}(t) \leq \beta_{d,2}(t) \leq \dots \leq \beta_{d,N}(t)$. Define also the function $\kappa_d : [\hat{V}_d(t), \hat{V}_c(t)] \times \mathbb{R}_+^N \mapsto \mathbb{N}$, such that $\kappa_d + 1$ corresponds to the minimum number of batteries that should be activated in order to generate a total discharging flexibility potential of $E(t) < 0$. In other words,

$$\kappa_d(E, \hat{\mathbf{v}}_d(t)) \doteq \max \left\{ k \in \{1, 2, \dots, N\} : 0 \geq \sum_{i=1}^k \hat{v}_{d,i}(t) > E \right\}. \quad (9)$$

Algorithm 4 Optimal activation for generating discharging flexibility $E(t) < 0$

```

1: procedure  $\mathcal{O}_{\text{ACT},d,t}(E(t), \hat{\mathbf{v}}_d(t))$ 
2:   order participants  $i = 1, 2, \dots, N$  in ascending order  $\beta_{d,1}(t) \leq \beta_{d,2}(t) \leq \dots \leq \beta_{d,N}(t)$ 
3:   for  $i = 1, 2, \dots, N$  do
4:     if  $i \leq \kappa_d(E(t), \hat{\mathbf{v}}_d(t))$  then
5:        $u_i^*(t) = -1$ 
6:     else
7:       if  $i = \kappa_d(E(t), \hat{\mathbf{v}}_d(t)) + 1$  then
8:          $u_i^*(t) = -\left(E(t) - \sum_{j=1}^{i-1} \hat{v}_{d,j}(t)\right) / \hat{v}_{d,i}(t)$   $\triangleright$  By definition of  $\kappa_d$ , the numerator
           defines the remaining energy that should be extracted from battery  $\kappa_d + 1$ .
9:       else
10:         $u_i^*(t) = 0$ 
11:   return  $\mathbf{u}^*(t)$ 

```

It is straightforward to show (using strong duality) that the optimal activation is given by Algorithm 4.

Proposition 1 (Proposition 7.2 in Chasparis et al. (2019)) *The activation \mathbf{u}^* computed by Algorithm 4 is an optimal solution to optimization (7).*

Day-ahead (DA) optimization

Optimization under perfect forecasts

In this section, we present the optimization of flexibility extraction $\{E(t)\}_t$ (positive or negative), during the next day $t = 1, \dots, T$, that would be optimal for the aggregator to provide to the DA market. The aggregator would like to exploit the variations in the DA price over the next day in order to increase its revenues.

In this subsection, we will assume that the aggregator has perfect knowledge over the dynamics of the day-ahead flexibility potential of the participating batteries, i.e., all the required forecasts are assumed perfectly known. In this case, the optimization problem for computing the optimal activations, given some exogenous sequence of the clearing DA prices, denoted by $\{p_{\text{DA}}(t)\}_t$, will be of the following form:

$$\mathcal{O}_{\text{DA}} \left(\{\mathbf{v}_d(t)\}_{t=1}^T, \{\mathbf{v}_c(t)\}_{t=1}^T \right) : \quad (10)$$

$$\left\{ \begin{array}{l} \arg \max \quad \sum_{t=1}^T g_t(E(t), \mathbf{u}^*(t), \mathbf{v}_d(t), \mathbf{v}_c(t), p_{\text{DA}}(t)) \\ \text{s.t.} \quad [V_d(t+1), V_c(t+1)] = \overline{\Phi}(V_d(t), V_c(t), \mathbf{u}^*(t), \Delta P(t)) \\ \quad \quad V_d(t) \leq E(t) \leq V_c(t) \\ \quad \quad \mathbf{u}^*(t) \in \mathcal{O}_{\text{act},t}(E(t), \mathbf{v}_d(t), \mathbf{v}_c(t)) \\ \text{var.} \quad E(t) \in \mathbb{R}, t = 1, \dots, T. \end{array} \right.$$

where,

$$g_t(E(t), \mathbf{u}^*(t), \mathbf{v}_d(t), \mathbf{v}_c(t), p_{\text{DA}}(t)) \doteq$$

$$\mathbb{I}_{(0,\infty)}(E(t)) \cdot [-p_{\text{DA}}(t) \cdot E(t) - C_{c,t}(\mathbf{u}^*(t), \mathbf{v}_c(t))] +$$

$$\mathbb{I}_{(-\infty,0]}(E(t)) \cdot [-p_{\text{DA}}(t) \cdot E(t) - C_{d,t}(\mathbf{u}^*(t), \mathbf{v}_d(t))]$$

represents the utility of the aggregator when provides flexibility $E(t)$ (positive/negative) at time t , where $\mathbb{I}_A : \mathbb{R} \mapsto \{0, 1\}$ denotes the index function for some set $A \subseteq \mathbb{R}$, i.e.,

$$\mathbb{I}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{else.} \end{cases}$$

Note that, when $E(t) > 0$, the aggregator is *charged* with $-p_{\text{DA}}(t)E(t)$ (since this energy is purchased from the DA market), and when $E(t) < 0$, the aggregator is *credited* with $-p_{\text{DA}}(t)E(t)$ (since this energy is sold to the DA market). The variable $\mathbf{u}^*(t)$ is a solution to the activation optimization $\mathcal{O}_{\text{act},t}$ presented in “[Optimal activation](#)” section, i.e., it is a function of the overall offered flexibility $E(t)$. In other words, in order to evaluate the utility of an offered flexibility $E(t)$, we should know the corresponding optimal activation (i.e., which batteries should contribute to generate $E(t)$). We write $\mathbf{u}^*(t)$ instead of $\mathbf{u}^*(E(t), \mathbf{v}_d(t), \mathbf{v}_c(t))$ to simplify notation.

Due to the equilibrium constraints that should be satisfied by \mathbf{u}^* summarized in the constraint $\mathbf{u}^*(t) \in \mathcal{O}_{\text{act},t}(E(t), \mathbf{v}_d(t), \mathbf{v}_c(t))$, optimization \mathcal{O}_{DA} corresponds to a mathematical program with equilibrium constraints (MPEC). The outcome of the DA optimization will be a sequence of optimal flexibilities $E^*(t)$ over the next day that the aggregator should commit to offer to the DA market.

Optimization under imperfect forecasts

In the presence of imperfect forecasts, the aggregator can still address the optimization problem (10) using forecasts of the flexibility potential, $\hat{\mathbf{v}}_d(t)$ and $\hat{\mathbf{v}}_c(t)$, and the DA price $\hat{p}_{DA}(t)$. However, possible discrepancies between actual and estimated quantities will lead to imbalances between the promised commitment $E^*(t)$ (calculated using forecasts) and the available one. This can lead to imbalance costs.

Assuming that the imbalance price is $p_{imb}(t)$ at time interval t , the *actual objective function* that the aggregator is facing when implementing $E^*(t)$ is instead:

$$\begin{aligned} \tilde{g}_t(E^*(t), \mathbf{u}^*(t), \mathbf{v}_d(t), \mathbf{v}_c(t), p_{DA}(t)) \doteq & \\ \mathbb{I}_{(0, V_c(t))}(E^*(t)) \cdot [-p_{DA}(t)E^*(t) - C_{c,t}(\mathbf{u}^*(t), \mathbf{v}_c(t))] + & \\ \mathbb{I}_{[V_d(t), 0]}(E^*(t)) \cdot [-p_{DA}(t)E^*(t) - C_{d,t}(\mathbf{u}^*(t), \mathbf{v}_d(t))] + & \\ \mathbb{I}_{(V_c(t), \infty)}(E^*(t)) \cdot [-p_{DA}(t)V_c(t) - C_{c,t}(\mathbf{u}^*(t), \mathbf{v}_c(t)) - p_{imb}(t)(E^*(t) - V_c(t))] + & \\ \mathbb{I}_{(-\infty, V_d(t))}(E^*(t)) \cdot [-p_{DA}(t)V_d(t) - C_{d,t}(\mathbf{u}^*(t), \mathbf{v}_d(t)) + p_{imb}(t)(E^*(t) - V_d(t))], & \end{aligned} \quad (11)$$

where recall that $\mathbf{u}^*(t)$ is the optimal activation under $E^*(t)$. In other words, the above utility function is the one realized by the aggregator after the completion of the optimization horizon and when the actual measured data ($\mathbf{v}_d(t)$, $\mathbf{v}_c(t)$ and $p_{DA}(t)$) are revealed. The last two terms of the actual utility function corresponds to the penalty that the aggregator pays due to the resulting energy imbalance.

Ideally, we would like that the actual utility of Eq. (11) replaces the perfect-forecast utility of optimization (10), so that we also incorporate the expected costs of faulty forecasts. However, in order to incorporate these costs, it is necessary that we have available an accurate distribution of the forecast errors. Usually, such distribution of errors is not available. For this reason, in the following section, we propose a reinforcement-learning methodology, where, due to an averaging effect, the trained policy will incorporate the possibility of forecast errors.

Approximate dynamic programming (ADP) for DA optimization

ADP background and algorithm

We will consider a version of ADP that can be used to compute optimal policies for dynamic optimization problems of the form (10). The proposed scheme is based upon Monte-Carlo simulations and least-squares approximation.

ADP methodologies are based on the notion of *cost-to-go*, or better here *utility-to-go*. That is, for each time-interval $t \in \{1, 2, \dots, T\}$, we consider the objective function of the sub-problem starting from t onwards (until the end of the optimization horizon T). The utility-to-go at time t will be denoted by $J_t^{\mu_t}$, defined as follows $J_t^{\mu_t} \doteq \sum_{\tau=t}^T \tilde{g}_\tau$, which is a function of the current *state variables* $V_d(t)$, $V_c(t)$ and the exogenously defined DA prices $\{p_{DA}(\tau)\}_{\tau=1}^T$ during the optimization horizon. In other words, $J_t^{\mu_t} = J_t^{\mu_t}(E(t), V_d(t), V_c(t), \{p_{DA}(\tau)\}_{\tau=1}^T)$. The superscript μ_t refers to the policy implemented at time t and it captures the reasoning based on which actions are selected as a function of the currently available information. The $\mathbf{u}^*(t)$, which corresponds to the optimal activation of the selected flexibility $E(t)$ will not be directly included as a parameter of $J_t^{\mu_t}$, but it is indirectly taken into account when $E(t)$ is implemented. The utility-to-go $J_t^{\mu_t}$ also depends on the forecast PV generation and load consumption of the remaining optimization horizon, however we suppress this dependence to simplify notation.

For finite-horizon dynamic programming problems (of the form (10)), Bellman’s Dynamic Programming principle (cf., (Bertsekas 2000, Proposition 1.3.1)) states that a policy μ_t^* that maximizes $J_t^{\mu_t}$ at time t , is also an optimal policy for the original optimization problem (10) at time t . As we discussed though in “[Optimization under imperfect forecasts](#)” section, due to forecast errors, the actual realized utility is not g but \tilde{g} (possibly incorporating imbalance costs). Thus, the explicit form of $J_t^{\mu_t}$ is not a-priori known, and therefore an *exact solution* (i.e., a closed-form expression for the optimal policy μ_t) cannot be computed in practice. To this end, we introduce instead approximation functions of the form $Q_t(E(t), \hat{V}_d(t), \hat{V}_c(t), \{\hat{p}_{DA}(\tau)\}_{\tau=1}^T)$ (so called, Q-factors) that approximate the utility-to-go function $J_t^{\mu_t}$ based on the available forecasts. A different Q-factor applies to each time interval t , and it can be used to generate an approximate policy $\hat{\mu}_t$ for this interval. In particular, the approximate policy $\hat{\mu}_t$ can be computed as:

$$\hat{\mu}_t \left(\hat{V}_d(t), \hat{V}_c(t), \{\hat{p}_{DA}(\tau)\}_{\tau=1}^T \right) \doteq \max_{E \in [\hat{V}_d(t), \hat{V}_c(t)]} Q_t \left(E, \hat{V}_d(t), \hat{V}_c(t), \{\hat{p}_{DA}(\tau)\}_{\tau=1}^T \right), \tag{12}$$

which is a function of the available information at time t , $\hat{V}_d(t), \hat{V}_c(t), \{\hat{p}_{DA}(\tau)\}_{\tau=1}^T$.

A common approach that is used to compute the Q-factors is Monte-Carlo simulations that proceeds as follows. Prior to any day-ahead optimization, we first generate a time-series forecast (*sample path*) for the overall PV generation, the overall load consumption and the DA price. Then, for each time interval t of this sample path, we compute the optimal flexibility commitment $E^*(t)$ using the approximate policy of Eq. (12). Then, at the end of the optimization horizon, we evaluate the performance of the approximate policy $\hat{\mu}_t$ by computing the actual utility-to-go $J_t^{\hat{\mu}_t}$ using the actual measurements. By utilizing the approximation error $|J_t^{\hat{\mu}_t} - Q_t|$, we can improve our approximation functions Q_t .

The specific form of the Q-factors considered here is the following:

$$\begin{aligned} Q_t(E, \hat{V}_d(t), \hat{V}_c(t), \hat{p}_{DA}(t)) &\doteq \alpha_{0,t} + \\ &\alpha_{1,t} \cdot [E]_+ \cdot [\hat{p}_{DA,f}(t) - \hat{p}_{DA}(t)]_+ + \alpha_{2,t} \cdot [E]_- \cdot [\hat{p}_{DA,f}(t) - \hat{p}_{DA}(t)]_+ - \\ &\alpha_{3,t} \cdot [E]_+ \cdot [\hat{p}_{DA}(t) - \hat{p}_{DA,p}(t)]_+ - \alpha_{4,t} \cdot [E]_- \cdot [\hat{p}_{DA}(t) - \hat{p}_{DA,p}(t)]_+ - \\ &\alpha_{5,t} \cdot [E]_+ \cdot [\hat{p}_{DA}(t) - \hat{p}_{DA,f}(t)]_+ - \alpha_{6,t} \cdot [E]_- \cdot [\hat{p}_{DA}(t) - \hat{p}_{DA,f}(t)]_+ \end{aligned} \tag{13}$$

for $E \in [\hat{V}_d(t), \hat{V}_c(t)]$, where

$$\hat{p}_{DA,f}(t) \doteq \begin{cases} \frac{1}{T-t} \sum_{\tau=t+1}^T \hat{p}_{DA}(\tau) & \text{if } t < T \\ \hat{p}_{DA}(T) & \text{if } t = T \end{cases}$$

is the average DA price in the future time intervals, and

$$\hat{p}_{DA,p}(t) \doteq \begin{cases} \hat{p}_{DA}(1) & \text{if } t = 1 \\ \frac{1}{t-1} \sum_{\tau=1}^{t-1} \hat{p}_{DA}(\tau) & \text{if } t > 1 \end{cases}$$

is the average DA price in the previous time intervals. Furthermore, for some real number x , we denote $[x]_+ \doteq \max\{x, 0\}$ and $[x]_- \doteq \min\{x, 0\}$. Finally, the constants $\{\alpha_{0,t}, \alpha_{1,t}, \dots, \alpha_{6,t}\}$ for each interval t are the unknown parameters. We employ also the constraint $\alpha_{1,t}, \dots, \alpha_{6,t} \geq 0$.

The Q-factors of Eq. (13) capture the anticipated utility/losses in the remaining optimization steps. In particular, the first part (multiplying $\alpha_{1,t}$) captures the anticipated revenues when purchasing electricity (i.e., $E > 0$), due to a future increase in the DA price. The fourth part (multiplying $\alpha_{4,t}$) captures the anticipated revenues when selling

electricity (i.e., $E < 0$) due to a current increase in the DA price. The remaining parts intend to minimize different forms of anticipated opportunity costs, which may or may not eventually be part of the actual utility-to-go $J_t^{\mu_t}$. For example, the sixth term (multiplying $\alpha_{6,t}$) captures the anticipated revenues of discharging the battery when anticipating a drop in the future electricity price (thus, giving the opportunity to charge the battery at later stages).

As discussed in detail in (Bertsekas and Tsitsiklis (1996), Section 6.2.1), an issue critical to the success of this method is whether the assumed form of the Q -factors is rich enough to capture the actual utility-to-go. Furthermore, the simulations should generate sample paths that are *persistently exciting* in order for the (least-squares) approximation to converge (cf., (Ljung 1999)). As we will see in the forthcoming experimental evaluation, the considered form (13) will be sufficient to generate accurate approximations.

The details of the ADP algorithm implemented for the DA optimization is presented by Algorithm 5. It consists of two main processes, namely a) *forward simulation* (Step 3), and b) *backward evaluation* (Step 4). In the forward simulation, and for each sample path $s \in \mathcal{S}$ (i.e., each simulation day), we simulate the action selection process (according to Step 3a) and using the current approximation factors $\{Q_t^{(s)}\}_t$. In the backward evaluation, and starting from the last interval of each sample day, we evaluate the utility-to-go performance of the offered flexibility on the actual/realized data. This evaluation can then be used to better approximate the Q factors (Step 4c). We should expect that as we increase the number of tested sample paths, the approximation error of the utility-to-go approaches zero.

After the Q -factors have been trained, as Algorithm 5 demonstrates, at the beginning of each day, we can employ the trained factors to compute the optimal flexibility bidding as Step 3 of Algorithm 5 describes.

Discussion

The reasons for considering the proposed hierarchical- and ADP-based optimization are the following: i) the original problem formulation of Eq. (10) constitutes a dynamic-programming (DP) optimization problem (due to the battery dynamics); ii) the original problem is subject to equilibrium constraints (for each quantity of extracted flexibility there is an optimal activation); iii) the original problem involves integer (activation) variables; iv) an exact DP solution to the combined (integer) optimization of optimal bidding and optimal activation of the batteries is practically infeasible; v) by separating the decisions over the optimal energy commitment (Step 3a of Algorithm 5) from the (precomputed) optimal activation (Step 3b of Algorithm 5), the overall scheme avoids the computational complexity of the combined scheduling problem; vi) the ADP optimization provides strategies (rather than explicit schedules), thus the solution can be updated whenever updated forecasts become available; vii) the learning-based nature of the algorithm allows for capturing the effects of imperfect forecasts due to averaging.

Experimental evaluation

Algorithm 5 was implemented on real-world data collected from $N = 30$ battery-storage systems located in the state of Upper-Austria, over the duration of approximately one year. The simulation days used for training the Q -factors correspond to approximately 4000 days (since a simulation day was used more than once to achieve better

Algorithm 5 Optimal utility-to-go training $\{Q_t\}_{t=1}^T$ 1: **procedure** ADP2: **Step 1.** Choose sample paths, denoted by $s \in \mathcal{S}$ (each of which corresponds to one day) of DA clearing prices $\{\hat{p}_{\text{DA}}^{(s)}(1:T)\}_s$, and PV power generation $\{\hat{p}_{\text{PV},i}^{(s)}(1:T)\}_s$, load consumption $\{\hat{p}_{\text{load},i}^{(s)}(1:T)\}_s$ and initial state-of-charge $\{\text{SOC}_i^{(s)}(0)\}_s$ from several households $i = 1, \dots, N$.3: **Step 2.** Define the following approximation factor of the utility-to-go,

$$Q_t^{(s)} = Q_t^{(s)} \left(E^{(s)}(t), \hat{V}_d^{(s)}(t), \hat{V}_c^{(s)}(t), \hat{p}_{\text{DA}}^{(s)}(1:T) \right),$$

$t = 1, \dots, T$, $s \in \mathcal{S}$, given by Eq. (13), with unknown constants $\alpha_{0,t}^{(s)}, \alpha_{1,t}^{(s)}, \dots, \alpha_{7,t}^{(s)}$, for $s = 1, \dots, |\mathcal{S}|$, and $E^{(s)}(t) \in [\hat{V}_d^{(s)}(t), \hat{V}_c^{(s)}(t)]$.

4: **Step 3. Computing optimal flexibility extraction using forecasts**5: Set $s = 0$.6: **for** $t = 1, 2, \dots, T$ **do**▷ Throughout the sample path of one day s 7: **Step 3a.** Compute the decision at time t using the approximations of $Q_t^{(s-1)}$ during the previous sample path $s - 1$ evaluated at the current state, i.e.,

$$E^{(s)}(t) \in \begin{cases} \arg \max_{E \in [\hat{V}_d^{(s)}(t), \hat{V}_c^{(s)}(t)]} Q_t^{(s-1)} \left(E, \hat{V}_d^{(s)}(t), \hat{V}_c^{(s)}(t), \hat{p}_{\text{DA}}^{(s)}(1:T) \right), & \text{w.p.} 1 - \lambda \\ \text{rand} \left[\hat{V}_d^{(s)}(t), \hat{V}_c^{(s)}(t) \right], & \text{w.p.} \lambda. \end{cases}$$

▷ We use our currently available approximations, evaluated at sample path $s - 1$, to compute our action at time t . For a small probability λ , we select a random (feasible) action, in order to create conditions of *persistence of excitation* in the identification of the Q -factors (cf., (Ljung 1999, Definition 13.2)).

8: **Step 3b.** For the selected flexibility extracted (positive or negative) $E^{(s)}(t)$, we compute the optimal activations of the participating batteries, i.e.,

$$\mathbf{u}^{(s)}(t) \in \mathcal{O}_{\text{act},t} \left(E(t), \hat{\mathbf{v}}_d^{(s)}(t), \hat{\mathbf{v}}_c^{(s)}(t) \right)$$

the solution of which is given in “Optimal activation” section.

9: **Step 3c.** Update the total energy potential state

$$\left[\hat{V}_d^{(s)}(t+1), \hat{V}_c^{(s)}(t+1) \right] = \bar{\Phi} \left(\hat{V}_d^{(s)}(t), \hat{V}_c^{(s)}(t), \mathbf{u}^{(s)}(t), \widehat{\Delta P}^{(s)}(t) \right)$$

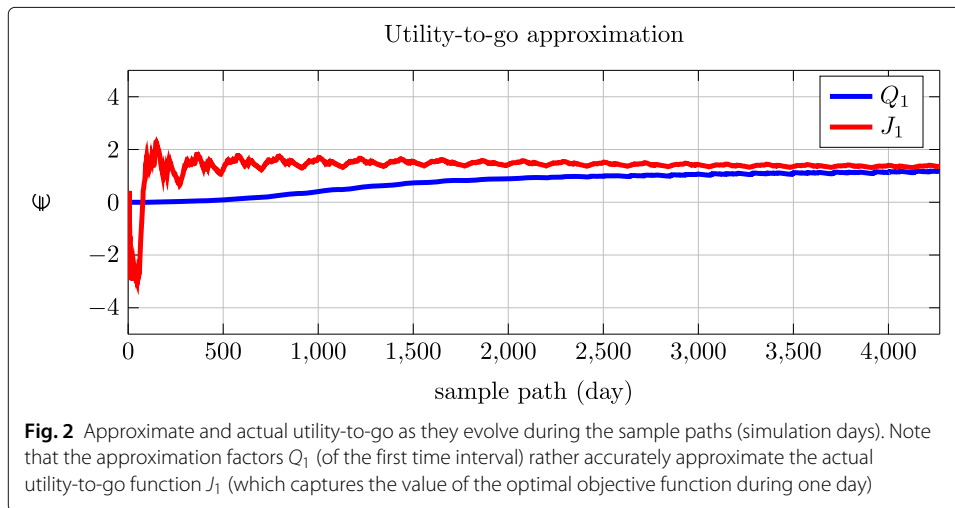
10: **Step 3d.** Implement the computed activation strategy $\mathbf{u}^*(t)$ during time interval t .11: **Step 4. Evaluating selected actions using actual data and going backwards**12: Set $J_T \doteq \tilde{g}_T$, which corresponds to the actual performance during the last time interval T of the sample path s .13: **for** $t = T - 1, T - 2, \dots, 1$ **do**14: **Step 4a.** At the end of interval t , and given the actual recorded data (of PV generation and load consumption), we evaluate the performance of the implemented activation strategy, $q_t = \tilde{g}_t \left(E^{(s)}(t), \mathbf{v}_d^{(s)}(t), \mathbf{v}_c^{(s)}(t), p_{\text{DA}}^{(s)}(t) \right)$.15: **Step 4b.** Compute a time- t sample of the utility-to-go, $J_t = r \cdot J_{t+1} + q_t$, with discount factor $r = 1$.16: **Step 4c.** Use this sample J_t to train the utility-to-go factor

$$Q_t^{(s)} \leftarrow \mathcal{M} \left(Q_t^{(s-1)}, J_t \right)$$

where \mathcal{M} is the training model that we use to train the parameters of the Q factors. In this case, it corresponds to a Recursive-Least-Squares filter (cf., (Sayed 2003)).

17: **Step 5.** Increment s . If $k \leq |\mathcal{S}|$ go to Step 3.18: **Step 6. return** the Q -factors of the final stage $|\mathcal{S}|$, $\{Q_t^{(|\mathcal{S}|)}\}_t$

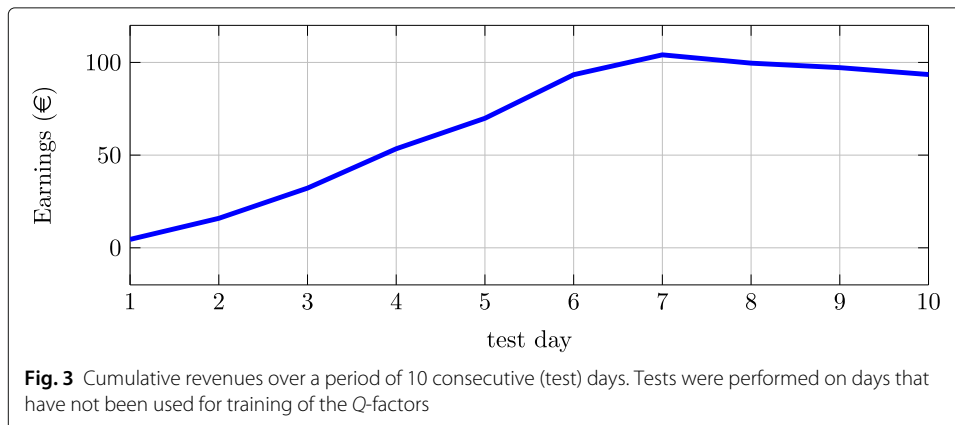
training performance of the recursive-least-squares filter implemented). For these evaluations, we considered $\Delta T = 15\text{min}$, which implies that each day consists of $T = 96$

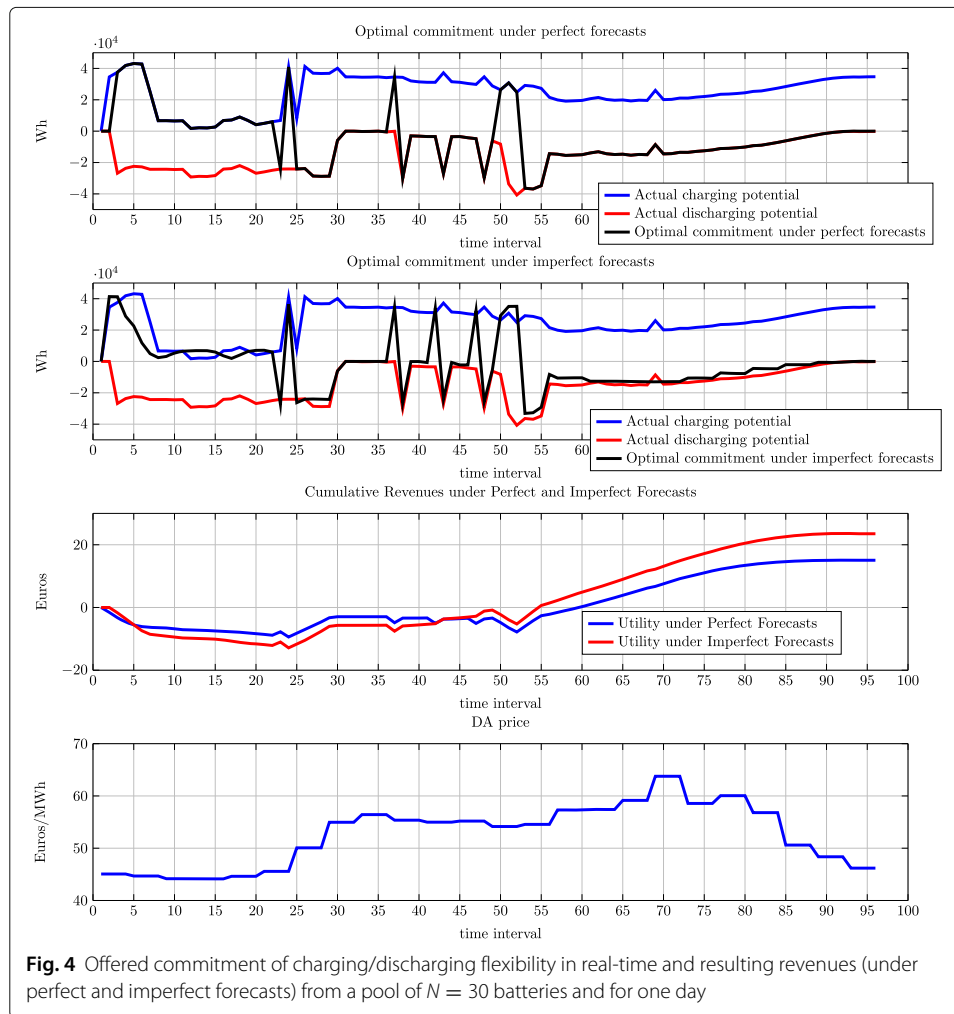


time intervals. The available data include recordings of PV power generation and load consumption.

At the beginning of each simulation day, the trained Q -factors are used to derive the optimal flexibility commitments over the upcoming day using forecast values of the unknown data ($\widehat{\Delta P}$ and \widehat{p}_{DA}). At the end of the upcoming day, when the actual data ΔP and p_{DA} have been revealed, the performance of these flexibility commitments are evaluated and the approximation errors over $t = 1, \dots, T$ are used to train the Q -factors. The approximation error of the utility-to-go function during training is depicted in Fig. 2, where we see that the considered Q_1 -factors (of the first time interval) can quite accurately approximate the actual utility-to-go J_1 .

When the training of the Q -factors has been completed, we evaluate the trained factors on a test data set of sample paths (days), which have not been used during training. The outcome of these test evaluations are depicted in Fig. 3. Finally, Fig. 4 demonstrates the real-time implementation of the computed optimal flexibility over the next day. Note that discrepancies between the offered flexibility and the actual flexibility potential can still be observed due to forecast inaccuracies.





Conclusions and future work

This work presented an optimization framework for optimal participation of an aggregator in the DA spot-market through the direct control of a set of residential battery-storage systems. The aggregator optimizes the amount of flexibility (the energy that can be charged/discharged in the participating batteries) that can be offered to the DA market during the next day. The optimization is based upon forecast PV generation and load consumption of the participating households over the next day. Given the expected forecast errors, as well as the complexity of the involved optimization, we proposed a reinforcement-learning methodology that trains over time to adapt to forecast inaccuracies and provides an optimal strategy for each time interval of the next day. Given that the outcome of the optimization is a strategy (rather than a specific schedule), we can immediately re-adjust the optimal schedules (e.g., in case of updated forecasts) without the need for a full re-optimization.

About this supplement

This article has been published as part of *Energy Informatics* Volume 2 Supplement 1, 2019: Proceedings of the 8th DACH+ Conference on Energy Informatics. The full contents of the supplement are available online at <https://energyinformatics.springeropen.com/articles/supplements/volume-2-supplement-1>.

Authors' contributions

GC and MP have designed and simulated the ADP-based optimization framework for the optimal activation and bidding of flexibility potential in the Day-ahead electricity market. JS and TE have provided feedback and data with respect to the operation of the Day-ahead electricity market and the participation of pools of flexible loads. All authors read and approved the final manuscript.

Funding

This work has been supported by the Austrian Research Agency FFG through the research project Flex+ (FFG # 864996). It has also been partially supported by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry of Science, Research and Economy, and the Province of Upper Austria in the frame of the COMET center SCCH. Publication of this supplement was funded by Austrian Federal Ministry for Transport, Innovation and Technology.

Availability of data and materials

The data used in this paper have been extracted from 30 battery-storage systems located in the state of Upper Austria. Due to privacy issues and non-disclosure agreements, these data are not publicly available.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Software Competence Center Hagenberg GmbH, Softwarepark 21, 4232 Hagenberg, Austria. ²AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Vienna, Austria.

Published: 23 September 2019

References

- Bertsekas D (2000) *Dynamic Programming and Optimal Control*. 2nd edn. Athena Scientific, Belmont
- Bertsekas D, Tsitsiklis J (1996) *Neuro-Dynamic Programming*. Athena Scientific, Belmont
- Chasparis G, Pichler M, Natschläger T (2019) A Demand-Response Framework in Balance Groups through Direct Battery-Storage Control. In: 18th European Control Conference, Napoli. pp 1392–1397
- Chen C, Wang J, Kishore S (2014) A Distributed Direct Load Control Approach for Large-Scale Residential Demand Response. *IEEE Trans Power Syst* 29(5):2219–2228
- Flex+. <https://www.flexplus.at>. <https://doi.org/10.1109/TPWRS.2009.2016607>
- Gomez-Villalva E, Ramos A (2003) Optimal energy management of an industrial consumer in liberalized markets. *IEEE Trans Power Syst* 18(2):716–723. <https://doi.org/10.1109/TPWRS.2003.811197>
- Herter K (2007) Residential implementation of critical-peak pricing of electricity. *Energy Policy* 35(4):2121–2130
- He G, Chen Q, Kang C, Pinson P, Xia Q (2016) Optimal Bidding Strategy of Battery Storage in Power Markets Considering Performance-Based Regulation and Battery Cycle Life. *IEEE Trans Smart Grid* 7(5):2359–2367. <https://doi.org/10.1109/TSG.2015.2424314>
- Iria JP, Soares FJ, Matos MA (2017) Trading small prosumers flexibility in the day-ahead energy market. In: 2017 IEEE Power & Energy Society General Meeting. IEEE, Chicago. pp 1–5. <https://doi.org/10.1109/PESGM.2017.8274488>
- Jiang DR, Powell WB (2015) Optimal Hour-Ahead Bidding in the Real-Time Electricity Market with Battery Storage Using Approximate Dynamic Programming. *INFORMS J Comput* 27(3):525–543. <https://doi.org/10.1287/ijoc.2015.0640>
- Kairies K, Haberschus D, Ouwerkerk J, Strebel J, Wessels O, Magnor D, Badeda J, Sauer U (2016) *Wissenschaftliches Mess- und Evaluierungsprogramm Solarstromspeicher 2.0. Jahresbericht*. Technical Report, Institut für Stromrichtertechnik und Elektrische Antriebe, RWTH Aachen
- Ljung L (1999) *System Identification: Theory for the User*. 2nd edn. Prentice Hall Ptr, Upper Saddle River
- Mohsenian-Rad H (2016) Optimal Bidding, Scheduling, and Deployment of Battery Systems in California Day-Ahead Energy Market. *IEEE Trans Power Syst* 31(1):442–453. <https://doi.org/10.1109/TPWRS.2015.2394355>
- Nan S, Zhou M, Li G (2018) Optimal residential community demand response scheduling in smart grid. *Appl Energy* 210:1280–1289. <https://doi.org/10.1016/j.apenergy.2017.06.066>
- Nguyen HK, Song JB, Han Z (2015) Distributed Demand Side Management with Energy Storage in Smart Grid. *IEEE Trans Parallel Distrib Syst* 26(12):3346–3357. <https://doi.org/10.1109/TPDS.2014.2372781>
- Parvania M, Fotuhi-Firuzabad M, Shahidehpour M (2013) Optimal Demand Response Aggregation in Wholesale Electricity Markets. *IEEE Trans Smart Grid* 4(4):1957–1965. <https://doi.org/10.1109/TSG.2013.2257894>
- Philpott AB, Pettersen E (2006) Optimizing Demand-Side Bids in Day-Ahead Electricity Markets. *IEEE Trans Power Syst* 21(2):488–498. <https://doi.org/10.1109/TPWRS.2006.873119>
- Ruiz N, Cobelo I, Oyarzabal J (2009) A Direct Load Control Model for Virtual Power Plant Management. *IEEE Trans Power Syst* 24(2):959–966
- Sayed A (2003) *Fundamentals of Adaptive Filtering*. John Wiley & Sons, Inc., New Jersey
- Triki C, Violi A (2009) Dynamic pricing of electricity in retail markets. *4OR* 7(1):21–36
- Xu Y, Li N, Low SH (2016) Demand Response With Capacity Constrained Supply Function Bidding. *IEEE Trans Power Syst* 31(2):1377–1394

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.