OR Spectrum (2016) 38:633–660
DOI 10.1007/s00291-016-0439-x

REGULAR ARTICLE

# Sell or store? An ADP approach to marketing renewable energy

**Jochen Gönsch**[1] · **Michael Hassler**[2]

Received: 14 August 2014 / Accepted: 8 March 2016 / Published online: 5 April 2016


**Abstract** In deregulated markets, electricity is usually traded in advance, and the advance commitments have a time lag of several periods. For example, in the German intraday market, the seller commits to providing electricity 45 min before the 15-min interval in which delivery has to be made. We consider the problem of a producer that generates energy from stochastic, renewable sources, such as solar or wind and uses a storage device with conversion losses. We model the problem as a Markov Decision Process and consider lagged commitments for the first time in the literature. The problem is solved using an innovative approximate dynamic programming approach. Its key elements are the analytical derivation of the optimal action based on the value function approximation and a new combination of approximate policy iteration with classical backward induction. The new approach is quite general with regard to the stochastic processes describing the energy production and price evolution. We demonstrate the application of our approach by considering a wind farm/storage combination. A numerical study using real-world data shows the applicability and performance of the new approach and investigates how the storage device's parameters influence profit.

**Keywords** Electricity · Renewable energy · Storage · Dynamic programming

✉ Jochen Gönsch
jochen.goensch@uni-due.de

Michael Hassler
michael.hassler@wiwi.uni-augsburg.de

1 Mercator School of Management, University of Duisburg-Essen, Lotharstraße 65, 47057 Duisburg, Germany

2 Chair of Analytics and Optimization, University of Augsburg, Universitätsstraße 16, 86159 Augsburg, Germany

## 1 Introduction

Today, the integration of a steadily increasing amount of electricity from renewable sources is a major challenge for many countries' power systems. However, energy production from these sources is often stochastic as it depends on, for example, the wind blowing and the sun shining. In practice, two approaches are predominantly employed to balance supply and demand at all times, which is essential for a stable power grid. The first approach seeks to temporally shift demand through demand side management (see, e.g., Strbac 2008 for an overview of the techniques, as well as the benefits and challenges). It uses, for example, smart meters and corresponding tariffs, with higher prices charged when energy is scarce. The second approach aims at the supply side and is based on storing excess energy for later usage. In fact, energy producers with renewable, stochastic sources increasingly consider adding a storage device to obtain higher prices for their energy and to avoid being constrained by power grid congestions limiting their feed-in into the network. For example, Bosch, the German technology and service supplier, recently added a storage device to a community wind farm in Northern Germany (Robert Bosch GmbH 2014).

This paper contributes to the second approach. We consider a profit-maximizing firm that generates electrical energy from stochastic, renewable sources, such as solar or wind. The energy is sold on a market and can be stored in a storage device with conversion losses. As in all modern energy markets—with the exception of special regulating markets that jump in if a party fails to comply with its contracts—there is a considerable time lag between the commitment and the actual delivery of energy. For example, at the EPEX SPOT intraday market, which is the major European power exchange covering France, Germany, Austria, and Switzerland, contracts can be traded until 45 min before the 15-min interval in which delivery has to be made. In the day-ahead market, the time lag may, at a minimum, comprise several hours. Hence, the commitment is decided well before the relevant energy production becomes known and the producer faces a tradeoff. On the one hand, if the firm overcommits and cannot deliver later as promised, because the energy production and the storage fall short of the commitment, the missing energy must be obtained at a possibly very high cost on the regulating market. On the other hand, if it undercommits and could have delivered more, the firm might forfeit its profit due to conversion losses or because the storage is full and cannot absorb the excess energy. Moreover, the power feed-in into the grid is constrained due to, for example, limited transmission line capacity.

In keeping with the trading process's structure, we formulate a Markov Decision Process (MDP) in discrete time with a finite time horizon. To the best of our knowledge, this is the first paper to consider real markets' lagged commitments. The resulting state space with numerous continuous dimensions inhibits the use of classical backward induction to solve this MDP to optimality, thus making approximate dynamic programming (ADP) techniques self-evident.

This paper's major contribution is the development of an innovative ADP framework based on autoregressive stochastic processes describing the energy production and price evolution. We analytically derive the optimal commitment for an approximation of the value function. Based on this, we develop an algorithm, called HAPI (Hybrid Approximate Policy Iteration), to approximate the value function, using a

new combination of approximate policy iteration and classical backward induction. In numerical experiments, we show that this framework can be used for a wind farm with a storage device. As there is no optimal solution available for the problem considered, we compare HAPI with another approximate solution of the MDP and with several intuitive heuristics based on the energy production's expected value. Conducted with real-world data, the numerical experiments indicate that our ADP approach shows a better profit and runtime performance.

The paper is organized as follows: In Sect. 2, we briefly review the literature. Sections 3 and 4 contain the general framework. The generic decision problem is modeled as an MDP in Sect. 3. We develop our new ADP approach regarding arbitrary stochastic processes in Sect. 4. Sections 5 and 6 contain the application to the wind farm/storage combination, with Sect. 5 presenting the algebra required for usage with the specific processes considered, and Sect. 6 presenting numerical experiments. Section 7 concludes the paper. The appendix contains the analytical derivations referred to in Sects. 4 and 5.

## 2 Literature review

In the following, we outline the stream of literature most relevant to our setting, i.e., publications that model problems involving the optimal control or valuation of energy storage as an MDP.

Mokrian and Stephen (2006) focus on the use of storage for intraday arbitrage, developing a stochastic optimization problem and a dynamic program for optimizing the operation of storage over a 24-h period. Costa et al. (2008) consider the combination of a storage device and a wind farm, formulating the optimal scheduling problem as a dynamic program. Lai et al. (2010) develop an ADP method that computes the lower and upper bounds of natural gas storage capacity's value. Hannah and Dunson (2011) apply ADP to the day-ahead commitment problem of a wind farm in combination with energy storage. Löhndorf et al. (2013) study the optimization of hydro-storage systems when participating in both the intraday and day-ahead markets. Sioshansi et al. (2014) determine the capacity value of energy storage by means of dynamic programming. Zhou et al. (2016) assess the value of storage for trading on a real-time market and derive structural results regarding the MDP's optimal policy. Surprisingly, in numerical experiments with real-world prices, they show that a very inefficient battery, which more or less only burns energy, has the highest value, because it exploits negative prices. Jiang et al. (2014) directly compare the performance of different ADP approaches for the optimal control of an energy storage device. Jiang and Powell (2015) focus on one of these approaches and exploit the monotonicity of the value function. In addition, the same group has published a number of (working) papers on similar topics (see, e.g., Nascimento and Powell 2009; Salas and Powell 2013; Scott and Powell 2012).

Closest to our paper are Zhou et al. (2014), Kim and Powell (2011), and Löhndorf and Minner (2010), all of whom address a profit-maximizing power producer with renewable sources. The producer utilizes a storage device and makes energy commitments on the market. In contrast to our work, none of these publications considers

a time lag between making the commitment and the delivery of energy. Zhou et al. (2014) consider transmission line capacity and analytically characterize the optimal policies for a discretized version of the problem. In contrast, Kim and Powell (2011), as well as Löhndorf and Minner (2010), do not incorporate a constrained feed-in into the grid and model the optimization problem by an MDP with continuous states. The former authors analytically derive an optimal policy for an infinite horizon under certain assumptions, whereas the latter use ADP to approximately solve the problem for a finite horizon. However, Löhndorf and Minner (2010) only consider perfectly efficient storage. An additional key difference between our study and that of Kim and Powell (2011) is that instead of deriving a solution for specific stochastic processes, our approach allows a broad range of processes for modeling market prices and energy production through the use of ADP. However, we are no longer able to analytically characterize the optimal policy for the exact MDP.

## 3 Generic model formulation

We assume a profit-maximizing power producer that generates electricity from a renewable power source and possesses a storage device with conversion losses. This firm operates in two markets. The energy produced is sold on the intraday market. Additionally energy may be bought on a regulating market to compensate for an unexpected shortfall of production.

On the intraday market, a commitment is made in advance. If the commitment cannot be met by means of actual energy production and energy stored in the storage device, energy has to be bought at a higher price on the regulating market. The producer thus pays an a priori unknown penalty for not meeting its commitment. On both markets, the producer is assumed to act as a price taker because the firm's trading volume is small—think of an individual wind farm operator.

The basic sequence of events is illustrated in Fig. 1: energy is produced continuously, but decisions are only taken at discrete points in time. Thus, we only consider the system at these points in time. At time $t$, the producer learns about the energy

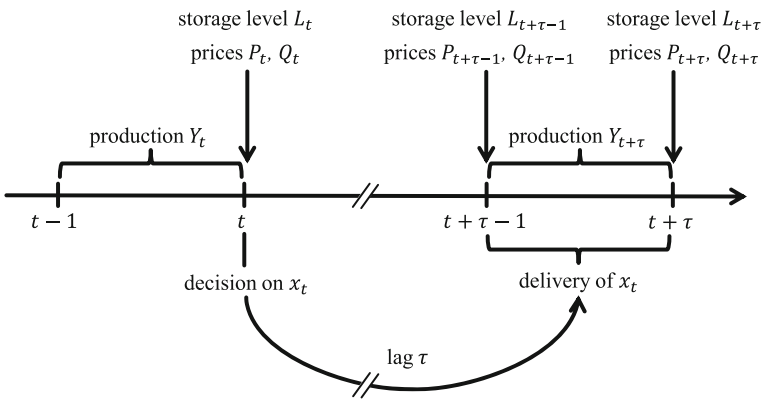


**Fig. 1** Sequence of events with lagged advance commitment

production $Y_t$ during the interval between time $t-1$ and time $t$ and the resulting storage level $L_t$. Moreover, the producer takes note of the current selling price $P_t$ and the penalty price $Q_t$. Using this information, the producer decides on $x_t$, the advance commitment sold now and delivered with a time lag of $\tau$ between time $t+\tau-1$ and $t+\tau$. If, in this time interval, supply $Y_{t+\tau}$ falls short of the commitment $x_t$, the producer first resorts to the storage, where a storage level of $L_{t+\tau-1}$ remains at time $t+\tau-1$. If the stored energy does not suffice, the producer incurs a penalty payment of $Q_{t+\tau}$ per unit, as energy is automatically bought on the regulating market. On the other hand, if the supply exceeds the commitment, the excess energy is stored. The new storage level at time $t+\tau$ is denoted by $L_{t+\tau}$.

In the following subsections, we model the decision problem outlined above as an MDP with the common elements from literature (see, e.g., Powell 2011, Chapter 5; Puterman 2005). We begin the description with the exogenous processes governing the prices and the energy production in Sect. 3.1. The state of the system and the transition function are defined in Sect. 3.2 and the evolution of the storage is stated in Sect. 3.3. In Sect. 3.4, the contribution earned from a commitment is formalized. Finally, the resulting value function is given in Sect. 3.5. Table 1 provides an overview of the notation introduced in this section.

### 3.1 Exogenous processes

As the producer is a price taker, we consider exogenous autoregressive processes for the total energy production $Y_t$ between time $t-1$ and $t$, for the price $P_t$ of selling energy on the intraday market at time $t$ and for the penalty price $Q_t$ incurred if energy is bought on the regulating market at time $t$. With these processes, the exogenous state of the system at time $t$ is $W_t = ((Y_{t'})_{t-h_Y+1\le t'\le t+1}, (P_{t'})_{t-h_P+1\le t'\le t+1}, (Q_{t'})_{t-h_Q+1\le t'\le t+1,})$, where the order of the respective autoregressive stochastic processes is given by $h_Y$, $h_P$, and $h_Q$.

Specifically,

$$\begin{aligned} Y_{t+1} &= \sum_{i=1}^{h_Y} \alpha_i^Y Y_{t-i+1} + \hat{y}_{t+1} \\ P_{t+1} &= \sum_{i=1}^{h_P} \alpha_i^P P_{t-i+1} + \hat{p}_{t+1} \\ Q_{t+1} &= \sum_{i=1}^{h_Q} \alpha_i^Q Q_{t-i+1} + \hat{q}_{t+1} \end{aligned} \tag{1}$$

with noise terms $\hat{y}_{t+1}$, $\hat{p}_{t+1}$. and $\hat{q}_{t+1}$ following a specific distribution. More complex autoregressive models are easily integrated into this framework. This includes, but is not limited to, the consideration of seasonalities by the integration of further deterministic components as well as allowing for heteroscedasticity. A detailed description of these extensions is excluded to keep the presentation concise but can be found in basic

**Table 1** Notation introduced in Sect. 3

| | |
|---|---|
| *Parameters* | |
| $L_{\max}$ | Storage capacity |
| $\rho_R$ | Efficiency when energy is stored |
| $\rho_E$ | Efficiency when energy is taken from the storage |
| $\Delta t$ | Length of a time interval |
| $\tau$ | Time lag between commitment and delivery of energy |
| $T$ | Number of time intervals |
| $\beta$ | Discount factor |
| $x_{\max}$ | Maximum commitment |
| $h_Y$ | Order of the autoregressive process describing the energy production |
| $h_P$ | Order of the autoregressive process describing the selling price for electricity |
| $h_Q$ | Order of the autoregressive process describing the penalty price for electricity |
| $(\alpha_i^Y)_{i=1,\dots,h_Y}$ | Parameters of the autoregressive process describing the energy production |
| $(\alpha_i^P)_{i=1,\dots,h_P}$ | Parameters of the autoregressive process describing the selling price for electricity |
| $(\alpha_i^Q)_{i=1,\dots,h_Q}$ | Parameters of the autoregressive process describing the penalty price for electricity |
| *Decision (action) variable* | |
| $x_t = \pi_t(S_t)$ | Commitment decided at time $t$ in state $S_t$ for delivery between time $t+\tau-1$ and $t+\tau$ |
| *State variables* | |
| $t$ | Discrete time index with actual time $t \cdot \Delta t$ |
| $L_t$ | Storage level at time $t$ |
| $Y_t$ | Total energy production between time $t-1$ and $t$ |
| $P_t$ | (Selling) price for electricity committed at time $t$ and delivered between time $t+\tau-1$ and $t+\tau$ |
| $Q_t$ | (Penalty) price for electricity bought at the regulating market at time $t$ |
| $W_t = \big((Y_{t'})_{t-h_Y+1\le t'\le t+1}, (P_{t'})_{t-h_P+1\le t'\le t+1}, (Q_{t'})_{t-h_Q+1\le t'\le t+1,}\big)$ | Exogenous state of the system at time $t$ |
| $S_t = (L_t, x_{t-\tau+1}, \dots, x_{t-1}, W_t)$ | State of the system at time $t$ |
| *Additional notation* | |
| $\hat{y}_t$ | Random variable capturing the noise of $Y_t$ |
| $\hat{p}_t$ | Random variable capturing the noise of $P_t$ |
| $\hat{q}_t$ | Random variable capturing the noise of $Q_t$ |

**Table 1** continued

| | |
|---|---|
| $L_t(L_{t-1}, Y_t, x_{t-\tau})$ | Evolution of the storage from time $t-1$ to time $t$ |
| $T(S_t, x_t, Y_{t+1}, P_{t+1}, Q_{t+1})$ | Transition function describing the transition from state $S_t$ to state $S_{t+1}$ |
| $\Phi(\cdot\|\mu, \sigma^2)$ | Cumulative distribution function of the normal distribution with mean $\mu$ and variance $\sigma^2$ |
| $C(S_t, x_t)$ | Contribution function: expected profit obtained from commitment $x_t$ in state $S_t$ |
| $\pi = (\pi_0, \ldots, \pi_{T-1})$ | Policy providing a decision for each time period and each state |
| $V_t(S_t)$ | Value function: total expected profit obtained in state $S_t$ from time $t$ onwards |

textbooks on time series analysis (see, e.g., Brockwell and Davis 2013 or Shumway and Stoffer 2013).

In the beginning of the time horizon, the exogenous state has either not reached its full size as there are not enough data yet (yielding the same result in the autoregressive models as setting the relevant values to zero) or, if the model is applied in rolling horizon based planning, past values from before the considered time horizon may be available and can be used. In the following, we assume full size of $W_t$ to keep the notation concise.

Some rather mild conditions are necessary to derive the analytical solution used in the ADP approach in Sect. 4:

- $(Y_t)$, $(P_t)$ and $(Q_t)$ need not be stochastically independent, but their joint distribution must be known.
- The first and second moment of each random variable have to exist. Moreover, this condition must also hold for the three pair-wise second lower partial moments, i.e., $\int_{[0,z]\times\mathbb{R}} yp \,\mathrm{d}F_{(Y_t,P_t)}(y,p)$, $\int_{\mathbb{R}^2} pq \,\mathrm{d}F_{(P_t,Q_t)}(p,q)$, and $\int_{[0,z]\times\mathbb{R}} yq \,\mathrm{d}F_{(Y_t,Q_t)}(y,q)$, where always $z > 0$.

### 3.2 State of the system and transition function

We represent the state of the system at time $t$ as $S_t = (L_t, x_{t-\tau+1}, \ldots, x_{t-1}, W_t) \in \mathbb{R}^{1+(\tau-1)+h_Y+h_P+h_Q}$. For $t < \tau - 1$, either past values are used for $x_t$ if available or $x_t$ is set to zero (analogous to the initialization of $W_t$ described in Sect. 3.1). The transition from state $S_t$ to state $S_{t+1}$ when committing $x_t$ is then given by the following transition function:

$$\begin{aligned} S_{t+1} &= T(S_t, x_t, Y_{t+1}, P_{t+1}, Q_{t+1}) \\ &= (L_{t+1}(L_t, Y_{t+1}, x_{t-\tau+1}), x_{t-\tau+2}, \ldots, x_t, W_{t+1}(W_t, Y_{t+1}, P_{t+1}, Q_{t+1})), \end{aligned}$$

where $L_{t+1}(L_t, Y_{t+1}, x_{t-\tau+1})$, the evolution of the storage, is given below in Sect. 3.3 and $W_{t+1}(W_t, Y_{t+1}, P_{t+1}, Q_{t+1}) = ((Y_{t'})_{t-h_Y+2\le t'\le t+1}, (P_{t'})_{t-h_P+2\le t'\le t+1}, (Q_{t'})_{t-h_Q+2\le t'\le t+1})$, i.e., the oldest information is discarded and new information revealed at time $t+1$ is included. Please note that the transition function yields the

next state $S_{t+1}$. The probability of this transition is given by the transition probability function (Puterman 2005), which we do not explicitly state here as it directly follows from the transition function and the exogenous processes.

### 3.3 Evolution of the storage

The storage level at time $t$ is given by

$$L_t(L_{t-1}, Y_t, x_{t-\tau}) = \begin{cases} \min\{L_{\max}, L_{t-1} + \rho_R(Y_t - x_{t-\tau})\} & \text{if } x_{t-\tau} < Y_t \\ \max\left\{0, L_{t-1} - \frac{1}{\rho_E}(x_{t-\tau} - Y_t)\right\} & \text{if } x_{t-\tau} \geq Y_t \end{cases} \tag{2}$$

If the total energy production between time $t-1$ and $t$ exceeds the commitment to be delivered in this time interval ($x_{t-\tau} < Y_t$), the excess energy $Y_t - x_{t-\tau}$ is stored with a conversion factor $\rho_R$, but the storage level $L_t$ cannot exceed the maximum capacity $L_{\max}$. On the other hand, if the energy production falls short of the commitment ($x_{t-\tau} \geq Y_t$), we have to extract $\frac{1}{\rho_E}(x_{t-\tau} - Y_t)$ from the storage, again with a conversion factor $\rho_E$. If the storage level is not sufficient, we fail to meet the commitment and energy is bought on the regulating market, which incurs a penalty cost of $Q_t$ as described in Sect. 3.4.

### 3.4 Contribution (revenue) function

We use the following contribution function:

$$C(S_t, x_t) = P_t x_t - \beta \mathbb{E}[Q_{t+1}[x_{t-\tau+1} - (\rho_E L_t + Y_{t+1})]^+ | S_t, x_t] \tag{3}$$

for $t \leq T-1$. At time $t$, this contribution includes the revenue obtained from selling the commitment $x_t$ minus an eventual penalty payment at time $t+1$. The revenue only depends on values known at time $t$ ($P_t, x_t$). The penalty payment is discounted with a factor $0 < \beta \leq 1$ and is not incurred after the end of the horizon. Its calculation necessitates an expectation, because—besides $L_t$ and $x_{t-\tau+1}$, which are known at $t$—it depends on values that realize after $t$ ($Y_{t+1}$, $Q_{t+1}$).

### 3.5 Objective function and value function

The power producer's goal is to maximize the expected discounted profit. Thus, the objective function is the sum of the contributions over the time horizon:

$$\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T-1} \beta^t C(S_t, \pi_t(S_t))\right]$$

Here, the maximization is over all policies $\pi = (\pi_0, \ldots, \pi_{T-1})$. A policy is a function of the current state $S_t$ that tells us what decision to take in that state, i.e., $x_t = \pi_t(S_t)$. A maximum feed-in into the grid is considered by $\pi_t(S_t) \in [0, x_{\max}]$.

Next, we formulate the value function (also known as the Bellman equation) to solve the problem in a dynamic programming framework. The maximum expected profit obtained in state $S_t$ from time $t$ onwards is given by

$$V_t(S_t) = \max_{x_t \in [0, x_{\max}]} \left\{ C(S_t, x_t) + \beta \mathbb{E}\left[ V_{t+1}(S_{t+1}) | S_t, x_t \right] \right\} \tag{4}$$

with the boundary conditions $V_T(S_T) = 0 \ \forall S_T$. The optimal commitment is $\pi_t^*(S_t) = \text{argmax}_{x_t \in [0, x_{\max}]} \{C(S_t, x_t) + \beta \mathbb{E}[V_{t+1}(S_{t+1}) | S_t, x_t]\}$. The value function captures the revenue from selling the current and all the future commitments, minus eventual penalty costs incurred in the future, including those due to past commitments.

Please note that we do not explicitly constrain the power flow into/out of the storage. However, the maximum production and the maximum feed-in into the grid $x_{\max}$ are implicit constraints. A tighter limit can be modeled by an adequate modification of Eq. (2) for inflows and Eqs. (2) and (3) for outflows.

## 4 Approximate solution of the generic model

Roughly speaking, three issues complicate solving the value function (4) (see also the three classical curses of dimensionality in Powell 2011): (i) The state space with several continuous dimensions complicates the representation and calculation of the value function $V_t(S_t)$. (ii) The expectation in $V_t(S_t)$ can be difficult to compute due to the continuous outcome space of $W_t$. Moreover, (iii) the continuous action space only allows for approximately determining the optimal action using the common hands-on discretization and by enumerating possible actions; further, the finer the discretization the more expensive the process.

In the following subsections, we tackle these challenges. We (i) start by simplifying the value function's representation through an approximation with basis functions, which (ii) also simplifies the calculation of the expectation (Sect. 4.1). This (iii) allows us to derive a closed-form solution for the commitment that is optimal with regard to the approximation (Sect. 4.2). Finally, we present the new policy iteration algorithm to fit this approximation in Sect. 4.3. Table 2 contains the notation additionally introduced in this section.

### 4.1 Approximation of the value function

Following a common technique in ADP, we approximate the value function at time $t$ using a linear architecture consisting of a linear combination of (time-dependent) basis functions (features) $(\phi_f)_{f \in F}$, i.e., $V_t(S_t) \approx \tilde{V}_t(S_t) = \sum_{f \in F} \theta_t^f \phi_f(S_t)$. Specifically, we use second degree polynomials. Set $N := 1 + (\tau - 1) + h_Y + h_P + h_Q$ so that $S_t \in \mathbb{R}^N$, let $S_t^i$ denote the $i$-th component of $S_t = (L_t, x_{t-\tau+1}, \ldots, x_{t-1}, W_t)$ as defined in Sect. 3.2, and denote by $\boldsymbol{\theta} = (\theta_t^{ij})$ the coefficients of the basis functions.

**Table 2** Notation introduced in Sect. 4

| | |
|---|---|
| *Parameters* | |
| $n_{\text{PI}}$ | Maximum number of policy iteration steps performed for each point in time |
| $n_{\text{paths}}$ | Number of sample paths evaluated in each policy evaluation step |
| $n_{\text{conv}}$ | Number of states sampled for convergence check |
| $\varepsilon_{\text{conv}}$ | Threshold for convergence check |
| *Additional notation* | |
| $F$ | Set of basis functions |
| $\tilde{V}_t(S_t)$ | Value function approximation: $\tilde{V}_t(S_t) \approx V_t(S_t)$ |
| $\boldsymbol{\theta} = (\theta_t^{ij})$ | Coefficients used in the approximation $\tilde{V}_t(S_t)$ |
| $\boldsymbol{v} = (v_t^m)$ | Accumulated contributions obtained in sample path $m$ from time $t$ onwards |

Then

$$\tilde{V}_t(S_t) = \theta_t^{00} + \theta_t^{N+1,0} t + \sum_{i=1}^{N} \left[ \theta_t^{i0} S_t^i + \theta_t^{ii} (S_t^i)^2 + \sum_{j=i+1}^{N} \theta_t^{ij} S_t^i S_t^j \right], \tag{5}$$

Note that time is included linearly with coefficient $\theta_t^{N+1,0}$ to allow the same set of coefficients to be used for several points in time.

### 4.2 Derivation of the best commitment

Now, the best action in state $S_t$ at time $t$ is to commit

$$\pi_t^*(S_t) = \underset{x_t \in [0, x_{\max}]}{\text{argmax}} \{C(S_t, x_t) + \beta \mathbb{E}[\tilde{V}_{t+1}(S_{t+1}) | S_t, x_t]\} \tag{6}$$

Given the approximation (5) and the conditions established in Sect. 3.1, this optimal commitment can be calculated as

$$\pi_t^*(S_t) = \begin{cases} \bar{x}_t^*(S_t) & \text{if } \theta_{t+1}^{\tau\tau} < 0 \wedge \bar{x}_t^* \in [0, x_{\max}] \\ \underset{x_t \in \{0, x_{\max}\}}{\text{argmax}} \{C(S_t, x_t) + \beta \mathbb{E}[\tilde{V}_{t+1}(S_{t+1}) | S_t, x_t]\} & \text{otherwise,} \end{cases} \tag{7}$$

with $\bar{x}_t^*(S_t) = -\frac{1}{2\theta_{t+1}^{\tau\tau}} (\frac{1}{\beta} P_t + \theta_{t+1}^{\tau 0} + \sum_{i \in \{1,\dots,N\} \setminus \tau} \theta_{t+1}^{\tau i} \mathbb{E}[S_{t+1}^i | S_t, x_t])$. See Appendices "Derivation of $x_t^*$" and "Derivation of $\mathbb{E}[L_{t+1} | S_t, x_t]$" for a respective detailed derivation of $\pi_t^*(S_t)$ and $\mathbb{E}[S_{t+1}^1 | S_t, x_t] = \mathbb{E}[L_{t+1} | S_t, x_t]$). If $\text{argmax}_{x_t \in \{0, x_{\max}\}} \{C(S_t, x_t) + \beta \mathbb{E}[\tilde{V}_{t+1}(S_{t+1}) | S_t, x_t]\}$ has to be computed, then $\mathbb{E}[(L_{t+1})^2 | S_t, x_t]$ is necessary, whose derivation is similar to that of $\mathbb{E}[L_{t+1} | S_t, x_t]$ given in Appendix

"Derivation of $\mathbb{E}[L_{t+1}|S_t, x_t]$". All the other expressions required for this case are known according to the conditions stated in Sect. 3.1.

The closed-form solution for $\pi_t^*(S_t)$ is key for an efficient ADP algorithm. The faster the calculation of the (partial) moments, the faster the evaluation of (7). This reduces the computational burden tremendously compared to the numerical solution approaches widely used in ADP, as we will show in our numerical experiments.

### 4.3 Approximate policy iteration

In this section, we present the new algorithm for approximately solving the MDP defined in Sect. 3. We call it Hybrid Approximate Policy Iteration (HAPI), because it blends elements from approximate policy iteration (API) with classical backward induction.

The basic structure of API algorithms (see, e.g., the textbooks by Powell 2011 and Bertsekas 2012 for a dynamic programming perspective, as well as that by Wiering and van Otterlo 2012 for a reinforcement learning perspective) is as follows: An inner loop evaluating a policy (policy evaluation step) is combined with an outer loop seeking to improve the policy (policy iteration step). Before starting the algorithm, the basis function coefficients are set to an initial value. This also determines the initial policy, which is obtained by calculating the optimal action, given the current value function approximation. Thereafter, in the policy evaluation step, the value of the current policy is estimated using a set of sample paths. Each sample path consists of an initial state of the system at the first point in time and the realizations of the exogenous processes until the end of the horizon. The system's evolution is simulated for each sample path. The process starts with the initial state. A decision is then made by following the current policy, and, given the decision and the realization of the exogenous process, the process moves on to the next state. Thereafter, a decision is again made, and so on. While this is done along the sample path, the accumulated contributions are calculated, representing the value of each state visited. Finally, these values are used to update the value function approximation, from which a new policy is derived in the policy iteration step. The next iteration starts with a new policy evaluation step.

In the following, we outline the key differences between the new HAPI algorithm shown in Fig. 2 and standard API algorithms. The parameters $n_{\mathrm{PI}}$ and $n_{\mathrm{paths}}$, respectively, denote the number of policy iteration steps and the number of sample paths used for each policy evaluation. The tolerance used to determine the convergence is $\varepsilon_{\mathrm{conv}}$ and the accumulated contributions obtained in the sample path $m$ from time $t$ onwards are stored in $\boldsymbol{v} = (v_{m,t'})_{\substack{t'=t,\ldots,T \\ m=1,\ldots,n_{\mathrm{paths}}}}$.

- Instead of starting every sample path in $t = 0$, accumulating the contributions, and updating $\theta_t^{ij}$ for all $t$ in each policy iteration, we employ a procedure working backwards in time, mimicking backward induction. This is reflected in the additional for-loop (step 2). Only after estimating the value function at time $t$ as best as possible (and, more importantly, obtaining a good policy), we move on to time $t-1$. Again, the initial states are sampled (now in $t-1$) and for each sample path $m$, the value $v_{m,t-1}$ of using the current policy is determined by following the sample

*1.* $\boldsymbol{\theta} = \mathbf{0}$
*2. for* $t = T - 1$ *to* 1
  *2.1. for* $k = 1$ *to* $n_{\text{PI}}$
    *2.1.1.* $\boldsymbol{v} = \mathbf{0}$
    *2.1.2. compute sample realizations* $Y_{m,t'}, P_{m,t'}, Q_{m,t'}\ \forall\ t' = t+1, \ldots, T,$
      $m = 1, \ldots, n_{\text{paths}}$
    *2.1.3. for* $m = 1$ *to* $n_{\text{paths}}$
      *2.1.3.1. sample initial state* $S_{m,t}$
      *2.1.3.2. for* $t' = t$ *to T-1*
        *2.1.3.2.1. determine optimal action* $x^*_{t'}\left(S_{m,t'}, \boldsymbol{\theta}\right)$
        *2.1.3.2.2.* $S_{m,t'+1} = T\left(S_{m,t'}, x^*_{t'}\left(S_{m,t'}, \boldsymbol{\theta}\right), Y_{m,t'+1}, P_{m,t'+1}, Q_{m,t'+1}\right)$
      *2.1.3.3. end for*
      *2.1.3.4.* $v_{m,t_1} = \sum_{t_2=t_1}^{T-1} \beta^{t_2-t_1} C\left(S_{m,t_2}, x^*_{t_2}\left(S_{m,t_2}, \boldsymbol{\theta}\right)\right)\ \forall t_1 = t, \ldots, T-1$
    *2.1.4.* $\boldsymbol{\theta}$ *=Update_Coefficients(*$\boldsymbol{v}$*,* $\boldsymbol{\theta}$*)*
    *2.1.5. if Policy_Converged(*$n_{\text{conv}}, \varepsilon_{\text{conv}}$*)*
      *2.1.5.1. break*
    *2.1.6. end if*
  *2.2. end for*
*3. end for*

**Fig. 2** Hybrid approximate policy iteration (HAPI) algorithm

path from $t-1$ until $T-1$. Moreover, we obtain $v_{m,t'}$ for all states visited on the path. All these values are then used to update the value function approximation. In doing so, we take advantage of the good policy already determined for all the following stages $t, t+1, \ldots, T-1$. Thus, the approach simultaneously ensures both broad exploration by sampling the initial states for every point in time $t$ and emphasis on the regions of the state space often visited.

- To fit the value function approximation (5), we use an approach initially proposed by Lagoudakis and Parr (2003). Specifically, we take advantage of the linear architecture and perform ordinary least squares to minimize the $L_2$-norm between the true (sampled) value function and the approximation. In this context, we use recursive least squares to minimize the computational burden incurred by updating the basis function coefficients in step 2.1.4. This technique is quite popular, but every new estimate of a state's value is usually immediately used to update the coefficients (see, e.g., Powell 2011). We speed up the process considerably by collecting the accumulated contributions for all the sample paths and performing recursive least squares in a batch after each policy evaluation, which is formally equivalent.
- The policy convergence is checked in step 2.1.5 and, if the policy has converged, the algorithm moves on to time $t-1$. The convergence check works as follows: A given number of $n_{\text{conv}}$ states is sampled. In these states, the policy at time $t-1$ resulting from the new coefficients is compared to the one resulting from the

**Table 3** Notation introduced in Sect. 5

| | |
|---|---|
| *Parameters* | |
| $\mu_P$ | Mean of the market price |
| $\kappa_P$ | Mean-reversion parameter of the market price |
| $\sigma_P$ | Standard deviation of the change in market price |
| $m$ | Slope of the penalty cost with regard to the market price |
| $s_{\text{ci}}$ | Cut-in speed |
| $s_{\text{co}}$ | Cut-out speed |
| $s_r$ | Rated speed |
| $r$ | Rated power |
| $\lambda$ | Scale parameter of the Weibull distribution for wind speed |
| $k$ | Shape parameter of the Weibull distribution for wind speed |
| *Additional notation* | |
| $\mathcal{N}(\mu, \sigma^2)$ | Normal distribution with mean $\mu$ and variance $\sigma^2$ |
| $\text{WB}(\lambda, k)$ | Weibull distribution with scale parameter $\lambda$ and shape parameter $k$ |
| $\text{WS}_t$ | Wind speed between time $t+\tau-1$ and $t+\tau$ |
| $a, b$ | Auxiliary parameters for wind turbine |
| $\text{sp}(y)$ | Wind speed providing the energy production $y$ |

previous coefficients as the coefficients $\theta_t^{ij}$ determine the policy at time $t-1$. If the mean absolute deviation of the policy over all the states is less than the parameter $\varepsilon_{\text{conv}}$, the policy is assumed to have converged and we move on to the next point in time $t-1$.

## 5 Optimizing a wind farm's commitments

In this section, we use our approach to optimize the policy of a wind farm with a storage device. In the following subsections, we explain the specific form of the exogenous processes used for $W_t$ in detail and show that the conditions in Sect. 3.1 are satisfied. We begin with the exogenous processes governing the prices (Sect. 5.1) and thereafter describe the process for the generation of electricity (Sect. 5.2). The model resulting from these specific stochastic processes is described in Sect. 5.3. Table 3 provides an overview of the notation newly introduced in this section.

### 5.1 Exogenous process for electricity prices

In the literature, the standard way to model electricity prices is by a mean-reverting process (Möst and Keles 2010). Accordingly, we assume a discrete-time version of the popular Ornstein–Uhlenbeck process (used, among others, by Kim and Powell 2011):

$$P_{t+1} = P_t + \kappa_P(\mu_P - P_t)\Delta t + \hat{P}_{t+1}. \tag{8}$$

The mean-reversion parameter $\kappa_P$ governs how fast the price returns to the mean price $\mu_P$ and is proportional to the expected frequency at which it crosses its mean per unit

time. The random variables $\hat{P}_t$ are i.i.d. with distribution $\mathcal{N}(0, \sigma_P^2)$ and capture the noise in the evolution of $P_t$. Transferred into the notation in Sect. 3.1 this leads to $h_P = 1, \alpha_1^P = 1 - \kappa_P \cdot \Delta t$ and $\hat{p}_{t+1} \sim \mathcal{N}(\mu_P \cdot \kappa_P \cdot \Delta t, \sigma_P^2)$.

Regarding the penalty payment incurred when a commitment cannot be met and energy is bought on the regulating market, we follow the simple and widely used assumption that the price depends linearly on the price on the intraday market with slope $m > 1$ (see, e.g., Kim and Powell 2011 or Löhndorf and Minner 2010):

$$P_t = m \cdot Q_t \tag{9}$$

However, our commitments are delivered with a time lag of $\tau$. Thus, the commitment $x_t$ is sold at price $P_t$, but the penalty price relevant when commitment $x_t$ cannot be met between time $t + \tau - 1$ and $t + \tau$ is $Q_{t+\tau}$, which in turn depends on $P_{t+\tau}$.

### 5.2 Exogenous process for energy production

For ease of presentation and in line with the literature (see, e.g., Kim and Powell 2011), we assume that the production of our small wind farm is independent of prices (wind has zero marginal cost). The wind turbine's production $Y_t$ depends only on the realized wind speed $\mathrm{WS}_t$ with $Y_t = Y(\mathrm{WS}_t)$, and is modeled using the following formula:

$$Y(\mathrm{WS}_t) = \begin{cases} 0 & \text{if } \mathrm{WS}_t < s_{\mathrm{ci}} \text{ or } \mathrm{WS}_t \geq s_{\mathrm{co}} \\ (a + b \cdot \mathrm{WS}_t^3) \cdot \Delta t & \text{if } \mathrm{WS}_t \in [s_{\mathrm{ci}}, s_r) \\ r \cdot \Delta t & \text{if } \mathrm{WS}_t \in [s_r, s_{\mathrm{co}}) \end{cases} \tag{10}$$

Since the production values of 0 and $r \cdot \Delta t$ are linked to intervals of wind speeds, the cdf of $Y_t$ is not continuous but exhibits jumps at these values.

The wind turbine only operates at wind speeds between its cut-in speed $s_{\mathrm{ci}}$ and the cut-out speed $s_{\mathrm{co}}$. Above its rated speed $s_r$, the turbine produces its rated power $r$. Between the cut-in speed $s_{\mathrm{ci}}$ and the rated speed, the power output at wind speed $\mathrm{WS}_t$ is determined by $a + b \cdot \mathrm{WS}_t^3$ for $\mathrm{WS}_t \in [s_{\mathrm{ci}}, s_r)$ with $a, b \in \mathbb{R}$. Here, the linear equations $a + b \cdot s_{\mathrm{ci}}^3 = 0$ and $a + b \cdot s_r^3 = r$ give the values $a$ and $b$.

We again follow the standard assumption of a Weibull distribution regarding the wind speed, i.e., $\mathrm{WS}_t \sim \mathrm{WB}(\lambda, k)$ (see, e.g., Justus et al. 1976 or Seguro and Lambert 2000) with pdf and cdf denoted by

$$\begin{aligned} f_{\mathrm{WB}(\lambda,k)}(\mathrm{WS}) &= \lambda k (\lambda \cdot \mathrm{WS})^{k-1} \mathrm{e}^{-(\lambda \cdot \mathrm{WS})^k}, \\ F_{\mathrm{WB}(\lambda,k)}(\mathrm{WS}) &= 1 - \mathrm{e}^{-(\lambda \mathrm{WS})^k} \forall \mathrm{WS} \geq 0 \end{aligned} \tag{11}$$

From (10) and (11), we obtain the cdf of $Y_t$:

$$F_{Y_t}(y) = \begin{cases} 0 & \text{if } y < 0 \\ (1 - \mathrm{e}^{-(\lambda s_{\mathrm{ci}})^k} + \mathrm{e}^{-(\lambda s_{\mathrm{co}})^k}) & \\ \quad + F_{\mathrm{WB}(\lambda,k)}(\mathrm{sp}(y)) - F_{\mathrm{WB}(\lambda,k)}(s_{\mathrm{ci}}) & \text{if } 0 \leq y < \Delta t \cdot r \\ 1 & \text{if } y \geq \Delta t \cdot r, \end{cases} \tag{12}$$

where the wind speed providing the energy production $y$ is denoted by $\text{sp}(y) = (\frac{\frac{y}{\Delta t}-a}{b})^{\frac{1}{3}} \forall y \in [0, \Delta t \cdot r)$. Note that we assume that the wind speeds $\text{WS}_t$, and, thus, the energy productions $Y_t$ are i.i.d. This assumption leads to a tractable benchmark mechanism for HAPI by reducing the dimension of the state space. However, in this framework, it is also possible to use more complex stochastic processes as shown in Sects. 3 and 4.

The derivations of the first and second partial moments of $Y_t$ can be found in Appendices "Derivation of $\int y \mathrm{d}F_{Y_t}(y)$" and "Derivation of $\int y^2 \, \mathrm{d}F_{Y_t}(y)$". Thus, the stochastic processes for prices and energy production satisfy the conditions stated in Sect. 3.1. All moments can be evaluated easily and quickly.

### 5.3 State space, optimal action, and contribution

As energy production is assumed to be i.i.d., we do not have to include its history in the state space, i.e., $h_Y = 0$. Market price $P_t$ is an autoregressive process of order 1, and $P_{t+1}$ and $Q_{t+1}$ only depend on $P_t$, because $Q_t = m \cdot P_t$, i.e., $h_P = 1$ and $h_Q = 0$. In our numerical experiments, we consider a time lag of $\tau = 4$. Therefore, we have to consider the last three commitments made and the state definition is $S_t = (L_t, x_{t-3}, x_{t-2}, x_{t-1}, P_t) \in \mathbb{R}^5$ as $W_t = (P_t)$.

Regarding the optimal action $\pi_t^*(S_t)$ given by (7), we now have

$$\pi_t^*(S_t) = \begin{cases} \bar{x}_t^*(S_t) & \text{if } \theta_{t+1}^{44} < 0 \wedge \bar{x}_t^* \in [0, x_{\max}] \\ \text{argmax}_{x_t \in \{0, x_{\max}\}} \{C(S_t, x_t) + \beta \mathbb{E}[\tilde{V}_{t+1}(S_{t+1}) | S_t, x_t]\} & \text{otherwise} \end{cases} \tag{13}$$

with $\bar{x}_t^*(S_t) = -\frac{1}{2\theta_{t+1}^{44}} \Big( \frac{1}{\beta} P_t + \theta_{t+1}^{40} + \theta_{t+1}^{41} \mathbb{E}[L_{t+1} | S_t, x_t] + \theta_{t+1}^{42} x_{t-2} + \theta_{t+1}^{43} x_{t-1} + \theta_{t+1}^{45} \mathbb{E}[P_{t+1} | S_t, x_t] \Big)$. Remember that the first moment of $L_{t+1}$, which is required to calculate (13), is given in Appendix "Derivation of $\mathbb{E}[L_{t+1} | S_t, x_t]$". Additional terms are needed to calculate the expected value in the second line: the first and second moments of $P_{t+1}$ and $Q_{t+1}$, as well as the second moment of $L_{t+1}$. We omit these straightforward calculations to keep our presentation concise. Additionally, we use the analytical expression for the contribution $C(S_t, x_t)$ stated in Appendix "Analytical expression for the contribution". As all these expressions are quickly computable, the optimal commitment $\pi_t^*(S_t)$ can be determined efficiently.

## 6 Numerical experiments

We performed numerical experiments to test the new HAPI approach using real-world data. All the implementations were done with MATLAB version R2013a, and were run on a PC with a 2.8 GHz Intel Core i7 processor and 8 GB of RAM, running on Microsoft Windows Server 2008 R2 64 bit. We did not use parallelization, although the structure of the algorithms clearly allows this. In the following, we first consider the setup

**Table 4** Notation introduced in Sect. 6

| | |
|---|---|
| *Parameters* | |
| ssl | Safety stock level of the certainty-equivalent heuristic CE |
| $\mu_{\text{WS}}$ | Mean of the wind speed |
| $\sigma_{\text{WS}}$ | Standard deviation of the wind speed |

(Sect. 6.1) and describe the tested methods (Sect. 6.1.1), as well as the data (Sect. 6.1.2). Thereafter, we turn to the results (Sect. 6.2) and analyze the methods' performance (Sect. 6.2.1). Finally, we investigate the influence of two major storage parameters: size and efficiency (Sect. 6.2.2). Again, the new notation is summarized in Table 4.

## 6.1 Simulation experiment design

### *6.1.1 Methods tested*

The following methods, of which several are benchmark procedures, were implemented to determine the policy (i.e., the commitments) to compare the new approach's performance:

- HAPI($\varepsilon_{\text{conv}}$) is the approach described in Sects. 4 and 5 with the tolerance used to determine convergence set to $\varepsilon_{\text{conv}}$. To improve the convergence, we used identical sets of coefficients for several points in time. Different coefficients were only used for the last $\tau + 1$ points in time to better capture the end-of-horizon effects. The states were sampled by adopting a space-filling design, i.e., a low-discrepancy Faure sequence.
- DISC is probably the most important benchmark and is also used in most of the literature cited in Sect. 2. Following the standard hands-on approach, the continuous dimensions of the state and action space of the original MDP described in Sect. 3 are discretized. The resulting discrete MDP is solved to optimality using the common backward induction. Thus, we have an approximation of the underlying continuous problem. In preliminary tests not reported here, we determined a compromise between the runtime and the solution quality and, thus, discretize each of the five dimensions with five equidistant points, resulting in 3125 states. The discretization of energy production does not influence the number of states and we used a finer equidistant grid with 100 points here.
- EV is a very simple heuristic that always commits the expected value of energy production. If the selling price is negative, the commitment is zero.
- CE(*ssl*) is a certainty-equivalent heuristic that aims at a safety stock level of ssl $\cdot$ $L_{\max}$. The evolution of the storage level is predicted for the next $\tau$ points in time under the assumption that energy production will be equal to the expected value. Taking past commitments into account, the commitment is then chosen such that the predicted storage after the commitment's delivery is equal to ssl $\cdot$ $L_{\max}$. Again, if the selling price is negative, the commitment is zero.

Please note that we did not investigate a purely myopic policy. This method only considers the immediate contribution on each stage, i.e., maximizes $P_t x_t$ in (3) as this

is the only part of the contribution depending explicitly on $x_t$. Obviously, this yields $x_{\max}$ for $P_t > 0$ and, thus, very high penalty payments in future periods. Preliminary tests confirmed this and indicated large negative overall profits for all test instances.

### *6.1.2 Data used*

Our numerical experiments are geared to the German intraday market, i.e., energy is traded in 15-min periods until 45 min before delivery ($\Delta t = 0.25$ and $\tau = 4$). Regarding the storage size and rated energy production, we emulate the wind farm/storage combination of the earlier mentioned German technology and service supplier Bosch (see Sect. 1). Accordingly, we consider a storage device with a capacity of $L_{\max} = 2.5$ MWh and a roundtrip efficiency of $\rho_R \cdot \rho_E = 90\ \%$. The properties of the wind turbine are based on the Siemens SWT-3.0-113 turbine with a hub height of 99.5 m. We assume a maximum power output of $r = 20$ MW and a 25-MW connection to the grid, resulting in a maximum commitment of $x_{\max} = 6.25$ MWh. All the parameter values are summarized in Table 5.

We next describe the parameters used for the stochastic processes. Price and weather data were available for the eight months from January to August 2013. Wind data were obtained from a meteorological station at our university in Augsburg, where a considerable number of wind farms already operate. The wind speeds were scaled to the hub height of 99.5 m of the considered turbine. The price data were obtained from EPEX SPOT for 15-min contracts with delivery in the German TSO zone. We split the data per month and estimated the stochastic processes for price and energy production for each month, obtaining a total of eight problem instances. The parameters estimated

**Table 5** Parameter values shared by all instances

| Parameter | Value |
|---|---|
| *Wind turbine* | |
| $s_{\text{ci}}$ | 3 m/s |
| $s_{\text{co}}$ | 25 m/s |
| $s_r$ | 12 m/s |
| $r$ | 20 MW |
| *Storage* | |
| $L_{\max}$ | 2.5 MWh |
| $\rho_R$ | $\sqrt{0.9}$ |
| $\rho_E$ | $\sqrt{0.9}$ |
| *Other parameters* | |
| $T$ | 20 |
| $\Delta t$ | 0.25 h |
| $\beta$ | 1 |
| $m$ | 2 |
| $x_{\max}$ | 6.25 MWh |

**Table 6** Price and wind speed parameters

| | Wind speed [m/s] | | | | Price [Euro/MWh] | | |
|---|---|---|---|---|---|---|---|
| | $\lambda$ | $k$ | $\mu_{\text{WS}}$ | $\sigma_{\text{WS}}$ | $\kappa_P$ | $\mu_P$ | $\sigma_P$ |
| January (Instance 1) | 0.127 | 1.430 | 7.145 | 5.072 | 1.035 | 40.712 | 12.693 |
| February (Instance 2) | 0.135 | 1.499 | 6.663 | 4.527 | 0.899 | 44.357 | 11.300 |
| March (Instance 3) | 0.143 | 1.598 | 6.289 | 4.030 | 1.150 | 39.112 | 18.475 |
| April (Instance 4) | 0.132 | 1.712 | 6.757 | 4.065 | 1.114 | 38.993 | 17.593 |
| May (Instance 5) | 0.129 | 1.677 | 6.943 | 4.257 | 1.441 | 34.981 | 16.075 |
| June (Instance 6) | 0.150 | 1.632 | 5.981 | 3.758 | 1.244 | 30.401 | 17.036 |
| July (Instance 7) | 0.148 | 1.653 | 6.052 | 3.760 | 1.700 | 38.202 | 16.199 |
| August (Instance 8) | 0.165 | 1.553 | 5.438 | 3.576 | 1.433 | 35.824 | 15.015 |

for each instance (rounded to three decimal places) are shown in Table 6, where the mean and standard deviation of the wind speed ($\mu_{\text{WS}}$ and $\sigma_{\text{WS}}$, respectively) are also included for illustration.

We evaluated the policies determined by the methods described in Sect. 6.1.1 by means of simulation and report average profits obtained from 10,000 simulation runs performed for each combination of method and problem instance described above. A simulation run mimics the real-world problem described in Sect. 3. At time $t$, the producer knows the total energy production between time $t-1$ and $t$ and the resulting storage level, as well as the past commitments that have to be delivered in the future. Moreover, the current selling price and the penalty price are observed. Using this information, the producer decides on the commitment. At time $t+1$, again new information is revealed; a new decision is made, and so on. Identical random numbers were used to evaluate all the methods on a particular problem instance.

## 6.2 Numerical results

### *6.2.1 Performance of HAPI*

A number of preliminary tests were performed to determine the parameters of the HAPI method. To avoid overfitting, these tests also included additional artificial problem instances. The maximum number of policy iterations was set to $n_{\text{PI}} = 50$, but this value was almost never limiting. The number of sample paths per policy evaluation was set to $n_{\text{paths}} = 500$ and convergence was determined using $n_{\text{conv}} = 100$ sample states. As in all iterative algorithms, the threshold value of convergence $\varepsilon_{\text{conv}}$ was the key to influencing the tradeoff between the runtime and the solution quality.

Figure 3 visualizes the tradeoff between the average profit and the runtime for determining the policy of HAPI relative to DISC. It considers different values of the convergence threshold $\varepsilon_{\text{conv}}$ and displays results averaged over all eight instances. To consider extreme cases, we also included fixed numbers of 1, 10 and 50 policy iterations (1 PI, 10 PI, 50 PI) without performing a convergence check. Overall, the results

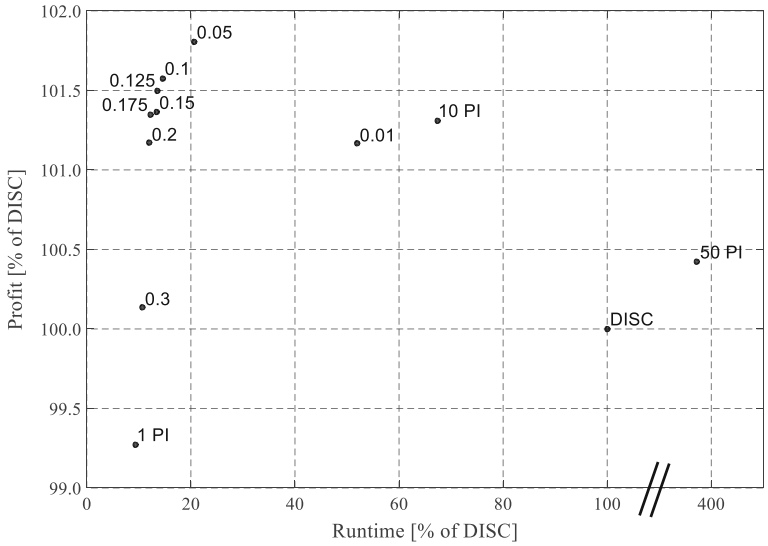


**Fig. 3** Runtime and profit of HAPI relative to DISC for different values of $\varepsilon_{\text{conv}}$, averaged over all eight instances

were as expected for $\varepsilon_{\text{conv}} \geq 0.05$: as this value increases, fewer policy iterations are performed, and the runtime decreases, but profits also tend to decrease. HAPI performs best for $0.05 \leq \varepsilon_{\text{conv}} \leq 0.2$. The runtime is between 12 and 20 % of DISC's and profits are about 1.5 % higher.

Using a fixed number of only 1 policy iteration is fast, but clearly not enough and yields a low profit. On the other hand, using a lower value of $\varepsilon_{\text{conv}} = 0.01$, or performing constant numbers of $n_{\text{PI}} = 10$ and $n_{\text{PI}} = 50$ policy iterations, takes much longer. However, surprisingly, profit is also low, with more policy iterations leading to lower profits in each of the eight instances considered. This counter-intuitive observation can be explained as follows: Mimicking backward induction, the current period's entire state space is sampled using a space-filling design to obtain a good exploration. When HAPI moves on and subsequently does the same for prior periods, these periods' state space is again sampled. Each sample is used as the initial state of a sample path corresponding to a possible evolution of the system and observations are obtained for all the periods until the end of the horizon. These observations are used to update the value function. In doing so, the approximation becomes better in areas of the state space that are more likely to be visited. While this generally improves the profit, the results depicted, as well as additional tests, hint that a certain balance is important and visiting similar states over and over again can offset the initial sampling.

Next, we compare in detail HAPI(0.125) and HAPI(0.175) with DISC, as well as the additional heuristics EV and CE (Fig. 4). Regarding CE, we considered the safety stock levels of ssl = 0 and ssl = 0.75, the value that performed best in a preliminary experiment using the eight instances.

Whereas Fig. 3 already shows that the profits of HAPI and DISC are, on average, very close, Fig. 4 shows that the corresponding profits are also close in each prob-

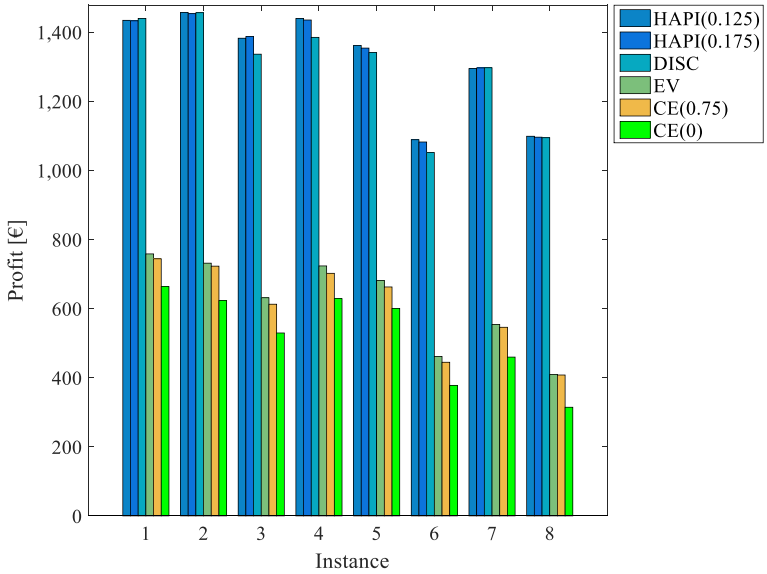


**Fig. 4** Comparison of HAPI, DISC, and additional heuristics EV and CE

lem instance. However, there is considerable variation between different instances. Instances 6 and 8 seem to be somehow worse for the wind farm than the others, probably due to a combination of low average prices and low average wind speeds. To a lesser extent, this also applies to Instance 7. The profit of the additional heuristics EV, CE(0), and CE(0.75) also reflect this general trend, although there is a huge gap between it and the dynamic programming approaches. In each instance, the heuristics obtain only about 29–53 % of the aforementioned approaches' profit. Of the additional heuristics, EV consistently obtains the highest profit (37–53 % of DISC), CE(0.75) follows closely (37–52 %), and CE(0) obtains the lowest profit (29–46 %).

Finally, we investigate the advantage of analytically determining the optimal action (see Sect. 4.2), a key contribution of this paper. We, therefore, modified HAPI and used Matlab's optimization routine `fmincon` (part of the Optimization Toolbox) to find the maximum of (6) instead of using (7). Figure 5 displays the runtime of HAPI(0.125) with *fmincon*. We also included the runtime of the original HAPI(0.125) using (7) and DISC for comparison. The runtime of the heuristics EV and CE is negligible and, thus, not given here. The figure is impressive, showing the 15- to 20-fold speedup due to the analytical maximization. Moreover, the figure depicts the eight problem instances' variations in runtime. With `fmincon`, the slowest instance takes about 1.3 times as long as the fastest. With the original algorithm, the factor is 1.5, but the absolute runtimes are also strikingly lower here. Finally, note that both the variants' profits are virtually the same.

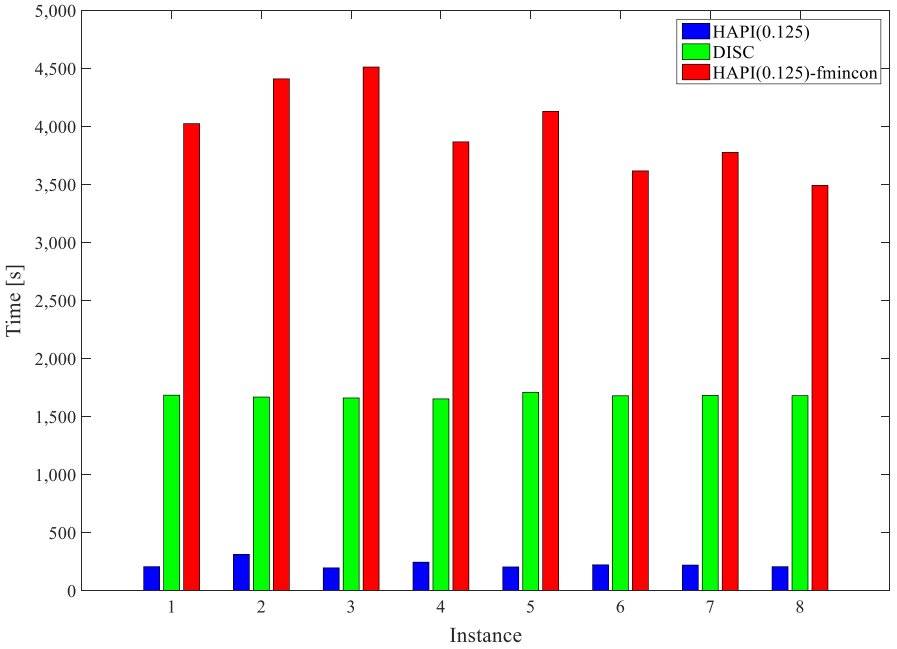


**Fig. 5** Runtime with analytical and numerical derivation of optimal action

### *6.2.2 Value added by the storage*

In this section, we investigate the influence of two important properties of an electricity storage on the profit: size and efficiency. In particular, we consider the storage sizes $L_{\max} \in \{0, 1.25, 2.5, 3.75, 5, 6.25, 7.5\}$ and the charge/discharge efficiencies $\rho_R = \rho_E$ with $\rho_R \rho_E \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1\}$.

Figure 6 shows the profit obtained by HAPI(0.125) for all combinations of $L_{\max}$ and $\rho_R \rho_E$, averaged over all eight problem instances. As expected, the profit increases with the storage size and efficiency, albeit at a decreasing rate, and the value of additional storage is higher if the efficiency is high. Combined with possible storage devices' parameters, or with the marginal cost of size vs. efficiency, this analysis can support the decision for appropriate power storage. But even without these specifications, we can conclude that a small storage with very poor efficiency (lower left corner of the figure) already increases profits considerably in the problem instances considered.

## 7 Conclusions

In this paper, we have developed an ADP approach to the problem of making advance commitments for a producer with a renewable, intermittent energy source and a finite storage device with conversion losses. In particular, we modeled the problem as an MDP and specifically considered the lagged advance commitments typical of modern electricity markets, where energy is traded in advance. To efficiently solve this MDP,

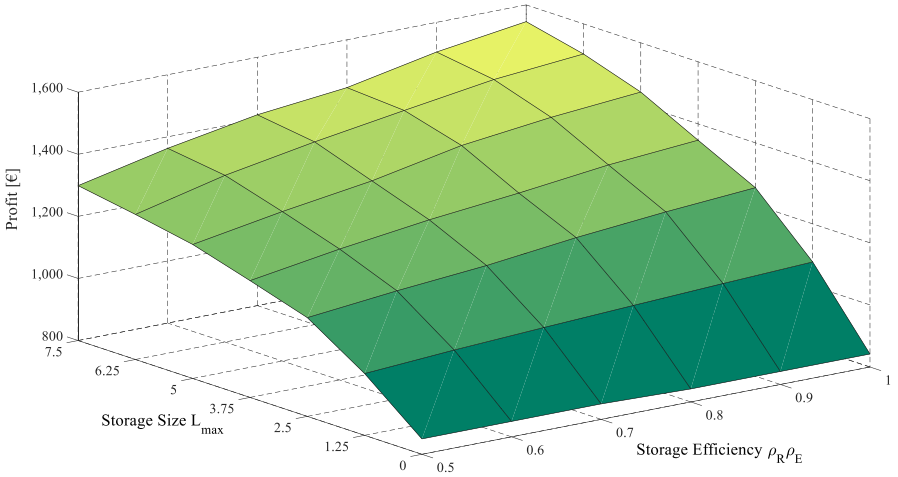


**Fig. 6** Profit influence of storage efficiency and size

we proposed an innovative ADP framework. Its major advantage is the efficient calculation of the best action, requiring only mild conditions on the stochastic processes describing the energy production and price evolution. We use the framework to consider a wind farm/storage combination with energy production sold on an intraday market where wind speed and prices, respectively, follow a Weibull distribution and an Ornstein–Uhlenbeck process. An extensive numerical study shows that our approach needs only a fraction of the runtime of the common MDP-based approximation with a discretized state and action space, and even obtains slightly higher profits. Moreover, a small storage with a very low efficiency can already increase the profit considerably.

We consider these results promising for the application of ADP approaches. Our approach scales much better than an MDP-based discrete approximation. This enables extensions of the state space and allows, for example, the consideration of higher-order autoregressive processes and an increased number of past commitments due to specific market structures.

There are basically two avenues for future research. The first avenue is application oriented. As this paper focused on developing the framework and the new HAPI approach, a key aspect of the numerical study was HAPI's performance. Future work could, for example, concentrate on the actual outcome/benefit and also consider more sophisticated processes, like non-i.i.d. wind conditions or constrained power flows into/out of the storage (see also Hassler 2016). The second avenue extends the methodology itself and combines the approach presented with building blocks that have performed well in similar problems. In particular, concepts and algorithms from the field of reinforcement learning might prove useful (see, e.g., Wiering and van Otterlo 2012 for an overview). This may include actor-critic algorithms that use separate approximations for the policy (the actor) and the value function (the critic) which were introduced by Barto et al. (1983) and Sutton (1984). These algorithms provably converge to a local optimum under certain conditions (for details see Konda and Tsitsiklis 1999) and could enhance the performance of HAPI. HAPI itself can be classified in this framework as a critic-only algorithm as it solely uses an

approximation of the value function but not a separate approximation of the policy. Other modifications of HAPI could include, for example, the usage of different basis functions, such as higher order polynomials (see, e.g., Löhndorf and Minner 2010) and Gaussian radial basis functions (see, e.g., Jiang et al. 2014). Likewise, researchers could investigate adapting this approach to a direct policy search (see, e.g., Scott and Powell 2012; Nascimento and Powell 2013) or to an approximate value iteration (AVI, see, e.g., Jiang and Powell 2015). Finally, future studies could consider the preservation and exploitation of the monotonicities inherent in the value function (Nascimento and Powell 2013; Jiang and Powell 2015).

**Acknowledgments** Dr. Joachim Rathmann (Chair of Physical Geography and Quantitative Methods, University of Augsburg) kindly provided weather data. Michael Hassler acknowledges the financial support of BMBF Project 03EK3015 (ENREKON).

## Appendix

### Derivation of $x_t^*$

In the following, we derive the optimal commitment $\pi_t^*(S_t)$ used in (7) in Sect. 4.2. In particular, we show that

$$\pi_t^*(S_t) = \underset{x_t \in [0, x_{\max}]}{\operatorname{argmax}} \{C(S_t, x_t) + \beta \mathbb{E}[\tilde{V}_{t+1}(S_{t+1}) | S_t, x_t]\}$$

is given by

$$\pi_t^*(S_t) = \begin{cases} \bar{x}_t^*(S_t) & \text{if } \theta_{t+1}^{\tau\tau} < 0 \wedge \bar{x}_t^* \in [0, x_{\max}] \\ \underset{x_t \in \{0, x_{\max}\}}{\operatorname{argmax}} \{C(S_t, x_t) + \beta \mathbb{E}[\tilde{V}_{t+1}(S_{t+1}) | S_t, x_t]\} & \text{otherwise} \end{cases}$$

with $\bar{x}_t^*(S_t) = -\frac{1}{2\theta_{t+1}^{\tau\tau}} (\frac{1}{\beta} P_t + \theta_{t+1}^{\tau 0} + \sum_{i \in \{1,\dots,N\} \setminus \tau} \theta_{t+1}^{\tau i} \mathbb{E}[S_{t+1}^i | S_t, x_t])$.

Set $N := 1 + (\tau - 1) + h_Y + h_P + h_Q$, so that $S_{t+1} \in \mathbb{R}^N$ and remember that $S_{t+1}^i$ is the $i$-th component of $S_{t+1} = (L_{t+1}(L_t, Y_{t+1}, x_{t-\tau+1}), x_{t-\tau+2}, \dots, x_t, W_{t+1}(W_t, Y_{t+1}, P_{t+1}, Q_{t+1}))$. We are searching for a global optimum of $C(S_t, x_t) + \beta \mathbb{E}[\tilde{V}_{t+1}(S_{t+1}) | S_t, x_t]$ on the interval $[0, x_{\max}]$. Taking the derivative with respect to $x_t$, we obtain $\frac{\partial}{\partial x_t} \{C(S_t, x_t) + \beta \mathbb{E}[\tilde{V}_{t+1}(S_{t+1}) | S_t, x_t]\} = \frac{\partial}{\partial x_t} C(S_t, x_t) + \beta \frac{\partial}{\partial x_t} \mathbb{E}[\tilde{V}_{t+1}(S_{t+1}) | S_t, x_t]$.

Regarding the first summand, we directly obtain $\frac{\partial}{\partial x_t} C(S_t, x_t) = P_t$ from Eq. (3).

Regarding $\frac{\partial}{\partial x_t} \mathbb{E}[\tilde{V}_{t+1}(S_{t+1}) | S_t, x_t]$, we have

$$\mathbb{E}[\tilde{V}_{t+1}(S_{t+1}) | S_t, x_t] = \theta_t^{00} + \theta_{t+1}^{N+1,0}(t+1) + \sum_{i=1}^{N} \theta_{t+1}^{i0} \mathbb{E}[S_{t+1}^i | S_t, x_t] + \sum_{i=1}^{N} \theta_{t+1}^{ii} \mathbb{E}[(S_{t+1}^i)^2 | S_t, x_t] + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \theta_{t+1}^{ij} \mathbb{E}[S_{t+1}^i S_{t+1}^j | S_t, x_t].$$

As $x_t = S^{\tau}_{t+1}$ is known at time $t$ we continue with

$$\begin{aligned}\mathbb{E}[\tilde{V}_{t+1}(S_{t+1})|S_t, x_t] = \theta^{00}_t + \theta^{\tau 0}_{t+1} x_t + \theta^{\tau\tau}_{t+1} x_t^2 + \sum_{i\in\{1,\dots,N\}\setminus\tau} \theta^{\tau i}_{t+1} x_t \mathbb{E}[S^i_{t+1}|S_t, x_t] \\ + \theta^{N+1,0}_{t+1}(t+1) + \sum_{i\in\{1,\dots,N\}\setminus\tau} \theta^{i0}_{t+1} \mathbb{E}[S^i_{t+1}|S_t, x_t] \\ + \sum_{i\in\{1,\dots,N\}\setminus\tau} \theta^{ii}_{t+1} \mathbb{E}[(S^i_{t+1})^2|S_t, x_t] \\ + \sum_{i\in\{1,\dots,N\}\setminus\tau} \sum_{j=i+1, j\neq\tau}^{N} \theta^{ij}_{t+1} \mathbb{E}[S^i_{t+1} S^j_{t+1}|S_t, x_t].\end{aligned}$$

Thus, $\frac{\partial}{\partial x_t}\mathbb{E}[\tilde{V}_{t+1}(S_{t+1})|S_t, x_t] = \theta^{\tau 0}_{t+1} + 2\theta^{\tau\tau}_{t+1} x_t + \sum_{i\in\{1,\dots,N\}\setminus\tau} \theta^{\tau i}_{t+1} \mathbb{E}[S^i_{t+1}|S_t, x_t]$ is obtained.

Summing up, we have $\frac{\partial}{\partial x_t}\{C(S_t, x_t) + \beta\mathbb{E}[\tilde{V}_{t+1}(S_{t+1})|S_t, x_t]\} = P_t + \beta\theta^{\tau 0}_{t+1} + 2\beta\theta^{\tau\tau}_{t+1} x_t + \beta\sum_{i\in\{1,\dots,N\}\setminus\tau} \theta^{\tau i}_{t+1} \mathbb{E}[S^i_{t+1}|S_t, x_t]$ and the second derivative is obviously $\frac{\partial^2}{\partial x_t^2}\{C(S_t, x_t) + \beta\mathbb{E}[\tilde{V}_{t+1}(S_{t+1})|S_t, x_t]\} = 2\beta\theta^{\tau\tau}_{t+1}$.

Next, depending on the sign of $\theta^{\tau\tau}_{t+1}$, two cases have to be distinguished:

- If $\theta^{\tau\tau}_{t+1} < 0$, $C(S_t, x_t) + \beta\mathbb{E}[\tilde{V}_{t+1}(S_{t+1})|S_t, x_t]$ is a strictly concave function of $x_t$. So, $\pi^*_t(S_t)$ is equal to $-\frac{1}{2\theta^{\tau\tau}_{t+1}}(\frac{1}{\beta}P_t + \theta^{\tau 0}_{t+1} + \sum_{i\in\{1,\dots,N\}\setminus\tau} \theta^{\tau i}_{t+1} \mathbb{E}[S^i_{t+1}|S_t, x_t])$ if this is a valid commitment or $\pi^*_t(S_t) \in \{0, x_{\max}\}$.
- If $\theta^{\tau\tau}_{t+1} \geq 0$, $C(S_t, x_t) + \beta\mathbb{E}[\tilde{V}_{t+1}(S_{t+1})|S_t, x_t]$ is either an affine linear, or a strictly convex function of $x_t$ for $\theta^{\tau\tau}_{t+1} = 0$ or $\theta^{\tau\tau}_{t+1} > 0$, respectively. In this case, $\pi^*_t(S_t) \in \{0, x_{\max}\}$.

### Derivation of $\mathbb{E}[L_{t+1}|S_t, x_t]$

The computation of the optimal action $\pi^*_t(S_t)$ in Sect. 4.2 necessitates determining the expected value of the storage level at time $t+1$. We derive $\mathbb{E}[L_{t+1}|S_t, x_t]$ in the following:

$$\begin{aligned}\mathbb{E}[L_{t+1}|S_t, x_t] = 0 \cdot F_{Y_{t+1}}(x_{t-\tau+1} - \rho_E L_t) + L_{\max} \\ \cdot \left(1 - F_{Y_{t+1}}\left(\frac{L_{\max} - L_t}{\rho_R} + x_{t-\tau+1}\right) + \mathbb{P}\left(Y_{t+1} = \frac{L_{\max} - L_t}{\rho_R} + x_{t-\tau+1}\right)\right) \\ + \int_{(x_{t-\tau+1} - \rho_E L_t, x_{t-\tau+1}]} \left(L_t - \frac{1}{\rho_E}(x_{t-\tau+1} - y)\right) \mathrm{d}F_{Y_{t+1}}(y) \\ + \int_{(x_{t-\tau+1}, \frac{L_{\max} - L_t}{\rho_R} + x_{t-\tau+1})} \left(L_t + \rho_R(y - x_{t-\tau+1})\right) \mathrm{d}F_{Y_{t+1}}(y)\end{aligned}$$

$$
\begin{aligned}
&= L_{\max} \cdot \left(1 - F_{Y_{t+1}}\left(\frac{L_{\max} - L_t}{\rho_R} + x_{t-\tau+1}\right)\right.\\
&\quad \left. + \mathbb{P}\left(Y_{t+1} = \frac{L_{\max} - L_t}{\rho_R} + x_{t-\tau+1}\right)\right)\\
&\quad + \left(L_t - \frac{1}{\rho_E} x_{t-\tau+1}\right)\left(F_{Y_{t+1}}\left(x_{t-\tau+1}\right) - F_{Y_{t+1}}\left(x_{t-\tau+1} - \rho_E L_t\right)\right)\\
&\quad + \frac{1}{\rho_E}\int_{(x_{t-\tau+1}-\rho_E L_t, x_{t-\tau+1}]} y \, \mathrm{d}F_{Y_{t+1}}(y)\\
&\quad + \left(L_t - \rho_R x_{t-\tau+1}\right)\left(F_{Y_{t+1}}\left(\frac{L_{\max} - L_t}{\rho_R} + x_{t-\tau+1}\right)\right.\\
&\quad \left. - \mathbb{P}\left(Y_{t+1} = \frac{L_{\max} - L_t}{\rho_R} + x_{t-\tau+1}\right) - F_{Y_{t+1}}(x_{t-\tau+1})\right)\\
&\quad + \rho_R \int_{(x_{t-\tau+1}, \frac{L_{\max}-L_t}{\rho_R} + x_{t-\tau+1})} y \, \mathrm{d}F_{Y_{t+1}}(y)
\end{aligned}
$$

Thus, $\mathbb{E}[L_{t+1}|S_t, x_t]$ is reduced to the marginal distributions and lower partial moments required in Sect. 3.1. Regarding the specific distributions considered in Sect. 5, the results in Appendix "Derivation of $\int y \mathrm{d}F_{Y_t}(y)$" and Sect. 5.2 can be plugged in directly to efficiently compute $\mathbb{E}[L_{t+1}|S_t, x_t]$. We do not state the derivation of $\mathbb{E}[(L_{t+1})^2|S_t, x_t]$ here as it is very similar to that of $\mathbb{E}[L_{t+1}|S_t, x_t]$.

### Derivation of $\int y \mathrm{d}F_{Y_t}(y)$

In this section, we derive the first partial moment of $Y_t$ ($\int_c^d y \mathrm{d}F_{Y_t}(y) \forall 0 \le c \le d < \infty$) under the assumption that energy production stems from a wind turbine and wind speed follows a Weibull distribution as defined in Sect. 5.2. This satisfies one requirement in Sect. 3.1.

First, assume $0 < c \le d < \Delta t \cdot r$. Then the power output of the wind turbine equals $c$ and $d$ at a wind speed of $\mathrm{sp}(c) = \sqrt[3]{\frac{\frac{c}{\Delta t} - a}{b}}$ and $\mathrm{sp}(d) = \sqrt[3]{\frac{\frac{d}{\Delta t} - a}{b}}$, respectively, and we have

$$
\begin{aligned}
\int_c^d y \, \mathrm{d}F_{Y_t}(y) &= \int_{\mathrm{sp}(c)}^{\mathrm{sp}(d)} \Delta t \cdot (a + bw^3) f_{\mathrm{WB}(\lambda,k)}(w) \, \mathrm{d}w\\
&= \left[a\Delta t \cdot \left(-\mathrm{e}^{(-\lambda \cdot \mathrm{sp}(d))^k} + \mathrm{e}^{(-\lambda \cdot \mathrm{sp}(c))^k}\right)\right] + b\Delta t \cdot \int_{\mathrm{sp}(c)}^{\mathrm{sp}(d)} w^3 f_{\mathrm{WB}(\lambda,k)}(w) \, \mathrm{d}w.
\end{aligned}
$$

We rewrite the integral as $\int_{\mathrm{sp}(c)}^{\mathrm{sp}(d)} w^3 \lambda k (\lambda w)^{k-1} \mathrm{e}^{-(\lambda w)^k} \mathrm{d}w$ and substitute $u = (\lambda w)^k$:

$$
\begin{aligned}
\int_{\mathrm{sp}(c)}^{\mathrm{sp}(d)} w^3 \lambda k (\lambda w)^{k-1} \mathrm{e}^{-(\lambda w)^k} \mathrm{d}w &= \int_{(\lambda \cdot \mathrm{sp}(c))^k}^{(\lambda \cdot \mathrm{sp}(d))^k} \left(\frac{1}{\lambda} \cdot u^{\frac{1}{k}}\right)^3 \mathrm{e}^{-u} du\\
&= \frac{1}{\lambda^3} \cdot \left[\gamma\left(1 + \frac{3}{k}, (\lambda \cdot \mathrm{sp}(d))^k\right) - \gamma\left(1 + \frac{3}{k}, (\lambda \cdot \mathrm{sp}(c))^k\right)\right],
\end{aligned}
$$

where $\gamma$ denotes the lower incomplete gamma function, i.e., $\gamma(s,x) = \int_0^x t^{s-1}\mathrm{e}^{-t}\mathrm{d}t$.

Summing up, for $0 < c \leq d < \Delta t \cdot r$, we have

$$\int_c^d y \, \mathrm{d}F_{Y_t}(y) = \Delta t \cdot \left[a\left(-\mathrm{e}^{(-\lambda \cdot \mathrm{sp}(d))^k} + \mathrm{e}^{(-\lambda \cdot \mathrm{sp}(c))^k}\right)\right] + \frac{\Delta t \cdot b}{\lambda^3} \cdot \left[\gamma\left(1+\frac{3}{k}, (\lambda \cdot \mathrm{sp}(d))^k\right) - \gamma\left(1+\frac{3}{k}, (\lambda \cdot \mathrm{sp}(c))^k\right)\right].$$

Second, if $d \geq \Delta t \cdot r$, then the atom of the distribution of $Y_t$ at $y = \Delta t \cdot r$ has to be taken into account, i.e., $\Delta t \cdot r \cdot \mathbb{P}(Y_t = \Delta t \cdot r) = \Delta t \cdot r \cdot (\mathrm{e}^{-(\lambda s_r)^k} - \mathrm{e}^{-(\lambda s_{\mathrm{co}})^k})$ has to be added to the expression above.

Finally, for $0 \leq c < d < \infty$, we have

$$\begin{aligned}\int_c^d y \, \mathrm{d}F_{Y_t}(y) = {} & \Delta t \cdot \left[a\left(-\mathrm{e}^{(-\lambda \cdot \mathrm{sp}(d))^k} + \mathrm{e}^{(-\lambda \cdot \mathrm{sp}(c))^k}\right)\right] + \frac{\Delta t \cdot b}{\lambda^3} \\ & \cdot \left[\gamma\left(1+\frac{3}{k}, (\lambda \cdot \mathrm{sp}(d))^k\right) - \gamma\left(1+\frac{3}{k}, (\lambda \cdot \mathrm{sp}(c))^k\right)\right] \\ & + \mathbf{1}_{[\Delta t \cdot r, \infty)}(d)\left(\Delta t \cdot r \cdot \left(e^{-(\lambda s_r)^k} - e^{-(\lambda s_{\mathrm{co}})^k}\right)\right)\end{aligned}$$

### Derivation of $\int y^2 \, \mathrm{d}F_{Y_t}(y)$

This derivation of the second partial moment of energy production, $\int_c^d y^2 \, \mathrm{d}F_{Y_t}(y) \forall 0 \leq c \leq d < \infty$, is very similar to Appendix "Derivation of $\int y \mathrm{d}F_{Y_t}(y)$". Therefore, we only state the final result here. This satisfies the remaining requirement in Sect. 3.1.

For $0 \leq c \leq d < \infty$, we have

$$\begin{aligned}\int_c^d y^2 \, \mathrm{d}F_{Y_t}(y) = {} & (\Delta t)^2\left[a^2\left(-\mathrm{e}^{(-\lambda \cdot \mathrm{sp}(d))^k} + \mathrm{e}^{(-\lambda \cdot \mathrm{sp}(c))^k}\right)\right] \\ & + 2ab\left[\frac{1}{\lambda^3} \cdot \left[\gamma\left(1+\frac{3}{k}, (\lambda \cdot \mathrm{sp}(d))^k\right) - \gamma\left(1+\frac{3}{k}, (\lambda \cdot \mathrm{sp}(c))^k\right)\right]\right] \\ & + b^2\left[\frac{1}{\lambda^6} \cdot \left[\gamma\left(1+\frac{6}{k}, (\lambda \cdot \mathrm{sp}(d))^k\right) - \gamma\left(1+\frac{6}{k}, (\lambda \cdot \mathrm{sp}(c))^k\right)\right]\right] \\ & + 1_{[\Delta t \cdot r, \infty)}(d)\left((\Delta t \cdot r)^2 \cdot \left(\mathrm{e}^{-(\lambda s_r)^k} - e^{-(\lambda s_{\mathrm{co}})^k}\right)\right)\end{aligned}$$

### Analytical expression for the contribution

In the following, we give the analytical expression of the contribution used in Sect. 5.3. Under the assumptions concerning the exogenous processes made in Sect. 5, the general contribution given by Eq. (3) in Sect. 3.4 reduces to

$$C(S_t, x_t) = P_t x_t - \beta \mathbb{E}\left[m \cdot P_{t+1}\left[x_{t-\tau+1} - (\rho_E L_t + Y_{t+1})\right]^+ | S_t, x_t\right]$$

for $t \leq T-1$ where

$$\begin{aligned}
&\mathbb{E}[m \cdot P_{t+1}[x_{t-\tau+1} - (\rho_E L_t + Y_{t+1})]^+ | S_t, x_t] \\
&= \mathbb{E}\left[\mathbf{1}_{[0, x_{t-\tau+1}-\rho_E L_t]}(Y_{t+1})(m \cdot P_{t+1} \cdot (x_{t-\tau+1} - \rho_E L_t) - m \cdot P_{t+1} \cdot Y_{t+1}) | S_t, x_t\right] \\
&= F_{Y_{t+1}}(x_{t-\tau+1} - \rho_E L_t)(m \cdot \mathbb{E}[P_{t+1} | S_t, x_t] \cdot (x_{t-\tau+1} - \rho_E L_t)) - m \\
&\quad \cdot \mathbb{E}[P_{t+1} | S_t, x_t] \int_{[0, x_{t-\tau+1}-\rho_E L_t]} y \, \mathrm{d}F_{Y_{t+1}}(y)
\end{aligned}$$

Expressions for $F_{Y_{t+1}}(x_{t-\tau+1} - \rho_E L_t)$ and $\int_{[0, x_{t-\tau+1}-\rho_E L_t]} y dF_{Y_{t+1}}(y)$, are given in Sect. 5.2, and Appendix "Derivation of $\int y \mathrm{d}F_{Y_t}(y)$", respectively.

## References


Barto AG, Sutton RS, Anderson CW (1983) Neuronlike adaptive elements that can solve difficult learning control problems. IEEE Trans Syst Man Cybern 5:834–846

Bertsekas DP (2012) Approximate dynamic programming, 4th edn. Athena Scientific, Belmont

Brockwell PJ, Davis RA (2013) Time series: theory and methods. Springer, New York

Costa LM, Juban J, Kariniotakis G (2008) Management of energy storage coordinated with wind power under electricity market conditions. In: Proceedings of the 10th International Conference on Probabilistic Methods Applied to Power Systems. Rincón, Puerto Rico

Hannah LA, Dunson DB (2011) Approximate dynamic programming for storage problems. In: Getoor, L. (ed.) Proceedings of the 28th International Conference on Machine Learning, Bellevue, WA

Hassler M (2016) Managing complexity: effective decision rules for trading renewable energy. Working Paper, University of Augsburg, Augsburg

Jiang DR, Powell WB (2015) Optimal hour-ahead bidding in the real-time electricity market with battery storage using approximate dynamic programming. Inf J Comput 27(3):525–543

Jiang DR, Pham TV, Powell WB, Salas DF, Scott WR (2014) A comparison of approximate dynamic programming techniques on benchmark energy storage problems: does anything work? In: 2014 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL), Orlando, FL

Justus CG, Hargraves WR, Yalcin A (1976) Nationwide assessment of potential output from wind-powered generators. J Appl Meteorol 15(7):673–678

Kim JH, Powell WB (2011) Optimal energy commitments with storage and intermittent supply. Oper Res 59(6):1347–1360

Konda VR, Tsitsiklis JN (1999) Actor-critic algorithms. In: Solla, S.A., Leen, T.K., Müller, K. (eds.) Advances in Neural Information Processing, vol. 12, pp. 1008–1014

Lagoudakis MG, Parr R (2003) Least-squares policy iteration. J Mach Learn Res 4:1107–1149

Lai G, Margot F, Secomandi N (2010) An approximate dynamic programming approach to benchmark practice-based heuristics for natural gas storage valuation. Oper Res 58(3):564–582

Löhndorf N, Minner S (2010) Optimal day-ahead trading and storage of renewable energies—an approximate dynamic programming approach. Energy Syst 1(1):61–77

Löhndorf N, Wozabal D, Minner S (2013) Optimizing trading decisions for hydro storage systems using approximate dual dynamic programming. Oper Res 61(4):810–823

Mokrian P, Stephen M (2006) A stochastic programming framework for the valuation of electricity storage. In: 26th USAEE/IAEE North American Conference. Ann Arbor, MI

Möst D, Keles D (2010) A survey of stochastic modelling approaches for liberalised electricity markets. Eur J Oper Res 207(2):543–556

Nascimento JM, Powell WB (2009) An optimal approximate dynamic programming algorithm for the lagged asset acquisition problem. Math Oper Res 34(1):210–237

Nascimento J, Powell WB (2013) An optimal approximate dynamic programming algorithm for concave, scalar storage problems with vector-valued controls. IEEE Trans Autom Control 58(12):2995–3010

Powell WB (2011) Approximate Dynamic Programming. Solving the Curses of Dimensionality, 2nd edn. Wiley-Blackwell, Hoboken

Puterman ML (2005) Markov decision processes. Discrete stochastic dynamic programming. Wiley, Hoboken

Robert Bosch GmbH (2014) Megawatt project near the North Sea. http://www.bosch-presse.de/presseforum/details.htm?txtID=6876&locale=en. Accessed 24 Jun 2015

Salas DF, Powell WB (2013) Benchmarking a scalable approximate dynamic programming algorithm for stochastic control of multidimensional energy storage problems. Working Paper, Princeton University, NJ

Scott WR, Powell WB (2012) Approximate dynamic programming for energy storage with new results on instrumental variables and projected Bellman errors. Working Paper, Princeton University, Princeton, NJ

Seguro JV, Lambert TW (2000) Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. J Wind Eng Ind Aerodyn 85(1):75–84

Shumway RH, Stoffer DS (2013) Time series analysis and its applications. Springer, New York

Sioshansi R, Madaeni SH, Denholm P (2014) A dynamic programming approach to estimate the capacity value of energy storage. IEEE Trans Power Syst 29(1):395–403

Strbac G (2008) Demand side management: benefits and challenges. Energy Policy 36(12):4419–4426

Sutton RS (1984) Temporal credit assignment in reinforcement learning. Ph.D. thesis, University of Massachusetts, Amherst, MA

Wiering M, van Otterlo M (2012) Reinforcement Learning. State-of-the-Art. Springer, Berlin

Zhou Y, Scheller-Wolf AA, Secomandi N, Smith S (2014) Managing wind-based electricity generation in the presence of storage and transmission capacity. Working Paper, Carnegie Mellon University, Pittsburgh, PA

Zhou Y, Scheller-Wolf AA, Secomandi N, Smith S (2016) Electricity trading and negative prices: storage vs. disposal. Manag. Sci. 62(3):880–898