Optimal Energy Management of a Residential Prosumer: A Robust Data-Driven Dynamic Programming Approach

Zhongjie Guo ©, Wei Wei ©, Senior Member, IEEE, Laijun Chen ©, Member, IEEE, Zhaojian Wang ©, Member, IEEE, João P. S. Catalão ©, Senior Member, IEEE, and Shengwei Mei ©, Fellow, IEEE

Abstract—Prosumers are agents that both consume and produce energy. This article studies the optimal energy management of a residential prosumer which consists of a renewable power plant and an energy storage unit. Energy could stream among power grid, renewable plant, storage unit, and demand, providing a highly flexible energy supply and the opportunity of arbitrage. To capture the uncertainty of renewable generation and electricity price, as well as the rolling horizon feature of the multiperiod energy management, the problem is formulated as a robust data-driven dynamic programming (RDDP). Kernel regression is utilized to build the empirical conditional distribution in a data-driven manner, and all candidates that reside in a Wasserstein metric-based ambiguity set are taken into account to tackle the inexactness of the empirical distribution. The RDDP can be transformed into a series of convex optimization problems with cost-to-go functions in their constraints. The piecewise linear expression of the cost-to-go function is retrieved from dual linear programs. Through such an analytical expression of cost-to-go functions, the RDDP can be solved via backward induction, unlike the popular stochastic dual dynamic programming technique that incorporates forward and backward passes. Case studies validate the performance and advantage of the proposed RDDP approach.

Index Terms—Energy storage, prosumer, robust data-driven dynamic programming, uncertainty, value function approximation.

NOMENCLATURE

A. Abbreviations

ADP Approximate dynamic programming.
DG Distributed generator.

B. Parameters and Indices

a Random vector of exogenous states.
N Number of historical observations.
H Number of sampled state variables.
Mₜ Ambiguity set of the conditional distribution in period t with radius ε.

C. Variables

xₑ Max charging power of ESU.
max Max discharging power of ESU.
sₘᵦ Min SoC of ESU.
max Max SoC of ESU.
ε Charging/discharging efficiency of ESU.
Δ Duration of a time slot.
ₜ Available renewable power, random parameter.
ₜ Demand in time slot t, random parameter.
ₚₜ Selling price of electricity, random parameter.
ₚₚ Buying price of electricity, random parameter.
ᵣ Random vector of exogenous states.
Mᵣ Ambiguity set of the conditional distribution in period t with radius ε.

I. INTRODUCTION

With the development of technology, renewable distributed generators (DGs) have been widely deployed at the demand side, bringing various benefits not only to the distribution system but also to end consumers [1]. Private-owned DGs, like rooftop solar panels, small wind turbines, and micro gas-fired units, endow traditional end consumers with the ability...
to produce, precipitating the advent of prosumers [2]. Energy storage unit (ESU) owned by the prosumer greatly improves the flexibility and reliability of maintaining power balance. On the one hand, the ESU helps to compensate the volatile output of renewable DGs; on the other hand, DG and ESU offer the prosumer a unique opportunity to actively participate in system operation through trading energy with the power grid [3]. Recently, prosumer has become a hot topic and attracted a lot of attention from researchers [4].

Because of low operation costs, renewable DGs have won the favor of prosumers. However, given the volatility of renewable generation, the energy management of prosumer becomes more challenging. Two-stage stochastic optimization and two-stage robust optimization are the mainstream methods for decision making under uncertainty. SO requires the probability distribution of uncertain factors and usually relies on scenario sampling [5] or chance-constraints [6]; robust optimization depicts uncertain parameters via an uncertainty set [7] and optimize the target under the worst-case scenario. Both of them have been widely applied in unit commitment [8], [9], economic dispatch [10], [11], optimal active-reactive power flow [12], [13], energy storage operation [14] and so on. A tutorial on two-stage robust optimization can be found in Appendix C of [15].

In the above two-stage optimization approaches, a decision is made with perfect knowledge of the realized uncertainty over the entire time horizon. However, in the energy management problem, renewable generation is observed period-by-period, and the dispatch strategy in period t cannot depend on the information after period t, which is called nonanticipativity; if nonanticipativity is neglected, the solution could be infeasible, as is illustrated in [16] by an example to demonstrate such phenomenon in the two-stage robust unit commitment problem. To overcome this difficulty, an affine policy-based multiperiod robust optimization method is proposed in [16] to enforcing nonanticipativity, in which the generator real-time adjustments are linear functions in the forecast errors. The method is applied to unit commitment with energy storage in [17]. Applying the affine policy restricts system flexibility and thus may give suboptimal strategies. More detailed discussions on the optimality performance of affine policy can be found in [18].

Another remedy is to perform multiperiod optimization or dynamic programming in a rolling horizon fashion. The main difficulty is the nested optimization structure that requires evaluating the optimal value for the remaining periods (also called the cost-to-go function) recursively. Two prevalent methods are approximate dynamic programming (ADP) and stochastic dual dynamic programming (SDDP) [19], depending on how the cost-to-go function is approximated.

ADP approximates the cost-to-go function with a smaller number of parameters; general strategies include aggregation and continuous value function approximation [20]. The authors in [21] proposed a heuristic interpolation-based method. However, interpolation-based regression does not guarantee the performance of approximation (optimistic or pessimistic). The method in [21] is applied to battery storage management in [22]. An inspiring method in [23] focuses on finding the slopes of a piece of function rather than the value, but this approach does not account for uncertainty. Machine learning methods are also found in value function approximation, such as deep recurrent neural network learning [24] and reinforcement learning [25]. ADP has been used in [26] for the strategic operation of energy hubs with uncertainties, in [27] for microgrid economic dispatch, in [28] for load management in smart grids, and so on.

SDDP approximates the cost-to-go function via multidimensional benders cuts. Lower bounds and upper bounds are generated in the forward and backward swapping process until a convergence criterion is met [29]. The applications of SDDP are also reported in [30] and [31] to economic dispatch, in [32] to transmission expansion planning and in [33] to ESU management strategy for microgrids. As variants of SDDP, stochastic dual dynamic integer programming [34], and distributionally robust SDDP [35] are also explored. In summary, SDDP is an elegant method that can solve dynamic programs under uncertainties with a provable guarantee on the optimality gap. However, as the forward pass and backward pass should repeat many times, the computation time may not meet the requirement of online use. Although some dedicated strategies have been proposed to accelerate SDDP, such as scenario reduction [36] and cut selection [37], it is still not clear how fast SDDP would converge in advance.

Except for ADP and SDDP, robust dual dynamic programming is proposed in [38]. It describes uncertain parameters using an uncertainty set and is a multistage robust optimization. The solution also relies on forward and backward swapping. A multiresolution dynamic programming is introduced in [39] for managing ESU. The optimization horizon is divided into a series of sub-horizons and discretized with different temporal resolutions, enabling a reduced computational complexity compared to the single-resolution approach. The authors in [40] discusses a data-driven multistage stochastic optimization approach for seasonal energy storage operation; the cost-to-go function is evaluated based on available historical data of uncertain parameters instead of modeling their underlying distributions.

In view of the limitation of two-stage optimization in the dynamic environment faced by prosumers, we resort to the dynamic programming modeling paradigm in [21]. Compared with existing work, the novelty of this article is twofold.

1) A robust data-driven dynamic programming (RDDP) formulation for energy management of a residential prosumer. To model the uncertainty in the dynamic environment, kernel regression is used to estimate the empirical conditional distribution using historical data following the paradigm in [21]: a Wasserstein metric-based ambiguity set is built accounting for the inexactness of conditional distribution. Unlike SDDP, the proposed method relies on data and is robust against the inexactness of conditional distribution. Compared with the robust dual dynamic programming in [38], the proposed RDDP is less conservative as the distributional property of uncertainty is taken into account.

2) A systematic approach to solve the RDDP, which entails solving a series of convex optimization problems with cost-to-go functions in its constraints. According to the multiparametric programming theory, we construct convex PWL expressions for the cost-to-go functions from dual variables extracted from
relying on historical data; the computational time of RDDP is mate the cost-to-functions at a very low computational expense not limited to residential users, microgrids, energy hubs, and industrial parks. Technically, the proposed method can approximate the cost-to-functions at a very low computational expense relying on historical data; the computational time of RDDP is \( K \times O(N) \), which is much shorter than that of SDDP, \( O(N^2) \). More importantly, the RDDP formulation enjoys a degree of robustness: even in case of the scarcity of historical data, the strategies can still guarantee an acceptable operating economy; results show that with only ten observations of historical data, the suboptimality of RDDP is less than 4\%.

This article proposes a paradigm for the optimal energy management of the prosumer-like energy systems, including but not limited to residential users, microgrids, energy hubs, and industrial parks. Technically, the proposed method can approximate the cost-to-functions at a very low computational expense relying on historical data; the computational time of RDDP is \( K \times O(N) \), which is much shorter than that of SDDP, \( O(N^2) \). More importantly, the RDDP formulation enjoys a degree of robustness: even in case of the scarcity of historical data, the strategies can still guarantee an acceptable operating economy; results show that with only ten observations of historical data, the suboptimality of RDDP is less than 4\%.

This article extends the work in [21] and [22] in two aspects. First, we consider the energy streams from the residential system to the main grid, improving the operational flexibility and capturing the arbitrage opportunity of the prosumer. Second, we develop a thorough scheme for approximating the cost-to-go function according to the multiparametric programming theory; compared with the interpolation-based approximation method in [21] and [22], which overly estimates the function value, such a method leads to an under-estimator and exhibits better accuracy without increasing complexity, especially when the sampling set is small.

The rest of this article is organized as follows: the architecture of the prosumer is established in Section II, followed by the RDDP formulation of the optimal energy management problem. The solution method for the RDDP is presented in Section III. The results of numerical tests are reported in Section IV. Finally, conclusions are drawn in Section V.

II. Mathematical Formulation

In this section, the architecture of the residential prosumer is introduced first; then the energy management problem in the deterministic case is presented; finally, the RDDP is set forth by incorporating uncertainty.

A. Architecture of the Residential Prosumer

The architecture of residential prosumer studied in this article is depicted in Fig. 1. We assume that the interfaces between wind turbine and storage, wind turbine and the load devices, wind turbine and grid, storage and grid (bidirectional), storage and load devices, as well as grid and load devices, are available. Demand is supplied by the DG, ESU, and the grid; ESU can store energy from renewable generation and grid; specifically, the energy in storage or generated from wind turbine can be sold back to the grid, which is practicable in a decentralized power grid. The ESU plays an important role in this residential-level energy system due to its ability to shift demand over time. The model of ESU is given as

\[
0 \leq x_{t}^{gs} + x_{t}^{ws} \leq x_{t}^{c}_{max}, \forall t \tag{1a}
\]

\[
0 \leq x_{t}^{sd} + x_{t}^{xd} \leq x_{t}^{dc}_{max}, \forall t \tag{1b}
\]

\[
s_{min} \leq s_{t} \leq s_{max}, \forall t \tag{1c}
\]

\[
s_{t+1} = s_{t} + (x_{t}^{gs} + x_{t}^{ws}) \Delta t - (x_{t}^{sd} + x_{t}^{xd}) \frac{\Delta s_{t}}{\eta_{dc}}, \forall t \tag{1d}
\]

where constraints (1a) and (1b) impose non-negativity and upper limits on charging power and discharging power of the ESU; constraint (1c) stipulates the feasible range of SoC; (1d) describes the SoC dynamics over time. Here we do not exert strict complementarity on charging and discharging power, which can be a barrier to unleash the flexibility of prosumer. The condition that the terminal SoC equals to the initial state is relaxed [41].

The total power supply from the renewable plant cannot exceed the available renewable generation \( w_{t} \), yielding

\[
x_{t}^{wd} + x_{t}^{ws} + x_{t}^{wg} \leq w_{t}, \forall t. \tag{2}
\]

Demand balance boils down to

\[
x_{t}^{sd} + x_{t}^{xd} + x_{t}^{gd} = d_{t}, \forall t. \tag{3}
\]

Such an architecture enjoys high flexibility of reconfiguration. For example, if a real prosumer system does not have the interface between wind turbine and grid, the variable \( x_{t}^{ws} \) is set to 0. Therefore, the proposed architecture and its model are very general.

B. Prosumer Energy Management in Stochastic Form

Given the rolling horizon feature and uncertain nature, the energy management problem can be cast as a stochastic DP. In period \( t \), storage SoC \( s_{t} \in S \) is the endogenous state, where \( S \) is defined in (1c). Vector \( \xi_{t} = \{w_{t}, \pi_{t}^{s}, \pi_{t}^{p}, d_{t}\} \in \mathbb{R}^{4} \) which encapsulates all external random parameters is the exogenous state, where \( \pi_{t}^{s}/\pi_{t}^{p} \) denotes selling and purchasing prices of electricity. \( \xi_{t} \) is observed at the beginning of period \( t \). Based on \( (s_{t}, \xi_{t}) \), the prosumer determines control action \( u_{t} = \{x_{t}^{gs}, x_{t}^{ws}, x_{t}^{sd}, x_{t}^{xd}, x_{t}^{wg}, x_{t}^{wd}, x_{t}^{gd}\} \in U_{t}(\xi_{t}) \) which includes energy flow variables in period \( t \), where feasible set \( U_{t} \) depending on \( w_{t} \) and \( d_{t} \) corresponds to constraints (1a), (1b), (2), and (3). After prosumer deploys action \( u_{t} \), the system moves to a new state \( s_{t+1} \) at the beginning of period \( t + 1 \) according to
the state transition equation \( s_{t+1} = q_t(s_t, u_t) \) described in (1d). Then, \( \xi_{t+1} \) is observed. The action \( u_t \) incurs an instantaneous payoff in period \( t \) which represents the cost for purchasing electricity from the power grid and the income for selling electricity to the power grid, i.e., \( c_t(\xi_t, u_t) = \pi^p_t(x^{gn}_t + x^{gd}_t) - \pi^t_t(x^{gn}_t + x^{gd}_t) \).

Based on the above notations and Bellman’s principle of optimality [19], the stochastic dynamic programming model of prosumer energy management is cast in a backward recursive manner

\[
V_t(s_t, \xi_t) = \min_{u_t} \left( c_t(\xi_t, u_t) + \mathbb{E}[V_{t+1}(s_{t+1}, \xi_{t+1}) | \xi_t] \right)
\]

s.t. \( u_t \in \mathcal{U}_t(\xi_t), s_{t+1} \in \mathcal{S} \)

\[
\text{for } t = T, \ldots, 1 \text{ with } V_{T+1} \equiv 0, \text{ where } V_t(s_t, \xi_t) \text{ is the optimal payoff for the remaining } T - t \text{ periods and is called the cost-to-go function or value function. It quantifies the optimal value based on } (s_t, \xi_t) \text{ revealed at the beginning of period } t. \text{ However, as the payoff depends on the future realization of uncertain parameters, we resort to minimize the conditional expectation in problem (4) for each period.}

C. RDDP Model

The stochastic DP model (4) encounters two difficulties in practical use. From the modeling perspective, the exact conditional probability distribution of \( \xi_{t+1} \) given \( \xi_t \) is not available; from the computational perspective, to perform backward induction, we need the analytical expression of cost-to-go function. In this section, we use the kernel regression method to construct the empirical conditional distribution based on historical data and formulate an RDDP model, inspired by the paradigm in [21]. Specifically, we consider all possible distributions in a Wasserstein-based ambiguity set which needs fewer data than the \( \chi^2 \)-distance based one in [21]. The computational issue is left for the next section.

Given a set of \( N \) historical trajectories of exogenous states \( \{\xi_t^{(i)}\}_{t=1}^{N}, \forall t \), the conditional expectation in (4) can be estimated via Nadaraya-Watson kernel regression method [21], [25]

\[
\sum_{i=1}^{N} q_t(\xi_t) V_{t+1}(s_{t+1}, \xi_{t+1}^{(i)}) \tag{5}
\]

where the weight coefficients \( q_t(\xi_t) \) are obtained from

\[
q_t(\xi_t) = \frac{K(\xi_t - \xi_t^{(i)})}{\sum_{k=1}^{N} K(\xi_t - \xi_t^{(k)})} \tag{6}
\]

where

\[
K(y) = \exp \left( -\frac{(\|y\|_2)^2}{2\sigma^2} \right) \tag{7}
\]

is the Gaussian kernel function; \( \| \cdot \|_2 \) stands for the Euclidean norm of a vector; \( \sigma > 0 \) is a bandwidth parameter. The kernel function guarantees that the observation nearer to \( \xi_t \) holds a larger weight. Substituting (5) into (4) we have

\[
\begin{align*}
[b]V_t(s_t, \xi_t) &= \min_{u_t} c_t(\xi_t, u_t) \\
&+ \sum_{i=1}^{N} q_t(\xi_t) V_{t+1}(s_{t+1}, \xi_{t+1}^{(i)}) \\
\text{s.t. } u_t &\in \mathcal{U}_t(\xi_t), s_{t+1} \in \mathcal{S} \\
&\text{for } t = T, \ldots, 1 \text{ with } V_{T+1} \equiv 0, \text{ where } V_t(s_t, \xi_t) \text{ is the optimal payoff for the remaining } T - t \text{ periods and is called the cost-to-go function or value function. It quantifies the optimal value based on } (s_t, \xi_t) \text{ revealed at the beginning of period } t. \text{ However, as the payoff depends on the future realization of uncertain parameters, we resort to minimize the conditional expectation in problem (4) for each period.}
\end{align*}
\]

Problem (8) is reminiscent of SDDP and can be solved via forward pass and backward pass algorithm in [29]. In practice, however, the historical data may be sparse, and the conditional probabilities \( q_t(\xi_t) \) could be inexact with the variance scalings with \( \mathcal{O}(\frac{1}{N}) \). [21].

To cope with this issue, when calculating the expectation in (5), not only the empirical distribution in (6), but also other distributions that are close to the empirical one, should be taken into account. To this end, we employ the Wasserstein metric to quantify the distance between two probability distributions. The Wasserstein metric for two discrete probability distributions \( P_1 \) and \( P_2 \) in probability space \( \mathcal{P}(\Xi) \) supported on \( \Xi \) is defined through an optimal transport problem

\[
d_W(P_1, P_2) = \inf_{\pi \in \Pi} \sum_{i,j} \pi_{ij} \|\xi_t^i - \xi_t^j\|_p
\]

s.t. \( \sum_j \pi_{ij} = P_1^i, \forall i \)

\[
\sum_i \pi_{ij} = P_2^j, \forall j
\]

\[
\sum_{i,j} \pi_{ij} = 1 \tag{9}
\]

where decision variable \( \pi \in \mathbb{R}^{N \times N} \) is a joint distribution of random variables \( \xi_t^i \) and \( \xi_t^j \) with marginal distributions \( P_1^i \) and \( P_2^j \), respectively; \( \| \cdot \|_p \) is the vector \( p \)-norm. We adopt \( p = 1 \) in this article, and \( d_W \) is called 1-Wasserstein metric or Kantorovich metric [42]. If we regard \( \pi_{ij} \) as the probability mass transported from slot \( i \) to slot \( j \), and \( \|\xi_t^i - \xi_t^j\|_p \) as the corresponding energy consumption, then problem (9) aims to find the energy optimal plan for reshaping distribution \( P_1 \) to distribution \( P_2 \), and the minimal energy consumption is defined as the distance between \( P_1 \) and \( P_2 \).

Equipped with the Wasserstein metric, we can build the ambiguity set which includes all probability distributions near the empirical distribution \( Q = \{q_t(\xi_t), \ldots, q_t(\xi_T)\} \), \( \forall t \)

\[
\mathcal{M}_\varepsilon^t = \{P \in \mathcal{P}(\{\xi_t^{(i)}\}_{t=1}^{N}) : d_W(P, Q) \leq \varepsilon \} \tag{10}
\]

where distribution \( P \) is described by the vector of probabilities \( \{p_t(\xi_t), \ldots, p_T(\xi_T)\} \) associated with historical trajectories \( \{\xi_t^{(i)}\}_{t=1}^{N} \). According to (10), the distance between empirical distribution \( Q \) and any \( P \in \mathcal{M}_\varepsilon^t \) is no greater than \( \varepsilon \) in the sense of 1-Wasserstein metric. Parameter \( \varepsilon \) is called the radius of the ambiguity set, and depends on the prosumer’s attitude on risk. With the increase of \( \varepsilon \), more distributions are contained in \( \mathcal{M}_\varepsilon^t \), and the model is more conservative, reflecting a risk-averse attitude. If the attitude towards risk is not clear, a proper value
of $\varepsilon$ is recommended by [43]
\[
\varepsilon = -\log(\alpha^*)/N
\]
where $N$ is the total number of historical observations. With such a selection, the probability for the true distribution belonging to $\mathcal{M}_T^\varepsilon$ is no less than $1 - \alpha^\varepsilon$.

With the ambiguity set $\mathcal{M}_T^\varepsilon$, the robust version of stochastic DP (8), which is called RDDP, can be written as
\[
[b]V_t(s_t, \xi_t) = \min c_t(\xi_t, u_t) + \max_{p_t(\xi_{t+1})} \sum_{i=1}^{N} p_{t,i}(\xi_{t}) V_{t+1}(s_{t+1}, \xi_{t+1}^i) \\
\text{s.t. } u_t \in \bigcup_{k} U_k(\xi_t), s_{t+1} \in S \\
s_{t+1} = g_t(s_t, u_t).
\]

In RDDP (12), the expectation is evaluated at the worst-case distribution in the ambiguity set $\mathcal{M}_T^\varepsilon$, therefore, the statistical performance of the optimal strategy is robust against the inexactness of conditional probability distribution. Compared to (8), an inner maximization over conditional distribution is performed accounting for uncertainty. Compared to a similar model in [21], the Wasserstein metric-based ambiguity set adapts to various cases with different data availability.

### III. Solution Method

In this section, we develop a systematic methodology to solve RDDP (12) via LP solver and backward induction. The key steps entail the elimination of inner maximization and the approximation of the cost-to-go function.

#### A. Eliminating the Inner Maximization

The maximization over $\mathcal{M}_T^\varepsilon$ is an infinite-dimensional optimization problem and must be reformulated. To this end, we summarize the observations from the concrete model provided as follows.

1) The cost function $c_t$ is linear in $u_t$.
2) The state transition equation $g_t$ is linear in $s_t$ and $u_t$.
3) The feasible set $U_k(\xi_t)$ is a polyhedron.

According to Proposition 2 in [44], problem (12) can be reduced to a finite-dimensional optimization problem
\[
V_t(s_t, \xi_t) = \min c_t(\xi_t, u_t) + \lambda \varepsilon - \mu + \sum_{i=1}^{N} q_{t,i}(\xi_{t}) y_{i,t} \\
\text{s.t. } u_t \in \bigcup_{k} U_k(\xi_t), s_{t+1} \in S \\
\mu \in \mathbb{R}, \lambda \in \mathbb{R}_+, z, y \in \mathbb{R}^N \\
s_{t+1} = g_t(s_t, u_t) \\
V_{t+1}(s_{t+1}, \xi_{t+1}^i) + \mu \leq z_i, \forall i = 1, ..., N \\
z_i - \lambda \|\xi_{t}^i - \xi_{t+1}^i\| \leq y_{i,j}, \forall i, j = 1, ..., N
\]

where $\mu$, $\lambda$, $z$, and $y$ are auxiliary variables; $q_{t,i}(\xi_{t}), \forall i, t$ and $\|\xi_{t}^i - \xi_{t+1}^i\| \forall i, j, t$ are constants calculated from historical observations. In the next section, we prove the cost-to-go function $V_{t+1}(s_{t+1}, \xi_{t+1}^i)$ can be expressed via a convex PWL function, so that problem (13) gives rise to an LP and can be solved very efficiently. Solve (13) via backward induction (from the last period to the first period), we obtain the solution of RDDP (12). The strategy in the first period is deployed. The remaining task is to construct the PWL expression of the cost-to-go function in period $t$ with respect to each observation.

#### B. PWL Expression of the Cost-to-go Function

We need an analytical expression of $V_{t+1}(s_{t+1}, \xi_{t+1}^i)$ to solve the problem (13). We know $V_{T+1} = 0$, so problem (13) in period $T$ is an LP. Suppose $V_{t+1}(s_{t+1}, \xi_{t+1}^i)$ is a convex PWL function in $s_{t+1}$, we prove $V_t(s_t, \xi_t)$ remains a convex PWL function in $s_t$. For notation brevity, let $x_t$ denote the parameter and $x$ represents all decision variables in (13). Under the assumption on $V_{t+1}(s_{t+1}, \xi_{t+1}^i)$, all constraints in (13) can be written as linear equalities and inequalities, resulting in an explicit LP. Proceeding one step back, we solve
\[
[b]V_t(s_t, \xi_t) = \min c^\top x \\
\text{s.t. } Ax \leq b + Bs_t
\]
where $A, B, b,$ and $c$ are constant coefficients corresponding to the concrete model. We aim to demonstrate that the optimal value function $V_t(s_t, \xi_t)$ can be analytically expressed as a convex PWL function in $s_t$.

Recalled the basic property of LP, the optimal solution can always be found at one of the vertices of the polyhedral feasible region. At the optimal solution, the constraints in (14) are categorized into active ones and inactive ones
\[
A'x^* = b' + B's_t \\
A''x^* < b' + B''s_t.
\]

Hence, the optimal solution and optimal value are
\[
x^* = A'^{-1}b' + A'^{-1}B's_t, s_t \in \Theta
\]
\[
[b]V_t^*(s_t, \xi_t) = c^\top A'^{-1}b' + c^\top A'^{-1}B's_t
\]
\[
= m_t + n_t s_t, s_t \in \Theta
\]
where $\Theta$ is called a critical region in which the set of active constraints remains unchanged. From (16) we can see that the optimal solution and the optimal value are linear functions of $s_t$. In fact, the feasible interval $S$ is covered by several disjoint subintervals, i.e., $S = \bigcup_{i=1}^{n_{\Theta}} \Theta_i$, where $\Theta_i \cap \Theta_j = \emptyset, \forall i \neq j$.

Therefore, the cost-to-go function must be PWL in $s_t$
\[
V_t(s_t, \xi_t) = \begin{cases}
    m_1 + n_1 s_t, \forall s_t \in \Theta_1 \\
    \vdots \\
    m_I + n_I s_t, \forall s_t \in \Theta_I
\end{cases}
\]
where the coefficients $m_i$ and $n_i$ depend on $\xi_t$. Computing the critical regions can be tricky in case of degeneracy. Here we use (17) to demonstrate the PWL structure of $V_t(s_t, \xi_t)$; and critical region is not actually used in the proposed method, once we prove the convexity of $V_t(s_t, \xi_t)$. 

---

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.
Consider the dual problem of LP (14)
\[
V_t(s_t, \xi_t) = \max_{\gamma} \gamma^T (b + B s_t) \\
\text{s.t. } A T \gamma = c, \gamma \leq 0
\]
where \( \gamma \) is the vector of dual variables. By strong duality, the optimal value of LP (18) coincides with \( V_t(s_t, \xi_t) \), the optimal value of its primal problem. For any given \( \gamma \), the objective function of (18) is a linear function in \( s_t \). Therefore, \( V_t(s_t, \xi_t) \) is the point-wise maximum of infinitely many linear functions, and thus convex in \( s_t \).

The dual problem (18) provides a more convenient way to build PWL expression (17) without the information on critical region. Let \( \Gamma = \{ \gamma | A T \gamma = c, \gamma \leq 0 \} \) be the polyhedral feasible set of dual variable and \( \text{vert}(\Gamma) \) the set of vertices. Because the optimum is finite, the optimal solution must be found at some \( \gamma \in \text{vert}(\Gamma) \), although \( \Gamma \) may contain extreme rays. In this regard
\[
V_t(s_t, \xi_t) = \max_i \{ \gamma_i^T b + \gamma_i^T B s_t \}, \forall \gamma_i \in \text{vert}(\Gamma).
\]
Compare (17) and (19), coefficients \( m \) and \( n \) can be retrieved from dual variable via
\[
m_i = \gamma_i^T b, n_i = \gamma_i^T B, \text{ for some } \gamma_i \in \text{vert}(\Gamma).
\]
However, we do not know in prior which vertex produces the linear functions in (19). We propose the following heuristic method: choose uniformly distributed points \( s^k \in S, k = 1, \ldots, K \); solve problem (14) for each \( s^k \), the corresponding dual variable is \( \gamma^k \), \( k = 1, \ldots, K \). After deleting duplicated elements in \( \{ \gamma_k \}^K_{k=1} \), we can retrieve the PWL expression according to (19).

Now we can claim: \( V_t(s_t, \xi_t) \) is a convex PWL function in \( s_t, \forall t \). The flowchart of constructing PWL expression for the cost-to-go functions is summarized in Algorithm 1. For clarity, an overall flowchart of the proposed RDDP approach is given in Fig. 2.

Additional discussions are given as follows.

1) Algorithm 1 performs backward induction from period \( T \) to period 2, unlike SDDP algorithm which performs forward pass and backward pass repeatedly. Hence, there is no convergence issue.

### IV. CASE STUDIES

A residential prosumer with a storage and a wind turbine is used to validate the proposed method. Some parameters are given in Table I; complete system data are available in [45]. LPs are solved by CPLEX 12.8 on a laptop with Intel i5-8250U CPU and 8 GB memory.
A. Performance of the PWL Expression

To conduct Algorithm 1, we collect ten couples of historical observations of wind power and demand from a real residential energy system [46]; so $N = 10$. The electricity price is collected from the PJM database from July 1st to July 10th, 2019 [46]. One observation is plotted in Fig. 3, whose corresponding cost-to-go functions over the entire time horizon are drawn in Fig. 3, from which we can see the following.

1) In period $t$, the cost-to-go function $V_t(\cdot)$ is monotonically decreasing in its argument $s_t$. The reason is clear: A higher SoC can supply more energy in the future for demand or for sale, and thus reduce the purchasing cost while circumventing potential risks brought by market price uncertainty.

2) With time rolling on, cost-to-go functions in each period shows a decreasing trend in general, because less energy is needed in the rest of the day.

3) Distinguished from the overall trend, the first exception marked by $\circ$ occurs at the end of a day. Referring to Fig. 5, the cost-to-functions in the last few periods have values higher than those in period 21 when the level of SoC is relatively high. Because we assume $V_{T+1} \equiv 0$, the prosumer will sell out the energy before a new day. Given a maximum discharging power, the energy is sold across multiple periods. Consequently, the prosumer enjoys a higher revenue (represented by negative cost) in periods 22 and 23 than in the last period.

B. Impact of Data Availability and Parameter $\varepsilon$

By Algorithm 1 with ten groups of historical observations mentioned above, all the cost-to-go functions are already calculated, with which we can solve the reformulated dynamic program (13) as time recedes from period 1–24 with new observations.

To this end, we employ another five sequences of uncertain renewable power, demand and electricity price, denoted by $S_1 \sim S_5$; they act as the observations, which are given period by period. The initial level of SoC is set to 10 kWh. The prosumer
TABLE II
COMPARISONS ON INDEX 1 AND 2 WITH DIFFERENT ε

<table>
<thead>
<tr>
<th>Test</th>
<th>ε</th>
<th>Index 1 ($)</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>Index 1 ($)</td>
<td>0.328</td>
<td>0.315</td>
<td>0.331</td>
<td>0.321</td>
<td>0.326</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>Index 2 ($)</td>
<td>0.264</td>
<td>0.345</td>
<td>0.384</td>
<td>0.375</td>
<td>0.269</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Index 1 ($)</td>
<td>0.585</td>
<td>0.572</td>
<td>0.587</td>
<td>0.578</td>
<td>0.582</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Index 2 ($)</td>
<td>0.294</td>
<td>0.330</td>
<td>0.368</td>
<td>0.381</td>
<td>0.270</td>
</tr>
</tbody>
</table>

TABLE III
COMPARISONS ON ENERGY STREAMS WITH DIFFERENT ε

<table>
<thead>
<tr>
<th>Test</th>
<th>ε</th>
<th>Energy Streams (kWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>e^w_g 14.97 6.046 5.053 16.75 15.66 0.528</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>e^w_g 14.68 6.171 4.914 16.64 16.07 0.109</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>e^w_g 14.48 6.224 4.874 16.81 16.10 0.093</td>
</tr>
</tbody>
</table>

The comparison results of Index 1 and 2 are displayed in Table II. Index 1 increases with the growth of ε, which defines the size of the ambiguity set. The larger the value of ε, the more distributions are taken into account, resulting in a worst extreme distribution as well as a higher cost. For Index 2, a risk-neutral prosumer with ε = 0 may enjoy a lower cost in scenario 1 and scenario 5; while he has to pay a price for the strategy without robustness: In scenario 2 or 3, the cost is higher not only than those corresponding to two risk-aversive cases, but also higher than its predictive expectation.

Furthermore, we explore the impact of ε on energy flows (in kWh, for example, \( e^w_g = \sum_{t=1}^{T} x_{wg} \)). Results are listed in Table III, from which we can observe the following.

1) The allocation of wind energy changes with the value of ε; specifically, as ε becomes larger, more volatile wind energy is sold to the grid, and more demand is supplied by either the grid power or the storage, which is controllable.

2) The energy exchange between storage and grid decreases as ε grows larger. Originally, the storage plays a prime role in arbitrage; when the uncertainty increases, more storage capacity is used to eliminate the negative impact of uncertainty and ensure a reliable energy supply.

The change of SoC from another perspective demonstrates how the robustness impacts the energy management strategy. The average SoC trajectories throughout the whole time horizon in three tests are drawn in Fig. 6; in the test with a larger ε, the SoC is kept at a higher level to avoid potential risks.

C. Comparison Between SDDP and RDDP

We compare the effectiveness and efficiency of the proposed RDDP and the popular SDDP. Please refer to [29] for a complete description of the forward–backward pass algorithm for SDDP. We use the same ten observations in SDDP. The algorithm terminates when the upper bound is close enough to the lower bound. Basically, LP in each step can be solved efficiently within 0.02 s. As we use \( N = 10 \) observations, the optimality gap between lower and upper bounds shrinks to \( 6 \times 10^{-4} \) after ten forward passes and ten backward passes. Because all the scenarios must be visited in every backward pass, the computational time of SDDP is \( O(N^2) \), which grows quickly if more historical observations are available. As for RDDP, the computational time is \( K \times O(N) \) where the number of discrete points \( K = 10 \) is a constant.

Finally, we compare Index 2 in the five newly observed trajectories. The forward–backward pass algorithm of SDDP constructs accurate cost-to-go functions. From Table V, we can see that the relative errors between Index 2 offered by RDDP and SDDP in all five scenarios are less than 5%, demonstrating that the PWL expressions used in RDDP are accurate enough. Furthermore, the values offered by RDDP are slightly smaller, because in Algorithm 1, the dual variables for constructing PWL expression are extracted from discrete samples of parameters, and some critical vertices [as in (19)] that produces binding linear functions in the PWL expression may be missed. In this regard, the proposed method offers an under estimator for the cost-to-go function in theory. Nevertheless, as the sampled parameters are well distributed, the approximation is very close.
to the accurate one. To improve accuracy, we can incorporate more sampled points of state variables to generate candidate dual variables.

V. CONCLUSION

This article proposes a robust data-driven dynamic programming method for prosumer energy management considering uncertainties of electricity price, renewable generation, and demand. The problem can be solved via backward induction using only linear programming solver, and the key step entails a piecewise linear approximation for the cost-to-go function, which is constructed from dual linear programs.

Case studies corroborate the effectiveness of the proposed RDDP method, revealing that the dispatch strategy of RDDP can keep a proper balance between optimality and computational expense: Compared to SDDP, the computational time of RDDP is $K \times O(N)$, which is much shorter than that of SDDP, $O(N^2)$; besides, results show the suboptimality of RDDP is less than 4%. Such results are based on only ten observations of historical data; therefore, the RDDP formulation enjoys a degree of robustness: even in case of the scarcity of historical data, the strategies can still guarantee an acceptable operating economy.

However, the disadvantage of the proposed method is that when the dimension of state variables is high, the computational complexity will exponentially grow.

In general, such a study establishes a paradigm for the optimal energy management of the prosumer-like energy systems, including but not limited to residential users, micro grids, energy hubs, and industrial parks. Our ongoing work is to apply RDDP to storage operation in large power systems with network flow constraints and develop more efficient solution algorithms.

REFERENCES

Zhongjie Guo received the B.S. degree in electrical engineering from Central South University, Changsha, China, in 2018. He is currently working toward the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China.

His research interests include energy storage, renewable power generation, and power system operation.

Wei Wei (Senior Member, IEEE) received the B.Sc. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2008 and 2013, respectively.

He was a Postdoctoral Research Associate with Tsinghua University from 2013 to 2015. He was a Visiting Scholar with Cornell University, Ithaca, NY, USA, in 2014, and a Visiting Scholar with Harvard University, Cambridge, MA, USA, in 2015. He is currently an Associate Professor with Tsinghua University. His research interests include applied optimization and energy system economics.

Laijun Chen (Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2006 and 2011, respectively.

He is currently an Associate Professor with Tsinghua University. His research interests include power system analysis and control, and renewable energy integration.

Zhaojian Wang (Member, IEEE) received the B.S. degree in electrical engineering from Tianjin University, Tianjin, China, in 2013, and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2018.

He was a Postdoctoral Research Associate with Tsinghua University from 2018 to 2020. From 2016 to 2017, he was a joint Ph.D. student at California Institute of Technology, CA, USA. He is currently an Assistant Professor with Shanghai JiaoTong University, Shanghai, China. His research interests include power system distributed control and optimization.

João P. S. Catalão (Senior Member, IEEE) received the M.Sc. degree in electrical engineering from the Instituto Superior Tecnico (IST), Lisbon, Portugal, in 2003, and the Ph.D. and Habilitation degrees in electrical engineering from the University of Beira Interior (UBI), Covilha, Portugal, in 2007 and 2013, respectively.

He is currently a Professor with the Faculty of Engineering, University of Porto (FEUP), Porto, Portugal, and a Research Coordinator at INESC TEC. He was also appointed as a Visiting Professor by North China Electric Power University, Beijing, China. His research interests include power system operations and planning, hydro and thermal scheduling, wind and price forecasting, distributed renewable generation, demand response, and smart grids.

Shengwei Mei (Fellow, IEEE) received the B.Sc. degree in mathematics from Xinjiang University, Urumqi, China, in 1984, the M.Sc. degree in operations research from Tsinghua University, Beijing, China, in 1989, and the Ph.D. degree in automatic control from the Chinese Academy of Sciences, Beijing, in 1996.

He is currently a Professor with Tsinghua University. His research interests include power system complexity and control, game theory, and its application in power systems.