Fair Energy Scheduling for Vehicle-to-Grid Networks Using Adaptive Dynamic Programming

Shengli Xie, Senior Member, IEEE, Weifeng Zhong, Student Member, IEEE, Kan Xie, Rong Yu, Member, IEEE, and Yan Zhang, Senior Member, IEEE

Abstract—Research on the smart grid is being given enormous supports worldwide due to its great significance in solving environmental and energy crises. Electric vehicles (EVs), which are powered by clean energy, are adopted increasingly year by year. It is predictable that the huge charge load caused by high EV penetration will have a considerable impact on the reliability of the smart grid. Therefore, fair energy scheduling for EV charge and discharge is proposed in this paper. By using the vehicle-to-grid technology, the scheduler controls the electricity loads of EVs considering fairness in the residential distribution network. We propose contribution-based fairness, in which EVs with high contributions have high priorities to obtain charge energy. The contribution value is defined by both the charge/discharge energy and the timing of the action. EVs can achieve higher contribution values when discharging during the load peak hours. However, charging during this time will decrease the contribution values seriously. We formulate the fair energy scheduling problem as an infinite-horizon Markov decision process. The methodology of adaptive dynamic programming is employed to maximize the long-term fairness by processing online network training. The numerical results illustrate that the proposed EV energy scheduling is able to mitigate and flatten the peak load in the distribution network. Furthermore, contribution-based fairness achieves a fast recovery of EV batteries that have deeply discharged and guarantee fairness in the full charge time of all EVs.

Index Terms—Adaptive dynamic programming (ADP), contribution-based fairness, residential distribution network, smart grid, vehicle-to-grid (V2G).

I. INTRODUCTION

Climatic change, gas emissions, and depletion of fossil fuel reserves pose direct threats to the global economy and community, thus arousing awareness of new revolutions in the industry for emission reduction and energy conservation. The smart grid is considered to be the next-generation power grids and will integrate renewable energy resources, intelligent control, real-time communications, high efficiency, and high reliability [1]–[5]. In addition, electric vehicles (EVs) are receiving increasingly more attentions, because they use clean energy with zero discharge. Currently, advanced technologies have improved the performance of EVs. People increasingly accept and purchase EVs. In 2013, more than 96,000 EVs were sold in the U.S., a year-on-year growth of more than 80% [6].

The penetration of EVs leads to both challenges and opportunities for the smart grid. The main challenge is the huge load from EV charging. In residential distribution networks, the household load peak usually occurs in the evening. This is the same time when the EVs reach home and are plugged in for battery recharge, resulting in another charge load peak. These two load peaks will impact the distribution network together and doubly cause the degradation of voltage, power loss, overload, and harmful harmonics. It is demonstrated in [7] and [8] that uncoordinated free charging can result in a variety of serious problems in distribution networks.

From the positive perspective, most vehicles spend approximately 22 h per day in the parking state. During this time, the batteries of EVs are considered to be idle assets in the smart grid [9]. Aggregating a large number of EVs can offer a potential backup supporting efficient integration of intermittent renewable resources [10], [11]. Furthermore, the concept of vehicle-to-grid (V2G) is proposed to enable two-way communications and bidirectional energy flow between EVs and the grid. With the V2G technology, EVs can provide the smart grid with auxiliary services, including frequency regulation [12], [13], and load shaving [14], [15]. Moreover, EV users can share energy with one another in V2G networks [16]. Smart pricing will be employed to achieve the demand response of EVs. EV users will adjust their charge or discharge behavior based on the states-of-charge (SOCs) of EVs and the price information [17].

V2G technology plays a significant role in the reliability of the smart grid in the future. Many V2G control schemes have been proposed to use EVs to provide auxiliary services. The V2G control can be categorized into unidirectional V2G [18]–[21] and bidirectional V2G [22]–[26]. Unidirectional V2G only includes EV charge control. EV charge load can be controlled to fill the valley of household load, which is called load valley filling. Bidirectional V2G refers to both charge and discharge controls, so we can achieve load...
peak shaving, which means that EVs can discharge energy to shave the peak of household load. Because EV mobility and EV demand dynamically change over time, the problems of both unidirectional V2G and bidirectional V2G are usually formulated as programming problems and solved via linear programming [18], [24], quadratic programming [19], [20], dynamic programming [22], [25], or particle swarm optimization [21]. By using the techniques described above, the basic household load profile and the EV model, including charge location, arrival time and charge energy, are assumed to be known in advance. To predict this information, day-ahead and historical data are used in [8], [23], and [26].

The literature referenced above formulates the EV energy scheduling problems as stationary problems. However, in real scenarios, the demand and mobility of each individual EV will be different each day. EV energy scheduling should be a dynamic process, and the corresponding prior knowledge is often unknown. Therefore, the conventional techniques with fixed prior knowledge are improper for solving the EV energy scheduling problem. Therefore, we propose fair energy scheduling by using the methodology of adaptive dynamic programming (ADP). During EV energy scheduling, new EVs plug in dynamically. The ADP technique can learn from the outside environment and estimate the long-term future system cost. By observing the information of EVs and the distribution network, the scheduler can perform optimized EV energy scheduling, even if the future information is unknown.

Given the limited capacity of the residential distribution network, the electricity supply may not fulfill the EV charge demand at any time. Therefore, fairness in EV charging should be considered to ensure the charge opportunity of each EV. Moreover, during load peak, it is also necessary to assign the discharge tasks to EVs and guarantee discharge fairness. In this paper, we consider two fairness operations during EV energy scheduling. In the discharge period, we adopt SOC-based fairness to select EVs for discharging. In this period, contribution-based fairness is proposed for charge energy allocation. We construct the cost function in the objective function. The penalty value is in inverse proportion to the energy consumption of other EVs that have deeply discharged, and ensure fairness in the full charge time of all EVs.

The remainder of this paper is organized as follows. Related work on the fairness models of V2G is given in Section II. The scenario of the V2G network and the proposed fair operation are presented in Section III. The energy scheduling problem is formulated as an MDP in Section IV. Section V discusses the ADP solution for the formulated MDP problem. The numerical results are shown in Section VI. Finally, the conclusions are drawn in Section VII.

II. RELATED WORKS

We introduce the existing fairness models for V2G problems in this section. Fairness problems arise in both charge and discharge periods. In the charge period, available energy should be fairly allocated to EVs. In the discharge period, the discharge jobs should be fairly assigned to EVs. Most existing fairness models aim at the EV charging or energy allocation problem. Chung et al. [27] consider the fairness of charge time allocation, which is calculated according to the mean and the standard deviation of EV charge time, and design an algorithm to maximize this fairness. In [28], the scheduler also considers the household loads of EV users while controlling EV charging. The available energy is allocated to households based on max–min fairness. The EV charge energy is the allocated energy minus the energy consumption of other plug loads in the household. Thus, one’s EV will obtain more charge energy if he has less household demand. In [29], a penalty is set to assess the unsatisfied demand of the EV in the objective function. The penalty value is in inverse proportion to the energy consumption of the EV. By doing so, the available energy will be fairly allocated to all EVs.

Some fairness concepts in communications theory are used in the EV charging problem. Proportional fairness is employed in [30] and [31] for EV charge energy scheduling. Proportional fairness is a compromise-based scheduling method, which is widely used in rate control in communications networks. It attempts to maximize resource utilization while allowing all users to have a minimum level of service. Its utility function has a logarithmic form, and the formulated problem is usually an NP-hard problem. From the concept of congestion control in the transmission control protocol, Liu and McLoone [32] and Liu et al. [33] adopt the additive increase multiplicative decrease (AIMD) algorithm to achieve fair coordinated charging of EVs. This algorithm allows EVs to increase their charge rates slowly. If the total charge load exceeds the expected level, EVs will be instructed to halve their charge rates. The energy allocation should be fair in long-term observation. The advantage of the AIMD algorithm is that the charge control can be implemented easily in a distributed manner.

In the smart grid, large-scale energy storage devices will be deployed to buffer the intermittent renewable energy. In the
case of an EV fast charge, a discharge of large-scale energy storage devices may be needed to support the high charge rates of EVs. The fair use of energy storage devices for EV charging is presented in [34]. If the upcoming EV charge demand is forecasted, the storage controller will restrict its current discharge rate to reserve enough energy for future charge demand. Paul and Aisu [34] address the fairness problem of energy storage discharge. Essentially, it is also a problem of energy allocation for charging.

The fairness problem is also discussed in EV discharge control. In a power system, the electricity frequency may fluctuate if there is imbalance between electricity generation and load in a short time scale. In [35], the EVs are aggregated to provide a frequency regulation service. EVs must discharge and recharge frequently to mitigate the frequency fluctuation. The aggregator assigns the regulation-up tasks to discharging EVs and then assigns the regulation-down tasks to charging EVs. The selection of EV is based on a fairness criterion, which is measured by the SOC of EVs. Minimum deviation and proportional fairness are considered as the fairness criteria in [35]. A distributed algorithm is proposed in [36] to fairly control the discharge rates of EVs. The discharge fairness defined in [36] is that EVs should have the same discharged energy to avoid the case in which some EVs obtain more money than others. In [37], the number of charge and discharge cycles of an EV battery is considered in the V2G control. The V2G system offers fair incentive to EVs based on the lifetime depreciation of their batteries.

As indicated above, the fair charge problems and the fair discharge problems are discussed separately in the literature. They usually divide EVs into charging EVs and discharging EVs. In the charge control period, energy is fairly allocated to charging EVs. In the discharge control period, discharge tasks are fairly assigned to discharging EVs. Their methods ignore the fact that an individual EV may perform discharge and recharge alternately and that both actions influence the fairness simultaneously.

In the smart grid scenario, EVs are regarded as flexible loads of household users. There is also a fairness problem that must be addressed when controlling the household load. Vuppala et al. [38] guarantee the fairness by considering the diversities of the user type and load type in the pricing scheme. Fixed load and optional load are separated and matched to different prices. In [39], users will be rewarded according to their contributions to ensure fairness. One can achieve a higher contribution if he can provide a more flexible load or allow the load to run in off-peak hours. Inspired by fair operations in the household load control problem, we define a new fairness criterion for the V2G problem.

This paper simultaneously considers both charge and discharge issues in the definition of fairness. We propose contribution-based fairness for EV charge control. If EVs make high contributions, they will have high priorities to obtain charge energy. The contribution value is determined by both the charge and discharge energy and the action time. The proposed fair operation in V2G energy scheduling is introduced in Section III.

III. SYSTEM MODEL

A. V2G Networks

The diagram of the V2G residential distribution network in the smart grid is shown in Fig. 1. This residential distribution network serves a number of houses equipped with EV chargers. After arriving home, EVs will plug in and charge their batteries. The original distribution network is designed to support the basic household load. If there is no extra facility investment, the old distribution network must feed the extra load from EV charging as well. V2G technologies are required to implement EV energy scheduling and prevent the negative impacts caused by free charging. The two-way communications enable the information exchange between EVs and the scheduler. The scheduler knows each household’s basic load and EV information (i.e., SOC and contribution) and, then, controls the charging and discharging of EVs centrally. This centralized V2G control manner is popular in the scenarios of household distribution networks and microgrids [40].

B. Fair Energy Scheduling

To promote grid reliability, the scheduler controls the EV load to ensure that the actual total load in the distribution network does not exceed the expected load level. When the network load exceeds the expected level, the scheduler selects EVs to discharge, mitigating the total network load. The discharge tasks are assigned to EVs based on SOC-based fairness. EVs with high SOC will be selected to discharge and obtain profit, which encourages the EV users to charge their EVs before they come home, and transport energy from outside to the household grid. When the network load lies below the expected level, the scheduler will allocate the available energy to EVs based on their contributions. The EVs with high contributions will have high priorities to obtain charge energy. This avoids long charge times of EVs that have deeply discharged. Moreover, the contribution is evaluated by the charge or discharge energy and its timing. The contribution values increase while EVs discharge and decrease...
while EVs charge. Discharging during the peak time results in more contributions than discharging in an off-peak time. In addition, charging during a heavy-load time reduces the contribution value more seriously than charging in a valley point. This operation avoids a scenario, in which EVs with little contribution occupy a large amount of charge energy. The contribution values of users will subsequently drop to a similar level. Thus, contribution-based fairness also achieves fairness of full charge time for all EVs, which will be illustrated in the simulation results.

The scheduler estimates and maximizes long-term fairness during energy scheduling. The EVs only provide the SOC information to the scheduler at each stage. The scheduler then launches the control commands of turns ON/OFF the chargers (or dischargers) to EVs. The computations of EV contribution values and the estimation of long-term fairness are completed by the scheduler. Little information is required to be exchanged during energy scheduling. Thus, the communication overhead is low. The cases of dynamical price control and fast charging are not involved in this paper. We assume that EVs with high demands will charge quickly by drawing energy from other resources. We consider that EV charging under the control of the scheduler can lead to a low electricity fee. The EV users who will not drive their EVs before tomorrow morning agree and participate in the energy scheduling.

IV. PROBLEM FORMULATION

The EV energy scheduling problem is formulated as an infinite-horizon MDP in this section. By observing the system state, the scheduler will execute energy scheduling at each stage to maximize the long-term fairness. The components of the MDP are provided as follows.

A. Stage

The energy scheduling process is naturally divided into a series of stages, indexed by \( t = 1, 2, \ldots \). Compared with the whole charging process, the duration of one stage \( \Delta t \) is regarded as significantly small, so the number of stages is observed as infinite. The scheduler executes energy scheduling at the beginning of each stage. New EVs arrive at the beginning of the stage and leave at the end of the stage (or the beginning of the next stage) dynamically. For brevity, the beginning of stage \( t \) is called stage \( t \) herein.

B. State

The scheduler observes the system state at each stage. We define the system state as comprising the state of EVs and the state of the distribution network. For the network state, \( L^N(t) \) presents the expected total load of the distribution network at stage \( t \), which is decided by the generation capacity and the distribution transformer rating. The network load includes the EV load and household load. Thus, we have

\[
L^N(t) = L^E(t) + L^H(t)
\]

where \( L^E(t) \) denotes the expected total demand of EVs and \( L^H(t) \) indicates the total household load at stage \( t \).

We have \( L^E(t) > 0 \) in the charge period and \( L^E(t) < 0 \) in the discharge period. The maximum number of EVs is \( K \). Each EV is indexed by \( k = 1, 2, \ldots, K \). The state of EVs is provided by the SOC of their batteries and their contribution values. The contribution value of EV \( k \) is denoted by \( C_k(t) \), and its definition is provided in Section IV-D. Let \( S_k(t) \) denote the SOC of EV \( k \) at stage \( t \). We have

\[
S_k^{\text{min}} \leq S_k(t) \leq S_k^{\text{max}}
\]

where \( S_k^{\text{min}} \) and \( S_k^{\text{max}} \) represent the minimum and maximum SOC levels of EV \( k \), respectively. Therefore, the system state at stage \( t \) is defined by

\[
x(t) = \{L^E(t), S_k(t), C_k(t)|k = 1, 2, \ldots, K\}.
\]

C. Action

At each stage, the scheduler will execute energy scheduling based on the system state \( x(t) \). The action at stage \( t \) is defined by

\[
u(t) = \{u_k(t)|k = 1, 2, \ldots, K\}
\]

where \( u_k(t) \) denotes the control action on EV \( k \) at stage \( t \). If \( u_k(t) = 1 \), EV \( k \) is permitted to charge. If \( u_k(t) = -1 \), EV \( k \) is selected to discharge. \( u_k(t) = 0 \) means that EV \( k \) is not allowed to charge or discharge, or EV \( k \) has not arrived yet. The actual total load from EVs at stage \( t \) is provided by \( L(t) = \sum_{k=1}^{K} L_k(t) \), where \( L_k(t) \) presents the actual load of EV \( k \), and we have \( L_k(t) = u_k(t)P \Delta t \). We define that the EV charge rate equals the discharge rate, denoted by \( P \) in kW. Let \( L^{\text{max}}(t) \) then denote the maximum load provided by EVs at stage \( t \). We have \( L^{\text{max}}(t) = \sum_{k=1}^{K} L_k^{\text{max}}(t) \) and \( L_k^{\text{max}}(t) = I_1(S_k(k) > 0)I_2(L^E(t) > 0)P \Delta t \). \( I_1(A) \) and \( I_2(A) \) are indicator functions. \( I_1(A) \) equals 1 if \( A \) is true or 0 if \( A \) is false. \( I_2(A) \) equals 1 if \( A \) is true or \(-1 \) if \( A \) is false. The following constraint shows the relationship among actual EV load \( L(t) \), maximum EV load \( L^{\text{max}}(t) \), and the expected EV load \( L^E(t) \):

\[
L(t) = \min\{L^{\text{max}}(t), L^E(t)\}.
\]

The action space is provided by \( U \). Constraint (5) indicates that the candidate actions at stage \( t \) should be selected in \( U[x(t)] \), which denotes the subset of \( U \) for a given system state \( x(t) \).

D. Cost Function

The scheduler executes energy scheduling \( u(t) \) based on the system state \( x(t) \) to minimize the system cost. The system cost is defined as the fairness of energy scheduling. The penalty \( \beta_k(t) \) is defined to assess the fairness provided by

\[
\beta_k(t) = \begin{cases} 
1/S_k(t), & u_k(t) < 0 \\
1/C_k(t), & u_k(t) > 0
\end{cases}
\]

\[
\beta_k(t) = \begin{cases} 
1/S_k(t), & u_k(t) < 0 \\
\beta_k(t-1), & u_k(t) = 0 \\
1/C_k(t), & u_k(t) > 0
\end{cases}
\]

where \( c_k(t) \) presents the actual total demand of EVs at stage \( t \). Thus, the EV with high \( S_k(t) \) will be selected first to discharge. In the
charge period, i.e., \( u_k(t) > 0 \), the scheduler controls the charging of EVs based on their contributions \( C_k(t) \). The EVs with high contributions will have high priorities to charge. The contribution \( C_k(t) \) of EV \( k \) is presented by

\[
C_k(t+1) = (1-\delta)C_k(t) - \delta \omega(t)L_k(t) \tag{7}
\]

where \( \delta \) is within \([0,1]\). If \( \delta \) is sufficiently small, \( C_k(t) \) is decided by the long-term average contribution. If \( \delta \) is large, the current weighted EV load \( \omega(t)L_k(t) \) influences the contribution significantly. We define the weight of the EV load to describe the importance of load timing provided by

\[
\omega(t) = |L^E(t)|. \tag{8}
\]

Low \( |L^E(t)| \) means that the actual total network load is near the expected level. High \( |L^E(t)| \) indicates that there is a large gap between the actual and expected network loads, which must be compensated by the EV load. Thus, the charging and discharging of EVs in high \( |L^E(t)| \) should lead to higher contributions than in low \( |L^E(t)| \). The \( C_k(t+1) \) increases while EVs discharge, i.e., \( L_k(t) < 0 \) and decreases while EVs charge, i.e., \( L_k(t) > 0 \). Thus, in the later charging period, all EVs will tend to have similar values of \( C_k(t) \).

Let \( \beta_k(t) \) represent the normalized value of \( \hat{\beta}_k(t) \) within \([0,1]\), denoted by

\[
\hat{\beta}_k(t) = \frac{\beta_k(t) - \min_k \beta_k(t)}{\max_k \beta_k(t) - \min_k \beta_k(t)}. \tag{9}
\]

The cost function at stage \( t \) is defined by

\[
U(t) = \sum_{k=1}^{K} \hat{\beta}_k(t)u_k(t) \tag{10}
\]

which calculates only the cost at stage \( t \). In the whole energy scheduling process, the long-term system cost is provided by

\[
J(t) = \sum_{j=1}^{\infty} \gamma^{j-t} U(j) \tag{11}
\]

where \( \gamma \) denotes the discount factor in the region of \([0,1]\). If \( \gamma = 0 \), we only consider the cost in the current stage and neglect the future costs. \( \gamma = 1 \) means that all future costs are equally considered for infinite stages. The goal of the scheduler is to obtain the optimal policy of energy scheduling \( \{u^*(t), u^*(t+1), \ldots\} \) to achieve the minimum long-term system cost, denoted by \( J^*(t) \).

V. ADAPTIVE DYNAMIC PROGRAMMING SOLUTION

In this section, the methodology of ADP is employed to solve the EV energy scheduling problem in the framework of MDP. The ADP approaches can achieve robust stabilization while controlling nonlinear systems with uncertainties [41]–[44]. Song et al. [45] proposed new multiple actor-critic structures for ADP to achieve optimal control for continuous-time unknown nonlinear systems. Value iteration [46], [47] and policy iteration [48] ADP algorithms are newly proposed to control infinite-horizon discrete-time nonlinear systems. For ADP application, the ADP approach is employed in [49] to implement optimal temperature control for water-gas shift reaction systems. In [50], the ADP algorithm achieves optimal multibattery control for home energy management systems.

A. Bellman Equation

In general, by solving the Bellman optimality equation, the optimal energy scheduling policy can be obtained. The optimal long-term system cost from stage \( t \) is presented by

\[
J^*(t) = \min_{u(t) \in U(t)} \{ U(t) + \gamma J^*(t+1) \}. \tag{12}
\]

The optimality principle is to minimize \( J \) in the immediate future and, subsequently, minimize the sum of \( U \) over all stages. If there is an available way to calculate the optimal long-term system cost \( J^*(t+1) \), the optimal energy scheduling action at stage \( t \) can be obtained by solving

\[
u^*(t) = \arg \min \{ U(t) + \gamma J^*(t+1) \}. \tag{13}\]

Because of the huge size of the state space and the overwhelming computational requirement, the accurate solution of \( J^*(t+1) \) in (13) may incur the curse of dimensionality. In addition, EV energy scheduling is formulated as an infinite-horizon MDP. The future system state and cost are unknown. Hence, we employ the ADP technique to produce an approximate optimal solution.

The conventional ADP architecture has three networks: action, model, and critic networks. The three networks can be implemented by neural network training. There is an obstacle in solving our energy scheduling problem by using ADP, because the scheduling actions are discrete variables, i.e., \( u_k(t) = -1, 0, 1 \). Action vector \( u(t) \) will be the output of the action network and the input of the model network, which prohibits the two networks from using neural network structures. However, in our energy scheduling problem, the cost function at the current stage is calculable, and the energy scheduling process is observable. Thus, the action and model networks have no need to use neural network structures to perform the function approximation. In this paper, we design a modified ADP architecture to solve the fair energy scheduling problem.

B. Modified ADP Architecture

As shown in Fig. 2, the modified ADP architecture includes three components: scheduler, system, and critic network.

1) Scheduler: The inputs of the scheduler component are the system state \( x(t) \) and the output of the critic network, and the output is the scheduling action \( u(t) \). The target of the scheduler is to regulate \( u(t) \) to minimize the long-term system cost \( \hat{J}(t) \). Given \( \hat{J}(t) = U(t) + \gamma \hat{J}(t+1) \), the operation of the scheduler can be described by

\[
u(t) = \arg \min \{ U(t) + \gamma \hat{J}(t+1) \} \tag{14}\]

where \( \hat{J}(t+1) \) is the output of the critic network. Given in (6)–(10), the calculation of the cost function \( U(t) \) requires the system state, so \( x(t) \) is also an input of the scheduler component. Because (14) is calculable, the scheduler component has no need to adopt a neural network for function approximation.
Fig. 2. Modified ADP architecture for the fair energy scheduling.

2) System: The system component in Fig. 2 describes the system dynamics. Its inputs are scheduling decision \( u(t) \) and system state \( x(t) \), and the output is the system state of the next stage \( x(t+1) \). The system state consists of the state of EVs and the state of the distribution network. The network state \( L^E(t) \) is decided by the daily household consumption and the expected network load, shown in (1), which cannot be controlled by the scheduler. For the EV state, newly arrived EVs can participate in the energy scheduling at the beginning of each stage. An EV whose SOC reaches its maximum value will stop charging. If the minimum SOC is reached, the EV will stop discharging. For charging or discharging EVs, their SOCs at the next stage are provided by

\[
S_k(t+1) = S_k(t) + L_k(t)
\]

which is subject to constraint (2). The update of the contribution value is provided in (7). Thus, the system state transition is observable.

3) Critic Network: The input of the critic network is \( x(t+1) \), and the output is the approximate long-term system cost \( \hat{J}(t+1) \). The critic network aims at \( \hat{J}(t+1) \) approaching the optimal system cost \( J^*(t+1) \). Because the exact solution of \( J^*(t+1) \) is unrealistic, the critic network adopts the backpropagation neural network (BPNN) for function approximation. Typically, the BPNN has three layers. Because the system state \( x(t+1) \) has \( 2K+1 \) elements, the input layer has \( 2K+1 \) neurons. The hidden layer has \( 3K \) neurons and uses the sigmoidal function provided by

\[
f(x) = \frac{1 - e^{-x}}{1 + e^{-x}}.
\]

The output layer has one neuron corresponding to one output of \( \hat{J}(t+1) \). The purelin function is used in the output layer.

C. Critic Network Training

In the training of the critic network, the gradient descent method is used, and the following error function will be minimized at stage \( t \):

\[
E_c(t) = \frac{1}{2}[\hat{J}(t) - U(t) - \gamma \hat{J}(t+1)]^2
\]

where \( \hat{J}(t) = \hat{J}[x(t), u(t), W_c(t)] \), and \( W_c(t) \) are the weight parameter of the critic network. If the network training achieves \( E_c(t) = 0 \), we have the following derivation:

\[
\hat{J}(t) = U(t) + \gamma \hat{J}(t+1) = U(t) + \gamma [U(t+1) + \gamma \hat{J}(t+2)] = \cdots = \sum_{j=t}^{\infty} \gamma^{j-t}U(j)
\]

which is the long-term system cost shown in (11). By minimizing the error function (17), the critic network training can estimate the approximate value of the long-term system cost for the next stage. The weight update iteration in the critic network is provided as follows:

\[
\Delta W_c(t) = l_c(t) \left[ \frac{\partial E_c(t)}{\partial W_c(t)} \right] = l_c(t) \left[ \frac{\partial E_c(t)}{\partial \hat{J}(t)} \frac{\partial \hat{J}(t)}{\partial W_c(t)} \right]
\]

\[
W_c(t+1) = W_c(t) + \Delta W_c(t)
\]

where \( l_c(t) \) is the positive learning rate. The method of the primarily gradient-based network training can be found in [51] and [52].

At each stage, the scheduler determines an action based on the current system state and the approximate function. After the action, the cost at the current stage is obtained. The critic network updates its weight parameter according to the cost per stage and produces a new approximate function. At the new stage, the same operation is executed again. Therefore, during EV energy scheduling, the critic network can observe increasingly more state transitions and learn the feature of the system. The network weights are refreshed every stage, so the approximate function increasingly resembles the objective function. Note that the scheduler has no need to well train the network before control. The weights of the network are updated stage by stage. The computational requirement for once iteration is low, which can be completed at one stage (0.2 h in simulation).

VI. NUMERICAL RESULTS

A. Simulation Setup

The V2G residential distribution network is modeled by using the IEEE 34 node test feeder [53], which is shown in Fig. 3. Two scenarios of 250 EVs and 1000 EVs are
considered, i.e., $K = 250$ and $K = 1000$. We select ten nodes in the network. Each node connects with 25 or 100 houses. Each house has its daily household load and is equipped with an EV charger. The EVs will arrive in the evening and then plug in for energy scheduling. We consider that household EV users are commuters who start their first trips from homes in the morning and finish the last trips by arriving home in the evening [8].

We model the arrival of household EVs as a normal distribution over time [11], [54], [55]. The SOC values are normalized within $[0, 1]$. The SOC of arrived EVs is initiated in $[20\%, 70\%]$ randomly. The duration of each stage is 0.2 h, so that 1 h has five stages, and one day has 120 stages. The charge or discharge rate $P$ is fixed at 4 kW. The EV battery size $S_{\text{max}}^k$ is 11 kWh [19]. The minimum value $S_{\text{min}}^k$ is 2 kWh. The critic network training and the simulation results are obtained by using the self-developed simulator based on MATLAB.

### B. Convergence Performance

First, we present the convergence performance of the proposed modified ADP for solving the EV energy scheduling problem. Fig. 4(a) shows the convergence of the critic network output with different EV populations. We find that both curves have periodic disturbances. The periodic disturbance is caused by the daily changes of household load and EV load. Thus, it is inappropriate to use an excessively low or high value of learning rate $l_c$. For a low $l_c$, the dynamical EV activities overwhelm the weak learning ability of critic network. A high $l_c$ causes the weight parameters to dramatically change and miss the optimal convergence value. We find that the convergence is lost easily if $l_c > 0.1$ or $l_c < 0.001$. In the simulation, the learning rate for online network training is set to $l_c = 0.02$. In Fig. 4(a), we find that the EV population mainly impacts the stability of convergence. The randomness of EV mobility and demand influences the estimation of the long-term system cost. The critic network output fluctuates more seriously when there are more EVs.

Then, we compare the proposed ADP with feature-based linear approximation [56] in the convergence performance in the case of 250 EVs. In the linear approximation, the system feature is extracted from the system state. Thus, the feature is a function of system state, i.e., $f[x(t)]$. Here, we simply use $f[x(t)] = x(t)$. Thus, the approximation of the system cost function is provided by $\hat{J}(t) = \theta^T x(t)$, where $\theta$ is the parameter vector for linear approximation. As shown in Fig. 4(b), the proposed ADP outperforms the feature-based linear approximation. In ADP, the critic network that adopts a neural network is more suitable for nonlinear function approximation. Compared with linear approximation, the critic network is able to achieve more stable convergence.

In the simulation, the discount factor is set to $\gamma = 0.9$. We find that the convergence is lost easily when $\gamma > 0.95$. This is because the long-term system cost will increase to infinity if $\gamma$ is too close to 1. As shown in Fig. 4(c), a smaller $\gamma$ value can improve the convergence. However, a small $\gamma$ value indicates that fewer costs from future stages will be considered in the system cost. Thus, the short-term system cost can be estimated more quickly. During energy scheduling in practice, we can set a low $\gamma$ value to improve the convergence of critic network output at the beginning of learning. $\gamma$ then increases stage by stage and stops increasing when it reaches...
a sufficiently high value. This operation can simultaneously achieve a fast convergence and long-term cost estimation. As shown in Fig. 4, our modified ADP architecture is able to achieve convergence and estimate the optimal system cost by online network training. Thus, an optimized fair energy scheduling policy can be obtained.

C. Fair Energy Scheduling Performance

We use the scenario of 1000 EVs to demonstrate the performance of the proposed fair energy scheduling in the V2G residential distribution network. As shown in Fig. 5, EVs reach home during the basic household load peak. Thus, uncoordinated EV charging causes an exceeded load peak over an existing basic load peak, which is unexpected by the smart grid. In the proposed method, the charging and discharging of EVs are controlled by the scheduler centrally. The total load of the distribution network is restricted under the expected level ($L^N = 650$ kWh), which mitigates and stabilizes the network load level.

Our approach can still perform well in the case of a large battery size that requires a long charge time. If we have higher $L^N$, the EV charging can be finished earlier. The scheduler, which knows the information about the EVs, can submit an application for higher $L^N$ to the main grid, so that the EVs with large batteries can finish charging before the following morning. Our method can perform well under any value of $L^N$. Moreover, EVs with large batteries can usually perform fast charging. Our proposed method is used only when the total network load exceeds the expected level. As shown in Fig. 5, after approximately 2 h, there is sufficient energy for EV free charging. At this time, there is no need to use our method, and EVs with large batteries can perform fast charging, so they can use the available energy sufficiently and shorten the charging time. The operation of fast charging is out of the scope of this paper.

Next, we present the performance of the contribution-based fairness in energy scheduling. We compare our contribution-based fairness with SOC-based fairness [29] and household-based fairness [28]. In the discharge period, all three approaches guarantee that EVs with high SOC have high priority to discharge and obtain profit. In the charge period, SOC-based fairness proposes that EVs with low SOC will have a higher chance of charging. In household-based fairness, the EVs can obtain more charging energy in the case of low household load.

Fig. 6(a) shows the SOC of an EV that has performed discharge. This EV arrives early and has sufficient SOC. Hence, it is selected to discharge during the load peak. After that, the scheduler allocates the available energy to this EV based on the three types of fair operation. In our contribution-based fairness, because this EV has made a discharge contribution, it has high priority to obtain energy in the charge period, and its SOC level increases faster than that in the other two fair operations. Fig. 6(b) shows the SOC of an EV that has not discharged. This EV has low SOC when plugging in. Thus, it is not selected for discharging and has no discharge contribution. Hence, it has a chance to charge later. Fig. 7 shows the SOC of two selected EVs under contribution-based fairness. EV 1 discharges at the peak time. EV 2 discharges at an off-peak time.
The fair energy scheduling problem is formulated as an infinite-horizon MDP. The methodology of ADP is employed to maximize long-term fairness by processing online network training. The numerical results illustrate that the proposed EV energy scheduling can mitigate and flatten the peak load in the distribution network. Furthermore, the proposed fair operation encourages EVs to provide discharging services. EV owners prefer to charge their batteries before arriving home and discharge in the residential grid to obtain profits and achieve higher contributions. Furthermore, contribution-based fairness achieves the fast recovery of EV batteries that have deeply discharged. In addition, the fairness of full charging time among all EVs is ensured.

REFERENCES


Kan Xie was born in Hubei, China. He received the M.S. degree in software engineering from the South China University of Technology, Guangzhou, China, in 2009. He is currently pursuing the Ph.D. degree in intelligent signal and information processing with the Guangdong University of Technology, Guangzhou. His current research interests include machine learning, nonnegative signal processing, blind signal processing, and smart grid.

Rong Yu (S’05–M’08) received the Ph.D. degree from Tsinghua University, Beijing, China, in 2007. He is currently a Full Professor with the School of Automation, Guangdong University of Technology, Guangzhou, China. He has co-authored over 100 international journal and conference papers, and holds over 30 patents. His current research interests include information network and data, including vehicular network, mobile cloud computing, smart grid, Internet of Things (IoT), and cognitive radio. Dr. Yu currently serves as the Deputy Secretary General of IoT Industry Alliance of Guangdong and the Deputy Head of the IoT Engineering Center of Guangdong.

Yan Zhang (M’05–SM’10) received the Ph.D. degree from the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore. He is currently the Head with the Department of Networks, Simula Research Laboratory, Fornebu, Norway, and a part-time Associate Professor with the Department of Informatics, University of Oslo, Oslo, Norway. His current research interests include wireless networks and reliable and secure cyberphysical systems, such as healthcare, transport, and smart grid.

Dr. Zhang is a Senior Member of the IEEE Communications Society, the IEEE Vehicular Technology Society, and the IEEE Computer Society. He serves as the Technical Program Committee (TPC) Member for numerous international conferences, including the IEEE Conference on Computer Communications, the IEEE International Conference on Communications, the IEEE Global Communications Conference, and the IEEE Wireless Communications and Networking Conference. He received eight best paper awards. He is an Associate Editor or on the Editorial Board of a number of well-established scientific international journals, such as Wiley Wireless Communications and Mobile Computing. He also serves as the Guest Editor of the IEEE TRANSACTIONS ON SMART GRID, the IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, the IEEE Communications Magazine, the IEEE WIRELESS COMMUNICATIONS, and the IEEE Network. He serves as the Chair in a number of conferences, including the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications in 2016, IEEE Cloudcom Conference in 2015 and 2016, the IEEE Consumer Communications and Networking Conference 2016, IEEE International Conference on Communications in China 2016, the International Wireless Internet Conference in 2016, and the IEEE International Conference on Smart Grid Communications 2015.