Adaptive Dynamic Programming-Based Optimal Control Scheme for Energy Storage Systems With Solar Renewable Energy

Qinglai Wei, *Member*, *IEEE*, Guang Shi, Ruizhuo Song, *Member*, *IEEE*, and Yu Liu

Abstract— **In this paper, a novel optimal energy storage control scheme is investigated in smart grid environments with solar renewable energy. Based on the idea of adaptive dynamic programming (ADP), a self-learning algorithm is constructed to obtain the iterative control law sequence of the battery. Based on the data of the real-time electricity price (electricity rate in brief), the load demand (load in brief), and the solar renewable energy (solar energy in brief), the optimal performance index function, which minimizes the total electricity cost and simultaneously extends the battery's lifetime, is established. A new analysis method of the iterative ADP algorithm is developed to guarantee the convergence of the iterative value function to the optimum under iterative control law sequence for any time index in a period. Numerical results and comparisons are presented to illustrate the effectiveness of the developed algorithm.**

Index Terms— **Adaptive critic designs, adaptive dynamic programming, approximate dynamic programming, solar renewable energy, energy storage system, optimal control, energy storage.**

I. Introduction

ALONG with the development of smart grid, more and more intelligence has been required in the design of smart energy storage systems [1]–[3]. Energy storage systems of the future will provide end users a better energy management, reducing waste and advanced optimization technology. Renewable energies, electricity load, and energy storage equipments (battery) can be defined as smart energy storage systems which are able to operate with physically islanded [4], [5]. The intelligent optimal utilization of energy storage, along with the development of computational intelligences, has received much attention in recent years [6]–[8].

Adaptive dynamic programming (ADP) is an important and powerful brain-like intelligent optimal control method [9]–[20], due to its strong abilities of self-learning and adaptivity, and has widely been applied to obtain the optimal control for energy storage systems [21]–[24]. In [25], an action-dependent heuristic dynamic programming (ADHDP) algorithm, which is also called Q-learning algorithm [26], was proposed to obtain the optimal energy storage (battery) control law in achieving minimization of the cost through neural network learning. In [27], a time-based Q-learning algorithm was proposed to obtain the optimal battery control law to minimize the total electricity cost for energy storage systems, where the wind and solar energies were taken into consideration. In [28], a particle swarm optimization (PSO) algorithm was introduced to pretrain the weights of the neural networks, which speeds up the training of neural networks in ADP. In [29], a new event-triggered ADP algorithm was proposed to obtain the optimal frequency control law of the load.

In previous ADP algorithms for energy storage systems, however, it is mainly focused on the structure improvements, while the properties of the proposed ADP structures are seldom analyzed. In this case, the optimality of the achieved control scheme cannot be guaranteed. Second, the previous ADP algorithms required that the time index t reach infinity to obtain the optimal performance index function, which means that the optimal performance index function and optimal control law are time-invariant functions as $t \to \infty$. As the electricity rate, the load demand, and the solar renewable energy are generally time-variant functions, the converged time-invariant functions cannot effectively approximate the optimal performance index function and optimal control law of the energy storage systems. In [30], without considering the renewable energy, a dual iterative Q-learning algorithm of ADP was proposed to obtain the optimal battery control policy for energy storage systems, where it was proven that the iterative Q function is convergent to the optimum, as the iteration index increases to infinity. Considering the renewable energy, it presents a more complex scheme comparing with [30], where the optimal decisions for the power flows should cover all the possibilities for the renewable energy resource, the load, and the battery. Hence, the ADP algorithm and the property analysis in [30] cannot directly be applied to energy storage systems with renewable energy. A new ADP algorithm with new property analysis methods is necessary for the optimal battery control of the energy storage systems with renewable energy. This motivates our research.

In this paper, inspired by [25], [27], [30], a new iterative ADP algorithm is developed to obtain the optimal battery control for energy storage systems with solar renewable energy. According to the data of the electricity rate, the load, and the solar energy, the energy storage system is described and the optimization objective is established. Based on the established

This work was supported in part by the National Natural Science Foundation of China under Grants 61374105, 61533017, 61273140, 61673054, 61503379, and 61233001.

Q. Wei and G. Shi are with The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China and also with the University of Chinese Academy of Sciences, Beijing 100049, China. (phone: +86-10-82544517; fax: +86-10-82544799; emails: qinglai.wei@ia.ac.cn, shiguang2012@ia.ac.cn).

R. Song is with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China (email: derong@ustb.edu.cn, ruishuosong@ustb.edu.cn).

Y. Liu is with Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: yu.liu@mail.ia.ac.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2017.2674581, IEEE Transactions on Industrial Electronics

system, the iterative ADP algorithm is implemented, where in each iteration, an iterative control law sequence of a period is obtained instead of a single control law. Next, the properties of the iterative ADP algorithm are analyzed. As the iteration index increases to infinity, we emphasize that the iterative value function for any time index in the period is proven to converge to the optimal performance index function. Finally, numerical experiments and comparisons are presented to show the effectiveness of the iterative ADP algorithm.

II. Preliminaries and Assumptions

In this section, the energy storage system with renewable energy, i.e., solar energy, is described and the optimization objective of our research will be defined.

A. Notation

The list of used notations is reported as follows.

k, t	Time indices.
i, j	Iteration indices.
$\mathcal{F}_{b,k}$	Battery energy (kWh).
$\mathcal{F}_b^{\min}/\mathcal{F}_b^{\max}$	Minimum/maximum storage energy of the battery (kWh).
$\vartheta(\cdot)$	Charging/discharging efficiency of battery.
$\mathcal{T}_{\text{rate}}$	Rated charing/discharging power of the battery (kW).
$\mathcal{T}_{G,k}$	Power supply of the grid (kW).
$\mathcal{T}_{L,k}$	Power of the load (kW).
$\mathcal{T}_{R,k}$	Power of the renewable resource (kW).
$\mathcal{T}_{RL,k}$	Power from the renewable resource to the load (kW).
$\mathcal{T}_{RB,k}$	Power from the renewable resource to the battery (kW).
$\mathcal{T}_{GL,k}$	Power from the power grid to the load (kW).
$\mathcal{T}_{GB,k}$	Power from the power grid to the battery (kW).
$\mathcal{T}_{BL,k}$	Power from the battery to the load (kW).
λ	Periods of the load, electricity rate, and the renewable resource.
GHI_k	Global horizontal irradiance received on a horizontal surface (kWh/m^2).
ϑ_{pv}	the efficiency of the PV.
A_{pv}	the total area of the PV panel (m^2).
$\mathcal{C}_k$	Electricity rate (cents/kWh).
$\mathcal{F}_b^o$	Middle of storage limit (kWh).
α, β, δ	Given positive constants in performance index function.
γ	Discount factor.
x_k	System state.
u_k	Control input.
$\underline{u}_k$	Control law sequence from k to ∞.
$\mathcal{U}_k$	Control sequence in a period.
$\mathcal{U}(\cdot)$	Control law sequence in a period.
$F(\cdot)$, $\mathscr{F}(\cdot)$	System functions.
$\mathscr{L}(\cdot)$, $\mathscr{U}(\cdot)$	Utility functions.
$J(\cdot)$	Performance index function.
$V_i^j(\cdot)$	Iterative value function.
$\Psi(\cdot)$	Initial iterative value function.

B. Energy Storage System With Solar Renewable Energy

The energy storage system with solar renewable energy is described in [27], [28], which is shown in Fig. 1. It is composed of the power grid, the solar renewable energy (solar energy in brief) resource, the load demand (load in brief), and the battery system. In the energy storage system, the solar energy can (i) meet the load; (ii) charge the battery. The battery can (i) charge from the grid; (ii) charge from solar energy resource; (iii) discharge to meet the load; (iv) idle. There are three power flows to meet the load, including the power grid, the renewable resource, and the battery, where the load balance can be defined as

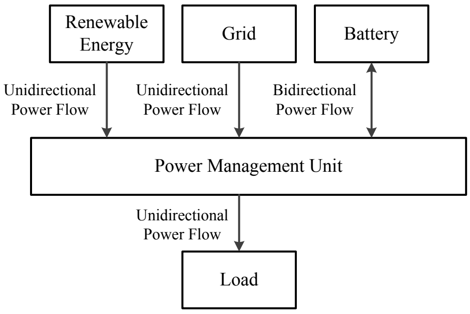


Fig. 1. Energy storage system with solar energy

$$\mathcal{T}_{L,k} = \mathcal{T}_{RL,k} + \mathcal{T}_{GL,k} + \mathcal{T}_{BL,k}. \tag{1}$$

The balance of the power grid can be defined as

$$\mathcal{T}_{G,k} = \mathcal{T}_{GL,k} + \mathcal{T}_{GB,k}. \tag{2}$$

In this paper, the renewable energy, i.e., solar energy, is considered. The power of the solar energy depends on position of the sun in the sky and hence the estimated total solar energy varies in each hour of a day. A typical solar panel characteristic is chosen as in [27], [28], [31]. The solar energy in first week of August 2014 in San Francisco is shown in Fig. 2(a). The average solar energy in a day is shown in Fig. 2(b). The power output for a photovoltaic (PV) panel [27], [32] at $k = 0, 1, \ldots$ can be expressed as $\mathcal{T}_{R,k} = \text{GHI}_k \cdot \vartheta_{\text{pv}} \cdot A_{\text{pv}}$, where ϑ_{pv} is included in the range $[0, 1]$. In this paper, as the solar energy

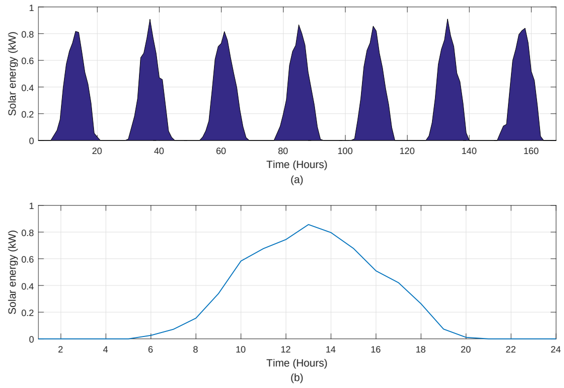


Fig. 2. Solar energy in San Francisco. (a) Solar energy in 168 hours. (b) Average solar energy in a day.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2017.2674581, IEEE Transactions on Industrial Electronics

can meet the load and charge the battery, the energy balance of the solar energy can be expressed as

$$\mathcal{T}_{R,k} = \mathcal{T}_{RL,k} + \mathcal{T}_{RB,k}. \tag{3}$$

C. Battery Model

The battery model used in this work is based on [25], [33], [34], where battery efficiency is considered to extend the battery's lifetime as far as possible. Under the situation that the battery cannot charge and discharge simultaneously, the battery model can be expressed as

$$\begin{aligned} \mathcal{F}_{b,k+1} =& \mathcal{F}_{b,k} - \mathcal{T}_{BL,k}(0.898 - 0.173\mathcal{T}_{BL,k}/\mathcal{T}_{\mathrm{rate}}) \\ & + (\mathcal{T}_{RB,k} + \mathcal{T}_{GB,k})(0.898 - 0.173(\mathcal{T}_{RB,k} \\ & + \mathcal{T}_{GB,k})/\mathcal{T}_{\mathrm{rate}}). \end{aligned} \tag{4}$$

In this paper, for the convenience of analysis, the battery self-discharge is not considered. The storage limits are defined as following

$$\mathcal{F}_b^{\min} \leq \mathcal{F}_{b,k} \leq \mathcal{F}_b^{\max}. \tag{5}$$

D. Assumptions and Optimization Objectives

For convenience of analysis, results of this paper are based on the following assumptions.

Assumption 1: The power flow from the energy storage system to the grid is not permitted.

Assumption 2: The electricity rate, the load, and the solar energy are periodic functions with the period $\lambda = 24$ hours.

From Assumption 1, we can get $\mathcal{T}_{BL,k} \geq 0$. According to Assumption 2, the parameters $\mathcal{C}_k$, $\mathcal{T}_{L,k}$, and $\mathcal{T}_{R,k}$ satisfy

$$\mathcal{C}_k = \mathcal{C}_{k+\lambda}, \mathcal{T}_{L,k} = \mathcal{T}_{L,k+\lambda}, \mathcal{T}_{R,k} = \mathcal{T}_{R,k+\lambda}. \tag{6}$$

Inspired by [30], the performance index function, which is expected to be minimized, is expressed as

$$\begin{aligned} \sum_{k=0}^{\infty} \gamma^k \Big(& \alpha(\mathcal{C}_k \mathcal{T}_{GL,k})^2 + \beta(\mathcal{F}_{b,k} - \mathcal{F}_b^o)^2 \\ & + \delta(\mathcal{T}_{RB,k} + \mathcal{T}_{GB,k} - \mathcal{T}_{BL,k})^2 \Big), \end{aligned} \tag{7}$$

where $\mathcal{F}_b^o = \frac{1}{2}(\mathcal{F}_b^{\min} + \mathcal{F}_b^{\max})$ is the middle of storage limit. The first term of the performance index function aims to minimize the total cost from the grid. The second term avoids fully charging/discharging of the battery and the third term is to prevent large charging/discharging power of the battery.

III. Iterative ADP Algorithm for Energy Storage Systems With Solar Energy

In this section, a novel iterative ADP algorithm is developed, which aims to maintain the energy storage systems at the optimum operating point under the solar energy.

A. System Establishment and Algorithm Derivations

As the solar energy is cost free, the energy of the renewable resource, first and foremost, is desired to meet the load and the rest of the energy is desired to charge the battery. Thus, we can derive that

$$\mathcal{T}_{RL,k} = \begin{cases} \mathcal{T}_{R,k}, & \mathcal{T}_{L,k} - \mathcal{T}_{R,k} \geq 0, \\ \mathcal{T}_{L,k}, & \mathcal{T}_{L,k} - \mathcal{T}_{R,k} < 0, \end{cases} \tag{8}$$

and

$$\mathcal{T}_{RB,k} = \begin{cases} 0, & \mathcal{T}_{L,k} - \mathcal{T}_{R,k} \geq 0, \\ \mathcal{T}_{R,k} - \mathcal{T}_{L,k}, & \mathcal{T}_{L,k} - \mathcal{T}_{R,k} < 0. \end{cases} \tag{9}$$

According to (2) and (9), the load balance (1) can be rewritten as $\mathscr{P}_{L,k} = \mathcal{T}_{G,k} + (\mathcal{T}_{BL,k} - \mathcal{T}_{GB,k})$. The renewable energy to charge the battery is $\mathscr{P}_{R,k} = \mathcal{T}_{R,k} - \mathcal{T}_{L,k}$ for $\mathcal{T}_{R,k} - \mathcal{T}_{L,k} \geq 0$, and $\mathscr{P}_{R,k} = 0$ for $\mathcal{T}_{R,k} - \mathcal{T}_{L,k} < 0$. For convenience of analysis, we introduce delays in $\mathscr{P}_{L,k}$, $\mathcal{T}_{BL,k}$ and $\mathcal{T}_{GB,k}$. Then, we can define the load balance as $\mathscr{P}_{L,k-1} = \mathcal{T}_{G,k} + \mathcal{T}_{BL,k-1} - \mathcal{T}_{GB,k-1}$. Let $x_{1,k} = \mathcal{T}_{G,k}$ and $x_{2,k} = \mathcal{F}_{b,k} - \mathcal{F}_b^o$ be the two system states. From the model of the battery, we know that the power flows $\mathcal{T}_{GB,k}$, and $\mathcal{T}_{BL,k}$ are nonnegative values, i.e., $\mathcal{T}_{GB,k} \geq 0$ and $\mathcal{T}_{BL,k} \geq 0$. As the battery is not permitted to charge and discharge simultaneously, we know $\mathcal{T}_{BL,k} = 0$, if $\mathcal{T}_{GB,k} \geq 0$ and $\mathcal{T}_{GB,k} = 0$, if $\mathcal{T}_{BL,k} \geq 0$. Then, we can define the control input as $u_k = \mathcal{T}_{BL,k} - \mathcal{T}_{GB,k}$. Letting $x_k = [x_{1,k}, x_{2,k}]^{\mathsf{T}}$, the equation of the energy storage system can be written as

$$\begin{aligned} x_{k+1} &= F(x_k, u_k, k) \\ &= \begin{pmatrix} \mathscr{P}_{L,k} - u_k \\ x_{2,k} - (u_k - \mathscr{P}_{R,k})\vartheta(u_k - \mathscr{P}_{R,k}) \end{pmatrix}, \end{aligned} \tag{10}$$

where $\vartheta(u_k - \mathscr{P}_{R,k}) = 0.898 - 0.173|u_k - \mathscr{P}_{R,k}|/\mathcal{T}_{\mathrm{rate}}$. Let $\underline{u}_k = (u_k, u_{k+1}, \ldots)$ denote the control sequence from k to ∞. Let $M_k = \begin{bmatrix} \alpha\mathcal{C}_k^2 & 0 \\ 0 & \beta \end{bmatrix}$ and let x_0 be the initial state. According to the definitions of the states and controls, the performance index function can be written as $J(x_0, \underline{u}_0, 0) = \sum\limits_{k=0}^{\infty} \gamma^k \mathscr{L}(x_k, u_k, k)$, and the optimal performance index function satisfies the following Bellman equation

$$J^*(x_k, k) = \inf_{u_k} \big\{ \mathscr{L}(x_k, u_k, k) + \gamma J^*(x_{k+1}, k+1) \big\}, \tag{11}$$

where $\mathscr{L}(x_k, u_k, k) = x_k^{\mathsf{T}} M_k x_k + \delta u_k^2$.

In the following, the detailed derivations of the iterative ADP algorithm is presented. First, according to (10), for $j = 0, 1, \ldots \lambda - 1$, we define a new system as

$$\begin{aligned} x_{k+1} &= \mathscr{F}(x_k, u_k, j) \\ &= \begin{pmatrix} \mathscr{P}_{L,\lambda-1-j} - u_k \\ x_{2,k} - (u_k - \mathscr{P}_{R,\lambda-1-j})\vartheta(u_k - \mathscr{P}_{R,\lambda-1-j}) \end{pmatrix}. \end{aligned} \tag{12}$$

For $j = 0, 1, \ldots, \lambda - 1$, we let

$$\mathscr{U}(x_k, u_k, j) = x_k^{\mathsf{T}} M_{\lambda-1-j} x_k + \delta u_k^2, \tag{13}$$

where $M_{\lambda-1-j} = \begin{bmatrix} \alpha(\mathcal{C}_{\lambda-1-j})^2 & 0 \\ 0 & \beta \end{bmatrix}$. Let the initial iterative value function $\Psi(x_k)$ be an arbitrary positive semi-definite function. For $i = 0, 1, \ldots$ and $j = 0, 1, \ldots, \lambda - 1$ be iteration indices. For $i = 0$ and $j = 0$, the initial iterative value function is defined as $V_0^0(x_k) = \Psi(x_k)$. Then, for $j = 0, 1, \ldots, \lambda - 1$, the iterative control law is obtained by

$$v_0^j(x_k) = \arg\min_{u_k} \{\mathscr{U}(x_k, u_k, j) + \gamma V_0^j(x_{k+1})\}, \tag{14}$$

where $x_{k+1} = \mathscr{F}(x_k, u_k, j)$ is defined in (12) and the utility function $\mathscr{U}(x_k, u_k, j)$ is defined in (13). According to the iterative control law $v_0^j(x_k)$, we can update the iterative value function by

$$V_0^{j+1}(x_k) = \mathscr{U}(x_k, v_0^j(x_k), j) + \gamma V_0^j(\mathscr{F}(x_k, v_0^j(x_k), j)). \tag{15}$$

For $i = 1, 2, \ldots$, we let $V_i^0(x_k) = V_{i-1}^{\lambda}(x_k)$. Then, for $i = 1, 2, \ldots$, $j = 0, 1, \ldots, \lambda - 1$, the iterative control law is obtained by

$$v_i^j(x_k) = \arg\min_{u_k} \{\mathscr{U}(x_k, u_k, j) + \gamma V_i^j(x_{k+1})\}, \tag{16}$$

and the iterative value function is updated by

$$V_i^{j+1}(x_k) = \mathscr{U}(x_k, v_i^j(x_k), j) + \gamma V_i^j(\mathscr{F}(x_k, v_i^j(x_k), j)). \tag{17}$$

Then, for any $i = 0, 1, \ldots$ and $j = 0, 1, \ldots, \lambda - 1$, we can construct the periodic iterative control law sequence by

$$\mathcal{U}_i^j(x_k) = \Big\{ v_i^j(x_k), v_i^{j-1}(x_k), \ldots, v_i^0(x_k), v_i^{\lambda-1}(x_k), v_i^{\lambda-2}(x_k), \ldots, v_i^{j+1}(x_k) \Big\}. \tag{18}$$

Remark 1: The developed iterative ADP algorithm (14)–(17) possesses inherent differences comparing with the iterative Q-learning algorithm in [30]. First, in [30], without considering the solar energy, the iterative Q-learning algorithm was developed to obtain the optimal battery control for the energy storage system. In this paper, the solar energy is clearly considered in the iterative ADP algorithm (14)–(17), which displays a more complex system. Second, in [30], the iterative Q function, which contains both state and control information, was constructed. In this paper, the iterative value function $V_i^j(x_k)$ is only the function of state. Hence, the computation quantity for updating the iterative value function is smaller than the iterative Q function in [30]. Third, in [30], it was required that the time index satisfied $k \in \{0, \lambda, 2\lambda, \ldots\}$ to construct the Q-learning algorithm, while in this paper, the time index is $k = 0, 1, \ldots$. Furthermore, in [30], the properties of the iterative Q function for $k \in \{0, \lambda, 2\lambda, \ldots\}$ were analyzed, which lacked analyzing the properties for other time index. Thus, we say that the analysis in [30] is not complete and may not be applicable. For the developed iterative ADP algorithm in this paper, a new analysis method will be presented.

B. Property Analysis

In this subsection, the properties of the iterative ADP algorithm (14)–(17) will be analyzed. It will be shown that the iterative value function will converge to the optimum as the iteration index increases to infinity. First, we will analyze the property of the optimal performance index function. It will be shown that the optimal performance index function is a periodic function under the assumptions in this paper.

Theorem 1: If Assumptions 1–2 hold, then for any state x_k, $k = 0, 1, \ldots$, the optimal performance index function $J^*(x_k, k)$ satisfies

$$J^*(x_k, k) = J^*(x_k, k + \lambda). \tag{19}$$

The proof of Theorem 1 is shown in Appendix I. From Theorem 1, we know that the optimal performance index function is a periodic function with the period $\lambda = 24$. Next, the convergence of the iterative value function will be analyzed. It will be proven that the iterative value function converges to the optimum as the iteration index increases to infinity.

Theorem 2: For $i = 0, 1, \ldots$ and $j = 0, 1, \ldots, \lambda - 1$, let $V_i^{j+1}(x_k)$ and $v_i^j(x_k)$ be obtained by (14)–(17). If Assumptions 1–2 holds, then for any $j = 0, 1, \ldots, \lambda - 1$, the iterative value function $V_i^{j+1}(x_k)$ converges to its optimal performance index function as $i \to \infty$, which satisfies

$$\lim_{i \to \infty} V_i^{j+1}(x_k) = J^*(x_k, \lambda - j - 1). \tag{20}$$

The proof of Theorem 2 is shown in Appendix II.

IV. Numerical Experiments

In this section, numerical experiments and comparisons will be displayed to show the performance of the iterative ADP algorithm. The profiles of the real-time electricity rate and the load are chosen from ComEd Company in [35] and NAHB Research report in [36], respectively. The trajectories of the electricity rate and the load in 168 hours (one week) are shown in Figs. 3(a) and (c), respectively. The average trajectories of the electricity rate and the load are shown in Figs. 3(b) and (d), respectively. In this paper, the average electricity rate, average load, and average solar energy are used as the periodic functions with the period $\lambda = 24$ to implement the iterative ADP algorithm.

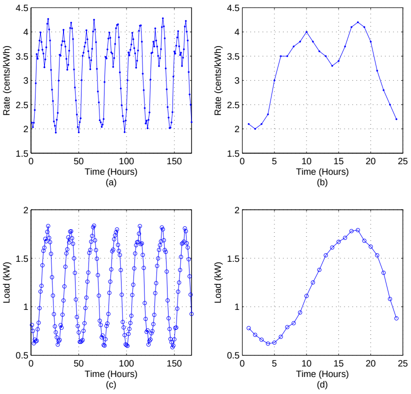


Fig. 3. Electricity rate and load power. (a) Electricity rate for 168 hours. (b) Average electricity rate. (c) Load for 168 hours. (b) Average load.

Choose the capacity of the battery as 16 kWh. The rated power of the battery is 3 kW. Let the lower and upper storage limits of the battery be $E_b^{\min} = 2$ kWh and $E_b^{\max} = 14$ kWh, respectively. Let $\gamma = 0.95$. The initial level of the battery is 9 kWh. Let the performance index function be expressed as in (7), where we set $\alpha = 1$, $\beta = 0.3$ and

$\delta = 0.2$. Choose the initial function as $\Psi(x_k) = x_k^{\mathsf{T}} P x_k$, where $P = [2.05,\ 0.11; 0.11,\ 8.07]$. Let the initial state be $x_0 = [1, 9]^{\mathsf{T}}$. After normalizing the data of the electricity rate, the load, and the solar energy [26], we implement the developed iterative ADP algorithm for $i = 15$ iterations to make the iterative value function convergent. The simulation plots of the iterative value function $V_i^j(x_k)$ for $i = 0, 1, \ldots, 15$ and $j = \lambda - 1$ are shown in Fig. 4. The trajectory of the iterative value function $V_i^j(x_k)$ at x_0 is shown in Fig. 5. From the simulation results, for any $j = 0, 1, \ldots, \lambda - 1$, the iterative value function is convergent after 15 iterations, which justifies the correctness of the developed algorithm.

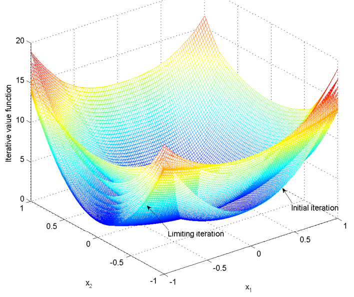


Fig. 4. The plots of the iterative value function $V_i^j(x_k)$ for $i = 0, 1, \ldots$ and $j = \lambda - 1$

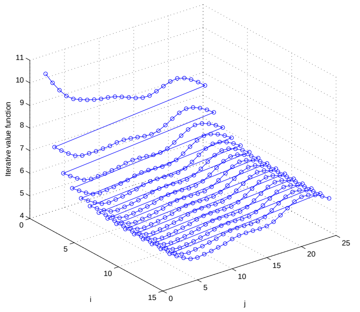


Fig. 5. The trajectory of the iterative value function $V_i^j(x_k)$ at x_0

After $i = 15$ iterations, the optimal battery control law for the energy storage system is achieved. The optimal battery control in 168 hours is shown in Fig. 6. The plot of the optimal battery energy is shown in Fig. 7. From Figs. 6 and 7, for different time indices, the optimal battery control law is different and the optimal charging/discharging power of the battery at each hour is obtained.

To show the superiority of the developed iterative ADP algorithm, three traditional optimal control methods, including time-based Q-learning (TBQL) algorithm [25], [26], particle swarm optimization (PSO) algorithm [28] and model predictive control (MPC) algorithm [37], [38], are employed for numerical comparisons. In the TBQL algorithm [25], [26], for $k = 1, 2, \ldots$, the iterative control law is designed to satisfy the following optimality equation

$$Q(x_{k-1}, u_{k-1}, k-1) = U(x_k, u_k, k) + Q(x_k, u_k, k). \quad (21)$$

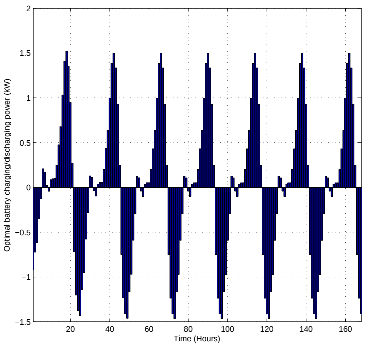


Fig. 6. Optimal battery control with solar energy

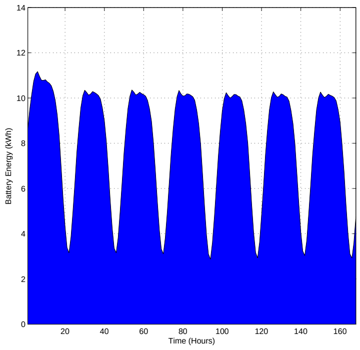


Fig. 7. Optimal battery energy in 168 hours

We implement the TBQL for 150 time steps, which makes the Q function converge.

In PSO algorithm [28], the movement of each particle naturally evolves to an optimal or near-optimal solution. Let $\mathcal{G}$ be the swarm size. The position of each particle is represented by x_k^ℓ, $\ell = 1, 2, \ldots, \mathcal{G}$ and its movement by the velocity vector υ_k^ℓ. Then, the update rule of PSO can be expressed as

$$\begin{aligned} x_k^\ell &= x_{k-1}^\ell + \nu_k^\ell \\ \nu_k^\ell &= \omega \nu_{k-1}^\ell + rand_1 \rho_1^{\mathsf{T}}(p^\ell - x_{k-1}^\ell) + rand_2 \rho_2^{\mathsf{T}}(p_g - x_{k-1}^\ell). \end{aligned} \quad (22)$$

Choose the swarm size $\mathcal{G} = 30$. Let the inertia factor be $\omega = 0.7$. Let the correction factors $\rho_1 = \rho_2 = [1, 1]^T$. Let $rand_1$ and $rand_2$ be random numbers in $[0, 1]$. Let p^ℓ be the best position of particles and let p_g be the global best position. We implement the PSO algorithm for 120 iterations, which makes the performance index function minimized.

Within the MPC framework [37], [38], it is to solve the optimization problem at the current time step $k = 0, 1, \ldots$. In the MPC, the receding horizon control algorithm [38] is employed to obtain the optimal battery control law. For $k = 0, 1, \ldots$, the finite horizon value function is expressed as

$$V(x_k, k) = \sum_{\tau=k}^{k+T} \gamma^\tau \mathscr{L}(x_\tau, u_\tau, \tau), \quad (23)$$

where we let $\gamma = 0.95$ and $T = 20$. In each receding horizon, the shooting method and sequential quadratic programming [39] are used to obtain the current optimal control law, where the terminal value function is set as $V(x_{k+T}, k+T) = 0$.

Based on the TBQL, PSO, MPC and the iterative ADP algorithm, the comparisons of the power supply from the grid is shown in Fig. 8(a). The comparisons of the battery power supply is shown in Fig. 8(b). The battery charging power and the real-time cost comparisons for TBQL, PSO, MPC and the iterative ADP algorithm are shown in Figs. 8(c) and (d), respectively.

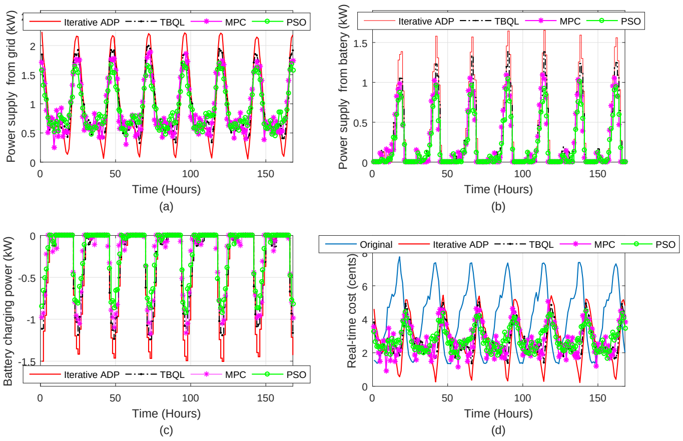


Fig. 8. Simulation comparisons. (a) Power supply from the grid. (b) Power supply from the battery. (c) Charging power of the battery. (b) Real-time cost.

From Figs. 8(a)–(d), when the electricity rate and the load demand are low, comparing with TBQL, PSO and MPC algorithms, the iterative ADP algorithm reaches the maximum power supply from the grid to charge the battery. When the electricity rate and the load demand are high, the iterative ADP algorithm supplies the maximum battery power to meet the load demand and thus, iterative ADP algorithm reaches the minimum real-time cost comparing with TBQL, PSO and MPC algorithms. On the other hand, when the solar energy is high, it directly meet the load demands, which prevents the power from the grid and the battery. Using the iterative ADP algorithm, the grid and battery supply powers are the minimium when the solar energy is high, comparing with TBQL, PSO and MPC algorithms. Thus, using the iterative ADP algorithm, the minimum cost can be achieved. T total cost comparisons of the TBQL, PSO, MPC, and t iterative ADP algorithm are shown in Table I, which verifi the superiority of the developed ADP algorithm.

TABLE I
COST COMPARISON

	Original	TBQL	PSO	MPC	Iterative AI
Total cost (cents) in San Francisco	682.46	454.32	477.99	467.93	441.27
Saving (%)		33.48	30.01	31.49	35.43
Total cost (cents) in Boston		459.19	474.14	465.34	438.95
Saving (%)		32.72	30.52	31.81	35.68

Now, we choose solar energy in first week of August 2014 in Boston [27], [28] to verify the effectiveness of the iterative ADP algorithm. The solar energy in a week is shown in Fig. 9(a) and the average one is displayed in Fig. 9(b). Let the rate and load data be the same as in Figs. 3(a) and (c), respectively. Implementing the iterative ADP algorithm for 15 iterations, such that the iterative value function converges to the optimum. The plots of the optimal battery control and optimal battery energy are shown in Fig. 9(c) and (d), respectively, where the optimal battery control law can be achieved. TBQL, PSO, and MPC methods are also employed for comparisons. Based on the new solar energy, the comparisons of the power supply from the grid is shown in Fig. 10(a). The comparisons of the battery power supply is shown in Fig. 10(b). The battery charging power and the real-time cost comparisons for TBQL, PSO, MPC and the iterative ADP algorithm are shown in Figs. 10(c) and (d), respectively. It can be seen that when the electricity rate and the load demand are low, the iterative ADP algorithm reaches the maximum power supply to charge the battery comparing with TBQL, PSO and MPC algorithms. The grid and battery supply powers are the minimum using the iterative ADP algorithm, when the solar energy is high. The total cost comparisons of the TBQL, PSO, MPC, and the iterative ADP algorithm under the new solar energy are shown in Table I, which can verify the superiority of the developed ADP algorithm.

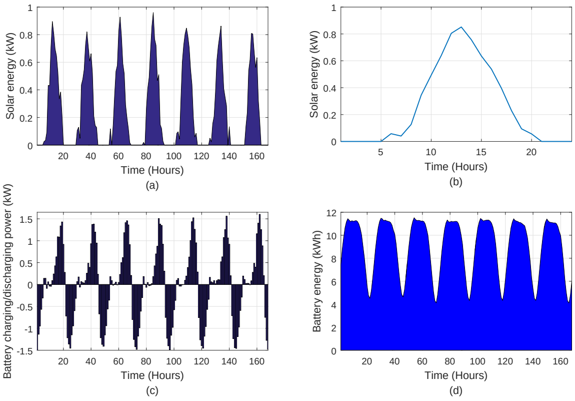


Fig. 9. Solar energy in Boston and control results. (a) Solar energy in Boston. (b) Average solar energy. (c) Optimal battery control. (d) Optimal battery energy.

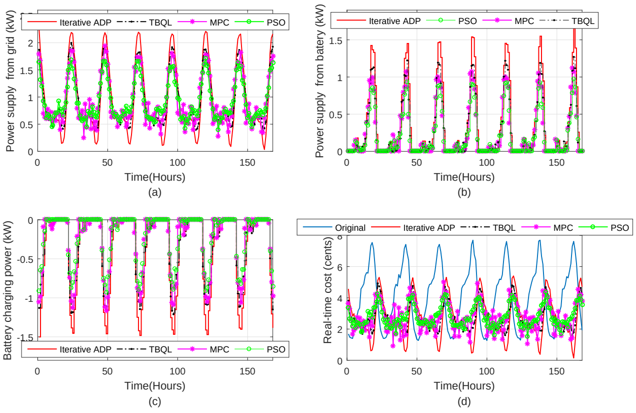


Fig. 10. Simulation results. (a) Optimal battery control by MPC. (b) Battery energy by MPC. (c) Real-time cost comparison. (d) Trajectory of $V_i^j(x_k)$ at x_0 without solar energy.

Next, we choose new simulation data to justify the correctness of the developed iterative ADP algorithm. The new solar energy in 168 hours is shown in Fig. 11(a), which is double power of the one in Fig. 9(a) with modifications. The average solar energy is shown in Fig. 11(d). The new electricity rate [40] and the load [25] are shown in Figs. 12(a) and (c), respectively. The average electricity rate and load are shown in Figs. 12(b) and (d), respectively. Define the parameters of the battery as in [25], where the capacity of the battery is defined as 100 kWh and the rated power of the battery is 16 kW. Let the upper and lower storage limits of the battery be $\digamma_b^{\min} = 20$ kWh and $\digamma_b^{\max} = 80$ kWh, respectively. Implementing the iterative ADP algorithm for 20 iterations based on the new data, the trajectory of the iterative value function is shown in Fig. 13, where the iterative value function is convergent to the optimum. The optimal battery control is shown in Fig. 14, where the battery charges in the hours that electricity rate, the load are low and the solar energy is high. The battery discharges in the hours that electricity rate and the load are high and the solar energy is low. Thus, the correctness of the developed iterative ADP algorithm can be verified.

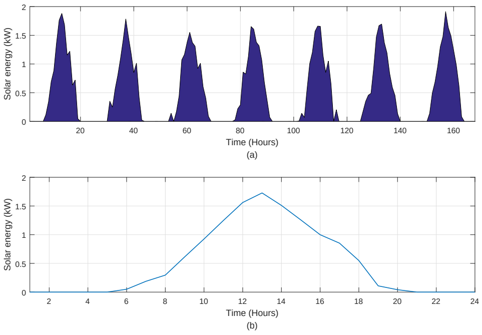


Fig. 11. Simulation results. (a) Optimal battery control without solar energy. (b) Optimal battery energy without solar energy. (c) Solar energy in 168 hours. (d) Average solar energy.

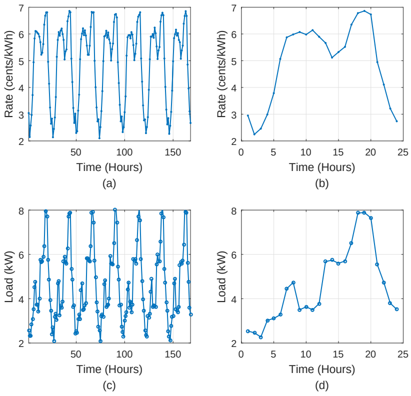


Fig. 12. Electricity rate and load. (a) Electricity rate for 168 hours. (b) Average electricity rate. (c) Load for 168 hours. (d) Average load.

V. Conclusions

In this paper, a new optimal battery control scheme is obtained for the energy storage systems with solar energy via an effective iterative ADP algorithm. The present iterative ADP algorithm is initialized by an arbitrary positive semi-definite function. In each iteration $i = 0, 1, \ldots$, the iterative control law sequence is obtained, instead of obtaining a single iterative control law. According to the data of the electricity rate, the load and the solar energy, it is proven that the iterative value function converges to the corresponding optimal performance index function as the iterative index increases to infinity. Finally, numerical experiments and comparisons are shown to justify the effectiveness of the developed iterative ADP algorithm.

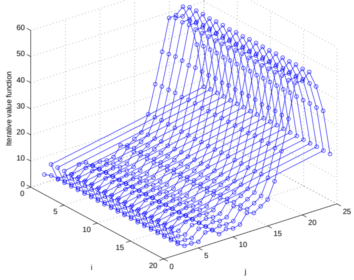


Fig. 13. The trajectory of the iterative value function $V_i^j(x_k)$ at x_0

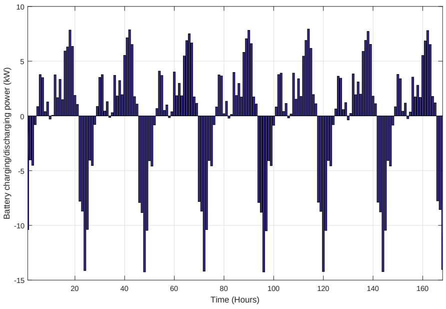


Fig. 14. Optimal control of the battery in 168 hours

Appendix I
Proof of Theorem 1

For any $k = 0, 1, \ldots$, let $\underline{u}_k^* = (u_k^*, u_{k+1}^*, \ldots)$ be the optimal control sequence, which is expressed as $\underline{u}_k^* = \arg\inf_{\underline{u}_k} \left\{ \sum_{t=k}^{\infty} \gamma^{t-k} \mathscr{L}(x_t, u_t, t) \right\}$. The optimal performance index function $J^*(x_k, k)$ in (11) can be expressed as

$$\begin{aligned} J^*(x_k, k) &= \inf_{\underline{u}_k} \left\{ \sum_{t=k}^{\infty} \gamma^{t-k} \mathscr{L}(x_t, u_t, t) \right\} \\ &= \sum_{t=k}^{\infty} \gamma^{t-k} \left(x_t^{*\mathsf{T}} \begin{bmatrix} \alpha(\mathcal{C}_t)^2 & 0 \\ 0 & \beta \end{bmatrix} x_t^* + \delta u_t^{*\mathsf{T}} u_t^* \right), \end{aligned} \tag{24}$$

where the system state x_{k+p}^*, $p = 0, 1, \ldots$, satisfies

$$\begin{aligned} & x_{k+p+1}^* \\ &= F(x_{k+p}^*, u_{k+p}^*, k+p) \\ &= \begin{pmatrix} \mathscr{P}_{L,k+p} - u_{k+p}^* \\ x_{2,k+p}^* - (u_{k+p}^* - \mathscr{P}_{R,k+p}) \vartheta(u_{k+p}^* - \mathscr{P}_{R,k+p}) \end{pmatrix}, \end{aligned} \tag{25}$$

and $x_k^* = x_k$. According to (8)–(10) and Assumption 2, we can derive $\mathscr{P}_{L,k+p} = \mathscr{P}_{L,k+p+\lambda}$, and $\mathscr{P}_{R,k+p} = \mathscr{P}_{R,k+p+\lambda}$.

On the other hand, the performance index function $J(x_k, k+\lambda)$ can be expressed as

$$\begin{aligned} J(x_k,&k+\lambda) \\ =&\sum_{t=k}^{\infty}\gamma^{t-k}\left(x_t^T\left[\begin{array}{cc}\alpha(\mathcal{C}_{t+\lambda})^2 & 0\\ 0 & \beta\end{array}\right]x_t+\delta u_t^Tu_t\right), \end{aligned} \tag{26}$$

where for $p = 0, 1, \ldots$, the state x_{k+p+1} satisfies

$$\begin{aligned} &x_{k+p+1} \\ &= F(x_{k+p}, u_{k+p}, k+\lambda+p) \\ &= \begin{pmatrix} \mathscr{P}_{L,k+p+\lambda} - u_{k+p} \\ x_{2,k+p} - (u_{k+p} - \mathscr{P}_{R,k+p+\lambda})\vartheta(u_{k+p} - \mathscr{P}_{R,k+p+\lambda}) \end{pmatrix}. \end{aligned} \tag{27}$$

Substituting $\underline{u}_k^*$ into (27), for $p = 0, 1, \ldots$, we can get

$$\begin{aligned} &x_{k+p+1} \\ &= \begin{pmatrix} \mathscr{P}_{L,k+p+\lambda} - u_{k+p}^* \\ x_{2,k+p}^* - (u_{k+p}^* - \mathscr{P}_{R,k+p+\lambda})\vartheta(u_{k+p}^* - \mathscr{P}_{R,k+p+\lambda}) \end{pmatrix} \\ &= \begin{pmatrix} \mathscr{P}_{L,k+p} - u_{k+p}^* \\ x_{2,k+p}^* - (u_{k+p}^* - \mathscr{P}_{R,k+p})\vartheta(u_{k+p}^* - \mathscr{P}_{R,k+p}) \end{pmatrix} \\ &= F(x_{k+p}^*, u_{k+p}^*, k+p) \\ &= x_{k+p+1}^*. \end{aligned} \tag{28}$$

As the optimal performance index function $J^*(x_k, k+\lambda)$ can be expressed as

$$J^*(x_k, k+\lambda) = \inf_{\underline{u}_k}\left\{\sum_{t=k}^{\infty}\gamma^{t-k}\mathscr{L}(x_t, u_t, t+\lambda)\right\}, \tag{29}$$

according to the definition of $\underline{u}_k^*$, we can get $\underline{u}_k^* = \arg\inf_{\underline{u}_k}\left\{\sum\limits_{t=k}^{\infty}\gamma^{t-k}\mathscr{L}(x_t, u_t, t)\right\} = \arg\inf_{\underline{u}_k}\left\{\sum\limits_{t=k}^{\infty}\gamma^{t-k}\mathscr{L}(x_t, u_t, t+\lambda)\right\}$, where for $p = 0, 1, \ldots$, $x_{k+p+1} = F(x_{k+p}, u_{k+p}, k+\lambda+p) = F(x_{k+p}, u_{k+p}, k+p)$. Thus, we know that the control sequence $\underline{u}_k^*$ is the optimal control sequence for the performance index function (26). Hence we can obtain

$$\begin{aligned} &J^*(x_k, k+\lambda) \\ &=\inf_{\underline{u}_k}\left\{\sum_{t=k}^{\infty}\gamma^{t-k}\left(x_t^{\mathsf{T}}\left[\begin{array}{cc}\alpha(\mathcal{C}_{t+\lambda})^2 & 0\\ 0 & \beta\end{array}\right]x_t+\delta u_t^{\mathsf{T}}u_t\right)\right\} \\ &=\inf_{\underline{u}_k}\left\{\sum_{t=k}^{\infty}\gamma^{t-k}\left(x_t^{\mathsf{T}}\left[\begin{array}{cc}\alpha(\mathcal{C}_{t})^2 & 0\\ 0 & \beta\end{array}\right]x_t+\delta u_t^{\mathsf{T}}u_t\right)\right\} \\ &=J^*(x_k, k). \end{aligned} \tag{30}$$

The proof is complete.

Appendix II
Proof of Theorem 2

According to Assumption 2, $\mathcal{C}_k$, $\mathcal{T}_{L,k}$, and $\mathcal{T}_{R,k}$ satisfy (6). According to (16) and (17), for $i = 1, 2, \ldots$ and $j = 0, 1, \ldots, \lambda-1$, we can derive the equation (31) in the next page, where $\underline{u}_k^{k+\lambda-1} = (u_k, u_{k+1}, \ldots, u_{k+\lambda-1})$.

From (31), the utility function $\sum\limits_{t=0}^{\lambda-1}\gamma^t\mathscr{L}(x_{k+t}, u_{k+t}, \lambda-j-1+t)$ is independent with i. For $j = 0, 1, \ldots, \lambda-1$, we let $\Pi(x_k, \mathcal{U}_k, j) = \sum\limits_{t=0}^{j}\gamma^t\mathscr{L}(x_{k+t}, u_{k+t}, j-t) + \sum\limits_{t=0}^{\lambda-j-1}\gamma^{t+j}\mathscr{L}(x_{k+j+t}, u_{k+j+t}, \lambda-t)$ and $\Gamma(x_k, \mathcal{U}_k, j) = \sum\limits_{t=0}^{\lambda-1}\gamma^t\mathscr{L}(x_{k+t}, u_{k+t}, \lambda-j-1+t)$, where we let $\mathcal{U}_k = \underline{u}_k^{k+\lambda-1}$. Then, equation (31) can be written as

$$\begin{aligned} V_{i+1}^{j+1}(x_k) &= \min_{\mathcal{U}_k}\{\Pi(x_k, \mathcal{U}_k, j) + \tilde{\gamma}V_i^{j+1}(x_{k+\lambda})\} \\ &= \min_{\mathcal{U}_k}\{\Gamma(x_k, \mathcal{U}_k, j) + \tilde{\gamma}V_i^{j+1}(x_{k+\lambda})\}, \end{aligned} \tag{32}$$

where $\tilde{\gamma} = \gamma^\lambda$. From (32), we can derive that

$$\begin{aligned} V_{i+1}^{j+1}(x_k) =&\min_{\mathcal{U}_k}\{\Gamma(x_k, \mathcal{U}_k, j) + \tilde{\gamma}V_i^{j+1}(x_{k+\lambda})\} \\ =&\min_{\mathcal{U}_k}\{\Gamma(x_k, \mathcal{U}_k, j) + \min_{\mathcal{U}_{k+\lambda}}\{\Gamma(x_{k+\lambda}, \mathcal{U}_{k+\lambda}, j+\lambda) \\ &+\cdots+\min_{\mathcal{U}_{k+i\lambda}}\{\Gamma(x_{k+i\lambda}, \mathcal{U}_{k+i\lambda}, j+i\lambda) \\ &+\tilde{\gamma}^{i+1}V_0^{j+1}(x_{k+(i+1)\lambda})\}\cdots\}\} \\ =&\min_{\underline{\mathcal{U}}_k^{k+i\lambda}}\left\{\sum_{\tau=1}^{i}\Gamma(x_{k+\tau\lambda}, \mathcal{U}_{k+\tau\lambda}, j) \right. \\ &\left. +\tilde{\gamma}^{i+1}V_0^{j+1}(x_{k+(i+1)\lambda})\right\}, \end{aligned} \tag{33}$$

where $\underline{\mathcal{U}}_k^{k+i\lambda} = \big(\mathcal{U}_k, \mathcal{U}_{k+1}, \ldots, \mathcal{U}_{k+i\lambda}\big)$. On the other hand, according to Theorem 1, we can obtain

$$\begin{aligned} &J^*(x_k, \lambda-j-1) \\ &\quad=\min_{u_k}\Big\{\mathscr{L}(x_k, u_k, \lambda-j-1) \\ &\qquad+\gamma\min_{u_{k+1}}\Big\{\mathscr{L}(x_{k+1}, u_{k+1}, \lambda-j)+\cdots \\ &\qquad+\gamma\min_{u_{k+\lambda-1}}\Big\{\mathscr{L}(x_{k+\lambda-1}, u_{k+\lambda-1}, 2\lambda-j-2) \\ &\qquad+\gamma J^*(x_{k+\lambda}, 2\lambda-j-1)\Big\}\cdots\Big\}\Big\} \\ &\quad=\min_{\mathcal{U}_k}\{\Gamma(x_k, \mathcal{U}_k, j) + \tilde{\gamma}J^*(x_{k+\lambda}, 2\lambda-j-1)\} \\ &\quad=\min_{\mathcal{U}_k}\{\Gamma(x_k, \mathcal{U}_k, j) + \tilde{\gamma}J^*(x_{k+\lambda}, \lambda-j-1)\} \\ &\quad=\min_{\underline{\mathcal{U}}_k^{k+i\lambda}}\left\{\sum_{\tau=1}^{i}\Gamma(x_{k+\tau\lambda}, \mathcal{U}_{k+\tau\lambda}, j)\right. \\ &\qquad\left.+\tilde{\gamma}^{i+1}J^*(x_{k+(i+1)\lambda}, \lambda-j-1)\right\}. \end{aligned} \tag{34}$$

From (33) and (34), it is desired that for any $j = 0, 1, \ldots, \lambda-1$, the iterative value function $V_{i+1}^{j+1}(x_k)$ converges to the optimal performance index function $J^*(x_k, \lambda-j-1)$ as the iteration index $i \to \infty$. As initial value function $\Psi(x_k)$ is finite for any state x_k, we can derive $V_0^{j+1}(x_k)$ is finite for any $j = 0, 1, \ldots, \lambda-1$. Then, for functions $J^*(x_k, \lambda-j-1)$, $\Gamma(x_k, \mathcal{U}_k, j)$, and $V_0^{j+1}(x_k)$, inspired by [41], [42], it is assumed that $\underline{\zeta}$, $\overline{\zeta}$, $\underline{\sigma}$ and $\overline{\sigma}$ are constants that satisfy $\underline{\zeta}\,\Gamma(x_k, \mathcal{U}_k, j) \leq \tilde{\gamma}J^*(x_{k+\lambda}, \lambda-j-1) \leq \overline{\zeta}\,\Gamma(x_k, \mathcal{U}_k, j)$, $\underline{\sigma}J^*(x_k, \lambda-1-j) \leq V_0^{j+1}(x_k) \leq \overline{\sigma}J^*(x_k, \lambda-1-j)$, where

$$\begin{aligned}
&V_i^{j+1}(x_k)\\
&\quad=\min_{u_k}\Big\{\mathscr{U}(x_k,u_k,j)+\gamma V_i^j(x_{k+1})\Big\}\\
&\quad=\min_{u_k}\Big\{\mathscr{U}(x_k,u_k,j)+\gamma\min_{u_{k+1}}\Big\{\mathscr{U}(x_{k+1},u_{k+1},j-1)+\cdots+\gamma\min_{u_{k+j}}\Big\{\mathscr{U}(x_{k+j},u_{k+j},0)+\gamma V_i^0(x_{k+j})\Big\}\cdots\Big\}\Big\}\\
&\quad=\min_{u_k}\Big\{\mathscr{U}(x_k,u_k,j)+\gamma\min_{u_{k+1}}\Big\{\mathscr{U}(x_{k+1},u_{k+1},j-1)+\cdots+\gamma\min_{u_{k+j}}\Big\{\mathscr{U}(x_{k+j},u_{k+j},0)+\gamma V_{i-1}^{\lambda}(x_{k+j})\Big\}\cdots\Big\}\Big\}\\
&\quad=\min_{u_k}\Big\{\mathscr{U}(x_k,u_k,j)+\gamma\min_{u_{k+1}}\Big\{\mathscr{U}(x_{k+1},u_{k+1},j-1)+\cdots+\gamma\min_{u_{k+j}}\Big\{\mathscr{U}(x_{k+j},u_{k+j},0)\\
&\qquad+\gamma\min_{u_{k+j+1}}\Big\{\mathscr{U}(x_{k+j},u_{k+j},\lambda-1)+\cdots+\min_{u_{k+\lambda-1}}\Big\{\mathscr{U}(x_{k+\lambda-1},u_{k+\lambda-1},j+1)+\gamma V_{i-1}^{j+1}(x_{k+\lambda})\Big\}\cdots\Big\}\Big\}\cdots\Big\}\Big\}\\
&\quad=\min_{u_k}\Big\{\mathscr{L}(x_k,u_k,\lambda-j-1)+\gamma\min_{u_{k+1}}\Big\{\mathscr{L}(x_{k+1},u_{k+1},\lambda-j)+\cdots+\gamma\min_{u_{k+j}}\Big\{\mathscr{L}(x_{k+j},u_{k+j},\lambda-1)\\
&\qquad+\gamma\min_{u_{k+j+1}}\Big\{\mathscr{L}(x_{k+j},u_{k+j},\lambda)+\cdots+\min_{u_{k+\lambda-1}}\Big\{\mathscr{L}(x_{k+\lambda-1},u_{k+\lambda-1},2\lambda-j-2)+\gamma V_{i-1}^{j+1}(x_{k+\lambda})\Big\}\cdots\Big\}\Big\}\cdots\Big\}\Big\}\\
&\quad=\min_{\underline{u}_k^{k+\lambda-1}}\Big\{\sum_{t=0}^{\lambda-1}\gamma^t\mathscr{L}(x_{k+t},u_{k+t},\lambda-j-1+t)+\gamma^{\lambda}V_{i-1}^{j+1}(x_{k+\lambda})\Big\}. \qquad (31)
\end{aligned}$$

$0<\underline{\zeta}\le\overline{\zeta}<\infty$ and $0\le\underline{\sigma}\le 1\le\overline{\sigma}<\infty$. Let $i=1$. For any $j=0,1,\ldots,\lambda-1$, according to the idea in [41], [42], we can obtain

$$\begin{aligned}
V_1^{j+1}(x_k)&=\min_{\mathcal{U}_k}\big\{\Gamma(x_k,\mathcal{U}_k,j)+\tilde{\gamma}V_0^{j+1}(x_{k+\lambda})\big\}\\
&\le\min_{\mathcal{U}_k}\big\{\Gamma(x_k,\mathcal{U}_k,j)+\tilde{\gamma}\overline{\sigma}J^*(x_{k+\lambda},\lambda-j-1)\big\}\\
&\le\min_{\mathcal{U}_k}\Big\{\Big(1+\frac{\overline{\sigma}-1}{(1+\overline{\zeta})}\overline{\zeta}\Big)\Gamma(x_k,\mathcal{U}_k,j)\\
&\qquad+\Big(\overline{\sigma}-\frac{\overline{\sigma}-1}{(1+\overline{\zeta})}\Big)\tilde{\gamma}J^*(x_{k+\lambda},\lambda-j-1)\Big)\Big\}\\
&=\Big(1+\frac{\overline{\sigma}-1}{(1+\overline{\zeta}^{-1})}\Big)J^*(x_k,\lambda-j-1). \qquad (35)
\end{aligned}$$

According to mathematical induction, we can obtain

$$\begin{aligned}
&\left(1+\frac{\underline{\sigma}-1}{(1+\underline{\zeta}^{-1})^i}\right)J^*(x_k,\lambda-j-1)\le V_i^{j+1}(x_k)\\
&\qquad\le\left(1+\frac{\overline{\sigma}-1}{(1+\overline{\zeta}^{-1})^i}\right)J^*(x_k,\lambda-j-1). \qquad (36)
\end{aligned}$$

Letting $i\to\infty$, we can obtain (20), which completes the proof.

References

[1] C. S. Lai and M. D. McCulloch, "Sizing of stand-alone solar PV and storage system with anaerobic digestion biogas power plants," *IEEE Transactions on Industrial Electronics*, article in press, 2016. DOI: 10.1109/TIE.2016.2625781

[2] H. Liu, P. C. Loh, X. Wang, Y. Yang, W. Wang, and D. Xu, "Droop control with improved disturbance adaption for a PV system with two power conversion stages," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 10, pp. 6073–6085, Oct. 2016.

[3] E. Chatzinikolaou and D. J. Rogers, "Cell SoC balancing using a cascaded full-bridge multilevel converter in battery energy storage systems," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 9, pp. 5394–5402, Sep. 2016.

[4] Q. Shafiee, C. Stefanovic, T. Dragicevic, P. Popovski, J. C. Vasquez, and J. M. Guerrero, "Robust networked control scheme for distributed secondary control of islanded microgrids," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 10, pp. 5363–5374, Oct. 2014.

[5] J. Appen, T. Stetz, M. Braun, and A. Schmiegel, "Local voltage control strategies for PV storage systems in distribution grids," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 1002–1009, Mar. 2014.

[6] S. Park, J. Lee, S. Bae, G. Hwang, and J. Choi, "Contribution based energy trading mechanism in micro-grids for future smart grid: A game theoretic approach," *IEEE Transactions on Industrial Electronics*, article in press, 2016. DOI: 10.1109/TIE.2016.2532842

[7] Y. Lee, W. Hsiao, C. Huang, and S. T. Chou, "An integrated cloud-based smart home management system with community hierarchy," *IEEE Transactions on Consumer Electronics*, vol. 62, no. 1, pp. 1–9, Jan. 2016.

[8] Q. Wei, D. Liu, Y. Liu, and R. Song, "Optimal constrained self-learning battery sequential management in microgrids via adaptive dynamic programming," *IEEE/CAA Journal of Automatica Sinica*, accept for publication, 2016. DOI: 10.1109/JAS.2016.7510262

[9] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Systems*, vol. 32, no. 6, pp. 76–105, Dec. 2012.

[10] Q. Yang, S. Jagannathan, and Y. Sun, "Robust integral of neural network and error sign control of MIMO nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 12, pp. 3278–3286, Dec. 2015.

[11] Y. Jiang and Z. P. Jiang, "Robust adaptive dynamic programming and feedback stabilization of nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 5, pp. 882–893, May 2014.

[12] Z. Ni, H. He, D. Zhao, X. Xu, and D. V. Prokhorov, "GrDHP: A general utility function representation for dual heuristic dynamic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 3, pp. 614–627, Mar. 2015.

[13] Q. Wei, D. Liu, and H. Lin, "Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems," *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 840–853, Mar. 2016.

[14] Q. Wei, R. Song, and P. Yan, "Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 2, pp. 444–458, Feb. 2016.

[15] Q. Wei, D. Liu, and X. Yang, "Infinite horizon self-learning optimal control of nonaffine discrete-time nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 4, pp. 866–879, Apr. 2015.

[16] Q. Wei, F. Wang, D. Liu, and X. Yang, "Finite-approximation-error based discrete-time iterative adaptive dynamic programming," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2820–2833, Dec. 2014.

[17] Q. Wei and D. Liu, "Data-driven neuro-optimal temperature control of water gas shift reaction using stable iterative adaptive dynamic programming," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6399–6408, Nov. 2014.

[18] Q. Wei, F. L. Lewis, D. Liu, R. Song, and H. Lin, "Discrete-time local value Iteration adaptive dynamic programming: Convergence analysis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, article in press, 2016. DOI: 10.1109/TSMC.2016.2623766

[19] R. Song, F. L. Lewis, Q. Wei, H. Zhang, Z. Jiang, and D. Levine, "Multiple actor-critic structures for continuous-time optimal control using input-output data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 4, pp. 851–865, Apr. 2015.

[20] R. Song, F. L. Lewis, Q. Wei, and H. Zhang, "Off-policy actor-critic structure for optimal control of unknown systems with disturbances," *IEEE Transactions on Cybernetics*, vol. 46, no. 5, pp. 1041–1050, Apr. 2016.

[21] Q. Wei, D. Liu, G. Shi, and Y. Liu, "Optimal multi-battery coordination control for home energy management systems via distributed iterative adaptive dynamic programming," *IEEE Transactions on Industrial Electronics*, vol. 42, no. 7, pp. 4203–4214, Jul. 2015.

[22] Q. Wei, D. Liu, F. L. Lewis, and Y. Liu, "Mixed iterative adaptive dynamic programming for optimal battery energy control in smart residential microgrids," *IEEE Transactions on Industrial Electronics*, accept, 2016. DOI: 10.1109/TIE.2017.2650872

[23] R. Song, W. Xiao, H. Zhang, and C. Sun, "Adaptive dynamic programming for a class of complex-valued nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 9, pp. 1733–1739, Sep. 2014.

[24] G. K. Venayagamoorthy, R. K. Sharma, P. K. Gautam, and A. Ahmadi, "Dynamic energy management system for a smart microgrid," *IEEE Transactions on Neural Networks and Learning Systems*, article in press, 2016. DOI: 10.1109/TNNLS.2016.2514358

[25] T. Huang and D. Liu, "A self-learning scheme for residential energy system control and management," *Neural Computing and Applications*, vol. 22, no. 2, pp. 259–269, Feb. 2013.

[26] J. Si and Y.-T. Wang, "On-line learning control by association and reinforcement," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 264–276, Mar. 2001.

[27] M. Boaro, D. Fuselli, F. D. Angelis, D. Liu, Q. Wei, and F. Piazza, "Adaptive dynamic programming algorithm for renewable energy scheduling and battery management," *Cognitive Computation*, vol. 5, no. 2, pp. 264–277, Jun. 2013.

[28] D. Fuselli, F. D. Angelis, M. Boaro, D. Liu, Q. Wei, S. Squartini, and F. Piazza, "Action dependent heuristic dynamic programming for home energy resource scheduling," *International Journal of Electrical Power and Energy Systems*, vol. 48, pp. 148–160, Jun. 2013.

[29] L. Dong, Y. Tang, H. He, and C. Sun, "An event-triggered approach for load frequency control with supplementary ADP," *IEEE Transactions on Power Systems*, article in press, 2016. DOI: 10.1109/TPWRS.2016.2537984

[30] Q. Wei, D. Liu, and G. Shi, "A novel dual iterative Q-learning method for optimal battery management in smart residential environments," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2509–2518, Apr. 2015.

[31] National renewable energy laboratory (NREL) of U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, operated by the Alliance for Sustainable Energy, LLC. [Online]. Available: http://www.nrel.gov/rredc/

[32] T. Markvart, *Solar electricity (2nd Edition)*. New York: Wiley, 2000.

[33] T. Y. Lee, "Operating schedule of battery energy storage system in a time-of-use rate industrial user with wind turbine generators: A multipass iteration particle swarm optimization approach," *IEEE Transactions on Energy Conversion*, vol. 22, no. 3, pp. 774–782, Mar. 2007.

[34] T. Yau, L. N. Walker, H. L. Graham, and R. Raithel, "Effects of battery storage devices on power system dispatch," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 1, pp. 375–383, Jan. 1981.

[35] Data of electricity rate from ComEd Company, USA. [Online]. https://rrtp.comed.com/live-prices/?.

[36] NAHB Research Center, Inc., "Review of residential electrical energy use data," 400 Prince George's Boulevard Upper Marlboro, MD, USA, July 16, 2001. [Online]. http://www.toolbase.org/PDF/CaseStudies /Res -Electrical-EnergyUseData.pdf

[37] T. Wang, H. Kamath, and S. Willard, "Control and otimization of grid-tied photovoltaic storage systems using model predictive control," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 1010–1017, Mar. 2014.

[38] E. F. Camacho and C. Bordons, *Model Predictive Control*. Berlin, Germany: Springer, 1999.

[39] J. T. Betts, *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*. Philadelphia: Society for Industrial and Applied Mathematics, 2010.

[40] ComEd, USA, http://www.thewattspot.com. Accessed 16 May 2010.

[41] B. Lincoln and A. Rantzer, "Relaxing dynamic programming," *IEEE Transactions on Automatic Control*, vol. 51, no. 8, pp. 1249–1260, Aug. 2006.

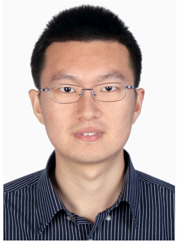
[42] A. Rantzer, "Relaxed dynamic programming in switching systems," *IET Control Theory and Applications* vol. 153, no. 5, pp. 567–574, Sep. 2006.

Qinglai Wei (M'11) received Ph.D. degree in control theory and control engineering, from the Northeastern University, Shenyang, China, in 2009. From 2009–2011, he was a postdoctoral fellow with The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is currently a Professor of the institute. He is also a Professor of the University of Chinese Academy of Sciences. He has authored two books, and published over 60 international journal papers. His research interests include adaptive dynamic programming, neural-networks-based control, optimal control, nonlinear systems and their industrial applications.

Dr. Wei is an Associate Editor of IEEE Transaction on Systems Man, and Cybernetics: Systems since 2016, Information Sciences since 2016, Neurocomputing since 2016, Optimal Control Applications and Methods since 2016, Acta Automatica Sinica since 2015, and has been holding the same position for IEEE Transactions on Neural Networks and Learning Systems during 2014–2015. He is the Secretary of IEEE Computational Intelligence Society (CIS) Beijing Chapter since 2015.

Guang Shi received the B.S. degree in automation from Zhejiang University, Hangzhou, China, in July, 2012. Currently, he is working towards the Ph.D. degree at The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include neural networks, adaptive dynamic programming, optimal control and energy management in smart grids.

Ruizhuo Song (M'11) received the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2012. She is currently an associate professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing. Her research interests include optimal control, neural-networks-based control, nonlinear control, wireless sensor networks, adaptive dynamic programming and their industrial application.



Yu Liu received the B.S. degree in automation, the M.S. degree in Pattern Recognition and Intelligent Systems, from Beijing Institute of Technology, Beijing, China, in 2001 and 2004 respectively, and the Ph.D. degree in Pattern Recognition and Inteligent Systems, from Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2010. His research interests include artificial intelligence, streaming data analysis, and energy management in smart grids.