Approximate Dynamic Programming for Control of a System of Electric Vehicle Charging Stations

Ying Chen

Department of Industrial, Manufacturing, and Systems Engineering, The University of Texas at Arlington, U.S.A., ying.chen@mavs.uta.edu

Jay Rosenberger

Department of Industrial, Manufacturing, and Systems Engineering, The University of Texas at Arlington, U.S.A., jrosenbe@uta.edu

Victoria C. P. Chen

Department of Industrial, Manufacturing, and Systems Engineering, The University of Texas at Arlington, U.S.A., vchen@uta.edu

Wei-Jen Lee

Department of Electrical Engineering, The University of Texas at Arlington, U.S.A., wlee@uta.edu

Taking the Dallas Fort-Worth metropolitan area as the assumed sample system, control of an electric vehicle (EV) quick charging infrastructure with 5 EV charging stations is performed. This system integrates renewable energy to supply electricity to charging stations as well as the main electrical grid. Batteries as electrical storage facilities are installed in each station. Through the usage of batteries, the control system is expected to be able to store the surplus electricity from renewable energy or buy electricity at a lower electricity market price (EMP) and sell the stored electricity back to the main grid when the EMP reaches to a higher level to decrease the operational cost or make profits for this system. In order to control this system, we formulate it as a Markov decision process problem. An infinite horizon approximate dynamic programming approach based on design and analysis of computer experiments is used to solve this high-dimensional, large-scale, EV charging station control problem over continuous

spaces. A 45-degree line correspondence stopping criterion specified is used to stop the DP iterations and select the ADP policy. From the results, it is clear that control from the selected ADP policy has a better performance than the benchmark policy.

Key words: electric vehicle charging infrastructure, approximate dynamic programming, renewable energy, design and analysis of computer experiments, large-scale

Subject classification: Dynamic programming: Markov-infinite state, Applications; Natural resources: Energy

Area of review: Environment, Energy, and Sustainability

---

1. Introduction

Environmental pollution is becoming more and more serious with the rapid development of our society (Chen et al. 2013). Traditional energy sources such as coal, gas, and oil are some of the main sources resulting in pollution such as greenhouse effects, and dust and ashes in the air. Consequently, research on renewable energy sources is becoming more prevalent. With developing technology, greenhouse gas emissions will be reduced to 17% of 2005 levels by 2020 as the U.S. government pledged (http://www.eia.doe.gov/emeu/aer/pdf/pages/sec12_4.pdf, 2010). After investigation and analysis, the transportation sector, causing up to 33.1% of all energy related emissions, has been identified as the largest producer of carbon dioxide emission in the U.S. (http://cta.ornl.gov/data/download29.shtml). Therefore, to abate such emissions, electric vehicles (EVs) have been given a lot of attention. This fundamental transformation from oil based vehicles to electric power ones will help decrease carbon dioxide emissions substantially. From Fell et al. (2010), there will be more than 1 million EVs in US by 2017 with the target growth rate and in the Dallas-Fort Worth (DFW) metro area, the amount of EVs will be increased to around 10,000 by that time.

In order to increase the penetration and usage of EVs, drivers' anxiety about the driving range needs be resolved. In the existing literature, the design of optimal EV charging profiles is a way to serve this purpose. For example, the authors take advantage of charging behaviors from demographical statistical data to assess the EV charging scenarios (Steen et al. 2012), vehicle usage data to predict and analyze the EV charging profile (Ashtari et al. 2012) and a dynamic game theoretic optimization to formulate the optimal EV charging problem (Zhu et a. 2012). Additionally, in Clement-Nyns et al. (2010), coordinated charging profile with the objective of minimizing power losses and maximizing the main grid load factor in residential distribution, where the optimal EV charging stations are included, is presented. Yao et al. (2017), considering demand response, formulated the optimal power profile problem with the objective of maximizing the number of charging EVs and minimizing the total charging cost simultaneously. Furthermore, other models for coordinated EV charging systems have been developed to improve power utilization (Wang et al. 2015), smooth real-time power fluctuation in a regulation service (Shaaban et al. 2014), avoid overload in the utility grid (Gan et al. 2013), and support the power system restoration as a black-start power source (Sun et al. 2016). In order to improve the penetration of sustainable energy in the charging systems, Badawy and Sozer (2017), Khodayar et al. (2012), Marano and Rizzoni (2008), and Guo et al. (2014) have employed wind or solar generation as a type of energy resources to supply electricity to the charging stations.

Though these studies have proposed various different effective methods to develop EV charging infrastructure, they focus on level 1 or level 2 charging systems, which still cannot meet the desire of fast charging from the customers. Considering this issue, Sarikprueck et al. (2017) proposed a novel regional EV DC level 3 fast charging system equipped with renewable resources, such as wind and solar energy, to serve EV demand within minutes. However, the decision making procedure in Sarikprueck et al. (2017) is deterministic, and when making decisions, the system does not consider the future state but only the current state. Kulvanitchaiyanunt et al. (2016) utilized this system, which makes a decision for each stage through linear programming without considering the uncertainty.

Therefore, for a public charging station development, we also focus on this EV DC level 3 fast charging system, which was initially well designed in Sarikprueck et al. (2017). However, we formulate it as a Markov decision process (MDP) so that the future state needs to be considered when making a decision. In this system, wind and solar energy as well as the main power grid provide electricity. Besides, in each station, local battery storage systems are used as a buffer to minimize the operational cost. However, different from Kulvanitchaiyanunt et al. (2016) and Sarikprueck et al. (2017), we apply approximate dynamic programming (ADP) to solving this large-scale, high-dimensional, dynamic control system so that the decision making procedure considers the uncertainty and future states.

In the rest of this paper, background and contribution are introduced in section II. Then, the details of this dynamic EV charging station control problem are introduced in section III. Computational results using the ADP algorithm are shown in section IV. Moreover, simulation results are discussed in section V. Finally, conclusions about this research are presented.

2. Background

In the literature, ADP algorithms such as Q-learning and post-decision state approaches, have been used in smart grid related research as a control technique. Namely, Wei et al. (2015) proposed a novel iterative Q-learning method named “dual iterative Q-learning algorithm” to solve the optimal battery management control problem in smart residential environments. Based on this research, Wei et al. (2017) introduced a mixed iterative adaptive DP approach to control the battery energy in smart residential micro-grids. Boaro et al. (2013) presented an intelligent management scheme using dynamic programming for smart grids using renewable energy combined with battery storage. Shi et al. (2016) optimized electricity consumption in office buildings with online learning control systems, which are based on reinforcement learning (RL). In addition, reactive power control of grid-connected wind farm was solved by an adaptive DP in Tang et al. (2014). Moreover, a heuristic DP (HDP) architecture was established to schedule the quality of service in cognitive-ratio-based smart grid networks in Yu et al. (2016). Furthermore, a

modified ADP based on an actor-critic network in Xie et al. (2016) was used to schedule fair energy to vehicle-to-grid networks. Jiang and Powell (2015) employed a convergent ADP algorithm that exploits monotonicity of the value function to find a revenue-generating bidding policy in a real-time electricity market with battery storage. According to this latest research, it is clear that ADP algorithms are playing a critical role in the control of smart grid problems.

However, large-scale, high-dimensional DP problems over continuous spaces are still a challenge for the algorithms used in the above studies. For example, although, in Yu et al. (2016), there were 28 state variables and 8 decision variables in the space, these variables were discrete as in Shi et al. (2016) and Xie et al. (2016). For continuous spaces, the systems are not considered very large. For example, Ernst et al. (2009) considered only two state variables and one decision variable, and Wei et al. (2017) only used two state variables and two decisions variables over continuous spaces. Even though a deep learning based RL algorithm is developed in Peng et al. (2016) aiming to solve the high-dimensional infinite horizon DP problem over a continuous state space, this methodology requires a large amount of data, including 300,000 iterations of training and 10 million tuples, which suggests that the algorithm requires substantial computation. In addition, they optimized over a discrete action space using a derivative-free evolutionary algorithm. As Lee and Lee (2004) summarized, the common limitations of using RL and neuro-dynamic programming (NDP) are: a) in DP problems with continuous state and action spaces, the discretization and common "incremental" update rule are impractical, and the function approximation errors can grow rapidly; b) complex dynamics of most chemical processes still bound the state space that can be explored, which results in regions of sparse data.

To solve the large-scale, high-dimensional, infinite horizon DP problem over continuous spaces, Chen et al. (2017a) proposed a design and analysis of computer experiments (DACE) based infinite horizon ADP algorithm to sample the state space with design of experiments and approximate the value function via statistical modeling methods. This algorithm is based on the finite horizon version proposed by Chen et al. (1999). With the approach introduced in Chen et al. (1999), several large-scale, high-dimensional,

finite horizon DP cases have been solved successfully, such as an ozone problem with more than 500 dimensions (Yang et al. 2009) and a 38-dimensional waste water problem (Cervellera et al. 2006). Hence, considering the limitations in RL/NDP and success achieved by the DACE based finite horizon algorithm, we employ this DACE based infinite horizon ADP algorithm introduced in Chen et al. (2017a) to solve this large-scale, high-dimensional, infinite horizon, EV charging station control problem over continuous spaces. Therefore, the contribution of this paper is listed below:

(1) This EV Level 3 DC charging station system is formulated as a MDP problem.
(2) State transition models are developed based on support vector regression (SVR) and martingale model of forecast evolution (MMFE) models.
(3) A large-scale, high-dimensional, infinite horizon, DP problem over a continuous space is solved with an ADP algorithm

3. ADP approach

In this section, first, we will briefly describe the infinite horizon stochastic DP (SDP) formulation and then overview the DACE based infinite horizon ADP algorithm that will be used in this study.

3.1. Infinite horizon SDP model

DP as a mathematical programming approach to optimize a system evolving over time was introduced by Bellman in 1957. There are two versions of DP: finite horizon and infinite horizon. Different from finite horizon DP, infinite horizon DP only has one true value function. A typical recursive SDP formulation for an infinite horizon problem can be written as

$$V(x) = \min_{u \in \Gamma} E\{c(x,u,\zeta) + \gamma V(f(x,u,\zeta))\} \quad (1)$$

In Eq. (1), x is vector of state variables, u is a vector of decision variables, ζ is a vector of stochastic variables, f is the transition function, Γ is a set of feasible decisions, γ is a discount factor, c is the cost

function, and V is the future value function (FVF). As described in Section 1, finding an FVF exactly is intractable for medium-sized or large-sized problems. Consequently, ADP attempts to find an approximate FVF (aFVF) ($\hat{V}$) using the following formulation.

$$\hat{V}(x) \approx \tilde{V}(x) = \min_{u \in \Gamma} E\{c(x,u,\zeta) + \gamma \hat{V}(f(x,u,\zeta))\} \quad (2)$$

3.2. Overview of DACE based infinite horizon ADP algorithm

In order to solve large-scale infinite horizon DP problems over continuous spaces, Chen et al. (2017a) proposed an algorithm as shown below, on the basis of the DACE concept and adaptive value function approximation (AVFA) approach (Fan et al. 2013).

Step 0: Initialization:

(a). Input a discount factor, γ, a state transition function, f, and a cost function, c.

(b). Choose the training data set X^{Train} and testing data X^{Test} set generated by low-discrepancy sequences, respectively.

(c). Set the iteration counter to $k \leftarrow 0$, set the initial aFVF, $\hat{V}_0 = 0$, and the set of evaluated state variables $X \leftarrow \emptyset$.

Step 1: Iteration of infinite horizon DP:

(a). Set $k \leftarrow k + 1$.

(b). For each state variable $x \in X^{\text{Train}} \cup X^{\text{Test}} - X$, solve $\tilde{V}_k(x) = \min_{u \in \Gamma} E\{c(x,u,\xi) + \gamma \hat{V}_{k-1}(f(x,u,\xi))\}$ and set $X \leftarrow X \cup \{x\}$;

(c). Fit a regression model using the data $\{(x, \tilde{V}_k(x))\}_{x \in X^{\text{Train}}}$ to obtain $\hat{V}_k$;

(d). If $\hat{V}_k$ fails the stopping criteria for the data loop on the data $\{(x, \hat{V}_k(x))\}_{x \in X^{\text{Test}}}$, a new set of state variables X' will be added to the training data set $X^{\text{Train}} \leftarrow X^{\text{Train}} \cup X'$ and go to the step 1 (b);

(e). If $\hat{V}_k(x) \approx \hat{V}_{k-1}(x)$, for each $x \in X^{\text{Test}}$, output $\hat{V}_k$; otherwise, $X \leftarrow \emptyset$ and go to Step 1(a).

In this algorithm, two loops are used to develop the aFVF: an inner data loop and an outer DP loop. The data loop follows the AVFA approach of Fan et al. (2013) to sample the state space sequentially in

order to control the amount of sampling needed. The DP loop follows the value iteration concept to generate a sequence of aFVFs. In the above description, Steps 1(a)-(e) are the DP loop and Steps 1(b)-(d) represent the data loop. As presented in Chen et al. (2017b), SVR has presented a more stable performance in the value function approximation than multivariate adaptive regression splines for the infinite horizon DP case. Therefore, in this study, SVR with a Gaussian radial basis function (RBF) kernel is used to approximate the FVF. Following Chen et al. (2017b), we also utilize least-square support vector machines toolbox downloaded from http://www.esat.kuleuven.be/sista/lssvmlab/ to create the aFVFs. Please refer to Brabanter et al. (2011) for details on the tuning process for the parameters of the RBF kernel.

4. Dynamic control problem formulation

In this EV charging station control problem, following the design in Sarikprueck et al. (2017), it is assumed that there are 11 stations located in the DFW metropolitan area. In this system, a remote wind farm provides wind power for the system through a long term contract and solar panels installed on the roof of each station supply the solar power for each station. The main grid also exchanges electricity with each station. A battery is installed in each station for electricity storage, and the priority of this control is to satisfy the demand from EVs and simultaneously, make optimal decisions at each time period for electricity trading with the main grid to minimize the operational cost for the system. The blueprint of this dynamic control system is shown blow. Ideally, the control strategy is that when the electricity market price (EMP) is low, the system will buy electricity and store the surplus in the battery, and sell electricity when the EMP is high. In this study, based on the technology of photovoltaic panels, we assume solar power only provides a very small amount of electricity to the system, and in addition, the maximum supply of wind power is no more than 30% of total demand from the EVs. Therefore, the EMP plays a critical role in this power trading system.

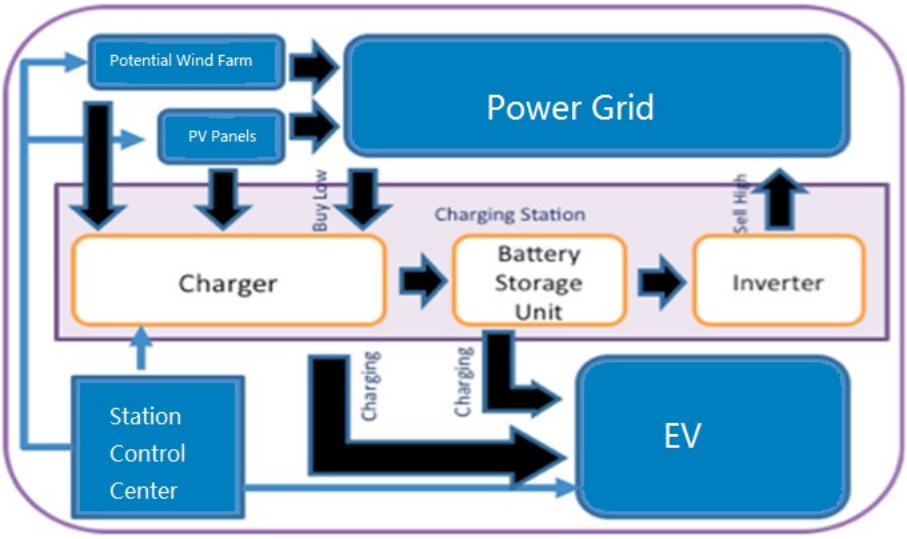


Figure 1. Blueprint of this EV charging station control system (Sarikprueck et al. 2017)

Historically, the EMP can spike, and the historical EMP range is from -150 $/MW•h to 3300 $/MW•h (Sarikprueck et al. 2015). However, the real EMP is distributed as a Cauchy distribution, and the occurrence of the negative or a spike EMP is rare as shown below.

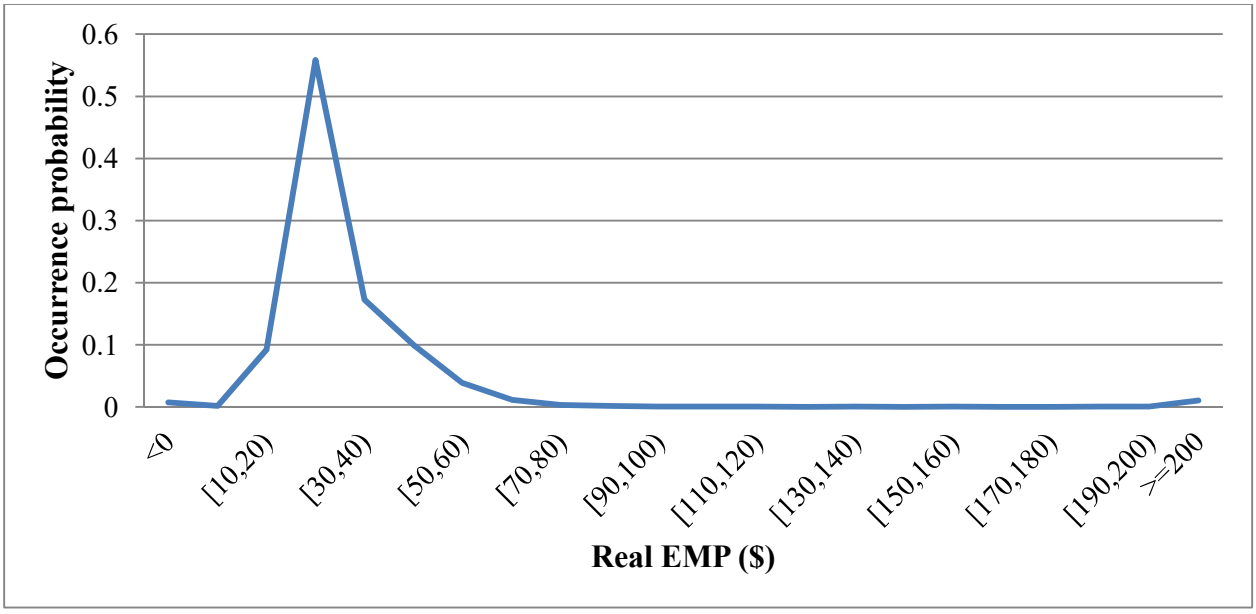


Figure 2. Real EMP distribution

In statistics, when building the model with such distribution, the accuracy of the model will be affected by the rare data with large values (Kutner et al. 2004). Hence, based on the control strategy mentioned above, it is straightforward to control the system to sell electricity back to the main grid when a positive spike EMP occurs and purchase electricity for making profits when a negative EMP occurs, which is defined as the first-type control strategy in this paper. The ADP policy is appropriate when EMPs are in the range [0, 200), which is referred to as the second-type control strategy. In this paper, we mainly focus on developing the second-type control strategy.

4.1. EV charging station control problem formulation

Kulvanitchaiyanunt et al. (2016) formulated the EV control problem as a deterministic linear programming problem. However, the uncertainty cannot be avoided in reality. Considering stochastic variables, when this EV charging station system evolves over time, we have the following process:

At each period, the control center first needs to determine how to satisfy the demands from EVs, allocate wind power, and store and manage electricity in the batteries. After that, random realizations of wind power, solar power and EMP are determined, and the recourse decisions, which are to trade between power grid and batteries accounting for surpluses and shortages, are recalculated based on the previous determined information in order to optimize the objective. After this, the process transitons into the next period.

With this process and the deterministic formulation in Kulvanitchaiyanunt et al. (2016), we formulate this EV charing station control problem as an MDP. In order to solve it, an ADP algorithm is conducted using the following formulation. It is noted that the ADP solution process has some small differences from the simulation process described above: in the ADP solution process, recourse decisions do not need to be recalculated since the solution process aims to achieve an convergent policy.

Table 1. Major nomenclature used

γ	Discount factor	I	Inventory of the battery
D	Total demand	$D2$	Demand satisfied by the battery of a station
e	Battery storage efficiency	R	Electricity sold back to the grid from the battery
W	Wind energy generated for the system	$\widetilde{W}$	Allocation of the wind energy to each station
n	Total number of open stations	PV	Photovoltaic (solar) production for each station
N	Number of recourse scenarios	$D1$	Demand satisfied by direct charge at each station
I_{min}	Minimum battery storage level at each station	I_{max}	Maximum battery storage level at each station
g^{+}	Electricity bought from the main grid	g^{-}	Electricity sold to the grid from direct charge
BC	Amount of battery charged electricity	EMP	Electricity market price
cr	Battery charge upper limit	dc	Battery discharge rate

$$\sum_{t=1}^{\infty}\gamma^{t}\left(\frac{1}{N}\left(\sum_{i=1}^{N}\sum_{j=1}^{n}\left(EMP_{t,i}^{j}g_{t,i}^{+,j}-EMP_{t,i}^{j}\left(g_{t,i}^{-,j}+R_{t}^{j}\right)\right)\right)\right), \qquad \forall j\in n,\forall j\in n,\forall i\in N \quad (3)$$

Eq. (3) is the objective function of this SDP problem, and the first constraint set (4) includes the battery level transition from period t-1 to period t for each open station j:

$$I_t^j = I_{t-1}^j + BC_t^j - \frac{R_t^j}{e} - \frac{D2_t^j}{e} \quad \forall j \in n, \forall i \in N \ . \tag{4}$$

The solar generation is the same for each station. Therefore, the amount of battery charged electricity is calculated with the following equation (5):

$$BC_t^j = \widetilde{W}_t^j W_{t,i} + PV_{t,i} + g_{t,i}^{+,j} - g_{t,i}^{-,j} - D1_t^j \qquad \forall j \in n, \forall i \in N \ . \tag{5}$$

The total demand consists of the demand satisfied by direct charge and demand satisfied by the battery as shown in the following constraint set (6):

$$D_t^j = D1_t^j + D2_t^j \qquad \forall j \in n \ . \tag{6}$$

The combination of electricity sold back to the grid from the battery and the demand satisfied by the battery together is less than or equal to the discharge rate (*dc*) multiplied by the storage efficiency, as shown below in (7):

$$R_t^j + D2_t^j \leq dc \times e \qquad \forall j \in n \ . \tag{7}$$

The battery charge must not be greater than its charge upper limit, and the battery level must be constrainted in between the minimum battery level and the battery capacity for each station, as shown below in (8) and (9):

$$BC_t^j \leq cr \quad \forall j \in n \quad , \tag{8}$$

$$I_{min} \leq I_t^j \leq I_{max} \qquad \forall j \in n \qquad . \tag{9}$$

The sum of the fraction of the wind allocation should be equal to 1 as in (10):

$$\sum_{j=1}^{n} \widetilde{W}_t^j = 1 \qquad \forall j \in n \quad , \tag{10}$$

$$I_t^j, \widetilde{W}_t^j, g_{t,i}^{+,j}, g_{t,i}^{-,j}, BC_{tj}, R_t^j, D1_t^j, D2_t^j \geq 0 \quad \forall j \in n, \forall i \in N \ . \tag{11}$$

As assumed, there are two types of decision variables: one is pre-realization decision variables, and the other includes post-realization decision variables, which are shown in Table 2. To solve this, stochastic programming is used, and multiple scenarios for wind power, solar power and EMP are implemented. It is noted that all of these decision variables are in the continuous space.

Table 2. Decision variables

Pre-Realization Decision Variables	R_t^j, $\widetilde{W}_t^j$, $D1_t^j$, $D2_t^j$, I_t^j, BC_t^j $\quad \forall j \in n$
Post-Realization Decision Variables	$g_{t,i}^{-,j}$, $g_{t,i}^{+,j}$ $\forall j \in n, \forall i \in N$

4.2. State transition model

Using the features selected in Sarikprueck (2015) for wind power, solar power and non-spike EMP, through SVR with its parameter tuning process for RBF kernel, we build the forecasting models for wind power, solar power and non-spike positive EMP. In addition, a Martingale model of forecast evolution (MMFE) (Heath and Jackson 1994) is implemented to explore the stochastic nature of these forecasting models. Therefore, combining SVR forecasting models and MMFE, we have the following state transition process. In details, at the time period *t*, we have the historical states d_{t-2}, d_{t-1} and forecasting state $d_{t,t}$; when evolving to time period *t* + 1 and keeping the same structure, the state vector becomes d_{t-1}, d_t and $d_{t+1,t+1}$. From time period *t* to *t* + 1, d_{t-2} changes to d_{t-1}, which is defined as an identity transition in Yang et al. (2009). For the other state variables:

$$d_t = d_{t,t} \times \varepsilon_{t,t} \quad , \tag{12}$$

$$d_{t+1,t+1} = F(d_{t-1}, d_t) \times \varepsilon_{t,t+1} \quad , \tag{13}$$

where ε is generated using MMFE methods (please refer to Health and Jackson for details), and F is the forecasting model generated by SVR mentioned above. Note that only wind power, solar power and EMP will use this transition process. The state space also includes the battery inventory level and demand level. Battery inventory level will transition using Eq. (4). In this study, we use the DFW EV daily demand profile shown in Fig. 4, designed by Khosrojerdi et al. (2013), which shows demand over 96 15-minute time periods for the random selected 5 open stations. In this demand profile, the demand usually remains the same from period to period but includes a few periods with either a sharp increase or decrease in demand. Hence, when transitioning from time period t to $t + 1$, we assume that the demand levels of stations are kept unchanged to approximate the FVFs. However, when simulating this whole system to investigate the performance of the selected ADP policy, we take advantage of Fig. 4 for the evolving demand level.

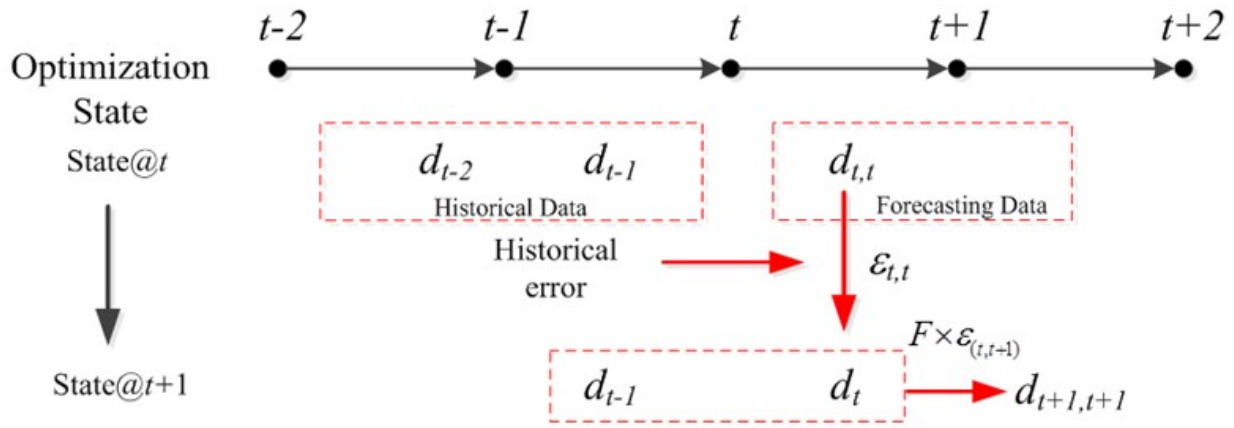


Figure 3. State Transition Process for wind, solar and EMP state variables

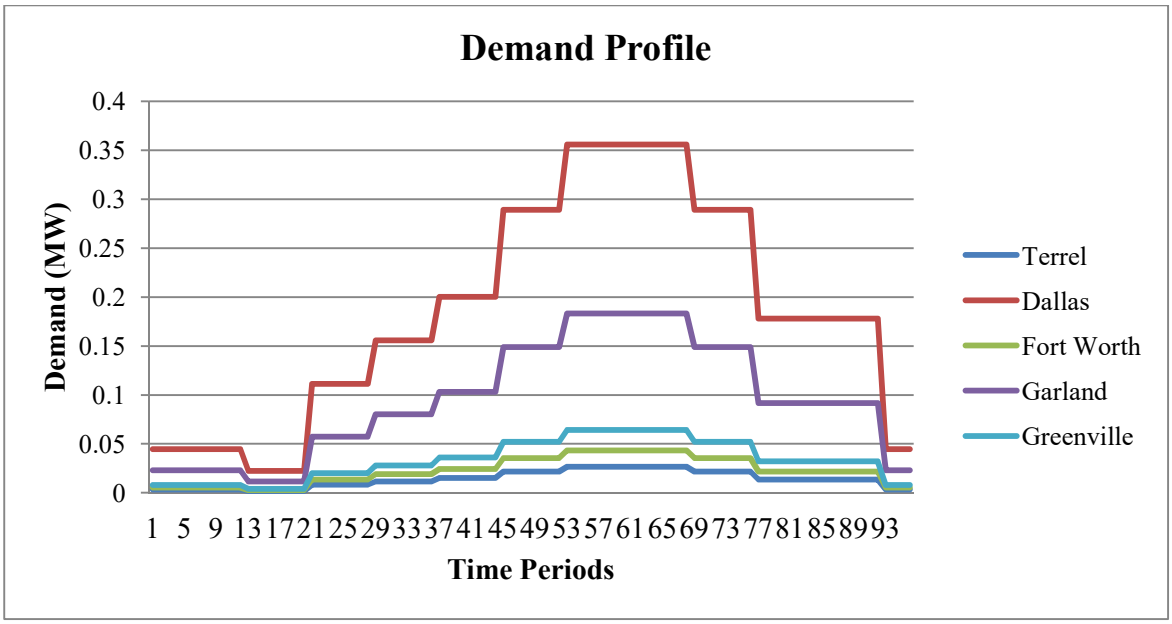


Figure 4. Demand profile of 5 charging stations located in DFW metro area (Khosrojerdi et al. 2013)

In wind power forecasting, wind speed is used as an important state variable, and in the non-spike EMP forecast, load profile and temperature are also included. As to these three state variables, we take advantage of historical data to represent them at time period *t* and shift them to the next period based on the data. Therefore, when implementing a DACE based infinite horizon ADP algorithm, these three state variables are not included in the state space. Regarding this, if there are 5 stations open in the state space, we have the following state variables: 3 for state variables wind, 3 for solar, 15 for EMPs in five stations, 5 for battery inventory levels of the five stations and 5 for the demand levels for each station. In total, the state space has 31 dimensions.

5. Computational Results

In this section, first, we will overview the 45-degree line stopping criterion proposed in Chen et al. (2017a) and specified in (2017b), and then quantify the stopping settings in advance. After that, we will use an ADP solution algorithm to build aFVFs iteratively until the stopping rule is satisfied. Moreover, the

selected aFVF will be applied in the simulation in order to explore its performance. Both the ADP policy and a greedy benchmark policy are simulated for comparison.

5.1. 45-degree line correspondence stopping criterion

45-degree line correspondence stopping criterion proposed in Chen et al. (2017a) is used to identify the shape of value function by observing the linear relationship between consecutive outputs of DP iterations. The fitted regression line is represented:

$$\hat{Y}_k = b_0 + b_1 X_{k-1} \quad , \tag{14}$$

where b_0 is the intercept, b_1 is the slope, X is the $\tilde{V}$ at iteration k-1, Y_k is the $\tilde{V}$ at iteration k, $\hat{Y}_k$ is the estimated Y_k.

To improve upon the 45-degree line correspondence stopping criterion, Chen et al. (2017b) presented an algorithm to specify a stopping rule. In this algorithm, $b_1^{'}$, named as b_1 with 1 lag, is the relationship between the consecutive outputs of $\tilde{V}$ at iteration k and k - 1. $b_1^{''}$, named as b_1 with 2 lags, is the relationship between the outputs of $\tilde{V}$ at iteration k and k - 2. This algorithm mainly takes advantage of the difference between $b_1^{'}$ and $b_1^{''}$ to identify the ADP policy and stop the DP iterations.

$$\delta_k = b_{1,k}^{''} - b_{1,k}^{'} \quad , \quad k = 3, 4, \ldots \tag{15}$$

where δ_k is the difference between b_1 with 2 lags and b_1 with 1 lag at k^{th} DP iteration.

According to this algorithm, if δ_k is less than or equal to ϵ, which is a specified error tolerance value, $b_{1,k}^{''}$ and $b_{1,k}^{'}$will be close enough. At this time, a potential "good" ADP policy might have been generated and we need to check the $|\delta|$ values from iteration k +1 to k +m to see if all $|\delta|$ values are also less than or equal to ξ. If so, output the aFVF ($\hat{V}_k$) at iteration k; otherwise, DP iterations should be continued. As to the detail of this algorithm, please refer Chen et al. (2017b).

5.2. FVF approximation

In this research, we randomly assume there are five stations open, and the number of charging slots in these stations is given in below:

Table 3. Operation condition of 5 open charging stations in DFW metro area

Order of stations	Terrel	Dallas	Fort Worth	Garland	Greenville
# of slots	2	9	6	4	8

The number of slots plays the key role in determining the size of battery in each station. Table 4 shows the battery size for each open station:

Table 1 Battery size of open stations

Order of stations	Terrel	Dallas	Fort Worth	Garland	Greenville
Min level (MW)	1.44	6.48	4.32	2.88	5.76
Max level (MW)	7.2	32.4	21.6	14.4	28.8

Since there are five stations open, the state space has 31 dimensions, which is much more than the related work in the literature. Following the algorithm proposed in Chen et al. (2017a), the size of training data set and testing data set is 250 and 250, respectively, and the increment size of the training data in the data loop is 50. Sobol sequence (Sobol 1967) is used to generate the training data, and Halton sequence (Halton 1960) is used to initialize the testing data. The storage efficiency is assumed to be 79.8% as in Kulvanitchaiyanunt et al. (2016) and Sarikprueck et al. (2017). The discount factor is set to be 0.995 since the time period is 15 minutes. Experiments on the DACE based infinite horizon ADP algorithm are conducted in MATLAB 2016b on a Lenovo computer with a Xeon, 16-core, 2.8 GHz CPU. For the optimization step 1(b), "fmincon" from MATLAB is used. In order to compute $\tilde{V}$ of each sampled point in the state space, eight recourse scenarios are generated from realizations of the stochastic variables. The

data loop stopping criterion is when the difference between consecutive data loop R^2 values from the testing data set is less than 0.01. We set the ϵ equal to 0, ξ equal to 0.005 and m equal to 5 as described in Section 5.1 for the DP loop stopping criterion. Consequently, the selected ADP policy occurs when the b_1'' and b_1' curves cross and then start to stabilize.

Figure 5. $\boldsymbol{\delta}$ value evolving pattern

Figure 5 shows the 61st to the142nd iterations of the DP loop, during which several potentially good ADP policies are saved. However, with aforementioned DP loop stopping criteria, the ADP algorithm stops at the 137th iteration, which requires almost 69 hours. The 137th aFVF is selected, and we simulate the MDP using it as well as other aFVF policies to evaluate performance.

5.3. Simulation Results

In this section, we compare aFVF policies in a simulated environment. Moreover, a myopic greedy algorithm, as described in Chen et al. (2017a, 2017b), is utilized as a benchmark in this study. At the

initial stage, we set the battery inventory of each station at the minimum level. As mentioned above, the demand profile in Fig. 4 is used in the simulation.

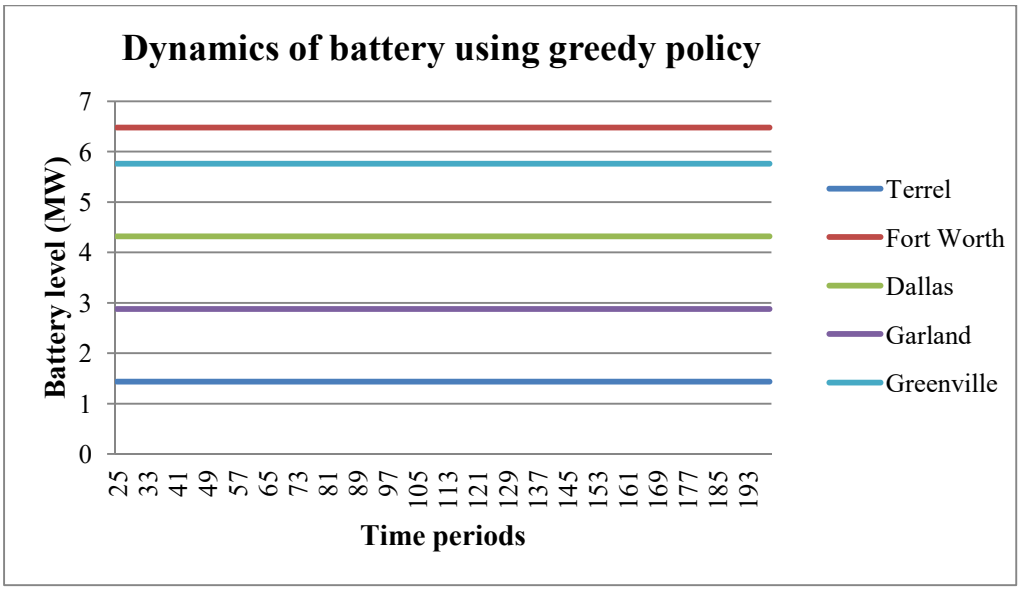
Following the simulation process proposed in Chen et al. (2017a), 15 scenarios are conducted with 200 stages for each scenario. Since we set the battery inventory of each station at the minimum level at time period 1, for the MDP simulation, the final results may be skewed by initial transient states (Taha, 2003). Therefore, in this study, we only consider the recurrent states in the simulation. After observing the dynamics of battery charging and discharging within the 200 stages, the states from 25th stage to 200th stage are regarded as recurrent states. Table 5 displays the simulation results of these 15 scenarios using the greedy policy and the 137th aFVF. The values in Table 5 are calculated with Eq. (16). The values in Table 5 are estimated total costs from 25th stage to 200th stage using the greedy myopic policy and the ADP policy. All estimated costs of the 15 scenarios are much smaller than those of greedy policy, which indicates the ADP policy is much better than the greedy policy. Fig. 6 shows an example of the dynamics of the battery in scenario 1 using the greedy and ADP policies.

$$\min_{u_1,\ldots,u_t} E\{\sum_{t=25}^{200} \gamma^t c(s_t, u_t, \varepsilon_t)\} \tag{16}$$

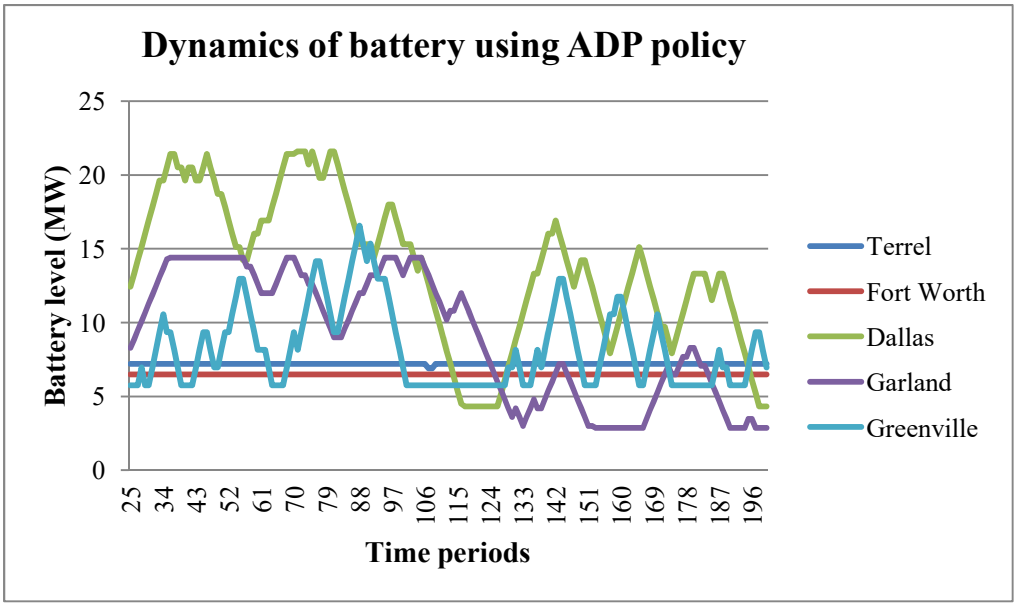
Table 5. Estimated total costs of 15 scenarios using the greedy policy and the 137^{th} aFVF

Order of Scenarios	Greedy policy	137^{th} aFVF
1	913.14	-1036.46
2	627.32	-1097.76
3	990.12	-1527.15
4	1087.69	-835.44
5	1161.63	-1008.51
6	936.19	-1032.07
7	1424.36	-391.89
8	995.70	-1333.42
9	1564.57	57.35
10	905.70	-580.08
11	1456.60	-240.29
12	1200.74	-679.61
13	1474.57	670.79
14	1338.57	-536.73
15	1230.17	-12.25

(Unit is $)



(a)

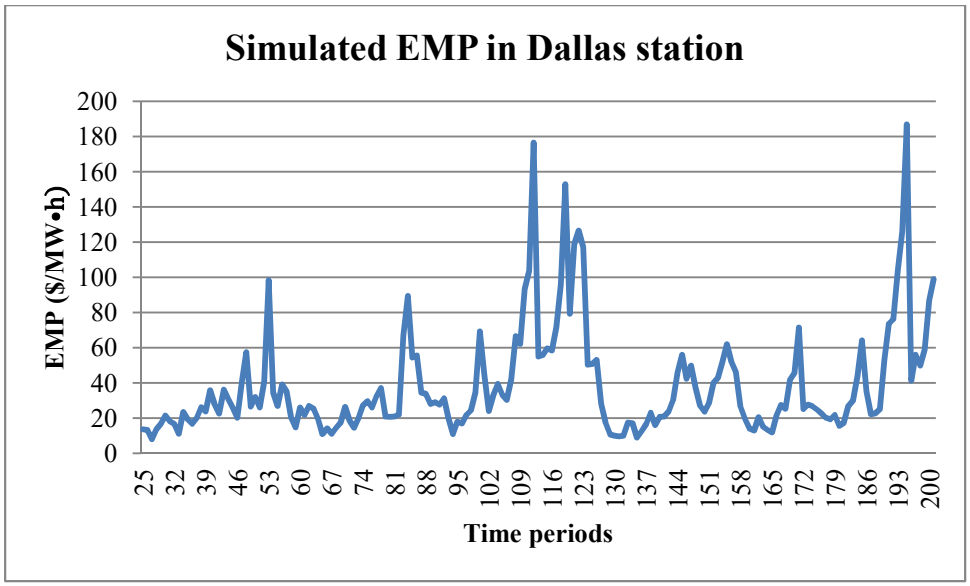


(b)

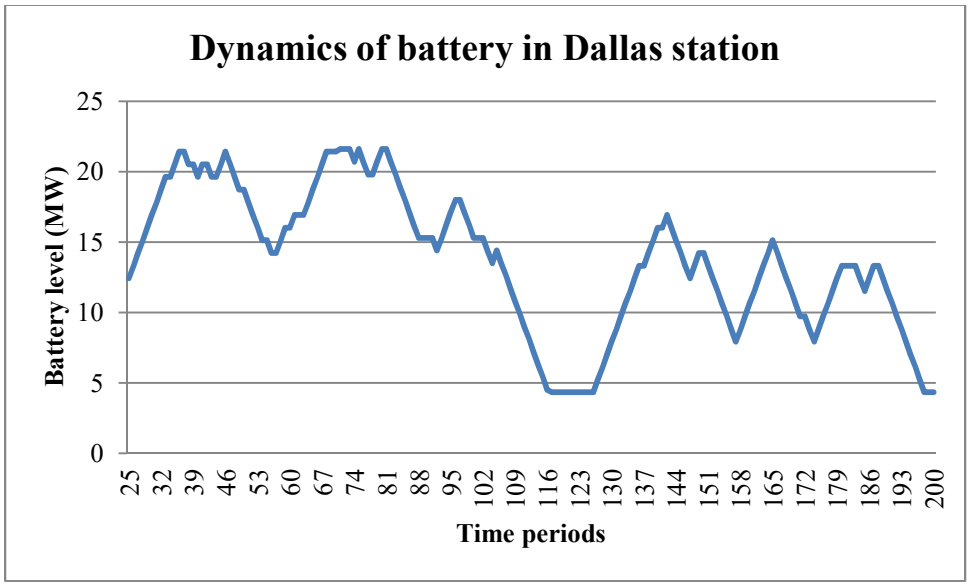
Figure 6. Dynamics of battery using greedy policy in (a) and using ADP policy in (b).

The difference between Fig. 6(a) and 6(b) is significant. The greedy policy never charges the batteries, because it does not consider future states in which the stored electricity could be used either to satisfy demand or sell back to the main grid during a period with higher EMP. By contrast, the ADP policy charges and discharges the batteries in four of five stations. In order to present the dynamic control more clearly, the simulated EMP from the Dallas station and the corresponding dynamics of battery using the ADP policy are shown in Fig. 7.

Fig. 7 shows that when the EMP is low, the battery is charged, and when the EMP is high, the battery is discharged, which indicates the ADP policy is able to control the system as expected. Based on the results from the 137^{th} aFVF and greedy policy, we can conclude the ADP policy performs much better than the greedy policy. It is noted that in the example shown in Figs. 6 and 7, the first-type control strategy is not used since all EMPs are positive and no more than 200. However, the first-type control strategy may be used in other scenarios.



(a)



(b)

Figure 7. Simulated EMP in station 5 in (a) and its corresponding dynamics of battery in (b)

6. Discussion

Before concluding the results, we further conduct another experiment to the 137th aFVF policy to an aFVF found running additional DP iterations. The L-infinity norm stopping rule, proposed by Powell (2007), stops based upon the criterion shown below:

$$\|V_k - V_{k-1}\|_\infty < \frac{\theta(1-\gamma)}{2\gamma}. \qquad (17)$$

The stopping criterion is reached when the maximum change in the value of any state is lower than the setting of the right-hand side in Eq. (17), where γ is the discount factor, and θ is a specified error tolerance. Fig. 8 shows the L-infinity norm over the first 200 iterations, which requires more than 4 days to compute.

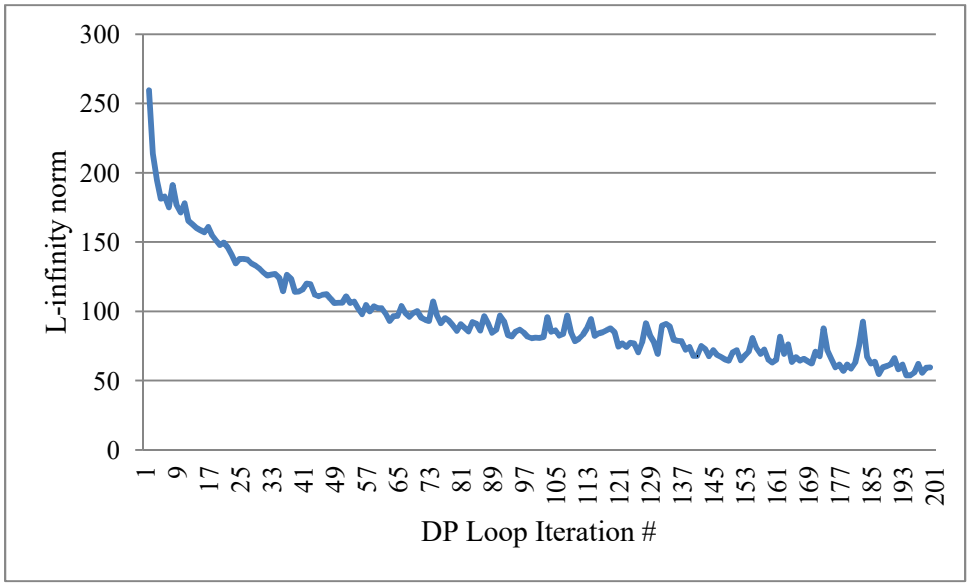


Figure 8. L-infinity norm value evolving pattern within 200 DP iterations

In Fig. 8, observe that the L-infinity norm is still decreasing slowly, which indicates that the aFVF is on the way to convergence. According to Fig. 8, all the norm values are more than 50, which suggests that more DP iterations should be executed in order to converge, even though the 45-degree correspondence line stopped at the 137th iteration. However, Table 6 compares the simulated estimated total cost of both

the 137th and the 200th aFVF policies. Using a paired t-test to compare the these policies, the *p*-value is 0.201, which suggests that there are no statistical differences between 137th aFVF and 200th aFVF when α is 0.05. Based on this result, we can conclude that the quantified 45-degree line correspondence stopping criteria in Chen et al. (2017b) is able to stop the DP loop reasonably and early, and select an ADP policy that has the same performance statistically as a policy found after additional iterations, which saves a lot of computational time.

Table 6. Estimated total costs of 15 scenarios using 137th aFVF and 200th aFVF in simulation

Order of Scenarios	137th aFVF	200th aFVF
1	-1036.46	-1058.14
2	-1097.76	-1011.96
3	-1527.15	-1505.09
4	-835.439	-1064.68
5	-1008.51	-848.721
6	-1032.07	-1120
7	-391.889	-167.266
8	-1333.42	-1113.03
9	57.34632	157.2903
10	-580.075	-474.113
11	-240.294	-315.418
12	-679.609	-594.316
13	670.7892	647.2279
14	-536.73	-558.852
15	-12.2482	80.00674

(Unit is $)

7. Conclusion

In this research, we apply the DACE based infinite horizon ADP algorithm proposed in Chen et al (2017a) to solve a large-scale, high-dimensional, infinite horizon, EV charging station control problem over a continuous state and decision space. In this system, renewable energy as resources provide electricity as well as the main power grid. In order to solve this control problem, first, we formulate it as a MDP problem, and then use MMFE and SVR forecasting models to formulate the transition model for the state variables such as wind power, solar power and EMP. For the wind speed, temperature and load profile, we use fixed trajectory from historical data to execute the transition. Therefore, with five stations, there are 31 state variables in the state space, which indicates this is a much higher dimensional DP problem than previous research. For such large-scale DP problems, computational time is a significant challenge. Following the research in Chen et al. (2017a and 2017b), we take advantage of the 45-degree line correspondence stopping criterion to stop the DP iterations after quantifying its parameters. The selected 137^{th} ADP policy presents a significantly better performance than a greedy myopic policy as shown in Table 5. In addition, from Fig 6, and Fig. 7, the ADP policy performs the way as we expected: when the EMP is low, the stations start to charge their batteries, and when the EMP is high, stored electricity in the batteries will be sold back to the main grid. Furthermore, we show that policies from the 137^{th} and the 200^{th} iterations have the same performance statistically, which indicates the quantified 45-degree line correspondence stopping rule is able to stop the DP iteration reasonably and select a good solution. In the future, multicollinearity among the state variables should be resolved, since this issue may result in transition models unstable. In this research, we make use of fixed trajectories from the historical data to represent temperature, wind speed and load profile when building the aFVF. In the future, a related transitioned model should be developed.

Acknowledgement

This research project was supported in part by National Science Foundation Grant ECCS-1128871.

Reference

Ashtari, A., Bibeau, E., Shahidinejad, S., Molinski, T. 2012. PEV charging profile prediction and analysis based on vehicle usage data. IEEE Transactions on Smart Grid, 3(1): 341-350.

Badawy, M. O., Sozer, Y. 2017. Power flow management of a grid tied PV-battery system for electric vehicles charging. IEEE Transactions on Industry Applications. 53(2): 1347-1357.

Boaro, M., Fuselli, D., De Angelis, F., Liu, D., Wei, Q., Piazza, F. 2013. Adaptive Dynamic Programming Algorithm for Renewable Energy Scheduling and Battery Management. Cognitive Computation. 5(2): 264-277.

Chen, V. C. P., Ruppert, D., Shoemaker, C. A. 1999. Applying experimental design and regression splines to high-dimensional continuous-state stochastic dynamic programming. Operations Research, 47(1): 38-53.

Chen, Y., Li, H., Jin, K., Song, Q. 2013. Wind farm layout optimization using genetic algorithm with different hub height wind turbines, Energy Conversion and Management, Vol. (70) 56-65.

Chen, Y., Liu, F., Kulvanitchaiyanunt, A., Chen, V. C. P., Rosenberger, J. 2017(a). Infinite Horizon Approximate Dynamic Programming Using Computer Experiments. COSMOS 17-02, University of Texas at Arlington.

Chen, Y., Liu, F., Chen, V. C. P., Rosenberger, J. 2017(b). Application of Support Vector Regression into Stochastic Infinite Horizon Dynamic Program. COSMOS 17-03, University of Texas at Arlington.

Clement-Nyns, K., Haesen, E., Driesen, J. 2010. The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. IEEE Transactions on Power Systems, 25(1): 371-380.

Cervellera, C., Chen, V. C. P., Wen, A. 2006. Optimization of a large-scale water reservoir network by stochastic dynamic programming with efficient state space discretization. European Journal of Operational Research, 171(3): 1139-1151.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., Vapnik, V. N. 1997. Support Vector Regression Machines, Advances in Neural Information Processing Systems 9, NIPS 1996, 155–161, MIT Press.

De Brabanter, K., Karsmakers, P., Ojeda, F., Alzate, C., De Brabanter, J., Pelckmans, K., De Moor, B., Vandewalle, J., Suykens, J. A. K. 2011. LS-SVM lab Toolbox User's Guide version 1.8. ESAT-SISTA Technical Report 10-146, August.

Ernst, D., Glavic, M., Capitanescu, F., Wehenkel, L. 2009. Reinforcement learning versus model predictive control: A comparison on a power system problem. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics; 39(2): 517-529.

Fell, K., Huber, K., Zink, B., Kalisch, R., Forfia, D., Hazelwood, D., Dang, N., Gionet, D., Musto, M., Johnson, W. 2010. Assessment of plug-in electric vehicle integration with ISO/RTO systems, KEMA, Inc. and ISO/RTO Council.

Fan, H., Tarun, P. K., Chen, V. C. P. 2013. Adaptive value function approximation for continuous-state stochastic dynamic programming. Computers & Operations Research, 40(4): 1076-1084.

Gan, L., Topcu, U., Low, S. H. 2013. Optimal decentralized protocol for electric vehicle charging. IEEE Transactions on Power Systems. 28(2): 940-951.

Guo, Y., Hu, J., Su, W. 2014. Stochastic optimization for economic operation of plug-in electric vehicle charging stations at a municipal parking deck integrated with on-site renewable energy generation, in Transportation Electrification Conference and Expo (ITEC), 2014 IEEE: 1-6.

Halton, J. H. 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik, 2: pp. 84-90.

Heath, D., Jackson, P. 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. IIE Transaction, 26(3): 17-30.

Jiang, D. R., Powell, W. B. 2015. Optimal hour-ahead bidding in the real-time electricity market with battery storage using approximate dynamic programming. INFORMS Journal on Computing. 27(3): 525-543.

Khodayar, M. E., Lei, W., Shahidehpour, M. 2012. Hourly coordination of electric vehicle operation and volatile wind power generation in SCUC, IEEE Transactions on Smart Grid: 3(3): 1271-1279.

Khosrojerdi, A., Xiao, M., Sariprueck, P., Allen, J., Mistree, F. 2013. Designing a system of plug-in hybrid electric vehicle charging stations, IDETC/CIE 2013, Portland, Oregon, USA, August 4-7.

Kulvanitchaiyanunt, A., Rosenberger, J., Lee, W., Chen, V., Sarikprueck, P. 2016. A Linear Program for Control of a System of PHEV Charging Stations, IEEE Transactions on Industry Applications. 52(3): 2046-2052.

Kutner, M. H., Nachtsheim, C. J. Neter, J., Li, W. 2004. Applied Linear Statistical Models. Fifth Ed. McGraw-Hill Irwin.

Lee, J., Lee, J. H. 2004. Approximate dynamic programming strategies and their applicability for process control: a review and future directions. International Journal of Control, Automation, and Systems, 2(3): 263-278.

Marano, V., Rizzoni, G. 2008. Energy and economic evaluation of PHEVs and their interaction with renewable energy sources and the power grid. IEEE International Conference on Vehicular Electronics and Safety, 2008. ICVES 2008: 84-89.

Peng, X., Berseth, G., van de Panne, M. 2016. Terrain-adaptive locomotion skills using deep reinforcement learning. ACM Transactions on Graphics (TOG). 35(4): No. 81.

Powell, W. B. 2007. Approximate dynamic programming: solving the curses of dimensionality. John Wiley, New York.

Sarikprueck, P., Lee, W., Kulvanitchaiyanunt, A., Chen, V., Rosenberger, J. 2017. Bounds for optimal control of a regional plug-In electric vehicle charging station system. IEEE Industrial and Commercial Power Systems Technical Conference, Niagara Falls, ON, Canada, May 6-11.

Sarikprueck, P., Lee, W., Kulvanitchaiyanunt, A., Chen, V., Rosenberger, J. 2015. Novel Hybrid Market Price Forecasting Method With Data Clustering Techniques for EV Charging Station Application.. IEEE Transactions on Industry Applications. 51(3).

Sarikprueck, P. 2015. Forecasting of wind, PV generation, and market price for the optimal operations of the regional PEV charging stations, Ph.D. dissertation. University of Texas at Arlington.

Shaaban, M. F., Ismail, M., El-Saadany, E. F., Zhuang, W. 2014. Real-time PEV charging/discharging coordination in smart distribution systems. IEEE Transactions on Smart Grid, vol. 5, pp. 1797-1807.

Shi, G., Wei, Q., Liu, D. 2016. Optimization of electricity consumption in office buildings based on adaptive dynamic programming. Soft Computing. pp 1-11.

Sobol, I. M. 1967. The distribution of points in a cube and the approximate evaluation of integrals. USSR Computational Mathematics and Mathematical Physics, 7: 784-802.

Steen, D., Tuan, L., Carlson O., Bertling, L. 2012. Assessment of electric vehicle charging scenarios based on demographical data. IEEE Transactions on Smart Grid, 3(3):1457-1468.

Sun, L., Wang, X., Liu, W., Lin, Z., Wen, F., Ang, S. P. Salam, M. A. 2016. Optimisation model for power system restoration with support from electric vehicles employing battery swapping. IET Generation, Transmission & Distribution. 10(2): 771-779.

Suykens, J. A. K., Van Gestel, T., De Brahanter, J., De Moor, B., Vandewalle, J. 2002. Least squares support vector machines. World Scientific Pub. Co. Singapore.

Taha, H. 2003. Operations research: an introduction (seventh ed.), Prentice Hall, New Jersey.

Tang, Y., He, H., Ni, Z., Wen, J., Sui, X. 2014. Reactive power control of grid-connected wind farm based on adaptive dynamic programming. Neurocomputing. 125(11): 125-133.

Wang, M., Ismail, M., Shen, X., Serpedin, E., Qaraqe, K. 2015. Spatial and temporal online charging/discharging coordination for mobile PEVs. IEEE Transactions on Wireless Communications. 22(1):112-121.

Wei, Q., Liu, D., Shi, G. 2015. A Novel Dual Iterative Q-Learning Method for Optimal Battery Management in Smart Residential Environments. IEEE Transactions on Industrial Electronics: 64(4).

Wei. Q., Liu, D., Lewis, F. L., Liu, Y., Zhang, J. 2017. Mixed Iterative Adaptive Dynamic Programming for Optimal Battery Energy Control in Smart Residential Microgrids. IEEE Transaction on Industrial Electronics. 64(5): 4110-4120.

Xie, S., Zhong, W., Xie, K., Yu, R., Zhang, Y. 2016. Fair energy scheduling for vehicle-to-grid networks using adaptive dynamic programming. IEEE Transaction on Neural Networks and Learning Systems. 27(8): 1697-1707.

Yao, L., Lim, W. H., Tsai, T. S. 2017. A real-time charging scheme for demand response in electric vehicle parking station. IEEE Transactions on Smart Grid, 8(1): 52-62.

Yu, R., Zhong, W., Xie, S., Zhang, Y., Zhang, Y. 2016. QoS Differential scheduling in cognitive-radio-based smart grid networks: an adaptive dynamic programming approach. IEEE Transaction on Neural Networks and Learning Systems. 27(2): 435-443.

Yang, Z., Chen, V. C. P., Chang, M. E., Sattler, M. L., Wen, A. 2009. A decision making framework for ozone pollution control. Operations Research, 57(2): 484 498.

Zhu, Z., Lambotharan, S., Chin, W. H., Fan, Z. 2016. A mean field game theoretic approach to electric vehicles charging. *IEEE Access,* 4: 3501-3510.