Experimental Validation of Approximate Dynamic Programming Based Optimization and Convergence on Microgrid Applications

Avijit Das[1], Zhen Ni[1,*], Xiangnan Zhong[1], Di Wu[2]
[1] Dept. of Computer, Electrical Engineering and Computer Science, Florida Atlantic University
[2] Energy and Environment Directorate, Pacific Northwest National Laboratory
[*] Correspondence: zhenni@fau.edu

Abstract—**Stochastic optimization can better model uncertainties in power system problems. However, when state space and action space become large, many existing approaches become computationally expensive and even infeasible to solve the problem. Approximate dynamic programming (ADP) attracts researchers' attention as a powerful tool for solving power system optimization problems with reduced computational cost. In this paper, in light of the existing literature, we investigate how the ADP approach with post-decision value function approximation converges to the nearly optimal solution with improved computational speed and experimentally validate the performance of the approach for a microgrid energy optimization problem. The approximation error versus the number of iteration is studied for convergence analysis of the post-decision ADP. A flowchart is provided to illustrate the proposed ADP algorithm for a microgrid energy optimization problem. The performance of ADP and dynamic programming (DP) is compared in terms of optimization error and computational time. It has found that the post-decision ADP approach can achieve competitive optimality with improved computational speed compared to the traditional DP.**

I. Introduction

In past decades, many optimization problems were solved as deterministic standard modeling frameworks developed in the fields of mathematical programming and optimal control [1]. However, most real-life problems contain uncertain parameters that are unknown at the time a decision is made. In this case, the deterministic optimization approaches may not ensure the optimal point. To solve this issue, stochastic optimization is prescribed involving uncertainties and probabilities [2].

In recent years, for stochastic optimization in power systems, scenario-based stochastic programming methods are reported in several papers [3], [4]. Other conventional stochastic programming techniques like backward dynamic programming, policy iteration, value iteration, etc. are also used [5], [6]. When the number of scenarios becomes large, the aforementioned techniques become computationally intractable and sometimes impose considerable computational cost. Though different scenario reduction techniques have been proposed to reduce the computational time cost of the existing scenario-based techniques [7], these scenario reduction techniques may overlook low probability but high-impact scenarios. The aforementioned situations are commonly referred as the "curse of dimensionality" [8]. Approximate dynamic programming (ADP) attracts a lot of researcher's attention and it approximates the optimal value with limited computational resources [9], [10]. The key idea of the ADP approach is to solve stochastic optimization problems based on the Bellman equation. In recent years, ADP based techniques are used to solve power system optimization problems considering the stochastic nature of the distributed energy sources [11], [12], [13].

Different from our previous works [13], [14], this paper aims to answer two new questions of ADP based optimization in microgrid energy systems. First, how ADP converges to the minimum operational cost of the microgrid system. Second, why ADP is computationally efficient in large-scale decision-making process. To answer these questions, we generalize the ADP approach with signal flowcharts and a probability-based sampling method. We study the convergence of the post-decision ADP value function for microgrid energy optimization problems and guarantee the optimality. The convergence analysis is presented with finite-time horizon fashion through an iterative process by introducing approximation error (AE) term in the value function to make it more realistic in power system optimization. A theorem is derived to show that the ADP approach can achieve the minimum operational cost of the microgrid after a finite number of iterations. For numerical studies, we use a stochastic energy optimization problem in an islanded microgrid with uncertain renewable energy sources. We analyze the performance of convergence in a stochastic environment. We examine the stochastic optimization performance (e.g., percentage of optimization error and computational time) of the ADP approach. The performance of the DP approach is also provided for reference.

The rest of this paper is organized as follows. The model description and problem formulation are presented in Section II. In Section III, the theoretical background of the post-decision ADP approach and its convergence proof are demonstrated. Simulation results and analysis are carried out in Section IV. Finally, the conclusions and future works are presented in

This work is supported in part by the U.S. Department of Energy, Office of Electricity through Energy Storage Program, and National Science Foundation under grant 1949921.

978-1-7281-5508-1/20/$31.00 ©2020 IEEE

Section V.

II. Model Description and Problem Formulation

According to [13], in the islanded microgrid, a wind turbine, a battery bank, and a diesel generator are considered as major power sources where the diesel generator serves as the backup power supply unit if the battery is not available in emergency cases. The system reference voltage and frequency are maintained by the battery bank and the diesel generator in accordance with the predetermined operation strategy. The wind turbine is used to meet the demand as well as to charge the battery. Charging and discharging modes of the battery depend on the state of charge of the battery. Considering the battery lifetime characteristics, the battery can discharge energy to a certain limit and needs to be charged when the battery state of charge goes below the minimum defined limit.

To formulate the optimization problem, a finite horizon of time is considered as $\tau = \{0, \Delta t, 2\Delta t, ..., T-\Delta t, T\}$, where $\Delta t = 1$ hour is the time step and $T = 24$ hours. At time instance t, the state variable of the power system can be written as

$$S_t = (B_t, W_t, D_t). \tag{1}$$

where B_t is the available energy in the battery in kWh, W_t is the available wind turbine power in kW, and D_t is the power demand in kW.

In this paper, five different actions are considered in a five-dimensional action set (decision vector) where the actions represent how much power is transferring from one unit to another unit. The action set at time t can be defined as

$$a_t = (a_t^{wd}, a_t^{gd}, a_t^{bd}, a_t^{wb}, a_t^{gb})^{\tau} \geq 0, a_t \in \chi_t, t \in \tau. \tag{2}$$

where a_t^{pq} represents the amount of transferred power from p to q at time t and χ_t is the feasible action space. The superscripts b, w, d, and g represent battery bank, wind turbine, power demand, and diesel generator, respectively. For example, a_t^{bd} represents an amount of power in kW transferring from the battery bank to the power demand subject to the operational constraints.

Transition Function for Exogenous Information: Transition function for exogenous information is used to determine the next-hour exogenous information values. In this paper, we have two exogenous inputs: W_t and D_t. Let the exogenous information E_t= (W_t,D_t) and the system state S_t = (B_t,E_t), where E_t is independent of B_t. The exogenous information transition can be expressed as, $E_{t+1} = E_t + e_{t+1}$. It means between time t and $t+1$, e_{t+1} = (w_{t+1},d_{t+1}) and e_{t+1} represents the change in E_t. The change in exogenous information e_{t+1} is independent of S_t and a_t.

Cost Functions: The operation cost of the battery can be calculated as

$$C_t^B = C_w p_t^B \Delta t \tag{3}$$

where C_w is the battery wear cost ($\$/kWh$) and p_t^B is the total amount of energy discharge from the battery at time t. The energy discharge from the battery p_t^B can represent as, $p_t^B = a_t^{bd}\lambda_{soc}$, where λ_{soc} is the effective weighting factor for the corresponding state of charge of the battery.

The operation cost of diesel generator can be written as

$$C_t^{gen} = C_t^{die-fuel} + C_{die-om} + C_{die-loss} \tag{4}$$

where $C_t^{die-fuel}$ is the fuel cost, C_{die-om} is operation and maintenance cost, and $C_{die-loss}$ is life loss cost of the diesel generator.

The Objective Function: The weighted cost function of the islanded microgrid is calculated by combining the cost function of the diesel generator and the battery. The operational cost of the microgrid at time t can be calculated as

$$C(S_t, a_t) = M_1 \times C_t^{gen} + M_2 \times C_t^B. \tag{5}$$

where C_t^{gen} and C_t^B represent the cost function of the diesel generator and battery at time t, respectively, and M_1 and M_2 are the weights. The weights of cost function can be determined based on the operator's objective. For example, if $M_1 = M_2 = 0.5$, two objectives are equally important to the operator. The overall system objective function can be expressed as

$$V = \min_{\pi} \mathbb{E}\Bigg[\sum_{t=0}^{T} C(S_t, \Pi_t^{\pi}(S_t))\Bigg]. \tag{6}$$

where $\mathbb{E}[.]$ is the expectation operator, $\Pi_t^{\pi}(S_t)$ represents the decision function, and π represents the type of policy. The system objective function is considered over a finite horizon of time. The goal is to find a proper action set $a_t = arg \min_{a_t \in \chi_t} V$, so that the overall system objective function V can be minimized over time.

III. Approximate Dynamic Programming for Solving Optimization Problem

A. Theoretical Background of Approximate Dynamic Programming

The Bellman's equation of optimality can be written as [15], [8],

$$V_t^*(S_t) = \min_{a_t \in \chi_t} [C(S_t, a_t) + \sum_{S_{t+\Delta t}} P_t(S_{t+\Delta t}|S_t, a_t) V_{t+\Delta t}^*(S_{t+\Delta t})], \tag{7}$$

where $P_t(S_{t+\Delta t}|S_t, a_t)$ is known as conditional transition probability for the decision a_t. The conditional transition probability is the probability of moving from current state S_t to the next state $S_{t+\Delta t}$. Here, $V_t^*(S_t)$ represents the optimal value function for the state S_t at time t.

Solving stochastic optimization problems efficiently with multi-dimensional state, action, and information spaces are usually challenging using traditional DP approaches. To overcome this challenge, the Bellman's equation can be rewritten using the post-decision state as,

$$V_t^*(S_t) = \min_{a_t \in \chi_t} [C(S_t, a_t) + V_t^a(S_t^a)]. \tag{8}$$

where S_t^a is the post-decision state which includes the information of the current hour state and decision set before

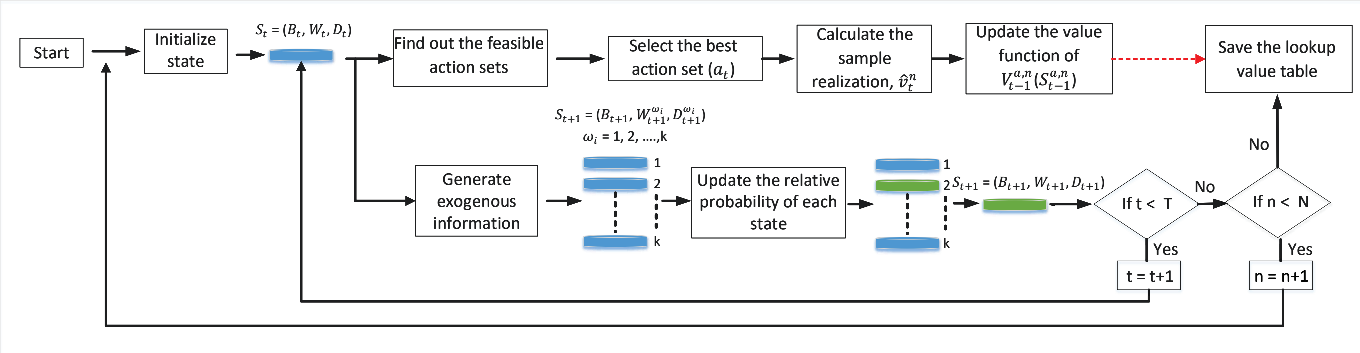

Fig. 1. The ADP flowchart for power system optimization. There are two stages in the flowchart. In the upper stage, the system calculates the sample realization, and updates the value function as well as stores in the data storage. In the lower stage, the system uses the present-hour state information to find out the next-hour state information using the probability and sends the state information to the next time-step.

arriving any new exogenous information. Using the post-decision state variable, the future expected value function can be replaced by a post-decision value function. The post-decision state consists of variables to compute the transition dynamics. In other words, a variable can be a part of the post-decision state if and only if its next-hour value is dependent on its current hour value.

The value function of the post-decision state $V_t^a(S_t^a)$ can be expressed as

$$V_{t-1}^a(S_{t-1}^a) = \mathbb{E}\{V_t^*(S_t)|S_{t-1}^a\}, \tag{9}$$

A sample realization of the value function at time t can be written as

$$\hat{v}_t{}^n = \min_{a_t \in \chi_t} [C(S_t, a_t) + V_t^{a,n-1}(S^{M,a}(S_t, a_t))], \tag{10}$$

where n represents the number of iteration. In any n iteration, the sample realization of the value function at time t represents the operational cost of the microgrid based on the prior knowledge of the post-decision value function approximation (VFA) at $(n-1)$ iteration. Using $\hat{v}_t{}^n$, the VFA of the post-decision state can be calculated as

$$V_{t-1}^{a,n}(S_{t-1}^a) = (1-\alpha_{n-1})V_{t-1}^{a,n-1}(S_{t-1}^a) + \alpha_{n-1}\hat{v}_t{}^n, \tag{11}$$

where α can be defined as a "step-size", "learning rate" or "stochastic filter", and generally takes on values between 0 and 1. The value function approximation is used to be smoothed by the step-size α over time.

The flowchart of the ADP approach is illustrated in Fig. 1. As can be seen, the algorithm starts by initializing the state information. Then, the state information goes through two different procedures. The upper level updates the post-decision value function, and the lower level determines the future hour state information.

- At the upper level, the state information is used to find the possible number of action sets that satisfy the operational constraints and the battery control strategy. Next, the system finds an action set that minimizes the cost function. The system sends the information of the cost function to the next block for determining the sample realization of the value function. Then, the system updates the post-decision value function using the sample realization and stores the data.
- At the lower level, the system sends the current hour state information to generate the next-hour exogenous information using a first-order Markov chain method. In this block, the system learns the noise information for the wind power and load demands (stochastic wind power and load demand models are described in Section IV), and generates a possible number of sample paths ω_i for the next-time step. At each time step, the next-hour battery information is updated using the battery transition function and the next-hour battery information is the same for all the possible next-hour states. Then, the relative probability of each state is calculated by the probability density function. Next, the state with the highest relative probability is selected as the next-hour state.

B. Convergence Study of the Post-Decision ADP Approach

The post-decision ADP approach is an iterative process. At each iteration, the sample realization of the value function is updated with the post-decision VFA of the previous iteration. The VFA process causes AEs in the value function. The algorithm is to reach the optimal value (daily operational cost of the microgrid) of each time step t by minimizing the AEs through the iterative process. Adding the AE, the equation (10) can be rewritten as

$$\hat{v}_t{}^{n+1} = \min_{a_t \in \chi_t} [C(S_t, a_t) + V_t^{a,n}(S^{M,a}(S_t, a_t))] + \epsilon^n(S_t). \tag{12}$$

where $\epsilon^n(.)$ represents the AE and $V_t^{a,n}(.)$ represents the post-decision VFA at the nth iteration.

To analyze the convergence of sequence for value function $\{\hat{v}_t{}^n\}_{n=0}^{\infty}$, we follow [12], [16] and define the bounding sequences as

$$\overline{v_t}^{n+1} = \min_{a_t \in \chi_t} [C(S_t, a_t) + \mu C(S_t, 0) + \overline{V_t}^{a,n}(S^{M,a}(S_t, a_t))]. \tag{13}$$

$$\underline{v_t}^{n+1} = \min_{a_t \in \chi_t} [C(S_t, a_t) - \mu C(S_t, 0) + \underline{V_t}^{a,n}(S^{M,a}(S_t, a_t))]. \tag{14}$$

where $\overline{v_t}^n : \mathbb{R} \rightarrow \mathbb{R}_+$ and $\underline{v_t}^n : \mathbb{R} \rightarrow \mathbb{R}_+$. It is assumed that both bounding sequences are initiated from some $\overline{v_t}^0$ and $\underline{v_t}^0$. If the AE $\epsilon^n(S_t)$ is constrained with an upper bound then the following results can be obtained. Here, $\overline{V_t}^{a,n}(S^{M,a}(S_t, a_t))$ and $\underline{V_t}^{a,n}(S^{M,a}(S_t, a_t))$ represent the upper bound and the lower bound of the post-decision VFA at the nth iteration.

Lemma 1. *Assume* $|\epsilon^n(S_t)| \leq \mu C(S_t, 0)$, $\forall n \epsilon \mathbb{N}$ *with* $\mu \epsilon [0, 1)$. *If the value functions are in equations (11), (12), (13) and (14) initialized such that* $\underline{v_t}^0 \leq \hat{v}_t{}^0 \leq \overline{v_t}^0$, *then,* $\underline{v_t}^n \leq \hat{v}_t{}^n \leq \overline{v_t}^n$, $\forall n \epsilon \mathbb{N}$. *Moreover, if* $\underline{v_t}^0 = \hat{v}_t{}^0 = \overline{v_t}^0$, *then* $\underline{v_t}^n$ *and* $\overline{v_t}^n$ *are the greatest lower bound and the least upper bound of* $\hat{v}_t{}^n$, *respectively.*

Proof. The mathematical induction process can be used to prove the lemma. Initially it is assumed that $\underline{v_t}^0 \leq \hat{v}_t{}^0 \leq \overline{v_t}^0$. Let $\underline{V_t}^{a,n}(S^{M,a}(S_t, a_t)) \leq V_t^{a,n}(S^{M,a}(S_t, a_t)) \leq \overline{V_t}^{a,n}(S^{M,a}(S_t, a_t))$, hold for some iteration n. If the equation (13) is compared with equation (12) then we get $\hat{v}_t{}^{n+1} \leq \overline{v_t}^{n+1}$ because $\epsilon^n(S_t) \leq \mu C(S_t, 0)$ and $V_t^{a,n}(S^{M,a}(S_t, a_t)) \leq \overline{V_t}^{a,n}(S^{M,a}(S_t, a_t))$. Therefore, for all $n \epsilon \mathbb{N}$, $\hat{v}_t{}^n \leq \overline{v_t}^n$. Similarly, it can also be proved that $\underline{v_t}^n \leq \hat{v}_t{}^n$ for all $n \epsilon N$ comparing equation (14) and equation (12). The last part of the lemma can be proved assuming $\epsilon^n(S_t) = \mu C(S_t, 0)$, $\forall n$ which leads to $\hat{v}_t{}^n = \overline{v_t}^n$. Similarly, assumption of $\epsilon^n(S_t) = -\mu C(S_t, 0)$, $\forall n$ leads to $\hat{v}_t{}^n = \underline{v_t}^n$. □

Here, $\overline{v_t}^n(.)$ and $\underline{v_t}^n(.)$ can be seen as the value functions at the nth iteration for the cost functions

$$\overline{J}_t^* = \frac{1}{K} \sum_{k=1}^{K} [C(S_t(\omega^k), \Pi_t^{\pi^*}(S_t(\omega^k))) + \mu C(S_t(\omega^k), 0)], \tag{15}$$

$$\underline{J}_t^* = \frac{1}{K} \sum_{k=1}^{K} [C(S_t(\omega^k), \Pi_t^{\pi^*}(S_t(\omega^k))) - \mu C(S_t(\omega^k), 0)], \tag{16}$$

respectively, considering the recursive relations of the value functions in equations (13) and (14). Here, ω and K represent the sample path and the total number of different sample paths, respectively [11], [13].

Lemma 2. *The value functions presented in equations (13) and (14) converge to the optimal value functions (15) and (16), respectively, if they are initialized by the value functions* $\underline{v_t}^0$ *and* $\overline{v_t}^0$ *such that* $0 \leq \underline{v_t}^0 \leq (1-\mu)C(S_t, 0)$, *and* $0 \leq \overline{v_t}^0 \leq (1+\mu)C(S_t, 0)$, *with* $\mu \epsilon [0, 1)$ *[17].*

Considering the above conditions, the following theorems prove the convergence of the value function $\{\hat{v}_t{}^n\}_{n=0}^{\infty}$ using the approximate value iteration (AVI).

Theorem 1. *Assume* $|\epsilon^n(S_t)| \leq \mu C(S_t, 0)$, $\forall n \epsilon \mathbb{N}$ *with* $\mu \epsilon [0, 1)$. *If the AVI presented in equation (12) is initialized such that* $0 \leq \hat{v}_t{}^0 \leq (1-\mu)C(S_t, 0)$, *then, the greatest lower bound of* $\hat{v}_t{}^n$ *converges to* $\underline{v}_t^*$ *and the least upper bound of* $\hat{v}_t{}^n$ *converges to* $\overline{v}_t^*$ *as* $n \rightarrow \infty$. *Here,* $\overline{v}_t^*$ *and* $\underline{v}_t^*$ *are the optimal value functions of the cost functions (15) and (16).*

Proof. The boundedness of $\{\hat{v}_t{}^n\}_{n=0}^{\infty}$ can be proved based on the Lemma 1 and the convergence of the bounds $\underline{v_t}^0$ and $\overline{v_t}^0$ which obeys $0 \leq \underline{v_t}^0 = \hat{v}_t{}^0 = \overline{v_t}^0 \leq (1-\mu)C(S_t, 0)$, can be proved by the Lemma 2. □

Theorem 2. *Let* $|\epsilon^n(S_t)| \leq \mu C(S_t, 0)$, $\forall n \epsilon \mathbb{N}$ *with* $\mu \epsilon [0, 1)$. *If the AVI presented in equation (12) is initialized such that* $0 \leq \hat{v}_t{}^0 \leq (1-\mu)C(S_t, 0)$, *then, the greatest lower bound and the least upper bound of* $\hat{v}_t{}^n$ *for* $n \rightarrow \infty$ *converge to the optimal value function at time* t *as* $\mu \rightarrow 0$.

Proof. If the optimal value function at time t is V_t^*, then $\tilde{V}_t^*$ can be defined as

$$\tilde{V}_t^* := \frac{1}{K} \sum_{k=1}^{K} C(S_t^{\pi^*}(\omega^k), 0), \tag{17}$$

Here, the summation in equation (17) is evaluated with the optimal trajectory of the optimal value function V_t^*. So,

$$V_t^* \leq \overline{v}_t^*, \tag{18}$$

According to the definition of $\overline{v}_t^*$, the relationship between the value functions with $\tilde{V}_t^*$ can be written as

$$\overline{v}_t^* \leq V_t^* + \mu \tilde{V}_t^*, \tag{19}$$

Using inequalities in equations (18) and (19),

$$|V_t^* - \overline{v}_t^*| \leq \mu \tilde{V}_t^*. \tag{20}$$

Since the state vectors are all finite, we can assume $\tilde{V}_{t,max}^* := sup \ \tilde{V}_t^*$ where $\tilde{V}_{t,max}^*$ is a finite constant. Therefore, the equation (20) can be rewritten as

$$|V_t^* - \overline{v}_t^*| \leq \mu \tilde{V}_{t,\max}^*. \tag{21}$$

According to equation (21), the value function $\overline{v}_t^*$ converges to the optimal value function V_t^* as $\mu \rightarrow 0$. Similarly, the convergence proof can also be shown for the value function $\underline{v}_t^*$ [18]. □

According to the convergence analysis, the daily operational cost of the microgrid calculated by the post-decision ADP approach should converge to the optimal value.

IV. Simulation Results and Analysis

For stochastic analysis, the first-order Markov chain process is used to generate noises. Pseudonormal probability distribution function is used to calculate the probability density function for each sample of the exogenous information.

The stochastic power system demand function is provided as $D_{t+1} = \min\{\max\{D_t + \mu^D, D_{\min}\}, D_{\max}\}$, where μ^D represents noise of the system. The range of the μ^D is defined as $\{0, \pm1, \pm2\}$ with an interval of 1. For the stochastic load demand, the pseudonormal probability density function is used as $N(0, 2^2)$ where the mean value and the variance value are 0 and 2, respectively. Similarly, the stochastic wind power function can be written as $W_{t+1} = \min\{\max\{W_t + \mu^W, W_{\min}\}, W_{\max}\}$, where μ^W represents noise of the system. The range of the μ^W is defined as $\{0, \pm1, \pm2, \pm3\}$ with the interval of 1. For the stochastic wind power, the

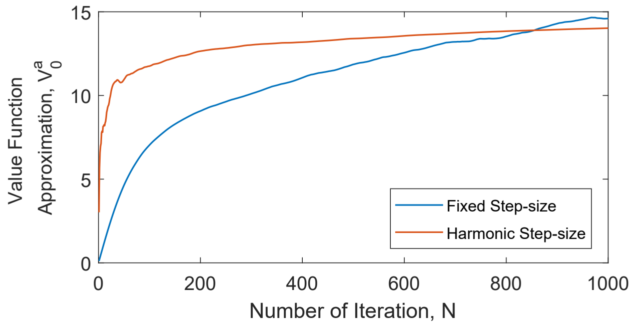

Fig. 2. The value of the VFA obtained from the ADP approach for different step-sizes after 1000 iterations.

TABLE I
PERCENTAGE OF OPTIMIZATION ERROR OF THE ADP APPROACH

Step-size (α)	Percentage of Optimization Error (%)
Fixed step-size	4.56
Harmonic step-size	3.07

pseudonormal probability density function is used as $N(0, 4^2)$. All the parameters of the islanded microgrid power supply units and the load profiles of the island, are taken from [13].

The step-size α plays an important role in the VFA process. In literature [8], different types of step-size (stochastic filter) are described. In this paper, we investigate two step-sizes such as fixed step-size and harmonic step-size. For the fixed step-size, the value of α sets as 0.2. For the harmonic step-size, we have used an equation as $\frac{a}{a+n}$ where a is the tuning parameter and is set to 5 in this study.

To test the performance of the ADP approach, we obtain the value table after $N = 1000$ iterations. We generate 500 test scenarios and solve the optimization problem for each scenario using the lookup table obtained from the ADP approach. Later, we calculate the statistical mean and obtain the value function. Similarly, we solve the optimization problem for each of the scenarios using the DP approach, calculate the statistical mean, and use the value to quantify the assessment of the ADP approach. The percentage of optimization error (OE) is calculated as

$$OE = \left| \frac{\bar{V}^N - \bar{V}^*}{\bar{V}^*} \right| \times 100\%. \tag{22}$$

where $\bar{V}^N$ and $\bar{V}^*$ represent the statistical mean value of the ADP approach and the DP approach, respectively. All the simulations are conducted in a computer with a configuration of $2.60GHz$ Intel Core $i7-6700HQ$ CPU and $8GB$ memory.

VFAs of the ADP approach versus the number of iteration using different step-size methods are plotted in Fig. 2. As can be seen, the harmonic step-size method provides a higher convergence rate than the fixed step-size method. For the 500 test scenarios studied using the ADP approach, the mean value of total microgrid operational cost with fixed and harmonic VFAs are 98.90 and 97.49, respectively. The 500 test scenarios are also studied using the DP approach, and the mean value of the cost is 94.59. The optimization error of the ADP can be calculated using (22) and results are provided in Table I. It takes 1059.61 seconds on average for training and testing using the ADP approach while the DP approach takes 1712.37 seconds to generate the statistical estimated value. The ADP approach is promising for the stochastic energy optimization of a microgrid.

V. CONCLUSION

This paper studies the convergence analysis of the post-decision ADP approach for solving time-dependent, finite-horizon stochastic microgrid energy optimization problem, where there uncertainties in wind power generation and load demand. The performance of the ADP approach is compared with the traditional DP approach in terms of result accuracy and computational time. It has found that the ADP approach is an efficient tool for stochastic optimization in smart grid.

REFERENCES

[1] W. B. Powell and S. Meisel, "Tutorial on stochastic optimization in energy—part I: Modeling and policies," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1459–1467, 2016.

[2] E. C. Finardi, B. U. Decker, and V. L. de Matos, "An introductory tutorial on stochastic programming using a long-term hydrothermal scheduling problem," *J. Contr. Auto. Elec. Sys.*, vol. 24, no. 3, pp. 361–376, 2013.

[3] W. Su, J. Wang, and J. Roh, "Stochastic energy scheduling in microgrids with intermittent renewable energy resources," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1876–1883, 2014.

[4] S. Mohammadi, S. Soleymani, and B. Mozafari, "Scenario-based stochastic operation management of microgrid including wind, photovoltaic, micro-turbine, fuel cell and energy storage devices," *Int. J. Elec. Power Energy Sys.*, vol. 54, pp. 525–535, 2014.

[5] R. Sioshansi, S. H. Madaeni, and P. Denholm, "A dynamic programming approach to estimate the capacity value of energy storage," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 395–403, 2014.

[6] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[7] J. Dupačová, N. Gröwe-Kuska, and W. Römisch, "Scenario reduction in stochastic programming," *Mathematical programming*, vol. 95, no. 3, pp. 493–511, 2003.

[8] W. B. Powell, *Approximate Dynamic Programming: Solving the curses of dimensionality*. John Wiley & Sons, 2007, vol. 703.

[9] J. Si, A. Barto, W. Powell, and D. Wunsch, Eds., *Handbook of learning and approximate dynamic programming*. John Wiley & Sons, 2004.

[10] Z. Ni, Y. Tang, X. Sui, H. He, and J. Wen, "An adaptive neuro-control approach for multi-machine power systems," *Int. J. Elec. Power Energy Sys.*, vol. 75, pp. 108–116, 2016.

[11] D. F. Salas and W. B. Powell, "Benchmarking a scalable approximate dynamic programming algorithm for stochastic control of multidimensional energy storage problems," *Dept Oper Res Financial Eng*, 2013.

[12] D. R. Jiang and W. B. Powell, "An approximate dynamic programming algorithm for monotone value functions," *Operations Research*, vol. 63, no. 6, pp. 1489–1511, 2015.

[13] A. Das and Z. Ni, "A computationally efficient optimization approach for battery systems in islanded microgrid," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6489–6499, 2017.

[14] A. Das, Z. Ni, and X. Zhong, "Near optimal control for microgrid energy systems considering battery lifetime characteristics," in *IEEE Sym. Series Comput. Intell.* IEEE, 2016, pp. 1–8.

[15] R. E. Bellman, *Dynamic programming*. Princeton Univ. Press, Princeton, NJ, USA, 1957.

[16] D. Liu and Q. Wei, "Finite-approximation-error-based optimal control approach for discrete-time nonlinear systems," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 779–789, 2013.

[17] A. Heydari, "Revisiting approximate dynamic programming and its convergence," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2733–2743, 2014.

[18] ——, "Theoretical and numerical analysis of approximate dynamic programming with approximation errors," *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 2, pp. 301–311, 2015.