A Novel Approach to Control of Autonomous Microgrid Systems

M. I. Abouheaf, M. S. Mahmoud, S. Azher Hussain

Systems Engineering Department, KFUPM, P. O. 5067, Dhahran 31261, Saudi Arabia
abouheaf@kfupm.edu.sa; msmahmoud@kfupm.edu.sa; g201105450@kfupm.edu.sa

***Abstract*-In this paper, reinforcement learning techniques are proposed for the control of autonomous microgrids. A type of approximate dynamic programming method is used to solve the Bellman equation, namely heuristic dynamic programming. The proposed control strategy is based on actor-critic networks. The control strategy is designed using a dynamic model of islanded microgrids and makes use of an internal oscillator for frequency control. The proposed control technique is based on a value iterations algorithm which is implemented online. Using only partial knowledge of the microgrid dynamics, the simulation results showed that the proposed control technique stabilizes the system and is robust to the load disturbances.**

Keywords- Distributed Generation; Microgrid; Reinforcement Learning; Approximate Dynamic Programming, Heuristic Dynamic Programming; Actor-Critic Networks

I. INTRODUCTION

Distributed Generation (DG) units such as photovoltaic arrays, wind turbine generators, and micro turbines are increasingly used to reduce the cost of energy prices and the environmental problems. High DG penetration levels has brought about the concept of 'microgrid'. A microgrid (MG) is an integrated energy system consisting of loads, distribution grid, and DG units that can operate in [1]:

- a grid-connected mode,
- an islanded (autonomous) mode, and
- transition between the two modes.

Under normal conditions, when a microgrid operates in a grid-connected mode, each DG unit within the microgrid utilizes the well-known dq-current control strategy [2] to regulate its real/reactive power components. Autonomous operation of a microgrid, however, requires sophisticated control strategies and protection systems. Depending on the electrical proximity of the DG units and their dedicated loads, several topologies for microgrids can be defined, e.g., parallel connection, ring, and radial connection of DGs. Each DG unit within a microgrid is connected to the point of common coupling (PCC) where the dedicated loads are also connected [3]. When all the PCCs are along a transmission line with nonzero impedance between the PCCs, a radial configuration is obtained. It turns out that the control of the MG is a key aspect which aims towards the stable operation of the microgrid [4], and has been the focus of researchers over the past few years. During the islanded operation, the main task of a MG is to deliver quality power by regulating the output voltage. A MG with its own control structure should be able to regulate any disturbances in its load towards zero to ensure the stability of the system [5]. The dq–current control strategy for multiple DG units in an islanded microgrid, which is based on frequency/power and voltage/reactive power droop characteristics of each DG unit, is well known and extensively reported [2]-[27]. In this approach, each DG unit is equipped with two droop characteristics:

1) Frequency as a linear function of real power, and

2) Voltage magnitude as a linear function of reactive power.

Based on these droop characteristics, frequency is dominantly controlled by real power flow, and voltage magnitude is regulated by reactive power flow of the DG unit. This approach does not directly incorporate load dynamics in the control loop. Thus, large and/or fast load changes can result in poor dynamic response or even voltage/frequency instability. A control strategy for autonomous operation of a DG unit and its dedicated load is introduced in [6]. This method is intended for a fairly fixed load and cannot accommodate large perturbations in the load parameters.

Multilevel control [17] of MGs is extensively studied in the literature. It is widely used and consists of primary, secondary and tertiary control levels [18]. A pseudo-decentralized control architecture is proposed in [19] that can be used for the optimization of Wireless Communication Networks (WCNs) with the help of a Global Supervisory Control (GSC) and local controllers. A networked control scheme based on a system of systems is proposed for microgrids in [20]. A MG with multiple DG units is treated as system of systems, and an output feedback control scheme is applied. A communication network that is subjected to packet dropouts and delays is used for the application of control. A two-level coordinating control approach for islanded MG is presented in [21]. An MG with *n* parallel connections of DG units connected to a common load is considered

DOI: 10.5963/IJEE0505001

for the study, and parallel connections are conveniently controlled in a two level coordinating scheme as the MG forms an interconnected control system. Control of autonomous MGs with a local load is introduced in [6]. It develops a dynamic model of MG and presents a classical control approach to design the controller. A robust servomechanism controller for autonomous MGs is presented in [23]. This approach uses the same dynamic model developed in [6] and uses an optimal control design procedure to guarantee robust stability.

On another research front, reinforcement learning (RL) is concerned with how an agent can pick its actions in a dynamic environment to transit to new states in such a way that optimizes the sum of cumulative reward [7]. RL has been successfully employed for several difficult problems such as control of inverted pendulums and adaptive control of chemical processes. In the context of power systems, RL has been used in security and stability control, automatic generation control [32, 33, 34], voltage and reactive power control [35]. It is an area of machine learning that allows development of online algorithms to obtain solutions to problems related to optimal control for dynamic systems that are described by difference equations [8, 9]. It involves two techniques known as Value Iteration (VI) and Policy Iteration (PI) [10, 11]. Policy iteration and value iteration algorithms have been developed for continuous time systems in [12, 13, 14]. Reference [9] proposed Adaptive Dynamic Programming (ADP) to solve dynamic programming problems. There are several different types of ADP, namely Heuristic Dynamic Programming (HDP), Dual Heuristic Dynamic Programming (DHP), Action-Dependent HDP (ADHDP) and Action-Dependent Dual HDP (ADDHP) [8, 9]. Actor-critic networks are one type of RL method. The actor component applies actions or control policies to their environment, while the critic component assesses the values of these actions. Based on this assessment, the actor policy is updated at each learning step [10, 11].

In this paper, a novel approach for the control of MG using RL techniques is proposed. The dynamic model of an islanded MG proposed in [6] is adopted to carry out the research. An HDP algorithm based on actor-critic value iteration is used to develop the control of MGs [24]. A value assessment of the control actions is performed by the critic component so that the actor action is updated at each learning step [7]. Online learning algorithms are used in the implementation of control techniques. To the best of author's knowledge, this is the first time that RL techniques have been applied to the concepts of MGs.

The paper is organized as follows. Section 2 provides a brief background of the dynamic model of islanded microgrid used to carry out this study. Section 3 formulates the control algorithm based on an on-line adaptive learning algorithm. Section 4 presents simulation results to verify the performance of the proposed controller.

II. AUTONOMOUS MICROGRID SYSTEM

The autonomous or islanded mode of MG can be caused by network faults/failures in the utility grid due to scheduled maintenance [25, 26] and economical optimization and management constraints [1].

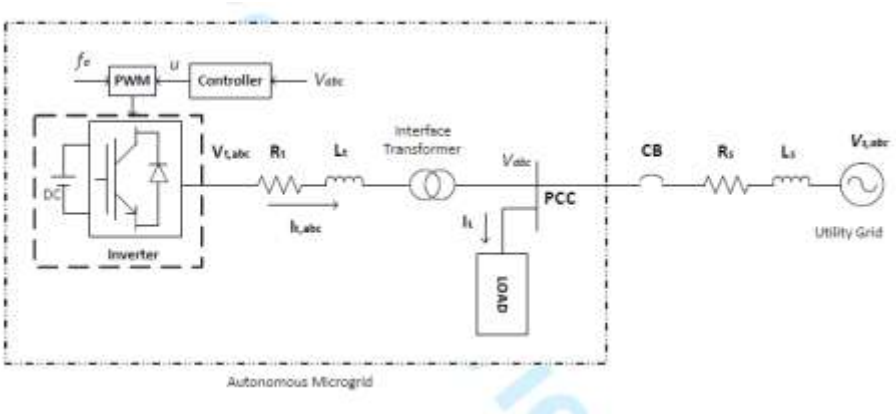

Fig. 1 Schematic diagram of microgrid

The schematic single-line diagram of an electronically coupled microgrid model is depicted in Fig. 1. A switch at a point of common coupling (PCC) will isolate the MG from the utility grid [27]. The islanded system consists of inverter based DG units supplying a load via series filters and step-up transformers. The dc voltage source represents the generating unit, and R_t and L_t represents the series filter. A local load, represented by a three-phase parallel RLC network, is connected at the PCC. The system parameters are given in Table 1, see the Appendix. During islanded operation, the main task of a MG is to deliver quality power by regulating any disturbances in the load. The MG with its own control structure should be able to maintain the load voltage level at a desired, prespecified set point.

A. State-Space Model

Consider the system described in Fig. 1. The dq standard state variables forms are given by [16]:

DOI: 10.5963/IJEE0505001

$$\frac{dI_{td}}{dt} = -\frac{R_t}{L_t}I_{t,d} + \omega_0 I_{tq} - \frac{1}{L_t}V_d + \frac{1}{L_t}V_{td} \quad (1)$$

$$\frac{dI_{tq}}{dt} = \omega_0 I_{td} - \frac{R_l}{L}I_{tq} - 2\omega_0 I_{Ld} + (\frac{R_l C\omega_0}{L} - \frac{\omega_0}{R})V_d \quad (2)$$

$$\frac{dI_{Ld}}{dt} = \omega_0 I_{tq} - \frac{R_l}{L}I_{Ld} + (\frac{1}{L} - \omega_0^2 C)V_d \quad (3)$$

$$\frac{dV_d}{dt} = \frac{1}{C}I_{td} - \frac{1}{C}I_{Ld} - \frac{1}{RC}V_d \quad (4)$$

$$V_{tq} = L_t[2\omega_0 I_{td} + (\frac{R_t}{L_t} - \frac{R_l}{L})I_{tq} - 2\omega_0 I_{Ld} + (\frac{R_l\omega_0 C}{L} - \frac{\omega_0}{R})V_d \quad (5)$$

Casting the foregoing autonomous MG system into the standard time state-space representation results in:

$$\dot{x}(t) = A_c x(t) + B_c u(t); \quad yt = C_c x(t); \quad u(t) = v_{td}.$$

It follows that the system matrices are given by:

$$A_c = \begin{bmatrix} -\frac{R_t}{L_t} & \omega_0 & 0 & \frac{1}{L_t} \\ \omega_0 & \frac{R_l}{L} & -2\omega_0 & \frac{R_l C\omega_0}{L} - \frac{\omega_0}{R} \\ 0 & \omega_0 & -\frac{R_l}{L} & \frac{1}{L} - \omega_0^2 C \\ \frac{1}{C} & 0 & -\frac{1}{C} & -\frac{1}{RC} \end{bmatrix}, \quad B_c = \begin{bmatrix} \frac{1}{L_t} \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad C_c^T = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (6)$$

where the state vector is

$$x^T = \begin{bmatrix} I_{td} & I_{tq} & I_{Ld} & V_d \end{bmatrix} \quad (7)$$

To facilitate further analytical development, model (6) will be discretized.

III. REINFORCEMENT LEARNING TECHNIQUES

Reinforcement Learning (RL) [7], also known as action-based learning, refers to interactions of an actor with its environment so as to improve its actions/control policies depending on the evaluative information received from the environment. RL can be implemented using an *Actor-Critic* architecture. The role of actor/critic is indicated before. The key feature of RL is that it provides an adaptive control which converges to the optimal control [29].

In the sequel, we will consider the Heuristic Dynamic Programming (HDP) algorithm to minimize a prescribed performance index. A simple HDP system consists of two sub-networks, namely, actor and critic networks. These networks have feedforward and feedback components.

A. Heuristic Dynamic Programming

Heuristic Dynamic Programming (HDP) is based on adaptive critics [30], which use value function approximation to solve dynamic programming problems. Fig 2 shows the structure of HDP design, which consists of a system to be controlled and two sub-networks, namely, *Actor* and *Critic* networks [31]. The control structure does not require the desired control signals to be known. Both the cost function and the control policy are approximated at each step by these two networks.

B. Discrete-Time Bellman Equation

Consider the following discrete-time system in state-space form:

$$x_{k+1} = Ax_k + Bu_k \quad (8)$$

which is an appropriate discrete version of system (6), where the states $x_k \in \mathbf{R}^n$ and control input $u_k \in \mathbf{R}^m$, and k is the discrete time index. It is assumed that system (8) is stabilizable on some set $\Omega \in \mathbf{R}^m$.

DOI: 10.5963/IJEE0505001

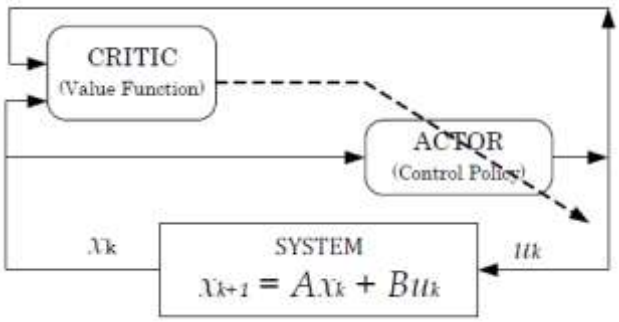

Fig. 2 Block diagram of actor-critic

Definition 1: Stabilizable System: A system is said to be stabilized on a set $\Omega \in \mathbf{R}^n$ *if there exists a control input* $u \in R^m$ *such that the closed loop system given by* x_{k+1} *is asymptotically stable on* Ω.

A function h (.): $\mathbf{R}^n \rightarrow \mathbf{R}^m$ from the state space to the control space is known as the control policy such that for every state x_k, there is a control $u_k = h(x_k)$. This describes the actor mathematically, as it is the one generating control policy in RL techniques. That is, the actor takes states x_k as input and gives control output u_k. It is desired to find the control policy $u(x_k)$ that minimizes the following performance measure/value function

$$V(x_k) = \sum_{j=k}^{\infty} \frac{1}{2}(x_j^T Q x_j + u_j^T R u_j), \tag{9}$$

where $0<Q_j=Q_j^T \in \mathbf{R}^{nxn}$ and $0<R_j=Rj^T \in \mathbf{R}^{mxm}$, so that the performance measure is well-defined.

Definition 2: Admissible Control [36]: A control policy $u_k = h(x_k)$ *is said to be admissible if it stabilizes system (8) and yields a finite performance* $V(x_k)$.

For any admissible control $u_k = h(x_k)$, $V(x_k)$ is known as cost or value and can be selected based on minimum energy, minimum cost requirements, etc. (9) can be written as follows:

$$V(x_k) = \frac{1}{2}(x_k^T Q x_k + u_k^T R u_k) + V(x_{k+1}), V(0) = 0 \tag{10}$$

Therefore, by using current control policy u_k the cost can be evaluated by solving the above difference equation. Strictly speaking, the Bellman equation is a functional equation consisting of dynamical systems state and a value or optimal return function.

According to Bellman's optimality principle [37], the optimal value and the optimal policy can be obtained by

$$\begin{aligned} V^*(x_k) &= \min_{u_k}[\frac{1}{2}(x_k^T Q x_k + u_k^T R u_k) + V^*(x_{k+1})] \\ u_k^* &= -R^{-1}B^T \nabla V^*(x_{k+1}) \end{aligned} \tag{11}$$

The key concept in developing RL techniques is to assess the current policy value by using the Bellman equation. To solve the above equation, one must know the policy at (k+1) to determine the state at k; therefore, the Bellman optimality equation is a dynamic programing scheme. In this paper, we are interested in value iteration techniques, an iterative method for determining optimal control. This technique does not require an initial stabilizing policy. It is significant to note that only partial microgrid dynamics are required.

C. Value Iteration Algorithm

In this section, an on-line value optimality iteration algorithm for autonomous microgrid systems is developed and used to solve the discrete-time Bellman equation (9). It can be considered as a simple backup operation that integrates policy improvement and truncated policy evaluation steps. The value iteration algorithm is summarized by **Algorithm 1** in the Appendix.

D. Actor-Critic Networks Implementation

The performance function (9) is now approximated by a critic network, and the control policy (11) is approximated by an actor network. Let $W_c \in \mathbf{R}^{nxn}$ and $W_a \in \mathbf{R}^{nxm}$ be the critic and actor weights, respectively. Therefore, the performance function and control policy approximations can be written as follows:

DOI: 10.5963/IJEE0505001

$$\hat{V}_k(W_c) = x_k^T W_c^T x_k \tag{12}$$

$$\hat{u}_k(W_a) = W_a^T x_k \tag{13}$$

Hence, the network approximation error of the actor is given as:

$$\zeta_{u_k}^{V(x_k)} = \hat{u}_k(W_a) - u_k \tag{14}$$

The control policy is given in terms of critic network, such that:

$$u_k = -R^{-1} B^T \nabla \hat{V}(x_{k+1}) \tag{15}$$

On expressing this target control in terms of critic weights, one obtains:

$$u_k = -R^{-1} B^T W_c^T x_k \tag{16}$$

The squared approximation error is given by:

$$\frac{1}{2}(\zeta_{u_k}^{V(x_k)})^T \zeta_{u_k}^{V(x_k)}.$$

The change in the actor weights is given by the gradient descent method. Therefore, the actor update rule is given as follows:

$$W_a^{(l+1)T} = W_a^{lT} - \lambda_a[(W_a^{lT} x_k - u_k^l)(x_k)^T] \tag{17}$$

where $0 < \lambda_a < 1$ is the actor learning rate. Let

$$\psi_{x_k}^{V(x_k)}.$$

be the target value of the critic network, and the value update is given by (11). Therefore, we have:

$$\psi_{x_k}^{V(x_k)} = \frac{1}{2}[(x_k^T Q x_k + u_k^{lT} R u_k^l)] + V^l(x_{k+1}) \tag{18}$$

The network approximation error of the critic is:

$$\zeta_{x_k}^{V(x_k)} = \psi_{x_k}^{V(x_k)} - \hat{V}_k(W_c).$$

Similarly, the squared approximation error is given by:

$$\frac{1}{2}(\zeta_{x_k}^{V(x_k)})^T \zeta_{x_k}^{V(x_k)}.$$

The change in the critic weights is given by the gradient descent method. Therefore, the critic update rule is given as follows:

$$W_c^{(l+1)T} = W_c^{lT} - \lambda_c[\psi_{x_k}^{V(x_k)} - x_k^T W_c^{lT} x_k] x_k x_k^T \tag{19}$$

where $0 < \lambda_c < 1$ is the critic learning rate.

Algorithm 2 in the appendix implements the actor-critic network solution foe value iteration of **Algorithm 1**.

Remark 1: It is important to note that the value iteration depends on the solution of the simply recursive equation (19), which is easy to compute and is called partial backup in reinforcement learning. Value iteration successfully mixes one sweep of policy evaluation and one sweep of policy improvement in each of its sweeps.

IV. ACTOR-CRITIC ONLINE IMPLEMENTATION

The following diagram is used for solving the microgrid control problem by *online* tuning of actor-critic networks. In this algorithm, we start with the given initial conditions for the system states. The algorithm makes use of real-time data measured along the system trajectories and tunes the actor-critic structure to generate the suitable control policy.

DOI: 10.5963/IJEE0505001

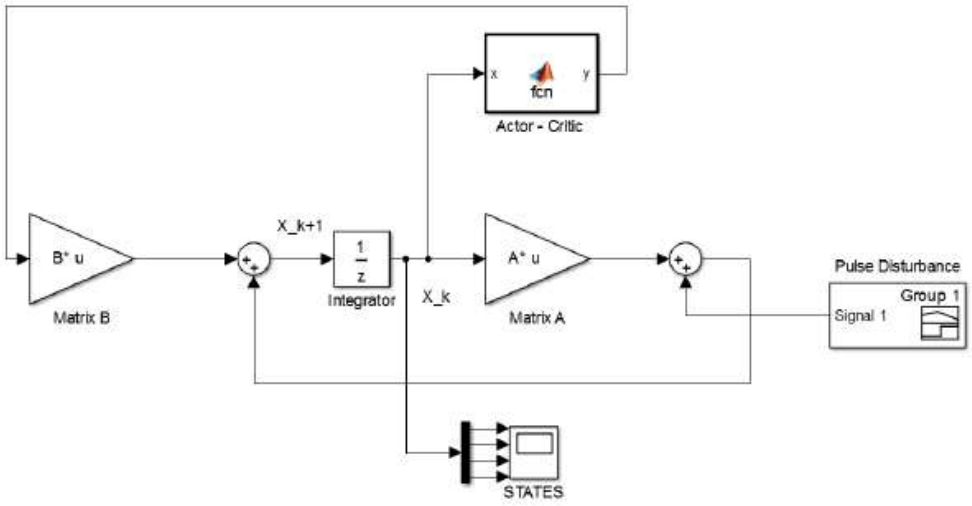

Fig. 3 Simulink blocks for algorithm-2 implementation

V. PERFORMANCE EVALUATION OF PROPOSED CONTROLLER

The parameters for the online **Algorithm 2** were chosen as $\lambda_a = 0.2$, $\lambda_c = 0.2$, $Q = I_{4x4}$ and R I. Fig. 3 describes the Simulink structure for implementation of **Algorithm 2** for the system. The generated control is fed online to the system, and at time t = 0.1 seconds, a pulse disturbance is introduced in the system states. Fig. 4 shows the response of all four states. By observing the figure, it can be concluded that the online **Algorithm 2** yields stability and proves synchronization of the weights.

DOI: 10.5963/IJEE0505001

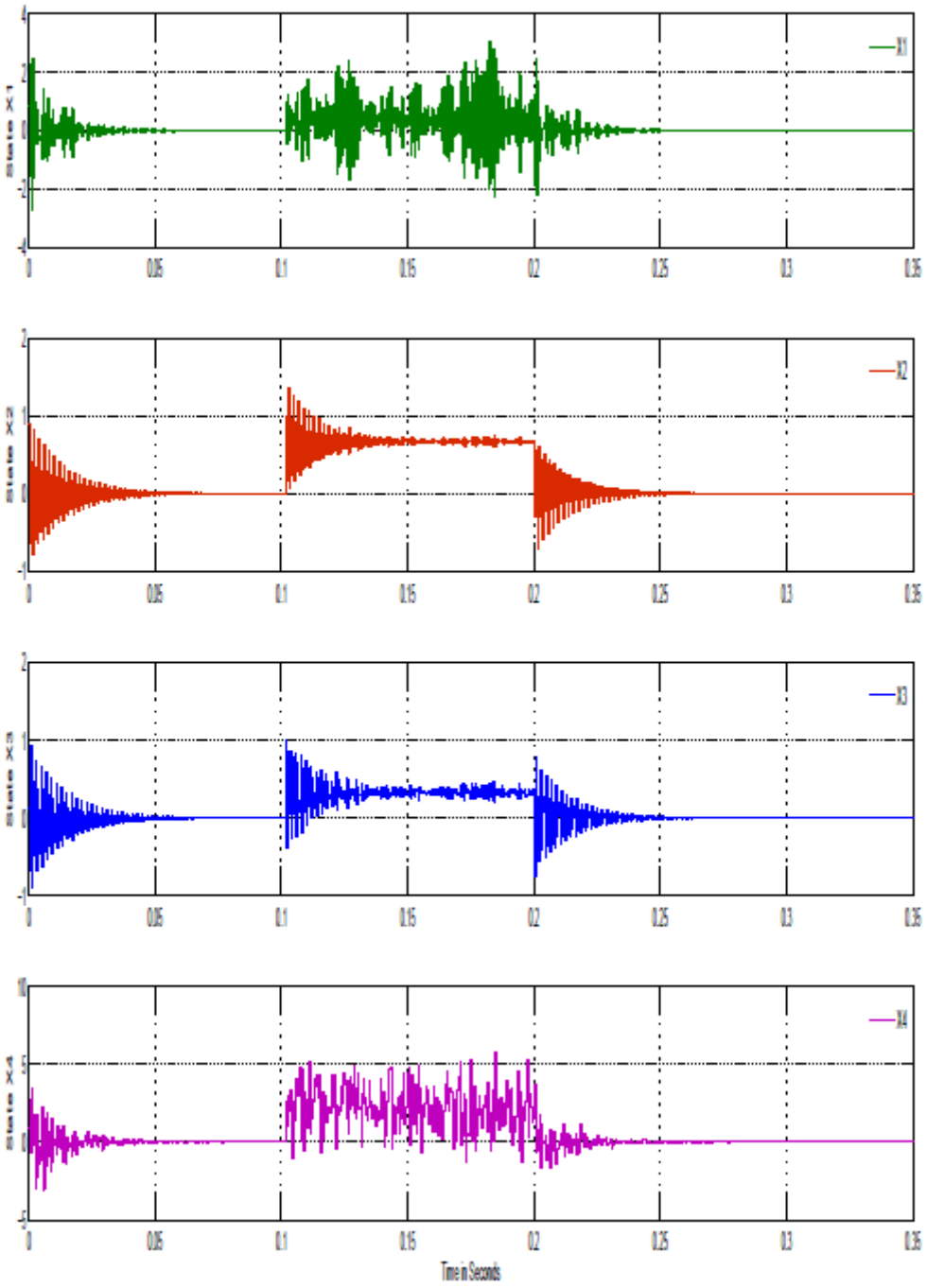

Fig. 4 Responses of the system states

Figs. 5-7 show the simulation results of **Algorithm 2**. Figs. 5 and 6 depict the tuning profile for the critic and actor weights, respectively. Fig. 7 shows the error dynamics of the microgrid system.

DOI: 10.5963/IJEE0505001

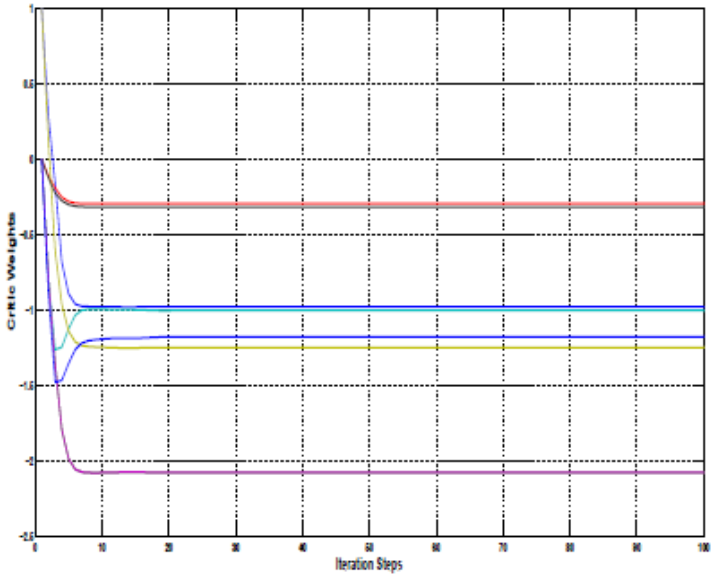

Fig. 5 On-line critic weights

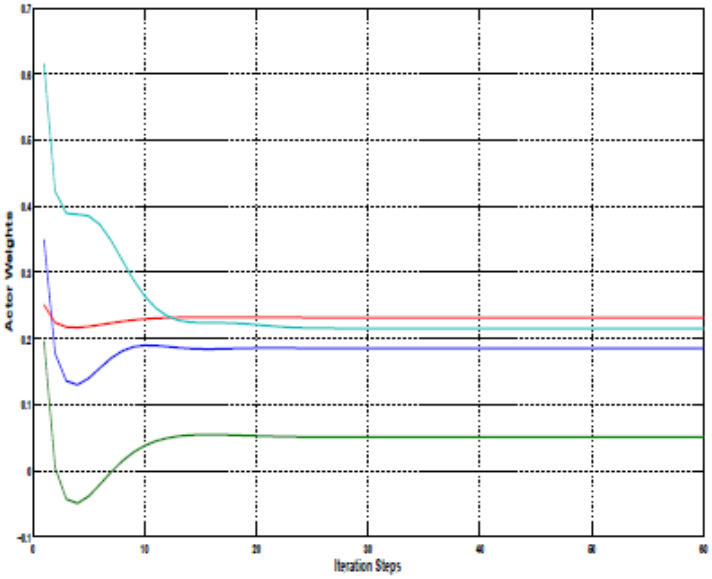

Fig. 6 On-line actor weights

DOI: 10.5963/IJEE0505001

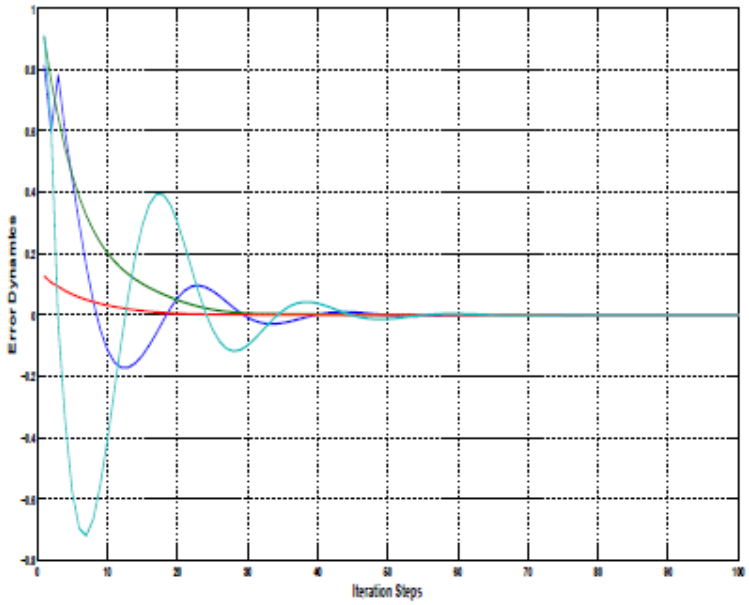

Fig. 7 On-line error dynamics

VI. CONCLUSIONS

A reinforcement learning technique for the control of autonomous microgrids based on heuristic dynamic programming is proposed. The strategy is based on a value iteration algorithm and is implemented using actor-critic networks. Based on this structure, an offline learning algorithm is developed to solve the bellman equation. From the simulation results, it is evident that the converging weights of the actor-critic system stabilizes the system and regulates the output voltage to a nominal value. The proposed control strategy is also robust against any disturbances in the states and load. Only partial knowledge about the dynamics is required. The input matrix is needed.

ACKNOWLEDGMENTS

The authors would like to thank the deanship for scientific research (DSR) at KFUPM for support through research group project **RG1316-1**.

REFERENCES

[1] F. Katiraei, M. R. Iravani, N. Hatziargyriou and A. Dimeas, "Microgrids management", *IEEE Trans. Power and Energy Magazine*, vol. 6(3), pp. 54-65, 2008.

[2] P. Piagi and R. H Lasseter. (2006), "Autonomous control of microgrids", *IEEE Power Engineering Society General Meeting*, 18-22 June, pp. 139-147, 2008.

[3] J. A, Peas Lopes, L. C., Moreira and A. G. Madureira, "Defining control strategies for microgrids islanded operation", *IEEE Trans. Power Systems*, vol. 21(2), pp. 916-924, 2006.

[4] M. S. Mahmoud, S., A. Hussain and M. A. Abido, "Modeling and control of microgrid: an overview", *J. the Franklin Institute*, vol. 351(5), pp. 2822-2859, 2014.

[5] A. M. Bouzid, et al, "A survey on control of electric power distributed generation systems for microgrid applications", *Renewable and Sustainable Energy Reviews*, vol. 44, pp. 751-766, 2015.

[6] H. Karimi, H., Nikkhajoei, and M. R. Iravani, "Control of an electronically-coupled distributed resource unit subsequent to an islanding event", *IEEE Trans. Power Delivery*, vol. 23, no. 1, 2008.

[7] R. S. Sutton and A. G. Barto (1998), Reinforcement Learning: An Introduction, MIT Press, Massachusetts, 1998.

[8] P. J. Werbos, "Neural networks for control and system identification", *IEEE Proc. Decision and Control*, 1989.

[9] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modelling", *Handbook of intelligent control: Neural, fuzzy, and adaptive approaches*, pp. 493-525, 1992.

[10] D. P. Bertsekas and J. N. Tsitsiklis, "Neuro-Dynamic Programming", Athena Scientific, Massachusetts, 1996.

[11] D. P. Bertsekas and J. N. Tsitsiklis, "Neuro-dynamic programming: an overview", *IEEE Proc. Decision and Control*, vol. 1, pp. 560-564, 1989.

[12] A. Al-Tamimi, F. L. Lewis and M. Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof", *IEEE Trans. Systems, Man, and Cybernetics*, vol. 38(4), pp. 943-949, 2008.

[13] D. Vrabie, D., et al, "Adaptive optimal control for continuous-time linear systems based on policy iteration", *Automatica*, vol. 45(2), pp.

DOI: 10.5963/IJEE0505001

477-484, 2009.

[14] K. G. Vamvoudakis and F. L. Lewis, "Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem", *Automatica*, vol. 46(5), pp. 878-888, 2010.

[15] P. J. Werbos, "A menu of designs for reinforcement learning over time", *Neural networks for control*, pp. 67-95, 1990.

[16] B. Lasseter, "Microgrids [distributed power generation]", *IEEE Power Engineering Society Winter Meeting*, vol. 1, pp. 146-149, 2001.

[17] M. S. Mahmoud, M. F. Hassan, and M. G. Darwish, *Large Scale Control Systems: Theories and Techniques*, Marcel Dekker Inc., New York, 1985.

[18] Y. R. Mohamed and A. A. Radwan, "Hierarchical control system for robust microgrid operation and seamless mode transfer in active distribution systems", *IEEE Trans. Smart Grid*, vol. 2(2), pp. 352-362, 2011.

[19] S. K. Mazumder, M. Tahir, and K. Acharya, "Pseudo-decentralized control-communication optimization framework for microgrid: a case illustration", *IEEE Proc. Transmission and Distribution Conference and Exposition (T & D)*, pp. 304-310, 2008.

[20] M. S. Mahmoud, M. S. Rahman, and F. M. AL-Sunni, "Networked control of microgrid system of systems", Int. J. Systems Science, DOI: 10.1080/00207721.2015.1005723, 2015.

[21] M. S. Mahmoud and O. Al-Buraiki, "Two-level control for improving the performance of microgrid in islanded mode", *IEEE Symposium Industrial Electronics (ISIE)*, June, 1-4, Istanbul, Turkey, pp. 254-259, 2014.

[22] H. Karimi, H. Nikkhajoei, and M. R. Iravani, "Control of an electronically-coupled distributed resource unit subsequent to an islanding event", *IEEE Trans. Power Delivery*, vol. 23, no. 1, pp. 493-501, 2008.

[23] H. Karimi, E. J. Davison, and M. R. Iravani, "Multivariable servomechanism controller for autonomous operation of a distributed generation unit: Design and performance evaluation", *IEEE Trans. Power Systems*, vol. 25(2), pp. 853-865, 2010.

[24] M. I. Abouheaf, "Multi-agent discrete-time graphical games and reinforcement learning solutions", *Automatica*, vol. 50(12), pp. 3038-3053, 2014.

[25] J. A. P. Lopes, C. L. Moreira, and A. G. Madureira, "Defining control strategies for microgrids islanded operation", IEEE *Trans. Power Systems*, vol. 21(2), pp. 916-924, 2006.

[26] H. Jiayi, J., Chuanwen, and X. Rong, "A review on distributed energy resources and microgrid", *Renewable and Sustainable Energy Reviews*, vol. 12(9), pp. 2472-2483, 2008.

[27] F. Katiraei and M.R. Iravani, "Power management Strategies for a microgrid with multiple distributed generation units", *IEEE Trans. Power Systems*, vol. 21, no.4, pp. 1821-1831, 2006.

[28] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control", *IEEE Circuits and Systems Magazine*, vol. 9(3), pp. 32-50, 2009.

[29] B. Widrow, N. K. Gupta, and S. Maitra, "Punish/reward: learning with a critic in adaptive threshold systems", *IEEE Trans. Systems, Man and Cybernetics*, vol. 5, pp. 455-465, 1973.

[30] A. G. Barto, G., Andrew, R. S. Sutton, and C. W. Anderson, "Neuron-like adaptive elements that can solve difficult learning control problems", *IEEE Trans. Systems, Man and Cybernetics*, vol. 5, pp. 834-846, 1983.

[31] J. Si, *Handbook of Learning and Approximate Dynamic Programming*, (2), John Wiley & Sons, New York, 2004.

[32] W. X. Liu, G. K. Venayagamoorthy, and D. C. Wunsch, "A heuristic–dynamic programming based power system stabilizer for a turbogenerator in a single–machine power system," *IEEE Trans. Industry Applications*, vol. 41(5), pp. 1377-1385, 2005.

[33] T. P. L. Ahmed, P. S., Nagendra Rao and P. S. Sastry, "A reinforcement learning approach to automatic generation control", *Electric Power Systems Research*, vol. 63(1), pp. 9-26, 2002.

[34] T. Yu and B. Zhou, "A novel self-tuning CPS controller based on Q-learning method", *Proc. IEEE PES General Meeting*, July 20-24, pp. 1-6, 2008.

[35] J. G. Vlachogiannis and N. D. Hatziargyriou, "Reinforcement learning for reactive power control", *IEEE Trans. Power Systems*, vol. 19(3), pp. 1317-1325, 2004.

[36] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach", *Automatica*, vol. 41(5), pp. 779-791, 2005.

[37] F. L. Lewis and V. L. Syrmos, *Optimal Control*, John Wiley & Sons, 1995.

DOI: 10.5963/IJEE0505001

APPENDIX

TABLE 1 SYSTEM PARAMETERS

QUANTITY	VALUE
R_t	1.5 m Ω
L_t	300 μ H
V_{dc}	1500 V
PWM Carrier Frequency	1980 Hz
LOAD PARAMETERS	
R	76 Ω
L	111.9 mH
C	62.855 μ F
R_l	0.3515 Ω
GRID PARAMETERS	
R_s	1 Ω
L_s	10 μ H
Nominal Frequency f_o	60 Hz
Nominal Voltage (rms)	13.8 kV
INTERFACE TRANSFORMER PARAMETERS	
Type	Wye/Delta
Rating	2.5 MVA
Voltage Ration	0.6/13.8 kV

Algorithm 1 (Value Iteration Algorithm for Autonomous Microgrid)

1) **Initialization:** *Select any arbitrary initial values for the policy* u_k *and* V (x_k), *not necessarily admissible or stabilizing.*

2) **Value Update:** *Solve the Bellman equation to get* $V^{\ell+1}$(xk) *as follows*

$$V^{\ell+1}(x_k) = \frac{1}{2}(x_k^T Q x_k + u_k^{\ell T} R u_k^{\ell}) + V^{\ell}(x_{k+1}) \tag{20}$$

where ℓ *is the iteration index.*

3) **Policy Improvement:** *The control policy* uk *is updated as follows*

$$u_k^{\ell+1} = -R^{-1}B^T \bigtriangledown V(x_{k+1})^{\ell+1} \tag{21}$$

where the gradient is defined as $\nabla V(x_{k+1}) = \dfrac{\partial V(x_{k+1})}{\partial x_{k+1}}$.

Algorithm 2 (Actor-Critic Online Implementation of Algorithm 1)

1) The weights of actor W_a and critic W_c are initialized randomly. Initializing $Wc = I_{4\times4}$ and $Wa = rand_{1\times4}$.

2) **Loop Iterations Begins**

- The iteration loop begins with l as iteration index
- Start with given initial values for the system states.
- The control policy $\hat{u}_k^l$
- k is evaluated using equation (13)
- The dynamics of the system x_{k+1}^l are evaluated using (8)
- The performance measure $\hat{V}^l(x_{k+1})$ is calculated using equation (12)
- The critic network is updated based on equation (19)
- The actor network is updated based on equation (17)

DOI: 10.5963/IJEE0505001

- On convergence of $\left\|\hat{V}(x_k)^{l+1} - \hat{V}(x_k)^{l}\right\|$, end loop.

Mohamad I. Abouheaf was born in Smanoud, Egypt. He received his B.Sc. and M.Sc. degrees in Electronics and Communication Engineering, Mansoura College of Engineering, Mansoura, Egypt (2000, 2006). He worked as an assistant lecturer with the Air Defense College, Alexandria, Egypt (2001–2002). He worked as a Planning Engineer for the maintenance department, Suez Oil Company (SUCO), South Sinai, Egypt, (2002–2004). He worked as an Assistant Lecturer with the Electrical Engineering Department, Aswan College of Energy Engineering, Aswan, Egypt (2004–2008). He received his Ph.D. degree in Electrical Engineering, University of Texas at Arlington (UTA), Arlington, Texas, USA (2012). He worked as a Postdoctoral Fellow with the University of Texas at Arlington Research Institute (UTARI), Fort Worth, Texas, USA (2012–2013). He worked as Adjunct Faculty with the Electrical Engineering Department, University of Texas at Arlington (UTA), Arlington, Texas, USA (2012–2013). He was a member of the Advanced Controls and Sensor Group (ACS) and the Energy Systems Research Center (ESRC), University of Texas at Arlington, Arlington, Texas, USA (2008–2012). Currently, he is Assistant Professor with the Systems Engineering Department, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia. His research interests include optimal control, adaptive control, reinforcement learning, fuzzy systems, game theory, microgrids, and economic dispatch.

Magdi S. Mahmoud obtained B. Sc. (Honors) in communication engineering, M. Sc. in electronic engineering and Ph. D. in systems engineering, all from Cairo University in 1968, 1972 and 1974, respectively. He has been a professor of engineering since 1984. He is now a Distinguished Professor at KFUPM, **Saudi Arabia.** He was on the faculty at different universities worldwide including **Egypt (CU, AUC)**, **Kuwait (KU), UAE (UAEU), UK (UMIST), USA (Pitt, Case Western), Singapore (Nanyang)** and **Australia (Adelaide)**. He lectured in **Venezuela (Caracas), Germany (Hanover), UK ((Kent), USA (UoSA), Canada (Montreal)** and **China (BIT, Yanshan)**.

He is the principal author of thirty-six (**36**) books, inclusive book-chapters and the author/co-author of more than **520** peer-reviewed papers. He is the recipient of two national, one regional and several university prizes for outstanding research in engineering and applied mathematics. He is Fellow, Institution of Electrical Engineers (IEE), UK, Senior Member, Institute of the Electrical and Electronic Engineers (IEEE), USA. Member, Council of Engineering Institutions, UK. Member, Egyptian Engineers Society, EGYPT, Member, Kuwaiti Engineers Society, KUWAIT. Member, Operations Research Society of Egypt, Egypt Member, Sigma XI, USA, Member, National Academy of Sciences, New York, USA.

He is currently actively engaged in teaching and research in the development of modern methodologies to distributed control and filtering, networked-control systems, triggering mechanisms in dynamical systems, fault-tolerant systems and information technology**.**

S. Azher Hussain was born in Hyderabad, India. He received a Bachelor of Engineering (BE) degree in Electrical & Electronics Engineering from Osmania University and Master of Science (M. Sc.) degree in Systems Engineering from King Fahd University of Petroleum and Minerals. His research interests include control of microgrid, neural networks and adaptive control.

DOI: 10.5963/IJEE0505001