A Reinforcement Learning Approach for Frequency Control of Inverted-Based Microgrids

Adibi, Mahya; van der Woude, Jacob

DOI
10.1016/j.ifacol.2019.08.164

Publication date
2019

Document Version
Final published version

Published in
IFAC-PapersOnLine

Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 52-4 (2019) 111–116

A Reinforcement Learning Approach for Frequency Control of Inverted-Based Microgrids ⋆

Mahya Adibi * **Jacob van der Woude** *

* *Delft Institute of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE, Delft, The Netherlands. (e-mail: {m.adibi, j.w.vanderwoude}@tudelft.nl).*

Abstract: In this paper, we present a reinforcement learning control scheme for optimal frequency synchronization in a lossy inverter-based microgrid. Compared to the existing methods in the literature, we relax the restrictions on the system, i.e. being a lossless microgrid, and the transmission lines and loads to have constant impedances. The proposed control scheme does not require a priori information about system parameters and can achieve frequency synchronization in the presence of dominantly resistive and/or inductive line and load impedances, model parameter uncertainties, time varying loads and disturbances. First, using Lyapunov theory a feedback control is formulated based on the unknown dynamics of the microgrid. Next, a performance function is defined based on cumulative rewards towards achieving convergence to the nominal frequency. The performance function is approximated by a critic neural network in real-time. An actor network is then simultaneously learning a parameterized approximation of the nonlinear dynamics and optimizing the approximated performance function obtained from the critic network. The performance of our control scheme is validated via simulation on a lossy microgrid case study in the presence of disturbances.

Keywords: reinforcement learning, microgrids, stability, frequency synchronization.

1. INTRODUCTION

A microgrid consists of loads, storage units and renewable energy sources. It forms a locally controllable system that can detach from the main grid and operate autonomously, Lasseter (2002), Guerrero et al. (2013). However, an imbalance between the generated power and the demand results in frequency instability. To regulate the frequency, primary droop controllers are widely employed, however, steady state deviations from the nominal frequency are observed due to load demand variations. Therefore, an additional control level, namely the secondary control, must be implemented to achieve the ultimate frequency regulation and power sharing; see Simpson-Porco et al. (2015), Guerrero et al. (2011), De Persis et al. (2016).

A conventional approach to deal with the frequency synchronization problem consists of using a primary droop controller enhanced by a secondary control scheme following the gain plus integral approach, e.g. Simpson-Porco et al. (2015). To deal with the uncertainties that impact a microgrid system and to further elevate the performance, more complex control frameworks have been designed, Ersdal et al. (2016), Zribi et al. (2005), Chang and Fu (1997), Dorfler and Grammatico (2017), De Persis and Monshizadeh (2018), Trip et al. (2018),Weitenberg et al. (2018), Adibi et al. (2017). In Ersdal et al. (2016) a model predictive control scheme is proposed which requires a nominal model of the system and it can handle tightly bounded disturbances. In Zribi et al. (2005), an adaptive controller is developed using a lineare approximation of the system and its performance is hence limited. A fuzzy controller is proposed in Chang and Fu (1997), however the convergece of the algorithm is slow and also stability is not guaranteed. In Dorfler and Grammatico (2017), semi-decentralized frequency synchronization schemes are presented without taking the transmission losses into account. To achieve frequency and voltage regulation, microgrid controllers are desiged in De Persis and Monshizadeh (2018) based on reduced-order models. However, such networks do not explicitly describe the loads. Hence, the controllers are not robust to load variations and model parametric uncertaities. In Trip et al. (2018), a sliding-mode controller is developed for the case of lossless microgrids and with the assumption of constant disturbances. Finally, in Weitenberg et al. (2018) an integral frequency control scheme, robust to disturbances, is proposed. However, similar to Trip et al. (2018), the power network is assumed to be purely inductive (lossless). However, this expectation is not generally met for the microgrids in the medium and low voltage levels.

In this paper, we propose a new control framework that handles lossy microgrids, has fast convergence and does not depend on a nominal model of the system. We present an actor-critic based reinforcement learning approach for frequency control of islanded microgrids with inverter-based DG units. The adaptive actor-critic control scheme

⋆ This work was supported by the NWO (Netherlands Organization for Scientific Research) program *Uncertainty Reduction in Smart Energy Systems* (URSES) under the support of the project ENBARK.

10.1016/j.ifacol.2019.08.164

presented here compensates for the uncertain dynamics of DG units and time-varying loads. Therefore, the necessity to know the nonlinear dynamics of the system is eliminated (as opposed to our previous work Adibi et al. (2017)). Hence, the controller can be integrated in a DG without the need to be initially tuned for the DG and furthermore, the closed-loop system's performance would not degrade by system parameters alteration due to e.g. aging, environmental effects and load variations. The proposed reinforcement learning approach appropriately reacts to changes in the nominal conditions of the system and rapidly tune the control parameters. For the frequency regulation problem, a long-term performance function is defined based on instantaneous rewards, but since the dynamics is unknown, we define a critic network to learn this performance function in real-time. On the other hand, an actor network aims at deriving an optimal control policy by approximating the unknown nonlinear dynamics and minimizing the learned performance function obtained from the critic network. Details of our proposed control design are presented in the following sections.

The remainder of the paper is arranged as follows. Section 2 describes the modeling for a lossy microgrid and formulates the frequency control problem, along with the closed loop stability of the error dynamics. Next, we present our proposed learning algorithm based on coupled critic and actor networks in Section 3. Simulation results are discussed in Section 4. Section 5 summarizes the paper.

2. PROBLEM STATEMENT AND THEORETICAL FOUNDATIONS

We assume a microgrid can be modeled as a graph $G = (N, E)$, with $N = \{1, 2, ..., n\}$ the nodes (buses that generate or consume power) and $E \subseteq N \times N$ the edges (network transmission lines) that connect the nodes. Each node $i \in N$ is a distributed generation source that has an inverter for interacting with the grid. Further, we consider a Kron-reduced lossy microgrid, in which the effect of the impedance loads is merged into the network impedances via the so-called Kron-reduction procedure (Kundur (1994), Dorfler and Bullo (2013)). Therefore, two nodes $\{i, j\} \in E$ are connected by a complex admittance $Y_{ij} = G_{ij} + iB_{ij} \in \mathbb{C}$ with conductance $G_{ij} \in \mathbb{R}$ and susceptance $B_{ij} \in \mathbb{R}$. Let $N_i = \{j \in N \mid j \neq i,\ \{i, j\} \in E\}$ denote the neighbors of node i. We assign a time-dependent voltage phase angle $\delta_i \in \mathbb{R}$ and a voltage amplitude $V_i \in \mathbb{R}_{\geq 0}$ to each node i in the grid. The relative voltage phase angles are denoted by $\delta_{ij} := \delta_i - \delta_j$, $\{i, j\} \in E$.

Based on the above notations, the active power flow coming to the grid at node $i \in N$ is formulated as (Kundur (1994))

$$P_i = G_{ii}V_i^2 - \sum_{j \in N_i} V_i V_j \Big(G_{ij} \cos(\delta_{ij}) + B_{ij} \sin(\delta_{ij}) \Big), \quad (1)$$

with $G_{ii} := \hat{G_{ii}} + \sum_{j \in N_i} G_{ij}$, where $\hat{G}_{ii} \in \mathbb{R}$ is the shunt conductance at the i^{th} node.

2.1 Microgrid Non-Linear Dynamical Model

We consider a microgrid model with discrete dynamics consisting of inverter-interfaced DG sources. The inverters have the conventional primary droop controllers that compromise between frequency and active power as in Schiffer et al. (2014)

$$\delta_i(k+1) = \delta_i(k) + \tau_1 \omega_i(k), \quad (2)$$

$$\omega_i(k+1) = \omega_i(k) - \frac{\tau_1}{\tau_2} \Big(\omega_i(k) + k_{P_i}(P_i(k) - P_i^\star) - u_i(k) \Big), \quad (3)$$

for $i \in N$. Here, $\omega_i \in \mathbb{R}$ is the inverter frequency and u_i is the secondary control input for which the design procedure will be presented in Section 2.2. The term P_i is the active power given by (1) and $P_i^\star$ represents the active power setpoint. The parameter $\tau_1 \in \mathbb{R}^+$ is the discretization step-size and $k_{P_i} \in \mathbb{R}^+$ is the frequency droop gain. We take into account that the power signals are measured with intermediate low-pass filters that have time constant $\tau_2 \in \mathbb{R}^+$. Moreover, we presume that the amplitude of voltage signals at each node are constant and consequently, the injected reactive for each node is zero.

To simplify notation we define

$$P^\star := \text{col}(P_i^\star) \in \mathbb{R}^n, \quad P := \text{col}(P_i) \in \mathbb{R}^n, \quad (4)$$

$$T_1 := \tau_1 \mathbb{I}_n \in \mathbb{R}^{n \times n}, \quad T_2 := \text{diag}(\frac{\tau_1}{\tau_2}) \in \mathbb{R}^{n \times n}, \quad (5)$$

$$K_P := \text{diag}(k_{P_i}) \in \mathbb{R}^{n \times n}, \quad (6)$$

$$x_1(k) := [\delta_1(k), \delta_2(k), \cdots, \delta_n(k)]^T \in \mathbb{R}^n, \quad (7)$$

$$x_2(k) := [\omega_1(k), \omega_2(k), \cdots, \omega_n(k)]^T \in \mathbb{R}^n, \quad (8)$$

$$u(k) := [u_1(k), u_2(k), \cdots, u_n(k)]^T \in \mathbb{R}^n, \quad (9)$$

$$x(k) := [x_1^T(k), x_2^T(k)]^T \in \mathbb{R}^{2n}, \quad (10)$$

and write the system (2)-(3) compactly as

$$x_1(k+1) = x_1(k) + T_1 x_2(k),$$

$$x_2(k+1) = x_2(k) - T_2 \Big(x_2(k) + K_P(P(k) - P^\star) - u(k) \Big). \quad (11)$$

We can write down the above system dynamics in the following form

$$x_1(k+1) = f_1(x(k)), \quad (12)$$

$$x_2(k+1) = f_2(x(k)) + g_2 u(k), \quad (13)$$

where $g_2 = T_2$ defined in (5) and

$$f_1(x(k)) := x_1(k) + T_1 x_2(k), \quad (14)$$

$$f_2(x(k)) := x_2(k) - T_2 \Big(x_2(k) + K_P(P(k) - P^\star) \Big). \quad (15)$$

We assume that the nonlinear dynamics of DGs, i.e. functions $f_1(x(k))$ and $f_2(x(k))$, are unknown. The aim is to develop a controller to compensate for frequency deviations, while being robust against parametric uncertainties resulted from the concealed dynamics and disturbances affecting the network. Therefore in Section 2.2, we will first design the control input in which the unknown dynamics are part of the overall input signal. In Section 3.2, we will then design actor-critic learning algorithms to estimate these unknown dynamics.

In the next section, the regulation error signal and the structure of the control input are defined which are the basis for our adaptive learning-based control design in Section 3.

2.2 Regulation Error Dynamic and Control Input Design

Consider system dynamics (12)-(13) with unknown nonlinear functions $f_1(x(k))$ and $f_2(x(k))$, and the control input $u(k)$ to be designed. For simplicity, we assume that τ_1 and τ_2 are known, hence, $g_2 = T_2$ is a known constant matrix. Let us define the nominal frequency of the system as $\omega^\star \in \mathbb{R}^+$ and the vector of the desired frequency signals as $x_2^\star := \omega^\star \mathbb{1}_n \in \mathbb{R}^n$. The control objective is to compensate the deviation of frequency signals (8) from their nominal value $\omega^\star$ and make frequencies converge to the desired signal $x_2^\star$. To accomplish this, we define the regulation error signal $e(k) \in \mathbb{R}^n$ as

$$e(k) = x_2^\star - x_2(k), \tag{16}$$

which results in the error dynamics

$$\begin{aligned} e(k+1) &= x_2^\star - x_2(k+1) \\ &= x_2^\star - f_2(x(k)) - g_2 u(k). \end{aligned} \tag{17}$$

To design $u(k)$ such that (17) is stabilized, we define the candidate Lyapunov function

$$L(k) = e^T(k)e(k). \tag{18}$$

Differentiating of $L(k)$ in discrete time results in

$$\Delta L(k) = e^T(k+1)e(k+1) - e^T(k)e(k). \tag{19}$$

Using the error dynamics (17) and substituting it in (19), we obtain

$$\begin{aligned} \Delta L(k) =& \Big(x_2^\star - f_2(x(k)) - g_2 u(k)\Big)^T \\ &\times \Big(x_2^\star - f_2(x(k)) - g_2 u(k)\Big) - e^T(k)e(k). \end{aligned} \tag{20}$$

In order to have $\Delta L(k) < 0$, we select the control input as

$$u(k) = g_2^{-1}\Big(x_2^\star - f_2(x(k)) + Ke(k)\Big), \tag{21}$$

where $K \in \mathbb{R}^{n\times n}$ is a constant diagonal positive definite gain matrix. If we assume $f_2(x(k))$ is known, substituting (21) in (20) yields

$$\Delta L(k) = \sum_{i=1}^{n} (K_i^2 - 1)e_i^2, \tag{22}$$

where e_i is the i^{th} element of $e(k)$ and K_i is the i^{th} eigenvalue of the diagonal matrix K for $i \in N$. Hence, $\Delta L(k) < 0$ and the error system (17) is asymptotically stable if

$$0 < K^{\max} < 1, \tag{23}$$

where $K^{\max} \in \mathbb{R}$ is the maximum eigenvalue of K.

However, the dynamics $f_2(x(k))$ is not known. Instead, we use the estimation of the function $f_2(x(k))$, i.e. $\hat{f}_2(x(k))$ ($\hat{f}_2(x(k))$ is approximated using the actor network and will be discussed in Section 3.2). We design the control input (21) as follows

$$u(k) = g_2^{-1}\Big(x_2^\star - \hat{f}_2(x(k)) + Ke(k)\Big), \tag{24}$$

which results in

$$\begin{aligned} \Delta L(k) =& \Big(\tilde{f}_2(x(k)) - Ke(k)\Big)^T \Big(\tilde{f}_2(x(k)) - Ke(k)\Big) \\ &- e^T(k)e(k), \end{aligned} \tag{25}$$

where $\tilde{f}_2(x(k)) = \hat{f}_2(x(k)) - f_2(x(k))$ is the error of function estimation. Therefore, $\Delta L(k) < 0$ if

$$\left\|\tilde{f}_2(x(k)) - Ke(k)\right\| < \|e(k)\|. \tag{26}$$

Let the known value $f_2^{\max} \in \mathbb{R}^+$ be the upper bound of the function estimation error $\tilde{f}_2(x(k))$, such that $\left\|\tilde{f}_2(x(k))\right\| \le f_2^{\max}$. Hence, $\Delta L(k) < 0$ provided that

$$\begin{aligned} \left\|\tilde{f}_2(x(k)) - Ke(k)\right\| &\le \left\|\tilde{f}_2(x(k))\right\| + \|Ke(k)\| \\ &\le f_2^{\max} + K^{\max}\|e(k)\|. \end{aligned} \tag{27}$$

Considering (26), the system of error dynamics is stable if

$$f_2^{\max} + K^{\max}\|e(k)\| < \|e(k)\|. \tag{28}$$

Defining $e^{\max} := \frac{f_2^{\max}}{1-K^{\max}}$, it follows that

$$\Delta L(k) < 0, \quad \forall \|e(k)\| > e^{\max}. \tag{29}$$

In other words, $\Delta L(k)$ is negative outside of the compact set $S_e := \{\|e(k)\| \le e^{\max}\}$, or equivalently, all the solutions that start outside of S_e will enter this set within a finite time, and will remain inside the set forever. This means that

$$\|e(k)\| < \frac{f_2^{\max}}{1 - K^{\max}}, \tag{30}$$

and therefore the estimation errors and the closed-loop system is bounded above with the ultimate bound $e^{\max}$.

3. ACTOR-CRITIC LEARNING ALGORITHM

We consider a neural network that has one hidden layer for both actor and critic networks. In order to measure the long-term performance of the system, the cost function $J(k) \in \mathbb{R}^n$ is defined using the instantaneous reward as (Lewis et al. (1998), Sokolov et al. (2015))

$$\begin{aligned} J(k) &= \sum_{m=k}^{\infty} \gamma^{m-k} r(m+1) \\ &= r(k+1) + \gamma r(k+2) + \gamma^2 r(k+3) + \cdots, \end{aligned} \tag{31}$$

where $0 < \gamma < 1$ is the discount factor and $r(k) = [r_1(k)r_2(k)\cdots r_n(k)]^T \in \mathbb{R}^n$ is the vector of instantaneous rewards (reinforcement learning signal) as follows (He and Jagannathan (2005))

$$r_i(k) = \begin{cases} 0 & \text{if} \quad |e_i(k)| \le c \\ 1 & \text{if} \quad |e_i(k)| > c \end{cases} \tag{32}$$

for $i \in N$ and $c \in \mathbb{R}^+$ is a fixed threshold. The instantaneous reward $r_i(k)$ is a measure of the current performance of the i^{th} DG. To be more precise, it quantifies how the control input has performed; $r_i(k) = 0$ indicates a success in the frequency regulation and $r_i(k) = 1$ shows a performance degradation.

Since the dynamics is unknown, we define a critic network to learn the cost function $J(k)$ in real-time in Section (3.1).

3.1 Adaptation of Critic Network

The critic neural network, with output $\hat{J}(k) \in \mathbb{R}^n$, learns to approximate the cost function $J(k) \in \mathbb{R}^n$. The output of the critic neural network can be described in the form

$$\hat{J}(k) = \hat{\psi_c}^T(k)\phi_c\Big(v_1^T(k)x(k)\Big) = \hat{\psi_c}^T(k)\phi_c(k), \tag{33}$$

such that $\hat{\psi_c}^T(k) \in \mathbb{R}^{n\times n_1}$ represents the matrix of weights between the hidden and output layer and $v_1^T \in \mathbb{R}^{n_1\times 2n}$ represents the matrix of weights between the input and hidden layer. We assume that the matrix of the weights,

v_1, is fixed and only the weights $\hat{\psi}_c$ between the hidden and output layer are being adapted. We fix the weights of the hidden layer in order to reduce the training time and to have faster learning. Moreover, $\phi_c(k) \in \mathbb{R}^{n_1}$ is the vector of basis functions and n_1 denotes the total number of nodes for the hidden layer.

Let $e_c(k) \in \mathbb{R}^n$ be the prediction error (Temporal-Difference error; see Sutton and Barto (1998)) of the critic network as

$$\begin{aligned} e_c(k) &= r(k) + \gamma \hat{J}(k) - \hat{J}(k-1), \\ &= r(k) + \gamma \hat{\psi}_c^{\ T}(k)\phi_c(k) - \hat{\psi}_c^{\ T}(k-1)\phi_c(k-1), \end{aligned} \tag{34}$$

and the cost function that is going to be minimized as

$$J_c(k) = \frac{1}{2} e_c^T(k) e_c(k). \tag{35}$$

Applying gradient descent algorithm for minimizing $J_c(k)$, and hence $e_c(k)$, results in

$$\begin{aligned} \hat{\psi}_c(k+1) &= \hat{\psi}_c(k) - \alpha_c \frac{\partial J_c(k)}{\partial e_c(k)} \frac{\partial e_c(k)}{\partial \hat{J}(k)} \frac{\partial \hat{J}(k)}{\partial \hat{\psi}_c(k)} \\ &= \hat{\psi}_c(k) - \alpha_c \gamma \phi_c(k) e_c(k), \end{aligned} \tag{36}$$

which leads to the following update rule for weights of the critic network

$$\begin{aligned} \hat{\psi}_c(k+1) = \hat{\psi}_c(k) - \alpha_c \phi_c(k) \times \\ \Big(r(k) + \gamma \hat{\psi}_c^{\ T}(k)\phi_c(k) - \hat{\psi}_c^{\ T}(k-1)\phi_c(k-1) \Big), \end{aligned} \tag{37}$$

where $\alpha_c \in \mathbb{R}^+$ is the critic learning rate.

In Section (3.2), the actor network is constructed to minimize both the function estimation error $\tilde{f}_2(x(k))$ and the cost function $\hat{J}(k)$.

3.2 Adaptation of Actor Network

The main purpose of the actor network is to generate the approximation of the unknown nonlinear function $f_2(x(k))$ and then plug the estimated $\hat{f}_2(k)$ into the control policy (24). The estimated function is parameterized as

$$\hat{f}_2(k) = \hat{\psi}_a^{\ T}(k)\phi_a\Big(v_2^T(k)x(k)\Big) = \hat{\psi}_a^{\ T}(k)\phi_a(k), \tag{38}$$

where $\hat{\psi}_a^{\ T}(k) \in \mathbb{R}^{n \times n_2}$ represents the matrix of weights between the hidden layer and the output layer and $v_2^T \in \mathbb{R}^{n_2 \times 2n}$ represents the matrix of weights between the input layer and the hidden layer. We assume that the matrix of the weight v_2 is fixed and only the weights $\hat{\psi}_a$ between the hidden layer and the output layer are being adapted. Moreover, $\phi_a(k) \in \mathbb{R}^{n_2}$ is the vector of basis function of the hidden layer and n_2 denotes the total units of the hidden layer.

We define the function estimation error $\tilde{f}_2(k) \in \mathbb{R}^n$ as

$$\tilde{f}_2(k) = \hat{f}_2(k) - f_2(k), \tag{39}$$

and the error between the desired cost function $J^\star(k) \in \mathbb{R}^n$ and the critic network output $\hat{J}(k)$ as

$$\tilde{J}(k) = \hat{J}(k) - J^\star(k). \tag{40}$$

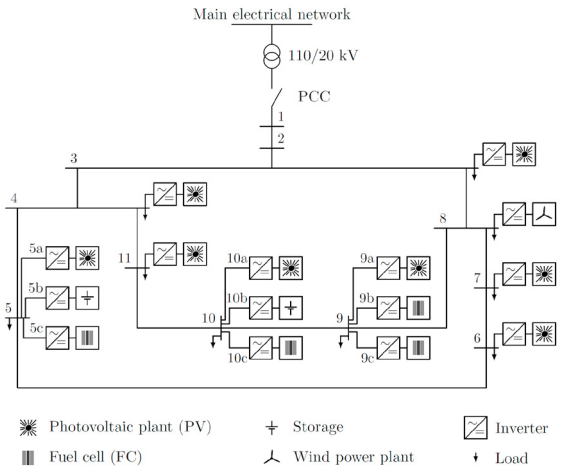

Fig. 1. The grid model taken from Schiffer (2015) has eleven buses and multiple storage and generation units.

The training of the actor network is done using $\tilde{f}_2(k)$ and $\tilde{J}(k)$ and defining the prediction error $e_a(k) \in \mathbb{R}^n$ as

$$e_a(k) = \tilde{f}_2(k) + \tilde{J}(k). \tag{41}$$

According to (31) and (32), the desired value for $J^\star(k)$ is 0. Thus, (41) becomes

$$e_a(k) = \tilde{f}_2(k) + \hat{J}(k). \tag{42}$$

We consider the cost function that is going to be minimized by the actor network in the form

$$J_a(k) = \frac{1}{2} e_a^T(k) e_a(k). \tag{43}$$

Using the gradient descent algorithm for minimizing $J_a(k)$ and subsequently for $e_a(k)$, we obtain

$$\begin{aligned} \hat{\psi}_a(k+1) &= \hat{\psi}_a(k) - \alpha_a \frac{\partial J_a(k)}{\partial e_a(k)} \frac{\partial e_a(k)}{\partial \tilde{f}_2(k)} \frac{\partial \tilde{f}_2(k)}{\partial \hat{\psi}_a(k)} \\ &= \hat{\psi}_a(k) - \alpha_a \phi_a(k) e_a(k), \end{aligned} \tag{44}$$

which results in

$$\hat{\psi}_a(k+1) = \hat{\psi}_a(k) - \alpha_a \phi_a(k)(\tilde{f}_2(k) + \hat{J}(k))^T, \tag{45}$$

where $\alpha_a \in \mathbb{R}^+$ is the actor learning rate. However, we can not use the weight update rule (45) in practice. This is due to the fact that the error function $\tilde{f}_2(k)$ defined in (39) consists of the *unknown* nonlinear function $f_2(k)$. This problem can be addressed by substituting (24) in (17), which yields

$$\begin{aligned} e(k+1) &= -f_2(x(k)) + \hat{f}_2(x(k)) - Ke(k) \\ &= \tilde{f}_2(x(k)) - Ke(k). \end{aligned} \tag{46}$$

Hence, the function estimation error becomes

$$\tilde{f}_2(k) = e(k+1) + Ke(k). \tag{47}$$

Substituting (47) in (45), results in the actor network weight update rule

$$\hat{\psi}_a(k+1) = \hat{\psi}_a(k) - \alpha_a \phi_a(k) \Big(e(k+1) + Ke(k) + \hat{J}(k) \Big)^T. \tag{48}$$

In the following section, we validate the performance of the proposed control scheme via simulation on a benchmark microgrid in the presence of disturbances.

4. CASE STUDY

The effectiveness of our reinforcement learning-based control scheme is evaluated on the isolated three-phase subnetwork of the CIGRE medium voltage benchmark network as described in Rudion et al. (2006) and Schiffer et al. (2014). The benchmark microgrid is illustrated in Fig. (1). The simulation is carried out by considering $n = 6$ controllable generation sources at buses 5b, 5c, 9b, 9c, 10b and 10c named by DG1 to DG6 from now on, respectively. All photovoltaic (PV) sources together with the wind turbine at bus 8 are considered as non-controllable units and are neglected. All of the generation units have integrated droop controllers. For each inverter $i \in N$, the active power rating $P_i^N \in \mathbb{R}^+$ is assigned. The associated active power rating P_i^N, active power setpoints $P_i^\star$ and droop controller gains k_{P_i} to each inverter $i \in N$ are given in Table 1. The industrial and household loads at nodes 3-11 are described in Rudion et al. (2006), see Table 1 in there. The other parameters such as the transmission line lengths and etc are summarized Rudion et al. (2006), see Table 3 in there. Furthermore, we discard the load at node 1.

The voltage amplitudes are set to $V_i = 1$ per unit for all $i \in N$. The nominal frequency, the time constant and the sampling time are taken as $f^\star = 50$ Hz, $\tau_2 = 0.5$ s and $\tau_1 = 50$ ms, respectively. The elements of the diagonal gain matrix K is selected as $K_i = 0.1$ for $i \in N$. The threshold value c is set to $c = 0.02$. We consider one hidden layer for both critic and actor neural networks, and we assume that each hidden layer contains 10 nodes, i.e. $n_1 = n_2 = 10$. For weight updating rules, the learning rates are selected as $\alpha_c = 0.1$, $\alpha_a = 0.1$ and the discount factor is set as $\gamma = 0.5$. All the weight parameters of the matrices v_1 and v_2, between the input layer and the hidden layer, are fixed as 1. The initial values for the adapting weights $\hat{\psi}_c$ and $\hat{\psi}_a$ are selected randomly between 0 and 1. Furthermore, we choose hyperbolic tangent functions as activation functions.

Table 1. Network parameters

Base values	$P_{\text{base}} = 4.75$ MVA, $V_{\text{base}} = 20$ kV
$P_i^N, i = 1, \cdots, 6$	[0.505, 0.028, 0.261, 0.179, 0.168, 0.012] p.u.
$P_i^\star, i = 1, \cdots, 6$	[0.202, 0.008, 0.078, 0.054, 0.067, 0.004] p.u.
$k_{P_i}, i = 1, \cdots, 6$	[0.396, 7.143, 0.766, 1.117, 1.191, 16.667] $\frac{\text{Hz}}{\text{p.u.}}$

In this case study, we demonstrate the effectiveness of the adaptive control scheme under load variations. The trajectories of the frequencies $f_i = \frac{\omega_i}{2\pi}$ in Hz for $i = 1, ..., 6$ for the controllable sources are presented in Fig. (2) and Fig. (3), with closer view. We choose the initial states arbitrarily. Further, we consider the microgrid to be in the islanded mode. Since we have developed an online learning algorithm, we do not have the entire training data set available at once as in batch neural network training approaches. Instead, the learning data becomes available in a sequential order and the new observed data at each time step is used to continuously train and update our control law. As seen in Fig. (2), during the initial phase

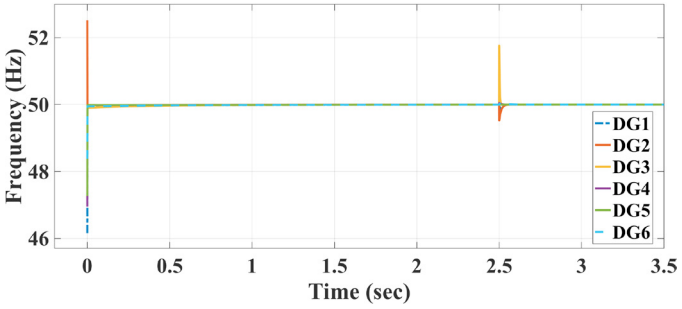

Fig. 2. Time trajectories of the frequency signals, considering a change in system parameters at $t = 2.5$ s.

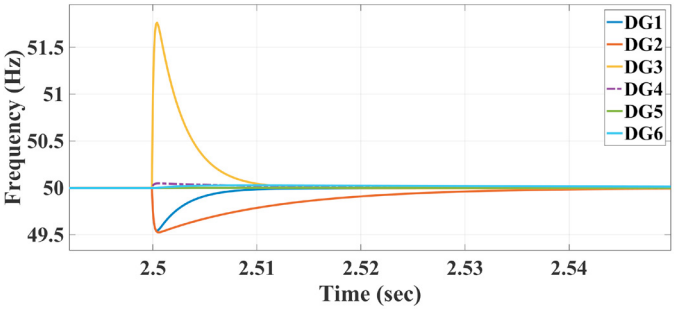

Fig. 3. Time trajectories of the frequency signals from a closer point of view at $t = 2.5$ s.

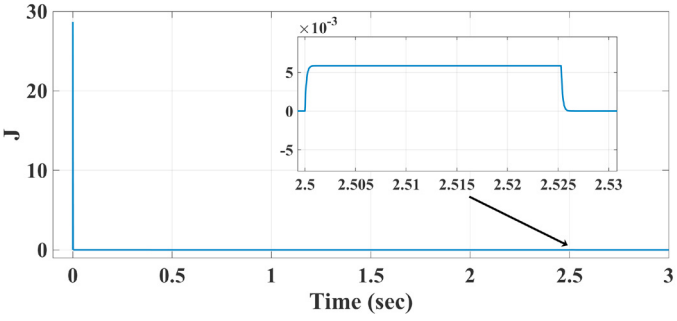

Fig. 4. Sum of cost functions among DGs, considering a change in system parameters at $t = 2.5$ s.

of the simulation, the critic and actor networks quickly learn the undisclosed dynamics, within a short transient. At time $t = 2.5$ s, the conductance and inductance in the system are changed. As one can observe in Fig. (2) and from a closer view in Fig. (3), after applying the changes, the frequency signals vary from 50 Hz due to sudden impedance changes. However, after some oscillations for a short period of time, the frequencies converge to the nominal frequency $f^\star = 50$ Hz. Hence, the reinforcement learning actor-critic based control scheme compensates for the deviation of frequencies and the frequency regulation errors quickly converge to zero. Note that without the proposed control strategy and by using only the primary droop control the lossy system has deterioration from the nominal frequency and in the presence of the mentioned disturbances the system becomes unstable.

Fig. (4) illustrates the sum of cost functions among DG sources. At time instant $t = 2.5$ s, when the load changes are applied, a rise in the total cost function is observed.

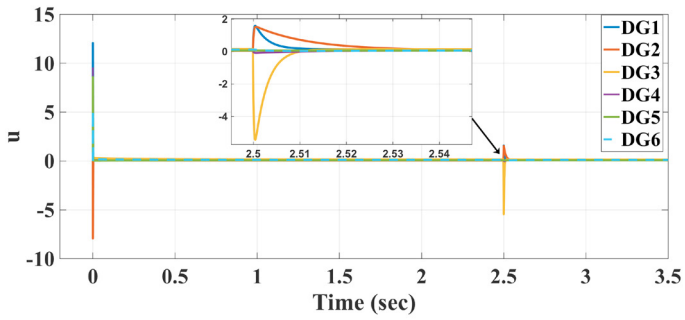

Fig. 5. The control input trajectories, considering a change in system parameters at $t = 2.5$ s.

However, the control input restores it back to the optimal value $J^\star = 0$ in less than a second. The time trajectory of the learned control law (24) that regulates the frequencies to the nominal frequency is illustrated in Fig. (5).

5. CONCLUSIONS AND FUTURE RESEARCH

A fast reinforcement learning control scheme has been proposed for optimal frequency synchronization of lossy microgrids. Our method is able to efficiently handle general cases of resistive and inductive line and load impedances, parameter uncertainties, time varying loads and disturbances. Using this adaptive control approach, no priori knowledge about the system dynamics is required. Adaptive critic and actor neural networks are exploited to approximate the nonlinear system dynamics, and approximate and minimize the cost function corresponding to the frequency errors. The simulation results have shown that the proposed control scheme provides fast convergence of frequency signals of DG sources to the nominal frequency in the presence of disturbances affecting the system. In case we discard the contribution from the neural network in our control scheme, the overall performance will be deteriorated. However, the regulation error will remain bounded based on (30). Further discussions on this situation will be presented in the longer version of this paper.

As next steps, we will extend our approach to deal with the voltage control and active/reactive power sharing problems. The convergence proof of the learning algorithms is currently under developments by the authors. Experimental validations of our proposed methods will be carried out as well.

REFERENCES

M. Adibi, J. van der Woude, and D. Jeltsema. A port-Hamiltonian approach to secondary voltage control of microgrids. In *IEEE PES Innovative Smart Grid Technologies Conference Europe*, 2017.

C. Chang and W. Fu. Area load frequency control using fuzzy gain scheduling of PI controllers. *electric power systems research*, 42(2):145–152, 1997.

C. De Persis and N. Monshizadeh. Bregman storage functions for microgrid control. *IEEE Transactions on Automatic Control*, 63(1):53–68, 2018.

C. De Persis, N. Monshizadeh, J. Schiffer, and F. Dorfler. A Lyapunov approach to control of microgrids with a network-preserved differential-algebraic model. In *55th IEEE Conference on Decision and Control*, 2016.

F. Dorfler and F. Bullo. Kron reduction of graphs with applications to electrical networks. *IEEE Transactions on Circuits and Systems*, 60(1):150–163, 2013.

F. Dorfler and S. Grammatico. Gather-and-broadcast frequency control in power systems. *Automatica*, 79: 296–305, 2017.

A. M. Ersdal, L. Imsland, and K. Uhlen. Model predictive load frequency control. *IEEE Transactions on power systems*, 31(1):777–785, 2016.

J. M. Guerrero, J. C. Vsquez, J. Matas, M. Castilla, L. G. d. Vicua, and M. Castilla. Hierarchical control of droop-controlled AC and DC microgrids: A general approach toward standardization. *IEEE Transactions on Industrial Electronics*, 58(1):158–172, 2011.

J. M. Guerrero, M. Chandorkar, T. L. Lee, and P. Chiang Loh. Advanced control architectures for intelligent microgrids. *Automatica*, 60(4):1254–1270, 2013.

P. He and S. Jagannathan. Reinforcement learning-based output feedback control of nonlinear systems with input constraints. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(1):150–154, 2005.

P. Kundur. *Power System Stability and Control*. McGraw-Hill, New York, United States, 1994.

R. Lasseter. Conditions for stability of droop-controlled inverter-based microgrids. *Automatica*, 1:305–308, 2002.

F. L. Lewis, A. Yesildirak, and S. Jagannathan. *Neural Network Control of Robot Manipulators and Nonlinear Systems*. Taylor and Francis, Inc., Bristol, PA, USA, 1998.

K. Rudion, A. Orths, Z. Styczynski, and K. Strunz. Design of benchmark of medium voltage distribution network for investigation of DG integration. In *IEEE PESGM*, 2006.

J. Schiffer. *Stability and power sharing in microgrids*. Ph.D. thesis, TU Berlin, Berlin, Germany, 2015.

J. Schiffer, R. Ortega, A. Astolfi, J. Raisch, and T. Sezi. Conditions for stability of droop-controlled inverter-based microgrids. *Automatica*, 50(10):2457–2469, 2014.

J. W. Simpson-Porco, Q. Shafiee, F. Dorfler, J. C. Vasquez, J. M. Guerrero, and F. Bullo. Secondary frequency and voltage control in islanded microgrids via distributed averaging. *IEEE Transactions on Industrial Electronics*, 62(11):7025–7038, 2015.

Y. Sokolov, R. Kozma, L. D. Werbos, and P. J. Werbos. Complete stability analysis of a heuristic approximate dynamic programming control design. *Automatica*, 59: 9–18, 2015.

R. Sutton and A. Barto. *Reinforcement learning: An introduction*. MIT Press, Cambridge, MA, USA, 1998.

S. Trip, M. Cucuzzella, C. De Persis, A. van der Schaft, and A. Ferrara. Passivity-based design of sliding modes for optimal load frequency control. *IEEE Transactions on control systems technology*, pages 2457–2469, 2018.

E. Weitenberg, Y. Jiang, C. Zhao, E. Mallada, C. De Persis, and F. Dorfler. Robust decentralized secondary frequency control in power systems: Merits and tradeoffs. In *European Control Conference*, 2018.

M. Zribi, M. Al-Rashed, and M. Alrifai. Adaptive decentralized load frequency control of multi-area power systems. *International journal of electrical power and energy systems*, 27(8):575–583, 2005.