2019

Cooperative secondary voltage control of static converters in a microgrid using model-free reinforcement learning

Edward Smith
University of Wollongong, ejs760@uowmail.edu.au

Duane A. Robinson
University of Wollongong, duane@uow.edu.au

Ashish P. Agalgaonkar
University of Wollongong, ashish@uow.edu.au

Cooperative secondary voltage control of static converters in a microgrid using model-free reinforcement learning

Abstract

Agent-based secondary voltage regulation in an islanded MicroGrid is complicated by non-linear system dynamics, state couplings and uncertain communication network topology information. This paper proposes an off-policy learning algorithm for cooperative secondary voltage control which can synthesize an optimal feedback controller in real-time without knowledge of the system model. A simulation model has been developed using MATLAB/Simulink, which demonstrates a working controller. Results from the simulations are included, and practical considerations regarding implementation on a real system discussed.

Keywords

model-free, microgrid, reinforcement, converters, learning, static, control, voltage, cooperative, secondary

Disciplines

Engineering | Science and Technology Studies

Publication Details

E. Smith, D. A. Robinson & A. Agalgaonkar, "Cooperative secondary voltage control of static converters in a microgrid using model-free reinforcement learning," in EPE'19 ECCE Europe, 2019, pp. 1-10.

Cooperative secondary voltage control of static converters in a microgrid using model-free reinforcement learning

Edward Smith, Duane A. Robinson, Ashish Agalgaonkar
Australian Power Quality & Reliability Centre
School of Electrical, Computer & Telecommunications Engineering
University of Wollongong, NSW 2522
Australia
E: ejs760@uowmail.edu.au

Keywords

«Adaptive control»,«Communication for Power Electronics»,«Converter control»,«Distributed power»,«micro-grids»,«power quality»,«Non-linear control»,«Neural network»,«Smart grids»,«Voltage Source Converter»,«Wireless Control»

Abstract

Agent-based secondary voltage regulation in an islanded MicroGrid is complicated by non-linear system dynamics, state couplings and uncertain communication network topology information. This paper proposes an off-policy learning algorithm for cooperative secondary voltage control which can synthesize an optimal feedback controller in real-time without knowledge of the system model. A simulation model has been developed using MATLAB/Simulink, which demonstrates a working controller. Results from the simulations are included, and practical considerations regarding implementation on a real system discussed.

1. Introduction

During autonomous operation of a MicroGrid, otherwise known as islanded mode, it is necessary that subsequent to a load step or network switching event, some form of secondary control action is required to restore voltage and frequency to utility prescribed steady-state norms. Methods for both primary and secondary local control in power converters have been extensively researched, and the reader is referred to [1]. Agent-based control offers several advantages due to its intrinsic resilience and well understood dynamic behavior. For cooperative control of agent-based systems, the problem can be described as tracking synchronization, wherein all agents synchronize output voltages to an exogenous reference value [2]. Agent communication using, for example, a wireless ad-hoc network supports such cooperative control.

To ensure convergence and stability of the output voltages, restrictive conditions are placed on the agent dynamics, coupling gains and network topology. In this regard, agent based synchronization strategies have been proposed in the control literature which guarantee convergence given practical implementation constraints [3]–[5]. Secondary voltage control using an agent-based coordination strategy is modelled in [6] by using input-output feedback linearization to guarantee stability and convergence. Similarly in [7] the large signal non-linearities are compensated for using a radial basis function network approximator. The problem of unknown global communication graph topology information, which is required to establish the local coupling gains, is addressed in [8] using an adaptive control law that adjusts the coupling gains to assure convergence. These control schemes require complete or at least partial knowledge of the system model dynamics.

Power electronic converters are now pervasive on electricity networks for grid-interfacing renewable energy generators and storage, embedded with intelligent electronic devices (IEDs) having fast information processing capability, consistent with Smart Grid functionality. This enables recent

advances in model-free adaptive control, in particular reinforcement learning, that are data-driven to be realized in a practical sense [9]. For example, see [10] for application to a synchronous generator. This allows for the benefits of a fully distributed agent based MicroGrid control system to be realized. In this regard, the paper proposes an alternative approach to cooperative output voltage synchronization in a MicroGrid which enhances existing schemes in several important aspects, specifically i) no dynamic model is required for the converter system with controller synthesis utilizing real-time data acquisition and reinforcement learning, ii) the dynamic response is optimal for an explicit cost function including transient response and reactive power usage, and iii) the adaptive scheme is off-policy which offers several practical advantages in this respect.

2. Preliminaries: Graph Theory and Multi-Agent Systems

The following notation is observed, the state vector for each Agent i is bold lowercase $\boldsymbol{x}_i \in \mathbb{R}^n$, and $f_i(\boldsymbol{x}_i)$ is a $\mathbb{R}^n$ column vector of functions of states. Matrices are given in uppercase as $A_i \in \mathbb{R}^{n\times m}$, $|.|$ deonotes the Euclidean norm and $\otimes$ the Kronecker product. The identity matrix of appropriate dimension is I, and for matrix inequalities the operator $>$ ($\geq$) means positive definite (positive semidefinite). The terms control law and policy are used interchangeably, and a control policy is said to be admissible if it is stabilizing in the sense of Lyapunov and has finite performance cost as $t \to \infty$. A directed graph, or digraph, $G = (V, E, A)$ is shown below, where the set of nodes are $V = \{v_1, v_2, \cdots, v_n\}$, the set of directed arcs or edges E from $v_i$to v_j given by (v_i, v_j), and the weights associated with each edge $0 \leq a_{ij} \leq 1$ described by an adjacency matrix $A = [a_{ij}]$ for G. Each node has an associated set of neighbours $N_j = (v_j : (v_j, v_i) \in E)$. The diagonal in-degree matrix D is defined by $D = \text{diag}\{d_i\}$ with $d_i = \sum_{j=1}^{N} a_{ij}$, and the graph Laplacian matrix L as $L = D - A$. The eigenvalues of L, denoted λ_n, are particularly significant since they determine the global dynamics of the multi-gent system on graph G [2]. As for in-degree, the out-degree matrix is defined as $D^o = \text{diag}\{d_o\}$ where $d_o = \sum_{j=1}^{N} a_{ji}$. A balanced node has in-degree equal to its out-degree, and all undirected graphs are balanced. A digraph is strongly connected if there is a continuous path a node to any other node. A graph is said to have a spanning tree if it is strongly connected.

In agent-based control each node is considered a dynamic sub-system, where a common control objective is to achieve consensus between node states or outputs. Where a single node provides the reference value for the group, it is known as the pinning node. The connection graph is abstracted, but may be for example an ad-hoc wireless network [11]. The pinning node is introduced as v_o to the bi-graph below with pinning gain g_1. Associated with each node is a dynamic system $\dot{\boldsymbol{x}}_i = f(\boldsymbol{x}_i, u_i)$, with state vector $\boldsymbol{x}_i$, and the leader subsystem at node v_0 has state vector $\boldsymbol{x}_o$.

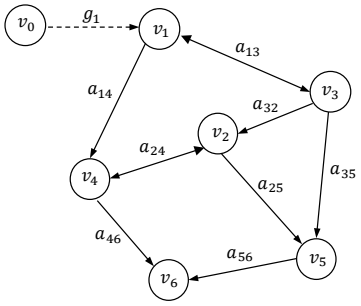


Figure 1. Communication graph

The leader dynamics describe a trajectory to which all nodes should synchronize. The cooperative tracking problem is to design a control input $u_i = F_i(\boldsymbol{x}_i, u_i)$ for each agent such that all nodes in the system synchronise to the leader trajectory. If the global synchronisation error, or disagreement vector, is defined as $\boldsymbol{\delta} = \boldsymbol{x} - \underline{\boldsymbol{x}}_0$, then the tracking synchronisation problem is solved if $\lim_{t\to\infty} \boldsymbol{\delta}(t) = 0$.

Define the local neighborhood synchronisation error as

$$\boldsymbol{e}_i = \sum_{j\epsilon N} a_{ij}(\boldsymbol{x}_i - \boldsymbol{x}_j) + g_i(\boldsymbol{x}_i - \boldsymbol{x}_o) \tag{1}$$

The objective of tracking synchronisation is to select a distributed control $u_i = -cK\boldsymbol{e}_i$, where $c > 0$ is the scalar coupling gain and $K \in \mathbb{R}^{m\times n}$ is the feedback control gain matrix, g_i is the pinning gain. It should be noted that the only information available to each node is the local neighborhood error.

The sufficient conditions for convergence of all node states, or subset of states, to the leader trajectory are well established [2]. For first order linear time invariant dynamics, the required coupling gain is given by

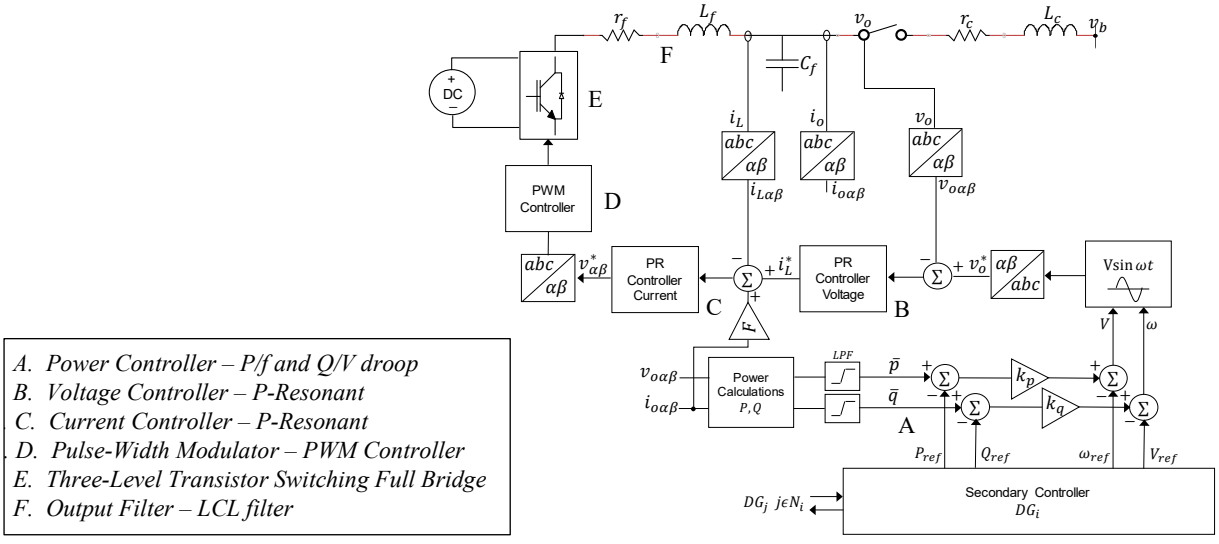
$$c \geq \frac{1}{2\lambda_{min}} \tag{2}$$

where $\lambda_{min} = \min_{i \epsilon N} Re(\lambda_i)$ and $Re(\lambda_i)$ is the real part of the eigenvalue for the Fig. 1 graph Laplacian matrix. The salient point here being that local calculation of λ_{min} for an Agent requies global knowledge of the communication Graph structure, information which may not be available for a practical implementation. A fully distributed consensus protocol which adaptively determines the coupling gain based only on neighbour information is proposed in [3], and applied to secondary voltage and frequency control in a MicroGrid [8].

3. The Output Voltage Tracking Problem in MicroGrids

The three-phase voltage-sourced converter model adopted for this paper is based on the large-signal stationary-frame state space model given in [12], [13], and the complete MicroGrid system model from [14]. The model constitutes the power conversion system for each distributed generator, or DG. The converter architecture is based on a current-controlled voltage-source inverter, using a fast-inner current control loop, and an outer voltage and frequency control loop for autonomous operation. Primary voltage control is normally fast acting and reaches steady state within a few cycles. Load sharing of power is based on P/f and Q/V droops, with an external voltage and frequency reference input to eliminate steady-state errors. Primary controllers in this implementation are of proportional-resonant (P-RES) design, which have particular advantages as described in [12], [13]. Several performance improvements for secondary controller design have been proposed, in particular with regards to droop response [15].

Previous controller enhancements do not address the output voltage tracking problem as it is framed below. As such, the converter model introduced here is intended to be simple enough to illustrate the nature of the control problem. The details of the converter model are described below. Referring to each converter as DG_i, the control block diagram is shown in Fig. 2 and a partial set of state equations for the power controller are given in (3)-(7). The internal state $\boldsymbol{x_i}$ of each DG_i is a vector of scalar values as shown in the figure below. As per [14], the equations are grouped according to the converter control hierarchy.



A. Power Controller – P/f and Q/V droop
B. Voltage Controller – P-Resonant
C. Current Controller – P-Resonant
D. Pulse-Width Modulator – PWM Controller
E. Three-Level Transistor Switching Full Bridge
F. Output Filter – LCL filter

Figure 2. DG_i Converter Reference Model

The converter model consists of functional blocks *A* to *F*, as listed in Fig. 2. Three-phase variables undergo a Clarke transformation shown as $abc \rightarrow \alpha\beta 0$ to a stationary reference frame, however for simplicity only three-wire power systems are considered here and subsequently there is no zero-sequence component.

The power controller provides a sinusoidal output voltage reference signal for the voltage controller, which provides the inductor current reference for the current controller. The inner primary voltage and current controllers (*B, C*) are based on proportional-resonant compensators, while each DG_i utilizes voltage and frequency droop controllers for the outer-loop (*A*) to self-regulate load sharing. The complete state space model of the converter is provided in [14]. The power controller equations are reproduced below, where V_{ni} and ω_{ni} are the external voltage and frequency reference inputs, P_i and Q_i the calculated active and reactive power based on measured outputs v_{oi} and i_{oi}, and m_{Pi} and n_{Qi} the droop gain coefficients. The voltage reference signal output is $v^*_{o\alpha\beta i}$.

$$\dot{P}_i = -\omega_c P_i + \omega_c v_{o\alpha i} i_{o\alpha i} + \omega_c v_{o\beta i} i_{o\beta i} \quad (3)$$

$$\dot{Q}_i = -\omega_c Q_i + \omega_c v_{o\beta i} i_{o\alpha i} - \omega_c v_{o\alpha i} i_{o\beta i} \quad (4)$$

$$\dot{\delta}_i = \omega_{ni} - m_{pi} P_i \quad (5)$$

$$v^*_{o\alpha i} = \left(V_{ni} - n_{qi} Q_i\right) \cos \delta_i \quad (6)$$

$$v^*_{o\beta i} = \left(V_{ni} - n_{qi} Q_i\right) \sin \delta_i \quad (7)$$

The full converter model consists of a non-linear differential equation describing the DG_i dynamics. The complete model can be summarized in (8) below, where $\boldsymbol{x}_i$ is the state vector and u_i the control input. The state vector $\boldsymbol{x}_i$ as per (9) contains auxiliary state variables $\varphi_{\alpha\beta i}$ and $\gamma_{\alpha\beta i}$ which capture the primary control loop internal dynamics. The measured bus voltage $v_{b\alpha\beta}$ is considered as an independent state variable or known disturbance for the converter. For cooperative control in a multi-agent system, the non-linear dynamics in particular present some difficulty which will be addressed in the next Section. The secondary output voltage tracking problem is therefore to find an admissible $u_i = V_{ni}$ for each DG_i such that all $v_{oi} \rightarrow V_{nom}$ as $t \rightarrow \infty$, where V_{nom} is the utility prescribed nominal grid voltage.

$$\dot{\boldsymbol{x}}_i = f_i(\boldsymbol{x}_i) + g_i(\boldsymbol{x}_i) u_i \quad (8)$$

$$\boldsymbol{x}_i = \left[\delta_i, P_i, Q_i, \varphi_{\alpha i}, \varphi_{\beta i}, \gamma_{\alpha i}, \gamma_{\beta i}, i_{L\alpha i}, i_{L\beta i}, v_{o\alpha i}, v_{o\beta i}, i_{o\alpha i}, i_{o\beta i}, v_{b\alpha}, v_{b\beta}\right]^T \quad (9)$$

Other non-linear effects could be included in order to accurately model a practical hardware implementation, for example inductor magnetics and transistor switching characteristics. Due to parametric and partial-model uncertainty it may be difficult to discover the full dynamics. The following section describes a data-driven approach based on state information processing for controller synthesis.

4. Output Voltage Synchronization with Unknown Nonlinear Dynamics

Given the above DG_i model, consider a MicroGrid consisting of a network of spatially interconnected inverters, loads and distribution lines and where each DG_i can communicate with neighbouring DG_j. Assume also that each DG_i output voltage must be controlled to within utility prescribed norms. Clarke-transformed state variable components from (9) now undergo a further transformation to a rotating reference frame shown as $\alpha\beta 0 \rightarrow dq0$ using δ_i. The state vector is $\boldsymbol{x}_i$ and let y_i be the output voltage, with the quadrature axis value of the output voltage assumed to be zero (v_{oqi}=0). As described above, the secondary voltage control problem for a cooperative controller is to design a distributed u_i in (8) for each DG_i such that $y_i \rightarrow V_{nom}$, that is all DG_i output voltages track the reference value.

The dynamics of the converter are expressed by the affine formulation (13) below, where u_i is the input voltage reference (12). An admissible control input can in principle be established using for example feedback linearization. With knowledge of the dynamic model, that is $f(x)$ and $g(x)$, a direct relationship between input u_i and output y_i can be established

$$y_i = h_i(\boldsymbol{x}_i) \quad (10)$$

$$y_i = v_{odi} \quad (11)$$

$$u_i = V_{ni} \quad (12)$$

as in [6], using the method of Lie derivatives. The problem satisfies the conditions for convergence and synchronizes to the external reference.

The above problem can be formulated as an output synchronization problem [16], [17], where the output is required to track an exogenous reference signal with dynamics given by (16)-(17). This requires design of a distributed observer given by (18), and the input protocol is described by an (as yet unknown) function in (19). The problem is equivalent to the voltage tracking problem described above, specifically, we must solve for gains c and K in (18) and determine the unknown control protocol $\varphi_i(x_i, w_i)$ that guarantees output synchronization over time as shown in (20). Refer to [3] for a method to determine the coupling gain c in (18).

The solution is the control protocol shown in (21), which requires solving the output regulator equations (22)-(23), which can be interpreted as the necessary conditions for output synchronization [18], [19]. The additional requirement is that feedback matrix K_i is stablizing (Hurwitz) around a small signal approximation for $f_i(x_i)$ and $g_i(x_i)$. Refer to [20] for the convergence conditions and proofs. Solving the output regulator equations is not straightforward and requires the system model.

$$\dot{x}_i = f_i(x_i) + g_i(x_i)u_i \tag{13}$$

$$y_i = C_i x_i \tag{14}$$

$$x_i(t) \in \mathbb{R}^{n_i} \tag{15}$$

$$\dot{x}_o = s(x_o) \tag{16}$$

$$y_o = C_o x_o \tag{17}$$

$$\dot{w}_i = s(w_i) - cK \sum_{j \in N_i} e_{ij}\big(w_i - w_j\big) + b_i(w_i - x_o) \tag{18}$$

$$u_i = \varphi_i(x_i, w_i) \tag{19}$$

$$\lim_{t \to \infty} (y_i - y_o) \to 0 \tag{20}$$

$$\varphi_i(x_i, w_i) = K_i\big(x_i - \pi_i(w_i)\big) + h_i(w_i) \tag{21}$$

$$\frac{\partial \pi_i}{\partial x_o} s(x_o) = f_i\big(\pi_i(x_o)\big) + g_i(\pi_i(x_o))h_i(x_o) \tag{22}$$

$$C_i \pi_i(x_o) - C_o x_o = 0 \tag{23}$$

An alternative formulation in [20] improves the transient response and suggests a model-free reinforcement learning algorithm based on policy iteration and temporal difference learning. Using this approach, the augmented system is given by (24) and the system dynamics by (25). The synchronization error e_i is given by (27) and the distributed feedback by (28), as above it is required that $e_i \to 0$ over time. In this case the controller performance is given explicitly by the value function in (29), which is now an infinite horizon optimal control problem and therefore conducive to application of the Hamilton-Jacobi-Bellman (HJB) Equation in order to find the optimal distributed control u_i.

$$X_i = \begin{bmatrix} x_i \\ w_i \end{bmatrix} \quad F_i(X_i) = \begin{bmatrix} f_i(x_i) \\ s(w_i) \end{bmatrix} \quad G_i(X_i) = \begin{bmatrix} g_i(x_i) \\ 0 \end{bmatrix} \quad D = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad C_i = [C_i - C_o] \tag{24}$$

$$X_i = F_i(X_i) + G_i(X_i)u_i + Dz_i \tag{25}$$

$$e_i = C_i X_i \tag{26}$$

$$e_i = y_i - C_o w_i \tag{27}$$

$$z_i = \sum_{j=1}^{N} a_{ij}\big(w_i - w_j\big) + b_i(w_i - x_o) \tag{28}$$

$$V_i(X_i) = \int_t^{\infty} e^{-\alpha_i(\tau - t)}\big(X_i^T Q_{iT} X_i + u_i^T R_i u_i\big) d\tau \quad Q_{iT} = [C_i - C_o]^T Q_i [C_i - C_o] \tag{29}$$

Theorem 1 [21]: Let $V_i(X_i)$ be the solution to the HJB equation (31), and the distributed control protocol given by (30). Assuming the discount factor α_i condition in (32) is met, then the output regulator equations (22) and (23) are implicitly solved. The proof is omitted, refer to [20].

$$u_i^* = -\frac{1}{2}R_i^{-1}G_i^T\nabla V_i \quad (30)$$

$$X_i^T Q_{iT} X_i + \nabla V_i^T F_i - \alpha_i V_i - \frac{1}{4}\nabla V_i G_i^T R_i^T G_i \nabla V_i = 0 \quad (31)$$

$$\alpha_i \leq 2\left\|(B_i^T R_i^{-1} B_i Q_i)^{\frac{1}{2}}\right\| \quad (32)$$

A solution to (31) is then required for optimal controller synthesis in a practical sense. Solving the HJB equation for nonlinear systems is generally not feasible using analytic techniques. Is also requires knowledge of $f_i(x_i)$ and $g_i(x_i)$ from (13). The next section describes an algorithm which allows for model-free controller synthesis by off-policy temporal difference learning.

5. Model-Free Reinforcement Learning Algorithm

Reinforcement learning encompasses a family of general iterative algorithms for finding optimal strategies in a range of applications. The concept is based on a staged process of policy evaluation and progressive policy improvement by a critic and actor on a system or external environment and assumes system behavior consistent with a Markov Decision Process (MDP). The advantages are that optimal strategies can be found, or at least approximated, with incomplete information and learnt either online or offline depending on the application [21]. The continuous time formulation of the algorithm, also known as integral reinforcement learning, is described in [22]. Practical learning schemes utilize a value function approximator based on for example radial or polynomial basis function networks [24], also known as approximate dynamic programming (ADP) or neuro-dynamic programming (NDP).

A more recent innovation utilizes model-free computational techniques based solely on data acquisition [24], [25]. This technique applies an exploring control policy, which is arbitrary though stabilizing, to generate a dynamic trajectory of sampled measurements over a suitable time period. The system dynamics are embedded in this information. The optimal control policy is then solved for as described below. Using approximate dynamic programming (ADP) as described in [10], the first step is to express the value function and control protocol in basis function network approximator form as shown in (33) and (34). Assuming suitable basis function selection, then the problem turns to the determination of appropriate weights γ_r and δ_r.

Given any admissible and stabilizing initial control policy $u = u_0 + e$ for DG_i, where e is exploration noise, equation (31) can be transformed into (35) below, refer to [18] or [21] for this procedure. This equation can be utilized to simultaneously approximate (solve) for V_i^n and u_i^{n+1}. This is normally an iterative process and requires a data acquisition stage following by an iterated least squares solver stage, to approximate an optimal control policy.

$$V_i^n(x) = \sum_{r=1}^{N_1} \gamma_{i,r}^n \emptyset_{i,r} \quad (33)$$

$$u_i^{n+1} = \sum_{r=1}^{N_2} \delta_{i,r}^n \emptyset_{i,r} \quad (34)$$

$$\begin{aligned} e^{-\alpha_i T} V_i^n\big(X_i(t+T)\big) - V_i^n\big(X_i(t)\big) \\ = \int_t^{t+T} e^{-\alpha_i(\tau-t)}\big(-X_i^T Q_{iT} X_i - u_i^{nT} R u_i^n\big)d\tau \\ + \int_t^{t+T} e^{-\alpha_i(\tau-t)}\left[(-2u_i^{(n+1)T} R(u-u_i^n)) + \Delta V_i^n D z_i\right]d\tau \end{aligned} \quad (35)$$

Using the function approximators as given in (33) and (34), we can substitute these into (35) to derive the basis function approximator form of the problem (36), which uses the gathered data and least squares regression to solve for the weights $\gamma_{i,r}^n$ and $\delta_{i,r}^n$. The least squares solver requires a sufficiently large sampled data set of state and input measurements, gathered along a single trajectory, to populate the full rank matrices required for a consistent solution of the weights. This data is stored in memory and then (36) is solved using a least-squares solver in repeated iterations until the weights converge such that i) the value functional V_i is identified and ii) the optimal control policy $u^*$is found and then applied. Once the sample time T is selected, each sequential dataset is required to be at least $N > N_1 + N_2$ in size.

$$e^{-\alpha_i T}\sum_{r=1}^{N_1}\gamma_{i,r}^{n}[\emptyset_{i,r}(t_{n+1})-\emptyset_{i,r}(t_n)] \tag{36}$$
$$=\int_{t_k}^{t_{k+1}} e^{-\alpha_i(\tau-t)}\left(-X_i^T Q_{iT} X_i - u_i^{nT} R u_i^n\right)d\tau$$
$$+\int_{t_k}^{t_{k+1}} e^{-\alpha_i(\tau-t)}\left[-2\left(\left[\sum_{r=1}^{N_2}\delta_{i,r}^{n}\emptyset_{i,r}(t)\right]^T R(u-u_i^n)\right)+\Delta V_i^n Dz_i\right]d\tau$$

The key advantage of the off-policy method is that there is no need to test and evaluate multiple sub-optimal control policies online, effectively the algorithm does this in a single calculation step. This circumvents the problem of inadequate exploration, which is critical for reinforcement learning. The optimal feedback control is computed at the end of the exploring phase. The control block diagram is shown in Fig. 3. In (38) the integrand terms are separable in the weights, so simple integrators can be applied [24]. For computing the ordinary least-squares (OLS) to solve for the weights, either MATLAB function "mldivide(A,B)" or "lscov(A,B)" are robust solvers which can be utilized. The OLS solver calculation is terminated according to an arbitrary convergence test for successive solutions. Polynomial basis functions are common, in the following example the control input u_i uses odd polynomials up to degree 3, while the value function V_i uses even polynomials up to degree 4.

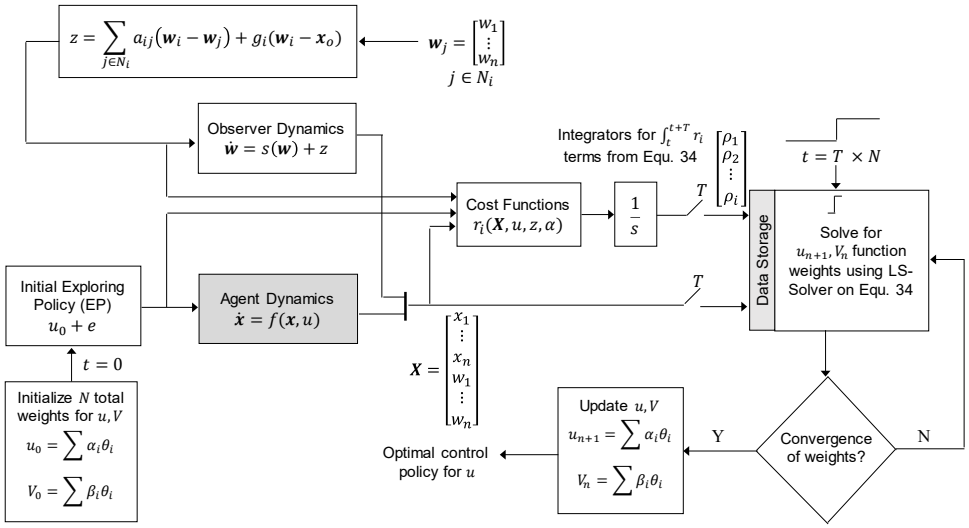


Figure 3. Reinforcement Learning (IRL) controller for Agent-based Output Synchronization

ADP RI-Algorithm 1.

1. Initialize $w_i(0)$, $x_i(0)$, and control policy $u=V_{ni}(0)$
2. Select a sample time T and total number of samples $N>N_1+N_2$.
3. Collect system state vector X_i and input trajectory data, including w_j from neighboring DG_j
4. Solve Equ. 34 for $\gamma_{i,r}^n$ and $\delta_{i,r}^n$, let $n=n+1$
5. Repeat Step 4 until

$$\sum_{r=1}^{N_1}\left|\gamma_{i,r}^{n}-\gamma_{i,r}^{n-1}\right|^2\le\varepsilon$$

6. Apply feedback control policy $V_{ni}=u_i^{n+1}$ to DG_i converter secondary control loop

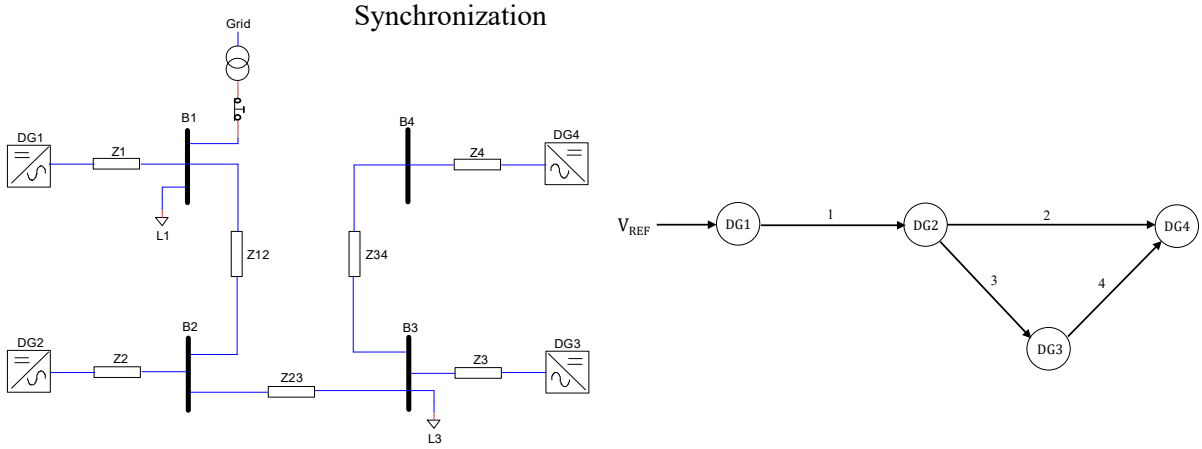


Figure 4. Reference 4-Bus MicroGrid

6. Simulation Results

A cooperative secondary controller using the described Reinforcement Learning (RL) algorithm has been simulated using MATLAB/Simulink on a 4-bus reference MicroGrid shown in Fig. 4. An external voltage reference value is provided to DG_1, and the communication Graph structure is also shown. The controller block diagram is shown in Fig. 3. The voltage reference is initially 0.95 p.u. and increased to 1.05 p.u. after 3 s. To better demonstrate the improvement in performance that the algorithm achieves, with regard to simultaneously tracking the voltage reference signal at each DG, the system is perturbed with a 75 kW load step at 3 s into the simulation, at the same time an exploration phase of 1 s duration commences and terminates at 4 s. At this point in the simulation, the iterated solver routine described in the previous section runs and calculates the optimal weights for a new control policy, which is then applied for repeated test conditions. The simulation results are shown in Fig. 5-6, while the simulation parameters are shown in Table 1. The dynamic response in Fig. 5c-d shows accurate tracking of the voltage reference using the synthesized controller, which approximates the HJB-optimal solution.

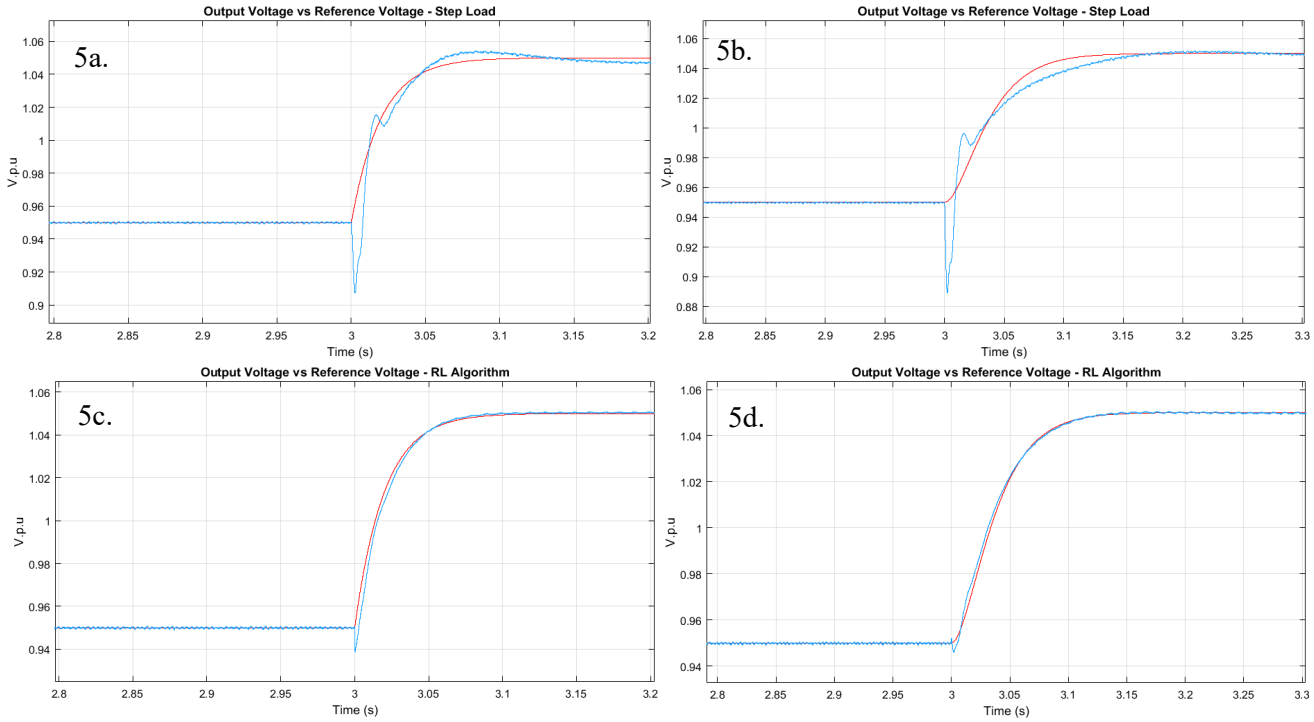


Figure 5. Dynamic response of output voltage to step change in V_{REF} for DG_1 and DG_2 before (upper plot) and after (lower plot) new control policy is applied.

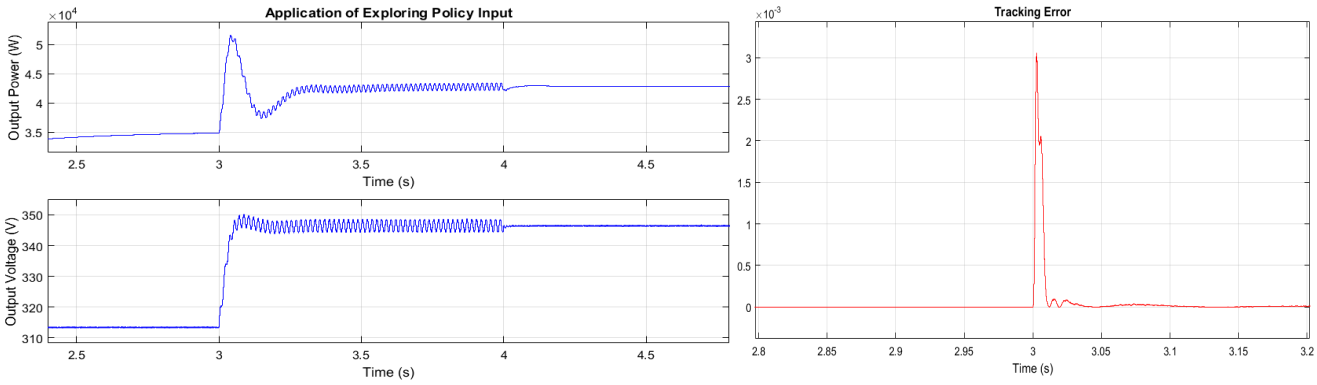


Figure 6. Application of exploring policy (left plot), and tracking error of output voltage response (right plot)

	DG1	DG2	DG3	DG4
DG_i converter parameters				
P_{nom}	100 kVA	90 kVA	80 kVA	70 kVA
V_{nom}	230/400 Vrms	230/400 Vrms	230/400 Vrms	230/400 Vrms
f_{nom}	50 Hz	50 Hz	50 Hz	50 Hz
L_f	1.0 mH	1.0 mH	1.0 mH	1.0 mH
C_f	5.53e-04 F	5.53e-04 F	5.53e-04 F	5.53e-04 F
f_{PWM}	1.35 kHz	1.35 kHz	1.35 kHz	1.35 kHz
K_{RES-V}	400	400	400	400
K_{P-V}	0.35	0.35	0.35	0.35
K_{RES-I}	100	100	100	100
K_{P-I}	0.7	0.7	0.7	0.7
m_{Pi}	2.5e-6	2.857e-6	3.33e-6	4.0e-6
n_{Qi}	2.0e-6	2.286e-6	2.664e-6	3.2e-6
DG_i distributed observer				
c_i	50.0	50.0	50.0	50.0
K_i	1.0	1.0	1.0	1.0
DG_i model-free controller parameters				
T_s	2.5 ms	2.5 ms	2.5 ms	2.5 ms
N_s	400	400	400	400
N_1	45	45	45	45
N_2	24	24	24	24
R	0.25	0.25	0.25	0.25
Q	100	100	100	100
α	-0.5	-0.5	-0.5	-0.5
Electrical network impedances				
$Z1$	$(0.03 - 0.11j)\Omega$	$Z12$	$(0.23 - 0.10j)\Omega$	$L1$	3.2Ω
$Z2$	$(0.03 - 0.11j)\Omega$	$Z23$	$(0.35 - 0.58j)\Omega$	$L3_{T<3s}$	6.4Ω
$Z3$	$(0.03 - 0.11j)\Omega$	$Z34$	$(0.23 - 0.10j)\Omega$	$L3_{T\geq 3s}$	1.6Ω
$Z4$	$(0.03 - 0.11j)\Omega$				

Table 1. Simulation parameters for the 4-bus MicroGrid

Basis Term	$x1$	$x2$	$x1x3^2$	$x1x3x4$	$x4^3$
Iteration No.	w_1	w_3	w_{11}	w_{18}	w_{24}
1	-3.9085	-18.1132	48.1561	16.2643	-41.5611
2	-2.0287	-7.2372	25.5231	10.3104	-22.5524
3	-1.1123	-1.3979	14.7396	7.7429	-13.6634
4	-0.6833	1.9698	11.4165	7.1597	-11.3746
5	-0.2700	2.8489	21.4599	-0.732	-21.7232
6	-0.0914	-5.9234	5.1576	0.7207	-3.6907
7	-0.3061	-1.0261	7.2703	0.4757	-6.8605
8	-0.0862	-5.4343	3.9014	2.0925	-2.5703
9	-0.2717	1.7305	5.2377	0.9565	-5.4078
10	-0.2706	1.8182	5.1272	0.9574	-5.3172
11	-0.2709	1.8146	5.1539	0.9561	-5.3427
12	-0.2708	1.8159	5.1489	0.9563	-5.3381
13	-0.2708	1.8156	5.1499	0.9563	-5.3391
14	-0.2708	1.8157	5.1496	0.9563	-5.3388
15	-0.2708	1.8157	5.1498	0.9563	-5.3388
16	-0.2708	1.8157	5.1498	0.9563	-5.3388
17	-0.2708	1.8156	5.1498	0.9563	-5.3388

Table 2. Sample of control function weights for the first 17 iterations of the algorithm. Convergence occurs after 10 iterations.

7. Practical Considerations for Hardware Implementation

Application of an exploring policy for controller tuning is not a recent development. Adaptive online tuning of PID-controller gains is an established method used in automation, for example the commissioning phase of industrial drive systems. However, it is natural to consider practical constraints for a power system application with respect to scheduling controller adjustments and system-wide impacts. For example, adding exploring noise at particular frequencies may be attenuated through an appropriate grid-coupling transformer. As an Agent-based control technique, there are important considerations with respect to communication links. Developments in wireless network protocols for cooperative control are an active research area. Wireless link delays, transmission rate and interference impact the dynamic response. MAC-level wireless protocols are particularly relevant, and can be probabilistic and contention-based, or deterministic scheduling methods for channel access [11]. SmartGrid communication research in areas such as M2M, and cognitive radio will determine the architecture of future wireless networks. Since the link performance and controller performance are coupled, joint optimization should be considered in the design of a communication system [27], though this topic is beyond the scope of this paper. Further work considers robust stability in the presence of for example switching graph topologies and packet delays.

A working controller has been implemented successfully using MATLAB/Simulink. The control law which is synthesized approximates the HJB-optimal solution for the non-linear dynamics. Further enhancements to the control policy with respect to robustness are considered in [10]. For practical considerations regarding basis function selection and exploring noise refer to [23] and [24]. The algorithm is computationally intensive for processor and memory resource usage but is amenable to implementation on recent digital power DSP and MCU hardware, while numerical methods to expedite a solution are suggested in [9], [24]. Voltage regulation requires the converter to generate some reactive power, which may degrade the active power capability. The design matrices Q and R are selected for the desired dynamic response.

References

[1] J. M. Guerrero, M. Chandorkar, T. Lee, and P. C. Loh, "Advanced Control Architectures for Intelligent Microgrids; Part I: Decentralized and Hierarchical Control," *Ind. Electron. IEEE Trans.*, vol. 60, no. 4, pp. 1254–1262, 2013.

[2] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[3] Z. Li, G. Wen, Z. Duan, and W. Ren, "Designing Fully Distributed Consensus Protocols for Linear Multi-agent Systems with Directed Graphs," vol. 60, no. 4, pp. 1152–1157, 2013.

[4] Z. Li, W. Ren, X. Liu, and M. Fu, "Consensus of multi-agent systems with general linear and lipschitz nonlinear dynamics using distributed adaptive protocols," *IEEE Trans. Automat. Contr.*, vol. 58, no. 7, pp. 1786–1791, 2013.

[5] A. Das and F. L. Lewis, "Automatica Distributed adaptive control for synchronization of unknown nonlinear networked systems ☆," *Automatica*, vol. 46, no. 12, pp. 2014–2021, 2014.

[6] A. Bidram, S. Member, A. Davoudi, F. L. Lewis, J. M. Guerrero, and S. Member, "Distributed Cooperative Secondary Control of Microgrids Using Feedback Linearization," vol. 28, no. 3, pp. 3462–3470, 2013.

[7] A. Bidram, S. Member, A. Davoudi, F. L. Lewis, and S. S. Ge, "of Inverter-Based Microgrids," vol. 29, no. 4, pp. 862–872, 2014.

[8] N. M. Dehkordi, N. Sadati, and M. Hamzeh, "and Voltage Control of Islanded Microgrids," vol. 32, no. 2, pp. 675–685, 2017.

[9] B. Luo, H. Wu, T. Huang, and D. Liu, "Automatica Data-based approximate policy iteration for affine nonlinear continuous-time optimal control design ☆," *Automatica*, vol. 50, no. 12, pp. 3281–3290, 2014.

[10] Y. Jiang, S. Member, and Z. Jiang, "Robust Adaptive Dynamic Programming and Feedback Stabilization of Nonlinear Systems," vol. 25, no. 5, pp. 882–893, 2014.

[11] H. Liang, B. J. Choi, W. Zhuang, X. Shen, A. S. A. Awad, and A. Abdr, "Multiagent coordination in microgrids via wireless networks," *IEEE Wirel. Commun.*, vol. 19, no. 3, pp. 14–22, 2012.

[12] J. C. Vasquez, J. M. Guerrero, M. Savaghebi, J. Eloy-Garcia, and R. Teodorescu, "Modeling, analysis, and design of stationary-reference-frame droop-controlled parallel three-phase voltage source inverters," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1271–1280, 2013.

[13] D. G. Holmes, T. A. Lipo, B. P. McGrath, and W. Y. Kong, "Optimized design of stationary frame three phase AC Current regulators," *IEEE Trans. Power Electron.*, vol. 24, no. 11, pp. 2417–2426, 2009.

[14] N. Pogaku, M. Prodanović, and T. C. Green, "Modeling, analysis and testing of autonomous operation of an inverter-based microgrid," *IEEE Trans. Power Electron.*, vol. 22, no. 2, pp. 613–625, 2007.

[15] K. De Brabandere, B. Bolsens, J. Van den Keybus, A. Woyte, J. Driesen, and R. Belmans, "A Voltage and Frequency Droop Control Method for Parallel Inverters," *IEEE Trans. Power Electron.*, vol. 22, no. 4, pp. 1107–1115, 2007.

[16] N. Chopra, "Output synchronization on strongly connected graphs," *IEEE Trans. Automat. Contr.*, vol. 57, no. 11, pp. 2896–2901, 2012.

[17] J. Xiang, W. Wei, and Y. Li, "Synchronized Output Regulation of w," *IEEE Trans. Automat. Contr.*, vol. 54, no. 6, pp. 1336–1341, 2009.

[18] C. I. Byrnes and A. Isidori, "Output Regulation for Nonlinear," *Proc. 37th IEEE Conf. Decis. Control*, no. December, 1998.

[19] A. Isidori, "Output Regulation of Nonlinear Systems [E]," vol. 35, no. 2, 1990.

[20] H. Modares, F. L. Lewis, and A. Davoudi, "Optimal output synchronization of nonlinear multi-agent systems using approximate dynamic programming," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2016-Octob, pp. 4227–4232, 2016.

[21] P. R. Montague, "Reinforcement Learning: An Introduction, by Sutton, R.S. and Barto, A.G.," *Trends Cogn. Sci.*, vol. 3, no. 9, p. 360, 1999.

[22] V. L. S. Lewis, Frank L. , Draguna Vrabie, *Optimal Control 3rd*, vol. XXXIII, no. 2. 2012.

[23] R. W. Beard, N. Saridiss, and J. T. Went, "Galerkin Approximations of the Generalized Equation *," vol. 33, no. 12, pp. 2159–2177, 1997.

[24] Y. Jiang and Z. P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, no. 10, pp. 2699–2704, 2012.

[25] T. Bian, Y. Jiang, and Z. Jiang, "Automatica Adaptive dynamic programming and optimal control of nonlinear," *Automatica*, vol. 50, no. 10, pp. 2624–2632, 2014.

[26] S. K. Mazumder, K. Acharya, and M. Tahir, "Joint Optimization of Control Performance and Network Resource Utilization in Homogeneous Power Networks," *IEEE Trans. Ind. Electron.*, vol. 56, no. 5, pp. 1736–1745, 2009.