# Frequency Control Using Online Learning Method for Island Smart Grid with EVs and PVs

Due to the intermittent power generation from renewable energy in the smart grid (i.e., photovoltaic (PV) or wind farm), large frequency fluctuation occurs when the load-frequency control (LFC) capacity is not enough to compensate the unbalance of generation and load demand. This problem may become worse when the system is in island operating.

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

### ABSTRACT

Due to the intermittent power generation from renewable energy in the smart grid (i.e., photovoltaic (PV) or wind farm), large frequency fluctuation occurs when the load-frequency control (LFC) capacity is not enough to compensate the unbalance of generation and load demand. This problem may become worse when the system is in island operating.

### SUBJECT TERMS

Online learning, intelligent grid, adaptive control, computational intelligence
Frequency Control Using On-line Learning Method for Island Smart Grid with EVs and PVs

ABSTRACT

Due to the intermittent power generation from renewable energy in the smart grid (i.e., photovoltaic (PV) or wind farm), large frequency fluctuation occurs when the load-frequency control (LFC) capacity is not enough to compensate the unbalance of generation and load demand. This problem may become worsen when the system is in island operating. Meanwhile, in the near future, electric vehicles (EVs) will be widely used by customers, where the EV station could be treated as dispersed battery energy storage. Therefore, the vehicle-to-grid (V2G) power control can be applied to compensate for inadequate LFC capacity, thus improving the island smart grid frequency stability. In this paper, an on-line learning method, called goal representation adaptive dynamic programming (GrADP), is adopted to coordinate control of units in an island smart grid. In the controller design, adaptive supplementary control signals are provided to proportional-integral (PI) controllers by online GrADP according to the utility function. Simulations on a benchmark smart grid with micro turbine (MT), EVs and PVs demonstrate the superior control effect and robustness of the proposed coordinate controller over the original PI controller and fuzzy controller.

Conference Name:  Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN 2014)
Conference Date:  July 06, 2014
Frequency Control Using On-line Learning Method for Island Smart Grid with EVs and PVs

Yufei Tang, Student Member, IEEE, Jun Yang, Member, IEEE, Jun Yan, Student Member, IEEE, Zhili Zeng, and Haibo He Senior Member, IEEE,

Abstract—Due to the intermittent power generation from renewable energy in the smart grid (i.e., photovoltaic (PV) or wind farm), large frequency fluctuation occurs when the load-frequency control (LFC) capacity is not enough to compensate the unbalance of generation and load demand. This problem may become worsen when the system is in island operating. Meanwhile, in the near future, electric vehicles (EVs) will be widely used by customers, where the EV station could be treated as dispersed battery energy storage. Therefore, the vehicle-to-grid (V2G) power control can be applied to compensate for inadequate LFC capacity, thus improving the island smart grid frequency stability. In this paper, an on-line learning method, called goal representation adaptive dynamic programming (GrADP), is adopted to coordinate control of units in an island smart grid. In the controller design, adaptive supplementary control signals are generated according to the utility function and then fed into PI controllers. Moreover, the goal representation network in the adopted GrADP algorithm could provide much more informative internal reinforcement signal to the critic network, which facilitate the learning process and improve the control performance.

I. INTRODUCTION

Due to environmental and energy security concerns, electric vehicles (EVs) and renewable energy from wind or solar have been widely deployed in the smart grid during the last decade. The large-scale integration of these new types of generation and load in power grids will have significant impact on power grid operation, planning and stability control [1]. One important issue is large frequency fluctuation caused by intermittent renewable energy in island operating smart grid, where load-frequency control (LFC) capacity is not enough to compensate the unbalance of generation and load demand. In this paper, we focus on the design of a power system frequency oscillation damping controller based on a new kind of adaptive dynamic programming (ADP) algorithm, the goal representation adaptive dynamic programming (GrADP).

In traditional load-frequency controller design, e.g. the proportional-integral (PI) controller, a linearized power system model near the operating point is used [2]. The nominal design model is obtained for a particular operating condition. After off-line tuning of the parameters, extensive field testing is done at the time of commissioning. A PI controller based on this approach can be well tuned to an operating condition and provide excellent frequency damping over a certain range around the design point. However, new type of generation and load in the smart grid requires us to relax this linearity assumption. In island smart grid with photovoltaics (PVs) and EVs, system state parameters and operating conditions are changing rapidly, therefore, the fixed parameters of a PI controller may not be optimal for the whole set of possible operating conditions and configurations. Considering the facts above, it is desirable to develop a stabilizer which considers the non-linear time-varying nature of a plant and has the ability to adjust its parameters on-line according to its environment.

Meanwhile, with vehicle-to-grid (V2G) technique, EVs can act as controllable loads and mobile storage devices, which will bring new solution for frequency regulation service [3]. By providing active power to load or absorbing extra active power from generation, EVs could quickly compensate the active power gap between generation and load, thus improving the power grid frequency stability. Intensive investigations on EVs have been carried out by the research community, such as load frequency control using V2G system considering the customer convenience of EVs [4], integration of V2G in a real power system in western Denmark [5], fuzzy logic controller based V2G for frequency regulation [6][7], supplementary LFC with both EVs and heat pump water heaters [8], and coordinated frequency control between wind power and V2G [9].

Based on the aforementioned discussions, a new on-line learning method, called GrADP, is adopted in this paper to provide supplementary control signal to the original PI controller. The advantage of ADP based design is that no system model or off-line training is needed for the supplementary controller [10]. Once a system state is observed and send to the GrADP controller, adaptive supplementary control signals are generated according to the utility function and then fed into PI controllers. Moreover, the goal representation network in the adopted GrADP algorithm could provide much more informative internal reinforcement signal to the critic network, which facilitate the learning process and improve the control performance.

The rest of this paper is organized as follows. Section II introduces the benchmark power system used in this paper. Section III presents the on-line learning GrADP algorithm. Section IV gives the detailed load frequency controller design.
based on the adopted method. Simulations are presented in Section V, and conclusions are drawn in Section VI.

II. BENCHMARK POWER SYSTEM DESCRIPTION

The benchmark power system used in this paper is shown in Fig. 1. In this paper, the benchmark power system is in island operating mode, where the system power flow are balanced by local generation and load. Micro turbines (MTs) and photovoltaic (PV) arrays provide active power to local load, such as smart homes. Two EV stations could be treated as dispersed battery energy storages to compensate the unbalance of generation and load demand. All the measured signals (i.e., system frequency, active power) are collected at the point of common coupling (PCC) by a distributed management system (DMS). Then control signals are sent back to each unit (i.e., micro turbine, EV station) to improve the system stability.

Since this system is in island operating, the load-frequency control (LFC) capacity is not enough to damp the frequency oscillation. After incorporating the EVs into LFC, the system inertial could be increased, thus improving the island smart grid frequency stability. Fig. 2 shows the system frequency dynamics after active power disturbance with EVs and without EVs. It is shown that with EVs, the system frequency performance after disturbance is largely improved. Moreover, the frequency performance with EVs could still be further improved by using advanced intelligent controller. The following section will briefly introduces the adopted on-line learning method, followed by the controller design for this benchmark system.

III. ON-LINE LEARNING METHOD: GRADP

The adopted on-line learning method is called goal representation adaptive dynamic programming (GrADP), which is a new reinforcement learning algorithm from the family of adaptive critic designs (ACDs) developed in recent years [11]. It requires three function approximation networks: a goal network, a critic network, and an action network. The critic network learns to approximate the cost-to-go function in Bellman’s equation; the action network learns to generate a control policy that minimizes the cost-to-go approximated by the critic network; the goal network provides an adaptive internal reinforcement signal along with the primary reinforcement signal to the critic network to improve generalization and learning capability [12]. Specifically, the cost-to-go function is defined as follows:

$$J[x(i), \hat{i}] = \sum_{t=i}^{\infty} \gamma^{t-i} U[x(t), u(t), i]$$  

where $x(t)$ is the state vector of the system, $u(t)$ is the control action, $U$ is the utility function, and $\gamma$ is a discount factor.

In this paper, all three networks are implemented as neural networks with a three-layer nonlinear architecture that includes only one hidden layer, in which the sigmoid function is used exclusively. Meanwhile, the learning principles can also be generalized and extended to any arbitrary function approximator by properly applying the back-propagation rule.
sequence \( u(t) \) so the cost function \( J \) is minimized:

\[
J^*(x(t)) = \min_{u(t)}\{ U(x(t), u(t)) + \gamma J^*(x(t+1)) \} \tag{2}
\]

This is the foundation for implementing dynamic programming by working backward in time domain. In this structure, \( J \) can be estimated by minimizing the following error over time:

\[
\| E_h \| = \frac{1}{2} \sum_t [J(t) - r(t) - \alpha J(t + 1)] \tag{3}
\]

When \( E_h = 0 \) for all \( t \), (3) indicates:

\[
J(t) = r(t) + \alpha J(t + 1) \tag{4}
\]

With one time step backward, we can write:

\[
J(t - 1) = r(t - 1) + \alpha J(t) \tag{5}
\]

From equation (4) and (5), the objective function to be minimized in the goal network [13][14] can be written as:

\[
\begin{cases}
    e_g(t) = \alpha J(t) - [J(t - 1) - r(t - 1)] \\
    E_g(t) = \frac{1}{2} e_g^2(t)
\end{cases} \tag{6}
\]

and the high-level conceptual back-propagation path is:

\[
\frac{\partial E_g(t)}{\partial \omega_g(t)} = \frac{\partial E_g(t)}{\partial J(t)} \frac{\partial J(t)}{\partial s(t)} \frac{\partial s(t)}{\partial \omega_g(t)} \tag{7}
\]

For the three-layer critic network, the weights of the output layer and hidden layer in the two parts as follows:

\[
\begin{cases}
    \Delta \omega_{g_{i}}^{(2)} = \eta_g(t) \left[ - \frac{\partial E_{g}(t)}{\partial \omega_{g_{i}}(t)} \right] \\
    \Delta \omega_{g_{i,j}}^{(1)} = \eta_g(t) \left[ - \frac{\partial E_{g}(t)}{\partial \omega_{g_{i,j}}(t)} \right] \tag{8}
\end{cases}
\]

2) Critic Network Training: Once the \( s(t) \) signal is obtained from the goal network, it will be used as an input to the critic network, and also be used to define the error function to adjust the parameters of the critic network:

\[
\begin{cases}
    e_c(t) = \alpha J(t) - [J(t - 1) - s(t)] \\
    E_c(t) = \frac{1}{2} e_c^2(t)
\end{cases} \tag{9}
\]

And the back-propagation path is:

\[
\frac{\partial E_c(t)}{\partial \omega_c(t)} = \frac{\partial E_c(t)}{\partial J(t)} \frac{\partial J(t)}{\partial u(t)} \tag{10}
\]

The weights of the output layer and the hidden layer of the critic network are updated respectively as follows:

\[
\begin{cases}
    \Delta \omega_{c_{i}}^{(2)} = \eta_c(t) \left[ \frac{\partial E_{c}(t)}{\partial \omega_{c_{i}}(t)} \right] \\
    \Delta \omega_{c_{i,j}}^{(1)} = \eta_c(t) \left[ \frac{\partial E_{c}(t)}{\partial \omega_{c_{i,j}}(t)} \right] \tag{11}
\end{cases}
\]

3) Action Network Training: The procedure of adapting the action network in this architecture is similar to the classic ADP approach to implicitly back-propagate the error between a desired ultimate object \( U_c \) and the approximate \( J \) function of the critic network [10], \( U_c \) is in accordance with the signal of the reinforcement when the state conducted by the action implies a success. Therefore, the error function used to update the parameters of the action network is:

\[
\begin{cases}
    e_a(t) = J(t) - U_c(t) \\
    E_a(t) = \frac{1}{2} e_a^2(t) \tag{12}
\end{cases}
\]

Since the action network is connected with both goal network and critic network, the back-propagation path will consist of two parts as follows:

\[
\begin{cases}
    \frac{\partial E_a(t)}{\partial \omega_a(t)} = P_{a,c}(t) + P_{a,g}(t) \\
    P_{a,c}(t) = \frac{\partial E_a(t)}{\partial \omega_a(t)} \frac{\partial \omega_a(t)}{\partial u(t)} = \frac{\partial E_a(t)}{\partial \omega_a(t)} \frac{\partial u(t)}{\partial \omega_c(t)} = \frac{\partial E_a(t)}{\partial \omega_a(t)} \frac{\partial \omega_c(t)}{\partial \omega_c(t)} \frac{\partial \omega_c(t)}{\partial u(t)} \tag{13}
\end{cases}
\]

And the weights of the output layer and hidden layer in the action network are updated as follows:

\[
\begin{cases}
    \Delta \omega_{a_{i}}^{(2)} = \eta_a(t) \left[ \frac{\partial E_{a}(t)}{\partial \omega_{a_{i}}(t)} \right] \\
    \Delta \omega_{a_{i,j}}^{(1)} = \eta_a(t) \left[ \frac{\partial E_{a}(t)}{\partial \omega_{a_{i,j}}(t)} \right] \tag{14}
\end{cases}
\]
IV. Design of the Controller

A. Input and reinforcement signal design

The proposed GrADP based controller with the island benchmark power system is shown in Fig. 4. It is modeled and implemented in Matlab/Simulink environment, where the MT fuel system, turbine and two EVs are represented by transfer functions. The active power fluctuation from the PVs are modeled as power disturbance $\Delta P_D$, which will cause system frequency oscillation. The deviation of the system active power and the inertia of the island smart grid $H_f$ are then used to generate the system frequency deviation $\Delta f$. In the original design, three PI controllers are used to control the MT and two EVs to damp the frequency oscillation. In this paper, the GrADP controller is modeled as a supplementary controller to help the original PI controllers to improve the damping performance.

The input of the GrADP controller is designed as follows:

$$x(t) = [\Delta f(t) \quad \Delta f(t - 1) \quad \Delta f(t - 2)]$$

(15)

where $\Delta f$ is the measured system frequency deviation. One-time and two-time delay are used here to provide the controller more system dynamic information under disturbance [15][16]. The output of the GrADP controller is $\Delta u_s$, which will be added by the output of three PI controllers to form new controller actions to the MT and EVs. In order to prevent to dominate the PI control, a limitation unit is imposed to $\Delta u_s$. The reinforcement signal of the GrADP controller is designed as follows [17]:

$$Q = \text{diag} [1 \quad 0.5 \quad 0.5^2]$$

$$r(t) = -x(t) * Q * x(t)'$$

(16)

Based on the aforementioned input and reinforcement signal design, the parameters used in the GrADP controller are shown in Tab. I.

<table>
<thead>
<tr>
<th>Parameters Used in GrADP Controller</th>
<th>Networks</th>
<th>Action</th>
<th>Goal</th>
<th>Critic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs Number</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Outputs Number</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Hidden Neurons</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Activation Function</td>
<td>Sigmoid</td>
<td>Sigmoid</td>
<td>Sigmoid</td>
<td></td>
</tr>
</tbody>
</table>

B. Work flow and parameter setting

Once the GrADP controller is initialized, it will be plugged into the system and works in the following procedure [18][19]:

1) The action network receives the measured plant state $x(t)$ and uses it to generate the control signal $\Delta u_s$ to the PI controllers.

2) The goal network uses the external reinforcement signal $r(t)$ and plant state $x(t)$ to generate the internal reinforcement signal $s(t)$.

3) Then the critic network uses the internal reinforcement signal $s(t)$, plant state $x(t)$, and control signal $\Delta u_s$ to estimate the cost function $J$.

4) The goal network will update its weights according to equation (6), (7) and (8) until the stop criterion is satisfied.

5) The critic network will update its weights according to equation (9), (10) and (11) until the stop criterion is satisfied.

6) The action network will update its weights according to equation (12), (13) and (14) until the stop criterion is satisfied.

7) Step (1) - (6) are repeated in each simulation time step until the end of the simulation.

V. Simulation studies

In this section, comparisons of the original PI controller, a fuzzy logic controller and the GrADP controller to damp frequency oscillation are carried out on the benchmark power system. Active power deviation from PVs are modeled as the power disturbance for the system.

A. Case 1: active power disturbance without EVs constraints

In this case, there are no output constraints of the two EVs. Three sequential active power disturbances are applied to the system with appropriate time interval. Specifically, a $+0.1p.u.$ step disturbance is applied at 5s, a $–0.1p.u.$ step disturbance is applied at 40s and a $+0.1p.u.$ step disturbance is applied at 80s. Under this sequential disturbance, the system frequency deviation with PI controller, fuzzy logic controller and GrADP controller is shown in Fig. 5, respectively. Three learning stages of the GrADP controller are observed from this figure, which correspond to the three sequential disturbances. Since the weights of the neural networks are randomly initialized in stage one (5s to 40s), the GrADP controller does not generate proper control strategy. The performances of the original PI controller and the GrADP controller are similar. In stage two (40s to 80s), GrADP approach utilizes the knowledge learned from stage one, which results in generating proper control action. This improves the control performance compared to the original PI controller. In stage three (80s to 120s), the control performance of the GrADP has been further improved. The peak value in the first swing has been largely decreased and the system frequency oscillation has been quickly damped. While it can also be observed that the fuzzy logic controller performances well in stage two, without the on-line learning ability, the performance in the other two stages is not as well as in stage two.

The outputs of the MT, EVs with PI controller, fuzzy logic controller and GrADP controller are shown in Fig. 6, respectively. Since there are no constraints of the two EVs, the outputs of the two EVs are the same. Three stages of output response are clearly shown in this figure. Fig. 7 shows the three sequential active power disturbances and the corresponding GrADP output. As mentioned before, a limitation of $\pm 0.02$ is imposed to $\Delta u_s$. We can see that, although stage one and stage three yield the same $+0.1p.u.$ step disturbance, the controller generated different control action, demonstrating the superior learning ability of the proposed GrADP controller.
B. Case 2: active power disturbance with EVs constraints

In case 2, two different output constraints are imposed to the EVs. Specifically, the output constraints of EV1 is ±0.04 and the output constraints of EV2 is ±0.02. The same three sequential active power disturbances in case 1 are also applied to the system with appropriate time interval. Due to the added output constraints, the system will need longer time to damp the frequency oscillation with the same disturbance. So the disturbances are sequentially applied at 5s, 80s and 140s, resulting in three learning stages (5s to 80s, 80s to 140s and 140s to 220s). A learning process is observed from the system frequency deviation with PI controller, fuzzy logic controller and GrADP controller, as shown in Fig. 8. After the first two learning stages, the GrADP controller demonstrates superior damping performance over the PI controller and the fuzzy logic controller.

The outputs of the MT, EVs with PI controller, fuzzy logic controller and GrADP controller are shown in Fig. 9, respectively. It can be seen that the outputs of the two EVs are different under the additional constraints. An interesting observation is that the output of the MT is increased in this case. Since the capacity of the EVs are constrained, the system will need the MT to output more active power to damp the system frequency oscillation. In reality, EV stations are usually relatively small, so the capacity of MT should be carefully considered based on the foreseeable disturbance from renewable energies. ∆u_s of the action network, J value of the critic network and r value of the reinforcement signal in case 2
are shown in Fig. 10, respectively, which illustrated the instant learning ability of the GrADP controller.

![Fig. 9. MT, EV 1 and EV 2 output with PI, fuzzy logic and GrADP controller in case 2](image)

![Fig. 10. GrADP output, J value and r value in case 2](image)

VI. CONCLUSIONS

This paper adopted a new computational intelligence algorithm, called GrADP, to design supplementary controller to improve system frequency stability in island smart grid with EVs and PVs. The proposed controller design, including the input signal selection and the reinforcement signal design, is introduced in details. Comparative studies with the original PI controller and a fuzzy logic controller are shown through two simulation study cases with three sequential active power disturbances, which demonstrate the superior instant learning ability and control effect of the proposed GrADP controller.

Optimal intelligent control methods have been introduced into power system operation for more than ten years. However, considerable efforts are still needed for real power engineering applications. This paper may shed some light on narrowing the gap between the theoretic research and realistic applications. Considering the adaptive ability of damping system oscillation of the proposed GrADP controller, it may be interesting to apply it to other power system oscillation problems, such as low-frequency oscillation problems in connected grids by long distance tie-lines in the future studies.

REFERENCES


[13] Z. Ni, H. He, and J. Wen, “Adaptive learning in tracking control based on the dual critic network design,” Neural Networks and Learning Systems, 


