Q-Learning with Eligibility Traces to Solve Non-Convex Economic Dispatch Problems

Mohammed I. Abouheaf, Sofie Haesaert, Wei-Jen Lee, Frank L. Lewis

Abstract—Economic Dispatch is one of the most important power system management tools. It is used to allocate an amount of power generation to the generating units to meet the load demand. The Economic Dispatch problem is a large scale nonlinear constrained optimization problem. In general, heuristic optimization techniques are used to solve non-convex Economic Dispatch problem. In this paper, ideas from Reinforcement Learning are proposed to solve the non-convex Economic Dispatch problem. Q-Learning is a reinforcement learning technique where each generating unit learns the optimal schedule of the generated power that minimizes the generation cost function. The eligibility traces are used to speed up the Q-Learning process. Q-Learning with eligibility traces is used to solve Economic Dispatch problems with valve point loading effect, multiple fuel options, and power transmission losses.

Keywords—Economic Dispatch, Non-Convex Cost Functions, Valve Point Loading Effect, Q-Learning, Eligibility Traces.

I. INTRODUCTION

THE operation of the power system is tightly controlled to achieve the efficient use of its capabilities [1]. The operation cost of the generating units highly depends on the fuel cost. The Economic Dispatch (ED) is a modern power system energy management tool. It results in the best economical use of the generating units and fuel sources [2].

The high nonlinearity of the power system imposes mathematical complexities in formulating the generation cost models necessary to solve the Economic Dispatch problem [1]. The sources of the mathematical complexities are due to the design specifications and operation constraints of the generating units such as the spinning reserve, transmission losses, prohibited operation zones, ramp rate limit, valve point loading effect, and multiple fuel options [1]. The spinning reserve determines how the generating units are robust to the unexpected outages or incorrect load allocation among the generating units [3]. The Prohibited zones are caused by faults in the machines itself or the associated auxiliaries [4]. Restrictions in the power system define the ramp rate limits constraints [3]. In addition, some turbines use multiple valves that are opened sequentially to satisfy the load requirement, which adds more non-convexity to the generation cost function [3]. Some generating units use multiple fuel types for different operation regions [5].

The conventional methods used to solve Economic Dispatch problem include Newton Raphson, gradient techniques, lambda iteration method, the base point and participation factors method [1], interior point algorithm, linear programming, dynamic programming and dual quadratic programming [6]-[8] where the generation cost functions are assumed to be monotonically increasing piecewise linear functions. Heuristic optimization techniques are used to find the optimal solution for the non-convex Economic Dispatch problem. These techniques include Evolutionary programming (EP) [9], Genetic Algorithm (GA) [4], Differential Evolution [10], Particle Swarm Optimization (PSO) [11], Simulated annealing (SA) [12], Tabu Search [13], Gravitational Search Algorithm (GSA) [2], and Biogeography method [14]. These Heuristic algorithms don’t always guarantee the global best solution.

In this paper, ideas from Reinforcement Learning (RL) are used to solve the non-convex Economic Dispatch problem. Reinforcement Learning is an area of machine learning, used to solve multi-stage decision making problems. It is concerned with how an agent will pick its actions in a dynamic environment to transit to new states in such a way that the optimization of the objective function can be achieved [15]-[20].

This paper is organized as follows. In Section II, the classical Economic Dispatch problem is introduced. In Section III, Q-Learning and eligibility traces are introduced. In Section IV, an algorithm based on Q-Learning with eligibility traces is developed to solve non-convex Economic Dispatch. In Section V, simulation is performed using the developed algorithm to solve the Economic Dispatch problem with valve point loading effect, multiple fuel options, and transmission losses.

II. FORMULATION OF THE ECONOMIC DISPATCH PROBLEM

In this section the classical Economic Dispatch problem is formulated using Lagrange dynamics [1]. The main operation constraints related to the generating units are mentioned. Furthermore, the different generation cost models are introduced.

A. Economic Dispatch Problem

Lagrange dynamics is used to formulate and solve the Economic Dispatch problem. The objective of the optimization problem is to minimize the fuel generation cost, so that
Minimize \( F_i = \sum_{i=1}^{n} F_i(P_i), \forall i \) \hspace{1cm} (1)

where \( F_i \) is the total fuel generation cost and it is given by \( F_i = F_i + F_2 + F_3 + \ldots + F_n \), \( F_i \) is the fuel generation cost of each unit \( i \), \( P_i \) is the power generated by each unit \( i \), and \( n \) is the number of generating units.

The generation cost function \( F_i \) is approximated by the quadratic function [1], so that

\[ F_i(P_i) = a_i + b_i P_i + c_i P_i^2 \] \hspace{1cm} (2)

where \( a_i, b_i, \) and \( c_i \) are the fuel cost coefficients of the generation unit \( i \).

Equation (2) states the basic generation cost model. The Lagrange operator is given so that

\[ L = F_i + \sum_{i=1}^{n} \lambda_i \phi_i \] \hspace{1cm} (3)

where \( \phi_i \) is the number of constraints, \( \lambda_i \) is the Lagrange multiplier associated with each constraint \( \phi_i \).

The Lagrange operator \( L \) is minimized with respect to the generated power, while the constraints \( \phi_i, \forall i \) are satisfied [1].

### B. Operation Constraints

The generating units’ constraints are classified into two types. The first is related to the design and operation specifications of the generating units such as the generation capacity, line maximum power flow, generation ramp limits, prohibited operation zones constraints, and spinning reserve. The second is related to an upper level of operation control such as unit commitment and other operation plans like maintenance. Here, only constraints related to the work are considered.

1) Generation-Demand Equality Constraints

The Generation–Demand equality constraint, states that the sum of the generated power is equal to the total active load demand plus the transmission losses so that

\[ \sum_{i=1}^{n} (P_i) = P_D + P_{\text{losses}}, \] \hspace{1cm} (4)

where \( P_D \) is the total active load demand, \( P_{\text{losses}} \) is the transmission losses. The transmission losses are given in terms of Kron’s loss formula [5] so that

\[ P_{\text{losses}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \left( P_{ij} T_{ij} \right) + \sum_{i=1}^{n} \left( B_{\text{th}} P_i + B_{\text{con}} \right) \] \hspace{1cm} (5)

where \( B_{\text{th}}, B_{\text{con}}, \) and \( B_{\text{con}} \) are the transmission network power losses coefficients. The B-loss coefficients represent the transmission line and the corona losses [5].

2) Generation Capacity

Each generating unit has maximum and minimum generation capacities so that

\[ P_i^{\text{min}} \leq P_i \leq P_i^{\text{max}}, \forall i \] \hspace{1cm} (6)

where \( P_i^{\text{min}} \) and \( P_i^{\text{max}} \) are the designed minimum and maximum generated power capacities of each unit \( i \).

3) Spinning Reserve Constraints

During the power system operation, the generating units are not working on the maximum designed capacity, instead those units keep about 5-10\% of their capacity unused [21]. This operation enhances the security of the power system in the case of emergencies. Here, the spinning reserve constraints are given only for the generating units without prohibition zones [21] so that

\[ SR_i = \min \left( \left( P_i^{\text{max}} - P_i \right), SR_{i^{\text{max}}(i)} \right), \forall i \text{ without POZ} \] \hspace{1cm} (7)

where \( SR_i \) is the spinning reserve of unit \( i \) (MW), \( SR_{i^{\text{max}}(i)} \) is the maximum spinning reserve of unit \( i \), SR is the total spinning reserve given by the generating units that do not have any Prohibited Operating Zones (POZ).

C. Practical Generation Cost Functions

The simplified generation cost function (2) does not include the valve point loading effect and the multiple fuel types’ effects. Accurate generation cost models are given as follows.

1) Economic Dispatch with Valve Point Loading Effect

The admission valves operate in a sequential manner in some turbines. This sequential operation causes ripples or non-differentiable points in the generation cost models [22]. This effect is modeled by a sinusoidal function so that, the generation cost function is given by

\[ F_i(P_i) = a_i + b_i P_i + c_i P_i^2 + e_i \sin(f_i \times \sin(P_i^{\text{min}} - P_i)) \] \hspace{1cm} (9)

where \( a_i, b_i, c_i, e_i, \) and \( f_i \) are the fuel cost coefficients for each unit \( i \) and \( P_i^{\text{min}} \) is the minimum generated power by each unit \( i \) with valve point loading effect.

2) Economic Dispatch with Multiple Fuel Options

The generating units can use multiple fuel options for different regions in the operation range. This adds more non-convexity to the generation cost function so that
are the fuel cost coefficients of each unit $i$ using fuel type $k$. 

3) Economic Dispatch with Valve Point Loading Effect and Multiple Fuel Options

The generation cost function due to valve point loading effect and multiple fuel options is denoted as “Hybrid cost function” [22]. The hybrid cost function models result from combining the generation cost models (9) and (10) so that

$$
F_i(P) = \begin{cases} 
    a_{i1} + b_{i1}P_i + c_{i1}P_i^2 + |e_{i1} \times \sin(f_{i1} \times \sin(P_{i\min} - P_i))|, & P_{i\min} \leq P_i \leq P_{i1} \\
    a_{i2} + b_{i2}P_i + c_{i2}P_i^2 + |e_{i2} \times \sin(f_{i2} \times \sin(P_i - P_{i1}))|, & P_{i1} \leq P_i \leq P_{i2} \\
    \vdots & \vdots \\
    a_{ib} + b_{ib}P_i + c_{ib}P_i^2 + |e_{ib} \times \sin(f_{ib} \times \sin(P_{i(k-1)} - P_i))|, & P_{i(k-1)} \leq P_i \leq P_{i(max)} 
\end{cases}
$$

(10)

where $a_{ia}, b_{ia}, c_{ia}$ are the fuel cost coefficients for each unit $i$ and fuel type $k$.

III. REINFORCEMENT LEARNING WITH ELIGIBILITY TRACES

In this section, the Reinforcement Learning (RL) is used to solve the Economic Dispatch problem with valve point loading effect, multiple fuel options, and transmission losses. Ideas from Reinforcement Learning, Markov Decision process, Q-Learning, and eligibility traces are introduced [16].

The Reinforcement Learning (RL) algorithm developed herein learns the optimal power distribution for the Economic Dispatch problem by interacting with the environment i.e. choosing the proper generating values to minimize the generation cost objective functions [23].

A. Markov Decision Process

Reinforcement Learning requires a mapping of the continuous Economic Dispatch problem structure to a discrete problem structure similar to the Markov Decision Process (MDP) [20].

The normal Markov Decision Process (MDP) is defined by the tuple $M \{ X, U, f, \rho \}$ where $X$ is the discrete set of all possible states, $U$ is the discrete set of all possible actions, $f : X \times U \rightarrow X$ is the state transition function, and $\rho : X \times P \rightarrow R^+$ is the penalty function. The actions are chosen based on a policy $\pi : X \rightarrow U$. This policy minimizes the sum of future costs, this sum is stored in a value function.

B. Q-Learning

Q-Learning is a reinforcement learning technique. The goal of each agent (generating unit) is to learn a policy (scheduling the generated power) that minimizes the penalty function (generation cost function) (1). One way to learn the optimal policy (optimal generation schedule) is by using Q-Learning with the sum of future costs to be defined by the Q-function $Q : X \times U \rightarrow R^1$. The Q-function gives the expected cost for a given state-action pair under a given policy $\pi$ so that

$$
Q(x,u) = \rho(x,u) + \min_{u'} Q(f(x,u),u')
$$

(12)

where $\rho(x,u)$ is the penalty function.

The Q-function is iterated by correcting the old value with the penalty so that

$$
Q^{t+1}(x_i,a_i) = Q^t(x_i,a_i) + \alpha(\rho(x_i,a_i) + \gamma \min_{a_i} \{Q(x_{i+1},\pi) - Q^t(x_i,a_i)\})
$$

(13)

where $\rho(x_i,a_i)$ is the penalty function (generation cost function), $\alpha$ is the learning coefficient, $q$ is the number of iterations, $k$ is the number of the state, $\gamma$ is the discount factor, and $A$ is the set of all possible actions.

The balance between the exploration and exploitation is important in the Q-Learning. Moreover, the proper selection of the action affects the performance of both the learning and evaluation of the agent’s policy [16]. The $\epsilon$-greedy (near-greedy) method is an effective strategy of choosing the best actions during Q-Learning. It acts very well in environments with noisier cost functions. The $\epsilon$-greedy Q-Learning selects the action with the lowest expected cost with probability $1-\epsilon$ and selects a random action from the feasible action set with probability $\epsilon \in [0,1]$ [16].

The $Q(\bar{X})$ learning with eligibility traces is used to speed up the Q-Learning process. As per Sutton and Barto, the eligibility trace $\bar{X}$ temporarily memorizes the parameters associated with an event to be eligible for learning changes in the Q-Learning process [16]. The state-action pairs are backed together and memorized as long as the greedy policy is followed.

IV. $Q(\bar{X})$ LEARNING WITH ELIGIBILITY TRACES ALGORITHM

In this section, an algorithm based on Q-Learning with eligibility traces is developed. Algorithm 1 explains how the Q-Learning algorithm with eligibility traces will learn the optimal generated powers for any applicable active load demand. Next, Algorithm 2 is used to extract the optimal actions (optimal generated power instances) for specified active load demands, taking into consideration the transmission losses. The Markov Decision Process implies that the different stages are seen as the different generation units. The state $x_i$ is defined as the residual power demand
and $k$ is the generator number. The power demand and the action spaces are discretized, whenever the generated power space is originally continuous. The discretization step will be an important factor that will impact the accuracy of the results. The action space $U(x_k)$ is the feasible generated power choices for the state $x_k$. The penalty function is given in terms of the generation cost functions $(9)-(11)$. The Q-Learning process Algorithm is given as follows

Algorithm 1: Q-Learning with Eligibility Traces (Learning Process)
- Identify the minimum and the maximum possible generated power for all the generation units $n_g$.
- Determine the power demand limits, the demand should not be greater than the summation of maximum generated power or not less than the summation of minimum generated power.
- Assume that every generating unit, generates at least its minimum power so the maximum amount of power that needs to be distributed over the generation units is given by

$$P_{D_{max}} = \sum_{i} (P_{max} - P_{min})$$

- Initialize the Q-function, the total number of trials (iterations), and the exploration rate $\epsilon$.

while ($t < \text{trial}_{\text{max}}$) Do 
1. Generate random power demand instance ($P_d$) picked from the uniform distribution over $[0, P_{D_{max}}]$.
2. Define the first state so that $x_1 = (1, P_d)$.
3. Determine optimal action for the first generation unit by
4. Identify the feasible discrete action space $P_1 \in U(x_1)$ so that

$$(x_1 - \sum_{j=2}^{n_g} (P_{max,j})), \sum_{j=2}^{n_g} (P_{min,j}) \leq P_1 \leq x_1$$

$$0 \leq P_1 \leq (P_{max}, -P_{min})$$

5. Retrieve the optimal action $P_{1}^* \in \arg \min Q(x_1, P_1)$ for the next states steps
For $k = 1, ..., n_g - 2$ Do 
4. Apply $\pi$-greedy action
If $\pi < (1 - \epsilon)$ do 
5. The remaining load to be distributed to the next stages or states.
6. Determine optimal action for $(k + 1)$th generator by
7. Identify the feasible action space $P_{k+1} \in U(x_{k+1})$ so that

$$(x_{k+1} - \sum_{j=k+2}^{n_g} (P_{max,j})), \sum_{j=k+2}^{n_g} (P_{min,j}) \leq P_{k+1} \leq x_{k+1}$$

$$0 \leq P_{k+1} \leq (P_{max}, -P_{min})$$

6.2. Retrieve the optimal action $P_{k+1}^*$ from the feasible space such that $P_{k+1}^* = \arg \min Q(x_{k+1}, P_{k+1})$.
7. Update Q-function including trace information
Define the error function ($\Delta_{k}$) so that

$$\Delta_{k} = F(P_{ng-1}) + F(P_{ng}) - Q^*(x_{ng-1}, P_{ng})$$

For $(x_j, p_j)$ with $l \in [k, ..., k]$ $Q^{\pi+1}(x_j, p_j) = Q^l(x_j, p_j) + \alpha \Delta_{l}^{-1}[\Delta_{k}]$$
where $\Delta_{k} = 1 - \exp(-\delta_{l} / \text{trial}_{\text{max}})$
End
8. Repeat steps (3) and (4) for $k = n_g - 1$
9. The feasible action space of last generator is given by

$$P_{ng} = P_{D} - \sum_{k=1}^{n_g} P_{k}$$
10. Update Q-function
Calculate the error function for the last stage (generation unit).

$$\Delta_{k} = F(P_{ng-1}) + F(P_{ng}) - Q^*(x_{ng-1}, P_{ng})$$

For $(x_j, p_j)$ with $l \in [k, ..., , k]$ $Q^{\pi+1}(x_j, p_j) = Q^l(x_j, p_j) + \alpha \Delta_{l}^{-1}[\Delta_{k}]$"
Initially the computational effort is done in the learning process. Once the learning process is done, it is easy to retrieve the optimal power distribution for the generating units for any active load demand [20]. To simplify the simulation of the Q learning with eligibility traces, the learning coefficient is picked so that $\alpha = 1$, where it is multiplied by the exponentially decreasing trace $\lambda^k$, and the discount factor is picked so that $\gamma = 1$.

V. Q-LEARNING: CASE STUDIES AND NUMERICAL SIMULATION

The advantages of the proposed algorithm to solve the Economic Dispatch problem are verified in this section. The Q-Learning with eligibility traces is compared to other Heuristic optimization techniques. Three study cases are considered for the simulation purposes. In case 1, the Q-Learning with eligibility traces is used to solve Economic Dispatch problem for 6 generating units with valve point loading effect and transmission losses. In case 2, the Q-Learning with eligibility traces is used to solve Economic Dispatch problem for 10 generating units with multiple fuel options. In case 3, Q-Learning with eligibility traces is used to solve Economic Dispatch for 15 generating units considering the transmission losses.

A. Case Study 1:

In this case, Q-Learning with eligibility traces algorithm results is compared to other published results for 6 generating units. The fuel cost coefficients and generation capacities of the 6 generating units with valve point loading effect are given in Table I. The simulation parameters (discrete step=7 MW, $\text{trail}_{\text{max}} = 10^5$, $\epsilon = 0.1$, and $\mu = 0.01$).

<table>
<thead>
<tr>
<th>Unit</th>
<th>$a(b)($/h)</th>
<th>$b(b$/MW)</th>
<th>$c(b$/MW $^2$)</th>
<th>$T$</th>
<th>$P_{\text{min}}$</th>
<th>$P_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>240</td>
<td>0.007</td>
<td>0.031</td>
<td>500</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>0.0095</td>
<td>0.042</td>
<td>200</td>
<td>200</td>
<td>300</td>
</tr>
<tr>
<td>3</td>
<td>220</td>
<td>0.009</td>
<td>0.086</td>
<td>300</td>
<td>300</td>
<td>500</td>
</tr>
<tr>
<td>4</td>
<td>200</td>
<td>0.009</td>
<td>0.086</td>
<td>300</td>
<td>300</td>
<td>500</td>
</tr>
<tr>
<td>5</td>
<td>220</td>
<td>0.008</td>
<td>0.086</td>
<td>500</td>
<td>500</td>
<td>700</td>
</tr>
<tr>
<td>6</td>
<td>190</td>
<td>0.0075</td>
<td>0.083</td>
<td>500</td>
<td>500</td>
<td>700</td>
</tr>
</tbody>
</table>

The transmission power losses are expressed in terms of Kron’s loss formula. The losses coefficients $B_1$, $B_2$, and $B_3$ are given as follows:

$$B_1 = \begin{bmatrix} 0.0017 & 0.00012 & 0.00007 & -0.00001 & -0.00005 & -0.00002 \\ 0.0012 & 0.00014 & 0.00001 & -0.00006 & -0.00001 & 0 \\ 0.0007 & 0.00009 & 0.00031 & 0 & -0.001 & -0.00006 \\ -0.0001 & 0.00001 & 0 & 0.0024 & -0.00006 & -0.00008 \\ -0.0005 & -0.00006 & -0.001 & -0.00006 & 0.0129 & -0.00002 \\ -0.0002 & -0.00001 & -0.00006 & -0.00008 & -0.00002 & 0.015 \end{bmatrix}$$

$B_0 = 0.0001[0.3906 -0.1297 0.7047 0.0591 0.2161 -0.6635], B_0 = 0.0056.$

The optimal generated power for different methods are given in Table II for active load demand (PD=1263 MW). The Q-Learning with eligibility traces achieved the lowest fuel generation cost (15439 $/h) compared to BBO [26], and BGA [14] and it was the same as IWD [27] as shown in Table III. Moreover, results for Q-Learning with eligibility traces are given in Table IV for active load demands (1080 MW, 1100 MW, 1220 MW, and 1240 MW).

A. Case Study 2:

In this case, Q-Learning with eligibility traces algorithm results is compared to other published results for 10 generating units. The fuel cost coefficients and generation capacities of the 10 generating units with valve point loading effect and transmission losses are given in Table IV for different active load demands (PD=1263 MW) (without valve effect loading point).

B. Case Study 2:

In this case, Q-Learning with eligibility traces algorithm results is compared to other published results for 10 generating units. The fuel cost coefficients and generation capacities of the 10 generating units with valve point loading effect and transmission losses.
multiple fuel options are given in Table V and Table VI. The simulation parameters (discrete step=7 MW, $trail_{max} = 3 \times 10^3$, $\epsilon = 0.1$, and $\mu = 0.01$).

### Table V

| Case Study 2: Fuel Cost Coefficients, Valve Point Loading Effect, and Fuel Types for 10 Generating Units |
|---|---|---|---|---|---|
| Unit | Fuel | a | b | c | e |
| 1 | 1 | 26.97 | -0.3975 | 0.002176 | 0.027 |
| 2 | 1 | 21.13 | -0.3059 | 0.003861 | -3.1 |
| 3 | 2 | 13.65 | -0.198 | 0.00162 | 0.037 |
| 4 | 1 | 18.4 | -1.269 | 0.004194 | -13 |
| 5 | 2 | 1.865 | -0.03988 | 0.001138 | -0.4 |
| 6 | 1 | 13.65 | -0.198 | 0.00162 | 0.037 |
| 7 | 2 | 13.65 | -0.198 | 0.00162 | 0.037 |
| 8 | 1 | 13.65 | -0.198 | 0.00162 | 0.037 |
| 9 | 2 | 13.65 | -0.198 | 0.00162 | 0.037 |
| 10 | 1 | 26.97 | -0.3975 | 0.002176 | 0.027 |

### Table VI

| Case Study 2: Fuel Options and Generation Units’ Capacities |
|---|---|---|---|---|---|
| Unit | Pmin | Fuel | Pmax |
| 1 | 100 | 196 | Fuel 2 |
| 2 | 50 | 114 | 157 |
| 3 | 200 | 332 | 388 |
| 4 | 99 | 138 | 200 |
| 5 | 190 | 338 | 407 |
| 6 | 85 | 138 | 200 |
| 7 | 200 | 331 | 391 |
| 8 | 99 | 138 | 200 |
| 9 | 130 | 213 | 370 |
| 10 | 200 | 362 | 407 |

C. Case study 3:

In this case, Q-learning with eligibility traces are compared to other published results for 15 generating thermal units, whose fuel cost characteristics and generation capacities are given in Table IX. Moreover, the power system transmission losses are considered. The B loss formula is used to express the transmission losses and the losses coefficients are given in the Appendix. The simulation parameters (discrete step=5 MW, $trail_{max} = 10^3$, $\epsilon = 0.1$, and $\mu = 0.01$).

### Table VII

| Comparison between Q-Learning with Eligibility Traces and Other Methods for Active Load Demand (PD=2700 MW) |
|---|---|---|
| Method | Cost, $/h |
| HM [28] | 625.18 |
| HNN [29] | 626.12 |
| AHNN [30] | 624.5178 |
| EP [31] | 624.5178 |
| CGA-MU [22] | 624.5178 |
| IGA-MU [22] | 624.5178 |
| DE [32] | 624.5178 |
| RGA [4] | 624.5178 |
| PSO [32] | 624.5178 |
| GA [33] | 624.5178 |
| Q-Learning | 624.3116 |

### Table VIII

| Optimal Generated Power and Respective Fuel Options by Q-Learning with Eligibility Traces for (PD=2700 MW) |
|---|---|---|---|---|---|---|
| Unit | P (MW) | Fuel | Unit | P (MW) | Fuel |
| 1 | 219.08 | 10 | 2 | 6 | 240.03 |
| 3 | 211.46 | 11 | 4 | 239.37 | 3 |
| 5 | 288.23 | 19 | 7 | 426.48 | 3 |
| 8 | 233.52 | 21 | 10 | 273.59 | 1 |
| Total Generation (MW) | 2700 | Total Cost ($/h) | 624.3116 |

For active load demand (PD=2700 MW), the Q-Learning algorithm achieved the lowest fuel generation cost (32676$/h) for active load demand (PD=2630 MW) compared to PSO [34], GA [34], SPSO [11], PC_PSO [11],...
TABLE X
COMPARISON BETWEEN Q-LEARNING WITH ELIGIBILITY TRACES AND OTHER METHODS FOR ACTIVE LOAD DEMAND (PD=2630 MW)

<table>
<thead>
<tr>
<th>Method</th>
<th>Cost, $/h</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSO [34]</td>
<td>32858</td>
</tr>
<tr>
<td>GA [34]</td>
<td>33063.54</td>
</tr>
<tr>
<td>SOH-PSO [34] [11]</td>
<td>32751.39</td>
</tr>
<tr>
<td>MTS [33]</td>
<td>32796.13</td>
</tr>
<tr>
<td>SA [35]</td>
<td>32786.4</td>
</tr>
<tr>
<td>SCA [35]</td>
<td>32867.025</td>
</tr>
<tr>
<td>APSO [36]</td>
<td>32742.77</td>
</tr>
<tr>
<td>CPSO [34]</td>
<td>32834</td>
</tr>
<tr>
<td>BF [34]</td>
<td>32784.5</td>
</tr>
<tr>
<td>MDE [34]</td>
<td>32704.9</td>
</tr>
<tr>
<td>TSA [33]</td>
<td>32917.87</td>
</tr>
<tr>
<td>DSPSO-TSA [33]</td>
<td>32715.06</td>
</tr>
</tbody>
</table>

TABLE XI
OPTIMAL GENERATED POWER AND LOSSES BY Q-LEARNING WITH ELIGIBILITY TRACES (PD=2630 MW)

<table>
<thead>
<tr>
<th>Unit</th>
<th>P (MW)</th>
<th>Unit</th>
<th>P (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>403.7229</td>
<td>6</td>
<td>427.4191</td>
</tr>
<tr>
<td>2</td>
<td>426.7635</td>
<td>7</td>
<td>459.4340</td>
</tr>
<tr>
<td>3</td>
<td>124.6209</td>
<td>8</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>121.7764</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>434.7174</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>45.2892</td>
<td>12</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Total Generation (MW)</td>
<td>2665.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total cost ($/h)</td>
<td>32676</td>
<td></td>
</tr>
</tbody>
</table>

VI. CONCLUSION
Q-Learning is used to solve the Economic Dispatch problem with non-convex cost function. Eligibility traces are used to speed up the learning process. The study cases included Economic Dispatch problems with valve point loading effect, multiple fuel options, and transmission losses. Simulation results showed that Q-Learning with eligibility traces achieved the lowest fuel generation cost compared to some Heuristic optimization techniques. The importance of the developed algorithm is that once the learning process is complete, the optimal generated power distribution for any active load demand can be retrieved without any addition efforts unlike other optimization techniques.

APPENDIX
The B loss coefficients are given as follows

\[
B = [-0.0001, -0.0002, 0.0028, -0.0001, 0.0001, -0.0003, -0.0002, 0.0006, 0.0039, -0.0017, 0 -0.0032, 0.0067, -0.0064],
\]

ACKNOWLEDGMENT
This work was supported by NSF grant ECCS-1128050, ARO grant W911NF-05-1-0314, AFOSR grant FA 9550-09-1-0278, China NNSF grant 61120106011, and China Education Ministry Project 111 (No.B08015)

REFERENCES

ISNI:0000000091950263


