A Reinforcement Learning approach to the Optimal Torque MPPT problem in wind turbines

A Reinforcement Learning approach to the Optimal Torque MPPT problem in wind turbines

E Choquehuanca[1]* and A Ortega[2]

[1]Universidad Nacional de Ingeniería, Av. Túpac Amaru 210, Rímac, Lima, Peru
[2]Universidad Autónoma del Perú, Panamericana Sur Km. 16.3, Villa El Salvador, Lima, Peru

*Email: echoquehuancap@uni.pe

Abstract. In this work, a torque controller for a variable rotational speed wind turbine has been modelled using Reinforcement Learning and considering the Optimal Torque - Maximum Power Point Tracking problem as one of optimization. The reward optimization function is designed as a non-linear function depending mainly on the rotor power variation. Based on this, an optimal action (electromagnetic torque variation) regulates the turbine rotational speed. A simulated 1.5 MW three bladed wind turbine operation is managed by the torque controller. It keeps the turbine working at optimal operational conditions after a successful training process, which is carried out using the Proximal Policy Optimization algorithm. For the controller training, the turbine confronts constant and then randomly staggered wind speed behaviour. Time series of rotor angular speed, torque and power are presented. Our results show that the modelled controller is able to reach and maintain the wind turbine operation at its optimal power generation conditions. This methodology avoids using some empirical parameter characteristic of the Optimal Torque - Maximum Power Point Tracking algorithm widely used in wind turbine control systems.

1. Introduction

The wind turbine is a promising technology for clean energy generation. Worldwide, deployments of wind turbine farms, onshore and offshore, are expected to continue to grow in the coming years. Nevertheless, a profitable operation of wind turbines depends on the weather condition, characterized by their stochastic behaviour. Therefore, it is necessary to model control systems aimed at making the wind turbines work at optimal operating conditions. The Optimal Torque - Maximum Power Point Tracking (OT-MPPT) technique is one of the most widely used control system methods. It is based on finding, by "tracking", an optimal turbine rotational speed that creates the state for the generation of an optimal rotor torque for a reference wind speed, thus allowing the turbine to operate at its condition of maximum power generation [1]. However, the OT-MPPT depends on an empirical parameter found after several laboratory tests and after the post-processing of large amounts of data.

On the other hand, technologies such as machine learning (ML) offer alternatives to overcome this challenge, as shown in the last efforts done by [2], [3] and [4]. Reinforcement Learning (RL) is an ML method based on an entity that learns a skill by experience, like a child learning to walk. Due to the nature of the problem (the OT-MPPT problem in wind turbines), this ML method offers a solution to this challenge. In this sense, using RL, a modelled controller (agent) interacts with the turbine (environment) by imposing a torque (action) in such a way that the controller receives a signal

(reward) indicating an increase or decrease in the generated power. This reward can be modelled as an optimization function, so the controller can only receive favourable signals and respond by sending only optimal torques oriented to the wind turbine to work at its optimal operating point.

In this work, a torque controller of a variable rotational speed wind turbine has been modelled. The modelling is based on RL, considering the OT-MPPT problem as one of optimization, allowing the wind turbine to work at optimal operation conditions.

2. Analytical model for wind turbines

This section describes the mathematical model of the 1.5 MW three bladed wind turbine utilised for the training of the controller [1], [5]. The rotor power of a wind turbine can be calculated by:

$$P_{\mathrm{r}} = \frac{1}{2}\rho\pi R^2 C_{\mathrm{p}} v_{\mathrm{w}}^3 \tag{1}$$

where ρ is the air density (1.25 kg/m^3), R is the rotor blade radius (38.5 m), C_{p} is the power coefficient, and v_{w} is the wind speed (m/s). For the calculation of the power coefficient, the following correlation was used [6]:

$$C_{\mathrm{p}}(\lambda,\beta) = 0.22\left(\frac{116}{m} - 0.4\beta - 5\right)\mathrm{e}^{-12.5/m} \tag{2}$$

where λ is the ratio of the speed at the blade tip to the wind speed, equation (3). The parameter β is the pitch angle of the blade profile (0 deg), and m is given by equation (4).

$$\lambda = \frac{w_{\mathrm{r}} R}{v_{\mathrm{w}}} \tag{3}$$

in this expression, the rotor angular speed (rad/s) is represented by w_{r}.

$$\frac{1}{m} = \frac{1}{\lambda + 0.08\beta} - \frac{0.035}{\beta^3 + 1} \tag{4}$$

The changes in the rotor angular speed can be calculated based on the relationship between the torque, inertia and damping of the mechanical and electrical components:

$$J_{\mathrm{t}}\dot{w}_{\mathrm{r}} = T_{\mathrm{r}} - K_{\mathrm{t}} w_{\mathrm{r}} - T_{\mathrm{g}} \tag{5}$$

in this expression, T_{r} and T_{g} are the rotor and electromagnetic torque (Nm), respectively. The parameters J_{t} and K_{t} represent the total contribution of the mechanical and electrical components to the inertia and damping, respectively. The total inertia can be calculated by:

$$J_{\mathrm{t}} = J_{\mathrm{r}} + n_{\mathrm{g}}^2 J_{\mathrm{g}} \tag{6}$$

where J_{r} represents the inertia due to the turbine rotor and blades (4456761 kg-m^2), and J_{g} represents the inertia due to the electric generator shaft and rotor (123 kg-m^2). The damping effects were neglected for this work. The angular speed ratio in the gearbox is represented by n_{g} (105.494). In this paper, the electromagnetic torque is represented by T_{g} and not by $n_{\mathrm{g}}T_{\mathrm{g}}$ in order to simplify the mathematical expressions for easy reading.

Finally, the relation between the rotor torque and power is represented by:

$$P_{\mathrm{r}} = T_{\mathrm{r}} w_{\mathrm{r}} \tag{7}$$

3. Reinforcement Learning modelling

In Reinforcement Learning, an “agent” learns to take “actions” to a yet unknown “environment” according to a defined goal. A numerical “reward” sent to the “agent” quantifies how good the imposed “action” was in contributing to reaching the goal. By experience, the “agent” learns which actions bring the most significant long-term “reward” even at the expense of short-term “reward” [7].

Figure 1, internal loop, presents the interaction of the RL characters considered for this work. At time t, the "agent" (controller) receives the "state" s_t of the "environment" (wind turbine) and uses it to formulate an "action" a_t based on a "policy" or criterion π. Thus, $\pi(a_t/s_t)$ is the probability that the action a_t is formulated according to the state s_t. Then the formulated "action" is imposed on the "environment", and it returns to the "agent" a new "state" s_{t+1} and "reward" r_{t+1}.

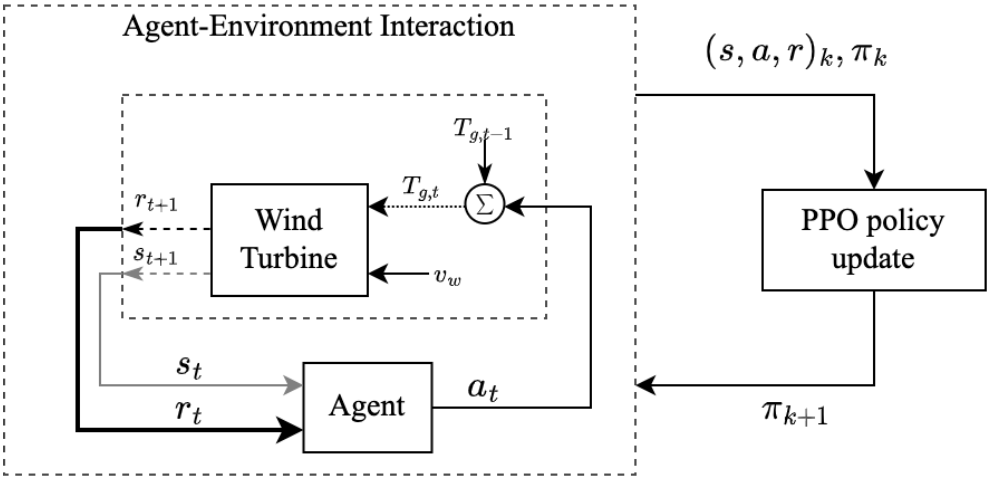

Figure 1. The "agent" - "environment" interaction and the PPO training modelled for this work.

In this work, a series of finite episodes were considered for each iteration (k). Each episode concludes when a time limit (t_{limit}) is reached. An episode may end before t_{limit} if the wind turbine reaches an unrealistic state, which must be avoided, for instance, negative values of w_r (a sudden stop and change of direction of the rotor angular speed). In that case, the agent learns to consider the actions that led the controller to that terminal state as "undesirable actions" since these actions do not allow the agent to obtain more rewards during the episode. The state is composed of:

$$\boldsymbol{s} = \left[v_{\mathrm{w}}, w_{\mathrm{r}}, \dot{w}_{\mathrm{r}}, T_{\mathrm{r}}, T_{\mathrm{g}}\right] \tag{8}$$

The chosen action is the variation of the electromagnetic torque a_t:

$$T_{\mathrm{g,t}} = T_{\mathrm{g,t-1}} + a_t \tag{9}$$

thus, the electromagnetic torque is able to regulate the change of the rotor angular speed according to equation (5). This action is considered to work on a continuous space [8] bounded by the designing characteristics of the turbine, assumed equal to ± 500 kNm/s.

The agent uses a stochastic policy to decide the value to take from the action space. This policy was parameterized using a two-layer neural network (NN) and a normal distributed probability density function [7]. The NN receives the state vector as input and estimates the mean of the probability function. The standard deviation of the probability function is a "trainable" parameter that is updated during the training process. It is considered the agent's exploration parameter for the action space. In case this parameter decreases, it makes the agent's actions progressively less random.

We designed a reward function (optimization function) as a non-linear function depending on two variables:

$$r_{\mathrm{t}} = k_{\mathrm{p}}\left(P_{\mathrm{r,t}} - P_{\mathrm{r,t-1}}\right) - k_u\left(T_{\mathrm{g,t}} - T_{\mathrm{g,t-1}}\right)^2 + k_{\mathrm{a}} \tag{10}$$

The first addend on the right side of this equation uses the Potential-Based Reward Shaping (PBRS) method [9] to speed up and guide the training process. It adds a positive reward to an increase in rotor

power. The second addend penalizes too high action signals, thus keeping a safe and stable controller. The last addend is a small constant reward that encourages the agent to stay alive, thus not "falling" into situations that end the episode before t_{limit}. This situation can result from negative rewards that can lead the agent to believe that terminating the episode is the best alternative to maximize the reward. The coefficients k_p (0.001), k_u (0.1) and k_a (0.05) are the tuning hyperparameters calculated by trial and error.

The Proximal Policy Optimization (PPO) algorithm was used for the training of the agent. It was selected due to its good performance in continuous action space for complex environments [10]. The advantage function needed for the PPO calculation was estimated using the Generalized Advantage Estimation (GAE) [11]. Figure 1, external loop, shows the PPO training process for this work. It starts with the agent-environment interaction for a horizon set equal to 8192 time steps to collect the current state, action, and reward. The collected data and the current policy are then used in the outer process (external loop) to update the policy for each k iteration.

4. Simulations and results

In order to evaluate the performance of the modelled controller, two simulations were carried out. The first considers a constant wind speed (Figure 2), and the second assumes a randomly staggered behaviour for the wind speed (Figure 3). As initial conditions, it was assumed that the turbine starts working with a T_g and a w_r equal to 105.494 kNm and 1 rad/s, respectively.

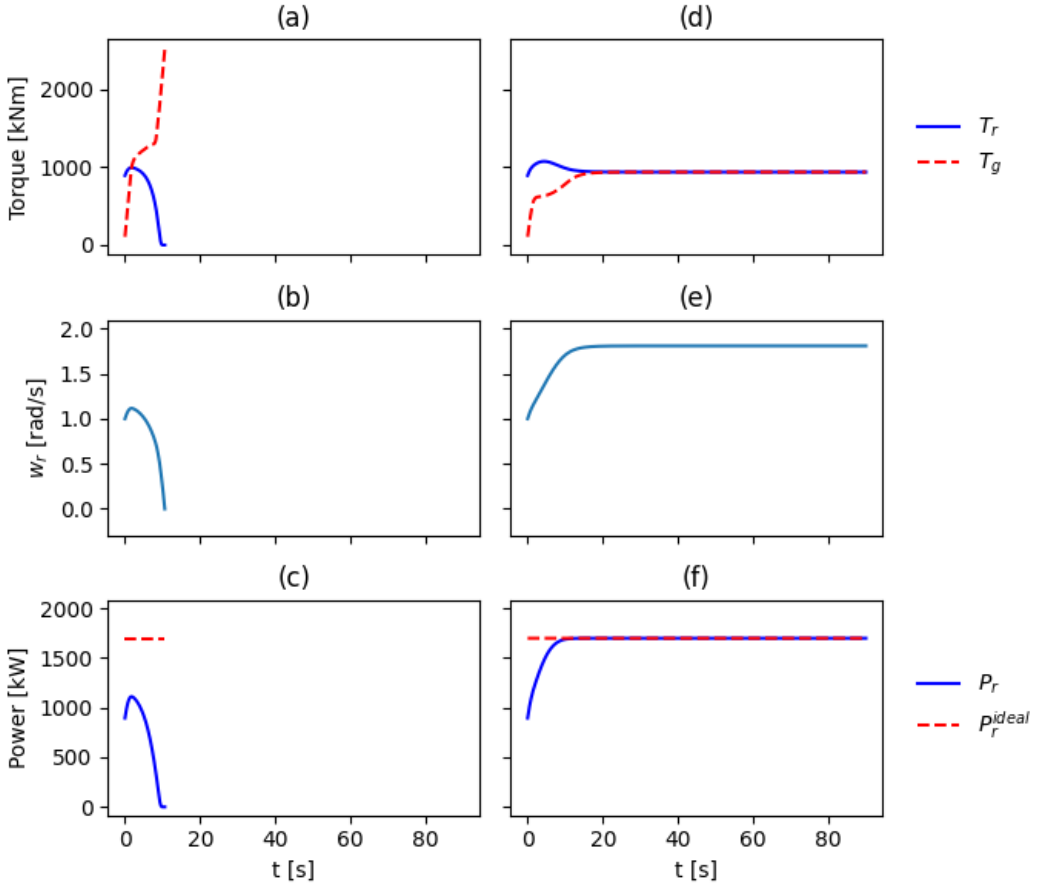

Figure 2. Response of the torque-controlled wind turbine for the iteration 50 (left) and 100 (right) - constant wind speed case.

4.1. Constant wind speed

As a proof of concept, the agent was trained in the environment with a constant wind speed equal to 11 m/s during 90 s. Figure 2 (left) shows the response of the turbine for the iteration 50. Similar results were obtained for the iteration 0. The dotted red lines represent the ideal behaviour that the turbine should have. Even though the turbine starts to increase its rotational speed (b) and power (c) during the first instants of time, the agent has a self-destructive behaviour that quickly takes the turbine to stop (undesirable state).

However, for the iteration 100, the controller was able to manage the turbine to work at its optimal operation condition, Figure 2 (right). This figure shows that after a short transient period and for a time close to 20 s, the turbine reaches its optimal w_r equal to 1.8 rad/s (e) and a maximum P_r equal to 1685 kW (f) for an optimal T_r equal to 936 kNm (d). A similar situation was presented in later iterations.

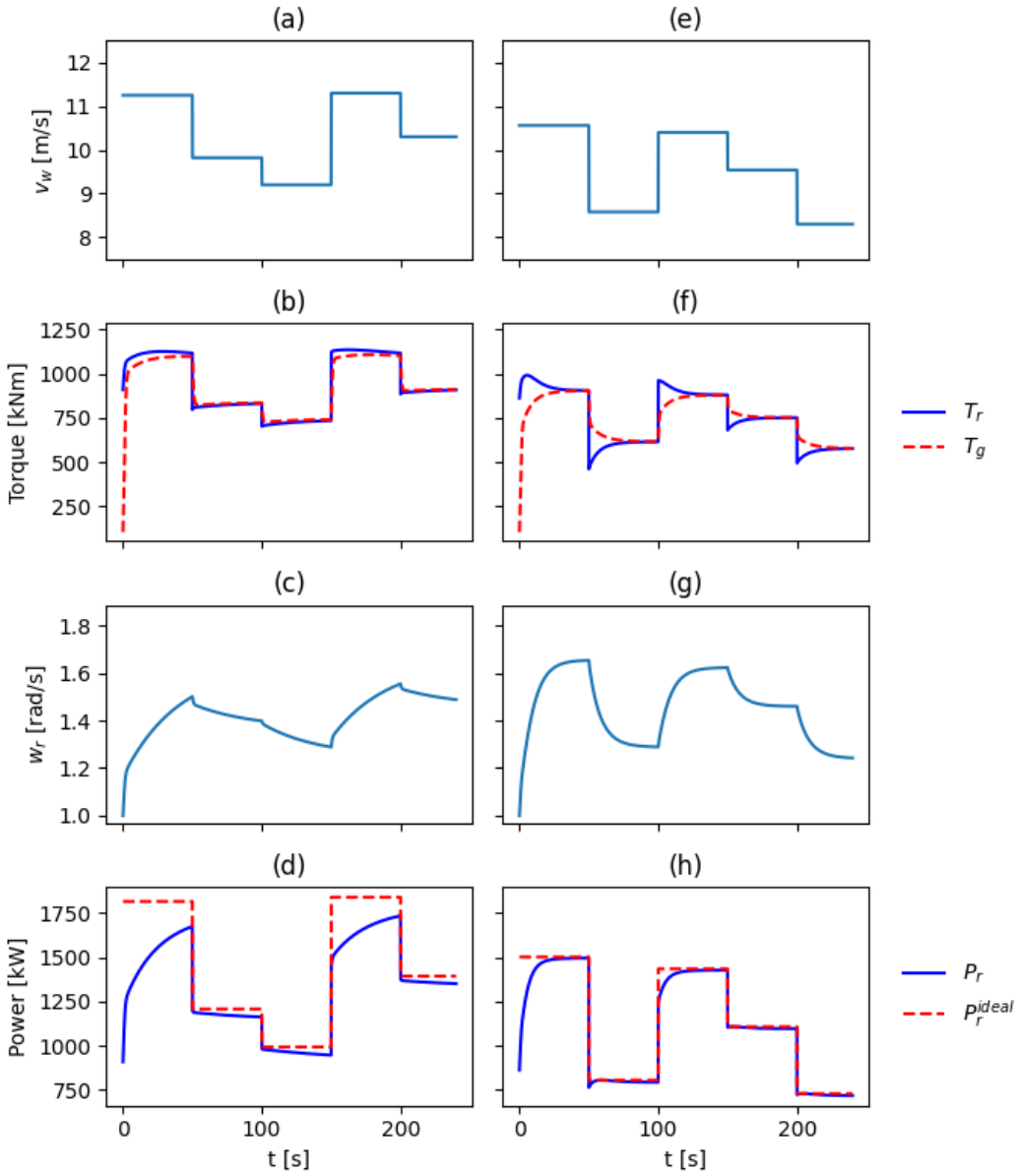

Figure 3. Response of the torque-controlled wind turbine for the iteration 100 (left) and 800 (right) - randomly staggered wind speed case.

4.2. Randomly staggered wind speed

In this case, the controller was trained for the wind turbine facing a randomly staggered wind speed. The v_w magnitude was set to vary between 8 m/s and 12 m/s each 50 s. Similar self-destructive behaviour of the controller was presented in this case for the iterations 0 and 50. However, for the iteration 100, the controller starts to manage the wind turbine response. Figure 3 (left) shows the imposed randomly staggered wind speed over time (a). The controller acts by imposing a dynamic T_g trying to reach the optimal torque for the current wind speed (b). The turbine responds by changing w_r (c) and T_r, the last one with values close to the T_g (b). Despite this, the turbine is not yet able to reach its maximum power (d). It is due to the sudden change of the wind speed magnitude, the inertia of the mechanical and electrical components, and because the controller is still in the learning process.

Now, for the iteration 800, Figure 3 (right), the controller entirely is able to manage the turbine operation. The modelled controller learned to keep the wind turbine working at optimal operation conditions. At each value of v_w (e), the controller calculates an optimal torque (f), with a short transient at the beginning of the v_w change. The turbine responds by generating an optimal w_r (g) and T_r (f), but the short transient is still present. For this iteration, the optimal w_r and T_r combination keeps the turbine working at its maximum rotor power generation (h).

5. Conclusion

In this work, a torque controller for a variable rotational speed wind turbine was modelled based on RL, PBRS for the reward function, and PPO for the training process. The reward function was designed as an optimization function, considering a linear contribution of positive reward for rotor power rising, a negative quadratic penalization for abrupt action change, and a constant live bonus reward. This designed reward function that contemplates the OT-MPPT goal as an optimization problem rather than a reference tracking procedure allowed the torque controller to find the best strategy on its own. By this strategy, our modelled torque controller was able to keep the wind turbine operating under the combination of optimal rotor angular speed and optimal rotor torque for maximum rotor power generation.

An obvious continuation of this work will be facing the torque-controlled wind turbine under an actual wind speed from field measurements or computational simulations. Once trained, the deployment of the torque controller into a complete wind turbine simulator and then the manufacture of the torque controller for laboratory tests will be the future tasks if the previous stages are successful.

References

[1] Kumar D and Chatterjee K 2016 *Renewable and Sustainable Energy Reviews* **55** pp 957–70
[2] Shuvo S S, Islam M M and Yilmaz Y 2022 *North American Power Symp*. (Salt Lake City) (IEEE) pp 1-6
[3] Vu N T-T, Nguyen H D and Nguyen A T 2022 *IEEE Access* **10** pp 95771–80
[4] Bustan D and Moodi H 2022 *J. Modern Power System and Clean Energy* **10** 2 pp 524–30
[5] Meng W, Yang Q and Sun Y 2014 *Proc. of the 33rd Chinese Control Conf.* (Nanjing) (IEEE) pp 8877-82
[6] Slootweg J G, Polinder H and Kling W L 2001 *Power Engineering Society Summer Meeting Conf. Proc.*(Vancouver) vol 1 (IEEE) pp 644-9
[7] Sutton R S and Barto A G 2018 *Reinforcement Learning: An Introduction* (Cambridge: The MIT Press)
[8] Lillicrap T P, Hunt J J, Pritzel A, Heess N, Erez T, Tassa Y, Silver D and Wierstra D 2016 *Proc. of the 4th Int. Conf. on Learning Representations* (San Juan de Puerto Rico) pp 1-14
[9] Ng A Y, Harada D and Russell S 1999 *Proc. of the 16th Int. Conf. on Machine Learning* (San Francisco: Morgan Kaufmann) pp 278–87
[10] Schulman J, Wolski F, Dhariwal P, Radford A and Klimov O 2017 *CoRR* **abs/1707.06347**
[11] Schulman J, Moritz P, Levine S, Jordan M and Abbeel P 2016 *Proc. of the 4th Int. Conf. on Learning Representations* (San Juan de Puerto Rico) pp 1-14