Automatic generation control of a two area power system using deep reinforcement learning

S. Reynolds (BMa., BFin.)

This interim report submitted as part of the assessment schedule for:

BACHELOR OF ENGINEERING HONOURS (ELECTRICAL)

Supervisors:
Dr. C. Yeo & Dr. S. Klaric

Charles Darwin University
College of Engineering

April 30, 2020

Declaration

I, *S. Reynolds (BMa., BFin.)*, declare that this interim report is submitted in partial fulfilment of the requirements for the conferral of the degree *BACHELOR OF ENGINEERING HONOURS (ELECTRICAL)*, from Charles Darwin University, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

S. Reynolds (BMa., BFin.)

April 30, 2020

Abstract

Acknowledgments

Contents

List of Figures

List of Tables

List of Abbreviations

List of Symbols

Chapter 1

Introduction

In 2018, approximately 261TW h of power was generated in the Australian electricity sector. Renewables contributed 19% of the total generation, an increase from 15% in 2017. The Department of Industry, Science, Energy and Resources have observed an increase in renewable energy generation year-on-year in the electricity generation market since 2008, as shown in Figure 1.1 [1].

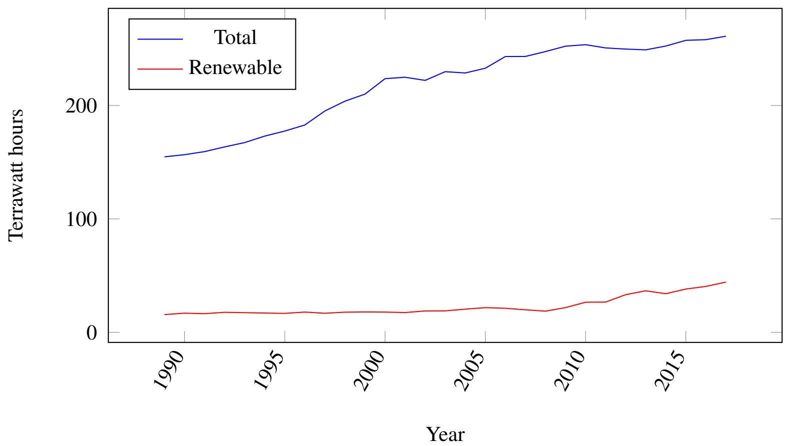

Figure 1.1: Power generation from renewable sources (light blue line), and total power generation (dark blue line) in Australia from 1977 to 2018.

One of the benefits of transitioning from thermal sources of power generation to renewable sources is reduced greenhouse gas emissions [2]; however, this transition is not without its drawbacks. With an increased reliance on renewable power generation sources posing challenges for power system stability owing to load management. A recent example is the system failure in Alice Springs, caused by an event cascade that was triggered by cloud cover shadowing a solar array. The system failure left residents in Alice Springs without power for approximately eight hours [3]. An independent investigation highlighted that poor control policies were one of the factors that contributed to the blackout. In this instance, the generator provisioned to ramp up in the event of cloud cover was unable to be controlled. Moreover, generators that were still under the control regime were issued operating set points above their rated capacity, that resulted in thermal overload and subsequent tripping from the protection system [4].

Current control methods use classical feedback loop techniques. These methods can be brittle when faced with system changes, or scenarios which they were not designed for. An improved controller would be one that can learn and adapt its controller to an unseen system or event, given some broad control objective. This research proposes a deep reinforcement learning (DRL) agent for controlling the frequency of a power system consisting of multiple generators, and multiple load demands with individual stochastic profiles.

1.1 Research Aim

The principle aim of this research is to compare the performance of known, optimised feedback loop controller architectures against a DRL based control system when tasked with performing load following ancillary services with regulating generators under AGC for a two area power system. This research will be undertaken in order to understand the feasibility of using DRL agents for two area power system management.

1.2 Structure of Thesis

Chapter 2

Background

2.1 Power Systems and Frequency

Interconnected power systems are comprised of power generating units and energy storage systems connected to transmission and distribution networks. Generated power is used to service load demand. A single line diagram of a power network can be seen in Figure 2.1. The diagram shows how thermal generation units (left-hand side), such as coal and nuclear, in addition to renewable sources of generation, like wind and solar provide a power generation mix that is transmitted by a network for the consumers of generated energy: industry and households (right-hand side).

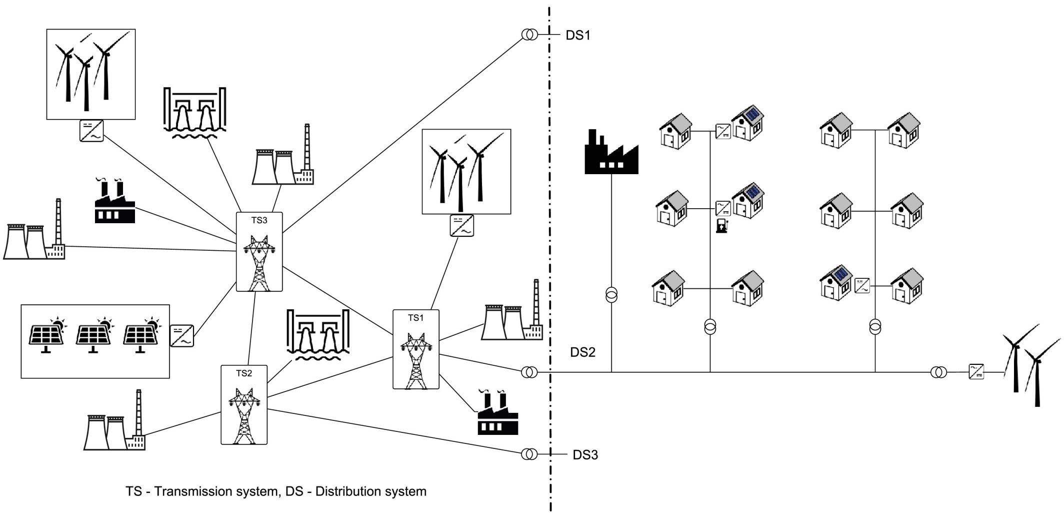

Figure 2.1: A single line diagram of a typical power system. The image shows points of generation from thermal and renewable sources, and the subsequent supply of generated energy to meet load demand through the transmission and distribution network [5].

One of the key elements to successful operation of interconnected power systems is ensuring total load demand is matched with total generation while taking into account power losses involved with generation, transmission, and distribution [6]. To understand why it is important to match generation with load demand consider the basic operation of a single thermal generator.

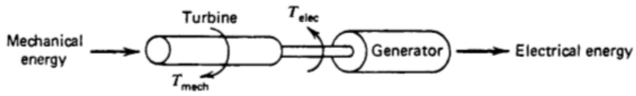

Figure 2.2: A thermal generation unit consisting of a prime mover (turbine), and a synchronous machine [6].

The essential elements of a thermal generator are a prime mover (such as a gas turbine) and a synchronous machine, as depicted in Figure 2.2. The prime mover provides mechanical torque, T_{mech}, which drives the synchronous machine producing electrical energy. In response, the synchronous machine creates an opposing torque that depends on the size of the load demand. This opposing torque is referred to as electrical torque and is denoted as T_{elec}. If α represents angular acceleration of the generator rotating mass, and I is its moment of inertia, then by Newton's second law:

$$T_{mech} - T_{elec} = I\alpha \tag{2.1}$$

When T_{mech} equals T_{elec} the system will be in a steady state of zero angular acceleration with a constant angular velocity, ω. Now, if $T_{mech} > T_{elec}$, then the system has an angular acceleration causing the angular velocity, ω, to increase. This results in a frequency increase in the system. Conversely, if $T_{mech} < T_{elec}$ then the angular velocity ω will decrease, resulting in a frequency decrease. It is important to note that, at any point in time, the total electrical load demand will fluctuate as businesses and households switch grid connected appliances or motors on and off. The result is that an uncontrolled system will have a continually changing frequency. Australia's electricity network is designed to operate at a frequency of 50Hz. In the majority of network scenarios, the Australian Energy Market Operator (AEMO) has a desired operating range for frequency which lies between 49.85 and 50.15Hz [7]. Similarly, the Power and Water Corporation (PWC) network technical code for the Northern Territory states that under normal operating conditions frequency should be maintained in the range of 49.80 to 50.20Hz [8]. Network operation outside of the specified range can cause damage to electrical equipment such as transformers or motors, which are designed to operate at specific frequencies [9]. Network designers engineer protection schemes so that sustained frequency excursions outside of the allowed range will cause equipment to trip from the network [10].

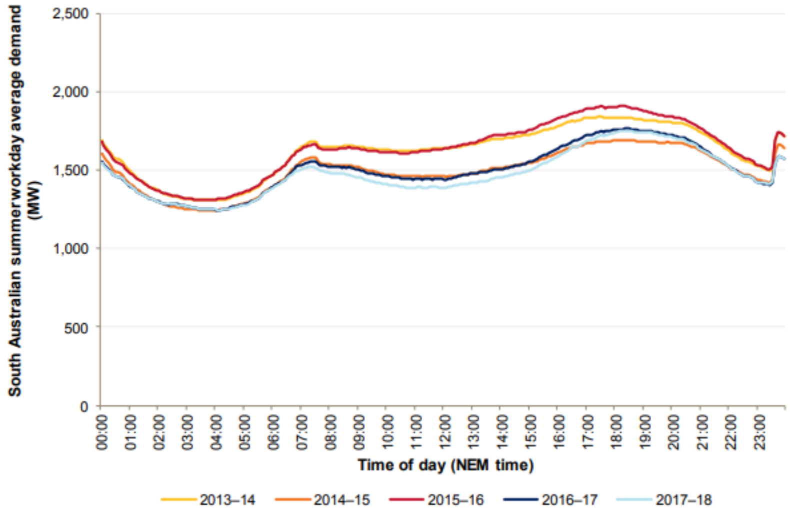

Figure 2.3: Weekday energy demand profile in South Australia during summer [11].

Protection schemes tripping equipment from the network is undesirable as this can leave households and industry without power, resulting in economic loss. Further, if disconnections are uncontrolled the system stability is reduced [10]. System controllers, such as the AEMO and PWC, are interested in being able to control the system to follow changes in load demand so that system frequency is maintained in the allowable range. Additionally, they are interested in control mechanisms to restore frequency excursions as a result of unexpected disturbances. System controllers can use historical data, like that shown in Figure 2.3, to forecast daily demand profiles with some reliability. This type of forecasting does not help when trying to predict the occurrence of random disturbances; however, it does provide a starting point for estimating required generation needed to meet demand. It is important to note that forecasting is not perfect. Inevitably mismatches in supply and demand will occur causing small imbalances between T_{mech} and T_{elec}, resulting in a change to angular velocity ω and the network frequency [12]. To perfectly match supply and demand, system controllers use generators referred to as regulating units, placed under Automatic Generation Control (AGC) [13]. A regulating unit is a generator that has the capacity to increase or decrease mechanical torque T_{mech}, and AGC is the name used for a system providing control over the mechanical torque output of regulating generators. If the system controller has a sufficient number of regulating units under AGC it can perform two functions: load following, and restoring the system to stable operating conditions in the event of a disturbance [14]. Using a regulating unit under AGC control to load follow is referred to, by AEMO, as load following ancillary services [15].

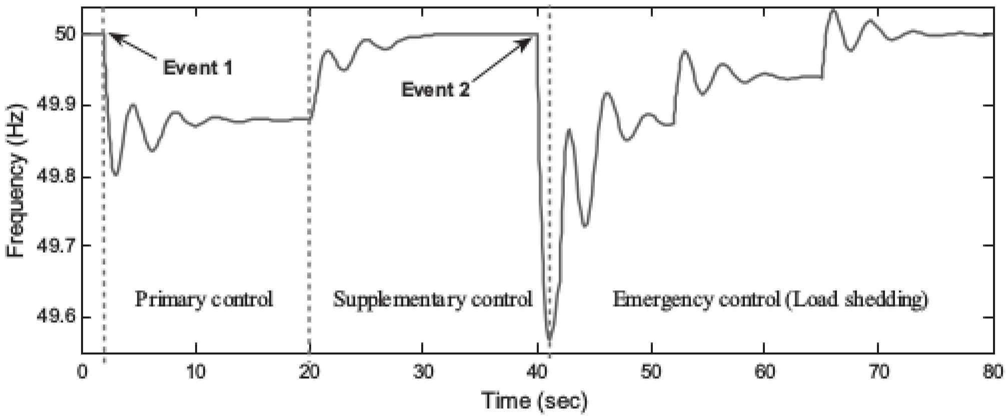

Figure 2.4: A minor frequency disturbance occurs at the 2 sec mark and primary control systems (governors) arrest the frequency drop. System frequency is adjusted to desired 50Hz operating level using AGC control of regulating units. This referred to as supplementary (or secondary) control in the literature. AEMO refers to this as load following ancilliary services. At the 40 sec mark the network experiences a major frequency disturbance which is arrested by emergency control measures such as under-frequency load shedding (UFLS). System restoration is aided using AGC control of regulating units, which AEMO refers to as spinning reserve ancillary services [16].

Load following control adjusts regulating units in order to match supply with a demand load profile, and maintain frequency in a normal operating range a shown in the first 40 seconds of Figure 2.4. Using a regulator under AGC control to restore the system after a major disturbance is referred to, by AEMO, as providing spinning reserve ancillary services. [15]. When used in either fashion it is important to note that the regulating unit is not responsible for arresting frequency excursions, rather, it is used to restore the system back to the allowable frequency operating range after the frequency excursion has been arrested. An example of a frequency excursion, arrest, and subsequent restoration for minor and major disturbances can be seen in Figure 2.4. AEMO and PWC do not require all generators on the network to act as regulating units since adequate frequency control can be achieved using a subset of the total available generators.

2.1.1 Frequency Control for a Single Area System

The power system model shown in Figure 2.1 depicts total generation coming from many generation assets — this is complex to model. Researchers often find it useful to divide generation assets into sub-groups referred to as control areas [13]. A control area is defined as a subset of generators that are in close proximity to each other and constitute a coherent group that speed up and slow down together, maintaining their relative power angles [13]. Therefore, the total network is comprised of many interconnected control areas. An example of this can be seen in Figure 2.5. Notice that for each area there is only one load and one generator.

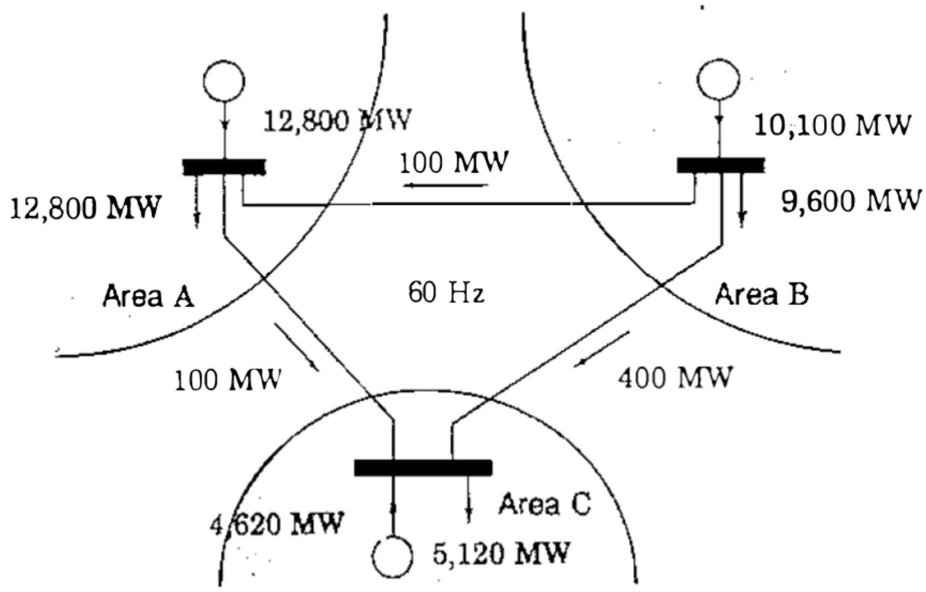

Figure 2.5: An example of three interconnected control areas in a 60Hz power system. The interconnections allow power to flow from one area to another, allowing generators to service loads from different areas. Each control area consists of several generators and loads, but are modelled with a single generator and single load for simplicity [14].

Typically, for each control area, researchers will aggregate many loads into a single load, and many generators into a single generator. This simplifies the model further [14]. The simplest power system to control is one that consists of a single control area. A single control area power system has no interconnections to any other control area. It is comprised of a consumer load demand, and a set of generators, some of which are acting as regulating units. As previously mentioned, for modelling simplicity, loads are aggregated to a single load, and generators can be aggregated to a single generator. This simple system is well understood. It is generally acknowledged that a speed droop governor feedback control regime will perform primary frequency control, and an AGC feedback loop is used to perform secondary frequency control when restoring a minor frequency excursion [6], [13], [14], [17]. A particularly well laid out approach to developing linear models for the turbine, generator, load, and governor was presented by Kundur [17]. The full model is shown in Figure 2.6. This particular model provides a model for a single regulating generator supplying a load. The governor block is a first-order linear model of the governor. The turbine block is a first-order linear model of the turbine. The final block is the generator-load, which is also a first-order linear model. The AGC feedback loop uses a proportional integral controller.

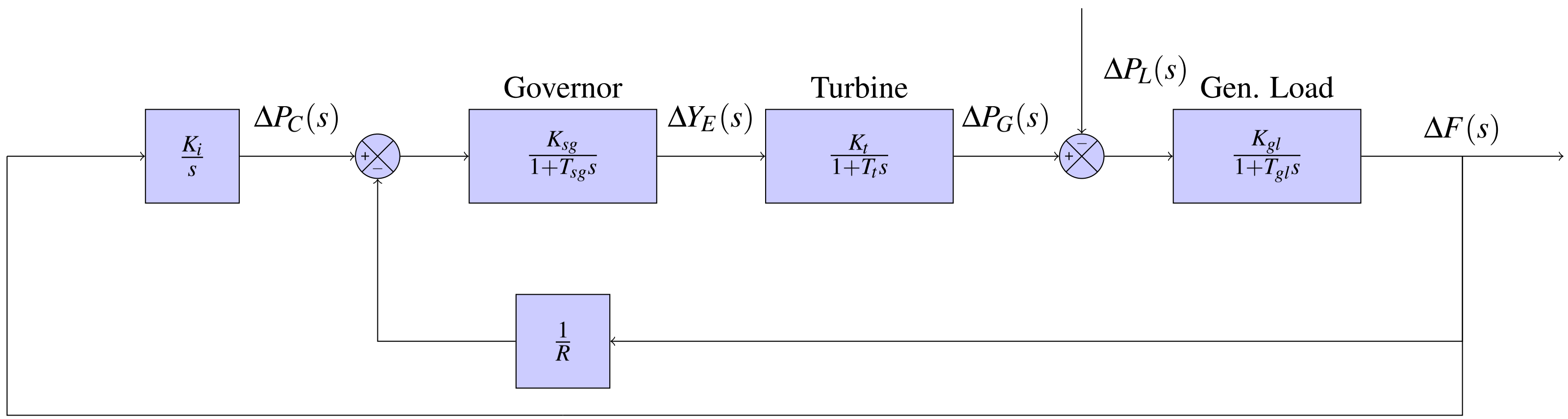

Figure 2.6: A classical feedback control approach for a single control area power system. The system is comprised of a first order models for both turbines, and generators. The governor controllers are also first order models. AGC is implemented using an integral control block in a feedback loop [17].

2.1.2 Frequency Control for Two Area System

The single area system presented in Section 2.1.1 is useful to help understand the role of governors and AGC in controlling power system frequency. In reality, power systems are comprised of many control areas connected by transmission lines (referred to in the literature as tie lines). Often it is the case that there is some net power transfer over the tie lines, enforceable by economic contract. Single area models do not provide for this additional complexity.

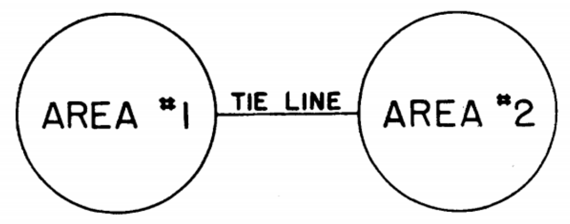

Figure 2.7: A two area power system comprised of generators and load connected via a tie line. Power flows from one area to the other depending on the power demands.

Distinct control areas are typically thought of as different participants in the generation market, or simply as different regions in which generation assets are based [13].

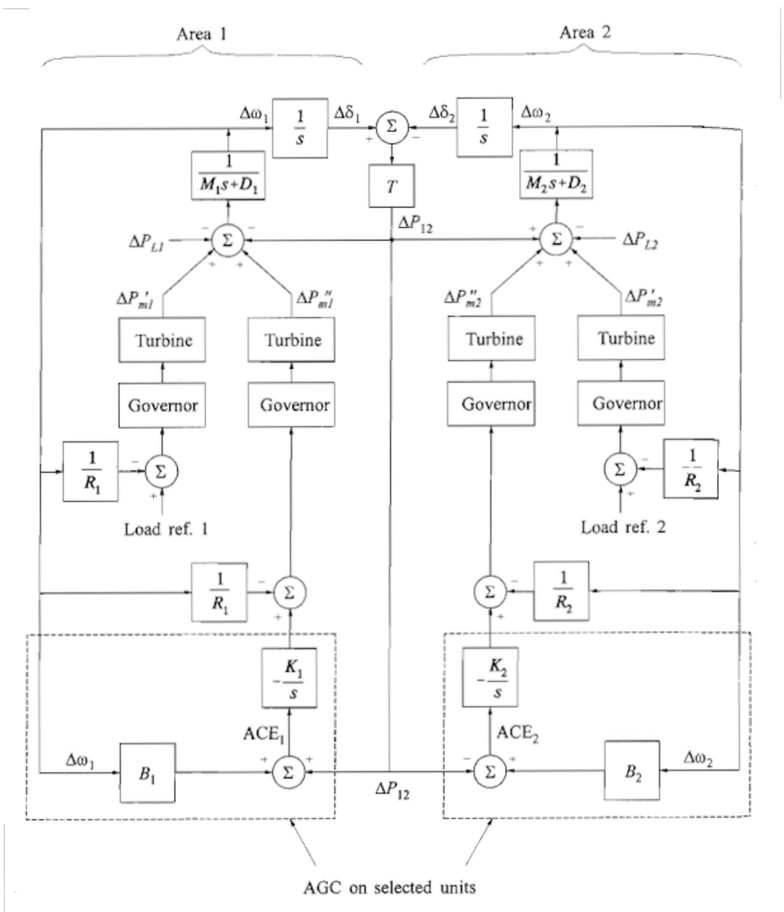

Figure 2.8: A classical feedback control approach to a two area power system [17].

The simplest model that includes tie lines is the two area power system, shown in Figure 2.7. The control objective with this system is to maintain the inter-area power transfer, whilst regulating the frequency of each area. An AGC integral feedback loop on regulating units, like that shown in Figure 2.6, would ensure that power system frequency is maintained, however, would not guarantee inter-area power transfer agreements are observed. Violation of power transfer contracts due to control issues does not allow for a stable market in which energy can be reliably traded. Fortunately, multi control area power systems are well understood. Linear models have been developed to simulate these systems, and classical control approaches have been successfully implemented to meet the new control objectives. In order to achieve this, a metric called Area Control Error (ACE) is used in the AGC feedback loop for each control area. ACE is a linear combination of the frequency deviations and the . The implementation of this control system is shown in Figure 2.8.

2.2 Reinforcement Learning

According to Sutton and Barto's seminal text, reinforcement learning (RL) is a branch of machine learning, based on trial-and-error, that is concerned with sequential decision making [18]. An RL agent exists in an environment where it can act and receive a reward. The environment is modelled as a set of probabilistic transitions between states, for a set of possible actions that can be selected by the agent. A state transition presents the agent with a reward signal that informs the agent whether an action taken was good or bad. This environmental architecture is referred to as a Markov Decision Process (MDP). It is the agent's objective to maximise the reward it will receive in the future. An agent can achieve this by learning an optimal policy which maps environment states to actions. Learning such a policy is key idea in RL, and the agent achieves this by experimentation.

2.2.1 Markov Decision Process

Bellman's pioneering work on the Markov Decision Process (MPD) provided the necessary architecture to develop RL algorithms [19]. His work considered an agent that exists in some environment comprised of many discrete states, $s \in S$, such that S denotes the state space. At any discrete point in time the agent can take action $a \in A$, where A denotes the action space. When the agent takes an action in a given state, the agent receives some reward, denoted with $r \in R$, where R is the set of rewards. Fundamental to Bellman's MDPs were the state transition dynamics which were defined by probabilities: if an agent is in a given state, s, and takes action, a, this will transition the agent to a new state, s', and yield reward, r, with some given probability. This set of probabilities are referred to as the state transition probabilities, and are denoted as follows:

$$P(S_{t+1} = s',\, R_{t+1} = r | S_t = s,\, A_t = a) = p(s', r | s, a) \tag{2.2}$$

The set of parameters outlined above, and expression 2.2, make up a framework referred to as an MDP.

2.2.2 Returns, Episodes, and Policy

In addition to developing the MDP framework, Bellman was also responsible for key developments in a field of research called dynamic programming (DP) [20]. Assuming that the agent has complete knowledge of the state transition probabilities of an environment, DP algorithms can be used to determine analytical solutions for the problem of how an agent should behave to maximise it's cumulative reward [20], [21]. This idea is distinct from RL but was critical in RL's development. The main difference is that DP provides the agent with complete knowledge of it's environment, whereas the RL agent has no knowledge of the environment dynamics and must learn them as well as how to maximise it's cumulative reward [18]. Many researchers made links between DP and RL [22]–[24], but it wasn't until 1989 that Watkins presented the first formal treatment of RL in an MDP framework, paving the way for modification of DP algorithms for use with RL problems [25]. Three of the central ideas used in DP algorithms are episodes, returns, and policies [18].

The duration of time that an agent will spend taking actions and transitioning states before encountering a terminal state is defined as an episode. It is the agent's goal to take actions such that it maximises the sum of all the rewards as it concludes an episode. The cumulative sum of rewards is called the return. Consider an agent taking an action at each discrete time step, t, and receiving reward, r_t, after each action. If there are N discrete time steps before the agent reaches a terminal state, Bellman defines the return as:

$$G_t = \sum_{k=0}^{N-1} r_{t+k} \tag{2.3}$$

Rewards received in the future are often perceived as less valuable than rewards received in the present. To account for this Bellman used a discount factor applied to each reward in the sequence. Letting $\gamma \in [0, 1]$ then 2.3 becomes:

$$G_t = \sum_{min}^{max} \gamma^k r_{t+k} \tag{2.4}$$

Finally, in order for the agent to take actions it must have a belief of what action it should take, given it's current state. This belief is called a policy and denoted as π [18]. Sutton and Barto define a policy as the mapping of states to actions i.e. a rule that determines what actions the agent should take for a given state. A policy can be deterministic, and depend only on the state, $\pi(s)$, or stochastic, $\pi(a|s)$, such that it defines a probability distribution over the actions, for a given state. An optimal policy, denoted $\pi*$, is a policy which will maximise the return an agent receives over an episode.

2.2.3 Value Function and the Bellman Equations

The basic principal of dynamic programming is to assign a value to each state that informs an agent how useful a state is to achieving a high cumulative reward. Watkins refers to the creation of systems to assign values to states as the credit assignment problem **Watkin1989**. Bellman's approach to solving credit assignment was to develop mathematical functions to assign values to states [20]. Bellman's *value function*, $V_\pi(s)$, is defined as the expected sum of the discounted return, G_t, that the agent will receive while following policy π from a particular state s. Mathematically, this is expressed as:

$$V_\pi(s) = \mathbb{E}_\pi\big(G_t|s_t - s\big) = \mathbb{E}_\pi\left(\sum_{k=0}^{\inf}\gamma^k r_{t+k}|s_t = s\right) \tag{2.5}$$

A slight variation of equation 2.5 is the *state-action value function*, $Q_\pi(s,a)$, which is defined as the expected sum of the discounted return, G_t, that the agent will receive if it takes action a in state s, and then follows policy π thereafter. Mathematically, this is expressed as:

$$Q_\pi(s,a) = \mathbb{E}_\pi\big(G_t|s_t - s, a_t = a\big) = \mathbb{E}_\pi\left(\sum_{k=0}^{\inf}\gamma^k r_{t+k}|s_t = s, a_t = a\right) \tag{2.6}$$

Bellman used the value functions presented in 2.5 and 2.6 to formulate recursive expressions which could then be used to solve the DP problem [19]. These are known as the *Bellman equations*. It is clear the agent would prefer policy π over some other policy π' provided the expected return from using policy π is greater than the expected return from using policy π' for all $s \in S$. Since the value function is defined by the expected return, Bellman expressed this idea in value function terms i.e. if policy π is preferred to π' then $V_\pi(s) \geq V_{\pi'}(s)$ for all $s \in S$. Thus, the optimal value function, $V*(s)$, can be defined as,

$$V*(s) = \max_\pi V_{pi}(s),\ \forall s \in S. \tag{2.7}$$

Similarly, the optimal state-action value function, $Q*(s,a)$, can be defined as,

$$Q*(s,a) = \max_\pi Q_\pi(s,a),\ \forall s \in S,\ a \in A. \tag{2.8}$$

Letting $A(s)$ be the set of actions available in state s, if the agent is operating under the optimal policy $\pi*$ then it is true that

$$V*(s) = \max_{a \in A(s)} Q_{\pi*}(s,a). \tag{2.9}$$

Using equation 2.6, equation 2.9 can be rewritten as

$$V*(s) = max_a \sum_{s',r} P(S_{t+1} = s', R_{t+1} = r | S_t = s, A_t = a)[r + \gamma V*(s')]. \tag{2.10}$$

Equation 2.10 is referred to as the Bellman optimality equation for $V*(s)$. The Bellman optimality equation for $Q*(s,a)$ is

Watkins is credited with the most influential integration of RL with MDPs, and DP. His work on an RL algorithm called Q-learning highlighted the importance of another type of value function called the action-value function. The action-value function is defined as the expected sum of rewards that the agent will receive while taking action a in state s and, thereafter, following policy π.

BELLMAN EQUATION FOR ACTION-VALUE FUNCTION

2.2.4 Value Based

Bellman invented these algorithms - where were they formaliesd though? Watkins adapted these DP algorithms for the use with RL

Monte Carlo Methods

Temporal Difference Methods

2.2.5 Policy Search Methods

2.2.6 Actor Critic Methods

2.3 Deep Neural Networks

introduction

2.3.1 A Perceptron

The inputs from the previous layer, and the non-linear activation of a node form a computational element called a neuron (also known as a perceptron) - these can be loosely thought of as decision making elements. An example of a neuron can seen in Figure 5.

artificial neuron rosenblatt

2.3.2 Activation Functions

sigmoid - who developed these relu - who developed these tanh - who developed these

2.3.3 Feedforward Networks

A typical fully connected feed-forward ANN consists of an input layer, one or more hidden layers, and an output layer, as shown in Figure 4. Hidden layers are made up of multiple nodes. The nodes themselves contain a non-linear activation function, such as a sigmoid or ReLU, and receive weighted input from the previous layers in the model.

2.3.4 Training the Network

Changing the weights in a neuron changes the neurons's contribution to the model, which in turn affects the overall model output. Weight changes occur during model training, which uses large volumes of labelled data to adjust the weights. Hidden layers are important because they allow highly non-linear models to be constructed, providing an approach for estimating complex phenomena which may be difficult to model with classical approaches, or computationally intractable. Generally, the more hidden layers, the more non-linear the model. Network architectures with multiple hidden layers have become so wide spread that the term Deep Neural Network (DNN) was coined to describe feed-forward ANNs which use two or more hidden layers. It must be noted that whilst increased non-linearity may allow us to model more complex phenomenon, making the ANN deeper does not guarantee increased model performance. This is mainly due to the fact that deeper models may over-fit the data during training, resulting in a failure to generalise on test and validation data sets.

werbos backpropagation

2.3.5 Regularisation

dropout

2.4 Deep Reinforcement Learning

2.4.1 Deep Q-Learning

2.4.2 Deep Deterministic Policy Gradient

Chapter 3

Literature Review

Chapter 4

Approach

4.1 Required Data Sources and Data Management

Training a DRL agent to change regulating generator set points in order to maintain system frequency and tie line requirements while load following will require realistic demand profiles. Similarly, performing system restoration after a disturbance will require realistic disturbance scenarios. Ideally this data would come from a major utility provider, such as PWC, in the form of a time series dataset with a large number of features, and high sample rate. To this end, data acquisition will be one of the principle objectives in the early stages of research. Should the acquisition of data from PWC or TGEN be viable, a data management plan will need to be developed which addresses concerns around the sensitivity and security of the data. The data management plan will outline where data the will be stored, and how the data will be treated (or disposed of) once the research is concluded.

In the event data cannot be acquired from a utility provider, a simulated data set may need to be used. This would be achieved by understanding key statistical parameters of a typical load demand profile, and using these to create a process which emulates the load demand signal. This could also be done for other system variables; however, care would need to be taken ensuring correlations are preserved between multiple variables in the simulated time series dataset.

4.2 Theoretical Approach

In order to establish the most effective way to approach this research problem, a clear understanding of the benefits and limitations of existing AGC approaches is needed. Determining justifications for practical AGC design choices will help to uncover important performance aspects the research should focus on. Equally important is exploring alternative approaches to AGC that researchers have investigated historically. This should have a particular focus on the use of Neural Networks and DRL agents for AGC. A literature review will be the main avenue for achieving this.

As discussed in §4.1, securing load demand profile datasets from a major utility provider, or developing simulated load profile datasets based on local load profile characteristics is an important aspect of this research and is a priority. Similarly, investigating suitable software packages to develop the power system simulation model, and investigating suitable programming languages to implement a DRL agent are necessary. Exploring the field literature and finding published examples will be a key stepping stone. It will be important to understand how other researchers integrated the DRL agent with the simulation environment.

A simulated model of the two area power system will be developed. The decision to use a linear or non-linear model will be informed by the literature review. It may be interesting to explore DRL agent performance on both linear and non-linear models since one of their advantages is that they have a demonstrated capacity for controlling non-linear systems. Classical engineering system modelling techniques will be employed for power system model development [26]. An area of interest is how sensitive a DRL agent control regime is to changes in key plant parameters — for a given set of parameter changes both DRL agent and classical control architecture performance could be compared to see which controller is more brittle.

A feedback loop controller will be developed for the two area power system using models presented in the literature. A DRL agent will be developed using an architecture that takes continuous input signals and provides continuous output signals. There are a number of established DRL models presented in texts like Sutton and Barto that will be explored to determine the most ideal approach [18]. Time permitting, a DRL model that uses discrete input and output signals will be developed. Discrete models offer lower performance due to errors introduced in the discretisation process, but can be computationally less expensive than continuous models. DRL models will be trained using data previously acquired either from a utility provider, or from simulation. Metrics will be selected to measure the performance of both controllers. Choice of metrics will be informed by earlier research (mentioned above). Control models differences in performance will be compared — this will be one of the major focuses of the research.

The full list of tasks for the research design are as follows:

1. Enquire with power utility provider to secure data.

2. Investigate ways to simulate data.

3. Develop data management plan.

4. Literature review centred on three central themes:

    (a) AGC

   (b) DRL
   (c) Applications of DRL to AGC.

5. Investigate suitable software package to conduct simulation.
6. Investigate suitable programming language to implement DRL agent and integrate with simulation.
7. Develop and test simulation of two area power system.
8. Develop feedback loop controller for two area power system.
9. Test classical controller.
10. Develop DRL model.
11. Train and test DRL model.
12. Execute control trials on both models for an unseen sequences of load demand data.
13. Compare controller performance on AGC task.

It is anticipated that there may be some issues in carrying out the aforementioned research design. The biggest risk would be the inability to successfully build the control models for both the classical engineering controller, and the DRL controller. For the classical engineering controller, the issue would be the inability to find the appropriate parameter settings to deliver stable control. With the DRL controller, the problem is selection of an appropriately sized neural network, and training hyper-parameters.

Chapter 5

Experiments

Chapter 6

Analysis and Discussion of Results

Chapter 7

Conclusion

Chapter 8

Future Work

Bibliography

[1] (2020). Electricity generation, Department of industry science energy and resources, [Online]. Available: https://www.energy.gov.au/data/electricity-generation.

[2] "Special report on renewable energy sources and climate change mitigation," Intergovernmental Panel on Climate Change, Tech. Rep., 2012. [Online]. Available: https://www.ipcc.ch/site/assets/uploads/2018/03/SRREN_FD_SPM_final-1.pdf.

[3] "Independent investigation of alice springs system black incident on 13 october 2019," Utilities commission of the northen territory, Tech. Rep., 2019. [Online]. Available: https://utilicom.nt.gov.au/__data/assets/pdf_file/0011/767783/Independent-Investigation-of-Alice-Springs-System-Black-Incident-on-13-October-2019-Report.pdf.

[4] D. Wilkey, "Alice springs system black 13 october 2019," Entura, Tech. Rep., 2019. [Online]. Available: https://utilicom.nt.gov.au/__data/assets/pdf_file/0012/767784/Advice-Entura-Alice-Springs-System-Black-13-October-2019-Report.pdf.

[5] M. Glavic, "Deep reinforcement learning for electric power system control and related problems: A short review and perspectives," *Annual reviews in control*, 2019.

[6] A. J. Wood, B. F. Wollenberg, and G. B. Shelbe, *Power generation, operation, and control*, 3rd Edition. Wiley, 2013.

[7] "Power system frequency and time deviation monitoring report - reference guide," Australian Energy Market Operator, Tech. Rep., Jul. 2012. [Online]. Available: https://aemo.com.au/-/media/files/electricity/nem/security_and_reliability/ancillary_services/frequency-and-time-error-reports/frequency_report_reference_guide_v2_0.pdf.

[8] *Network technical code and network planning criteria*, Power and Water Corporation, 2013. [Online]. Available: https://www.powerwater.com.au/__data/assets/pdf_file/0022/5962/Power-and-Water-Corporation-Network-Technical-Code-and-Network-Planning-Criteria.pdf.

[9] P. C. Sen, *Principles of Electric Machines and Power Electronics*, 3rd Edition. Wiley, 2014.

[10] “Power system frequency risk review - final report,” Australian Energy Market Operator, Tech. Rep., Apr. 2018. [Online]. Available: https://aemo.com.au/-/media/files/electricity/nem/planning_and_forecasting/psfrr/2018_power_system_frequency_risk_review-final_report.pdf?la=en&hash=1684259023A1FA274D7F3B8CE855D0BA.

[11] “South australian electricity report,” Australian Energy Market Operator, Tech. Rep., Nov. 2019. [Online]. Available: https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/SA_Advisory/2019/2019-South-Australian-Electricity-Report.pdf.

[12] J. D. Glover, S. S. Mulukutla, and T. J. Overbye, *Power system analysis and design*, 5th Edition. Cengage Learning, 2012.

[13] D. P. Kothari and I. J. Nagrath, *Modern Power System Analysis*, 4th Edition. McGraw Hill India, 2011.

[14] J. J. Grainger and W. D. Stevenson, *Power System Analysis*. McGraw Hill, 1994.

[15] (2020). Ancilliary services, Australian Energy Market Operator, [Online]. Available: https://aemo.com.au/en/energy-systems/electricity/wholesale-electricity-market-wem/system-operations/ancillary-services.

[16] H. Bevrani and T. Hiyama, *Intelligent Automatic Generation Control*. CRC Press, 2011.

[17] P. Kundur, *Power System Stability and Control*. McGraw-Hill Inc., 1994.

[18] R. S. Sutton and A. G. Barto, *Reinforcement Learning*, M. Press, Ed. 2018.

[19] R. Bellman, “A markovian decision process,” *Journal of Mathematics and Mechanics*, vol. 6, no. 5, pp. 679–684, 1957.

[20] ——, “The theory of dynamic programming,” *Bulletin of the American Mathematical Society*, vol. 60, no. 6, pp. 503–515, 1954. [Online]. Available: https://projecteuclid.org/download/pdf_1/euclid.bams/1183519147.

[21] R. A. Howard, *Dynamic Programming and Markov Processes*. Cambridge, MA: MIT Press, 1960.

[22] R. Bellman and S. E. Dreyfus, “Functional approximations and dynamic programming,” *Mathematical Tables and Other Aids to Computation*, 1959.

[23] I. H. Witten, "An adaptive optimal controller for discrete-time markov environments," *Information and Control*, vol. 34, no. 4, pp. 286–295, 1977, ISSN: 0019-9958. DOI: `https://doi.org/10.1016/S0019-9958(77)90354-0`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0019995877903540`.

[24] P. J. Werbos, "Building and understanding adaptive systems: A statistical/numerical approach to factory automation and brain research," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 17, no. 1, pp. 7–20, 1987.

[25] C. Watkins, "Learning from delayed rewards," PhD thesis, King's College, Cambridge, UK, May 1989. [Online]. Available: `http://www.cs.rhul.ac.uk/~chrisw/new_thesis.pdf`.

[26] K. Ogata, *Modern Control Engineering*, 5th ed. Pearson, 2010.

Appendix A

Your first appendix

A.1 The title of the first section

The appendices work exactly the same way as chapters, they are numbered with letters rather than numbers though [26].

Appendix B

Your second appendix

B.1 The title of the first section