University of Strathclyde

Department of Electronic and Electrical Engineering

Learning to Trade Power

by

Richard W. Lincoln

A thesis presented in fulfilment of the
requirements for the degree of

Doctor of Philosophy

2010

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Acknowledgements

This research was funded by the UK Engineering and Physical Sciences Research Council through the Supergen Highly Distributed Power Systems project under grant GR/T28836/01.

I take this opportunity to thank my supervisors, Prof. Graeme Burt and Dr Stuart Galloway, for their guidance and scholarship. Many thanks also to my parents for their support and help in editing this thesis.

This research made extensive use of software projects by researchers from other institutions, made available as open source. Optimal power flow solvers were translated from Matpower, which is developed and maintained under the direction of Ray Zimmerman at Cornell University. Reinforcement learning algorithms and artificial neural networks were imported from PyBrain, which is developed by researchers from the Dalle Molle Institute for Artificial Intelligence (IDSIA) and the Technical University of Munich. The Roth-Erev learning method was translated from the Java Reinforcement Learning Module (JReLM), developed by Charles Gieseler from Iowa State University.

Abstract

Reinforcement learning methods that use connectionist systems for value function approximation offer few convergence guarantees, even in simple systems. Table-based value function reinforcement learning methods have been used previously for the simulation of electricity markets, but they operate only in discrete action and sensor domains. If learning algorithms are to deliver on their potential for application in operational settings then it will be necessary for them to operate in continuous domains. The principle contribution of this thesis is the demonstration of policy-gradient reinforcement learning algorithms being applied to continuous representations of electricity trading problems, showing that superior use of sensor data results in improved overall performance when compared with previously applied value-function methods. From this it follows that learning methods which search directly in the policy space will be better suited to decision support applications and automated electric power trade.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis presents a comparison of algorithms which learn to trade power. This first chapter explains the motivation for the research presented below, defines the problem that has been tackled and states the principle research contributions that have been made. A reading guide and outline of the remaining chapters is provided at the end.

1.1 Motivation for Electricity Market Research

The average total demand for electricity in the United Kingdom (UK) is approximately 45GW and the cost of buying 1MW for one hour is around £40 (Department of Energy and Climate Change, 2009). This equates to yearly transaction values of £16 billion. The value of electricity to society becomes especially apparent when supply fails. The New York black-out in August 2003 involved a loss of 61.8GW of power supply to approximately 50 million consumers. The majority of supplies were restored within two days, but the event is estimated to have cost more than $6 billion and to have contributed to 11 deaths (Minkel, 2008; ICF Consulting, 2003).

Quality of life for a person has been shown to be directly proportional to that person's level of electricity usage (Alam, Bala, Huo, & Matin, 1991). The world population is currently 6.7 billion and forecast to pass 9 billion by the year 2050 (United Nations, 2003). Electricity production currently demands over 1/3 of the annual primary energy extracted[ref]. As people endevour to improve their quality of life, finite primary energy fuel resources are becoming increasingly scarce and markets are a proven economic device for efficient allocation of scarce resources.

Commercialisation of electricity supply industries is a relatively new practice, having begun in the early 1990s. The inability to store electricity, once generated, in a commercially viable quantity prevents trade as a conventional commodity. Trading mechanisms must allow shortfalls in electric energy to be purchased at short notice from quickly dispatchable generators. Numerous mechanisms for this have been implemented in countries and states around the world and how best to structure them remains an important and unanswered question.

1.2 Problem Statement

An indivdual interacting with an electricity market environment, be they representing a generating company, load serving entitiy, trader etc., is presented with state and action spaces of many dimensions and which are mostly continuous in nature. Moreover, certain state information, such as demand forecasts, exhibit a degree of uncertainty and other data, such as a competitor's bid, is hidden.

Traditional value-function based reinforcement learning methods (defined in Appendix B.2, below) offer few convergence guarantees in partially observable Markov decsison processes [ref]. Without the use of function approximation techniques, these methods are restricted by Bellman's Curse of Dimentionality (Bellman, 1961) and can not be applied to complex problems with high-dimensional state and actions space. With value function approximation, feedback between policy updates and value function changes can result in oscillations or divergence in these methods (Peters & Schaal, 2008).

Policy gradient reinforcement learning methods (defined in Appendix B.3, below) do not suffer from many of the problems that mar value-function based methods in high-dimensional domains. They offer strong convergence guarantees, do not require that all states be continuously visited and work with state and action spaces that are continuous, discrete or mixed. Policy performance may be degraded by uncertainty in state data, but the learning methods need not be altered. Importantly, they have been sucessfully applied in many operational settings (Sutton, McAllester, Singh, & Mansour, 2000; Peters & Schaal, 2006; Moody & Saffell, 2001; Peshkin & Savova, 2002).

It is proposed that agents which learn using policy gradient methods may outperform those using value function based based methods in a simulated competitive electricity trading environment. It is proposed that policy gradient methods may learn faster, achieve greater profitability and better exploit constraints in the power system to their advantage.

1.3 Research Contributions

The research presented in this thesis pertains to the academic fields of Electric Power Engineering, Artificial Intelligence and Economics. The principle contributions made by this thesis in these fields are:

- The proof that policy gradient reinforcement learning algorithms outperform value-function algorithms when applied to the power trade problem,
- A novel coupling of power system models and optimal power flow algorithm results with agents capable of handling discrete and continuous sensor and action spaces,
- Implementations of Roth-Erev reinforcement learning algorithms and continuous versions of Q-learning and Q(λ) for the open source PyBrain library,
- Open source implementations of power flow and optimal power flow algorithms in the Python programming language.

1.4 Reader's Guide

This thesis applies standard and advanced reinforcement learning methods in the domain of electric power trade. The reader will require a certain degree of prior knowledge and may need to read Chapter 2 and much of the referenced material, to fully understand the methodology used. This thesis is written for several kinds of readers. A student who has taken an energy economics class or two may appreciate it as an introduction to electricity markets and their simulation. Research students embarking upon postgraduate study of electricity markets may find the ideas for further work in Section 8.2 of particular interest. Researchers experienced in adpative control and machine learning, looking for new application domains for their methods, may find the electricity market model definition in Chapter 4 to be of value.

1.5 Thesis Outline

The presentation is organised into nine chapters. Chapter 2 provides a introduction to electric power supply and wholesale electricity markets in the UK. The research in this thesis is described in the context of related work from the fields of Power Enginneering, Machine Learning and Computer Science in Chapter 3. Chapter 4 defines the electricity market model and a multi-agent system used to coordinate simulated electricity trade. Reinforcement learning methods are compared in a series of increasingly daring experiments in Chapter 5. Finally, a summary of conclusions drawn from this research is given in 8 along with several ideas for building upon the tools developed.

Chapter 2

Background

Competitive power trade is a relatively new practice and is one that has a number of unique characteristics among commodities. This chapter provides background information on electric power supply and the associated wholesale electricity markets. It also explains how such markets are simulated and, in particular, the agent-based approach to this.

2.1 Electric Power Systems

Generation and bulk movement of electricity in the UK takes place in a three-phase alternating current (AC) power system. Each of the phases is high voltage, sinusoidal electrical waveform, 120 degrees offset in time and oscillating at a constant frequency of 50Hz. Synchronous generators (or alternators), typically rotating at 3600rpm or 1800rpm, generate apparent power S at a line voltage V_l typically between 11kV and 25kV. One of the prinicpal reasons that alternating current, and not direct current (DC), systems are common in electricity supply is that they allow power to be transformed between voltages with very high efficiency. The apparent power conducted by a line l is the product of the line current I_l and the line voltage

$$S = \sqrt{3} V_l I_l. \tag{2.1}$$

Thus increasing the voltage has an inverse effect on the current. Ohmic heating losses are proportional to the square of line current

$$P_r = 3 I_l^2 R \tag{2.2}$$

where R is the resistance of the line. Hence reducing the line current causes a dramatic reduction in heating losses. One consequence of higher voltages is the larger extent and integrity of conductor insulation required between one another, neutral and earth. This results in necessity for large transmission towers and in high cable costs when undergrounding systems.

The UK transmission system operates at 400kV and 275kV (plus 132kV in Scotland), but systems with voltages upto and beyond 1000kV are used in certain larger countries[1]. The ability to transform power between voltages and transmit large volumes over long distances allows for generation to take place at high capacity stations (located away from large load centres and closer to fuel sources) which offer economies of scale and lower operating costs. It allows electricity to be transmitted across country borders and from renewable energy plant such as hydro power stations located in remote areas.

For delivery to most consumers, electric energy is transferred, at a substation, from the transmission system to the grid supply point of a distribution system. Distribution networks are also three-phase AC systems, but typically operate at lower voltages and differ in their general structure/topology from transmission networks. Transmission networks are typically highly interconnected, providing several paths for power flow. Whereas distribution networks, in rural areas, typically consist of long radial feeders (usually overhead lines) or, in urban areas, consist of many ring circuits. Three-phase transformers, that step the voltage down to levels more convenient for general use (typically from 11kV or 33kV to 400V), are spaced along the branches/rings. All three-phases at 400V may be provided for industrial and commercial loads or individual phases at 230V supply typical domestic and other commercial loads. Splitting of phases is usually planned so that each is loaded equally. As explained in Appendix C this produces a balanced symetrical system that may be analysed as a single phase circuit. Figure [] illustrates the basic sturcture of a typical national electric power system.

2.2 Electricity Markets

The UK was the first large country to privatise its electricity supply industry. The market structures that have since been adopted implement some of the main princi-

[1]For transmission over very long distances or undersea, high voltage DC systems have become econmically viable in recent years.

ples behind electricity markets.

The England and Wales Electricity Pool was created in 1990 to break apart the monolithic Central Electricity Generating Board (CEGB) and to gradually introduce competition in generation and retail supply. Early adoption of electrcity markets by the UK has lead to the country hosting many of the main European power and gas exchanges and the UK boasts a high degree of consumer switching, an important factor in any competitive marketplace. The Pool has since been replaced by trading arrangements in which market outcomes are not centrally determined, but arise largely from bilateral agreements between producers and suppliers.

2.2.1 The England and Wales Electricity Pool

The Electric Lighting Act 1882 began the development of the UK's electricity supply industry by allowing persons, companies and local authorities to set up supply systems, principally at the time for the purposes of street lighting and trams. Under The Electricity Supply Act 1926 the Central Electricity Board started operating the first grid of regional networks interconnected and synchronised at 132kV, 50Hz in 1933. This began operation as a national system five years later in 1938 and was nationalised under The Electricity Act 1947 with the merger of over 600 electricity companies and the creation of the British Electricity Authority. This was dissolved and replaced with the CEGB and the Electricity Council under The Electricity Act 1957. The CEGB was responsible for planning the network and generating sufficient electricity until the start of privatisation in 1990.

The industry was privatised under Prime Minister Margaret Thatcher and The England and Wales Electricity Pool was created in March 1990. Control of the transmission system was transferred from the CEGB to The National Grid Company, which was originally owned by twelve regional electricity companies and is now publically listed. The Pool was a multilateral contractual arrangement between generators and suppliers and did not itself buy or sell electricity. Competition in generation was introduced gradually, by entitling customers with consumption greater than or equal to 1MW (approximately 45% of the non-domestic market (Department of Energy and Climate Change, 2009)) to purchase electricity form any listed supplier. This limit was lowered in April 1994 to included customers with peak loads of 100kW or more. Finally, between September 1998 and March 1999 the market was opened to all customers.

Scheduling of generation was on a merit order basis (cheapest first) at a day ahead stage and set a wholesale electricity price for each half-hour period of the schedule day. Forecasts of total demand in MW, based on historic data and adjusted for factors such as the weather, for each settlement period were used by generating companies and organisations with interconnects to the England and Wales grid to formulate bids that had to be submitted to the grid operator by 10AM on the day before the schedule day.

Figure X illustrates the four of the five price parameters that bids consisted of. A start-up price would also be included, representing the cost of turning on the generator from cold. The no-load price c_{noload} represents the cost in pounds of keeping the generator running regardless of output. Three incremental prices c_1, c_2 and c_3 specify the cost per MWh of generation between set-points p_1, p_2 and p_3.

A settlement computer program was used to calculate an unconstrained schedule (with no account being taken for the physical limitations of the transmission system), meeting the forecast demand and requirements for reserve while minimising cost. Cheapest bids up to the marginal point would get accepted first and the bid price from the marginal generator would generally determine the system marginal price for each settlement period. The system marginal price would determine the prices paid by consumers and paid to generators, which would be adjusted such that that the costs of transmission are covered by the market and that the availability of capacity is encouraged at certain times.

Variations in demand and changes in plant availability would be adjusted for by the grid operator, producing a constrained schedule. Generators having submitted bids would be instructed to increase or reduce production appropriately. Alternatively, the grid operator could instruct large customers with contracts to curtail their demand to do so or instruct generators contracted to provide ancillary services to adjust production.

2.2.2 British Electricity Transmission and Trading Arrangements

Concerns over exploitation of market power in The England and Wales Electricity Pool and its effectiveness in reducing consumer electricity prices prompted the introduction of New Electricity Trading Arrangements (NETA) in March 2001 (D. Bunn & Martoccia, 2005). The aim was to improve efficiency and provide greater choice to participants. Control of the Scottish transmission system was handed to England when the industry was integrated into the nationwide British Electricity Transmission and Trading Arrangments (BETTA) in April 2005 under The Energy Act 2004. While The Pool operated a single daily auction and dispatched plant centrally, under the new arrangements participants became self-dispatching and market positions became determined through continuous bilateral trading between generators, suppliers, traders and consumers.

The majority of power is traded under the BETTA through long-term contracts that are customised to the requirements of each party (Kirschen & Strbac, 2004). These suit participants responsible for large power plants or those purchasing large volumes of power for many customers. Considerable amounts of time and effort are required to form these long-term contracts and this results in a high associated transaction cost. However, they reduce risk for large players and a degree of flexibility can be provided through option contracts.

Power is also traded directly between participants through over-the-counter contracts that are usually of a standardised form. Such contracts typically concern smaller volumes of power and have much lower associated transaction costs. Often they are used by participants to refine their market position ahead of delivery time.

Trading facilities, such as power exchanges, provide a means for participants to fine-tune their positions further, through short-term transactions for relatively small quantities of energy. Modern exchanges are computerised and accept anonymous offers and bids submitted electronically. A submitted offer/bid will be paired with any outstanding bids/offers in the system with compatible price and quantity values. The details are then displayed for traders to observe and use to educate their decisions.

All bilateral trading must be completed before "gate-closure" which is a point in time, before delivery time, that gives the system operator an opportunity to balance supply and demand and mitigate potential breaches of system limits. In keeping with the UK's free market philosphy, a competitive spot market (Schweppe, Caramanis, Tabors, & Bohn, 1988) is used in the balancing process. A generator that is not fully loaded may offer a price at which it is willing to increase its output by a specified quantity, stating the rate at which it is capable of doing so. Certain loads may also offer demand reductions at a price which can typically be implemeted very quickly. Longer-term contracts for balancing services are also struck between the system operator and generators/suppliers in order to avoid the price volatility often associated with spot markets.

2.3 Electricity Market Simulation

Previous sections have shown the importance of electricity to modern societies and have explained how its supply is trusted to unadministered bilateral trading arrangements. Electricity supply involves technology, money, people, natural resources and the environment. These aspects are all changing and the discipline must be constantly researched to ensure that systems such as electricity markets are fit for purpose. The value of electricity to society means that it is not feasible to experiment with radical changes to trading arrangements on real systems. A practical alternative is to create an abstract mathematical model with a set of simplifying approximations and assumptions and find analytical solutions by simulating the model using computer programs.

Game theory is the branch of applied mathematics in which behaviour in strategic situations is captured. A common approach to doing this is to model the system and players as a mathematical optimisation problem. Appendix A defines the optimal power flow problem, which is a classic optimisation problem in the field of electric power Engineering. Electricity markets are commonly modelled using variations on the optimal power flow problem with player strategies integrated[ref]. However, this thesis concerns *agent-based* simulation, which is an alternative approach to the mathematics of games.

2.3.1 Agent-Based Simulation

Social systems, such as electricity markets, are inherently complex and involve interactions between different types of individuals and between individuals and collective entities, such as organisations or groups, the behaviour of which is itself the product of individual interactions. This complexity drives classical monolithic equilibrium models to their limits. Models are often highly stylised and limited to small numbers of players with strong constraining assumptions made about their behaviour.

Agent-based simulation involves modelling simultaneous operations and interactions between adaptive agents and assessing their effect on the system as a whole. Macro-level system properties arise from agent interactions, even those with simple behavioural rules, that could not be deduced by simply aggregating the agent's properties.

Following (Tesfatsion & Judd, 2006), the objectives of agent-based modelling research fall roughly into four strands: empirical, normative, heuristic and methodological. The *empirical* objectives are to understand how and why macro-level regularities have evolved from micro-level interactions when little or no top-down control is present. Reaserch with *normative* goals aims to relate agent-based models to an ideal standard or optimal design. The objective being to evaluate proposed designs for social policy, institutions or processes in their ability to produce socially desirable system performance. The *heuristic* strand aims to generate theories on the fundamental causal mechanisms in social systems that can be observed, even in simple systems, when there are alternative initial conditions. This thesis aims to provide *methodologial* advancement in the field. Improvements in the tools and methods available aid research with the former objectives.

2.4 Summary

Chapter 3

Related Work

This chapter describes the research presented in this thesis in the context of similar work, with particular emphasis on the simulation methodology used. For a similar review with greater detail on the simulation results and the conclusions that have been drawn from them, the interested readers is referred to (?, ?).

In the interests of repeatability, the software developed for this thesis has been released as open source under a project named Pylon. The end of this chapter describes the software project in the context of other open source electric power Engineering tools and explains the contribution that has been made.

3.1 Custom Learning Methods

Early agent-based electricity market simulations in the literature do not utilise traditional learning methods from Artificial Intelligence, but rely upon custom heuristic methods. They are typically formulated using the author's intuition and encapsulate basic trading rules, but disregard many of the key concepts from reinforcement learning theory.

3.1.1 Market Power

Under Professor Derek Bunn, researchers from the London Business School performed some the first and most reputable agent-based electricity market simulations in the literature. Their research was initially motivated by proposals in 1999 to transform the structure of The England and Wales Electricity Pool with the aim of combating generator market power, that was widely believed to be causing elevated market prices.

In (Bower & Bunn, 2001) a detailed model of electricity trading in England and Wales is used to compare day-ahead and bilateral contract markets under uniform price and discriminatory settlement. Twenty generating companies operating in the Pool during 1998 are modelled as agents endowed with a portfolio of generating plants. Plant capacities, costs and expected availabilities are synthesised from public and private data sources and the author's own estimates. In simulations of the day-ahead market, each agent submits a single price for the following simulated trading day, for each item of plant in its portfolio. Whereas, under the bilateral contract model, 24 bids are submitted for each generator, cooresponding to each hour of the following simulated day. Revenues are calculated at the end of each trading day and are determined either by the bid price of the marginal unit or the generator's own bid price. Each generating plant is characterised in part by an estimated target utilisation rate that represents its desire for forward contract cover. The agents learn to achieve this utilisation rate and then improve profitability.

Algorithm X defines the rules followed by each agent. If the utilisation rate is not achieved, a random percentage from a uniform distribution with a range of $\pm 10\%$ and 0% mean is substracted from the bid price of all generators in the agent's portfolio. Agents with more than one generator transfer successful bidding strategies between items of plant by setting the bid price for a generator to the level of the next highest submitted bid price if the generator sold at a price lower than that of other generators in the same portfolio. If an agent's total profit does not increase, a random percentage from the same distribution as above is added or subtracted from the bid price from the previous day for each of its generators. A cap on bid prices is imposed at £1000 in each period. Demand follows a 24-hour profile based on the 1997/98 peak winter load pattern. The response of the load schedule to high prices is modelled as a reduction of 25MW for every £1/MWh that the system marginal price rises above £75/MWh.

750 trading days are simulated for each of the four combinations of a day-ahead market and the bilateral model under uniform pricing and discriminatory settlement. Prices are found to generally be higher under pay-as-bid pricing for both market models. Agents with larger portfolios are shown to have a significant advantage over smaller generators due to their greater ability to gather scarce market price information and distribute it among generators.

In (Bower, Bunn, & Wattendrup, 2001) a more sophisticated custom learning method, resembling the Roth-Erev method described in Appendix B.4, is applied to a more detailed model of the New Electrcity Trading Arragements. The balancing mechanism is modelled as a one-shot market, that follows the contract market, to which increment and decrement bids are submitted. Active demand side participation is modelled and generator dynamic constraints are represented by limiting the number of off/on cycles per day, but again, transmission constraints and regional variations are ignored.

Supplier and generator agents are assigned an optimal value for exposure to the balancing mechanism that is typically low due to high price and volume uncertainty. The agents learn to maximise profit, but profits are penalised if the objective for balancing mechanism exposure is not achieved. They learn policies for pricing markups on the bids submitted to the power exchange and the increments and decrements submitted to the balancing mechanism. Markups in the power exchange are relative to to prices from the previous day and markups on balancing mechanism bids are relative to power exchange bid prices on the same day. Different markup ranges are specified for generators and suppliers in the power exchange and balancing mechanism and each is partitioned into ten discrete intervals.

As with the Roth-Erev method, a probability for the selection of each markup is calculated by the learning method. Daily profits and acceptance rates for bids/offers from previous trading days are extrapolated out to determine expected values and thus the expected reward for each markup. The markups are then sorted by expected reward in decending order. The perceived utility of each markup j is

$$U_j = \mu\left(\frac{\phi - n}{\phi}\right)^{i_j - 1} \tag{3.1}$$

where i is the index of j in the ordered vector of markups and ϕ is a search parameter. High values of ϕ cause the agent to adopt a more exploratory markup selection policy. For all of the experiments $\mu = 1000$, $\phi = 4$, $n = 3$ and the probability of selecting markup j is

$$Pr_j = \frac{U_j}{\sum_{k=1}^{K} U_k} \tag{3.2}$$

for K possible markups.

A representative model of the England and Wales system with 24 generator agents, associated with a total of 80 generating plants, and 13 supplier agents is analysed over 200 simulated trading days. The authors draw conclusions on the importance accurate forecasts, greater risk for suppliers than generators, the value of flexible plant and the influence of capacity margin on opportunities for collusive behaviour.

The same learning method is applied in (D. W. Bunn & Oliveira, 2003) as part of an inquiry by the Competition Commision into whether two specific companies in the England and Wales electricity market had enough market power to operate against the public interest.

Another early publication on agent-based simulation of electricity markets in which a custom learning method is used is (Visudhiphan & Ilic, 1999). The simulations comprise only three generators, market power is assumed, and the authors analyse the mechanisms by which the market power is exercised. Two bid formats are modelled. The *single-step supply function* model requires each generator to submit and price and a quantity, where the quantity is determined by the generator's marginal cost function. The *linear supply function* model requires each generator to submit and value corresponding to the slope of its supply function. The bid price or slope value for generator i after simulation period t is

$$x_i(t+1) = x_i(t) + b_i(p_m(t))u_i(t) \tag{3.3}$$

where $b_i \in \{-1, 0, 1\}$ is the reward as a function of the market clearing price p_m from stage t and u_i is a reward gain or attenuation parameter. The reward b_i is defined according to strategies for estimated profit maximistaion and competition to be the base load generator. Both elastic and inelastic load models are considered. Using the single-step supply function model, the two strategies are compared in a day-ahead market setting, using a case where there is sufficient capacity to meet demand and a case where there is excessive capacity to the point where demand can be met by just two of the generators. The linear suuply model is analysed using moth day-ahead and hour-ahead markets with inelastic load. The hour-ahead simulation is repeated with elastic demand response.

The first coauthor goes on to compare a similar custom learning method with two other algoithms in her thesis (Visudhiphan, 2003). The custom method is designed specifically for the power pool model that is implemented and uses separate policies for selecting bid quantities and prices according to a slew of if-then rules that attempt to capture capacity withholding behaviour. The method is compared with algorithms developed in (Auer, Cesa-Bianchi, Freund, & Schapire, 2003) for application to the n-armed bandit problem (Robbins, 1952; Sutton & Barto, 1998, §2.1) and a method based on evaluative feedback with softmax action selection (Sutton & Barto, 1998, §2).

In the algorithms from (Auer et al., 2003) each action $i = 1, 2, \ldots K$ for K possible actions is associated with a weight $w_t(i)$ in simulation period $t \in T$, where T is the total number of simulation periods, that is used in determining the action's probability of selection

$$p_i(t) = (1-\gamma)\frac{w_i(t)}{\sum_{j=1}^{K} w_j(t)} + \frac{\gamma}{K} \tag{3.4}$$

where γ is a tuning parameter, with $0 < \gamma \leq 1$, that is initialised such that

$$\gamma = \min\left\{\frac{3}{5}, 2\sqrt{\frac{3}{5}\frac{K \ln K}{T}}\right\}. \tag{3.5}$$

Using the received reward $x_t(i_t)$, the weight for action j in period $t+1$ is

$$w_{t+1}(j) = w_t(i)\exp\left(\frac{\gamma}{3K}\left(\hat{x}_t(i) + \frac{\alpha}{p_t(i)\sqrt{KT}}\right)\right) \tag{3.6}$$

where

$$\hat{x}_t(i) = \begin{cases} x_t(j)/p_t(i) & \text{if } j = i_t \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

and

$$\alpha = 2\sqrt{\ln(KT/\gamma)}. \tag{3.8}$$

In the evaluative feedback method from (Sutton & Barto, 1998, §2) each action i has a value $Q_t(i)$ in simulation period t equal to the expected average reward if that action is selected. The softmax method uses a Boltzman distribution to select actions with probability

$$p_t(i) = \frac{e^{Q_t(i)/\tau}}{\sum_{j=1}^{K} e^{Q_t(j)/\tau}} \tag{3.9}$$

where τ is a *temperature* parameter with $\tau > 0$. The value of action i in the $(t+1)^{th}$

period is

$$Q_{t+1}(i) = \begin{cases} (1-\alpha)Q_t(i) + \alpha r_t(i) & \text{if } i_{t+1} = i \\ Q_t(i) & \text{otherwise} \end{cases} \tag{3.10}$$

where α is a constant *step-size* parameter with $0 < \alpha \leq 1$.

An extensive suite of simulation reuslts are reported and the choice of learning method is found to have a significant impact on agent performance, but no quantitative comparison measure is provided and no conclusions as to which method is the superior are drawn.

3.1.2 Financial Transmission Rights

In (Ernst, Minoia, & Ilic, 2004) a custom learning method is defined and used to study generator and supplier profits where financial transmission rights are part of the market. A two node transmission system is defined with one lossless transmission line of limited capacity, endowed to a transmission operator agent. Generator agents submit bids for their respective generating plants and the transmission owner submits a bid representing the cost per MW of transmitting power between the nodes. The market operator clears the bids, minimising costs while balancing supply and demand and not violating the line capacity. Prices at each node are calculated to provide a signal for both energy and transmission costs.

Each agent selects its bid according to a calculation of the reward it would expect to receive if all other agents were to bid as they did in the previous stage. If multiple bids are found to have the same value then the least expensive is selected. In the first period, previous bids are assumed to be at marginal cost. Several case studies are examined with different numbers of generators and line capacities, but few concrete conclusions are drawn.

3.2 Simulations Applying Q-learning

Agent-based simulation of electricity markets has been carried out with participants behavioral aspects modelled using the Q-learning methods described in Appendix B.2.3.

3.2.1 Nash Equilibrium Convergence

The most prominent work in which Q-learning is applied was conducted at the Swiss Federal Institutes of Technology in Zurich and Lausanne. The foundations for this work were laid in (Krause et al., 2004) with a comparison of agent-based modelling using reinforcement learning and Nash equilibrium analysis when assessing network constrained power pool market dynamics. Parameter sensistivity of comparison results were later analysed in (Krause et al., 2006).

The authors model a mandatory spot market which is cleared using a DC optimal power flow formulation. A five bus power system model is defined with three generators and four inelastic and constant loads. Linear marginal cost functions

$$C_{g,i}(P_{g,i}) = b_{g,i} + s_{g,i}P_{g,i} \tag{3.11}$$

are defined for each generator i where $P_{g,i}$ is the active power output, $s_{g,i}$ is the slope of the cost function and $b_{g,i}$ is the cost when $P_{g,i} = 0$. Suppliers are given the option to markup their bids to the market not by increasing $s_{g,i}$, but increasing $b_{g,i}$ by either 0, 10, 20 or 30%.

Nash equilibrium is computed by clearing the market for all possible markup combinations and determining the actions for which no player is motivated to deviate from, as it would result in a decrease in expected reward. Experiments are conducted in which there is a single Nash equilibrium and where there are two Nash equilibria.

An ϵ-greedy strategy is applied for action selection and a *stateless* action value function is updated at each time step t according to

$$Q(a_t) \leftarrow Q(a_t) + \alpha(r_{t+1} - Q(a_t)) \tag{3.12}$$

where α is the learning rate. Further to (Krause et al., 2004), simulations with discrete sets of values for the parameters α and ϵ were carried out in (Krause et al., 2006). While parameter variations affected the frequency of equilibrium oscillations, Nash equilibrium was still approached and the oscillatory behaviour observed for almost all combinations.

The significance of this research is that is verifies that the agent-based approach settles at the same theoretical optimum as with closed-form equilibrium approaches and that exploratory policies result in the exploitation of multiple equlibria if they exist.

Convergence to a Nash equilibrium is also confirmed in (Naghibi-Sistani, Akbarzadeh-T., Javidi-D.B., & Rajabi-Mashhadi, 2006). Boltzman (soft-max) exploration is used for action selection with the temperature parameter adjusted during the simulations. A modified version of the IEEE 30 bus test system is used with the number of generators reduced from nine to six. No optimal power flow formulation or details of the reward signal used are provided. Generators are given a three step action space where the slope of a linear supply function may be less then, equal to or above marginal cost. The experimental results show that with temperature parameter adjustment Nash equilibrium is achieved and the oscillations associated with ϵ-greedy action selection are avoided.

3.2.2 Congestion Management Techniques

Having validated the suitability of an agent-based, bottom-up, approach to assessing evolution of market characteristics, the authors applied the same technique in a comparison of congestion management schemes (Krause & Andersson, 2006). The first scheme considered was locational marginal pricing, or nodal pricing, where congestion is managed by optimising the output of generators with respect to maximum social welfare. The "market splitting" scheme they considered is similar to locational marginal pricing, but the system gets subdivided into zones, within which the nodal prices are uniform. The final "flow based market coupling" scheme also features uniform zonal pricing, but requires a simplified representation of the network. Power flows within zones are not represented and all lines within zones are aggregated into one equivalent interconnector.

As an alternative to the conventional DC optimal power flow formulation, line power flows computation is done using a power transfer distribution factor (PTDF) matrix. The $(i, j)^{th}$ element of this corresponds to the change in active power flow on line j given an additional injection of 1MW at the slack bus and corresponding withdrawl of 1MW at node i.

The congestion management schemes get evaluated under perfect competition, where suppliers bid at marginal cost, and under oligopolistic competition, in which markups of 5% and 10% can be added to marginal cost. The benefits obtained between reward at marginal cost and a maximum markup are used to assess market power. The experimental results show different market power allocations under the three constraint management schemes.

3.2.3 Gas-Electricity Market Integration

The Q-learning method from (Krause et al., 2004, 2006) is used to analyse strategic behaviour in integrated electricity and gas markets in (Kienzle, Krause, Egli, Geidl, & Andersson, 2007). Again, power flows are computed using a PTDF matrix. Pipeline losses in the gas network are approximated using using a cubic function of flow and three combined gas and electricity models are compared.

In the first model, operators of gas-fired power plant submit separate bid functions for gas and electricity. Bids are then cleared as a single optimisation problem. In model two, operators submit one offer for their capacity to convert gas to electricity. In the third model, bids are submitted only to the electricity market, after which gas is purchased regardless of price. Gas supply offers are modelled as a linear function with no strategic involvement. The models are compared in terms of social welfare, using a three bus power system model with three non-gas-fired power plants and one gas-fired plant.

The experimental results show little difference between electricity prices and social welfare prices between the models. However, this research illustrates the interest in and complexity associated with modelling relationships between markets. The authors recognise the need for further and more detailed simulation in order to improve evaluation of market coupling models.

3.2.4 Electricity-Emissions Market Interactions

Researchers at the Argonne National Laboratory have published results from a preliminary study of interactions between *emission* and electricity markets (J. Wang, Koritarov, & Kim, 2009). A cap-and-trade system for emissions is modelled where generator companies are allocated with CO_2 allowances that may subsequently be traded. Generator companies are assumed to have negligible influence on market clearing prices in the emissions market and allowance prices from the European Energy Exchange were used. In the electricity market, an oligopoly is assumed and bids are cleared using a DC optimal power flow formulation.

To improve selection of the ϵ parameter for exploratory action selection, a simulated annealing (SA) Q-learning method based on the Metropolis criterion is used (Guo, Liu, & Malec, 2004). Under this method ϵ is changed at each simulation step to allow solutions to escape from local optima. A two bus system is used to study cases in which allowance trading is not used, allowances can be exchanged in the emission market and with variations in the allowance allocations. A one year, hourly load profile with a summer peak is used to model changes in demand. The electricity market is cleared each hour and the emissions market gets cleared at the end of each simulated week.

The agents learn, when they have a defecit of allowances, to borrow future allowances in the summer when load and allowance prices are high. Conversly, when having a surplus, they learn to sell at this time. In the third case, the authors show the sensitivity of profits to initial allocations and conclude that the experimental results can not be generalised. The authors cite further model validation and agent learning method improvements as necessary future work.

3.2.5 Tacit Collusion

The SA-Q-learning method was previously used in (Tellidou & Bakirtzis, 2007) by researchers from the University of Thessaloniki to study capacity withholding and tacit collusion among electricity market participants. A mandatory spot market is implemented, where bid quantities may be less than net capacity and bid prices may be marked up upon marginal cost by increasing the slope of a linear cost function. Again market clearing is achieved using DC optimal power flow and locational marginal prices are used to calculate profits and reinforce the learning process. Demand is assumed to be inelastic and transmission system parameters constant between simulation periods.

A simple two node power system model containing two generators is applied in three test cases. In a reference case, each generator bids full capacity at marginal cost. In the second case, generators bid quantities in steps of 10MW and price markups in steps of €2/MWh. In the third case, the same generation capacity is split among eight identical generators to increase the level of competition. The experimental results show that generators learn to withhold capacity and develop tacit collusion strategies to capture congestion profits.

3.3 Simulations Applying Roth-Erev

Roth and Erev's reinforcement learning method (defined in Appendix B.4, below) has received considerable attention from the agent-based electricity market simulation community.

3.3.1 Market Power

In (Nicolaisen, Petrov, & Tesfatsion, 2002) an agent-based model of a wholesale electricity market with both supply and demand side participation is constructed. It is used to study market power and short-run market efficiency under discriminatory pricing through systematic variation of concentration and capacity conditions.

Modelling the power system, each trader is assigned values of available transmission capability (ATC) with respect to each other trader. Offers from buyers and sellers are matched on a merit order basis, with quantities restricted by ATC values. Two issues with the original Roth-Erev method are observed and the modified version defined in Appendix B.4.1 is proposed.

A maximum markup (markdown) of \$40/MWh is specified for each seller (buyer). Traders are not permitted to make negative profits and the feasible price range is divided into 30 offer prices for 1000 auction rounds cases and 100 offer prices for 10000 aution round cases. The parameters of the Roth-Erev method are calibrated using direct search within reasonable ranges. Nine combinations of buyer and seller numbers and total trading capacities are tested using the calibarated parameter values and *best-fit* values determined empirically in (Erev & Roth, 1998).

The experimental results show that good market efficiency is achieved under all configurations and sensitivity to method parameter changes is low. Levels of market power are found to be strongly predictive and little difference is found between cases in which opportunistic price offers are permitted and when traders are forced to bid at marginal cost. The results are compared with those from (Nicolaisen, Smith, Petrov, & Tesfatsion, 2000), in which genetic algorithms are used. The authors conclude that the reinforcement learning approach leads to higher market efficiency due their adaption according to *individual* profits.

Further research from Iowa State University, involving the modified Roth-Erev method, has been based upon the AMES wholesale electricity market test bed. A detailed description of AMES is provided in Section 3.5.8 below. In (Li & Tesfatsion, 2009b) it is used to investigate strategic capacity withholding in a wholesale electricity market design proposed by the U.S. Federal Energy Regulatory Commission in April 2003. A five bus power system model with five generators and three dispatchable loads is defined and capacity withholding is represented by premitting traders to bid lower than true operating capacity and higher than true marginal costs.

Comparing results from a benchmark case (in which true production costs are reported, but higher than marginal cost functions may be reported) and cases in which reported production limits may be less than the true values, the authors find that with sufficient capacity reserve there is no evidence to suggest potential for inducing higher net earnings through capacity withholding in the market design.

3.3.2 Italian Wholesale Electricity Market

Researchers from the University of Genoa have used the modified Roth-Erev method to study strategic behaviour in the Italian wholesale electricity market (Rastegar, Guerci, & Cincotti, 2009). The actual clearing procedure is modelled and a model of the Italian transmission system, including an interconnector to Sicily, with zonal subdivision, is defined. Within each of the 11 zones, thermal plant is combined according to technology (coal, oil, combined cycle gas, turbo gas and repower) and associated with one of 16 generation companies according to the size of the companies share. The resulting 53 agents are assumed to bid full capacity and may markup bid prices in steps of 5%, with a maximum markup of 300%.

Bids are cleared using a DC optimal power flow formulation with generation capacity constraints and zone interconnector flow limits. Agents are rewarded according to a uniform national price, computed as a weighted average of zonal prices with respect to zonal load. Using real hourly load data it is shown that in experiments in which agents learn their optimal strategy, histrorical trends can be replicated in all but certain hours of peak load. The authors state a desire to test different learning methods and perform further empirical validation.

3.3.3 Vertically Related Firms and Crossholding

In (Micola, Banal-Estañol, & Bunn, 2008) a multi-tier model of wholesale natural gas, wholesale electricity and retail electricity markets is studied using another variant of the Roth-Erev method. Coordination between strategic business units (SBU) within the same firm, but participating in different markets, is varied systematically and profit differences are analysed.

An initial two-tier model involves firms with two associated agents whose rewards, r^1 and r^2, are initially independant. A *reward independance* parameter α is used to control the fraction of profit from the other market that is used in rewarding the agent. The total rewards are

$$R^1(t) = (1-\alpha)r^1(t) + \alpha r^2(t) \tag{3.13}$$

and

$$R^2(t) = (1-\alpha)r^2(t) + \alpha r^1(t) \tag{3.14}$$

Each action a is a single price bid between zero and the clearing price from the preceeding market. The Roth-Erev method is modified such that similar actions, $a-1$ and $a+1$, are reinforced also. For each agent i, the action selection propensities in auction round t are

$$p_a^i(t) = \begin{cases} (1-\gamma)p_a^i(t-1) + R^i(t) & \text{if } s = k \\ (1-\gamma)p_a^i(t-1) + (1-\delta)R^i(t) & \text{if } s = k-1 \text{ or } s = k+1 \\ (1-\gamma)p_a^i(t-1) & \text{if } s \neq k-1,\ s \neq k \text{ or } s \neq k+1 \end{cases} \tag{3.15}$$

where δ, with $0 \leq \delta \leq 1$, is the local experimentation parameter, γ is the discount parameter and $i \in \{1, 2\}$. Actions whose probability of selection fall below a specified value are removed from the action space.

The initial simulation consists of two wholesalers and three retailers and α is varied from 0 to 0.5 in 51 discrete steps. The experiment is repeated using a three tier model in which two natural gas shippers supply three electricity generators who in turn sell to four electricity retailers. The results show a rise in market prices as reward interdependance is increased and greater profits for integrated firms.

The same alternative formulation of the Roth-Erev method is also used in (Micola & Bunn, 2008) to analyse the effect on market prices of different degrees of producer crossholding[1] under private and public bidding information. Crossholding is represented with the introduction of a factor to each agent's reward function that controls the fraction of profit from the crossowned rival that the agent receives. Public information availability is modelled using a vector of probabilities for selection of each possible action that is the average of each agent's private probability and is available to all agents. The degree to which the public probabilities influence the agent's action selection probability from equation (B.13) is varied systematically in a series of experiments, along with crossholding levels and buyer numbers. The results

[1] Crossholdings occur when one publically traded firm owns stock in another such firm.

are illustrated using three-dimensional plots and show a direct relationship between crossholding and market price. The conclusions drawn on market concentration by the authors are dependant upon the ability to model both the demand and supply side participation in the market and the authors state that this shows to a certian extent the value of agent-based simulation.

3.3.4 Two-Settlement Markets

In (Weidlich & Veit, 2006) the modified Roth-Erev method is used to study interrelationships between contracts markets and balancing markets. Bids on the day-ahead contracts market consist of a price and a volume, which are assumed to be the same for each hour of the day. Demand is assumed to be fixed and inelastic. Bids on the balancing market consist of a reserve price, a *work* price and an offered quantity. The reserve price is that which must be paid for the quantity to be kept on standby and the work price must be paid if that quantity is called upon for transmission system stabilisation.

No optimal power flow formulation or power system model is defined. At the day-ahead stage contract market and balancing market (according to reserve price) bids are cleared by stacking in order of ascending price until the forecast demand is met. On the following day, accepted balancing bids are cleared according to work price such that requirements for reserve dispatch are met.

Bid prices on the contracts market are stratified into 21 discrete values between 0 and 100 and bid quantities into six discrete values between 0 and maximum capacity, giving 126 possible actions. Bid quantities on the balancing market equal to the capacity remaining after contract market participation. 21 discrete capacity prices between 0 and 500 and 5 work prices between 0 and 100 are permitted, giving 105 possible actions in the balancing market. Separate instances of the modified Roth-Erev method are used to learn bidding strategies for each agent in each of the markets.

Interrelationsships between the markets are studied using four scenarios in which the order of market execution and the balancing market pricing mechanism (discriminatory or pay-as-bid) are changed. Clearing prices in the market executed first are shown to have a marked effect on prices in the following market. The authors find agent-based simulation to be a suitable tool for reproducing realistic market outcomes and recognise a need for more detailed models with larger action domains.

In the same year, the authors collaborated with Jian Yao and Shmuel Oren from the University of California to study the dynamics between two settlement markets using the modified Roth-Erev method. The markets are a forwards contract market, in which transmission constraints are ignored, and a spot market that is cleared using a DC optimal power flow formulation with line flows calculated using a PTDF matrix.

Zonal prices get set in the forward market as weighted averages of nodal prices with repect to historical load shares. Profits are determined using these zonal prices and nodal prices from optimisation of the spot market. Demand is assumed inelastic to price, but different contingency states with peak and low demand levels are examined. A 53 bus stylised model of the Belgian electricity system from (Yao, Oren, & Adler, 2007; Yao, Adler, & Oren, 2008) is used to validate the results against those obtained using equilibrium methods. The nineteen generators are divided among two firms which learn strategies for bid price and quantity selection using the modified Roth-Erev method with a set of fixed parameter values taken from (Erev & Roth, 1998). The results show that the presence of a forward contracts market produces lower overall electricity prices and lower price volatility. The authors mention that risk aversion is to be included in suppliers utility functions in future work.

3.4 Policy Gradient Reinforcement Learning

The direct policy search reinforcement learning methods defined in Appendix B.3 have been sucessfully applied in both laboratory and operational settings (Sutton et al., 2000; Peters & Schaal, 2006; Peshkin & Savova, 2002). This section summarises market related applications of these methods.

3.4.1 Finincial Decision Making

Conventionally, supervised learning techniques are used in financial decision making problems to minimise errors in price forecasts through training on sample data. In (Moody, Wu, Liao, & Saffell, 1998) a recurrent reinforcement learning method is used to optimise investment performance without forecasting prices. The method is “recurrent” since it uses information from past decisions as input. The authors compare direct profit and the Sharpe ratio (Sharpe, 1966, 1994) as reward signals.

The Sharpe ratio is a measure of risk adjusted return defined as

$$S_t = \frac{\text{Average}(r_t)}{\text{Standard Deviation}(r_t)} \tag{3.16}$$

where r_t is the return for period t.

The parameters θ of the trading system are updated in the direction of the steepest accent of the gradient of some performance function U_t with repect to θ

$$\Delta\theta_t = \rho\frac{dU_t(\theta_t)}{d\theta_t} \tag{3.17}$$

where ρ is the learning rate. Direct profit is the simplest performance function defined, but assumes traders are insensitive to risk. Investors being sensitive to losses are, in general, willing to sacrifice potential gains for reduced risk of loss. To allow on-line learning and parameter updates at each time period, the authors define a *differential* Sharpe ratio. By maintaining an exponential moving average of the Sharpe ratio, the need to compute return averages and standard deviations for the entire trading history at each simulation period is avoided. Alternative performance ratios, including the Information ratio, Appraisal ratio and Sterling ratio, are mentioned.

Simulations are conducted using artificial price data, equivalent to one year of hourly trade in a 24-hour market, and using 45 years of monthly data from the Standard & Poor (S&P) 500 stock index and 3 month Treasury Bill (T-Bill) data. In a portfolio management simulation, in which trading systems invest proportions of their wealth among three different securities, it was shown that trading systems maximising the differential Sharpe ratio, produced more consistent results and achieved higher risk adjusted returns than those trained to simply maximise profit. This result is of interest as the majority of reinforcement learning applications to electricity market simulation use direct profit for the reward signal and may benefit from using measures of risk adjusted return.

In (Moody & Saffell, 2001) the recurrent reinforcement learning method from (Moody et al., 1998) is contrasted with value function based methods. In addition to the Sharpe ratio, a Downside Deviation ratio is defined. Results from trading systems trained on half-hourly United States Dollar-Great British Pound foreign exchange rate data and, again, learning switching strategies between the S&P 500 index and T-Bills are presented. They show that the recurrent reinforcement learning method outperforms Q-learning in the S&P 500/T-Bill allocation problem. The authors observe also that the recurrent reinforcement learning method has a much simpler functional form, that the output, not being discrete, maps easily to real valued actions and that the algorithm is more robust to noise in the financial data and adapts quickly to non-stationary environments.

3.4.2 Grid Computing

In (Vengerov, 2008) a marketplace for computational resources in envisioned. The authors propose a market in which grid service suppliers offer to execute jobs submitted by customers for a price per CPU-hour. The problem formulation requires customers to request a quote for computing a job k for a time τ_k on n_k CPUs. The quote returned specifies a price P_k at which k would be charged and a delay time d_k for the job. The service provider's goal is to learn a policy for pricing quotes that maximises long term revenue when competing in a market with other providers. Price differentiation is implemented though provision of a standard service, priced at \$1/CPU-hour and a premium sercice a \$$P$/CPU-hour, with premium jobs prioratised over standard jobs. The state of the market environment is defined by the current expected delays in the standard and premium service classes and by $n_k\tau_k$ – the product of the number of CPUs requested and the job execution time. The reward $r(s,a)$ for action a in state s is the total price paid for the job. The policy gradient method employed is a modified version of Williams' REINFORCE where

$$Q(s_t,a_t) = \sum_{t=1}^{T} r(s_t,a_t) - \overline{r}_t \tag{3.18}$$

and $\overline{r}_t$ is the current average reward.

The authors recognise that their grid market model could be generalised to other multi-seller retail markets. The experimental results show that if all grid service providers simultaneously use the learning algorithm then the process converges to a Nash equilibrium. The results also showed that significant increases in profit were possible by offering both standard and premium services.

3.5 Open Source Power Engineering Software

To couple existing implementations of policy gradient reinforcement learning methods from the PyBrain machine learning library with scalable and extensible optimal power flow formulations, the Matlab[2] source code from MATPOWER was translated to the Python programming language for this thesis. With permission from the MATPOWER developers, the resulting package was released under the terms of the Apache License version 2.0 as a project named PYLON (Lincoln, Galloway, & Burt, 2009). This section describes the project in the context of other open source electric power engineering software and demonstrates the contribution made.

3.5.1 MATPOWER

Since 1996, a team of researchers at the Power Systems Engineering Research Center at Cornell University have been developing MATPOWER – a package of Matlab workspace files for solving power flow and optimal power flow problems (R. Zimmerman, Murillo-Sánchez, & Thomas, 2009). Initial development was part of the PowerWeb project in which a power exchange auction market simulator was created, that could be accessed by multiple users simultaneously through a web-based interface. It is available under a custom license that permits it to be used for any purpose providing the project and authors are cited correctly. MATPOWER is highly popular in education and research and has an active mailing list that is moderated by Ray Zimmerman.

MATPOWER includes five power solvers for both AC and DC problems. The default solver uses Newton's method (Tinney & Hart, 1967) with a full Jacobian matrix updated in each iteration. Two variations on the fast decoupled method (Stott & Alsac, 1974) described in (Amerongen, 1989) provide quicker convergence for certain networks. The standard Gauss-Seidel method (Glimn & Stagg, 1957) is provided largely for academic reasons and the DC solver provides a non-iterative solutions. The properties of Matlab sparse matrices are fully exploited to allow the solvers to scale well to very large systems. All functions are run from the Matlab command-line or from within users programs and no graphical user interface is provided.

Starting with version 4.0, MATPOWER includes the Matlab Interior Ppoint Solver (MIPS) that can be used for solving DC and AC optimal power flow problems

[2] Matlab is a registered tradeamark of The Mathworks, Inc.

Package	Language	Licence	PF	DCOPF	ACOPF	CPF	SSSA	TDS	SE	SP	GUI	RL
AMES	Java	GPL		•							•	•
DCOPFJ	Java	GPL		•								
MatDyn	Matlab									•		
MATPOWER	Matlab		•	•	•	•			•	•		
OpenDSS	Pascal	BSD	•							•	•	
PSAT	Matlab	GPL	•		•	•	•	•		•	•	
PYLON	Python	Apache	•	•	•				•	•	•	•
TEFTS	C					•		•		•		
VST	Matlab		•			•	•	•		•	•	
UWPFLOW	C					•				•		

Table 3.1: Open source electric power engineering software feature matrix.

(H. Wang, Murillo-Sanchez, Zimmerman, & Thomas, 2007). Previously, FMINCON from the Matlab Optimization Toolbox[3] was required or one of a suite of high performance closed-source solvers. TSPOPF is a collection of three AC optimal power flow solvers, implemented in the C programming language and released as Matlab MEX files. It includes the original implementation of the step-controlled interior point method from which MIPS was derived. MINOPF provides an interface to the Fortran based MINOS[4] solver, developed at the Systems Optimization Laboratory at Stanford University, and is available only for educational and research purposes. DC optimal power flow problems can be solved be solved with a QP interface to MIPS or using a MEX interface to BPMPD – a commercial interior point method for linear and quadratic programming.

MATPOWER has an *extensible* optimal power flow formulation that allows additional optimisation variables and problem constraints to be introduced by the user. It is used internally to extend the standard optimisation formulation to support piecewise linear cost functions, dispatchable loads, generator PQ capability curves and branch angle difference limit constraints. Examples of possible additional extensions include: reserve requirements, environmental costs and contingency constraints.

MATPOWER currently requires Matlab , version 6.5 or greater, which is a commercial software product from The Mathworks that is supported on all major platforms. However, with the recent inclusion of MIPS there is the posibility to that MATPOWER could be run on GNU/Octave[5] with minimal alteration.

3.5.2 MATDYN

MATDYN is an extension to MATPOWER developed by Stijn Cole from the Katholieke Universiteit Leuven for dynamic analysis of electric power systems. It was first released in 2009 under the same license as MATPOWER and has the same programming style. The MATPOWER case format is extended with structs for dynamic and event data. MATDYN uses MATPOWER to obtain a power flow solution that is then used in solving the system of differential algrbraic equations representing the power system. Results for MATDYN are validated against those obtained from PSS/E[6] and PSAT

[3]Optimization Toolbox is a registered trademark of The Mathworks, Inc.
[4]MINOS is trademark of Stanford Business Software, Inc.
[5]GNU/Octave is an free program for numerical computation with strong Matlab compatibility.
[6]PSS/E is a registered trademark of Siemens Power Transmission & Distribution, Inc. Power Technologies International.

(The Power System Analysis Toolbox described in Section 3.5.3, below) and show good correspondance.

3.5.3 Power System Analysis Toolbox

The Power System Analysis Toolbox (PSAT) is a Matlab toolbox for static and dynamic analysis of electric power systems developed by Federico Milano who is currently an Assistant Professor at the University of Castilla in Spain. It is released under the terms of the GNU General Public License (GPL) version 2 and offers routines for

- power flow,
- bifurcation analysis,
- optimal power flow,
- small signal stability anlysis,
- time domain simulation and
- phasor measurement unit placement.

A large number of input data formats are supported through Perl scripts and simulation reports can be exported as plain text, Excel spreadsheets or $\LaTeX\,2_\varepsilon$ code. PSAT may be run from the Matlab command-line or from a Matlab based graphical user interface. The graphical interface can be used with Simulink[7] to construct networks such as that shown in Figure X. A slightly modified version of PSAT that can be run from the GNU/Octave command-line is also available.

Optimal power flow problems are solved via an interface to the General Algebraic Modeling System (GAMS). GAMS defines optimisation problems using a high-level modelling language and has a large solver portfolio, including all of the major commercial and academic solvers. The interface can be used for solving single period optimal power flow problems where the objective function can model maximisation of social benefit, maximisation of the distance to the maximum loading condition or multi-objective of a combination of these. Multi-period optimal power flow is formulated as a mixed integer problem with linearised power balance constraints. The

[7]Simulink is a registered trademark of The Mathworks, Inc.

objective function models maximisation of social welfare, but is extended to include startup and shutdown costs.

Power flow and dynamic data are typically separated in electric power simulation tools, but in PSAT they are integrated. This combined with the large number of routines supported by PSAT can make the code base difficult to understand and modify. However, comprehensive documentation is included with PSAT and the mailing list is highly active. The majority of correspondance on this list concerns PSAT's dynamic simulation features. The price of GAMS licenses and the need for optimal power flow problems to be converted to the GAMS language before being solved could be considered barriers to its selection for certain projects.

3.5.4 UWPFLOW

UWPFLOW is a research tool for voltage stability analysis developed at the University of Waterloo, Ontario, and the University of Wisconsin-Madison. It is written in ANSI-C and is available as open source for research purposes only. The program can be run with the terminal command

```
$ uwpflow [-options] input_file
```

where `input_file` is the path to a data file in the IEEE common data format (CDF) (IEEE Working Group, 1973) that may contain High-Voltage Direct Current (HVDC) and Flexible Alternating Current Transmission System (FACTS) device data. Output is also in CDF and can include additional data for post-processing, including values for nose curve plots. An interface to UWPFLOW is provided with PSAT and can be used for bifurcation analysis.

3.5.5 TEFTS

The University of Waterloo also hosts TEFTS – a transient stability program for studying energy functions and voltage stability phenomena in AC/HVDC dynamic power system models. It too is written in ANSI-C and is licensed for research purposes only. An executable file for DOS is provided and the source package contains a simple example.

3.5.6 Voltage Stability Toolbox

The Voltage Stability Toolbox (VST) is a Matlab toolbox, developed at the Center for Electric Power Engineering at Drexel University in Philidelphia, for investigating stability and bifurcation issues in power systems. The source is available for any purpose providing that the authors are suitably cited. VST features routines for

- power flow,
- time domain simulation,
- static and dynamic bifurcation analysis,
- singularity analysis and
- eigenvalue analysis.

The feature matrix in Table 3.1 shows the similar capabilities of VST and PSAT. It was developed around the same time and has the same goals for educational and research applications. However it does not have the same quality of documentation nor such an active community of users and developers as PSAT.

3.5.7 Distribution System Simulator

In November 2008, the Open Distribution System Simulator (OpenDSS) was released by the Electric Power Research Institute (EPRI) as open source. Development of OpenDSS began in April 1997 and it has been used extensively in distributed generation impact assessments. It is the only open source program designed for both distribution and transmission system simulation.

OpenDSS supports steady-state analysis in the frequency domain, including power flow, harmonics and dynamics. Arbitrary n-phase unbalanced circuit analysis is supported using an object orientated data model. Circuit elements are defined in Object Pascal and solutions are found using a linear sparse matrix solver written in C and C++. OpenDSS is available under the Berkeley Software Distribution (BSD) license, which allows use for almost any purpose. Circuits are defined in scripts, using a domain specific language, that may be executed through a graphical user interface or a Common Object Model (COM) interface. The user interface also provides circuit data editing, plotting and power flow visualisation tools.

The power flow solver is fast and can be configured for repeated studies using daily, yearly or duty-cycle data. The multi-phase circuit model allows complex fault conditions to be defined and three short-circuit analysis methods are provided. The heritage of OpenDSS is in harmonics and dynamics analysis and it does not support system optimisation.

3.5.8 Agent-based Modelling of Electricity Systems

The AMES (Agent-based Modeling of Electricity Systems) power market test bed is a software package that models core features of the Wholesale Power Market Platform – a market design proposed by the Federal Energy Regulatory Commission (FERC) in April 2003 for common adoption in regions of the U.S. (Sun & Tesfatsion, 2007a). The market design features:

- a centralised structure managed by an independent market operator,
- parallel day-ahead and real-time markets and
- locational marginal pricing.

Learning agents represent load serving entities or generating companies and learn using Roth-Erev methods (described in Appendix B.4) implemented with the Repast agent simulation toolkit (Gieseler, 2005). Agents learn from the solutions of hourly bid/offer based DC-OPF problems formulated as quadratic programs using the DCOPFJ package (Sun & Tesfatsion, 2007b) described in Section 3.5.9, below.

The capabilities of AMES are demonstrated using a 5-bus network model in (Li & Tesfatsion, 2009a). The model is provided with AMES and a step-by-step tutorial describes how it may be used. AMES comes with a Swing-based graphical user interface with plotting and table editor tools and is released under the the GNU GPL version 2.

3.5.9 DCOPFJ

To solve market problems defined in AMES, researchers at Iowa State University developed a stand-alone DC optimal power flow solver in Java named DCOPFJ. It formulates optimal power flow problems as convex quadratic programs and which are solved using QuadProgJ. The same researcher developed QuadProgJ as an independant solver that uses the dual active set strictly convex quadratic programming algorithm (Goldfarb & Idnani, 1983). DCOPFJ requires generator costs to be modelled as polynomial functions, of second order or less and no sparse matrix techniques are employed to allow application to large systems.

3.6 Summary

Chapter 4

Modelling Power Trade

The present chapter defines the models used to simulate electric power trade. An electricity market model is defined using an optimal power flow formulation, unit decommitment algorithm and an auction interface derived from (R. Zimmerman et al., 2009). Market participants are modelled as agents, with associated reinforcement learning methods, whose interactions with the auction interface are coordinated using a multi-agent system.

4.1 Electricity Market Model

Computation of the generator dispatch points is executed using parts of the of the optimal power flow formulation from MATPOWER. This section describes parts of the optimal power flow formulation, unit-decommitment algorithm and auction interface from MATPOWER that were used to represent a centralised electricity market. Notable components of the full optimal power flow forumlation that have been ignored are generator P-Q capability curves and dispatchable loads. The power flow equations associated with a network of these components are subsequently defined. The constrained cost variable approach to modelling generator cost functions from (H. Wang et al., 2007) is introduced, from which the optimal power flow formulation follows.

Since the optimal power flow formulations do not facilitate shutting down expensive generators, the unit-decommitment algorithm from MATPOWER is defined. Finally, to provide an interface to agent participants that resembles that of real electricity market, MATPOWER's auction wrapper for the optimal power flow routine is described.

4.1.1 Auction Interface

Solving the optimisation problem defined in section A is intended to represent the function of a pool market operator. To present agents participating in this market with an interface more representative of a real pool market, an auction clearing mechanism is implemented (R. D. Zimmerman & Murillo-Sánchez, 2009, p.31). The interface formulates optimal power flow problems from lists of offers to sell and bids to buy blocks of power.

An offer/bid specifies a quantity of power in MW and a price for that power in $/MWh, to be traded over a particular period of time. The market accepts sets of offers and bids and uses the solution of the unit de-commitment algorithm to clear the offers and bids as appropriate. The cleared offers/bids are then be used to compute values of revenue from which earnings/losses may be determined.

The interface allows maximum offer price limits and minimum bids price limits to be set. The clearing process the market begins by withholding offers/bids outwith these limits, along with those specifying non-positive quantities. Valid offers/bids for each generator are then sorted into non-decreasing/non-increasing order and used to form new piecewise-linear cost functions and adjust the generator's active power limits.

The dispatch points and nodal prices from solving the unit de-commitment optimal power flow with the newly configured generators as input are used in to determine the proportion of each offer/bid block that should be cleared and the associated price for each. Pricing may be uniform or discriminatory (pay-as-bid).

4.2 Multi-Agent System

This section describes the implementation of agents and the coordination of their interactions in multi-agent systems. A generic market environment, with which agents interact regardless of the learning method employed, is defined along with tasks that associate a purpose with an environment. The design of connectionist systems and tables, used to represent agent policies, are given and the process by which they are modified by the agent's learning algorithm is explained. Finally, the collection of agents and tasks into a multi-agent system and the sequence of interactions is illustrated.

4.2.1 Agent, Task & Environment

Environment

Each generator/dispatchable load in the power system model (See Section C.2, above) is associated with an agent[1] via the agent's environment. Each environment maintains an association with a singular market instance for submission of offers/bids. Two main operations are supported by an agent's environment.

For a power system with n_b buses, n_l and n_g generators, the `getSensors` method returns a $n_s \times 1$ vector of sensor values s_e^i for generator i where $n_s = 2n_b + 2n_l + 3n_g$. s_g^i represents the visible state of the environment for the agent associated with generator i. s_e^i is composed of sensor values for all buses, branches and generators.

$$s_{e,l}^i = \begin{bmatrix} P_f \\ Q_f \\ P_t \\ Q_t \\ \mu_{S_f} \\ \mu_{S_t} \end{bmatrix}, \quad s_{e,b}^i = \begin{bmatrix} V_m \\ V_a \\ \lambda_P \\ \lambda_Q \\ \mu_{v_{min}} \\ \mu_{v_{max}} \end{bmatrix}, \quad s_{e,g}^i = \begin{bmatrix} P_g \\ \mu_{p_{min}} \\ \mu_{p_{max}} \\ \mu_{q_{min}} \\ \mu_{q_{max}} \end{bmatrix} \quad s_e^i = \begin{bmatrix} s_{e,b}^i \\ s_{e,b}^i \\ s_{e,g}^i \end{bmatrix} \tag{4.1}$$

Not all values are used by the agent and the filtration is done according to the agent's task.

The `performAction` method takes $n_a \times 1$ vector of action values a_e if $s_{bid} = 0$, otherwise a $2n_a \times 1$ vector. If $s_{bid} = 0$, the i-th element of a_e is the offered/bid price in

[1]Management of a portfolio of generators is also supported by the architecture used, but this feature has not been exploited.

\$/MWh, where $i = 1, 2, \ldots n_{in}$. If $s_{bid} = 1$, the j-th element of a_e is the offered/bid price in \$/MWh, where $j = 1, 3, 5, \ldots n_{in} - 1$ and the k-th element of a_e is the offered/bid quantity in MW where $j = 2, 4, 6, \ldots n_{in}$. The action vector is separated into offers/bids and submitted to the market. If $s_{bid} = 0$, then $qty = p_{max}/n_{in}$.

Task

An agent does not interact directly with it's environment, but is associated with a particular task. A task associates a purpose with an environment and defines what constitutes a reward. Regardless of the learning method employed, the goal of an agent participant is to make a financial profit and the rewards are thus defined as the sum of earnings from the previous period t as calcualted by the market. Sensor data from the environment is filtered according to the task being performed. Agents using the value-function methods under test have a tabular representation of their policy with one row per environment state. Thus, observations consist of a single integer value s_v, where $s_v \leq n_s$ and $s_v \in \mathbb{Z}^+$. Agents using the policy-gradient methods under test have policy functions represented by connectionist systems that use an input vector w_i of arbitrary length where the i-th element $\in \mathbb{R}$. Before input to the connectionist policy function approximator, sensor values are scaled to be between -1 and 1. Outputs from the policy are denormalised using action limits before the action is performed on the environment.

Agent

Agent i is defined as an entity capable of producing an action a_i based on previous observations of its environment s_i, where a_i and s_i are vectors of arbitrary length. As illustrated in Figure X, each agent is associated with a *module*, a *learner* and a *dataset*. The module represents the agent's policy for action selection and returns an action vector a_m when activated with observation s_t. The value-function methods under test use modules which represent a $N \times M$ table, where N is the total number of states and M is the total number of actions.

$$\begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,m} \\ v_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ v_{n,1} & \cdots & \cdots & v_{n,m} \end{bmatrix} \tag{4.2}$$

Whereas for the policy gradient methods, the module is a connectionist network of other modules as illustrated in Figure X. The learner can use any reinforcement learning algorithm and modifies the values/parameters of the policy module to increase expected future reward. The dataset stores state-action-reward tuples for each interaction between the agent and its environment. The stored history is used by value-function learners when computing updates to the policy values. Policy gradient learners search directly in the space of the policy network parameters.

Value-function learners have an association with an explorer module which returns an explorative action a_e when activated with the current state s_t and action a_m from the policy module. For example, the ϵ-greedy explorer has a randomness parameter ϵ and a decay parameter d. When the ϵ-greedy explorer is activated, a random number x_r is drawn where $0 \leq x_r \leq 1$. If $x_r < \epsilon$ then a random vector of the same length as a_e is returned, otherwise $a_e = a_m$.

4.2.2 Simulation Event Sequence

In each simulation of a system consisting of one or more task-agent pairs a sequence of interactions is coordinated, as illustrated in Figure X.

At the beginning of each step/period the market is initialised and all offers/bids removed. From each task-agent tuple (T, A) an observation s_t is retrieved from T and integrated into agent A. When an action is requested from A its module is activated with s_t and the action a_e is returned. a_e is performed on the environment of A via its associated task T. Recall, this process involves the submission of offer/bids to the market. Once all actions have been performed the offer/bids are cleared using the auction mechanism. Each task T is requested to return a reinforcement reward r_t. All cleared offers/bids associated with the generator in the environment of T are retrieved from the market and r_t is computed from the difference between revenue and cost values.

$$r_t = \text{revenue} - (c_{fixed} + c_{variable}) \tag{4.3}$$

The reward r_t is given to agent A and the value is stored under a new sample is the dataset, along with the last observation s_t and the last action performed a_e. Each agent is instructed to learn from its actions using r_t, at which point the values/parameters of the module of A are updated according to the algorithm of the learner.

This constitutes one step of the simulation and the process is repeated until the specified number of steps are complete. Unless agents are reset, the complete history of states, actions and received rewards is stored in the dataset of each agent.

Chapter 5

Learning to Trade Power

To the best of the author's knowledge, this thesis presents the first case of policy gradient reinforcement learning methods being applied to electricity trading problems. It must first be proven that these methods are capable of learning a basic power trading policy. This section describes the method used to compare methods in their ability to do so.

5.1 Aims & Objectives

The purpose of this first experiment is to compare the relative abilities of value-function and policy gradient methods in learning a basic policy for trading power. The objective of the exercise is to examine:

- Speed of convergence to an optimal policy,
- Magnitude and variance of profit and,
- Sensitivity to algorithm parameter changes.

5.2 Method of Simulation

Each learning method is tested individually using a range of parameter configurations. A power system model with one bus, one generator k and one dispatchable load l, as illustrated in Figure X is used. In this context, the market clearing process is equivalent to creating offer and bids stacks and finding the point of intersection. A passive agent is associated with the dispatchable load. This agent bids for $-p_{g,l}^{min}$

at marginal cost each period regardless of environment state or reward signal. A dispathcable load is used instead of a constant load to allow a price to be set. Generator k is given sufficient capacity to supply the demand of the dispatchable load, $p_{g,k}^{max} > -p_{g,l}^{min}$, and the marginal of the k is half that of the load l. The generator and dispatchable load attributes are given in Table X. A price cap for the market is set to twice the marginal cost of the l at full capacity, $p_{g,l}^{min}$. The DC optimal power flow formulation (See Section A, above) is used to clear the market and reactive power trade is omitted. The Python code used to conduct the simulations is provided in Listing X.

5.3 Results

5.4 Discussion

5.5 Critical Analysis

Chapter 6

Competitive Power Trade

Having compared the learning methods in a one-player context, this section describes the method used to pit them against one and other and compare their performance.

6.1 Aims & Objectives

Competition is fundamental to markets and this experiment aims to compare learning methods in a complex dynamic market environment with multiple competing participants. The objective is to compare:

- Performance, in terms of profitability, over a finite number of periods,
- Profitability when trading both active and reactive power.
- Consistency of profit making and,
- Sensitivity to algorithm parameter changes.

6.2 Method of Simulation

Figure X illustrates the structure of the six bus power system model, from (Wood & Wollenberg, 1996), with three generators and fixed demand at three of the buses used to provide a dynamic environment with typical system values. Bus, branch and generator attribute values are stated in Tables X, Y, Z, respectively. Three learning methods are compared in six simulations encapsulating all method–generator combinations.

A price cap c_{cap} of twice the marginal cost of the most expensive generator at full capacity is set by the market. The simulations are repeated for with agents actions composing both price and quantity and with just price. For the value-function methods, the state is defined by the market clearing price from the previous period, divided equally into x_s discrete states between 0 and c_{cap}. The state vector s_t for the policy gradient methods consists of the market clearing price and generator set-point from the previous period.

$$s_t = \begin{bmatrix} c_{mcp} \\ p_g \end{bmatrix} \tag{6.1}$$

The script used to conduct the simulation is provided in Listing X.

6.3 Results

6.4 Discussion

6.5 Critical Analysis

Chapter 7

System Constraint Exploitation

One of the main features of agents using policy gradient learning methods and artifical neural networks for policy function approximation is their ability to accept many signals of continuous sensor data. This section describes an experiment in which the power system is severly constrained for certain periods, resulting in elevated nodal marginal prices in particular areas. The methods are tested in their ability to exploit these constraints and improve their total accumulated reward.

7.1 Aims & Objectives

7.2 Results

7.3 Discussion

7.4 Critical Analysis

Chapter 8

Conclusions and Further Work

8.1 Summary Conclusions

8.2 Further Work

8.2.1 AC Optimal Power Flow

8.2.2 Decentralised Trade

8.2.3 Standarisation

8.2.4 Blackbox optimisation

Bibliography

Alam, M. S., Bala, B. K., Huo, A. M. Z., & Matin, M. A. (1991). A model for the quality of life as a function of electrical energy consumption. *Energy*, *16*(4), 739–745.

Amerongen, R. van. (1989, May). A general-purpose version of the fast decoupled load flow. *Power Systems, IEEE Transactions on*, *4*(2), 760–770.

Auer, P., Cesa-Bianchi, N., Freund, Y., & Schapire, R. E. (2003). The nonstochastic multiarmed bandit problem. *SIAM Journal of Computing*, *32*(1), 48–77.

Bellman, R. E. (1961). *Adaptive control processes - a guided tour.* Princeton, New Jersey, U.S.A.: Princeton University Press.

Bower, J., & Bunn, D. (2001, March). Experimental analysis of the efficiency of uniform-price versus discriminatory auctions in the england and wales electricity market. *Journal of Economic Dynamics and Control*, *25*(3-4), 561-592.

Bower, J., Bunn, D. W., & Wattendrup, C. (2001). A model-based analysis of strategic consolidation in the german electricity industry. *Energy Policy*, *29*(12), 987-1005.

Bunn, D., & Martoccia, M. (2005). Unilateral and collusive market power in the electricity pool of England and Wales. *Energy Economics*.

Bunn, D. W., & Oliveira, F. S. (2003). Evaluating individual market power in electricity markets via agent-based simulation. *Annals of Operations Research*, 57–77.

Carpienterm, J. (1962, August). Contribution à l'étude du Dispatching Economique. *Bulletin de la Society Francaise Electriciens*, *3*(8), 431–447.

Department of Energy and Climate Change. (2009). Digest of United Kingdom Energy Statistics 2009. In (chap. 5). National Statistics – Crown.

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The*

American Economic Review, *88*(4), 848–881.

Ernst, D., Minoia, A., & Ilic, M. (2004, June). Market dynamics driven by the decision-making of both power producers and transmission owners. In *Power Engineering Society General Meeting, 2004. IEEE* (p. 255-260).

Gieseler, C. (2005). *A Java reinforcement learning module for the Repast toolkit: Facilitating study and implementation with reinforcement learning in social science multi-agent simulations.* Unpublished master's thesis, Department of Computer Science, Iowa State University.

Glimn, A. F., & Stagg, G. W. (1957, april). Automatic calculation of load flows. *Power Apparatus and Systems, Part III. Transactions of the American Institute of Electrical Engineers*, *76*(3), 817–825.

Goldfarb, D., & Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, *27*, 1–33.

Grainger, J., & Stevenson, W. (1994). *Power system analysis.* New York: McGraw-Hill.

Guo, M., Liu, Y., & Malec, J. (2004, October). A new q-learning algorithm based on the metropolis criterion. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *34*(5), 2140–2143.

ICF Consulting. (2003, August). *The economic cost of the blackout: An issue paper on the northeastern blackout.* (Unpublished)

IEEE Working Group. (1973, November). Common format for exchange of solved load flow data. *Power Apparatus and Systems, IEEE Transactions on*, *92*(6), 1916–1925.

Kienzle, F., Krause, T., Egli, K., Geidl, M., & Andersson, G. (2007, September). Analysis of strategic behaviour in combined electricity and gas markets using agent-based computational economics. In *1st European workshop on energy market modelling using agent-based computational economics* (pp. 121–141). Karlsruhe, Germany.

Kirschen, D. S., & Strbac, G. (2004). *Fundamentals of power system economics.* Chichester: John Wiley & Sons.

Krause, T., & Andersson, G. (2006). Evaluating congestion management schemes in liberalized electricity markets using an agent-based simulator. In *Power Engineering Society General Meeting, 2006. IEEE.*

Krause, T., Andersson, G., Ernst, D., Beck, E., Cherkaoui, R., & Germond, A. (2004). Nash Equilibria and Reinforcement Learning for Active Decision Maker

Modelling in Power Markets. In *Proceedings of 6th IAEE European Conference 2004, modelling in energy economics and policy.*

Krause, T., Beck, E. V., Cherkaoui, R., Germond, A., Andersson, G., & Ernst, D. (2006). A comparison of Nash equilibria analysis and agent-based modelling for power markets. *International Journal of Electrical Power & Energy Systems*, *28*(9), 599 – 607.

Li, H., & Tesfatsion, L. (2009a, July). The ames wholesale power market test bed: A computational laboratory for research, teaching, and training. In *IEEE Proceedings, Power and Energy Society General Meeting.* Alberta, Canada.

Li, H., & Tesfatsion, L. (2009b, March). Capacity withholding in restructured wholesale power markets: An agent-based test bed study. In *Power systems conference and exposition, 2009* (pp. 1–11).

Lincoln, R., Galloway, S., & Burt, G. (2009, May). Open source, agent-based energy market simulation with Python. In *Proceedings of the 6th International Conference on the European Energy Market, 2009. EEM 2009.* (pp. 1–5).

Micola, A. R., Banal-Estañol, A., & Bunn, D. W. (2008, August). Incentives and coordination in vertically related energy markets. *Journal of Economic Behavior & Organization*, *67*(2), 381–393.

Micola, A. R., & Bunn, D. W. (2008). Crossholdings, concentration and information in capacity-constrained sealed bid-offer auctions. *Journal of Economic Behavior & Organization*, *66*(3-4), 748-766.

Minkel, J. R. (2008, August 13). The 2003 northeast blackout–five years later. *Scientific American*.

Moody, J., & Saffell, M. (2001, July). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks*, *12*(4), 875–889.

Moody, J., Wu, L., Liao, Y., & Saffell, M. (1998). Performance functions and reinforcement learning for trading systems and protfolios. *Journal of Forecasting*, *17*, 441–470.

Naghibi-Sistani, M., Akbarzadeh-T., M., Javidi-D.B., M., & Rajabi-Mashhadi, H. (2006, November). Q-adjusted annealing for q-learning of bid selection in market-based multisource power systems. *Generation, Transmission and Distribution, IEE Proceedings*, *153*(6), 653–660.

Nicolaisen, J., Petrov, V., & Tesfatsion, L. (2002, August). Market power and efficiency in a computational electricity market with discriminatory double-auction pricing. *Evolutionary Computation, IEEE Transactions on*, *5*(5), 504–

523.

Nicolaisen, J., Smith, M., Petrov, V., & Tesfatsion, L. (2000). Concentration and capacity effects on electricity market power. In *Evolutionary Computation. Proceedings of the 2000 Congress on* (Vol. 2, pp. 1041–1047).

Peshkin, L., & Savova, V. (2002). Reinforcement learning for adaptive routing. In *Neural networks, 2002. IJCNN 2002. Proceedings of the 2002 international joint conference on* (Vol. 2, p. 1825-1830).

Peters, J., & Schaal, S. (2006, October). Policy gradient methods for robotics. In *Intelligent robots and systems, 2006 IEEE/RSJ international conference on* (pp. 2219–2225).

Peters, J., & Schaal, S. (2008). Natural actor-critic. *Neurocomputing*, *71*(7-9), 1180–1190.

Rastegar, M. A., Guerci, E., & Cincotti, S. (2009, May). Agent-based model of the italian wholesale electricity market. In *Energy market, 2009. 6th international conference on the european* (pp. 1–7).

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin American Mathematical Society*, *58*(5), 527–535.

Roth, A. E., Erev, I., Fudenberg, D., Kagel, J., Emilie, J., & Xing, R. X. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, *8*(1), 164–212.

Schweppe, F., Caramanis, M., Tabors, R., & Bohn, R. (1988). *Spot pricing of electricity*. Dordrecht: Kluwer Academic Publishers Group.

Sharpe, W. F. (1966, January). Mutual fund performance. *Journal of Business*, 119–138.

Sharpe, W. F. (1994). The Sharpe ratio. *The Journal of Portfolio Management*, 49–58.

Stott, B., & Alsac, O. (1974, May). Fast decoupled load flow. *Power Apparatus and Systems, IEEE Transactions on*, *93*(3), 859–869.

Sun, J., & Tesfatsion, L. (2007a). Dynamic testing of wholesale power market designs: An open-source agent-based framework. *Computational Economics*, *30*(3), 291–327.

Sun, J., & Tesfatsion, L. (2007b, June). Open-source software for power industry research, teaching, and training: A DC-OPF illustration. In *Power engineering society general meeting, 2007. ieee* (pp. 1–6).

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* MIT Press. Gebundene Ausgabe.

Sutton, R. S., McAllester, D., Singh, S., & Mansour, Y. (2000). Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems* (Vol. 12, pp. 1057–1063).

Tellidou, A., & Bakirtzis, A. (2007, Novemeber). Agent-based analysis of capacity withholding and tacit collusion in electricity markets. *Power Systems, IEEE Transactions on*, *22*(4), 1735–1742.

Tesfatsion, L., & Judd, K. L. (2006). *Handbook of computational economics, volume 2: Agent-based computational economics (handbook of computational economics).* Amsterdam, The Netherlands: North-Holland Publishing Co.

Tinney, W., & Hart, C. (1967, Novemeber). Power flow solution by Newton's method. *Power Apparatus and Systems, IEEE Transactions on*, *86*(11), 1449–1460.

United Nations. (2003, December 9). World population in 2300. In *Proceedings of the United Nations, Expert Meeting on World Population in 2300.*

Vengerov, D. (2008). A gradient-based reinforcement learning approach to dynamic pricing in partially-observable environments. *Future Generation Computer Systems*, *24*(7), 687–693.

Visudhiphan, P. (2003). *An agent-based approach to modeling electricity spot markets.* Unpublished doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Visudhiphan, P., & Ilic, M. (1999, February). Dynamic games-based modeling of electricity markets. In *Power Engineering Society 1999 Winter Meeting, IEEE* (Vol. 1, pp. 274–281).

Wang, H., Murillo-Sanchez, C., Zimmerman, R., & Thomas, R. (2007, Aug.). On computational issues of market-based optimal power flow. *Power Systems, IEEE Transactions on*, *22*(3), 1185–1193.

Wang, J., Koritarov, V., & Kim, J.-H. (2009, July). An agent-based approach to modeling interactions between emission market and electricity market. In *Power energy society general meeting, 2009. PES 2009. IEEE* (pp. 1–8).

Weidlich, A., & Veit, D. (2006, July 7-10). Bidding in interrelated day-ahead electricity markets - insights from an agent-based simulation model. In *Proceedings of the 29th IAEE International Conference.*

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. In *Machine learning* (pp. 229–256).

Wood, A. J., & Wollenberg, B. F. (1996). *Power Generation Operation and Control* (second ed.). New York: Wiley, New York.

Yao, J., Adler, I., & Oren, S. S. (2008). Modeling and computing two-settlement oligopolistic equilibrium in a congested electricity network. *Operations Research*, *56*(1), 34–47.

Yao, J., Oren, S. S., & Adler, I. (2007). Two-settlement electricity markets with price caps and cournot generation firms. *European Journal of Operational Research*, *181*(3), 1279–1296.

Zimmerman, R., Murillo-Sánchez, C., & Thomas, R. J. (2009, July). MATPOWER's extensible optimal power flow architecture. In *IEEE PES General Meeting.* Calgary, Alberta, Canada.

Zimmerman, R. D., & Murillo-Sánchez, C. E. (2009, December 24). MATPOWER 4.0b1 User's Guide (Version 4.0b1 ed.) [Computer software manual]. School of Electrical Engineering, Cornell University, Ithaca, NY 14853.

Appendix A

Optimal Power Flow

Nationalised electricity supply industries are typically planned operated and controlled centrally. A system operator determines which generators must operate and the required output of the operating units such that demand and reserve requirements are met and the overall cost of production is minimised. In electric power Engineering, this is termed the *unit commitment* and *economic dispatch* problem.

In 1962 a unit commitment formulation was published with network constraints included (Carpienterm, 1962). *Optimal power flow* is this integration of economic and power flow aspects of power systems into a mathematical optimisation problem. Its ability to solve centralised power system operation problems and determine prices in power pool markets has led to optimal power flow being one of the most widely studied subjects in the power systems community.

A.1 Power Flow Formulation

Optimal power flow derives its name from the *power flow*, or load flow, steady-state power system analysis technique. Given sets of generator data, load data and a nodal admittance matrix Y_{bus}, a power flow study determines the voltage

$$V_i = |V_i|\angle\delta_i = |V_i|(\cos\delta_i + j\sin\delta_i) \tag{A.1}$$

at each node i in the power system (Grainger & Stevenson, 1994). Y_{bus} describes the electrical network and its formulation is dependant upon the transmission line, transformer and shunt models employed[1]. Importantly, the relationship between nodal voltages and power entering the network is non-linear. For a network of n_b nodes, the current injected at node i is

$$I_i = \sum_{j=1}^{n_b} Y_{ij} V_j \tag{A.2}$$

where $Y_{ij} = |Y_{ij}|\angle\theta_{ij}$ is the $(i,j)^{th}$ element if the, $n_b \times n_b$, Y_{bus} matrix. Hence, the apparent power entering the network at bus i is

$$S_i = P_i + Q_i = V_i I_i^* = \sum_{n=1}^{n_b} |Y_{ij} V_i V_j| \angle(\delta_i - \delta_j - \theta_{ij}) \tag{A.3}$$

Converting to polar coordinates and separating the real and imaginary parts, the active power

$$P_i = \sum_{n=1}^{n_b} |Y_{ij} V_i V_j| \cos(\delta_i - \delta_j - \theta_{ij}) \tag{A.4}$$

and the reactive power entering the network

$$Q_i = \sum_{n=1}^{n_b} |Y_{ij} V_i V_j| \sin(\delta_i - \delta_j - \theta_{ij}) \tag{A.5}$$

at bus i are non-linear functions of V_i, as indicated by the presence of the sine and cosine terms. Kirchoff's Current Law requires that the net complex power injection (generation - load) at each bus equals the sum of complex power flows on each branch connected to the bus. The power balance equations

$$P_g^i - P_d^i = P^i \tag{A.6}$$

and

$$Q_g^i - Q_d^i = Q^i \tag{A.7}$$

where the subscripts g and d indicate generation and demand respectively, form a key constraint in the optimal power flow problem.

[1] The Y_{bus} formulation used in this thesis is defined in Section C.4.

A.2 Optimal Power Flow Formulation

Optimal power flow is a mathematical optimisation problem in which the complex power balance equations (A.6) and (A.7) must be satisfied. Optimisation problems have the general form

$$\min_{x} f(x) \tag{A.8}$$

subject to

$$\begin{aligned} g(x) &= 0 \qquad &(A.9)\\ h(x) &\leq 0 \qquad &(A.10)\end{aligned}$$

where x is the optimisation variable, f is the objective function and equations (A.9) and (A.10) are sets of equality and inequality constraints on x, respectively. Typical inequality constraints are bus voltage magnitude contingency state limits, generator output limits and branch power or current flow limits. The optimisation variable x may be made up of generator set-points, bus voltages, transformer tap settings etc. If the optimisation variable x is empty then the formulation reduces to a power flow problem as described in Section A.1, above.

A common objective of optimal power flow is total system cost minimisation. For and network of n_g generators the objective function is

$$\min \sum_{k=1}^{n_g} C_k(P_{g,k}) \tag{A.11}$$

where C_k is a cost function (typically quadratic) of the set-point $P_{g,k}$ for generator k. Alternative objectives may be to minimise losses, maxmimise the voltage stability margin or minimise deviation of an optimisation variable from a particular schedule.

A.3 Interior-Point Methods

Appendix B

Reinforcement Learning

Reinforcement learning is learning from reward by mapping situations to actions when interating with an uncertain environment (Sutton & Barto, 1998). An agent learns *what* to do in order to achieve a task through trial-and-error using a numerical reward or penalty signal without being instructed *how* to achieve it. In challenging cases, actions may not yield immediate reward or may affect the next situation and all subsequent rewards. A compromise must be made between exploitation of past experiences and exploration of the environment through new action choices. A reinforcement learning agent must be able to:

- sense aspects of its environment,
- take actions that influence its environment and,
- have an explicit goal or set of goals relating to the state of its environment.

In the classical model of agent-environment interaction, at each time step t in a sequence of discrete time steps $t = 1, 2, 3 \ldots$ an agent receives as input some form of the environment's state $s_t \in \mathcal{S}$, where $\mathcal{S}$ is the set of possible states. From a set of actions $\mathcal{A}(s_t)$ available to the agent in state s_t, the agent selects an action a_t and performs it upon its environment. The environment enters a new state s_{t+1} in the next time step and the agent receives a scalar numerical reward $r_{t+1} \in \mathbb{R}$ in part as a result of its action. The agent then learns from the state representations s_t and s_{t+1}, the chosen action a_t and the reinforcement signal r_{t+1} before beginning its next interaction. Figure X diagrams the agent-environment interaction event sequence.

B.1 Markov Decision Processes

For a finite number of states $\mathcal{S}$, if all states are Markov, the agent interacts with a finite Markov decision process (MDP). Informally, for a state to be Markov it must retain all relevant information about the complete sequence of positions leading up to the state, such that all future states and expected rewards can be predicted as well as would be possible given a complete history. A particular MDP is defined for a discrete set of time steps by a state set $\mathcal{S}$, an action set $\mathcal{A}$, a set of state transition probabilities $\mathcal{P}$ and a set of expected reward values $\mathcal{R}$. Given a state s and an action a, the probability of transitioning to each possible next state s' is

$$\mathcal{P}^a_{ss'} = \Pr\big\{s_{t+1} = s' | s_t = s, a_t = a\big\}. \tag{B.1}$$

Given the next state s', the expected value of the next reward is

$$\mathcal{R}^a_{ss'} = E\big\{r_{t+1} | s_t = s, a_t = a, s_{t+1} = s'\big\}. \tag{B.2}$$

In practice not all state signals are Markov, but should provide a good basis for predicting subsequent states, future rewards and selecting actions.

If the state transition probabilities and expected reward values are not known, only the states and actions, then samples from the MDP must be taken and a value function approximated iteratively based on new experiences generated by performing actions.

B.2 Value Function Methods

Any method that can optimise control of a MDP may be considered a reinforcement learning method. All search for an optimal policy π^* that maps state $s \in \mathcal{S}$ and action $a \in \mathcal{A}$ to the probability $\pi^*(s, a)$ of taking a in s and maxmimises the sum of rewards over the agents lifetime.

Each state s under policy π may be associated with a *value* $V^\pi(s)$ equal to the expected return from following policy π from state s. Most reinforcement learning methods are based on estimating the state-value function

$$V^\pi(s) = E\left\{ \sum_{t=0}^{\infty} \gamma^t r_t \middle| s_0 = s \right\} \tag{B.3}$$

where γ is a discount factor, with $0 \leq \gamma \leq 1$. Performing certain actions may result in no state change, creating a loop and causing the value of that action to be infinite for certain policies. The discount factor γ prevents values from going unbounded and represents reduced trust in the reward r_t as discrete time t increases. Many reinforcement learning methods estimate the action-value function

$$Q^\pi(s,a) = E\left\{ \sum_{t=0}^{\infty} \gamma^t r_t \middle| s_0 = s, a_0 = a \right\} \tag{B.4}$$

which defines the value of taking action a in state s under fixed policy π.

B.2.1 Temporal-Difference Learning

Temporal Difference (TD) learning is a central idea in reinforcement learning. TD methods do not attempt to estimate the state transition probabilities and expected rewards of the finite MDP, but estimate the value function directly. They learn to *predict* the expected value of total reward returned by the state-value function (B.3). For an exploratory policy π and a non-terminal state s, an estimate of $V^\pi(s_t)$ at any given time step t is updated using the estimate at the next time step $V^\pi(s_{t+1})$ and the observed reward r_{t+1}

$$V^\pi(s_t) \leftarrow V^\pi(s_t) + \alpha\big[r_{t+1} + \gamma V^\pi(s_{t+1}) - V^\pi(s_t)\big] \tag{B.5}$$

where α is the learning rate, with $0 \leq \alpha \leq 1$, which controls how much attention is paid to new data when updating V^π. TD learning evaluates a particular policy and offers strong convergence guarantees, but does not learn better policies.

B.2.2 Sarsa

Sarsa (or modified Q-learning) is an on-policy TD control method that approximates the state-action value function in (B.4). Recall that the state-action value function for an agent returns the total expected reward for following a particular policy for selecting actions as a function of future states. The function is updated according to the rule

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\big[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)\big] \tag{B.6}$$

This update also uses the action from the next time step a_{t+1} and the requirement to transition through state-action-reward-state-action for each time step derives the algorithm's name. Sarsa is referred to as an on-policy method since it learns the same policy that it follows.

B.2.3 Q-Learning

Q-learning is an off-policy TD method that does not estimate the finite MDP directly, but iteratively approximates a state-action value function which returns the value of taking action a in state s and following an *optimal* policy thereafter. The same theorems used in defining the TD error also apply for state-action values.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\big[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)\big] \tag{B.7}$$

The method is off-policy since the update function is independant of the policy being followed and only requires that all state-action pairs be continually updated.

B.2.4 Eligibility Traces

With the TD methods described above, only the value for the immediately preceeding state or state-action pair is updated at each time step. However, the prediction $V(s_{t+1})$ also provides information concerning earlier predictions and TD methods can be extended to update a set of values at each step. An eligibility trace $e(s)$ represents how eligible the state s is to receive credit or blame for the TD error

$$\delta = r_{t+1} + \gamma V^{\pi}(s_{t+1}) - V^{\pi}(s_t) \tag{B.8}$$

When extended with eligibility traces TD methods update values for all states

$$\Delta V_t(s) = \alpha \delta_t e_t(s) \tag{B.9}$$

For the current state $e(s) \leftarrow e(s) + 1$ and for all states the $e(s) \leftarrow \gamma\lambda e(s)$ where λ is the eligibility trace attenuated factor from which the extended TD methods TD(λ), Q(λ) and Sarsa(λ) derive their names. For $\lambda = 0$ only the preceeding value is updated is updated, as in the unextended definitions, and for $\lambda = 1$ all preceeding state-values or state-action values are updated equally.

B.2.5 Action Selection

Action selection may be acomplished using a form of the *softmax* method (Sutton & Barto, 1998) using the Gibbs, or Boltzmann, distribution to select action k for the $(t+1)^{th}$ interaction with probability

$$p_{jk}(t+1) = \frac{e^{q_{jk}(t+1)/\tau}}{\sum_{l=0}^{K} e^{q_{jl}(t+1)/\tau}} \tag{B.10}$$

where τ is the *temperature* parameter. This parameter may be lowered in value over the course of an experiment since high values give all actions similar probability and encourage exploration of the action space, while low values promote exploitation of past experience.

B.3 Policy Gradient Methods

Value function based methods have been successfully applied with discrete lookup table parameterisation to many problems [ref]. However, the number of discrete states required increases exponentially as the dimensions of the state space increase and if all possibly relevant situations are to be covered then these methods become subject to Bellman's Curse of Dimensionality (Bellman, 1961). Value function based methods can be used in conjunction with function approximators, artificial neural networks are popular, to work with continuous state and action space. However, when used with value function approximation they have been shown to offer poor convergence and even divergence characteristics, even in simple systems (Peters & Schaal, 2008).

These convergence problems have motivated research into policy gradient methods which make small incremental changes to the parameters θ of a policy function approximator. With artificial neural networks the parameters are the weights of the network connections. Policy gradient methods update θ in the direction of the gradient of some policy performance measure Y with respect to the parameters

$$\theta_{i+1} = \theta_i + \alpha \frac{\partial Y}{\partial \theta_i} \tag{B.11}$$

where α is a positive definite step size learning rate.

Aswell as working with continuous state and actions space, policy gradient methods offer strong convergence guarantees, do not require all states to be continually updated and although uncertainty in state data can degrade policy performance, the techniques need not be altered.

Policy gradient methods are differentiated largely by the techniques used to obtain an estimate of the policy gradient $\partial Y/\partial\theta$. The most successful real-world robotics results have been yielded using Williams' REINFORCE likelihood ratio methods (Williams, 1992) and natural policy gradient methods such as Natual Actor-Critic (Peters & Schaal, 2008).

B.4 Roth-Erev Method

The reinforcement learning method formulated by Alvin E. Roth and Ido Erev is based on empirical results obtained from observing how humans learn decision making strategies in games against multiple strategic players (Roth et al., 1995; Erev & Roth, 1998). It learns a stateless policy in which each action a is associated with a value q for the propensity of its selection. In time period t, if agent j performs action a' and receives a reward $r_{ja'}(t)$ then the propensity value for action a at time $t+1$ is

$$q_{ja}(t+1) = \begin{cases} (1-\phi)q_{ia}(t) + r_{ja'}(t)(1-\epsilon), & a = a' \\ (1-\phi)q_{ia}(t) + r_{ja'}(t)(\frac{\epsilon}{A-1}), & a \neq a' \end{cases} \tag{B.12}$$

where A is the total number of feasible actions, ϕ is the *recency* parameter and ϵ is the *experimentation* parameter. The recency (forgetting) parameter degrades the propensities for all actions and prevents propensity values from going unbounded. It is intended to represent the tendency for players to forget older action choices and to prioritise more recent experience. The experimentation parameter prevents the probability of choosing an action from going to zero and encourages exploration of the action space.

Erev and Roth proposed action selection according to a discrete probability distribution function, where action k is selected for interaction $t+1$ with probability

$$p_{jk}(t+1) = \frac{q_{jk}(t+1)}{\sum_{l=0}^{K} q_{jl}(t+1)} \tag{B.13}$$

Since $\sum_{l=0}^{K} q_{jl}(t+1)$ increases with t, a reward $r_{jk}(t)$ for performing action k will have a greater effect on the probability $p_{jk}(t+1)$ during early interactions while t is small. This is intended to represent Psychology's Power Law of Practice in which it is qualitatively stated that, with practice, learning occurs at a decaying exponential rate and that a learning curve will eventually flatten out.

B.4.1 Modified Roth-Erev Method

Two shortcomings of the basic Roth-Erev algorithm have been identified and a modified formulation proposed (Nicolaisen et al., 2002). The two issues are that

- the values by which propensities are updated can be zero or very small for certain combinations of the experimentation parameter ϵ and the total number of feasible actions A and
- all propensity values are decreased by the same amount when the reward, $r_{jk'}(t)$ is zero.

Under the variant algorithm, the propensity for agent j to select action a for interaction $t+1$ is:

$$q_{ja}(t+1) = \begin{cases} (1-\phi)q_{ia}(t) + r_{ja'}(t)(1-\epsilon), & a = a' \\ (1-\phi)q_{ia}(t) + q_{ja}(t)(\frac{\epsilon}{A-1}), & a \neq a' \end{cases} \tag{B.14}$$

As with the original Roth-Erev algorithm, the propensity for selection of the action that the reward is associated with is adjusted by the experimentation parameter. All other action propensities are adjusted by a small proportion of their current value.

Appendix C

MATPOWER OPF Formulation

Power systems are modelled as three-phase AC circuits operating in the steady-state, under balanced conditions that can be represented by an equivalent single phase nodal graph of busbars connected by branches (Grainger & Stevenson, 1994).

C.1 Branches

Following (R. D. Zimmerman & Murillo-Sánchez, 2009, p.11), each branch is modelled as a medium length transmission line in series with a regulating transformer at the "from" end. A nominal-π model with total series admittance $y_s = 1/(r_s + jx_s)$ and total shunt capacitance b_c represents the transmission line. The transformer is assumed to be ideal, phase-shifting and tap-changing, with the ratio between primary winding voltage v_f and secondary winding voltage $N = \tau e^{j\theta_{ph}}$ where τ is the tap ratio and θ_{ph} is the phase shift angle. Figure X diagrams the branch model. From Kirchhoff's Current Law the current in the series impedance is

$$i_s = \frac{b_c}{2} v_t - i_t \tag{C.1}$$

and from Kirchhoff's Voltage Law the voltage across the secondary winding of the transformer is

$$\frac{v_f}{N} = v_t + \frac{i_s}{y_s} \tag{C.2}$$

Substituting i_s from equation (C.1), gives

$$\frac{v_f}{N} = v_t - \frac{i_t}{y_s} + v_t \frac{b_c}{2y_s} \tag{C.3}$$

and rearranging in terms if i_t, gives

$$i_t = v_s \left(\frac{-y_s}{\tau e^{\theta_{ph}}} \right) + v_r \left(y_s + \frac{b_c}{2} \right) \tag{C.4}$$

The current through the secondary winding of the transformer is

$$N^* i_f = i_s + \frac{b_c}{2} \frac{v_f}{N} \tag{C.5}$$

Substituting i_s from equation(C.1) again, gives

$$N^* i_f = \frac{b_c}{2} v_t - i_t + \frac{b_c}{2} \frac{v_f}{N} \tag{C.6}$$

and substituting $\frac{v_f}{N}$ from equation (C.3) and rearranging, gives

$$i_s = v_s \left(\frac{1}{\tau^2} \left(y_s + \frac{b_c}{2} \right) \right) + v_r \left(\frac{y_s}{\tau e^{-j\theta}} \right) \tag{C.7}$$

Equations (C.4) and (C.7) are used in section C.4 below to define the system admittance matrices that describe the electrical network.

C.2 Generators

Each generator k is modelled as an apparent power injection $s_g^k = p_g^k + jq_g^k$ at a bus i, where p_g^k is the active power injection, q_g^k is the reactive power injection and each are expressed in per-unit to the system base MVA. Upper and lower limits on p_g^k are specified by p_{max}^k and p_{min}^k, respectively, where $p_{max}^k > p_{min}^k \geq 0$. Similarly, upper and lower limits on q_g^k are specified by q_{max}^k and q_{min}^k, respectively, where $q_{max}^k > q_{min}^k$.

C.3 Buses and Loads

At each bus i, constant active power demand is specified by p_d^i and reactive power demand by q_d^i. Upper and lower limits on the voltage magnitude at the bus are defined by $v_m^{i,max}$ and $v_m^{i,min}$, respectively. One generator bus $i \in \mathcal{I}_{ref}$ in the circuit is designated the *reference* bus and has voltage angle θ_k^{ref}. Dispatchable loads are modelled as generators with negative p_g^i and $p_{min}^i < p_{max}^i = 0$.

C.4 AC Power Flow Equations

Following (R. D. Zimmerman & Murillo-Sánchez, 2009, p.13), for a network of n_b buses, n_l branches and n_g generators, let Cg be the $n_b \times n_g$ bus-generator connection matrix such that the $(i, j)^{th}$ element of C_g is 1 if generator j is connected to bus i. The $n_b \times 1$ vector of complex power injections from generators at all buses is

$$S_{g,bus} = C_g \cdot S_g \tag{C.8}$$

where $S_g = P_g + jQ_g$ is the $n_g \times 1$ vector with the i^{th} element is equal to s_g^i.

Combining equations (C.4) and (C.7), the *from* and *to* end complex current injections for branch l are

$$\begin{bmatrix} i_f^l \\ i_t^l \end{bmatrix} = \begin{bmatrix} y_{ff}^l & y_{ft}^l \\ y_{tf}^l & y_{tt}^l \end{bmatrix} \begin{bmatrix} v_f^l \\ v_t^l \end{bmatrix} \tag{C.9}$$

where

$$y_{ff}^l = \frac{1}{\tau^2}\left(y_s + \frac{b_c}{2}\right) \tag{C.10}$$

$$y_{ft}^l = \frac{y_s}{\tau e^{-j\theta_{ph}}} \tag{C.11}$$

$$y_{tf}^l = \frac{-y_s}{\tau e^{j\theta_{ph}}} \tag{C.12}$$

$$y_{tt}^l = y_s + \frac{b_c}{2} \tag{C.13}$$

Let Y_{ff}, Y_{ft}, Y_{tf} and Y_{tt} be $n_l \times 1$ vectors where the l^{th} element of each corresponds to y_{ff}^l, y_{ft}^l, y_{tf}^l and y_{tt}^l, respectively. Furthermore, let C_f and C_t be the $n_l \times n_b$ branch-bus connection matrices, where $C_{f_{i,j}} = 1$ and $C_{t_{i,k}} = 1$ if branch i connects from bus j to bus k. The $n_l \times n_b$ branch admittance matrices are

$$Y_f = \mathbf{diag}(Y_{ff})C_f + \mathbf{diag}(Y_{ft})C_t \tag{C.14}$$

$$Y_t = \mathbf{diag}(Y_{tf})C_f + \mathbf{diag}(Y_{tt})C_t \tag{C.15}$$

and relate the complex bus voltages V to the branch “from” and “to” end current vectors

$$
\begin{aligned}
I_f &= Y_f V && \text{(C.16)} \\
I_t &= Y_t V && \text{(C.17)}
\end{aligned}
$$

The $n_b \times n_b$ bus admittance matrix

$$
Y_{bus} = C_f^{\mathsf{T}} Y_f + C_t^{\mathsf{T}} Y_t \tag{C.18}
$$

relates the complex bus voltages to the nodal current injections

$$
I_{bus} = Y_{bus} V \tag{C.19}
$$

The complex bus power injections are expressed as a non-linear function of V

$$
\begin{aligned}
S_{bus}(V) &= \mathbf{diag}(V) I_{bus}^* \\
&= \mathbf{diag}(V) Y_{bus}^* V^* && \text{(C.20)}
\end{aligned}
$$

As are the complex power injections at the "from" and "to" ends of all branches

$$
\begin{aligned}
S_f(V) &= \mathbf{diag}(C_f V) I_f^* \\
&= \mathbf{diag}(C_f V) Y_f^* V^* && \text{(C.21)} \\
S_t(V) &= \mathbf{diag}(C_t V) I_t^* \\
&= \mathbf{diag}(C_t V) Y_t^* V^* && \text{(C.22)}
\end{aligned}
$$

The net complex power injection (generation - load) at each bus must equal the sum of complex power flows on each branch connected to the bus. Hence the AC power balance equations are

$$
S_{bus}(V) + S_d - S_g = 0 \tag{C.23}
$$

C.5 DC Power Flow Equations

Following (R. D. Zimmerman & Murillo-Sánchez, 2009, p.14), the same power system model is used in the formulation of the linearised DC power flow equations, but the following additional assumptions are made:

- The resistance r_s and shunt capacitance b_c of all branches can be considered negligible.

$$y_s \approx \frac{1}{jx_s}, \quad b_c \approx 0 \tag{C.24}$$

- Bus voltage magnitudes $v_{m,i}$ are all approximately 1 per-unit.

$$v_i \approx 1e^{j\theta_i} \tag{C.25}$$

- The voltage angle difference between bus i and bus j is small enough that

$$\sin\theta_{ij} \approx \theta_{ij} \tag{C.26}$$

Applying the assumption that branches are lossless from equation (C.24), the quadrants of the branch admittance matrix in equations (C.10), (C.11), (C.12) and (C.13), approximate to

$$y_{ff}^l = \frac{1}{jx_s\tau^2} \tag{C.27}$$

$$y_{ft}^l = \frac{-1}{jx_s\tau e^{-j\theta_{ph}}} \tag{C.28}$$

$$y_{tf}^l = \frac{-1}{jx_s\tau e^{j\theta_{ph}}} \tag{C.29}$$

$$y_{tt}^l = \frac{1}{jx_s} \tag{C.30}$$

respectively. Applying the uniform bus voltage magnitude assumption from equation (C.25) to equation (C.9), the branch "from" end current approximates to

$$i_f \approx \frac{e^{j\theta_f}}{jx_s\tau^2} - \frac{e^{j\theta_t}}{jx_s\tau e^{-j\theta_{ph}}} \tag{C.31}$$

$$= \frac{1}{jx_s\tau}(\frac{1}{\tau}e^{j\theta_f} - e^{j(\theta_t+\theta_{ph})}) \tag{C.32}$$

and the branch "from" end complex power flow $s_f = v_f \cdot i_f^*$ approximates to

$$s_f \approx e^{j\theta_f} \cdot \frac{j}{x_s\tau}(\frac{1}{\tau}e^{-j\theta_f} - e^{j(\theta_t+\theta_{ph})}) \tag{C.33}$$

$$= \frac{1}{x_s\tau}\left[\sin(\theta_f - \theta_t - \theta_{ph}) + j\left(\frac{1}{\tau} - \cos(\theta_f - \theta_t - \theta_{ph})\right)\right] \tag{C.34}$$

Applying the voltage angle difference assumption from equation (C.26) yields the approximation

$$p_f \approx \frac{1}{x_s \tau}(\theta_f - \theta_t - \theta_{ph}) \tag{C.35}$$

Let B_{ff} and $P_{f,ph}$ be the $n_l \times 1$ vectors where $B_{ff_i} = 1/(x_s^i \tau^i)$ and $P_{f,ph_i} = -\theta_{ph}^i/(x_s^i \tau^i)$. Then if the system B matrices are

$$B_f = \mathbf{diag}(B_{ff})(C_f - C_t) \tag{C.36}$$

$$B_{bus} = (C_f - C_t)^{\mathsf{T}} B_f \tag{C.37}$$

then the real power bus injections are

$$P_{bus}(\Theta) = B_{bus}\Theta + P_{bus,ph} \tag{C.38}$$

where Θ is the $n_b \times 1$ vector of bus voltage angles and

$$P_{bus,ph} = (C_f - C_t)^{\mathsf{T}} + P_{f,ph} \tag{C.39}$$

The active power flows at the branch "from" ends are

$$P_f(\Theta) = B_f\Theta + P_{f,ph} \tag{C.40}$$

and $P_t = -P_f$ since all branches are assumed lossless.

C.6 AC OPF Formulation

Following (R. D. Zimmerman & Murillo-Sánchez, 2009, p.26), generator active and, optionally, reactive power output costs are defined by a convex n-segment piecewise linear cost function

$$c^{(i)}(x) = m_i p + c_i \tag{C.41}$$

for $p_i \leq p \leq p_{i+1}$ with $i = 1, 2, \ldots n$ where $m_{i+1} \geq m_i$ and $p_{i+1} > p_i$ as diagramed in Figure X. Since these costs are non-differentiable the constrained cost variable approach from (H. Wang et al., 2007) is used to make the optimisation problem smooth. For each generator i a helper cost variable y_i added to the objective function.

The inequality constraints

$$y_i \geq m_{i,j}(p - p_j) + c_j, \quad j = 1 \dots n \tag{C.42}$$

require that y_i lies on the epigraph[1] of $c^{(i)}(x)$. The objective of the optimal power flow problem is to minimise the sum of the cost variables for all generators.

$$\min_{\theta, V_m, P_g, Q_g, y} \sum_{i=1}^{n_g} y_i \tag{C.43}$$

Equation (C.23) forms an equality constraint which enforce the balance between the net complex power injection and injections into the network. Branch complex power flow limits S_{max} are enforced by the inequality constraints

$$|S_f(V)| - S_{max} \leq 0 \tag{C.44}$$

$$|S_f(V)| - S_{max} \leq 0 \tag{C.45}$$

and the reference bus voltage angle θ_i is fixed with the equality constraint

$$\theta_i^{ref} \leq \theta_i \leq \theta_i^{ref}, \quad i \in \mathcal{I}_{ref} \tag{C.46}$$

Upper and lower limits on the optimisation variables V_m, P_g and Q_g are enforced by the inequality constraints

$$v_m^{i,min} \leq v_m^i \leq v_m^{i,max}, \qquad i = 1 \dots n_b \tag{C.47}$$

$$p_g^{i,min} \leq p_g^i \leq p_q^{i,max}, \qquad i = 1 \dots n_g \tag{C.48}$$

$$q_g^{i,min} \leq q_g^i \leq q_q^{i,max}, \qquad i = 1 \dots n_g \tag{C.49}$$

C.7 DC OPF Formulation

Piecewise linear cost functions are also used to define generator active power costs in the DC optimal power flow formulation. Since the power flow equations are linearised, following the assumptions in equations (C.24), (C.25) and (C.26), the optimal power flow problem simplifies to a linear program. The voltage magnitude variables V_m and generator reactive power set-point variable Q_g are eliminated fol-

[1]Informally, the epigraph of a function is a set of points lying on or above its graph.

lowing the assumption in equation (C.26) since branch reactive power flows depend on bus voltage angle differences. The objective function reduces to

$$\min_{\theta, P_g, y} \sum_{i=1}^{n_g} y_i \tag{C.50}$$

Combining the nodal real power injections, expressed as a function of Θ, from equation (C.38), with active power generation P_g and active demand P_d, the power balance constraint is

$$B_{bus}\Theta + P_{bus,ph} + P_d - C_g P_g = 0 \tag{C.51}$$

Limits on branch active power flows $B_f\Theta$ and $B_t\Theta$ are enforced by the inequality constraints

$$B_f\Theta + P_{f,ph} - F_{max} \leq 0 \tag{C.52}$$

$$-B_f\Theta + P_{f,ph} - F_{max} \leq 0 \tag{C.53}$$

The reference bus voltage angle equality constraint from equation (C.46) and the p_g limit constraint from (C.48) are also applied.

C.8 Optimal Power Flow Solution

C.9 Unit De-commitment

The optimal power flow formulation defined in Section A above requires generators are dispatched within their upper and lower power limits. Expensive generators can not be completely shutdown, even if doing so would result in a lower total system cost. Algorithm 1 defines the unit de-commitment algorithm from (R. D. Zimmerman & Murillo-Sánchez, 2009, p.20) which allows a least cost commitment and dispatch to be determined using the optimal power flow formulation. The algorithm finds the least cost dispatch by solving repeated optimal power flow problems with different combinations of generating units that are at their minimum active power limit deactivated. The lowest cost solution is returned when no further improvement can be made and no candidate generators remain.

Algorithm 1 Unit de-commitment

initialise $N \leftarrow 0$
solve initial OPF
$L_{tot} \leftarrow$ total load capacity
while total min gen. capacity $> L_{tot}$ **do**
    $N \leftarrow N + 1$
end while
repeat
    for c in candidates **do**
        solve OPF
    end for
until done = True