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.

Signed: Date: September 5, 2010

Acknowledgements

I wish to thank Professor Jim McDonald for giving me the opportunity to study at The Institute for Energy and Environment and for giving me the freedom to pursue my own research interests. I also wish to thank my supervisors, Professor Graeme Burt and Dr Stuart Galloway, for their guidance and scholarship. I wish to offer very special thanks to my parents, my big brother and my little sister for all of their support throughout my PhD.

This thesis makes extensive use of open source software projects developed by researchers from other institutions. I wish to thank Dr Ray Zimmerman from Cornell University for his work on optimal power flow, researchers from the Dalle Molle Institute for Artificial Intelligence (IDSIA) and the Technical University of Munich for their work on reinforcement learning algorithms and artificial neural networks and Charles Gieseler from Iowa State University for his implementation of the Roth-Erev reinforcement learning method.

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

Abstract

In Electrical Power Engineering, learning algorithms can be used to model the strategies of electricity market participants. The objective of this work is to establish if *policy gradient* reinforcement learning algorithms can be used to create participant models superior to those using previously applied *value function* based methods.

Supply of electricity involves technology, money, people, natural resources and the environment. All of these aspects are changing and electricity market designs must be suitably researched to ensure that they are fit for purpose. In this thesis electricity markets are modelled as non-linear constrained optimisation problems, which are solved using a primal-dual interior point method. Policy gradient reinforcement learning algorithms are used to adjust the parameters of multi-layer feed-forward artificial neural networks that approximate each market participant's policy for selecting power quantities and prices that are offered in the simulated marketplace.

Traditional reinforcement learning methods that learn a value function have been previously applied in simulated electricity trade, but are largely restricted to discrete representations of a market environment. Policy gradient methods have been proven to offer convergence guarantees in continuous environments and avoid many of the problems that mar value function based methods.

Contents

Abstract **iv**

List of Figures **viii**

List of Tables **ix**

1 Introduction **1**
1.1 Research Motivation 1
1.2 Problem Statement 2
1.3 Research Contributions 3
1.4 Thesis Outline 4

2 Background **5**
2.1 Electric Power Supply 5
2.2 Electricity Markets 7
2.2.1 The England and Wales Electricity Pool 9
2.2.2 British Electricity Transmission and Trading Arrangements 11
2.3 Electricity Market Simulation 12
2.3.1 Agent-Based Simulation 12
2.3.2 Optimal Power Flow 13
2.4 Reinforcement Learning 19
2.4.1 Value Function Methods 20
2.4.2 Policy Gradient Methods 23
2.4.3 Roth-Erev Method 25
2.5 Summary 27

3 Related Work **29**
3.1 Custom Learning Methods 29
3.1.1 Market Power 29
3.1.2 Financial Transmission Rights 34
3.2 Simulations Applying Q-learning 34
3.2.1 Nash Equilibrium Convergence 34
3.2.2 Congestion Management Techniques 36
3.2.3 Gas-Electricity Market Integration 36
3.2.4 Electricity-Emissions Market Interactions 37
3.2.5 Tacit Collusion 38
3.3 Simulations Applying Roth-Erev 38

3.3.1 Market Power . . . 39
3.3.2 Italian Wholesale Electricity Market . . . 40
3.3.3 Vertically Related Firms and Crossholding . . . 42
3.3.4 Two-Settlement Markets . . . 43
3.4 Policy Gradient Reinforcement Learning . . . 44
3.4.1 Financial Decision Making . . . 45
3.4.2 Grid Computing . . . 46
3.5 Summary . . . 47

4 Modelling Power Trade 48
4.1 Electricity Market Model . . . 48
4.1.1 Optimal Power Flow . . . 49
4.1.2 Unit De-commitment . . . 50
4.2 Multi-Agent System . . . 51
4.2.1 Market Environment . . . 51
4.2.2 Agent Task . . . 54
4.2.3 Market Participant Agent . . . 55
4.2.4 Simulation Event Sequence . . . 56
4.3 Summary . . . 57

5 Nash Equilibrium Analysis 52
5.1 Introduction . . . 52
5.2 Aims and Objectives . . . 53
5.3 Method of Simulation . . . 53
5.4 Simulation Results . . . 56
5.5 Discussion and Critical Analysis . . . 57
5.6 Summary . . . 62

6 System Constraint Exploitation 63
6.1 Introduction . . . 63
6.2 Aims and Objectives . . . 63
6.3 Method of Simulation . . . 64
6.4 Simulation Results . . . 69
6.5 Discussion and Critical Analysis . . . 73
6.6 Summary . . . 74

7 Conclusions and Further Work 76
7.1 Summary Conclusions . . . 76
7.2 Further Work . . . 77
7.2.1 Parameter Sensitivity and Delayed Reward . . . 78
7.2.2 Alternative Learning Algorithms . . . 78
7.2.3 UK Transmission System . . . 79
7.2.4 AC Optimal Power Flow . . . 79
7.2.5 Multi-Market Simulation . . . 81

Bibliography 82

A Open Source Electric Power Engineering Software **90**
A.1 MATPOWER 90
A.2 MATDYN 93
A.3 PSAT 93
A.4 UWPFLOW 96
A.5 TEFTS 96
A.6 VST 97
A.7 OpenDSS 97
A.8 GridLAB-D 98
A.9 AMES 99
A.10 DCOPFJ 99
A.11 PYLON 99
A.12 Summary 100

B Case Data **102**
B.1 6-Bus Case 102
B.2 IEEE Reliability Test System 103

List of Figures

2.1 Basic structure of a three phase AC power system. 6
2.2 UK power station locations. 8
2.3 Pool bid structure. 10
2.4 Nominal-π medium length transmission line model in series with a phase shifting, tap changing transformer model. 15
2.5 Sequence diagram for the basic reinforcement learning model. . . 20
2.6 Multi-layer feed-forward perceptron with bias nodes. 24

3.1 Single-line diagram for a stylised Italian grid model. 41

4.1 Piecewise linear active power cost function with constrained cost variable minimsation illustrated. 50
4.2 Agent environment UML class diagram. 52
4.3 Learning agent UML class diagram. 55
4.4 Market experiment UML class diagram. 57
4.5 Sequence diagram for action selection process. 58
4.6 Sequence diagram for the reward process. 59
4.7 Sequence diagram for the SARSA learning process. 60

5.1 Single-line diagram for six bus power system model. 54
5.2 Average markup for agent 1 and standard deviation over 10 runs. 58
5.3 Average markup for agent 2 and standard deviation over 10 runs. 59
5.4 Average markup for agent 1 and standard deviation. 60
5.5 Average reward for agent 1 and standard deviation. 61

6.1 Generator cost functions for the IEEE Reliability Test System . . 66
6.2 Hourly, daily and weekly load profile plots from the IEEE Reliability Test System 67
6.3 IEEE Reliability Test System 68
6.4 Modified Roth-Erev compared with Stateful Roth-Erev. 69
6.5 Average rewards for agent 1 under two state configurations. . . . 70
6.6 Average rewards for agent 4 under two state configurations. . . . 71
6.7 Average rewards for two offers per generator. 72

7.1 UK transmission system. 80

A.1 UKGDS EHV3 model in PSAT Simulink network editor. 95

List of Tables

4.1 Example discrete action domain. 53

5.1 Generator cost configuration 1. 54
5.2 Generator cost configuration 2. 55
5.3 Agent rewards under cost configuration 1 56
5.4 Agent rewards under cost configuration 2 56

6.1 Generator types and cost parameters for the simplified IEEE Reliability Test System. 65
6.2 Agent portfolios. 65

A.1 Open source electric power engineering software feature matrix. . 91

B.1 6-bus case bus data. 102
B.2 6-bus case generator data. 102
B.3 6-bus case branch data. 103
B.4 IEEE RTS bus data. 104
B.5 IEEE RTS generator data. 105
B.6 IEEE RTS branch data. 106

Chapter 1

Introduction

This thesis examines reinforcement learning algorithms in the domain of electric power trade. In this chapter the motivation for research into electricity trade is explained, the problem under consideration is defined and the principle research contributions are stated.

1.1 Research Motivation

Quality of life for a person is directly proportional to his or her electricity usage (Alam, Bala, Huo, & Matin, 1991). The world population is currently 6.7 billion and forecast to exceed 9 billion by 2050 (United Nations, 2003). Electricity production currently demands over one third of the annual primary energy extracted (The International Energy Agency, 2010) and as people endeavour to improve their quality of life, finite fuel resources will become increasingly scarce. Market mechanisms, such as auctions, where the final allocation is based upon the claimants' willingness to pay for the goods, provide a device for efficient allocation of resources in short supply.

Commercialisation of large electricity supply industries began two decades ago in the UK. 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. Designed correctly, a competitive electricity market promotes efficiency and drives down costs to the consumer, while design errors can allow market power abuse and elevated market prices.

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 is particularly 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 (Minkel, 2008; ICF Consulting, 2003).

The value of electricity to society makes it impractical to experiment with radical changes to trading arrangements on real systems. An alternative is to study abstract mathematical models with sets of simplifying approximations and assumptions and, where possible, to find analytical solutions using digital computer programs. Competition is a fundamental part of all markets, but the strategies of human participants are difficult to model. Reinforcement learning methods can be used to represent adaptive behaviour in competing players and are capable of learning complex strategies (Tesauro, 1994).

1.2 Problem Statement

Individuals participating in an electricity market (be they representing generating companies, load serving entities, firms of traders etc.) must utilise noisy, mostly continuous, multi-dimensional data to their advantage. Certain types of data, e.g. demand forecasts, are uncertain and other types, e.g. the bids of competitors, are hidden. Reinforcement learning algorithms must operate with data of this kind if they are to successfully model participant strategies.

Traditional reinforcement learning methods attempt to find the *value* of each available action in a given state. When discrete state and action spaces are defined, these methods become restricted by Bellman's Curse of Dimensionality (Bellman, 1961) and can not be applied to highly complex problems. When used with function approximation techniques (e.g. artificial neural networks) they can be applied to continuous representations of an environment. However, the greedy updates used by most techniques have been shown to cause algorithms approximating a value function to not converge or even diverge (Tsitsiklis & Roy, 1994; Peters & Schaal, 2008; Gordon, 1995; Baird, 1995).

Policy gradient reinforcement learning methods do not attempt to approximate a value function, but to approximate a *policy-function* that, given the current perceived state of the environment, returns an action. They do not suffer from many of the problems that mar value-function based methods in high-dimensional problems. They have strong convergence properties, 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 do not need to be altered. They have been successfully applied in many operational settings, including: robotic control (Peters & Schaal, 2006), financial trading (Moody & Saffell, 2001) and network routing (Peshkin & Savova, 2002) applications.

It is proposed in this thesis that agents which learn using policy gradient methods may outperform those using value function based methods in simulated competitive electricity trade. It is further proposed that policy gradient methods may operate better under dynamic electric power system conditions, achieving greater profit by exploiting constraints to their financial benefit. This thesis will use electricity market simulation techniques to compare value function based and policy gradient learning methods to explore these proposals.

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 are:

- The first application of policy gradient reinforcement learning methods in simulated electricity trade.
- The first application of a non-linear optimal power flow formulation in agent based electricity market simulation.
- A new Stateful Roth-Erev reinforcement learning method.
- Simulation results comparing the convergence to a Nash equilibrium of policy gradient and value function based reinforcement learning methods.
- Simulation results that examine the exploitation of electric power system constraints by policy gradient reinforcement learning methods.
- An implementation of a power exchange auctions market model and multi-learning-agent system for simulating electricity trade.
- The idea of applying Neuro-Fitted Q-Iteration and GQ(λ) in simulations of competitive energy trade.
- A model of the UK transmission system derived from data in the National Grid Seven Year Statement.

The publications that have resulted from this thesis are: Lincoln, Galloway, and Burt (2009, 2007); Lincoln, Galloway, Burt, and McDonald (2006).

1.4 Thesis Outline

The presentation of this thesis is organised into nine chapters. Chapter 2 provides background information on electricity supply, wholesale electricity markets and reinforcement learning. It describes how optimal power flow formulations can be used to model electricity markets and defines the reinforcement learning algorithms that are later compared.

In Chapter 3 the research in this thesis is described in the context of previous work that is related in terms of application field and methodology. Publications on agent based electricity market simulation are reviewed with emphasis on the reinforcement learning methods used. Previous applications of policy gradient learning methods in other types of market setting are reviewed also.

Chapter 4 describes the power exchange auction market model and the multi-agent system used to simulate electricity trade. It defines the association of learning agents with portfolios of generators, the process of offer submission and the reward process.

Simulations that examine the convergence to a Nash equilibrium of systems of multiple electric power trading agents is reported in Chapter 5. A six bus test case is used and results for four learning algorithms under two cost configurations are presented and analysed.

Chapter 6 examines the ability of agents to learn policies for exploiting constraints in simulated power systems. The 24 bus model from the IEEE Reliability Test System provides a complex environment with dynamic loading conditions.

The primary conclusions drawn from the results in this thesis are summarised in Chapter 7. Shortcomings of the approach are noted and the broader implications are addressed. Some ideas for further work are also outlined, including alternative reinforcement learning methods and uses for a model of the UK transmission system.

Chapter 4

Modelling Power Trade

This chapter defines the model used in chapters 5 and 6 to simulate competitive electric power trade and compare learning algorithms. The first section describes how optimal power flow solutions are used to clear offers submitted to a simulated power exchange auction market. The second section defines how market participants are modelled as agents that use the reinforcement learning algorithms to adjust their bidding behaviour. It explains the modular structure of a multi-agent system that coordinates interactions between the auction model and participant agents.

4.1 Electricity Market Model

A power exchange auction market, based on SmartMarket by Zimmerman (2010, p.92), is used in this thesis as a trading environment for comparing reinforcement learning algorithms. In each trading period the auction accepts offers to sell blocks of power from participating agents[1]. A clearing process begins by withholding offers above the price cap, along with those specifying non-positive quantities. Valid offers for each generator are sorted into non-decreasing order with respect to price and converted into corresponding generator capacities and piecewise linear cost functions (See Section 4.1.1 below). The newly configured units form an optimal power flow problem, the solution to which provides generator set-points and nodal marginal prices that are used to determine the proportion of each offer block that is cleared and the associated clearing price. The cleared offers determine each agent's revenue and hence the profit used as a reward signal.

A nodal marginal pricing scheme is used in which the price of each offer is

[1] A double-sided auction, in which bids to buy blocks of power may be submitted by agents associated with dispatchable loads, has also been implemented, but this feature is not used.

cleared at the value of the Lagrangian multiplier on the power balance constraint for the bus at which the offer's generator is connected. An alternative discriminatory pricing scheme may be used in which offers are cleared at the price at which they were submitted (pay-as-bid). The advanced auction types from MATPOWER that scale nodal marginal prices are not used, but could be used in a detailed study of pricing schemes.

4.1.1 Optimal Power Flow

Bespoke implementations of both the DC and AC optimal power flow formulations from MATPOWER are used in the auction clearing process. The trade-offs between DC and AC formulations have been examined by Overbye, Cheng, and Sun (2004). DC models were found suitable for most nodal marginal price calculations and are considerably less computationally expensive to solve. The AC optimal power flow formulation is used to examine the exploitation of voltage constraints, that are not part of the DC formulation.

As in MATPOWER, generator active power, and optionally reactive power, output costs may be defined by convex n-segment piecewise linear cost functions

$$c^{(i)}(p) = m_i p + b_i \tag{4.1}$$

where p is the generator set-point for $p_i \leq p \leq p_{i+1}$ with $i = 1, 2, \ldots n$, m_i is the variable cost for segment i in \$/MWh where $m_{i+1} \geq m_i$ and $p_{i+1} > p_i$, and b_i is the y-intercept in \$, also for segment i.

Since these cost functions are non-differentiable, the constrained cost variable approach from H. Wang, Murillo-Sanchez, Zimmerman, and Thomas (2007) is used to make the optimisation problem smooth. For each generator j a helper cost variable y_j is added to the vector of optimisation variables. Figure 4.1 (Zimmerman, 2010, Figure5-3) illustrates how the additional inequality constraints

$$y_j \geq m_{j,i}(p - p_i) + c_i, \quad i = 1 \ldots n \tag{4.2}$$

ensure that y_j lies on or above $c^{(i)}(p)$ as the objective function minimises the sum of cost variables for all generators:

$$\min_{\theta, V_m, P_g, Q_g, y} \sum_{j=1}^{n_g} y_j \tag{4.3}$$

The extended optimal power flow formulations from MATPOWER with user-

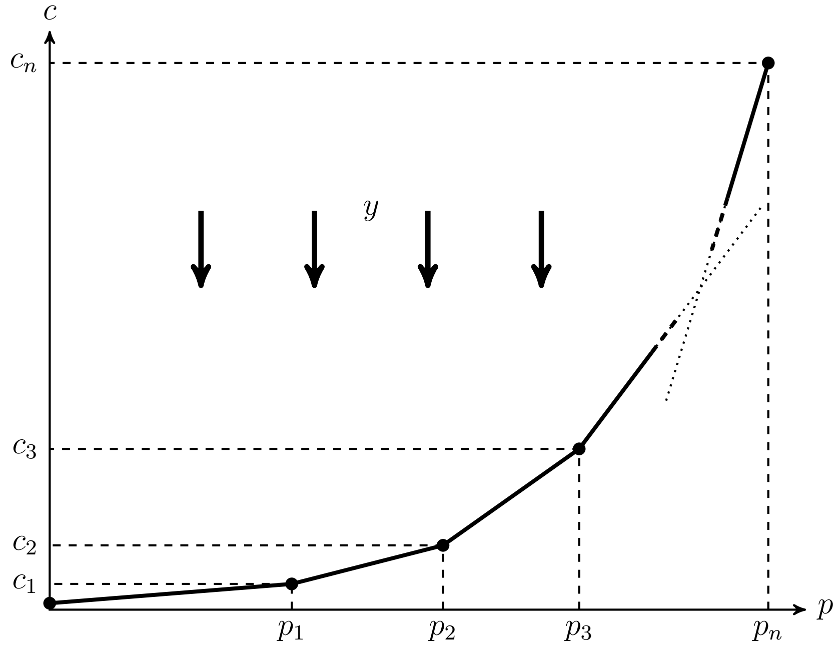

Figure 4.1: Piecewise linear active power cost function with constrained cost variable minimsation illustrated.

defined cost functions and generator P-Q capability curves are not used, but could be applied in further development of this work.

4.1.2 Unit De-commitment

The optimal power flow formulations constrain generator set-points between upper and lower power limits. The output of expensive generators can be reduced to the lower limit, but they can not be completely shutdown. The online status of generators could be added to the vector of optimisation variables, but being Boolean the problems would be mixed-integer non-linear programs which are typically very difficult to solve.

To compute a least cost commitment and dispatch the unit de-commitment algorithm from Zimmerman (2010, p.57) is used. The algorithm involves shutting down the most expensive units until the minimum generation capacity is less than the total load capacity and then solving repeated optimal power flow problems with candidate 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.

4.2 Multi-Agent System

Market participants are modelled using PyBrain software agents that use reinforcement learning algorithms to adjust their behaviour (Schaul et al., 2010). Their interaction with the market is coordinated in multi-agent simulations, the structure of which is derived from PyBrain's single player design.

This section describes: discrete and continuous market *environments*, agent *tasks* and *modules* used for policy function approximation and storing state-action values or action propensities. The process by which each agent's policy is updated by a *learner* is explained and the sequence of interactions between multiple agents and the market is described and diagrammed.

4.2.1 Market Environment

Each agent has a portfolio of n_g generators associated their environment. Figure 4.2 illustrates the association and how the environment references an instance of the auction market for offer submission. Each environment is responsible for (i) returning a vector representation of its current state and (ii) accepting an action vector which transforms the environment into a new state. To facilitate testing of value function based and policy gradient learning methods, both discrete and continuous representations of an electric power trading environment are defined.

Discrete Market Environment

An environment with n_s discrete states and n_a discrete action possibilities is defined for agents operating learning methods that make use of look-up tables. The environment produces a state s, where $s \in \mathbb{Z}^+$ and $0 \leq s < n_s$, at each simulation step and accepts an action a, where $a \in \mathbb{Z}^+$ and $0 \leq a < n_a$.

To keep the size of the state space reasonable, discrete states are derived only from the total system demand $d = \sum P_d$ where P_d is the vector of active power demand at each bus. Informally, the state space is n_s states between the minimum and maximum demand and the current state for the environment is the index of the state to which the current demand relates. Each simulation episode of n_t steps has a demand profile vector U of length n_t, where each element $0 \leq u_i \leq 1$. The load at each bus $P_{dt} = u_t P_{d0}$ in simulation period t, where P_{d0} is the initial demand vector. The state size $d_s = d(\max U - \min U)/n_s$ and the state space vector is $\mathcal{S} = d_s i$ for $i = 1 \ldots n_s$. At simulation step t, the state returned by the environment $s_t = i$ if $\mathcal{S}_i \leq P_{dt} \leq \mathcal{S}_{i+1}$ for $i = 0 \ldots n_s$.

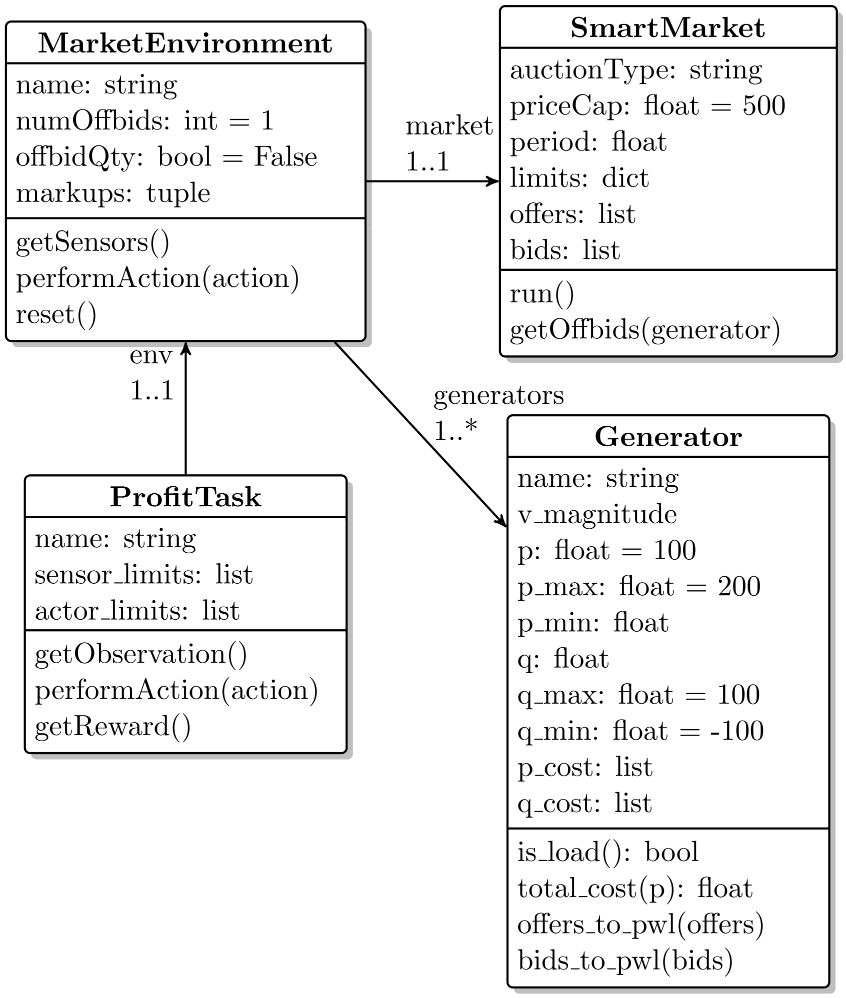

Figure 4.2: Agent environment UML class diagram.

Table 4.1: Example discrete action domain.

a	m_1	w_1	m_2	w_2
0	0	0	0	0
1	0	0	10	0
2	0	0	20	0
3	10	0	0	0
4	10	0	10	0
5	10	0	20	0
6	20	0	0	0
7	20	0	10	0
8	20	0	20	0

The action space for a discrete environment is defined by a vector m, where $0 \leq m_i \leq 100$, of percentage markups on marginal cost with length n_m, a vector w, where $0 \leq w_i \leq 100$, of percentage capacity withholds with length n_w and a scalar number of offers n_o, where $n_o \in \mathbb{Z}^+$, to be submitted for each generator associated with the environment.

A $n_a \times 2n_g n_o$ matrix with all permutations of markup and withhold for each offer that is to be submitted for each generator is computed. As an example, Table 4.1 shows all possible actions when markups are restricted to 0, 10% or 20%, $m = \{0, 10, 20, 30\}$, and 0% of capacity may be withheld, $w = \{0\}$, from two generators, $n_g = 2$, with one offer submitted for each, $n_o = 1$. Each row corresponds to an action and the column values specify the percentage price markup and the percentage of capacity to be withheld for each of the $n_g n_o$ offers. The size of the permutation matrix grows rapidly as n_o, n_g, n_m and n_w increase.

Continuous Market Environment

A continuous market environment that outputs a state vector s, where $s_i \in \mathbb{R}$, and accepts an action vector a, where $a_i \in \mathbb{R}$, is defined for agents operating policy gradient methods. Scalar variables m_u and w_u define the upper limit on the percentage markups on marginal cost and the upper limit on the percentage of capacity that can be withheld, respectively. Again, n_o defines the number of offers to be submitted for each generator associated with the environment.

The state vector can be any set of variables from the power system or market model. For example: bus voltages, branch power flows, generator limit Lagrangian multipliers etc. Each element of the vector provides one input to the neural network used for policy function approximation.

The action vector a has length $2n_gn_o$. Element a_i, where $0 \leq a_i \leq m_u$, corresponds to the percentage price markup and each element a_{i+1}, where $0 \leq a_{i+1} \leq w_u$, to the percentage of capacity to be withheld for the $(i/2)^{th}$ offer, where $i = 0, 2, 4, \ldots, 2n_gn_o$.

Not having to discretize the state space and compute a matrix of action permutations greatly simplifies the implementation of a continuous environment and increases in n_g and n_o only impact the number of output nodes on the neural network.

4.2.2 Agent Task

To allow alternative goals (such a profit maximisation or the meeting some target level for plant utilisation) to be associated with a single type of environment, an agent does not interact *directly* with its environment, but is paired with a particular *task*. A task defines the reward returned to the agent and thus defines the agent's purpose.

For all simulations in this thesis the goal of each agent is to maximise direct financial profit. Rewards are defined as the sum of earnings from the previous period t as determined by the difference between the revenue from cleared offers and the generator marginal cost at its total cleared quantity. Using some measure of risk adjusted return (as in (Moody & Saffell, 2001)) might be of interest in the context of simulated electricity trade and this would simply involve the definition of a new task and would not require any modification of the environment.

Agents with policy-gradient learning methods approximate their policy functions using artificial neural networks that are presented with an input vector s_n of length n_s where $s_{n,i} \in \mathbb{R}$. To condition the environment state before input to the connectionist system, where possible, a vector s_l of lower sensor limits and a vector s_u of upper sensor limits is defined. These are used to calculated a normalised state vector

$$v = 2\left(\frac{s - s_l}{s_u - s_l}\right) - 1 \tag{4.4}$$

where $-1 \leq s_{n,i} \leq 1$.

The output from the policy function approximator y is denormalized using vectors of minimum and maximum action limits, a_{min} and a_{max} respectively, to give an action vector

$$a = \left(\frac{y+1}{2}\right)(a_u - a_l) + a_l \tag{4.5}$$

where $0 \leq a_i \leq m_u$ and $0 \leq a_{i+1} \leq w_u$ for $i = 0, 2, 4, \ldots, 2n_gn_o$.

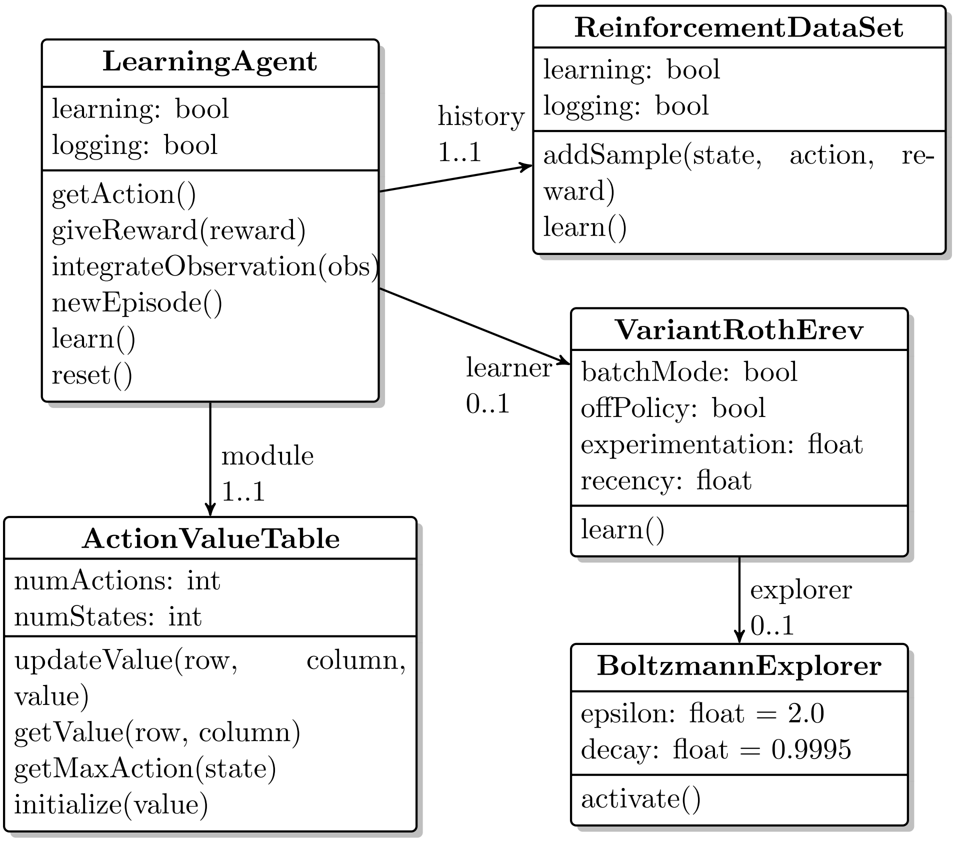

Figure 4.3: Learning agent UML class diagram.

4.2.3 Market Participant Agent

Each agent is defined as an entity capable of producing an action a based on a previous observation s of its environment. The UML class diagram in Figure 4.3 illustrates how each agent in PyBrain is associated with a *module*, a *learner* (modified Roth-Erev in the diagram), a *dataset* and an *explorer*.

The module is used to determine the agent's policy for action selection and returns an action vector a when activated with a state vector. When using value function based methods the module is a $n_s \times n_a$ table of the form

$$\begin{array}{c} \\ s_0 \\ s_1 \\ \\ s_{n_s} \end{array} \begin{array}{c} \begin{array}{cccc} a_0 & a_1 & \quad & a_{n_a} \end{array} \\ \begin{bmatrix} v_{0,0} & v_{0,1} & \cdots & v_{0,m} \\ v_{1,0} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ v_{n,0} & \cdots & \cdots & v_{n_s,n_a} \end{bmatrix} \end{array} \tag{4.6}$$

where each element $v_{i,j}$ is the value in state i associated with selecting action j. When using a policy gradient method, the module is a multi-layer feed-forward artificial neural network that outputs a vector a when presented with observation s_n.

The learner can be any reinforcement learning algorithm that modifies the values/propensities/parameters of the module to increase expected future reward. The dataset stores state-action-reward triples for each interaction between the agent and its environment. The stored history is used by a learners when computing updates to the module.

Each learner has an association with an explorer that returns an explorative action a_e when activated with action a from the module. Softmax and ϵ-greedy explorers are implemented for discrete action spaces. Policy gradient methods use a module that adds Gaussian noise to a_m. The explorer has a parameter σ that relates to the standard deviation of the normal distribution. The actual standard deviation

$$\sigma_e = \begin{cases} \ln(\sigma + 1) + 1 & \text{if } \sigma \geq 0 \\ \exp(\sigma) & \text{if } \sigma < 0 \end{cases} \tag{4.7}$$

to prevent negative σ values from causing an error if automatically adjusted during back-propagation.

4.2.4 Simulation Event Sequence

Each simulation consists of one or more task-agent pairs. Figure 4.4 shows the class associations for a simulation experiment. At the beginning of each simulation step (trading period) t the market is initialised and all previous offers are removed. Figure 4.5 is a UML sequence diagram that illustrates the process of choosing and performing an action that follows. For each task-agent tuple an observation s_t is retrieved from the task and integrated into the agent. When an action is requested from the agent its module is activated with s_t and the action $a_{e,t}$ is returned. Action $a_{e,t}$ is performed on the environment associated with the agent's task.

When all actions have been performed the offers are cleared by the market using the solution to a newly formed optimal power flow problem. Figure 4.6 illustrates the subsequent reward process. The cleared offers associated with the generators in the task's environment are retrieved from the market and the reward r_t is computed from the difference between revenue and marginal cost at the total cleared quantity. The reward r_t is given to the associated agent and the value is stored, along with the previous state s_t and selected action $a_{e,t}$, under a new sample is the dataset.

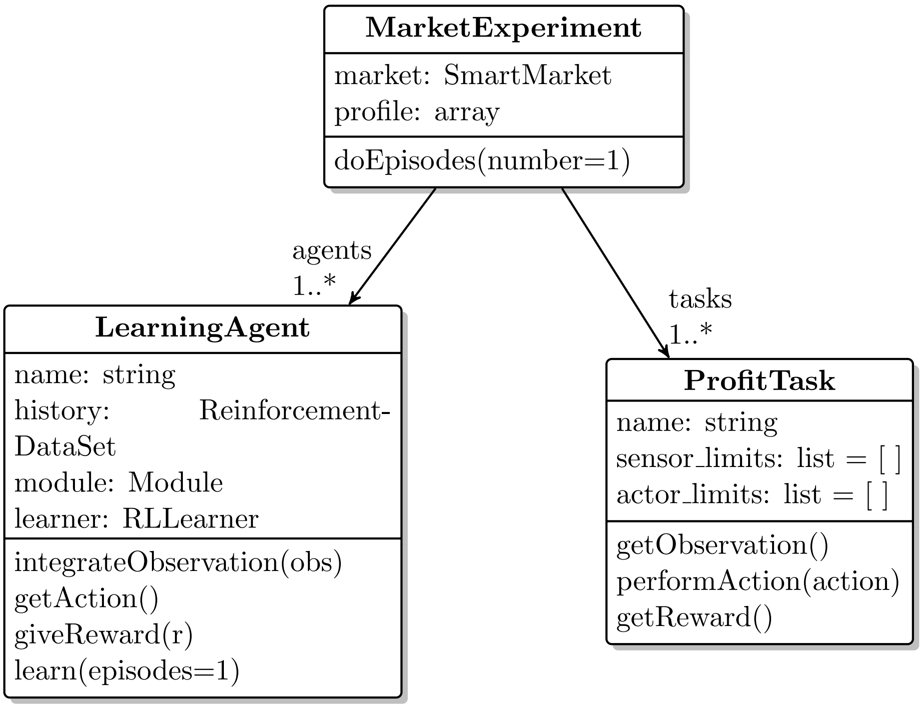

Figure 4.4: Market experiment UML class diagram.

The learning process is illustrated by the UML sequence diagram in Figure 4.7. Each agent learns from its actions using r_t, at which point the values or parameters of the module associated with the agent are updated according to the output of the learner's algorithm. Each agent is then reset and the history of states, actions and rewards is cleared.

The combination of an action, reward and learning process for each agent constitutes one step of the simulation and the processes are repeated until a specified number of steps are complete.

4.3 Summary

The power exchange auction market model defined in this chapter provides a layer of abstraction over the underlying optimal power flow problem and presents agents with a simple interface for selling power. The modular nature of the simulation framework described allows the type of learning algorithm, policy function approximator, exploration technique or task to be easily changed. The framework can simulate competitive electric power trade using almost any conventional bus-branch power system model with little configuration, but provides the facilities for adjusting all of the main aspects of a simulation. The framework's modularity

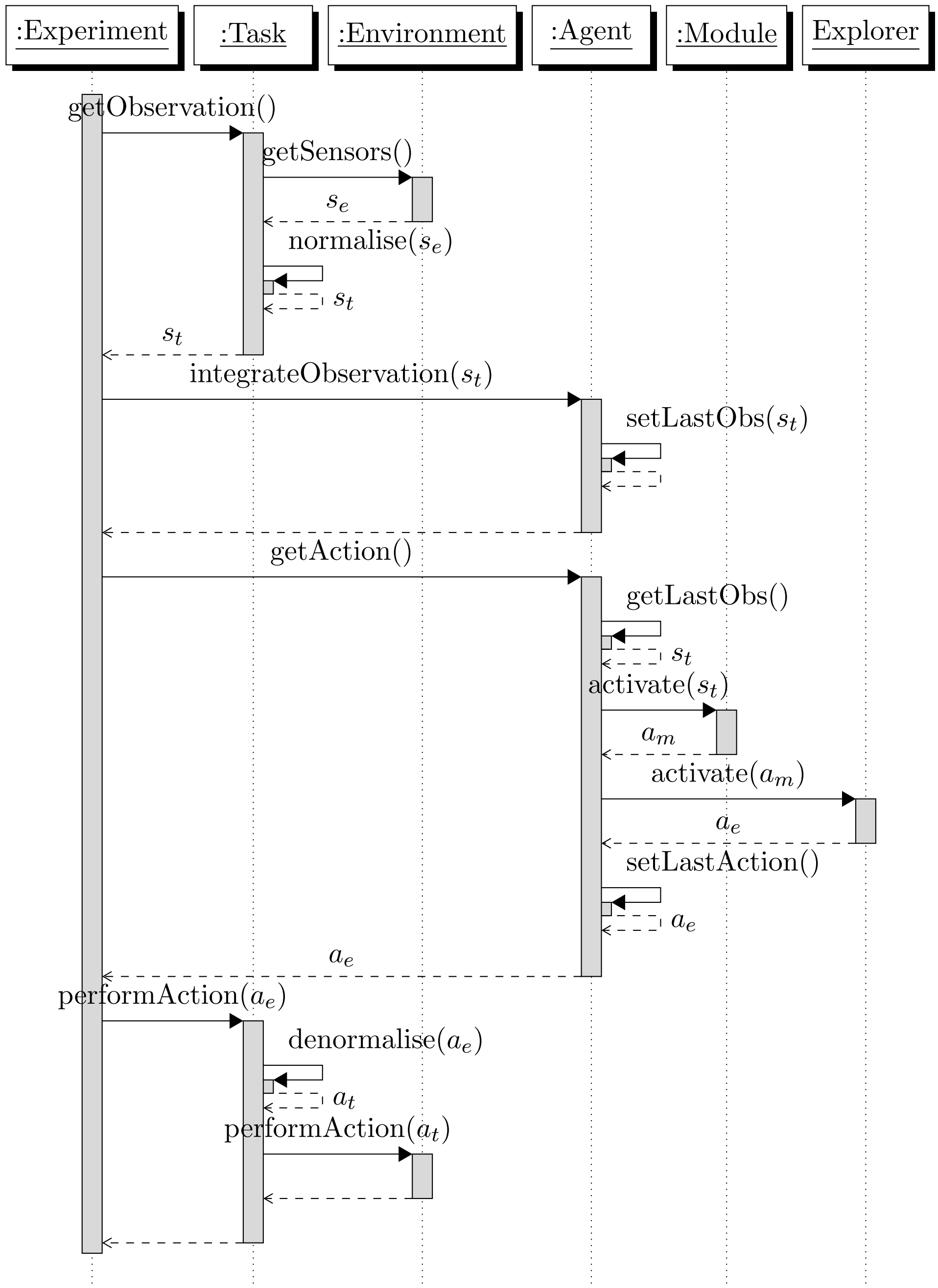

Figure 4.5: Sequence diagram for action selection process.

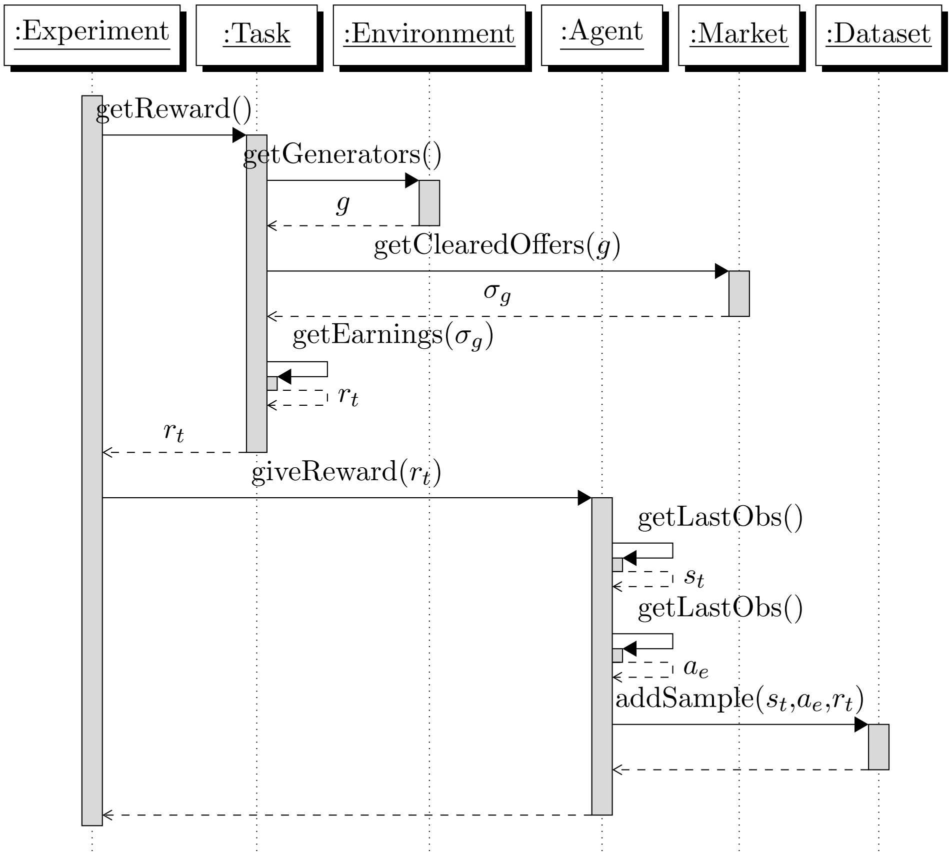

Figure 4.6: Sequence diagram for the reward process.

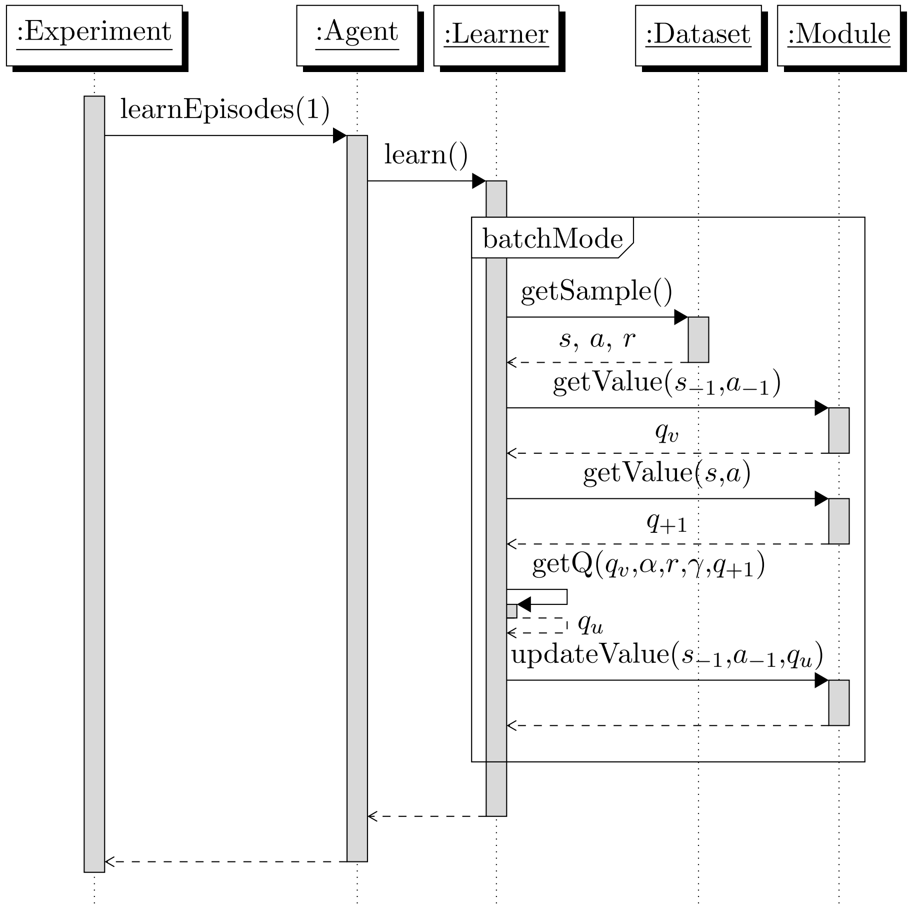

Figure 4.7: Sequence diagram for the SARSA learning process.

and support for easy configuration is intended to allow transparent comparison of learning methods under a wide variety of different scenarios.

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.

Aleksandrov, V., Sysoyev, V., & Shemeneva, V. (1968). Stochastic optimization. Engineering Cybernetics, 5, 11-16.

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

Application of Probability Methods Subcommittee. (1979, November). IEEE reliability test system. Power Apparatus and Systems, IEEE Transactions on, PAS-98(6), 2047-2054.

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

Baird, L. (1995). Residual algorithms: Reinforcement learning with function approximation. In Proceedings of the Twelfth International Conference on Machine Learning (p. 30-37). Morgan Kaufmann.

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

Bertsekas, D. P., & Tsitsiklis, J. N. (1996). Neuro-dynamic programming. Belmont, MA: Athena Scientific.

Bhatnagar, S., Sutton, R. S., Ghavamzadeh, M., & Lee, M. (2009). Natural actor-critic algorithms. Automatica, 45(11), 2471–2482.

Bishop, C. M. (1996). Neural networks for pattern recognition (1st ed.). Oxford University Press, USA. Paperback.

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.

Boyd, S., & Vandenberghe, L. (2004). Convex optimization. Cambridge University Press. Hardcover.

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.

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

Cole, S. (2010, February 4). MatDyn [Computer software manual]. Katholieke Universiteit Leuven.

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).

Fausett, L. (Ed.). (1994). Fundamentals of neural networks: architectures, algorithms, and applications. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

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.

Glynn, P. W. (1987). Likelilood ratio gradient estimation: an overview. In Wsc '87: Proceedings of the 19th conference on winter simulation (p. 366-375). New York, NY, USA: ACM.

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

Gordon, G. (1995). Stable function approximation in dynamic programming. In

Proceedings of the Twelfth International Conference on Machine Learning (p. 261-268). Morgan Kaufmann.

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.

Kallrath, J., Pardalos, P., Rebennack, S., & Scheidt, M. (2009). Optimization in the energy industry. Springer.

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 (p. 121-141). Karlsruhe, Germany.

Kietzmann, T. C., & Riedmiller, M. (2009). The neuro slot car racer: Reinforcement learning in a real world setting. Machine Learning and Applications, Fourth International Conference on, 0, 311-316.

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 (p. 1-11).

Lincoln, R., Galloway, S., & Burt, G. (2007, May 23-25). Unit commitment and system stability under increased penetration of distributed generation. In Proceedings of the 4th International Conference on the European Energy Market, 2007. EEM 2007. Cracow, Poland.

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. (p. 1-5).

Lincoln, R., Galloway, S., Burt, G., & McDonald, J. (2006, 6-8). Agent-based simulation of short-term energy markets for highly distributed power systems. In Proceedings of the 41st international universities power engineering conference, 2006. UPEC '06. (Vol. 1, p. 198-202).

Maei, H. R., & Sutton, R. S. (2010). Gq(λ): A general gradient algorithm for temporal-difference prediction learning with eligibility traces. In In proceedings of the third conference on artificial general intelligence. Lugano, Switzerland.

McCulloch, W., & Pitts, W. (1943, December 21). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biology, 5(4), 115-133.

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.

Momoh, J., Adapa, R., & El-Hawary, M. (1999, Feb). A review of selected optimal power flow literature to 1993. I. Nonlinear and quadratic programming approaches. Power Systems, IEEE Transactions on, 14(1), 96-104.

Momoh, J., El-Hawary, M., & Adapa, R. (1999, Feb). A review of selected optimal power flow literature to 1993. II. Newton, linear programming and interior point methods. Power Systems, IEEE Transactions on, 14(1), 105-111.

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-Tootoonchi, 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.

Nash, J. F. (1950, January). Equilibrium points in n-person games. Proceedings of the National Academy of Sciences of the United States of America, 36(1), 48-49.

Nash, J. F. (1951, September). Non-cooperative games. The Annals of Mathematics, 54(2), 286-295. Available from http://dx.doi.org/10.2307/1969529

National Electricity Transmission System Operator. (2007, September). Large combustion plant directive (Tech. Rep.). National Grid Electricity Transmission plc. (GCRP 07/32)

National Electricity Transmission System Operator. (2010, May). 2010 National Electricity Transmission System Seven Year Statement (Tech. Rep.). National Grid Electricity Transmission plc.

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, p. 1041-1047).

Overbye, T., Cheng, X., & Sun, Y. (2004, Jan.). A comparison of the AC and DC power flow models for LMP calculations. In System sciences, 2004. Proceedings of the 37th annual Hawaii international conference on (p. 9-).

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. (2010). Policy gradient methods. (Available online: www.scholarpedia.org)

Peters, J., & Schaal, S. (2006, October). Policy gradient methods for robotics. In

Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on (p. 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 (p. 1-7).

Riedmiller, M. (2005). Neural fitted Q iteration - first experiences with a data efficient neural reinforcement learning method. In In 16th European conference on machine learning (pp. 317–328). Springer.

Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: the rprop algorithm.

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.

Schaul, T., Bayer, J., Wierstra, D., Sun, Y., Felder, M., Sehnke, F., et al. (2010). PyBrain. Journal of Machine Learning Research, 11, 743-746.

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 (p. 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, p. 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.

Tesauro, G. (1994). TD-Gammon, a self-teaching backgammon program, achieves master-level play. Neural Computation, 6(2), 215-219.

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.

The International Energy Agency. (2010, Septemeber). Key world energy statistics 2010. Paris.

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

Tsitsiklis, J. N., & Roy, B. V. (1994). Feature-based methods for large scale dynamic programming. In Machine learning (p. 59-94).

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

U.S.-Canada Power System Outage Task Force. (2004, April). Final report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations (Tech. Rep.). North American Electric Reliability Corporation.

Veit, D., Weidlich, A., Yao, J., & Oren, S. (2006). Simulating the dynamics in two-settlement electricity markets via an agent-based approach. International Journal of Management Science and Engineering Management, 1(2), 83-97.

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, p. 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 (p. 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.

Weidlich, A., & Veit, D. (2008, July). A critical survey of agent-based wholesale electricity market models. Energy Economics, 30(4), 1728-1759.

WG 31.04. (1983). Electric power transmission at voltages of 1000 kV and above plans for future AC and DC transmission. Electra. (ELT_091_3)

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. In Machine learning (p. 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. (2010, March 19). MATPOWER 4.0b2 User's Manual [Computer software manual]. School of Electrical Engineering, Cornell University, Ithaca, NY 14853.

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.