

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.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date: August 16, 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 methods can provide superior participant models than 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 that are solved with a primal-dual interior point method. Policy gradient reinforcement learning algorithms are used to adjust the parameters of multi-layer feed-forward neural networks that approximate each market participant's policy for selecting power quantities and prices that are offered in a 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, such as in robotic control applications, 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	6
2.1 Electric Power Supply	6
2.2 Electricity Markets	8
2.2.1 The England and Wales Electricity Pool	10
2.2.2 British Electricity Transmission and Trading Arrangements	12
2.3 Electricity Market Simulation	13
2.3.1 Agent-Based Simulation	14
2.3.2 Optimal Power Flow	14
2.4 Reinforcement Learning	20
2.4.1 Value Function Methods	21
2.4.2 Policy Gradient Methods	24
2.4.3 Roth-Erev Method	26
2.5 Summary	28
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	39

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	45
3.4.1	Financial Decision Making	45
3.4.2	Grid Computing	46
3.5	Summary	47
4	Modelling Power Trade	49
4.1	Electricity Market Model	49
4.1.1	Optimal Power Flow	49
4.1.2	Unit De-commitment	50
4.1.3	Power Exchange	51
4.2	Multi-Agent System	53
4.2.1	Environment	53
4.2.2	Task	55
4.2.3	Agent	56
4.2.4	Simulation Event Sequence	56
4.3	Summary	57
5	Nash Equilibrium Analysis	64
5.1	Introduction	64
5.2	Aims and Objectives	65
5.3	Method of Simulation	65
5.4	Simulation Results	67
5.5	Discussion and Critical Analysis	75
5.6	Summary	76
6	System Constraint Exploitation	77
6.1	Introduction	77
6.2	Aims and Objectives	77
6.3	Method of Simulation	78
6.4	Simulation Results	82
6.5	Discussion and Critical Analysis	82
6.6	Summary	82
7	Conclusions and Further Work	83
7.1	Further Work	83
7.1.1	Alternative Learning Algorithms	83
7.1.2	UK Transmission System	84
7.1.3	AC Optimal Power Flow	86
7.1.4	Multi-Market Simulation	86
7.2	Summary Conclusions	87
	Bibliography	88

A	Open Source Power Engineering Software	96
A.1	MATPOWER	96
A.2	MATDYN	99
A.3	Power System Analysis Toolbox	99
A.4	UWPFLOW	101
A.5	TEFTS	101
A.6	Distribution System Simulator	102
A.7	Agent-based Modelling of Electricity Systems	103
A.8	DCOPFJ	104
A.9	PYLON	104
B	Case Data	106
B.1	6-Bus Case	106
B.2	IEEE Reliability Test System	106

List of Figures

2.1	Basic structure of a three phase AC power system.	7
2.2	UK power station locations.	9
2.3	Pool bid structure.	11
2.4	Piecewise linear active power cost function with constrained cost variable minimisation illustrated.	11
2.5	Nominal- π transmission line model in series with a phase shifting transformer model.	16
2.6	Sequence diagram for the basic reinforcement learning model. . .	21
2.7	Multi-layer feed-forward perceptron with bias nodes.	25
3.1	Single-line diagram for a stylised Italian grid model.	41
5.1	Single-line diagram for six bus power system model.	66
5.2	Average markup for agent 1 and standard deviation over 10 runs.	69
5.3	Average markup for agent 2 and standard deviation over 10 runs.	70
5.4	Average markup for agent 1 and standard deviation.	71
5.5	Average markup for agent 2 and standard deviation.	72
5.6	Average reward for agent 1 and standard deviation.	73
5.7	Average reward for agent 2 and standard deviation.	74
6.1	Generator cost functions for the IEEE Reliability Test System . .	79
6.2	Hourly, daily and weekly load profile plots from the IEEE Relia- bility Test System	80
6.3	IEEE Reliability Test System	81
7.1	UK transmission system.	85
A.1	UKGDS EHV3 model in PSAT Simulink network editor.	100
B.1	Single-line diagram for six bus power system model.	107

List of Tables

4.1	Example discrete action domain.	54
5.1	Generator cost configuration 1 for 6-bus case.	66
5.2	Generator cost configuration 2 for 6-bus case.	66
5.3	Agent rewards under cost configuration 1	68
5.4	Agent rewards under cost configuration 2	68
6.1	Cost parameters IEEE RTS generator types.	78
6.2	Agent portfolios by type and location.	82
A.1	Open source electric power engineering software feature matrix. .	97
B.1	6-bus case bus data.	106
B.2	6-bus case generator data.	107
B.3	6-bus case branch data.	108
B.4	IEEE RTS bus data.	108
B.5	IEEE RTS generator data.	109
B.6	IEEE RTS branch data.	110
B.7	IEEE RTS generator cost data.	111

Chapter 4

Modelling Power Trade

The present chapter defines the model used in this thesis to simulate electric power trade. The first section describes how optimal power flow solutions are used to clear offers and bids submitted to a simulated power exchange auction. The second section defines how market participants are modelled as agents that use 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 market participants.

4.1 Electricity Market Model

A model of a power exchange auction market is used in this thesis to compare reinforcement learning algorithms. It accepts offers of and bids for blocks of power at specified prices. The auction withholds invalid offers and bids and determines the cleared quantities and prices.

4.1.1 Optimal Power Flow

Bespoke implementations of the optimal power flow formulations from MATPOWER are used in the clearing process (Zimmerman, 2010, §5). Both the DC and AC formulations are used! The trade-offs between DC and AC models 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. The AC optimal power flow formulation is used in this thesis to examine the exploitation of voltage constraints that are not part of a DC formulation. A class diagram in the Unified Modelling Language (UML) for the object-orientated power system model that is used to compute optimal power

flow solutions is shown in Figure ??.

As in MATPOWER (Zimmerman, 2010, p.26), generator active power, and optionally reactive power, output costs may be defined by convex n -segment piecewise linear cost functions

$$c^{(i)}(x) = m_i p + c_i \quad (4.1)$$

for $p_i \leq p \leq p_{i+1}$ with $i = 1, 2, \dots, n$ where $m_{i+1} \geq m_i$ and $p_{i+1} > p_i$, as diagramed in Figure 2.4 (Zimmerman, 2010, Figure5-3). Since these costs are non-differentiable, the constrained cost variable approach from (H. Wang, Murillo-Sanchez, Zimmerman, & Thomas, 2007) is used to make the optimisation problem smooth. For each generator i a helper cost variable y_i added to the vector of optimisation variables. The additional inequality constraints

$$y_i \geq m_{i,j}(p - p_j) + c_j, \quad j = 1 \dots n \quad (4.2)$$

ensure that y_i lies on or above $c^{(i)}(x)$. The objective of the optimal power flow problem becomes the minimisation of the sum of cost variables for all generators:

$$\min_{\theta, V_m, P_g, Q_g, y} \sum_{i=1}^{n_g} y_i \quad (4.3)$$

The extensions to the optimal power flow formulations defined in MATPOWER for user-defined cost functions and generator P-Q capability curves are not utilised.

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 incorporated into the vector of optimisation variables, but as they are Boolean the problems would become 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. Algorithm 1 shows how this 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.

Algorithm 1 Unit de-commitment

```

1: while  $\sum P_g^{min} > \sum P_d$  do
2:   shutdown most expensive unit
3: end while
4:  $f \leftarrow$  initial total system cost
5: repeat
6:    $c \leftarrow$  generators at  $P_{min}$ 
7:   for  $g$  in  $c$  do
8:      $d \leftarrow$  true
9:     shutdown  $g$ 
10:     $f' \leftarrow$  new total system cost
11:    if  $f' < f$  then
12:       $f \leftarrow f'$ 
13:       $g_c \leftarrow g$ 
14:       $d \leftarrow$  false
15:    end if
16:    startup  $g$ 
17:  end for
18:  shutdown  $g_c$ 
19: until  $d = \text{true}$ 

```

4.1.3 Power Exchange

To simulate electric power trade a model is used in which agents representing market participants do not provide cost functions for the generators in their portfolio, but submit offers to sell and/or bids to buy blocks of active or reactive power. The offers/bids are submitted to a power exchange auction market model based on SmartMarket from Zimmerman (2010, p.92).

The clearing process begins by withholding offers/bids outwith maximum offer and minimum bid price limits, along with those specifying non-positive quantities. Valid offers/bids for each generator are then sorted into non-decreasing/non-increasing order and are converted into corresponding generator/dispatchable load capacities and piecewise linear cost functions. The newly configured units are used in a unit de-commitment optimal power flow problem, the solution of which holds generator set-points and nodal marginal prices which are used to determine the proportion of each offer/bid block that should be cleared and the cleared price for each.

A basic nodal marginal pricing scheme is used in which the price of each offer/bid is cleared at the value of the Lagrangian multiplier on the power balance

constraint for the bus at which the associated generator is connected. Alternatively, a discriminatory pricing scheme may be used in which offer/bids are cleared at the price at which they were submitted (pay-as-bid). Cleared offers/bids are returned to the agents and used to determine revenue values from which each agent's earnings or losses are derived.

4.2 Multi-Agent System

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

This section describes the environment of each agent, their tasks and the modules used for policy function approximation and storing state-action values in tables. The process by which each agent’s policy is updated by a learning algorithm is explained and the sequence of interactions between multiple agents and the market is described and illustrated.

4.2.1 Environment

In each experiment, agents are endowed with a portfolio of generators from the electric power system model (See Figure ??). As illustrated by the UML class diagram in Figure ??, the generators are contained within each agent’s *environment*. The environment also holds an association to an instance of the auction market that allows the submission of offers/bids. 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 Environment

For operation with learning methods that use look-up tables to store state-action values, an environment with n_s discrete states and n_a discrete actions is defined. An agent can not observe offers/bids submitted by competitor agents, but may sense other aspects of the power system model. To ensure that the size of the environment state space is kept reasonable, the agent is limited to observing a demand forecast. The initial demand at each bus P_{d0} , as defined in the original power system model, is assumed to be peak and the demand at each bus can follow a profile at each step t of the simulation (See Chapter 6). The state space is divided into discrete steps of size $P_{step} = (P_{d0} - P_d^{min})/n_s$, where P_d^{min} is the total demand at the lowest point of the profile. The environment computes the total system demand P_{dt} at step t and returns an integer representation of the

a_i	m_1	m_2
1	0	0
2	0	10
3	0	20
4	10	0
5	10	10
6	10	20
7	20	0
8	20	10
9	20	20

Table 4.1: Example discrete action domain.

state

$$s_t = \frac{(P_{dt} - P_d^{min})}{P_{step}} + 1. \quad (4.4)$$

To define the action space, a vector of percentage markups on marginal cost m_e is defined for each environment e along with a variable $n_o \in \mathbb{Z}^+$ which denotes the number of offers/bids to be submitted by the agent. (A similar vector of percentage markdowns on total capacity has also been implemented, but is not used in this thesis.) A set of all unique permutations of markup for n_o offers/bids of length n_a is formed, from which the agent must select. The action vector that the discrete environment is passed consists of a single integer value, corresponding to the column index in the agent’s action value table. The quantity and price for each offer/bid submitted to the market is taken from the vector of permutations using the a_t as the index. An example of the possible permutations of 0, 10 and 20% markups for a portfolio of two generators is given in Table 4.1. It should be clear how quickly the number of possible actions can grow as the number of possible markups and the size of the portfolio increases.

Continuous Environment

For agents operating policy gradient methods, continuous environments that output n_s sensors and accept n_a actions are defined. Each environment may be configured for actions that specify just price or price and quantity. If $q_e^i = 0$ where $q_e^i \in (0, 1)$ then the agent’s action is price selection and the offer/bid quantity is determined by the maximum rated capacity of the generator in question divided by the number of offers being submitted for it. The environment accepts an action vector a_e of length n_a if $q_e^i = 0$, otherwise of length $2n_a$. If $q_e^i = 0$, the i -th element of a_e is the offered/bid price in \$/MWh, where $i = 1, 2, \dots, n_{in}$. If $q_e^i = 1$, the i -th

element of a_e is the offered/bid price in \$/MWh, where $i = 1, 3, 5, \dots, n_{in} - 1$ and the j -th element of a_e is the offered/bid quantity in MW where $j = 2, 4, 6, \dots, n_{in}$. The action vector passed to the environment is converted into sets of offers/bids that are submitted to the market model.

4.2.2 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 experiments in this thesis the goal of each agent is to maximise financial profit and the rewards are thus defined as the sum of earnings from the previous period t as determined by the revenue from the market and any incurred costs. As explained in Section 3.4.1, utilising some measure of risk adjusted return might be of interest in the context of simulated electricity trade and this would simply involve the definition of a new task without any need for modification of the environment.

Sensor data from the environment is filtered according to the task being performed. Agents with value-function learning methods use a table to store state-action values, with one row per environment state. Thus, observations consist of a single value s_v , where $s_v \leq n_s$ and $s_v \in \mathbb{Z}^+$.

Agents with policy-gradient learning methods approximate their policy functions using artificial neural networks that are presented with input vector w of length n_s where $w_i \in \mathbb{R}$. To condition the environment state before input to the connectionist system, where possible, each sensor i in the state vector s is associated with a minimum value $s_{i,min}$ and a maximum value $s_{i,max}$. The state vector is normalised to:

$$w = 2 \left(\frac{s - s_{min}}{s_{max} - s_{min}} \right) - 1 \quad (4.5)$$

such that $-1 \leq w_i \leq 1$.

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

$$a = \left(\frac{y + 1}{2} \right) (a_{max} - a_{min}) + a_{min} \quad (4.6)$$

with valid values for price, and optionally quantity.

4.2.3 Agent

Each 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 length n_a and n_s respectively, where n_s is the total number of states and n_a is the total number of actions. In PyBrain each agent is associated with a *module*, a *learner*, a *dataset* and an *explorer*. The UML class diagram in Figure ?? illustrates the associations.

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

$$\begin{matrix} & a_0 & a_1 & & a_n \\ \begin{matrix} s_0 \\ s_1 \\ \vdots \\ s_n \end{matrix} & \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,m} \\ v_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ v_{n,1} & \cdots & \cdots & v_{n,m} \end{bmatrix} \end{matrix} \quad (4.7)$$

When using a policy gradient method, the module is a multi-layer feed-forward artificial neural network.

The learner can be any reinforcement learning algorithm that modifies the values/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 value-function learners when computing updates to the table values. Policy gradient learners search directly in the space of the policy network parameters.

Each learner has an association with an explorer that returns an explorative action a_e when activated with the current state s_t and action a_m from the module.

4.2.4 Simulation Event Sequence

Each experiment consists one or more agent-task pairs. At the beginning of each simulation step (trading period) the market is initialised and all existing offers/bids are removed. From 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 is returned. Action a_e is performed on the environment associated with the agent's task. Figure ?? provides a UML sequence diagram that illustrates the process of performing an action and Figure ?? shows the class associations for an experiment.

When all actions have been performed the offers/bids are cleared by the market using the solution of an optimal power flow problem. Each task returns a reward r_t . The cleared offers/bids associated with the generators in the task's environment are retrieved from the market and r_t is computed from the difference between revenue and cost values. 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 , under a new sample in the dataset. The reward process is illustrated by the UML sequence diagram in Figure ??.

Each agent learns from its actions using r_t , at which point the values/parameters of the module associated with the agent is 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 learning process is illustrated by the UML sequence diagram in Figure ??.

All of this constitutes one step of the simulation and the process is repeated until a set 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 and buying power. The modular nature of the simulation framework described allows the type of learning algorithm, policy function approximator, exploration technique or the task to be easily changed. The framework can simulate competitive electric power trade using any conventional bus-branch power system model, requiring little configuration, but provides the facility to adjust all of the simulation's main aspects. The modular framework and its support for easy configuration is intended to allow transparent comparison of learning methods in the domain of electricity trade under a number 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.
- 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.
- 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.
- 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.
- 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.
- 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). $G_q(\lambda)$: 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 portfolios. Journal of Forecasting, 17, 441-470.
- Naghbi-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.
- 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., & 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., Guerri, 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, November). 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.
- Tinney, W., & Hart, C. (1967, November). 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.
- 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). MAT-

POWER's extensible optimal power flow architecture. In IEEE PES General Meeting. Calgary, Alberta, Canada.