Learning Optimal Strategies in a Stochastic Game with Partial Information Applied to Power Markets

N. Chrysanthopoulos, G. P. Papavassilopoulos

School of Electrical & Computer Engineering
National Technical University of Athens

10th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion
6-9 November 2016, Belgrade, Serbia
Table of Contents

1 Introduction
   - Overview
   - Influential Literature

2 Model Formulation
   - Market Structure
   - Incomplete Information
   - Reinforcement Learning

3 Simulations
   - Overview
   - Demand & Production Side
   - Results

4 Conclusion
   - Conclusion
   - Relevant work
What we do?

We use an agent based simulation model to replicate the market outcome for two different informational concepts under two different market formations w.r.t. ownership.

Key points:
- Market modeled as a Stochastic Game
- State Space Transformation technique used
- Players adopt Reinforcement Learning
- Comparative Study of the different cases
Contributions

Main points:

- **State space transformation technique**
  - Adapted to concepts of incomplete information
  - Incorporate processed information

- **R-Learning algorithm**
  - Temporal difference (TD) control method
  - Off-policy generalized policy iterations (GPI) method

- **Comparative study**
  - Two informational Concepts
    (Players consider a simple or an extended information set)
  - Two different cases of ownership
    (3 identical firms Vs a small and a large one)
**Basic Elements — Market Structure**

**Day Ahead (DA) market**

Most power markets rely on a central **day-ahead auction** in which generators submit individual supply curves and the system operator uses these to determine the market price.

The Independent System Operator (ISO) is responsible for its operation and performs the following:

- Informs Power Producers of next day’s demand
- Collects bidding schedules of all participating Power Producers
- Performs the market clearance for each hour
- Determines Power Producers’ payments
Optimal Power Flow

Centralized determination of the production levels that minimize the total cost of production to meet the given load, respecting the network’s physical constraints.

Auctions:
- Single-side
- Uniform
- LAO or FRO
- Marginal price
Locational Marginal Price (LMP)

The locational marginal price is the marginal surplus of an extra megawatt of generation needed to serve the unit increase of the demand at that bus, given all the physical constraints.

Market Power

Market power is the ability to profitably alter prices away from competitive levels.

- Ask higher price than marginal cost
- Withhold output that could be produced
Learning Optimal Strategies in a Partial Information Environment

N. Chrysanthopoulos, G. P. Papavassilopoulos

Literature

Relevant Papers:

i Skoulidas, Vournas, Papavassilopoulos: "An adaptive learning game model for interacting electric power markets"
   ▶ Effects of interconnection’s capacity to coupled markets

ii Tellidou, Bakirtzis: "Multi-agent reinforcement learning for strategic bidding in power markets"
   ▶ Examine some variations of a sample network with constrains

iii Bach, Yao, Wang, Shengjie: "Research and application of the Q-learning for wholesale power markets"
   ▶ Study three cases which differ at the adopted learning technique

iv Ragupathi, Das: "A stochastic game approach for modeling wholesale energy bidding in deregulated power markets"
   ▶ Analyze the impact of constraints to producers’ financial results
Table of Contents

1 Introduction
   - Overview
   - Influential Literature

2 Model Formulation
   - Market Structure
   - Incomplete Information
   - Reinforcement Learning

3 Simulations
   - Overview
   - Demand & Production Side
   - Results

4 Conclusion
   - Conclusion
   - Relevant work
Market Structure

We consider:

- \( N \) individual production units, the players, \( \mathcal{N} = \{1, \ldots, N\} \)
- \( M \) available action functions, forming \( \mathcal{A} = \{a_1, \ldots, a_M\} \)
- \( K \) nodes, the transmission network’s buses, \( \mathcal{K} = \{1, \ldots, K\} \)

For player \( n \in \mathcal{N} \), action \( a^h \in \mathcal{A} \) and bus \( k \in \mathcal{K} \) we have:

**Actions**

\[ \alpha_n = [\alpha^1_n, \ldots, \alpha^{24}_n] \]
- daily bidding vector
- player’s choice variables

**State**

\[ x_k = (q_k, p_k) \] formed by
- load vector \( q_k = [q^1_k, \ldots, q^{24}_k] \)
- price vector \( p_k = [p^1_k, \ldots, p^{24}_k] \)
Market Operation

Daily Operation:

1. ISO provides a forecast for load & price vectors ▶ (State $x$)
2. Players submit their bidding vectors to the ISO ▶ (Action $\alpha_n$)
3. ISO clears the market given the faced demand ▶ (Transition)
4. Payments result from new load & price vectors ▶ (Reward $r_n$)

$n^{th}$ Player

State $x$

Reward $r_n$

Action $a_n$

System Operator

Demand

Learning Optimal Strategies in a Partial Information Environment

N. Chrysanthopoulos, G. P. Papavassilopoulos
Market Operation

Daily Operation:

1. ISO provides a forecast for load & price vectors
2. Players submit their bidding vectors to the ISO
3. ISO clears the market given the faced demand
4. Payments result from new load & price vectors

\( n^{th} \text{ Player} \)

\( \text{State } x \) \hspace{1cm} \text{Reward } r_n \)

\( \text{System Operator} \)

\( \text{Action } a_n \) \hspace{1cm} \text{Demand}
Market Operation

Daily Operation:

1. ISO provides a forecast for load & price vectors
2. Players submit their bidding vectors to the ISO
3. ISO clears the market given the faced demand
4. Payments result from new load & price vectors

- (State $x$)
- (Action $\alpha_n$)
- (Transition)
- (Reward $r_n$)
Assumptions:

- Demand is Exogenous, Inelastic, Stochastic
- Players behave Non-cooperatively
- Markov Property imposed
  - ISO provides the current state as the forecast
  - Players make decision given only current state
  - \( p(x' \mid x, a) = \Pr\{X_{t+1} = x' \mid X_t = x, A_t = a\} \)
  - \( p(x' \mid x, a) \) is independent of time, previous states & actions

Competitive Markov Decision Process (CMDP)

Since market’s operation recurs daily, the discrete process observed at \( t = 0, 1, 2, \ldots \), with state \( X_t \), constitutes a Competitive Markov Decision Process, namely \( \{\Gamma\}_t \).
Incomplete Information

The system’s current state is $X_t = [x_{1,t}, \ldots, x_{K,t}]$ where $x_{k,t} = (q_{k,t}, p_{k,t})$ is the state of the $k^{th}$ bus.

We assume that each player has his own comprehension about the state, so we define the vector $\tilde{X}_t^n$ to be the transformation of the original state vector $X_t$ that the $n^{th}$ player uses as information set in decision making.

$$\varphi_n : X_t \rightarrow \tilde{X}_t^n$$

Linear Examples ($\tilde{X}_t^n = X_t A_n$):
- $A_n$ identity matrix (original state)
- $A_n$ projection matrix (part of state)

Non-Linear Examples:
- The maximum price is included at the state
Reinforcement Learning (Algorithm)

Implemented R-Learning algorithm:

Initialization of learning parameters \((\lambda, \gamma)\), action-value function \(Q_n(\tilde{x}_n, \alpha_n)\) and average reward \(\bar{r}_n\).

Repeat:

\[\tilde{x}_n \leftarrow \text{linear transformation of the current state}\]
Player chooses action \(\alpha_n\) under a policy
System transitions to the new state \(x'\)
Immediate reward \(r(x, \alpha_n, x')\) is received
\[D \leftarrow r_n(x, \alpha_n, x') - \bar{r}_n + \max_b Q_n(\tilde{x}', b) - Q_n(\tilde{x}_n, \alpha_n)\]
\[Q_n(\tilde{x}_n, \alpha_n) \leftarrow Q_n(\tilde{x}_n, \alpha_n) + \lambda_t \cdot D\]
\[\bar{r}_n \leftarrow \bar{r}_n + \gamma_t \cdot [r_n(x, \alpha_n, x') - \bar{r}_n]\]
Update the policy

The update rule:

\[Q_n(\tilde{x}_n, \alpha_n) \leftarrow Q_n(\tilde{x}_n, \alpha_n) + \lambda \left[ r_n(x, \alpha_n, x') - \bar{r}_n + \max_b Q_n(\tilde{x}', b) - Q_n(\tilde{x}_n, \alpha_n) \right]\]
Reinforcement Learning (Policy)

Implemented learning policy:

- As the learning policy we define a sequence of probabilities \( \{c^n_t\}_{t \in \mathbb{N}} \) for selecting a random action among the non-greedy available actions.

\[
c^n_t = \Pr\left\{ a_n \neq \arg \max_b Q_n (x', b) \right\}
\]  

(1)

\[
c^n_t = \left\{ \mathcal{F}(t) : \lim_{t \to \infty} c^n_t = L \right\}
\]  

(2)

- \( L \) is the weakened exploring rate occurred at the end.
- The effect of further exploitation controlled by \( \lambda_t, \gamma_t \in [0, 1] \).
- Step size parameters follow a descending course over time.
Table of Contents

1 Introduction
   - Overview
   - Influential Literature

2 Model Formulation
   - Market Structure
   - Incomplete Information
   - Reinforcement Learning

3 Simulations
   - Overview
   - Demand & Production Side
   - Results

4 Conclusion
   - Conclusion
   - Relevant work
Simulations’ Overview

- For the implementation we used a six-bus power network
- Three power plants which serve three standalone load buses
- Network’s topology resembles one of Wood & Wollenberg’s
- Simulations carried out with MATLAB (MATPOWER for OPF)

Cases of Ownership

- A: Symmetric (3 firms)
- B: Non-Symmetric (2 firms)

Informational Concepts

- 1: Simplest Information Set
- 2: Enriched Information Set
### Simulations’ Overview

- For the implementation we used a six-bus power network
- Three power plants which serve three standalone load buses
- Network’s topology resembles one of Wood & Wollenberg’s
- Simulations carried out with MATLAB (MATPOWER for OPF)

### Cases of Ownership

- **A**: Symmetric (3 firms)
- **B**: Non-Symmetric (2 firms)

### Informational Concepts

- **1**: Simplest Information Set
- **2**: Enriched Information Set
Simulations’ Overview

- For the implementation we used a six-bus power network
- Three power plants which serve three standalone load buses
- Network’s topology resembles one of Wood & Wollenberg’s
- Simulations carried out with MATLAB (MATPOWER for OPF)

Cases of Ownership

- **A**: Symmetric (3 firms)
- **B**: Non-Symmetric (2 firms)

Informational Concepts

- 1: Simplest Information Set
- 2: Enriched Information Set
Simulations’ Overview

- For the implementation we used a six-bus power network
- Three power plants which serve three standalone load buses
- Network’s topology resembles one of Wood & Wollenberg’s
- Simulations carried out with MATLAB (MATPOWER for OPF)

Cases of Ownership

- A: Symmetric (3 firms)
- B: Non-Symmetric (2 firms)

Informational Concepts

- 1: Simplest Information Set
- 2: Enriched Information Set
Simulations’ Overview

- For the implementation we used a six-bus power network
- Three power plants which serve three standalone load buses
- Network’s topology resembles one of Wood & Wollenberg’s
- Simulations carried out with MATLAB (MATPOWER for OPF)

Cases of Ownership

- A: Symmetric (3 firms)
- B: Non-Symmetric (2 firms)

Informational Concepts

- 1: Simplest Information Set
- 2: Enriched Information Set
Simulations’ Overview

- For the implementation we used a six-bus power network
- Three power plants which serve three standalone load buses
- Network’s topology resembles one of Wood & Wollenberg’s
- Simulations carried out with MATLAB (MATPOWER for OPF)

Cases of Ownership

- A: Symmetric (3 firms)
- B: Non-Symmetric (2 firms)

Informational Concepts

- 1: Simplest Information Set
- 2: Enriched Information Set

\[ X_t = [x_{1,t}, ..., x_{n,t}, ..., x_K,t] \]
\[ \hat{X}_t^n = [q_{n,t}, p_{n,t}, q_{total,t}, P_{max,t}] \]
\[ \sum_{i \in N} q_{i,t} \]
\[ \max (p_{i,t}) \]
Mean Loads
- for each bus
- for every hour

Indicative Sample
- Bounded Normal distribution
- 32% at the boundaries

Bounding Function
\[ \varphi(x) = \begin{cases} 
\mu - \sigma, & x < \mu - \sigma \\
\mu + \sigma, & x > \mu + \sigma \\
x, & \text{otherwise}
\end{cases} \]
\[ x \sim \mathcal{N}(\mu, \sigma^2) \]
There is a lower and an upper bound in generation capacity, namely $Q_{\text{min}}^i = 50\, \text{MW}$ and $Q_{\text{max}}^i = 150\, \text{MW}$.

Constant marginal cost, equal with $4\, \text{€/MWh}$.

Available Actions:

- Piece-wise linear bidding functions
Results - Case A (3 Firms - Symmetric)

Average Daily Profits & RL Average Reward

Simple Information Set - Enriched Information Set

Symmetric Outcome due to Symmetric Market Formation
Simple Information Set found to be More Profitable
Results - Case B (2 Firms - Non-Symmetric)

Average Daily Profits & RL Average Reward

Simple Information Set - Enriched Information Set

▶ The **Large Firm** has always a plant in the dispatches’ schedule
▶ The other plants compete for the **Residual Demand**

Learning Optimal Strategies in a Partial Information Environment

N. Chrysanthopoulos, G. P. Papavassilopoulos
Results - Greedy Action Plans

Contribution of Actions to Greedy Action Plans

Firm 1

Firm 2

Firm 3

Firm 1

Firm 2

Firm 3

Action 1 | Action 2 | Action 3 | Action 4
Table of Contents

1 Introduction
   • Overview
   • Influential Literature

2 Model Formulation
   • Market Structure
   • Incomplete Information
   • Reinforcement Learning

3 Simulations
   • Overview
   • Demand & Production Side
   • Results

4 Conclusion
   • Conclusion
   • Relevant work

Learning Optimal Strategies in a Partial Information Environment
N. Chrysanthopoulos, G. P. Papavassilopoulos
Different **market structures** & **Different informational concepts**

We studied the implementation of

- State space transformation technique
- R-Learning algorithm

under

- two informational concepts
  - only private information
  - private information + aggregated demand + max price
- two different cases of ownership
  - 3 firms own 3 units
  - 2 firms own 3 units
Relevant work

**Market Power** under different levels of **Network Transmission Constraints**

- Three different Cases, **three levels** of transmission constraints offer **thorough benchmark**
- **State space transformation** (incomplete information) examined from a **sufficiency** and **efficiency** perspective.
- **R-Learning algorithm** enables players to **identify greedy action plans** and **exert market power**.
Thank you for your attention!!

Any questions?

Chrysanthopoulos Nikos

nikoschryys@mail.ntua.gr