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

About the paper

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

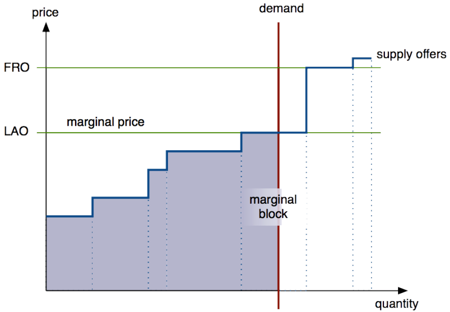
Basic Elements — Market Clearance

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:

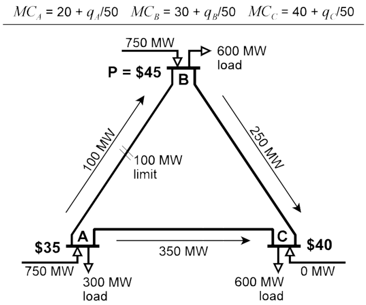
- Single-side
- Uniform
- LAO or FRO
- Marginal price



Basic Elements — LMP & Market Power

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

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

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

$\alpha_n = \left[\alpha_n^1, \ldots, \alpha_n^{24}\right]$

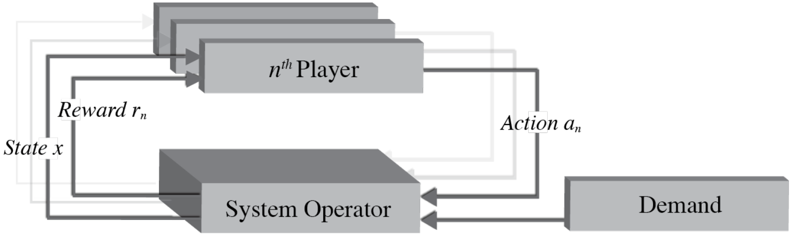
- daily bidding vector
- player's choice variables

State

$x_k = (q_k, p_k)$ formed by

- load vector $q_k = [q_k^1, \ldots q_k^{24}]$
- price vector $p_k = [p_k^1, \ldots p_k^{24}]$

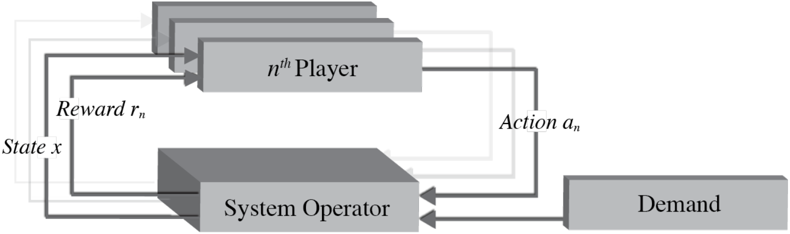
Market Operation



Daily Operation:

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

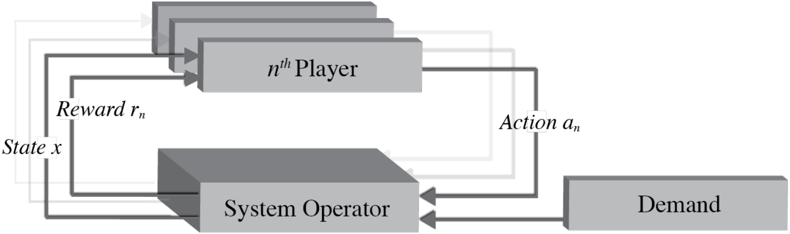
Market Operation



Daily Operation:

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

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 α_n)
3. ISO clears the market given the faced demand ▶ (Transition)
4. Payments result from new load & price vectors ▶ (Reward r_n)

Market Operation

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 $\widetilde{X}_t^n$ to be the **transformation** of the original state vector X_t that the n^{th} player uses as information set in decision making.

$$\varphi_n : X_t \to \widetilde{X}_t^n$$

Linear Examples ($\widetilde{X}_t^n = X_t A_n$) :

- 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 (λ, γ), action-value function $Q_n(\widetilde{x}_n, \alpha_n)$ and average reward $\bar{r}_n$.

Repeat:

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

The update rule:

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

Reinforcement Learning (Policy)

Implemented learning policy :

- As the learning policy we define a sequence of probabilities $\{c_t^n\}_{t\in N}$ for selecting a random action among the non-greedy available actions

$$c_t^n = \Pr\left\{a_n \neq \arg\max_b Q_n\left(x', b\right)\right\} \tag{1}$$

$$c_t^n = \left\{\mathcal{F}\left(t\right) \ : \ \lim_{t\to\infty} c_t^n = L\right\} \tag{2}$$

- 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

Simulations' Overview

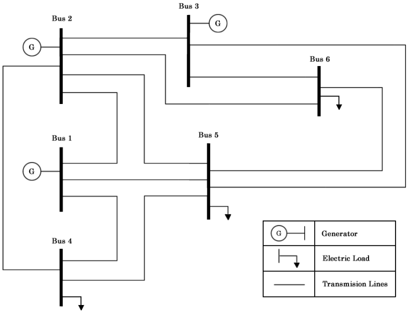
- For the implementation we used a six-bus power network
- Three power plants which serve three standalone load buses
- Network's topology resembles one of Wood & Wollenberg's
- Simulations carried out with MATLAB (MATPOWER for OPF)

Cases of Ownership

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

Informational Concepts

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



Simulations' Overview

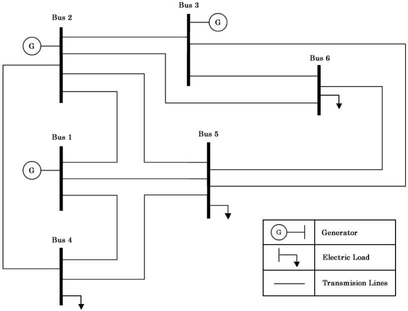
- For the implementation we used a six-bus power network
- Three power plants which serve three standalone load buses
- Network's topology resembles one of Wood & Wollenberg's
- Simulations carried out with MATLAB (MATPOWER for OPF)

Cases of Ownership

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

Informational Concepts

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



Simulations' Overview

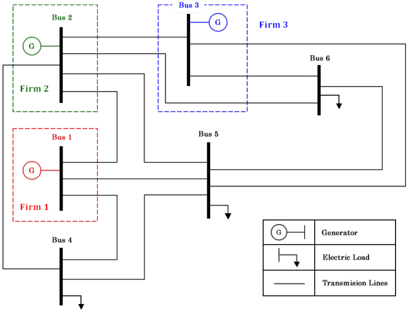
- For the implementation we used a six-bus power network
- Three power plants which serve three standalone load buses
- Network's topology resembles one of Wood & Wollenberg's
- Simulations carried out with MATLAB (MATPOWER for OPF)

Cases of Ownership

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

Informational Concepts

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



Simulations' Overview

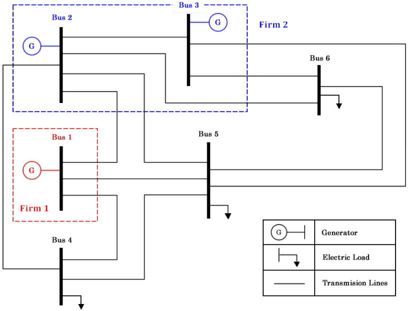
- For the implementation we used a six-bus power network
- Three power plants which serve three standalone load buses
- Network's topology resembles one of Wood & Wollenberg's
- Simulations carried out with MATLAB (MATPOWER for OPF)

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

Actual System's State $X_t = [x_{1,t}, \dots, \boxed{x_{n,t}}, \dots, x_{K,t}]$

State Perceived by n^th Player $\tilde{X}_t^n = [x_{1,t}, \dots, \boxed{x_{n,t}}, \dots, x_{K,t}]$

$(q_{n,t}, p_{n,t})$

Simulations' Overview

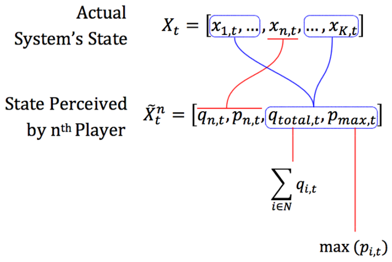
- For the implementation we used a six-bus power network
- Three power plants which serve three standalone load buses
- Network's topology resembles one of Wood & Wollenberg's
- Simulations carried out with MATLAB (MATPOWER for OPF)

Cases of Ownership

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

Informational Concepts

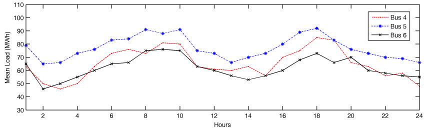
- 1: Simplest Information Set
- 2: Enriched Information Set



Stochastic Demand

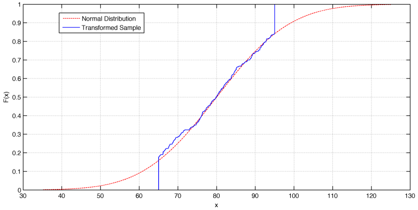
Mean Loads

- for each bus
- for every hour



Indicative Sample

- Bounded Normal distribution
- 32% at the boundaries



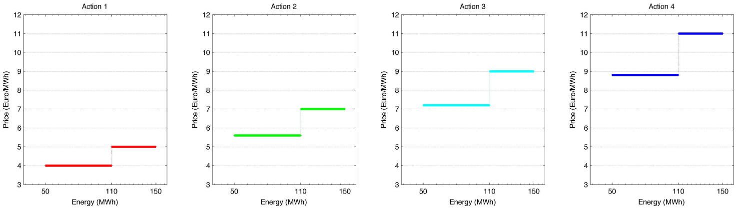
Bounding Function

$$\varphi(x) = \begin{cases} \mu - \sigma & , x < \mu - \sigma \\ \mu + \sigma & , x > \mu + \sigma \\ x & , otherwise \end{cases} \qquad x \sim \mathcal{N}(\mu, \sigma^2)$$

Production side

- There is **a lower and an upper bound** in generation capacity, namely $Q^i_{min} = 50MW$ and $Q^i_{max} = 150MW$
- **Constant marginal cost**, equal with $4€/MWh$.

Available Actions :

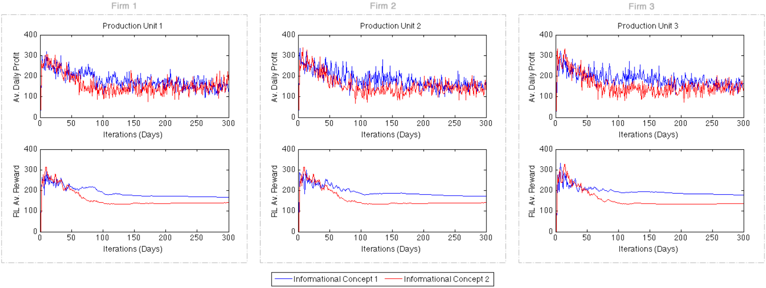


- **Piece-wise linear** bidding functions

Results - Case A (3 Firms - Symmetric)

Average Daily Profits & RL Average Reward

Simple Information Set - Enriched Information Set

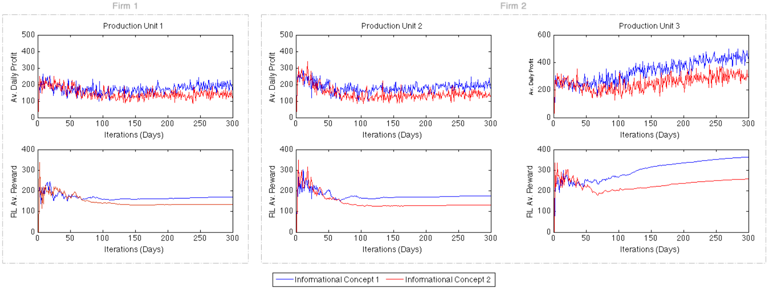


- **Symmetric Outcome** due to **Symmetric Market Formation**
- **Simple Information Set** found to be **More Profitable**

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**

Results - Greedy Action Plans

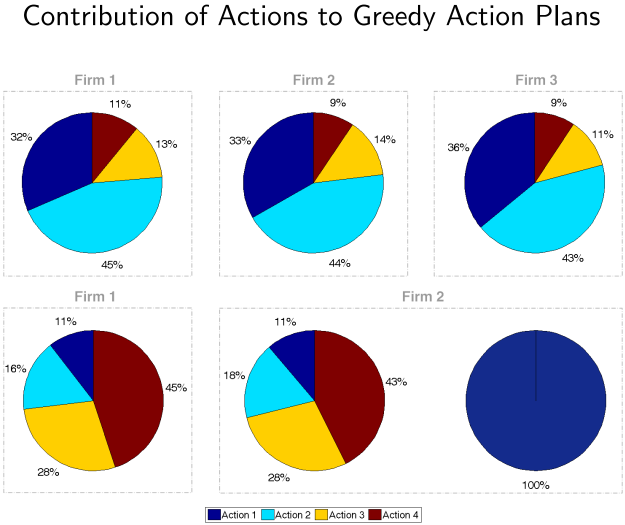


Table of Contents

Conclusion

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

nikoschrys@mail.ntua.gr