Reinforcement Learning for Optimal Day-Ahead Electricity Trading with Battery Storage

Peyman Kor[1], Aojie Hong[1]

[1] Department of Energy Resources, University of Stavanger, Stavanger, Norway

Introduction

- The coastal region of Europe has seen a growing interest in wind energy. A challenge with wind energy sources is handling the variability.
- A simple solution to handle this variability is to sell any electricity from a renewable source into the grid and use the tremendous power of the grid to handle this variability.
- However, there has been considerable interest in using storage (particularly battery storage) to handle variability in renewable energies. The battery storage provides flexibility to the energy provider.

Objectives

1. Illustrate the introduce the workflow for Solving sequential Decision Problesm
2. Provides a consistent policy for decision in the context of energy storage
3. Present Value Function Approximation (VFA) based on the Bellman optimality equation as a policy for decision in hand a day ahead.

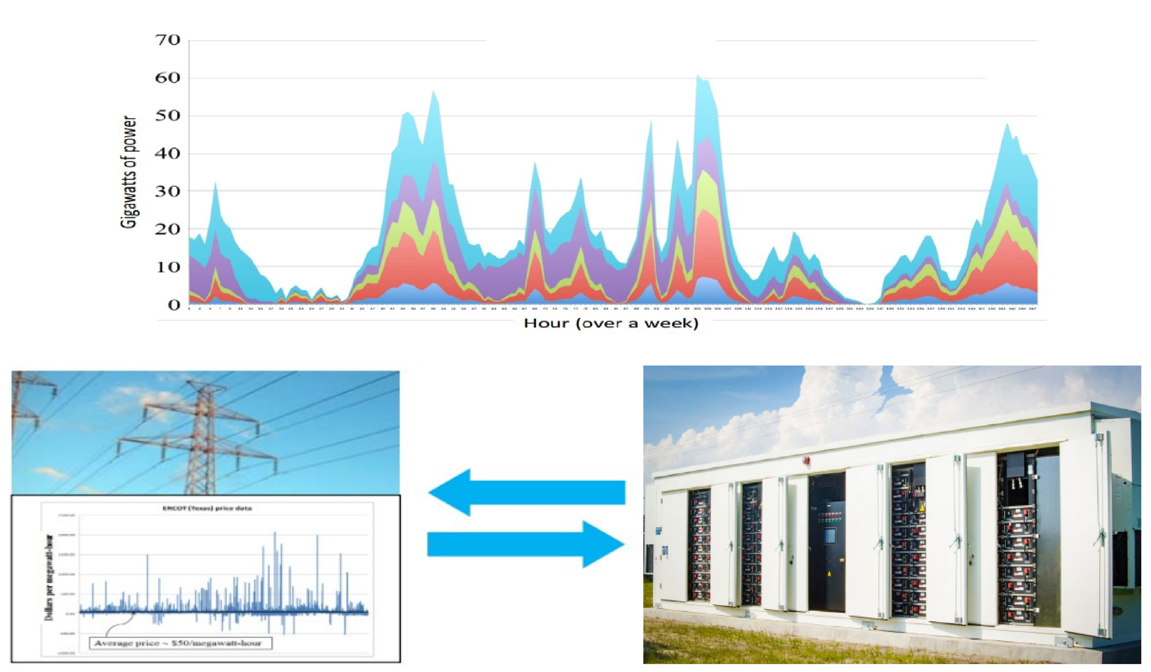

Figure 1: The Electricty Storage Decision Schematics

Methods

`Lookahead Policy` is optimal:

$$X_t^*(S_t) = \arg\max_{x_t}\left(C(S_t,x_t) + E\left\{\max_{\pi\in\Pi}\left\{E\sum_{t'=t+1}^{T} C(S_{t'},X_{t'}^{\pi}(S_{t'}))\,|\,S_{t+1}\right\}|\,S_t,x_t\right\}\right)$$

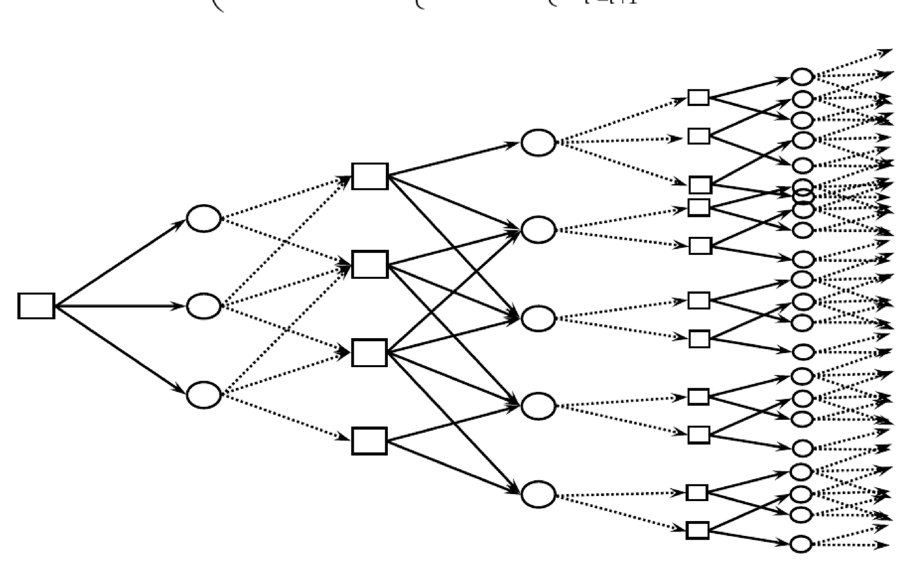

Figure 2: The Electricty Storage Decsion Schematics

Results

Visual Illustration of Concept:

- Battery could be fully charged Or Discharged, R t = [1,0]
- The Electricity Price could have three possible scenario, p=[1,2,3]
- The Electricity price could decrease or increase with ,Δp = [−0.5,0.5] each with chance.
- The number of time steps is 3, T = [0,1,2]

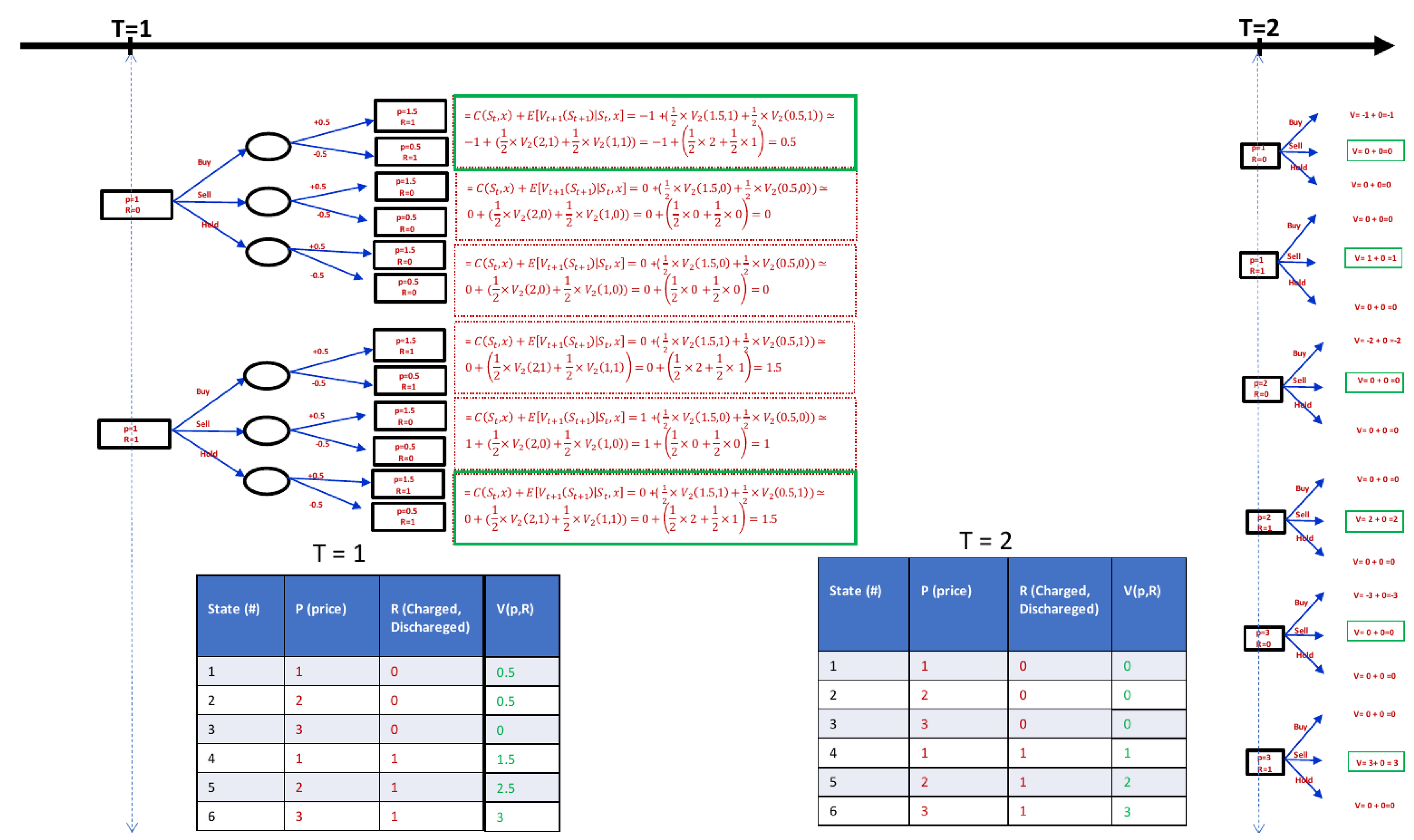

Figure 3: The Electricty Storage Decsion Schematics

Norway Electricity data:

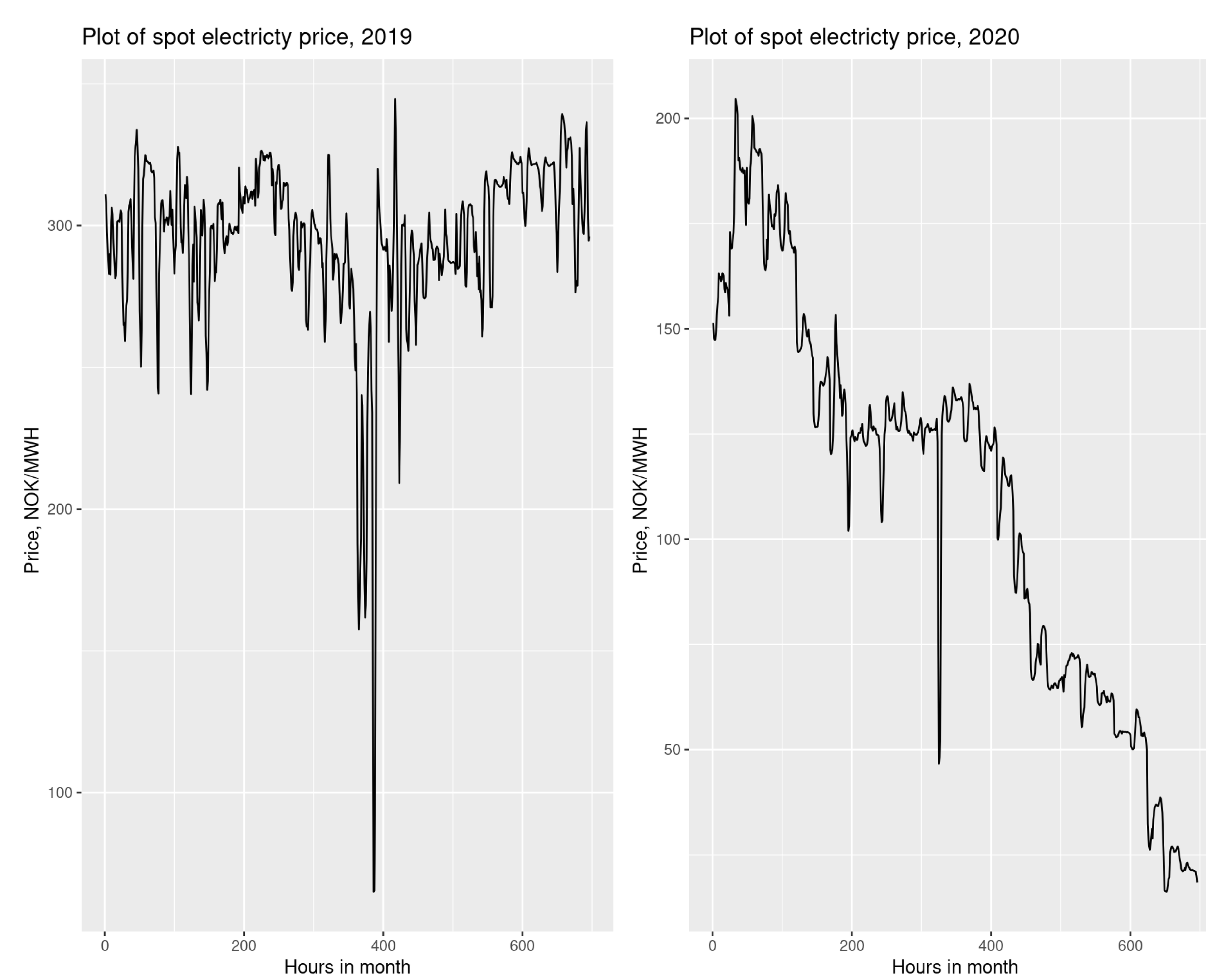

Figure 4: The Electricty Storage Decsion Schematics

Next Steps

- In this work, we discretized the state and decision space in a manner that the MDP problem can be solved in a computationally tractable way.
- Future work will apply the workflow to the high dimensional state-decision space in a computationally efficient way.

Conclusion

- We show that policy trained using VFA on historical price data from the Norway market data is effective.
- The trained agent exploits the monotonicity of the value function to find a profit-generating policy for trading.
- Finally, the VFA policies consistently generated more revenue than the rule-based heuristic strategies that we considered, confirming that a VFA approach approximating the current decision on future decisions is worthwhile.