Optimization in energy communities

Reinforcement learning and mathematical optimization model

TH-Köln – Cologne University of Applied Sciences, Masterproject supervised by
Prof. Thorsten Schneiders and Sascha Birk (M.Sc.)
Cologne, Germany
Syed Ali Mashood Naqvi (B.Sc.), Yannik Weber (B.Eng.)
Contact: syed_ali_mashood_naqvi.naqvi@smail.th-koeln.de, yannik.weber@smail.th-koeln.de

Abstract – **Intelligent use of energy is one of the keys to success for an energy revolution. With growing use of renewable energy sources and growing tariffs for electricity production, the need for optimization in small scale grids is omnipresent. This paper evaluates two approaches to optimize the usage and production of energy in small communities.**

I. Introduction – About the Project

Continuously growing tariffs for electricity production bring about the need of search for alternatives [1]. A great amount of research has been conducted not only to optimize the electricity generation but also regarding the distribution of electricity in order to reduce line losses and to lower the cost of electricity transportation. Distributed electricity generation or optimal power flow problems can be solved by several methods. Some of them are using conventional mathematical models, others are using Machine Learning techniques including Deep Neural Networks, fuzzy logic and RL [2]. In this project, these two approaches are compared. Therefore, different models were prepared, run and evaluated on the same set of data.

Reinforcement Learning is a sub-branch of Machine Learning which is based on a trial-and-error approach. An RL-agent interacts with the environment dynamically, trying different actions and consequently evaluating the path on the basis of rewards [3]. Reinforcement Learning is one of the emerging fields that have been largely incorporated in the energy sector to optimize the efficiency of energy systems. There are many applications of Reinforcement Learning in energy systems ranging from Multi-Agent Reinforcement Learning to Deep Q Learning. However, this project only involves a simple Reinforcement Learning strategy that employs an agent to interact with an environment and then estimate the best possible action based on rewards.

Pandapower is an open-source tool for the modelling, analysis and optimization of energy systems. Built on the data analysis library *pandas* and the energy system analysis toolbox PYPOWER, it enables and easy use of network calculations [4]. In this program, a simple linear timeseries optimization was implemented where the objective and constraints are calculated for each timestep. The framework of the network was taken from the pandapower library.

II. Issue and Objective

With the growing use of renewable energy sources, the energy market is evolving into a more decentralized system. Within this market, new energy transfer models gain interest, such as peer-to-peer-transactions through microgrid architectures [5]. These "energy communities" provide new challenges regarding energy distribution and grid stability, while different approaches to optimize those networks can be applied. The methods applied in this paper aim to minimize the resulting costs in the community by optimizing the energy production and distribution.

Both models were built very similarly to enable a later comparison. However, certain parameters that were used to describe the environments had to differ. The focus of the models was set on houses with PV-generation and battery storage units. The maximum power of the PV-production was taken from the data provided. The storage capacity as well as the charging power were sized adequately [6, 7]. Furthermore, the prices used for the calculation were adapted to current energy prices [8]. The parameters that were used are shown in the table below.

Parameter	**Reinforcement learning**	**Mathematical optimization**
Grid capacity	15 kWh	infinite
Storage capacity	10 kWh	10 kWh
Charging power	/	5 kW
Initial storage charge	0 kWh	0 kWh
Energy cost (grid)	40 cents	40 cents
Energy cost (storage)	15 cents	15 cents
Energy cost (PV)	10 cents	10 cents

Additionally, the datasets for the demand in the single nodes and the PV-production are the same for both models. This given input data can be seen in figure 1.

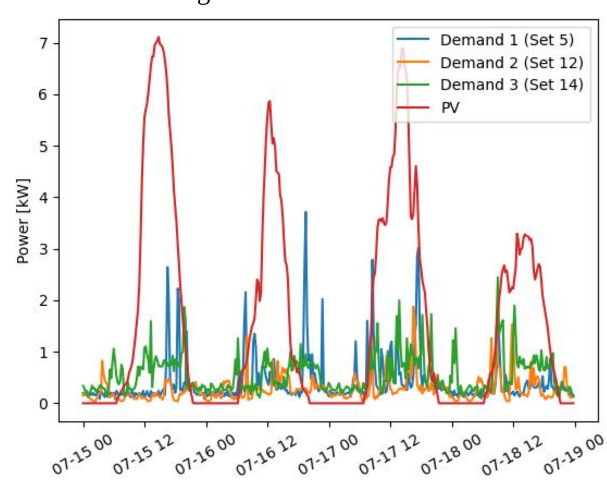

Figure 1: Input data for demand and PV-production

Both models were used to evaluate three different houses with different demand datasets. All houses, however, are considered to have the same PV-production. For the demand datasets, data from the HTW-Berlin was used [9]. For the PV-production the data was taken from a pvlib simulation for Cologne based on data of the DWD [10]. Within both models, the demand can be covered using either PV, storage or the grid.

The main difference between the methods discussed in this paper is that while the Reinforcement model only considered one node at a time, the mathematical optimization model considered all three nodes in a connected network. This way the results may be influenced by the possibility of interaction between the nodes and the limits of the connecting lines of the network.

III. Methodology

A. Reinforcement Learning

The first method employed in this research is a Reinforcement Learning model, which operates within a custom environment. The custom environment considers one house whose source if electricity is either of the three sources.

The custom environment has an agent that is provided to take an action out of a set of three predefined actions. Action:0 makes the agent take all the electricity from the main grid. Action:1 makes the agent take the maximum possible amount of electricity from the PV-source and the rest from the grid to cover its demand completely. Action:2 makes the agent fulfil its demand from the battery storage. It stores energy from the PV-source when there is excess PV available and discharges the energy when the demand is too high and cannot be covered by either grid or PV. If the agent does not choose any of the above-mentioned actions, it is severely penalized, so that the agent always chooses one of the given actions. The other type of penalization is based on the total power consumption. The agent is heavily penalized if the total maximum power it attains exceeds the capacity from the grid. The observation space is defined as a box space with a dictionary of two-dimensional integers with a range from 0 – the minimum value – to the maximum value of one less than the episode length. This range corresponds to the load demand at each time-step. The agent was trained and tested on this custom environment using two Reinforcement Learning algorithms, namely A2C Actor Critic Method and PPO Algorithm (Proximal Policy Optimization), which were alternately employed. These algorithms were incorporated into the custom environment using the Open AI gym library, a widely used and robust library for Reinforcement Learning.

This Reinforcement Learning model was trained on this custom environment using a load dataset "zero" taken from the recorded sets of data. After successful training on this dataset, this Reinforcement Learning model was then tested on three different datasets, namely datasets 5,12 and 14, the results of which are discussed in the results section.

B. Mathematical optimization model

The second method is using linear mathematical correlations to optimize the network. The objective for this method is defined in formula (1).

$$c = \sum_{1}^{t} c_1 \cdot m_p + c_s \cdot m_{s\_\mathrm{use}} + c_\mathrm{n} \cdot m_{pn} \quad (1)$$

Within this formula, *c* is used for the cost parameters and *m* is used for the variables in the model. The index *p* refers to the PV-production, while the index *s* refers to the storage units. The grid is referred to with the indices *n* for cost and *pn* for energy. It should be noted, that energy stored to the battery storage is not considered in this formula, as it would be considered twice – once for the usage and once for the storage. According to the cost structure, the model favours the PV-generated energy over the stored energy and over the grid energy.

To correctly evaluate the defined objective, several constraints were introduced to the model. First, the PV-generation was coupled with the PV-production data (formula (2)).

$$0 = gen_{\min\_p} \leq m_p \leq pv\_production \quad (2)$$

Additionally, the power generated by or stored in the battery storage units must be limited to assure a plausible model (formula (3) and (4)). Nevertheless, in this model, both usage and storage of energy are not considered to have any losses.

$$m_{s\_\mathrm{use}} \leq charging\ power \quad (3)$$

$$m_{s\_\mathrm{store}} \leq charging\ power \quad (4)$$

It is not plausible that a storage unit is simultaneously charged and discharged. Therefore, the product of used and stored energy must be zero. Unfortunately, this seemed to cause a critical error in the model, which is why the constraint in formula (5) was adapted to the limit of 0.1 or smaller.

$$m_{s\_\mathrm{use}} \cdot m_{s_\mathrm{store}} \leq 0.1 \quad (5)$$

Furthermore, the storage units each need to change the state of charge (*soc*) according to the used or stored energy. Therefore, formula (6) is used, where *T* is the total amount of calculated timesteps, and *S* is the amount of storage units within the model. Combined with the min and max criteria in formula (7) the storage is completely defined.

$$m_{soc}[s, t+1] = m_{soc}[s, t] - m_{s\_\mathrm{use}}[s, t] + m_{s\_\mathrm{store}}[s, t] \quad (6)$$
$$(for\ t\ in\ T\ and\ for\ s\ in\ S)$$

$$0 \leq m_{soc} \leq Storage\ capacity \quad (7)$$

Additional to the constraints for each generator, the connection between each unit is limited by the characteristics of the connecting lines between the generators. As these lines need to be limited by a maximum amount of energy transferred (*max_i*) in each direction, formula (8) and (9) differ only in one being positive and the other being negative. The transferred energy is hereby defined as a product of the susceptance of the line (*B*) and the voltage angle (*ϑ*) of the two connecting buses.

$$\mathrm{Line}_{B\_\mathrm{max}\_i} \geq line_B \cdot \left(m_{\vartheta\_\mathrm{from\_bus}} - m_{\vartheta\_\mathrm{to\_bus}}\right) \quad (8)$$

$$-\mathrm{line}_{B\_\mathrm{max}\_i} \leq line_B \cdot \left(m_{\vartheta\_\mathrm{from\_bus}} - m_{\vartheta\_\mathrm{to\_bus}}\right) \quad (9)$$

Ultimately, in the model, each generator is coupled with a load or demand (*d*) that needs to be covered, using either energy from the PV-system, the battery or the grid. This constraint is defined in the two formulas (10) and (11), where formula (11) sums up the loads of the lines (*l*), generators (*g*) and demands (*d*).

$$m_p + m_{s\_\text{use}} - m_{s\_\text{store}} + m_{pn} = d \quad (10)$$

$$\sum_1^l line_B \cdot m_\vartheta = \sum_1^g (m_p + m_{s\_\text{use}} - m_{s\_\text{store}}) - \sum_1^d d \quad (11)$$

To enable a theoretically infinite amount of energy taken directly from the grid, the voltage angle of the reference bus is set to zero (formula (12)) [11].

$$m_{\vartheta\_\text{refbus}} = 0 \quad (12)$$

IV. Results

Both models analyse the loads of the datasets 5, 12 and 14 regarding the optimal power flow and usage. Figure 2 shows the PV-production and the state of charge of the battery storage from the Reinforcement Learning model. These can be compared to the data from the mathematical optimization model in Figure 3.

In the Reinforcement Learning model the storages all get fully charged with excess PV for each cycle and get discharged when there is no excess PV available for the community, with only one exception. In the mathematical model on the other hand, only one of the storages gets charged and discharged in a similar manner. This may have different reasons, such as the Reinforcement model only optimizing one house at a time and therefore using the optimal solution for one house and the mathematical model optimizing three houses combined.

It is noteworthy that two of the storage units in the mathematical model are not charged and discharged regularly, which would have been anticipated. For the storages one and three, only a fraction of their capacity is used, with storage one even being charged from the grid. While this behaviour was not expected, there may be reasonable explanations for it, such as a stabilization of the voltage angle between two houses or the possibility to cover load spikes in the later course of the simulation. While this presents a great possibility for further investigation in the mathematical model, almost no similar events can be recorded in the Reinforcement model.

Regarding the PV-production on the other hand, it can be noted that the mathematical model uses the maximum amount of power for two of the three generators, while the Reinforcement model only uses the PV-power of the third dataset to a similar extent. Again, this may be an issue of the Reinforcement model only evaluating one house at a time and therefore not having the need to use more PV-power than the needed energy to cover the demand of that house, while the mathematical model uses the PV-power to reduce costs of the overall model.

Furthermore, both models not necessarily use the full amount of PV-power provided (eg. Figure 3 PV 2). This way of shutting down the power generation provided to optimize the models can be a very interesting topic for further investigation.

While both models are at some point not using the PV-power provided to its full potential, the grid connection still shows energy exportation for both models (figure 4 and 5).

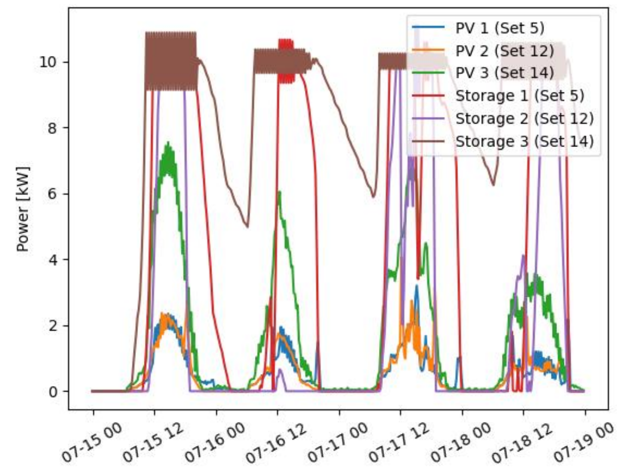

Figure 2: PV-production and storage soc (Reinforcement model)

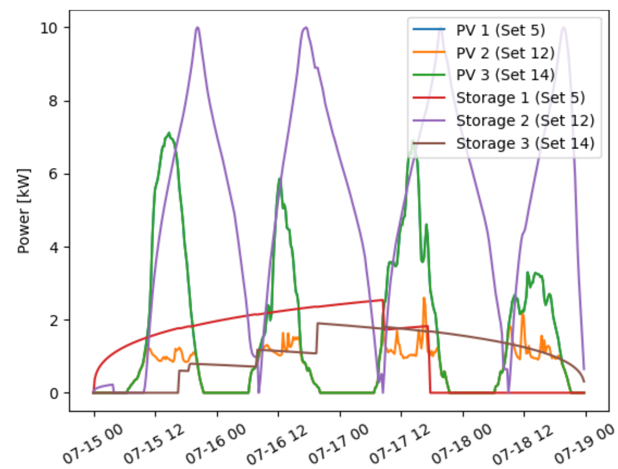

Figure 3: PV-production and storage soc (mathematical model)

In the Reinforcement model, only the dataset 14 has a negative grid value, which indicates energy exportation. This is rewarded in the model with a less penalized outcome for the agent, and therefore this outcome was expected. Especially as the PV-power is used to a very high extent, this outcome offers possibilities to integrate PV-production into the wider energy grid.

In the mathematical model on the other hand, every house exports energy into the grid. The two fully used PV-generators seem to cover the demand and export the rest completely, but even from the second generator, which is shut down at a smaller fraction of its potential, energy is exported into the grid. There is a very simple explanation for this behaviour, as the energy exported into the grid is considered as negative m_{pn} in the cost function and therefore drastically reducing the overall cost in the model.

With the current use of the battery storage units, the Reinforcement model offers a higher rate of self-consumption, as the batteries get charged and discharged every cycle. However, a valid statement regarding which model optimizes the cost more efficiently cannot yet be made, as the cost of the mathematical model is lowered by the energy exportation.

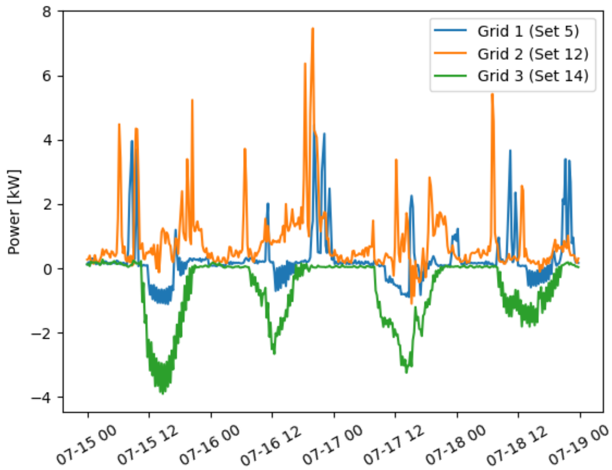

Figure 4: Energy exchange with the grid (Reinforcement model)

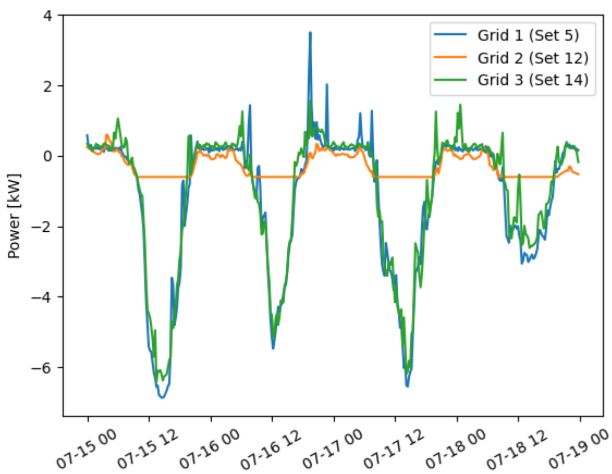

Figure 5: Energy exchange with the grid (mathematical model)

V. DISCUSSION

The Reinforcement Learning model sufficiently optimizes the cost of electricity across the local grid. However, there are many improvements that can be implemented in the model. Hyperparameter tuning is a key step that can be implemented to enhance and improve the model performances. For this project, both models (A2C, PPO) were implemented with their default parameters due to system constraints. Other approaches such as epsilon greedy methods can be implemented, which can be of vital importance in making the agent accumulate more rewards and to get the agent to make more precise decisions. This project uses a simple Reinforcement Learning method while Multi-Agent Reinforcement Learning can be used to make different agents interact with each other and extract electricity at different intervals from different respective resources. These different approaches can enhance the model's performance.

The mathematical optimization model on the other hand works well with a small network of three generators. So far, the results resemble the values we would expect in this network. Only the usage of storage units cannot be fully analyzed, as the model gives preference to exporting the energy into the grid. As a result, energy that is exported into the grid reduces the cost drastically. The easiest way to solve this for future calculations would be to use the absolute value of the energy put into or taken from the grid. This would force the system to minimize the interaction with the grid completely. Another way to solve this problem would be to split the variable of energy taken from the grid in a similar way as the batterie units are handled and only take the power used from the grid into account in the cost function.

VI. CONCLUSION AND OUTLOOK

The Reinforcement Learning model works more consistently regarding the self-consumption for single generator-storage-combinations. Although at the current state no statement regarding the functionality in small networks can be made, it can be a good model to implement in single household optimizations.

Further, the Reinforcement Learning model should be adapted to a small network like the one examined in the mathematical model, to test its functionality for energy communities. The preferred way to do this would be with a Multi-Agent Reinforcement Model.

The mathematical model however can be further developed to optimize networks with different combinations of generators and storage units. Furthermore, it can be of great interest to force the model to use all the available PV-power, as currently the PV-generators can be shut down. Additionally, the interaction with the grid can be further investigated, to improve the self-consumption.

REFERENCES

[1] A. V. Varganova, V. R. Khramshin, and A. A. Radionov, "Improving Efficiency of Electric Energy System and Grid Operating Modes: Review of Optimization Techniques," *Energies*, vol. 15, no. 19, p. 7177, 2022, doi: 10.3390/en15197177.

[2] I. Kim, B. Kim, and D. Sidorov, "Machine Learning for Energy Systems Optimization," *Energies*, vol. 15, no. 11, p. 4116, 2022, doi: 10.3390/en15114116.

[3] S. Touzani *et al.*, "Controlling distributed energy resources via deep reinforcement learning for load flexibility and energy efficiency," *Applied Energy*, vol. 304, p. 117733, 2021, doi: 10.1016/j.apenergy.2021.117733.

[4] Jekyll & Minimal Mistakes, *About pandapower.* [Online]. Available: http://www.pandapower.org/about/ (accessed: Mar. 12 2023).

[5] F. Imbault, M. Swiatek, R. de Beaufort, and R. Plana, "The green blockchain: Managing decentralized energy production and consumption," in pp. 1–5.

[6] N. Orth, J. Weniger, L. Meissner, I. Lawaczeck, and V. Quaschning, "Stromspeicher-Inspektion 2022," Mar. 2022. Accessed: Mar. 10 2023.

[7] J. Weniger, V. Quaschning, and T. Tjaden, "Optimale Dimensionierung von PV-Speichersystemen," *magazine*, 01/2013, pp. 70–75, 2013.

[8] U. Bemmann, *Was kosten Stromspeicher?: Preise und Vergleiche* (accessed: Mar. 10 2023).

[9] Hochschule für Technik und Wirtschaft (HTW) Berlin, *ELEKTRISCHE LASTPROFILE FÜR WOHNGEBÄUDE IN DEUTSCHLAND* (accessed: Mar. 13 2023).

[10] Deutscher Wetterdienst, *CDC (Climate Data Center)* (accessed: Mar. 13 2023).

[11] V. Dvorkin, P. van Hentenryck, J. Kazempour, and P. Pinson, "Differentially Private Distributed Optimal Power Flow," Oct. 2019. [Online]. Available: https://arxiv.org/pdf/1910.10136