Multi-agent reinforcement learning for renewable integration in the electric power grid

Vincent Mai **Tianyu Zhang**
Université de Montréal & Mila
Montréal, Canada
vincent.mai@umontreal.ca
tianyu.zhang@mila.quebec

Antoine Lesage-Landry
Polytechnique Montréal & GERAD
2500 de Polytechnique road
Montréal, Canada, H3T 1J4
antoine.lesage-landry@polymtl.ca

Abstract

As part of the fight against climate change, the electric power system is transitioning from fuel-burning generators to renewable sources of power like wind and solar. To allow for the grid to rely heavily on renewables, important operational changes must be done. For example, novel approaches for frequency regulation, i.e., for balancing in real-time demand and generation, are required to ensure the stability of a renewable electric system. Demand response programs in which loads adjust in part their power consumption for the grid's benefit, can be used to provide frequency regulation. In this proposal, we present and motivate a collaborative multi-agent reinforcement learning approach to meet the algorithmic requirements for providing real-time power balancing with demand response.

1 Introduction

Climate change mitigation brings about important transformations in the electric power system. In 2019, the United States (U.S.) Environmental Protection Agency (EPA) reported that 25% of the country's greenhouse gas emission came from electricity generation [2]. The EPA further noted that 62% of the U.S. electricity was generated from burning fossil fuel [1]. In an effort to cut down and ultimately eliminate greenhouse gas emissions produced by the electricity sector, the power grid is moving from a conventional, fuel-burning to a renewable, natural phenomenon-based generation, e.g., wind turbine and solar photovoltaics. Apart from the immediate financial consequence of deploying new generators, electric grid operations must undergo an important paradigm shift as future energy systems will rely on uncertain, fast-ramping, and intermittent sources of power, i.e., renewables [27, 6]. Without a renewed way of operating the power grid, the transition toward a renewable energy grid is hardly conceivable. Specifically, to ensure the stability of the electric grid, a near-perfect balance between the power demand and generation is critical [14]. Hence, trading a constant, deterministic generation for an intermittent, uncertain one – the now renewable generation – imposes operational challenges to system operators as the need for power balancing is exacerbated. At the second timescale, this balancing task is referred to as frequency regulation [6, 27] because a grid frequency at its nominal value, viz. 60 Hz in North America and 50 Hz in Europe, indicates a balanced load-generation [14].

1.1 Demand response for renewable integration

Demand response [26, 20, 25] (DR) of flexible loads is a greenhouse gas emission-free, and renewable alternative to provide frequency regulation services [7, 27] and hence to cope with renewable intermittency instead of relying, for example, on conventional power plants [22, 21] or expansive battery energy storage [11, 9, 34]. DR refers to programs in which loads, e.g. residential households, modulate their non-critical power consumption following an aggregator's or system operator's

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2021.

instructions (e.g., turning on or off air conditioning units) or signals (e.g., increasing/decreasing electricity rates). For example, this can represent a house that accepts to have its air conditioning unit be turned OFF and ON successively to track the output of a wind turbine as long as its temperature is within an acceptable range, e.g., between 20°C and 22°C. Loads constrained by temperature requirements are referred to as thermostatically controlled loads (TCLs). DR is an attractive prospect for power balancing because of its fast response and of its low deployment required investments [27]. Fast timescale DR like frequency regulation requires algorithmic approaches that provide actions within a few seconds or less. These actions must account for the dynamic behavior of the controllable loads and the uncertainty of both the environment, e.g. changing weather, and the loads, e.g. erratic costumer behavior. Moreover, DR requires a large number of loads to provide a sustained level of controllable power and hence to be able to track regulation setpoints, i.e. the power imbalances. Frequency regulation DR is, therefore, intrinsically a cooperative multi-agent problem.

1.2 Motivation for a machine learning-based approach

In sum, four critical design aspects impose tough constraints on the algorithmic design of frequency regulation DR: (i) adaptivity under uncertainty, (ii) computational efficiency for real-time decision-making, (iii) sequential decision-making for dynamic environments, and (iv) scalability to and cooperativity in multi-agent settings. Several frameworks have been considered for this problem but were all insufficiently addressing at least one aspect. For example, model predictive control (MPC) approaches [13, 18, 17] integrate the sequential decision-making component of the problem and adequately model system dynamics but are subject to an important computational burden due to the receding horizon formulation. The resulting is, therefore, not suitable for real-time implementation. Online optimization approaches [15, 16] address this issue; they offer very efficient algorithms to deploy DR resources in real-time but they are mostly greedy, i.e., they do not account for the system's dynamics. Finally, the multi-agent aspect of the problem renders dynamic programming-based [3] and standard reinforcement learning-based [23] approaches intractable.

In this work, we propose to implement a multi-agent reinforcement learning (MARL) method, namely multi-agent PPO [33], to coordinate the power consumption of residential households equipped with an air conditioning unit. Our approach accounts for all critical design aspects to provide high-performance frequency regulation using demand response of TCLs. Specifically, the policy-based implementation tackles the two first aspects as (i) the policy adapts to the current problem's state, i.e., no static trajectory is computed a priori, thus providing adaptivity under uncertainty, and (ii) the policy only requires a straightforward evaluation to provide an action, i.e., no multi-period problem needs to be solved in each round, hence allowing real-time decision-making. The Bellman equation used in reinforcement learning inherently accounts for (iii) the dynamics of the environment. Finally, the (iv) scalability is made possible via the centralized training and decentralized execution (CT-DE) framework, which allows to add a new TCL in the network as a new instance of the agent, without additional training. Additionally, networked or aggregated communications allow to add a new TCL without any change to the other agents, or any additional computing burden except for the new agent.

Our expected contributions are:

- We propose a frequency regulation demand response method based on multi-agent reinforcement learning, namely multi-agent PPO [33]. Our approach is, to the authors' knowledge, the only one that addresses all identified algorithmic requirements for frequency regulation.
- We build a thermostatically controlled load coordination Gym environment based on the second-order thermal model from [8, 12]. The environment together with our approach is to be made readily available to the community for (i) collaborative multi-agent reinforcement learning algorithmic development and (ii) for evaluating the performance of demand response methods. Frequency regulation demand response is a timely problem, but also presents interesting challenges which are of interest for the machine learning community.

2 Demand response as a multi-agent reinforcement learning problem

To train and test our multi-agent algorithm, we are developing a simulator based on the Ray framework [19], and compatible with OpenAI Gym.

We consider a system $\mathcal{T}$ of controllable TCLs, each of them controlled by a different agent A_i. Each TCL $i \in \mathcal{T}$ has unique characteristics: air conditioner's power consumption and coefficient of performance, etc. – and is simulated in a building with its own unique thermal properties.

On the simulator side, at each time step, a second-order thermal model [12], inspired from GridLAB-D [8], is used to update the state of each house. This is based on the previous state, the actions of the agents, and some random elements modelling, for example, an user opening a window. The variation in the grid's power generation and the change in outdoors temperature are also modelled.

On the agent side, each agent A_i observes a unique state $\mathbf{X}_{i,t}$, takes an action $a_{i,t}$, and receives a reward $r_{i,t}$. The **state** $\mathbf{X}_{i,t}$ contains: (1) the indoor temperature for TCL i as well as the temperature preference bounds set by the TCL user, (2) the status (ON or OFF) of its cooling unit, its power consumption, and its remaining lock-out time, (3) the calendar date, time, outdoors temperature, and optionally weather predictions, (4) the regulation signal common to all TCLs, (5) the unique thermal and power parameters of TCL i, and (6) the information communicated by other TCLs in the network as to coordinate the aggregation's response. For each agent A_i, the possible **action** at time t is to change the current status of the TCL cooling unit: if $a_{i,t} = 1$, TCL i switches ON of OFF. Otherwise, if $a_{i,t} = 0$, it keeps its current status. To protect the hardware, a lockout constraint is enforced: after $a_{t,i} = 1$, the TCL must wait a given amount of time before having access to $a_{t,i} = 1$ again [35, 30]. The negative **reward** is a combination of two penalties representing a dual objective: (1) every single agent must satisfy its indoors temperature constraints, which is regulated proportionally to the square distance between the current indoor temperature and the closest limit when the indoor temperature is outside an accepted range around a desired temperature, and (2) the combined power consumption of all agents at the point of connection with the grid should track the regulation signal: all agents are penalized based on the square distance between these two quantities. To collaborate, the agents can **communicate** their current status, temperature, constraints, and thermal and power properties. For scalability and privacy concerns, we will also experiment with networked communication, i.e. communicating locally with the $2n$ direct neighbours, for example A_{i-n} to A_{i+n}, so that the resulting graph stays fully connected. Another interesting option to explore is aggregated data – through a human-designed pipeline or a learned latent space.

3 MARL proposed method

Collaborative MARL is a challenging task. Common challenges are the non-stationarity of the environment as all agents learn simultaneously, its partial observability for a given agent which may not be aware of the reality of other agents, and the necessity for the agents to learn to communicate and to coordinate [10].

As a first step, we would train a DQN with CT-DE [28]. We will then experiment with more complex, dedicated MARL methods. For example, we propose to implement multi-agent PPO (MAPPO) [33] also with CT-DE, as it has shown great performance in cooperative MARL benchmarks like StarcraftII [24, 29] and Hanabi [5]. Additionally, we will test the role encoder-decoder from ROMA [31] and the role action space from RODE [32], as they reduce primitive action-observation spaces by clustering agents' roles. This will enhance both efficiency and policy generation.

Finally, having multiple instances of the same agent with different individual goals can be formulated as a single agent multi-goal problem, enabling the use of Hindsight Experience Replay [4].

4 Conclusion

Multi-agent demand response is a promising answer to the problem of frequency regulation, which appears in the context of the transition of energy generation means from fossil fuel to renewable sources. We propose to tackle this problem using multi-agent reinforcement learning methods, such as MAPPO with CT-DE, which can cope with the uncertain and dynamic environment, and can be deployed in a scalable way for real-time decision-making. To test and train our method, we are developing a simulator which will be made publicly available.

Once these algorithms are confirmed to work in a simulator, the next step in the deployment of such methods will be to work on multi-agent sim-2-real transfer where the policies trained in simulator are applied to real TCLs, in real buildings.

Acknowledgements

This work was funded by the Institute for Data Valorization (IVADO), by the National Sciences and Engineering Research Council of Canada, as well as Microsoft and Samsung.

References

[1] U.S. Energy Information Administration. Electricity explained - electricity in the united states.

[2] U.S. Environmental Protection Agency. Sources of greenhouse gas emissions.

[3] Maria Alejandra Zuniga Alvarez, Kodjo Agbossou, Alben Cardenas, Sousso Kelouwani, and Loıc Boulon. Demand response strategy applied to residential electric water heaters using dynamic programming and k-means clustering. *IEEE Transactions on Sustainable Energy*, 11(1):524–533, 2019.

[4] Marcin Andrychowicz, Filip Wolski, Alex Ray, Jonas Schneider, Rachel Fong, Peter Welinder, Bob McGrew, Josh Tobin, Pieter Abbeel, and Wojciech Zaremba. Hindsight experience replay, 2018.

[5] Nolan Bard, Jakob N. Foerster, Sarath Chandar, Neil Burch, Marc Lanctot, H. Francis Song, Emilio Parisotto, Vincent Dumoulin, Subhodeep Moitra, Edward Hughes, Iain Dunning, Shibl Mourad, Hugo Larochelle, Marc G. Bellemare, and Michael Bowling. The hanabi challenge: A new frontier for ai research. *Artificial Intelligence*, 280:103216, 2020.

[6] Hassan Bevrani, Arindam Ghosh, and Gerard Ledwich. Renewable energy sources and frequency regulation: survey and new perspectives. *IET Renewable Power Generation*, 4(5):438–457, 2010.

[7] Duncan S Callaway. Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy. *Energy Conversion and Management*, 50(5):1389–1400, 2009.

[8] David P. Chassin, Jason C. Fuller, and Ned Djilali. Gridlab-d: An agent-based simulation framework for smart grids. *Journal of Applied Mathematics*, 2014:1–12, 2014.

[9] AB Gallo, JR Simões-Moreira, HKM Costa, MM Santos, and E Moutinho Dos Santos. Energy storage in the energy transition context: A technology review. *Renewable and sustainable energy reviews*, 65:800–822, 2016.

[10] Sven Gronauer and Klaus Diepold. Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review*, Apr 2021.

[11] Cody A Hill, Matthew Clayton Such, Dongmei Chen, Juan Gonzalez, and W Mack Grady. Battery energy storage for enabling integration of distributed solar power generation. *IEEE Transactions on smart grid*, 3(2):850–857, 2012.

[12] Betelle Memorial Institute. Gridlab-d wiki. `http://gridlab-d.shoutwiki.com/wiki/Main_Page`.

[13] Stephan Koch, Johanna L Mathieu, Duncan S Callaway, et al. Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services. In *Proc. PSCC*, pages 1–7. Citeseer, 2011.

[14] Prabha Kundur. Power system stability. *Power system stability and control*, pages 7–1, 2007.

[15] Antoine Lesage-Landry and Joshua A Taylor. Setpoint tracking with partially observed loads. *IEEE Transactions on Power Systems*, 33(5):5615–5627, 2018.

[16] Antoine Lesage-Landry, Joshua A Taylor, and Duncan S Callaway. Online convex optimization with binary constraints. *IEEE Transactions on Automatic Control*, 2021.

[17] Mingxi Liu and Yang Shi. Model predictive control of aggregated heterogeneous second-order thermostatically controlled loads for ancillary services. *IEEE transactions on power systems*, 31(3):1963–1971, 2015.

[18] Mehdi Maasoumy, Borhan M Sanandaji, Alberto Sangiovanni-Vincentelli, and Kameshwar Poolla. Model predictive control of regulation services from commercial buildings to the smart grid. In *2014 American Control Conference*, pages 2226–2233. IEEE, 2014.

[19] Philipp Moritz, Robert Nishihara, Stephanie Wang, Alexey Tumanov, Richard Liaw, Eric Liang, Melih Elibol, Zongheng Yang, William Paul, Michael I. Jordan, and et al. Ray: A distributed framework for emerging ai applications. *arXiv:1712.05889 [cs, stat]*, Sep 2018. arXiv: 1712.05889.

[20] Peter Palensky and Dietmar Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE transactions on industrial informatics*, 7(3):381–388, 2011.

[21] Eric Pareis and Eric Hittinger. Emissions effects of energy storage for frequency regulation: Comparing battery and flywheel storage to natural gas. *Energies*, 14(3):549, 2021.

[22] David Appleyard GE Power. What every generation executive should know about the impact of ancillary services on plant economics.

[23] Frederik Ruelens, Bert J Claessens, Stijn Vandael, Bart De Schutter, Robert Babuška, and Ronnie Belmans. Residential demand response of thermostatically controlled loads using batch reinforcement learning. *IEEE Transactions on Smart Grid*, 8(5):2149–2159, 2016.

[24] Mikayel Samvelyan, Tabish Rashid, Christian Schröder de Witt, Gregory Farquhar, Nantas Nardelli, Tim G. J. Rudner, Chia-Man Hung, Philip H. S. Torr, Jakob N. Foerster, and Shimon Whiteson. The starcraft multi-agent challenge. *CoRR*, abs/1902.04043, 2019.

[25] Pierluigi Siano. Demand response and smart grids—a survey. *Renewable and sustainable energy reviews*, 30:461–478, 2014.

[26] Goran Strbac. Demand side management: Benefits and challenges. *Energy policy*, 36(12):4419–4426, 2008.

[27] Josh A Taylor, Sairaj V Dhople, and Duncan S Callaway. Power systems without fuel. *Renewable and Sustainable Energy Reviews*, 57:1322–1336, 2016.

[28] Justin K. Terry, Nathaniel Grammel, Ananth Hari, Luis Santos, Benjamin Black, and Dinesh Manocha. Parameter sharing is surprisingly useful for multi-agent deep reinforcement learning. *CoRR*, abs/2005.13625, 2020.

[29] Oriol Vinyals, Timo Ewalds, Sergey Bartunov, Petko Georgiev, Alexander Sasha Vezhnevets, Michelle Yeo, Alireza Makhzani, Heinrich Küttler, John P. Agapiou, Julian Schrittwieser, John Quan, Stephen Gaffney, Stig Petersen, Karen Simonyan, Tom Schaul, Hado van Hasselt, David Silver, Timothy P. Lillicrap, Kevin Calderone, Paul Keet, Anthony Brunasso, David Lawrence, Anders Ekermo, Jacob Repp, and Rodney Tsing. Starcraft II: A new challenge for reinforcement learning. *CoRR*, abs/1708.04782, 2017.

[30] Evangelos Vrettos, Charalampos Ziras, and Göran Andersson. Fast and reliable primary frequency reserves from refrigerators with decentralized stochastic control. *IEEE Transactions on Power Systems*, 32(4):2924–2941, 2016.

[31] Tonghan Wang, Heng Dong, Victor R. Lesser, and Chongjie Zhang. ROMA: multi-agent reinforcement learning with emergent roles. *CoRR*, abs/2003.08039, 2020.

[32] Tonghan Wang, Tarun Gupta, Anuj Mahajan, Bei Peng, Shimon Whiteson, and Chongjie Zhang. RODE: learning roles to decompose multi-agent tasks. *CoRR*, abs/2010.01523, 2020.

[33] Chao Yu, Akash Velu, Eugene Vinitsky, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of ppo in cooperative, multi-agent games, 2021.

[34] Behnam Zakeri and Sanna Syri. Electrical energy storage systems: A comparative life cycle cost analysis. *Renewable and sustainable energy reviews*, 42:569–596, 2015.

[35] Wei Zhang, Jianming Lian, Chin-Yao Chang, and Karanjit Kalsi. Aggregated modeling and control of air conditioning loads for demand response. *IEEE transactions on power systems*, 28(4):4655–4664, 2013.