Combined use of grid reconfiguration and redispatch in congested power systems

1st Arvid Rolander
Electric power & energy systems
KTH - Royal Institute of Technology
Stockholm, Sweden
arvidro@kth.se

2nd Xavier Weiss
Electric power & energy systems
KTH - Royal Institute of Technology
Stockholm, Sweden
xavierw@kth.se

3rd Mohammad Reza Hesamzadeh
Electric power & energy systems
KTH - Royal Institute of Technology
Stockholm, Sweden
mrhesa@kth.se

4th Emil Hillberg
Electric Power Systems
RISE - Research Institutes of Sweden
Göteborg, Sweden
emil.hillberg@ri.se

5th Lars Nordström
Electric power & energy systems
KTH - Royal Institute of Technology
Stockholm, Sweden
larsno@kth.se

***Abstract*—This paper proposes two methods for combining optimization based redispatch and reinforcement learning for grid reconfiguration. A relaxed ACOPF formulation of the redispatch problem is introduced, as well as an iterative method for finding the highest feasible line capacity margin. The performance of the combined redispatch and grid reconfiguration methods is evaluated through a case study using the IEEE 14-bus system. It is found that the combined methods are able to significantly reduce redispatch costs while maintaining maintaining security levels comparable to OPF-based redispatch alone.**

***Index Terms*—Semidefinite programming, deep reinforcement learning, optimal power flow, convex optimization.**

I. Introduction

Increased electrification is widely viewed as a key step toward reducing carbon emissions and reaching world climate goals. In order to accommodate increased utilization, electric grids need to be reinforced, or operated closer to their security limits, which poses new challenges to grid operators [1]. Grid reinforcement is a long and expensive process, with average lead times for constructing new transmission lines being close to ten years in both Europe and the United States [2]. In articles 20-23 of the System Operations (SO) network code [3], the European Network of Transmission System Operators for Electricity (ENTSO-e) together with the Agency for the Cooperation of Energy Regulators (ACER) define the use and categories of remedial actions in the European Union. Remedial actions are manual or automatic control actions that can be used by the system operator to alleviate problems in the grid, such as overloads, voltage problems or congestion, and can free up capacity to be used by actors on the electricity market. Furthermore, the Capacity Allocation & Congestion Management (CACM) [4] specifies how the potential capacity added by remedial actions should be taken into account during capacity allocation.

This work was supported by the project System Integrity Protection Schemes for added transfer capacity, funded by Energimyndigheten under project number P2022-00679.

For this work, special consideration is given to two types of remedial actions specified by the SO network code: redispatch or countertrading and topology control or grid reconfiguration. Redispatch and countertrading are common types of remedial actions for congestion management but can be very costly for the operator. Topological remedial actions are a low cost alternative, but can be challenging to use as the potential grid topologies that need to be considered grow combinatorially with the number of grid components.

Reinforcement learning (RL) has been proposed as a promising method for dealing with the large set of possible topological actions by e.g. [5], [6], but is less suitable for redispatch than traditional convex optimization methods, that come with optimality and feasibility certificates. There is a community studying how RL can be used to find topological remedial actions, illustrated e.g. by the Learn to Run a Power Network (L2RPN) competitions organized by RTÉ [7]. However, there is a gap in the research concerning how different remedial actions as well as solution methods, e.g. machine learning and optimization, can be combined.

A. Contributions

In this work, we propose two methods to combine RL-based grid reconfiguration with optimization based redispatch. We formulate a semidefinite relaxation of the AC redispatch problem based on [8] together with an algorithm for determining the lowest feasible maximum line loading. The performance of the proposed combined methods and redispatch algorithm is evaluated through a case study on the IEEE 14-bus system using Grid2Op [9] as a test platform.

II. Method

A. Redispatch Algorithm

In this section, we give the details of the ACOPF formulation based on [8]. Let U be the complex bus voltage vector and define $\mathbf{X} = \begin{pmatrix}\mathrm{Re}\{U\} & \mathrm{Im}\{U\}\end{pmatrix}^T$. Further, define the set

979-8-3315-8138-1

of buses $\mathcal{N} := \{1, 2, \ldots, n\}$, the set of generator buses $\mathcal{G} \subseteq \mathcal{N}$ and the set of lines $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$. We denote the admittance of line (l, m) by y_{lm} for $(l, m) \in \mathcal{L}$, the shunt element associated with the $\pi-$model of that line with y_{lm}^{shunt}, and the basis vectors of $\mathbb{R}^n$ by e_i. Further, following [8], we define the matrices:

$$Y_k := e_k e_k^T Y \tag{1}$$

$$Y_{lm} := (y_{lm}^{\text{shunt}} + y_{lm}) e_l e_l^T - y_{lm} e_l e_m^T, \tag{2}$$

$$\mathbf{Y}_k := \frac{1}{2} \begin{bmatrix} \operatorname{Re}\left\{Y_k + Y_k^T\right\} & \operatorname{Im}\left\{Y_k^T - Y_k\right\} \\ \operatorname{Im}\left\{Y_k - Y_k^T\right\} & \operatorname{Re}\left\{Y_k + Y_k^T\right\} \end{bmatrix}, \tag{3}$$

$$\mathbf{Y}_{lm} := \frac{1}{2} \begin{bmatrix} \operatorname{Re}\left\{Y_{lm} + Y_{lm}^T\right\} & \operatorname{Im}\left\{Y_{lm}^T - Y_{lm}\right\} \\ \operatorname{Im}\left\{Y_{lm} - Y_{lm}^T\right\} & \operatorname{Re}\left\{Y_{lm} + Y_{lm}^T\right\} \end{bmatrix}, \tag{4}$$

$$\bar{\mathbf{Y}}_k := -\frac{1}{2} \begin{bmatrix} \operatorname{Im}\left\{Y_k + Y_k^T\right\} & \operatorname{Re}\left\{Y_k^T - Y_k\right\}, \\ \operatorname{Re}\left\{Y_k - Y_k^T\right\} & \operatorname{Im}\left\{Y_k + Y_k^T\right\} \end{bmatrix}, \tag{5}$$

$$\bar{\mathbf{Y}}_{lm} := -\frac{1}{2} \begin{bmatrix} \operatorname{Im}\left\{Y_{lm} + Y_{lm}^T\right\} & \operatorname{Re}\left\{Y_{lm} - Y_{lm}^T\right\} \\ \operatorname{Re}\left\{Y_{lm}^T - Y_{lm}\right\} & \operatorname{Im}\left\{Y_{lm} + Y_{lm}^T\right\} \end{bmatrix}, \tag{6}$$

$$M_k := \begin{bmatrix} e_k e_k^T & 0 \\ 0 & e_k e_k^T \end{bmatrix}, \tag{7}$$

$$\tag{8}$$

and the injections $P_{k,inj}$ and $Q_{k,inj}$:

$$P_{k,inj} := P_{G_k} - P_{D_k}, \ \forall k \in \mathcal{G}, \tag{9}$$

$$Q_{k,inj} := Q_{G_k} - Q_{D_k}, \ \forall k \in \mathcal{G}, \tag{10}$$

$$P_{k,inj} := -P_{D_k}, \ \forall k \in \mathcal{N} \setminus \mathcal{G}, \tag{11}$$

$$Q_{k,inj} := -Q_{D_k}, \ \forall k \in \mathcal{N} \setminus \mathcal{G}, \tag{12}$$

Here, P_{G_k}, P_{D_k}, Q_{G_k} and Q_{D_k} represent the active and reactive powers generated and consumed at bus k. As is shown in [8], the following relations hold:

$$P_{k,inj} = \operatorname{Tr}\{\mathbf{Y}_k X X^T\}, \tag{13}$$

$$Q_{k,inj} = \operatorname{Tr}\{\bar{\mathbf{Y}}_k X X^T\}, \tag{14}$$

$$P_{lm} = \operatorname{Tr}\{\mathbf{Y}_{lm} X X^T\}, \tag{15}$$

$$Q_{lm} = \operatorname{Tr}\{\bar{\mathbf{Y}}_{lm} X X^T\}, \tag{16}$$

$$|S_{lm}|^2 = (\operatorname{Tr}\{\mathbf{Y}_{lm} X X^T\})^2 + (\operatorname{Tr}\{\bar{\mathbf{Y}}_{lm} X X^T\})^2, \tag{17}$$

$$|U_k|^2 = \operatorname{Tr}\{M_k X X^T\}. \tag{18}$$

The objective function of the redispatch problem differs from that of the economic-dispatch problem in that the generator active power setpoints are already known, and we instead want to minimize the cost of changing the active power setpoints such that the system satisfies the constraints. To this end, we define the setpoint change $\Delta P_{G_i} \in \mathbb{R}$ with associated monetary cost c_{G_i} of generator i. Using this definition, the redispatch objective function can be defined as:

$$J(\Delta P_G) = \sum_{k=0}^{N} c_{G_k} |\Delta P_{G_k}|, \tag{19}$$

where ΔP_G is the vector of all setpoint changes. Using the setpoint changes, the injected active power at the generator buses can now be rewritten as

$$P_{k,inj} = P_{G_k} + \Delta P_{G_k} - P_{D_k}, \tag{20}$$

where ΔP_{G_k} is the sum of the setpoint changes for the generators at bus k. Using what we have so far and defining $W = XX^T$, we can write out the redispatch problem with constraints as:

$$\underset{\Delta P_{G_k}, X, W}{\text{minimize}} \quad \sum_{k \in \mathcal{G}} c_{G_k} |\Delta P_{G_k}| \tag{21a}$$

subject to

$$\Delta P_{G_k}^{min} \leq \Delta P_{G_k} \leq \Delta P_{G_k}^{max} \ \forall k \in \mathcal{G}, \tag{21b}$$

$$P_{G_k}^{min} \leq \Delta P_{G_k} + P_{G_k} \leq P_{G_k}^{max} \ \forall k \in \mathcal{G}, \tag{21c}$$

$$\operatorname{Tr}\{\mathbf{Y}_k W\} = P_{G_k} + \Delta P_{G_k} - P_{d_k} \ \forall k \in \mathcal{N}, \tag{21d}$$

$$\operatorname{Tr}\{\bar{\mathbf{Y}}_k W\} \leq Q_{G_k}^{max} - Q_{d_k} \ \forall k \in \mathcal{N}, \tag{21e}$$

$$|U_k^{min}|^2 \leq \operatorname{Tr}\{M_k W\} \leq |U_k^{max}|^2 \ \forall k \in \mathcal{N}, \tag{21f}$$

$$(\operatorname{Tr}\{\mathbf{Y}_{lm} W\})^2 + (\operatorname{Tr}\{\bar{\mathbf{Y}}_{lm} W\})^2 \leq S_{lm,max}^2 \ \forall (l, m) \in \mathcal{L}, \tag{21g}$$

$$W = XX^T, \tag{21h}$$

The constraint $W = XX^T$ is equivalent to $\operatorname{rank}\{W\} = 1$, $W \succeq 0$, which is not convex. However, the rank constraint can be relaxed to yield a semidefinite relaxation of the original problem. We also prefer to have a linear objective to conform to solver conventions. The objective can be linearized as

$$J(t) = \sum_i t_i c_{G_i}, \tag{22}$$

subject to

$$t_i \geq \Delta P_{G_i}, \tag{23}$$

$$t_i \geq -\Delta P_{G_i}. \tag{24}$$

Additionally, due to practical concerns, we prefer constraining the maximum current between buses l and m rather than the apparent power flow. By introducing the thermal rating $I_{lm,max}$ and the relation

$$S_{lm,max}^2 = |U_l|^2 I_{lm,max}^2, \tag{25}$$

the maximum power flow constraint can be rewritten as

$$(\operatorname{Tr}\{\mathbf{Y}_{lm} W\})^2 + (\operatorname{Tr}\{\bar{\mathbf{Y}}_{lm} W\})^2 \leq \operatorname{Tr}\{M_l W\} \rho^2 I_{lm,max}^2. \tag{26}$$

Here, we substituted U_l^2 with (18) and introduced ρ to allow an additional safety margin. This constraint is also problematic as it is not quadratic in X, but can be rewritten as[1]

$$\begin{bmatrix} -\operatorname{Tr}\{M_l W\} \rho^2 I_{lm,max}^2 & \operatorname{Tr}\{\mathbf{Y}_{lm} W\} & \operatorname{Tr}\{\bar{\mathbf{Y}}_{lm} W\} \\ \operatorname{Tr}\{\mathbf{Y}_{lm} W\} & -1 & 0 \\ \operatorname{Tr}\{\bar{\mathbf{Y}}_{lm} W\} & 0 & -1 \end{bmatrix} \preceq 0 \tag{27}$$

using the Schur complement formula.

[1]Let $A = -\operatorname{Tr}\{M_l W\} \rho^2 I_{lm,max}^2$, $B = \begin{pmatrix} \operatorname{Tr}\{\mathbf{Y}_{lm} W\} & \operatorname{Tr}\{\bar{\mathbf{Y}}_{lm} W\} \end{pmatrix}$, $C = -I$ where I is the 2×2 identity matrix. By the Schur complement formula, we have $M \preceq 0 \Leftrightarrow C \prec 0, \ A - BC^{-1}B^T \preceq 0$ for $M = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$.

979-8-3315-8138-1

B. Iterative ACOPF Solution

Introducing ρ in (26) adds flexibility by adding the possibility of having a safety margin compared to the actual line rating $I_{lm,max}$. However, there are cases when the desired margin is too tight, resulting in infeasibility, due to e.g. ramp constraints on generators. In such cases, it is desirable to relax the safety margin, which introduces the additional problem of finding the smallest feasible loading ρ^*. To this end, we introduce an iterative solution algorithm given in Algorithm 1, using the desired maximum loading ρ_0 and known feasible maximum loading ρ_f as bounds. The following lemma gives the convergence properties of Algorithm 1:

Algorithm 1 Iterative algorithm for finding lowest feasible loading ρ^* given a desired maximum loading ρ_0 and the base case maximum loading ρ_f

```
ρ_ ← ρ0, ρ̄ ← ρf
Solve relaxed ACOPF given ρ0 as input
if Solution (ΔP*_G|ρ0, W*|ρ0) is found then
    return (ΔP*_G|ρ0, W*|ρ0)
else
    while i ≤ i_max and |ρ̄ − ρ_| > Δρ_max do
        ρi ← (ρ̄ + ρ_)/2
        Solve ACOPF given ρi as input
        if Solution (ΔP*_G|ρi, W*|ρi) is found then
            ρ̄ ← ρi, (ΔP*_G, W*) ← (ΔP*_G|ρi, W*|ρi)
        else
            ρ_ ← ρi
        end if
    end while
    return (ΔP*_G, W*)
end if
```

Lemma 1. *The difference $|\overline{\rho_i} - \underline{\rho_i}|$ between the lower and upper bound for the lowest feasible loading ρ^* after i iterations of Algorithm 1 is $2^{-i}|\overline{\rho_0} - \underline{\rho_0}|$.*

Proof. Consider the cases

1) A solution to the relaxed ACOPF problem was found for input ρ_{i-1},
2) The relaxed ACOPF problem is infeasible for input ρ_{i-1}.

In 1), the bounds $\underline{\rho_i}$ and $\overline{\rho_i}$ can be written $\underline{\rho_i} = \underline{\rho_{i-1}}$, $\overline{\rho_i} = \frac{\overline{\rho_{i-1}}+\underline{\rho_{i-1}}}{2}$ in terms of the bounds from the previous iteration, and similarly in 2) we have $\underline{\rho_i} = \frac{\overline{\rho_{i-1}}+\underline{\rho_{i-1}}}{2}$ and $\overline{\rho_i} = \overline{\rho_{i-1}}$. Calculating the difference between the bounds yields

1) $|\overline{\rho_i} - \underline{\rho_i}| = |\frac{\overline{\rho_{i-1}}+\underline{\rho_{i-1}}}{2} - \overline{\rho_{i-1}}| = \frac{1}{2}|\overline{\rho_{i-1}} - \underline{\rho_{i-1}}|$,
2) $|\overline{\rho_i} - \underline{\rho_i}| = |\frac{\overline{\rho_{i-1}}+\underline{\rho_{i-1}}}{2} - \underline{\rho_{i-1}}| = \frac{1}{2}|\overline{\rho_{i-1}} - \underline{\rho_{i-1}}|$,

that is, the distance between the upper and lower bound is halved every iteration, and the result of the lemma follows. □

C. Reinforcement Learning

The RL-agents used in this work use the Implicit Quantile Network (IQN) architecture introduced in [10], incorporating some improvements of [11], such as N-step learning [12].

1) Action space: In this work, network nodes are modeled as substations with two independent busbars, to which generators, loads and lines can be connected. By choosing which elements are connected to each busbar, the network node can be split into two new nodes, thereby changing power flows. At every timestep, the agent is allowed to choose the configuration of one substation among a set of safe configurations, where safe is taken to mean a configuration where no component is disconnected from the rest of the grid.

2) Observation space: The observation space consists of bus voltage magnitudes and angles, generator outputs, loading of all lines as a fraction of thermal capacity, load active and reactive power consumption, as well as a topology vector which for each component in the grid describes which busbar within its substation it is connected to.

3) Reward function: The per timestep reward is given by

$$r_t = \frac{|\mathcal{L}| - \max(0, \min(|\mathcal{L}|, \sum_{i\in\mathcal{L}} \rho_i))}{|\mathcal{L}|} \tag{28}$$

where $\mathcal{L}$ is the set of lines and ρ_i is the loading of line i as a fraction of its thermal capacity.

4) Training setup: Most timesteps in the training and evaluation scenarios do not have any congestion and do not require any remedial actions. In order to facilitate learning, agents are therefore only allowed to act during timesteps where at least one line in the network is loaded above an activation threshold denoted ρ_c. To provide a stronger learning signal for agents, rewards are allowed to accumulate between activations, so that the effective reward seen by the agent at the current timestep t_c is given by

$$r_{t_c}^{eff} = \sum_{t=t_c}^{t_c'} r_t, \tag{29}$$

where t_c' is the next timestep where the agent is allowed to act and r_t given by (28).

D. Combined DRL and ACOPF algorithm

There are several ways in which one could combine the DRL agents with a redispatch algorithm. The methods we consider here are:

- static method,
- adaptive method.

In the static method, redispatch is always applied after the DRL agent has taken an action, provided at least one line is operating above its thermal limit. In the adaptive method, four outcomes are simulated:

- only topological action,
- only redispatch action,
- combined topological and redispatch action,
- no intervention,

and the one yielding the lowest maximum thermal loading is chosen. One could also consider other approaches such as training the DRL agent with the optimizer as a shield, similar to [13]. However, there are some issues with such an approach, e.g. that training times increase and that the DRL

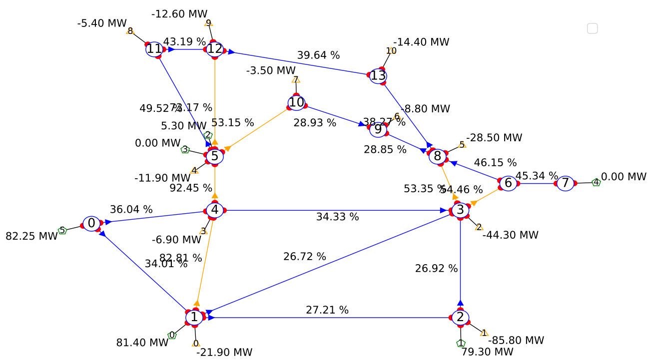

Fig. 1. Visualization of the grid used in the case study.

agent could become overly reliant on the optimizer. Redispatch and simulation is always performed based on an imperfect one-step ahead forecast of the grid, and ramp rates are adjusted to reflect the available ramp of the generators based on the output difference between the current state and the forecast.

It is worth noting that since the target redispatch is calculated based on an imperfect forecast, there will be some discrepancy between the calculated dispatch and the realized dispatch. Because the RL-agents are only allowed to act if a line is loaded above ρ_c, the choice of ρ_0 in Algorithm 1 can affect how many actions the agents take which in turn can affect final performance.

III. Case Study

In order to test the performance of the algorithm under realistic conditions, we perform a case study using the IEEE 14-bus system and the Grid2Op platform [9]. Grid2Op is a platform for modeling sequential decision making in power systems, providing a number of power system models and different scenarios and a way to model decision making in the power system as a Markov Decision process (MDP) [12]. The grid is illustrated in Fig. 1. The grid consists of a 138kV high voltage part and a 20kV low voltage part, with transformers between substations 4 and 5 and substations 3 and 8, as well as an auxiliary 14kV part comprising substations 6 and 7.

For the case study, we use 100 test scenarios that were not seen by the reinforcement learning agent during training. The scenarios are one month long and divided into 5-minute timesteps, and consist of timeseries for loads and generators, including realistic seasonal variations.

During the course of a scenario, some rules are enforced:

- If a line is above 100% of its thermal rating for more than 4 timesteps, it is automatically disconnected.
- If a line is above 200% of its thermal rating, it is immediately disconnected.
- A tripped line cannot be reconnected for 10 timesteps.
- If a generator or load is disconnected from the rest of the grid (e.g. due to a line being disconnected), the scenario is ended early.
- If any islands are created, the scenario is ended early.
- If Grid2Op's power flow solver is not able to find a solution for any reason, the scenario is ended early.

In addition to these rules, a static value of $\rho_c = 0.95$ is used, meaning remedial actions are performed if at least one line is loaded above 95% of its thermal capacity.

For tests using DRL agents, results are aggregated over three agents that were trained with different random seeds.

A. Results

Fig. 2 shows a violin plot of the mean episode duration grouped by the methods described in Sec. II-D, as well as a no-intervention policy for comparison. Additionally, results are split based on the target loading ρ_0 constraining the redispatch. Both OPF and RL on their own outperformed the no-intervention benchmark by a wide margin, and using the more conservative value $\rho_0 = 0.94$ let the OPF based method fully complete all scenarios. For the combined methods, Fig. 2 shows slight survivability improvements when using the static method with $\rho_0 = 0.95$ compared to RL on its own, while the adaptive method performed even better, with half as many prematurely terminated scenarios. Decreasing the target loading to $\rho_0 = 0.94$ further improved survivability for all methods using OPF, resulting in only two and one prematurely terminated scenarios, for the passive and adaptive methods respectively.

It should be noted here that since redispatch is calculated based on an imperfect forecast, there will be small discrepancies between the predicted and realized outcomes when using OPF. Therefore, choosing a target loading equal to the activation threshold for the RL-agent could cause unwanted behavior, where the redispatch is too conservative to prevent the RL-agent from acting during the next timestep. This could lead to the intended action sequence of the RL-agent no longer being valid under the new grid state, leading to suboptimal action sequences. Adding the extra safety margin during redispatch could effectively prevent this from happening, since the added margin would prevent the RL-agent from acting during the next timestep.

Fig. 3 takes a different perspective, showing per episode redispatch costs normalized by scenario duration. This shows a substantial cost reduction when using the combined RL and OPF methods compared to just using redispatch on its own, with approximately an order of magnitude difference in mean cost over all scenarios. Moreover, Fig. 3 shows a wider distribution of per episode costs when using the static method compared to the adaptive method, resulting in a slightly higher mean cost. Interestingly, the mean final cost is not necessarily increased when using the more conservative target value $\rho_0 = 0.94$. In summary, the results of this case study show that combining redispatch with topological remedial actions can be a valid approach, as it can maintain security while significantly reducing costs compared to only using redispatch.

979-8-3315-8138-1

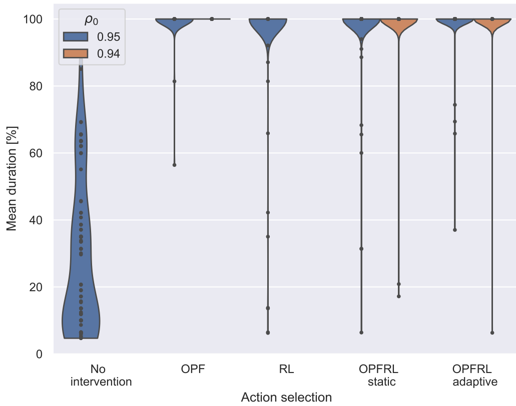

Fig. 2. Performance of the different methods for combining topology reconfiguration and redispatch in terms of relative scenario duration, including the no intervention benchmark as well as pure OPF and pure RL. Results are shown for two choices the target loading ρ_0, and the inside of the violins show individual datapoints.

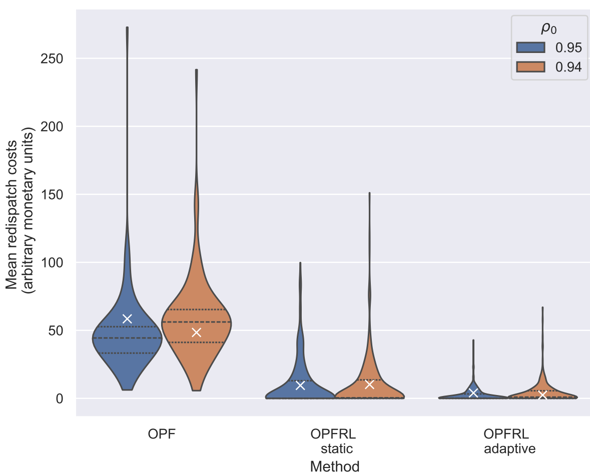

Fig. 3. Per scenario redispatch costs normalized by scenario duration when using only redispatch and RL combined with redispatch. The insides of the violins shows dashed quartile lines and crosses indicating means.

IV. Conclusions

In this paper, we formulate an AC version of the redispatch problem as a semidefinite program, together with a search algorithm for minimally relaxing tight line loading constraints. We formulate two methods of combining redispatch optimization with reinforcement learning based topology switching, and evaluate the methods performance in a case study on the IEEE 14-bus system using Grid2Op. We find that combined redispatch optimization and topology reconfiguration is a viable method to solve grid congestion, with both combined methods reaching almost the same performance as redispatch optimization from a security perspective, while reducing mean redispatch costs by up to an order of magnitude.

A. Future work

As mentioned in Sec. III-A, some unwanted behavior seems to arise when using the same value for the activation threshold ρ_c as the target loading ρ_0, most likely due to the redispatch being based on an imperfect forecast. It is possible that performance could be further improved by using an adaptive activation threshold. Thus, a promising route for further research could be to investigate what would be suitable criteria for remedial action activation. While traditional OPF already has real world applications, the question of how to scale RL-based topology switching is still open, although efforts have been made to address it in e.g. the L2RPN competitions. Importantly, future studies addressing scalability could also consider realistic modeling of substation topology, as this can be used to determine upper bounds on action space cardinality as well as to determine which contingencies can effectively be solved through topological remedial actions.

References

[1] IEA, "Electrification." https://www.iea.org/energy-system/electricity/electrification, 2024. [Accessed 2024-04-23].

[2] IEA, "Average lead times to build new electricity grid assets in Europe and the United States, 2010-2021 – Charts – Data & Statistics." https://www.iea.org/data-and-statistics/charts/average-lead-times-to-build-new-electricity-grid-assets-in-europe-and-the-united-states-2010-2021.

[3] "Commission Regulation (EU) 2017/1485 of 2 August 2017 establishing a guideline on electricity transmission system operation (Text with EEA relevance.)," Aug. 2017. Legislative Body: ENER, COM.

[4] "Commission Regulation (EU) 2015/1222 of 24 July 2015 establishing a guideline on capacity allocation and congestion management (Text with EEA relevance)," July 2015. Legislative Body: COM.

[5] M. Subramanian, J. Viebahn, S. H. Tindemans, B. Donnot, and A. Marot, "Exploring grid topology reconfiguration using a simple deep reinforcement learning approach," Apr. 2021. arXiv:2011.13465.

[6] M. Dorfer, A. R. Fuxjäger, K. Kozak, P. M. Blies, and M. Wasserer, "Power Grid Congestion Management via Topology Optimization with AlphaZero," Nov. 2022. arXiv:2211.05612 [cs].

[7] A. Marot, B. Donnot, G. Dulac-Arnold, A. Kelly, A. O'Sullivan, J. Viebahn, M. Awad, I. Guyon, P. Panciatici, and C. Romero, "Learning to run a Power Network Challenge: a Retrospective Analysis," Oct. 2021. arXiv:2103.03104 [cs].

[8] J. Lavaei and S. H. Low, "Zero Duality Gap in Optimal Power Flow Problem," *IEEE Transactions on Power Systems*, vol. 27, pp. 92–107, Feb. 2012.

[9] B. Donnot, "Grid2op- A testbed platform to model sequential decision making in power systems.," 2020.

[10] W. Dabney, G. Ostrovski, D. Silver, and R. Munos, "Implicit Quantile Networks for Distributional Reinforcement Learning," June 2018. arXiv:1806.06923.

[11] M. Hessel, J. Modayil, H. v. Hasselt, T. Schaul, G. Ostrovski, W. Dabney, D. Horgan, B. Piot, M. Azar, and D. Silver, "Rainbow: Combining Improvements in Deep Reinforcement Learning," Oct. 2017. arXiv:1710.02298.

[12] R. S. Sutton and A. G. Barto, *Reinforcement Learning, Second Edition: An Introduction*. Cambridge, UNITED STATES: MIT Press, 2018.

[13] X. Weiss, S. Mohammadi, P. Khanna, M. R. Hesamzadeh, and L. Nordström, "Safe Deep Reinforcement Learning for Power System Operation under Scheduled Unavailability," in *2023 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, July 2023. ISSN: 1944-9933.