Master thesis proposal

Viktor Todosijevic

August 2023

Introduction

The optimal power flow problem (OPF) is a nonlinear, non-convex optimization problem[1, 2] whose algebraic formulation is dictated by a graph, specifically the graph representation of a power grid. The most basic formulation of the problem consists of three distinct parts:

- the objective function - an economic cost function of the active powers produced by the generators in the grid and the corresponding costs of running those generators.
- the equality constraints - representing the physical law of conservation of energy.
- the inequality constraints - representing the technical limitations of running the grid which dictate that the voltages and powers of specific nodes(buses) must be within some predefined ranges.

The problem boils down to minimizing the operating cost while fulfilling the physical and technical constraints. In practice, grid operators need to solve the OPF problem for thousands of variations of the node parameters and even for variations of the underlying topology of the grid. For example, beginning with house owners or office buildings wanting to install heat pumps or a rooftop photovoltaic array, to larger entities installing wind and solar farms, each of these events constitutes a change of the underlying graph structure for which a large number of OPF calculations need to be run to ensure that the stability of the grid and technical constraints are maintained. Today, these calculations are performed with solvers, like the interior point method, whose run time is slow and optimality of the solution isn't guaranteed[1]. Graph machine learning methods can't improve the optimality guarantees, but they can improve the run time by orders of magnitude by learning a heuristic that approximately solves the OPF problem. The mathematical formulation, keeping in mind that $\underline{\boldsymbol{V}} = \boldsymbol{V} e^{j\boldsymbol{\theta}}$, is the following:

$$\begin{aligned} \min_{\boldsymbol{P}} \quad & f(\boldsymbol{P}; \boldsymbol{c}) \\ \text{s.t.} \quad & \boldsymbol{P} + j\boldsymbol{Q} = \underline{\boldsymbol{V}} \cdot (\underline{\boldsymbol{Y}}\underline{\boldsymbol{V}})^* \\ & \\ & \boldsymbol{V}_{min} \leq \boldsymbol{V} \leq \boldsymbol{V}_{max} \\ & \boldsymbol{P}_{min} \leq \boldsymbol{P} \leq \boldsymbol{P}_{max} \\ & \boldsymbol{Q}_{min} \leq \boldsymbol{Q} \leq \boldsymbol{Q}_{max} \end{aligned} \tag{1}$$

where, for a grid consisting of N nodes, $\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{V}, \boldsymbol{\theta} \in \mathbb{R}^N$ are vectors containing the active and reactive powers and voltage magnitudes and angles, respectively, for each node and $\boldsymbol{P}_{min}, \boldsymbol{P}_{max}, \boldsymbol{Q}_{min}, \boldsymbol{Q}_{max}, \boldsymbol{V}_{min}, \boldsymbol{V}_{max} \in \mathbb{R}^N$ are the corresponding technical constraints of the physical variables. The matrix $\underline{\boldsymbol{Y}} \in \mathbb{C}^{N \times N}$ is called the admittance matrix of the grid and represents a weighted adjacency matrix of the graph because the entry $\underline{\boldsymbol{Y}}_{ij}$ is equal to zero where there is no connection between nodes i and j, and contains physical properties of the line between the two otherwise. The function f parameterized by $\boldsymbol{c}$ is an arbitrary cost function but is often a polynomial i.e.

$$f(\boldsymbol{P};\boldsymbol{c}) = \sum_{i=0}^{N-1}\sum_{k=0}^{n} c_{ik}P_i^k$$

where n is the degree of the polynomial.

Of the values in the vectors $\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{V}$ and $\boldsymbol{\theta}$, which, along with the admittance matrix, fully describe the state of a grid, not all are variables in the optimization problem. In fact, some are fixed, and which value is fixed and which isn't is determined on the node level by the type of node the value belongs to. Nodes can be one of three types: Slack(reference), PQ and PVAR. **??** contains the binary masks for each node type and value where 1 indicates a fixed quantity and 0 indicates a variable in the optimization problem.

Node type	P	Q	V	θ
Slack	0	0	1	1
PQ	1	1	0	0
PVAR	0	0	0	0

Table 1: Value mask in relation to the node type.

Literature overview

This is a literature overview.

- Refine the already existing literature overview.

Methodology

This is the methodology.

- To correct for the lackings of the existing literature I will work on a fixed split public dataset, add interpretable metrics etc.
- I still wanna use GNNs for their ability to adapt to multi-topology data but to improve their performance I plan to use something that extends their reach like graph transformers, maybe jhop, probably not k-WL and the likes unless it can incorporate node and edge attribute
- Could add the definitions of the metrics I want to use in order to clarify the quality of the results, that will raise the page count and is also useful.

Conclusion

This is the conclusion.

Bibliography

[1] Mary B Cain, Richard P O'neill, Anya Castillo, et al. "History of optimal power flow and formulations". In: *Federal Energy Regulatory Commission* 1 (2012), pp. 1–36.

[2] Daniel Bienstock and Abhinav Verma. "Strong NP-hardness of AC power flows feasibility". In: *Operations Research Letters* 47.6 (2019), pp. 494–501.