

---

# Machine Learning for AC Optimal Power Flow

---

Neel Guha<sup>1</sup> Zhecheng Wang<sup>2</sup> Arun Majumdar<sup>2</sup>

## Abstract

We explore machine learning methods for AC Optimal Powerflow (ACOPF) - the task of optimizing power generation in a transmission network according while respecting physical and engineering constraints. We present two formulations of ACOPF as a machine learning problem: 1) an *end-to-end prediction* task where we directly predict the optimal generator settings, and 2) a *constraint prediction* task where we predict the set of active constraints in the optimal solution. We validate these approaches on two benchmark grids.

## 1. Introduction

The Optimal Power Flow problem (OPF) consists of determining the optimal operating levels for different generators within a transmission network in order to meet the demand that is changing over space and time. An established area of research in both power systems and operations, OPF is applied every day in the management and regulation of power grids around the world. In this work, we hope to obtain real-time approximate solutions to the OPF problem using machine learning.

The classical formulation of ACOPF (presented in Section 3) is a challenging non-convex and NP-hard problem (Bienstock & Verma, 2015). In addition to minimizing generator costs, solutions must adhere to physical laws governing power flow (i.e. Kirchhoff's voltage law) and respect the engineering limits of the grid. As a result, ACOPF is computationally intractable under the demands of daily grid management. In order to account for rapid fluctuations in power demand and supply, grid operators must solve ACOPF over the entire grid (comprising of tens of thousands of nodes) every five minutes<sup>1</sup> (Cain et al., 2012). Most traditional approaches (genetic algorithms, convex relaxations, etc) either fail to converge within this time frame

<sup>1</sup>Carnegie Mellon University <sup>2</sup>Stanford University. Correspondence to: Neel Guha <neelguha@gmail.com>.

*Presented at the Climate Change Workshop at ICML 2019*

<sup>1</sup>The addition of renewable sources of energy (wind, solar, etc) adds more unpredictability and is a motivation for improved techniques for ACOPF

or produce suboptimal solutions. In order to practically manage the grid, operators solve a linearized version of ACOPF practice known as DC Optimal Power Flow (DCOPF). However, DCOPF presents a number of issues. True grid conditions can deviate from the linear assumptions imposed by DCOPF, increasing the likelihood of instability and grid failure (Frank & Rebennack, 2016). Relying on DCOPF also has significant implications for climate change. A 2012 report from the Federal Energy Regulatory Commission estimated that the inefficiencies induced by approximate-solution techniques may cost billions of dollars and release unnecessary emissions (Cain et al., 2012). Determining an efficient solution for ACOPF could also be adapted to *combined economic emission dispatch* (CEED) - a variant of OPF which incorporates a per-generator emissions cost into the classic objective function (Venkatesh et al., 2003).

In this paper, we observe that it should be possible to learn a model that can predict an accurate solution over a fixed grid topology/constraint set. Intuitively, we expect some measure of consistency in the solution space - similar load distributions should correspond to similar generator settings. This suggests an underlying structure to the ACOPF problem, which a machine learning model can exploit.

Machine learning present several advantages. Neural networks have demonstrated the ability to model extremely complicated non-convex functions, making them highly attractive for this setting. A model could be trained off-line on historic data and used in real-time to make predictions on an optimal power setting. In this work, we explore two applications of machine learning for OPF:

**1. End-to-end:** Train a model to directly predict the optimal generator setting for a given load distribution. This is challenging, as the model's output must be adherence with physical laws/engineering limits.

**2. Constraint prediction:** Train a model to predict which constraints are active (i.e at equality) in the optimal solution. Knowing this active set can be used to warm start existing approaches (i.e. interior point methods) and reduce solution time.

## 2. Related Work

Prior work has explored different applications of machine learning on the grid. This includes work on estimating active constraints for DCOPF (Ng et al., 2018; Misra et al., 2018), predicting grid failures (Rudin et al., 2012), or choosing between traditional solvers (King et al., 2015). Machine learning has also been applied to related variants of the OPF problem, including automated grid protection (Donnot et al., 2017), price proxy prediction (Canyasse et al., 2016), or private information recovery (Donti et al.). To the extent of our knowledge, there has been limited work on direct applications of deep learning towards ACOPF.

## 3. Method

We now present the traditional ACOPF problem, and describe how to formalize it as a machine learning task (Frank et al., 2012). For a fixed grid topology  $\mathcal{G}$ , let  $\mathbf{N}$  denote the set of buses (nodes),  $\mathbf{L}$  denote the set of branches (edges), and  $\mathbf{G} \subseteq \mathbf{N}$  denote the set of controllable generators. For bus  $i$ , we enumerate  $P_i^G$  (real power injection),  $Q_i^G$  (reactive power injection),  $P_i^L$  (real power demand),  $Q_i^L$  (reactive power demand),  $V_i$  (voltage magnitude), and  $\delta_i$  (voltage angle). the power demand at AC OPF can be framed as:

$$\underset{P_i^G}{\text{minimize}} \quad \sum_{i \in \mathbf{G}} C_i(P_i^G) \quad (1a)$$

subject to

$$P_i(V, \delta) = P_i^G - P_i^L, \quad \forall i \in \mathbf{N}, \quad (1b)$$

$$Q_i(V, \delta) = Q_i^G - Q_i^L, \quad \forall i \in \mathbf{N}, \quad (1c)$$

$$P_i^{G,\min} \leq P_i^G \leq P_i^{G,\max}, \quad \forall i \in \mathbf{G}, \quad (1d)$$

$$Q_i^{G,\min} \leq Q_i^G \leq Q_i^{G,\max}, \quad \forall i \in \mathbf{G}, \quad (1e)$$

$$V_i^{\min} \leq V_i \leq V_i^{\max}, \quad \forall i \in \mathbf{N}, \quad (1f)$$

$$\delta_i^{\min} \leq \delta_i \leq \delta_i^{\max}, \quad \forall i \in \mathbf{N} \quad (1g)$$

Where (1a) typically represents a polynomial cost function, (1b)-(1c) corresponds to the power flow equations, and (1d)-(1g) represent operational limits on real/reactive power injections, nodal voltage magnitude, and nodal voltage angles<sup>2</sup> respectively. More recent settings of OPF - including ours- also include limits on branch currents. These are outlined in more detail by Frank et al. (2012). We now present two formalizations of AC OPF as a machine learning problem. In our setting, we assume that  $P_i^L$  and  $Q_i^L$  (real and reactive demand) are known across all  $N$  buses.

<sup>2</sup>A single reference bus ("slack" bus) is fixed to  $\tilde{V} = 1.0\angle 0$

### 3.1. End-to-end Prediction

In this setting, we pose the AC OPF problem as a regression task, where we predict the grid control variables ( $P_i^G$  and  $V_i^G$ ) from the grid demand ( $P_i^L$  and  $Q_i^L$ ). These fix a set of equations with equal number of unknowns, which can be solved to identify the remaining state values for the grid. Formally, given a dataset of  $n$  solved grids with load distributions  $\mathbf{X} = \{[P_0^L, \dots, P_N^L, Q_0^L, \dots, Q_N^L]\}_{i=1}^n$  and corresponding optimal generator settings  $\mathbf{Y} = \{[P_0^G, \dots, P_G^G, V_0^G, \dots, V_G^G]\}_{i=1}^n$ , our goal is to learn  $f_\theta : \mathcal{X} \rightarrow \mathcal{Y}$  which minimizes the mean-squared error between the optimal generator settings  $\mathbf{Y}$  and the predicted generator settings  $\tilde{\mathbf{Y}}$ . Solving for the remaining state variables can be posed as a *power flow* problem, and reduces to finding  $V_i^L$ ,  $Q_i^G$ , and  $\delta_i$  such that (1b)-(1g) are satisfied.

The central challenge in this setting is ensuring that the neural network's solution respects physical laws and engineering limits. Though provable guarantees may be difficult to make, we experiment by incorporating soft penalties into our loss function that encourage predictions to fall within legal limits. These correspond to linear penalties that activate when (1d) and (1f) are violated. In future work we hope to explore more sophisticated (and robust) techniques for enforcing legality.

### 3.2. Optimal Constraint Prediction

Given that neural networks may learn solutions that violate physical constraints, and are thus untrustworthy in practical settings, we explore *optimal constraint prediction* as formulated by Misra et al. (2018). In this setting, our model is trained to predict the set of constraints that are active in the optimal solution for some load distribution. A constraint is active if the corresponding state/control variable is at the maximum or minimum allowed value. As Misra et al. (2018) describe, knowing the active set of constraints can be used to warm start a more traditional optimization method, and reduce time to convergence.

Formally, for each grid we define a constraint vector  $y \in \mathbb{R}^{2G+2N}$  corresponding to an enumeration of constraints (1d)-(1e), where  $y_i = 1$  if the  $i$ -the constraint is active in the optimal solution, and  $y_i = 0$  otherwise. We learn  $f_\theta$  which maps from the load distribution  $[P_0^L, \dots, P_N^L, Q_0^L, \dots, Q_N^L]$  to this constraint vector. This corresponds to a multi-label classification problem.

Optimal constraint prediction presents several advantages over end-to-end prediction.

1. **Solver Speedup:** From an optimization perspective, knowing the set of active constraints equates to warm-starting, and can significantly speed-up more traditional

algorithms like interior point methods, active set methods, simplex methods, and others. Quantifying this speedup is the focus of ongoing work.

2. **Reliability:** This setting reduces the risk of a neural network producing a solution which violates physical laws/engineering limits. Because the physical and engineering constraints are enforced by the solver, an incorrect prediction will at worst increase solution time or lead to a suboptimal solution. In the end-to-end setting described in Section 3.1, incorrect predictions could destabilize the grid.
3. **Task complexity:** Classifying the set of active constraints is significantly easier than predicting a set of real valued targets.

## 4. Results

We validated approaches for end-to-end prediction and constraint prediction on IEEE 30-bus<sup>3</sup> and 118-bus test cases<sup>4</sup>. These test cases include predetermined constraints.

### 4.1. Dataset Generation

The IEEE test cases include a pre-calculated load distribution (denoted as  $x^*$ ). In order to construct a dataset for each case, we repeatedly sample candidate load distributions  $x' \sim \text{Uniform}((1 - \delta) \cdot x^*, (1 + \delta) \cdot x^*)$ , for some fixed  $\delta$ . We identify  $y'$  by solving the OPF problem for  $x'$  via Matpower (Zimmerman et al., 2011). In some cases, the solver fails to converge, suggesting that the sampled  $x'$  has no solution given the grid constraints. In this case, we discard  $x'$ .

We generated 95000 solved grids for case118 and 812888 solved grids for case30 with  $\delta = 0.1$  (a 10% perturbation to the IEEE base demand). Interestingly, we observe that while 100% of the samples generated for case118 were successfully solved, only 81.2% of the samples for case30 were successfully solved. For all prediction tasks, we used a 90/10 train-test split and report results on the test set.

### 4.2. End to end prediction

We evaluate task performance along two metrics:

- **Legality Rate:** The proportion of predicted grids which satisfy all engineering and physical constraints.
- **Avg. Cost Deviation:** The average fractional difference

between the cost of the predicted grid, and the cost of the true grid:  $\frac{1}{n} \sum_i^n |1 - \frac{\text{pred cost}_i}{\text{true cost}_i}|$  over legal grids.

Roughly, this captures the reliability and optimality of a particular model. We examine a range of different architectures and training strategies. We performed a grid search considering models with 1-2 hidden layers, 128/256/512 hidden neurons, ReLU/Tanh activations. We also experimented with vanilla MSE loss, and a variant with linear penalties for constraint violations (described in Section 3.1). Each model was trained with Adam ( $\text{lr} = 0.001$ ) until loss convergence, for a maximum of 2000 epochs.

Grid	Legality Rate	Avg. Cost Deviation
case30	0.51	0.002
case118	0.70	0.002

Table 1. End-to-end prediction performance. Average cost deviation is only reported for legal grids.

Table 1 reports the best performance for each grid type. For case30, the optimal model was a two layer neural network with tanh activations, and no loss penalty. For case118, the optimal model was a three layer network with 512 hidden neurons, ReLU activations, and a constraint loss penalty. Interestingly, we observe better performance on case118 than case30. Though we would intuitively expect task difficulty to scale with grid size, this result suggests that other factors could affect a model’s generalization ability. In particular, smaller grids could be less stable, and thus more likely to produce a wide range of (less predictable) behavior under varying demand distributions. We also observe that the cost of the optimal model predictions were within 1% of the optimal cost.

### 4.3. Constraint Prediction

For constraint prediction, we evaluate performance in terms of accuracy (i.e. the proportion of constraints classified successfully). We perform a similar hyperparameter grid search and report the best results in Table 2.

Grid	% Accuracy
case30	0.99
case118	0.81

Table 2. Constraint prediction performance

<sup>3</sup><https://electricgrids.enr.tamu.edu/electric-grid-test-cases/ieee-30-bus-system/>

<sup>4</sup><https://electricgrids.enr.tamu.edu/electric-grid-test-cases/ieee-118-bus-system/>

In general, we find neural networks to be highly successful at determining which the active constraint set.

## 5. Conclusion

In this work, we presented two approaches that leverage machine learning for solving ACOPF. Preliminary experiments present promising results in both settings. In next steps, we hope to evaluate our methods on more complex grid architectures, and explore different approaches for incorporating grid constraints into our models.

## References

Bienstock, D. and Verma, A. Strong np-hardness of ac power flows feasibility. *arXiv preprint arXiv:1512.07315*, 2015.

Cain, M. B., O’neill, R. P., and Castillo, A. History of optimal power flow and formulations. *Federal Energy Regulatory Commission*, pp. 1–36, 2012.

Canyasse, R., Dalal, G., and Mannor, S. Supervised learning for optimal power flow as a real-time proxy. *CoRR*, abs/1612.06623, 2016. URL <http://arxiv.org/abs/1612.06623>.

Donnot, B., Guyon, I., Schoenauer, M., Panciatici, P., and Marot, A. Introducing machine learning for power system operation support. *arXiv preprint arXiv:1709.09527*, 2017.

Donti, P. L., Azevedo, I. L., and Kolter, J. Z. Inverse optimal power flow: Assessing the vulnerability of power grid data.

Frank, S. and Rebennack, S. An introduction to optimal power flow: Theory, formulation, and examples. *IIE Transactions*, 48(12):1172–1197, 2016.

Frank, S., Rebennack, S., et al. A primer on optimal power flow: Theory, formulation, and practical examples. *Colorado School of Mines, Tech. Rep*, 2012.

King, J. E., Jupe, S. C., and Taylor, P. C. Network state-based algorithm selection for power flow management using machine learning. *IEEE Transactions on Power Systems*, 30(5):2657–2664, 2015.

Misra, S., Roald, L., and Ng, Y. Learning for constrained optimization: Identifying optimal active constraint sets. *arXiv preprint arXiv:1802.09639*, 2018.

Ng, Y., Misra, S., Roald, L. A., and Backhaus, S. Statistical learning for dc optimal power flow. In *2018 Power Systems Computation Conference (PSCC)*, pp. 1–7. IEEE, 2018.

Rudin, C., Waltz, D., Anderson, R. N., Boulanger, A., Salleb-Aouissi, A., Chow, M., Dutta, H., Gross, P. N., Huang, B., Jerome, S., et al. Machine learning for the new york city power grid. *IEEE transactions on pattern analysis and machine intelligence*, 34(2):328–345, 2012.

Venkatesh, P., Gnanadass, R., and Padhy, N. P. Comparison and application of evolutionary programming techniques to combined economic emission dispatch with line flow constraints. *IEEE Transactions on Power systems*, 18(2):688–697, 2003.

Zimmerman, R. D., Murillo-Sánchez, C. E., and Thomas, R. J. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on power systems*, 26(1):12–19, 2011.