Using AI to Speed Up Traditional Power Systems Optimization Problems

Presented by Dr. Kyri Baker

Associate Professor

University of Colorado Boulder, Fellow of the Renewable and Sustainable Energy Institute

Why use AI for Power Systems Optimization?

The use cases are endless when you shift the computational burden offline to make a model that's faster than a conventional solver (run more scenarios! Solve ACOPF in real time! Provide better warm-starts to Gurobi!)

Combining Deep Learning and Optimization for Security-Constrained Optimal Power Flow

DC3: A LEARNING METHOD FOR OPTIMIZATION WITH HARD CONSTRAINTS

LEARNING WARM-START POINTS FOR AC OPTIMAL POWER FLOW

CANOS: A Fast and Scalable Neural AC-OPF Solver Robust To N-1 Perturbations

Learning Optimal Solutions for Extremely Fast AC Optimal Power Flow

A Multi-Stage Warm-Start Deep Learning Framework for Unit Commitment

Learning to accelerate distributed ADMM using graph neural networks

GridSFM: A new, small foundation model for the electric grid

Learning to optimize meets neural-ODE: Real-time, stability-constrained AC OPF

Bottlenecks and Open Questions

Bottleneck 1: Even if you can solve a (AC OPF? SCOPF?) problem in 1 millisecond, you are solving **steady-state** power flows. We can't apply those actions every 1 millisecond anyway. Extremely fast OPF is good for running many scenarios simultaneously, but can we actually use their solutions in **real-time?**

Bottleneck 2: Diagnosing infeasibilities/nonconvergence is often one of the most time-consuming tasks for an actual grid operator, but a lot of these models are solving **cases that are too easy** and **do not provide a way of handling infeasible system states.**

There are more bottlenecks and open questions but I don't want to give away alllll my research ideas ;)

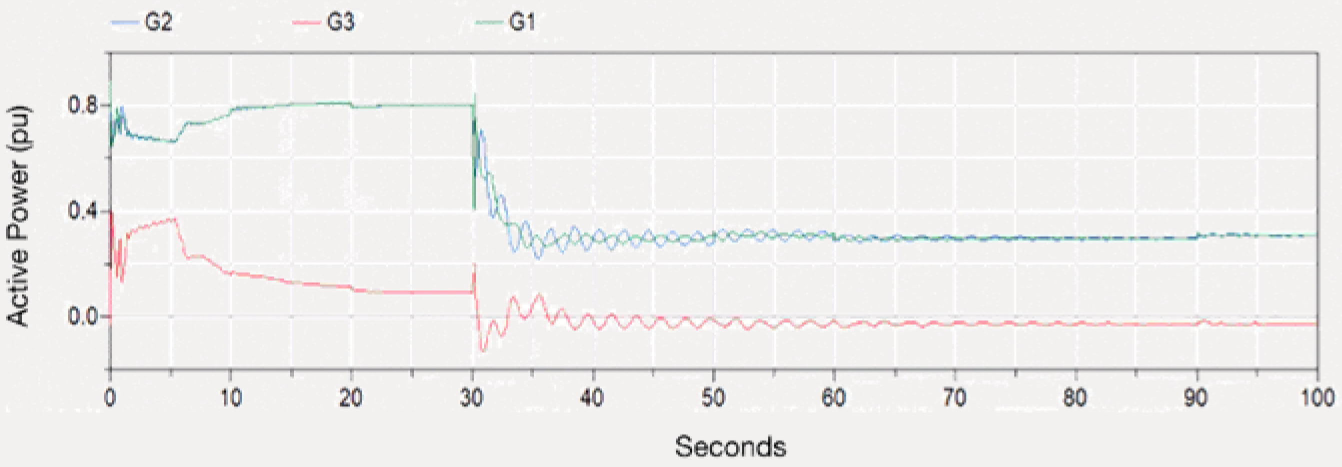
Non-Technical Bottleneck: Getting Grid Operators to adopt these AI-based solutions!

We have the *inputs* on fast timescales, but AC OPF is steady-state…

Most SCADA systems return measurements on a 2-second scale. PMUs are 100x+ faster, but measurements are more scattered

SCADA and PMUs measure voltage, current, real and reactive power, but phase angles must be calculated from a state estimator[5] – we just need the real and reactive demand at each bus

If we have these every 1-2 seconds, we are now exiting the validity of the steady-state assumption …



Modelica 9-bus dynamic simulation where generator setpoints are changed too quickly, resulting in unstable rotor angles

[5]M. Rice and G. Heydt, Phasor Measurement Unit Data in Power System State Estimation, PSERC report, 2005

Comparing other learning-for-OPF methods...

Changing generator setpoints too rapidly with high magnitudes of change may cause instability, even if it's "optimal"

We generated random loading samples in the 57-bus network and tried three of these methods:

Method	Metric							
	MSE (p^g) $\times 10^{-3}$	MSE (q^g) $\times 10^{-3}$	MSE ($\lvert V \rvert$) $\times 10^{-3}$	MSE (θ) $\times 10^{-3}$	Flow Vio. $\times 10^{-3}$	Boundary Vio. $\times 10^{-4}$	**Stability Vio.** $\mathbf{\times 10^{3}}$	Optimality Gap $\times 10^{-1}$
MSE	3.48 ± 0.321	3.86 ± 1.512	0.77 ± 0.148	1.42 ± 0.237	12.65 ± 2.281	6.44 ± 1.434	$\mathbf{2.33 \pm 0.206}$	1.66 ± 0.243
DC3	3.31 ± 0.579	6.74 ± 0.580	0.51 ± 0.078	0.64 ± 0.081	0.00	0.00	$\mathbf{2.86 \pm 0.232}$	1.64 ± 0.049
LD	3.97 ± 0.279	3.52 ± 2.427	0.34 ± 0.012	0.95 ± 0.054	6.23 ± 0.125	0.00	$\mathbf{2.31 \pm 0.219}$	1.68 ± 0.125

Here, we're defining dynamical stability in terms of generator rotor angle stability, as defined in the classical machine model. Other measures of stability can also be incorporated.

Despite low constraint violations and optimality gaps, stability is not guaranteed

Introducing… Dynamic OPF Net (DynOPF-Net)

Learning optimal (AC OPF) generator setpoints while better ensuring stability

Method	Metric							
	MSE (p^g) $\times 10^{-3}$	MSE (q^g) $\times 10^{-3}$	MSE ($\|V\|$) $\times 10^{-4}$	MSE (θ) $\times 10^{-4}$	Flow Vio. $\times 10^{-3}$	Boundary Vio. $\times 10^{-4}$	**Stability Vio. $\times 10^{3}$**	Optimality Gap $\times 10^{-1}$
MSE	1.90 ± 0.272	1.65 ± 1.018	0.32 ± 0.153	0.43 ± 0.149	10.45 ± 2.183	9.72 ± 4.93	$\mathbf{2.26 \pm 0.189}$	1.29 ± 0.008
DC3	1.86 ± 0.217	1.56 ± 0.326	0.26 ± 0.195	0.48 ± 0.343	0.00	0.00	$\mathbf{2.45 \pm 0.205}$	1.17 ± 0.006
LD	1.77 ± 0.163	1.72 ± 0.248	0.16 ± 0.099	0.55 ± 0.051	7.19 ± 0.425	0.00	$\mathbf{2.13 \pm 0.175}$	1.12 ± 0.008
DynOPF-Net	2.41 ± 0.253	3.18 ± 1.273	2.55 ± 0.354	3.82 ± 0.924	8.32 ± 0.596	0.410 ± 0.243	**0.00**	1.32 ± 0.032

Slightly higher optimality gaps but significantly fewer stability issues

How does it work…? →

We want to now solve the Stability-Constrained AC OPF

Model 1 The AC Optimal Power Flow Problem (AC-OPF)

variables: $S_i^g, V_i \;\; \forall i \in \mathcal{N}, \;\; S_{ij} \;\; \forall (i,j) \in \mathcal{L}$

minimize: $$\sum_{i \in \mathcal{G}} c_{2i}(\mathrm{Re}(S_i^g))^2 + c_{1i}\,\mathrm{Re}(S_i^g) + c_{0i} \quad (1a)$$

subject to:

$$v_i^l \le |V_i| \le v_i^u \quad \forall i \in N \quad (1b)$$

$$-\theta_{ij}^{\Delta} \le \angle(V_i V_j^*) \le \theta_{ij}^{\Delta} \quad \forall (i,j) \in \mathcal{L} \quad (1c)$$

$$S_i^{gl} \le S_i^g \le S_i^{gu} \quad \forall i \in \mathcal{N} \quad (1d)$$

$$|S_{ij}| \le s_{ij}^u \quad \forall (i,j) \in \mathcal{L} \quad (1e)$$

$$S_i^g - S_i^d = \textstyle\sum_{(i,j)\in L} S_{ij} \quad \forall i \in \mathcal{N} \quad (1f)$$

$$S_{ij} = Y_{ij}^* |V_i|^2 - Y_{ij}^* V_i V_j^* \quad \forall (i,j) \in \mathcal{L} \quad (1g)$$

$$\theta_{\text{ref}} = 0 \quad (1h)$$

Model 2 The Stability Constrained AC-OPF Problem

variables: $S_i^g, V_i \;\; \forall i \in \mathcal{N}, \;\; \delta^g, \omega^g \;\; \forall i \in \mathcal{G}, \;\; S_{ij} \;\; \forall (i,j) \in \mathcal{L}$

minimize: $$\sum_{i \in \mathcal{G}} c_{2i}(\mathrm{Re}(S_i^g))^2 + c_{1i}\,\mathrm{Re}(S_i^g) + c_{0i} \quad (9a)$$

subject to:

$$(1b) - (1h) \quad (9b)$$

$$\frac{d\delta^g(t)}{dt} = \omega_s(\omega^g(t) - \omega_s) \quad \forall g \in \mathcal{G} \quad (9c)$$

$$\frac{d\omega^g(t)}{dt} = \frac{1}{m^g}\left(p_m^g - d^g(\omega^g(t) - \omega_s)\right) - \frac{e_q'^g(0)|V_g|}{x_d'^g m^g}\sin(\delta^g(t) - \theta_g) \quad \forall g \in \mathcal{G} \quad (9d)$$

$$\frac{e_q'^g(0)|V_g|\sin(\delta^g(0) - \theta_g)}{x_d'^g} - p^g = 0 \quad \forall g \in \mathcal{G} \quad (9e)$$

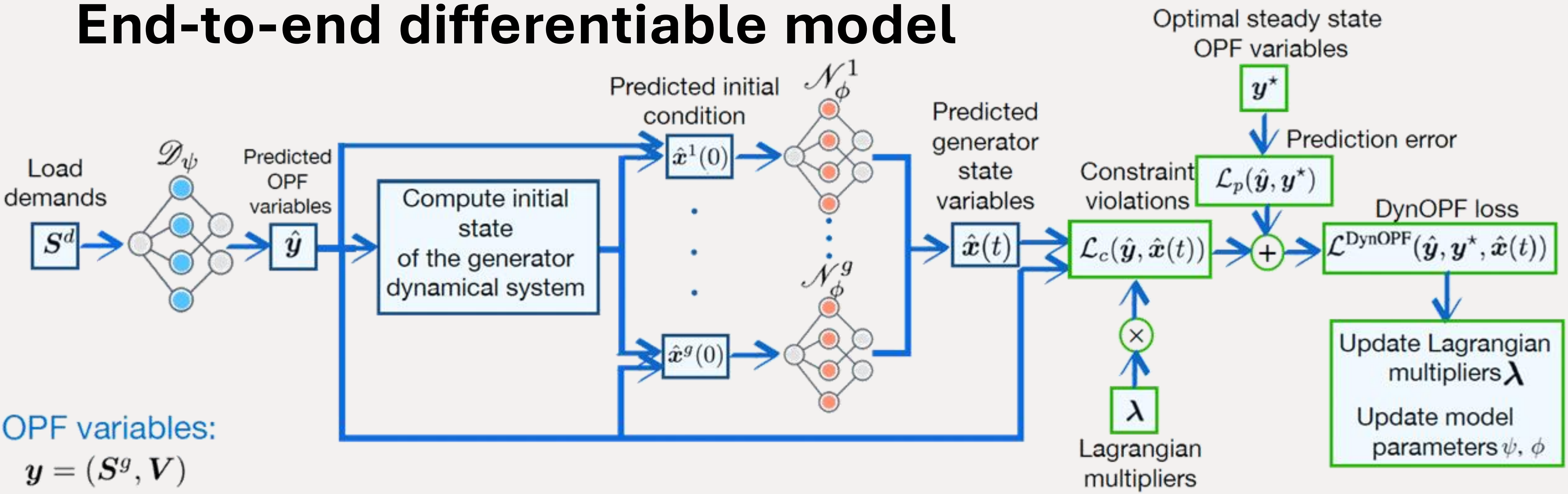
$$\frac{e_q'^g(0)|V_g|\cos(\delta^g(0) - \theta_g) - |V_g|^2}{x_d'^g} - q^g = 0 \quad \forall g \in \mathcal{G} \quad (9f)$$

$$\omega^g(0) = \omega_s \quad \forall g \in \mathcal{G} \quad (9g)$$

$$\delta^g(t) \le \delta^{\max} \quad \forall g \in \mathcal{G}. \quad (9h)$$

Synchronous Generator Dynamic Model (Single Machine Equivalent – SIME)

End-to-end differentiable model



OPF variables:

$y = (S^g, V)$

Generator state variables:

$x^g(t) = (\delta^g(t), \omega^g(t))$

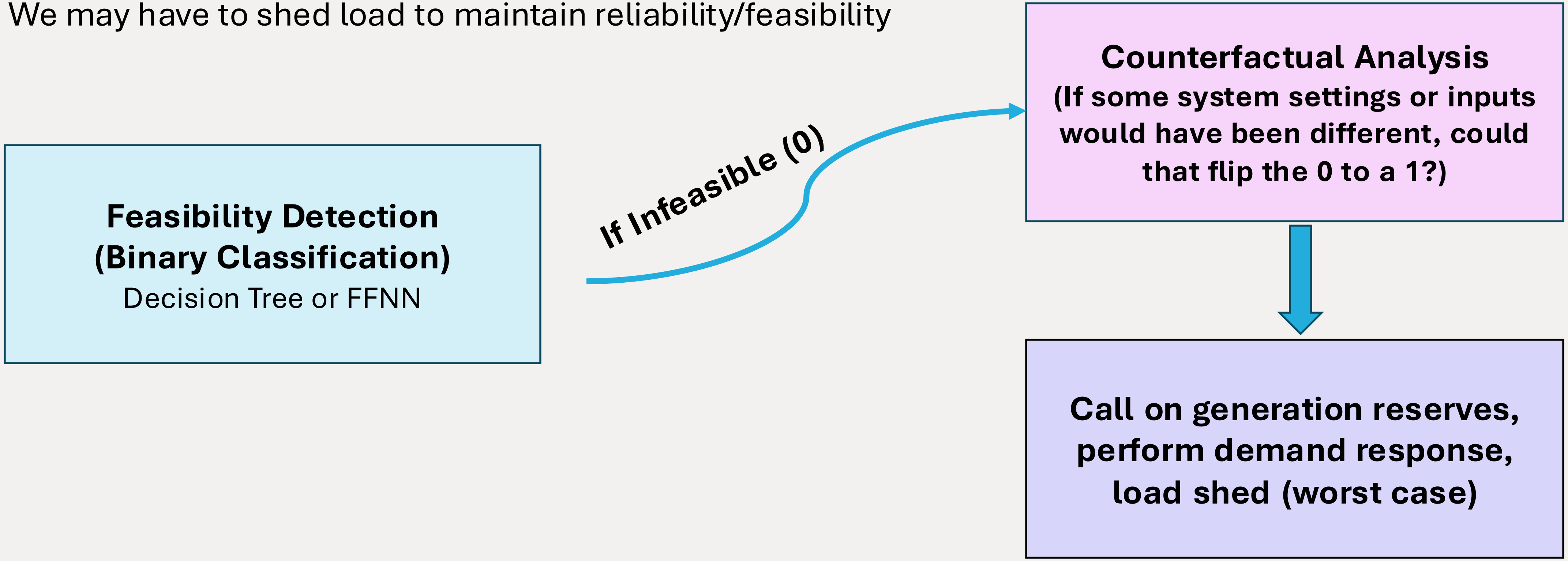
(green box) represents operations performed only at training time.

- Leverages feed-forward NN architecture for OPF, Neural ODE architecture for dynamical stability
- Constraints are incorporated into the loss function via a Lagrangian Relaxation based approach

Challenge 2: Infeasible Problems

→ Sometimes in power systems, infeasibility actually means we don't have enough generation capacity
→ Or lines or transformers are overloaded

→ We may have to shed load to maintain reliability/feasibility



A suite of options that can push us to feasibility

Isn’t there an “optimal”?

$$\min_{\boldsymbol{x},\boldsymbol{s}} \; f(\boldsymbol{x}) + g(\boldsymbol{s})$$

$$h(\boldsymbol{x}) \leq \boldsymbol{s}$$

$$\mathrm{s} \geq \mathbf{0}$$

Isn’t the ‘optimal’ decision the one which minimizes the amount of slack added to the original problem?

Not necessarily.

Isn't there an "optimal"?

$$\min_{\boldsymbol{x},\boldsymbol{s}} \; f(\boldsymbol{x}) + g(\boldsymbol{s})$$

$$h(\boldsymbol{x}) \leq \boldsymbol{s}$$

$$\text{s} \geq \boldsymbol{0}$$

Isn't the 'optimal' decision the one which minimizes the amount of slack added to the original problem?

Not necessarily.

NATIONAL

California power outages highlight economic disparity

by: CATHY BUSSEWITZ, Associated Press
Posted: Oct 13, 2019 / 12:19 PM EDT
Updated: Oct 13, 2019 / 12:19 PM EDT

This Sunday, Oct. 6,
brother Colton Awal
Read More

SHARE

This is an archived article and the information in the article may be outdated. Please look at the time stamp on the st
to see when it was last updated.

When the nation's largest utility warned customers that it would cut power to nearly 2 million people across Northern California, many rushed out to buy portable generators, knowing the investment could help sustain them during blackouts.

The "optimal" way to load shed, for example, may hurt more communities more than others

Mathematical optimality $\neq$ Social optimality

(Also, for large systems, adding hundreds or thousands of new variables/constraints is not ideal!)

The Counterfactual

What modifications to the load vector will result in the classifier prediction changing from a 0 to a 1?

Set of *k* counterfactuals (load/gen modification vectors)

$$C(\boldsymbol{x}) = \arg \underbrace{\min_{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_k}}\Big(\frac{1}{k}\sum_{i=1}^{k} \mathrm{yloss}(f(\boldsymbol{c}_i), y) + \frac{\lambda_1}{k}\sum_{i=1}^{k} \mathrm{dist}(\boldsymbol{c}_i, \boldsymbol{x}) - \lambda_2 \, \mathrm{dpp\_diversity}(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k)\Big)$$

We can generate a *set* of possible counterfactuals that an operator can choose from

(should I shed load at bus 34? Should I bring online additional generation at bus 12?)

Explainable AI

The Counterfactual

What modifications to the load vector will result in the classifier prediction changing from a 0 to a 1?

Set of k counterfactuals (load/gen modification vectors)

$$C(\boldsymbol{x}) = \arg\min_{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_k} \Big(\frac{1}{k} \sum_{i=1}^{k} \mathrm{yloss}(f(\boldsymbol{c}_i), y) + \frac{\lambda_1}{k} \sum_{i=1}^{k} \mathrm{dist}(\boldsymbol{c}_i, \boldsymbol{x}) - \lambda_2 \,\mathrm{dpp\_diversity}(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k) \Big)$$

The loss function here describes how effectively the counterfactuals change the classification from a 0 to a 1

f is a differentiable model (e.g. our classifier)

The Counterfactual

What modifications to the load vector will result in the classifier prediction changing from a 0 to a 1?

Set of *k* counterfactuals (load/gen modification vectors)

$$C(\boldsymbol{x}) = \arg\min_{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_k} \Big(\frac{1}{k}\sum_{i=1}^{k} \text{yloss}(f(\boldsymbol{c}_i), y) + \frac{\lambda_1}{k}\sum_{i=1}^{k} \text{dist}(\boldsymbol{c}_i, \boldsymbol{x}) - \lambda_2 \,\text{dpp\_diversity}(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k) \Big)$$

A distance metric that penalizes too much change from our original feature vector ***x***

We don't want to change the system more than we have to!

The Counterfactual

What modifications to the load vector will result in the classifier prediction changing from a 0 to a 1?

Set of *k* counterfactuals (load/gen modification vectors)

$$C(\boldsymbol{x}) = \arg\min_{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_k} \Big(\frac{1}{k}\sum_{i=1}^{k} \text{yloss}(f(\boldsymbol{c}_i), y) + \frac{\lambda_1}{k}\sum_{i=1}^{k} \text{dist}(\boldsymbol{c}_i, \boldsymbol{x}) - \lambda_2 \, \text{dpp\_diversity}(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k) \Big)$$

A diversity metric that promotes a wider *variety* of counterfactuals to yield diverse options (rather than a bunch of counterfactuals that are slight perturbations of each other)

$\text{dpp\_diversity} = \det(\boldsymbol{K})$ where $\boldsymbol{K}_{i,j} = \frac{1}{1 + \text{dist}(\boldsymbol{c}_i, \boldsymbol{c}_j)}$

Example scenario

	0	1	2	3	4	5	6	7	8	9	Total load reduction or gen increase
CF Option 1	-	-	-	-	27.630195	-	-	3.7318735	-	-	**31.36 MW**
CF Option 2	-	-	-	-	41.550417	-	-	-	-	-	**41.55 MW**
CF Option 3	-	-	-	-	17.093741	-	-	-	9.58467	-	**26.67 MW**

Hmm…option 3 results in the lowest change, but customers at bus 8 have experienced a lot of outages lately.

I could call on a lot of demand response for the industrial customer at bus 4, or use CF Option 1 and call on a smaller amount of DR for customer 4 and use some storage reserves at bus 7.

→ Any of these options will help keep the lights on!

Conclusion

The community has done a great job with the foundation work for solving AC OPF quickly

But there are a lot of other challenging power systems optimization problems to be solved

And not just the pretty ones which make the results look good, but the hard ones: the ones close to Jacobian singularities, the ones with no feasible solution to begin with, the ones where we're close to voltage collapse…

It's time to move on to the hard ones now that we've proven that AI can be used to solve optimizations quickly

Thank you!

Questions?