WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Optimal Power Flow Sensitivities

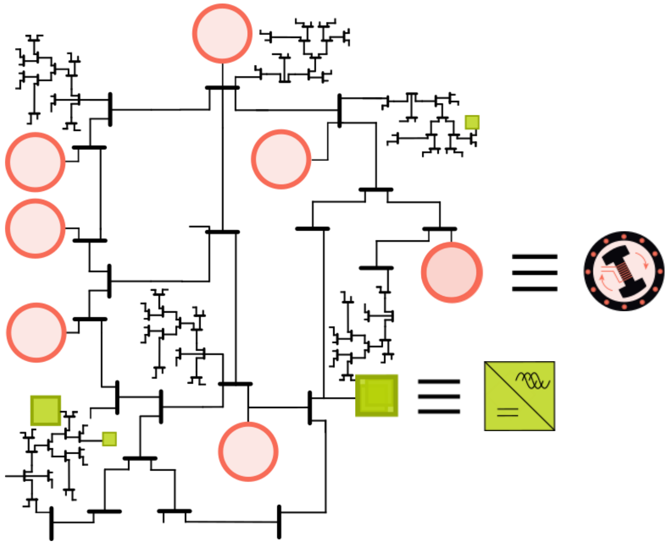
Analysis & Applications

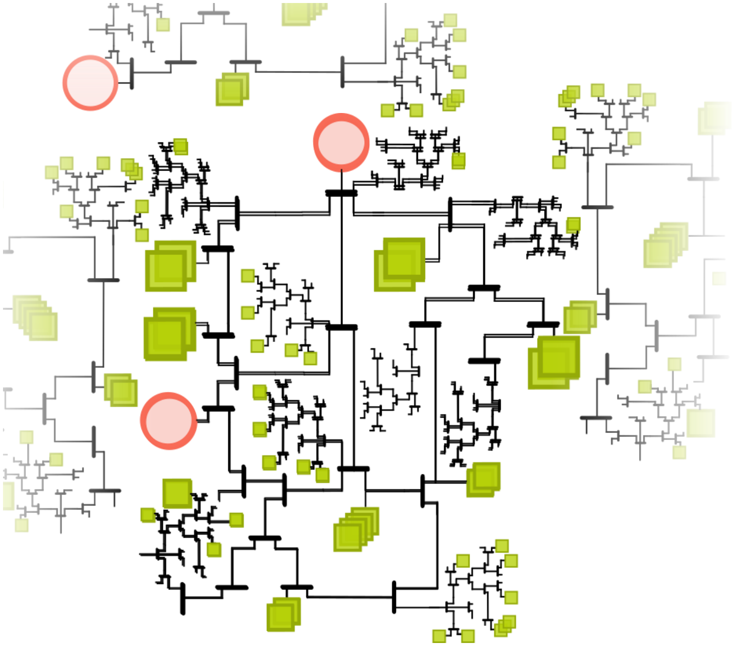
Manish K. Singh

manish.singh@wisc.edu

Sep 6, 2024

Motivation: Power-system transition



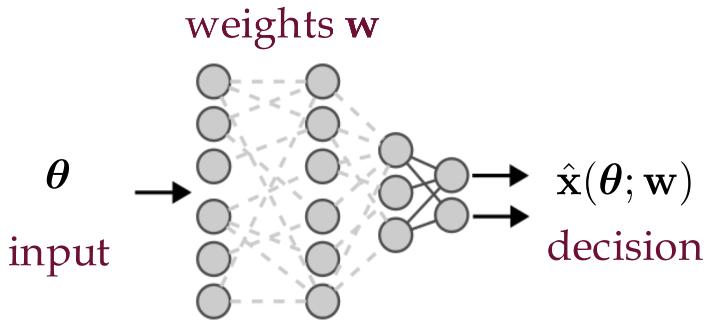


Increasing renewables, DERs, prosumers…

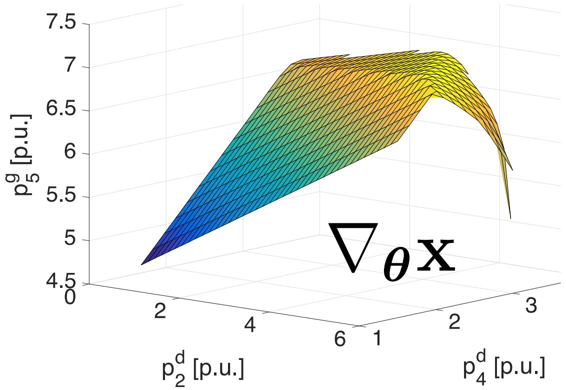
Need to solve problems *faster* over a *broader range* of operating conditions…

Outline

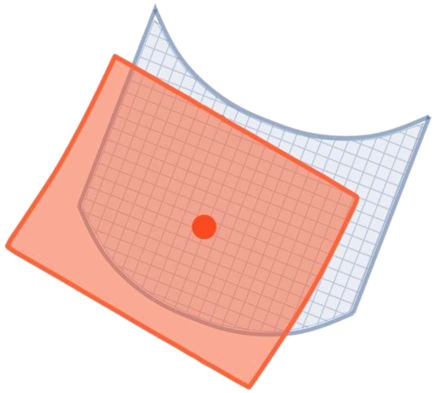
$$\mathbf{x}(\boldsymbol{\theta}) = \arg\min_{\mathbf{x}} \; f(\mathbf{x};\boldsymbol{\theta}) \quad \text{s.to} \;\; \mathbf{g}(\mathbf{x};\boldsymbol{\theta}) \leq \mathbf{0}$$



Learning to optimize



Sensitivity analysis



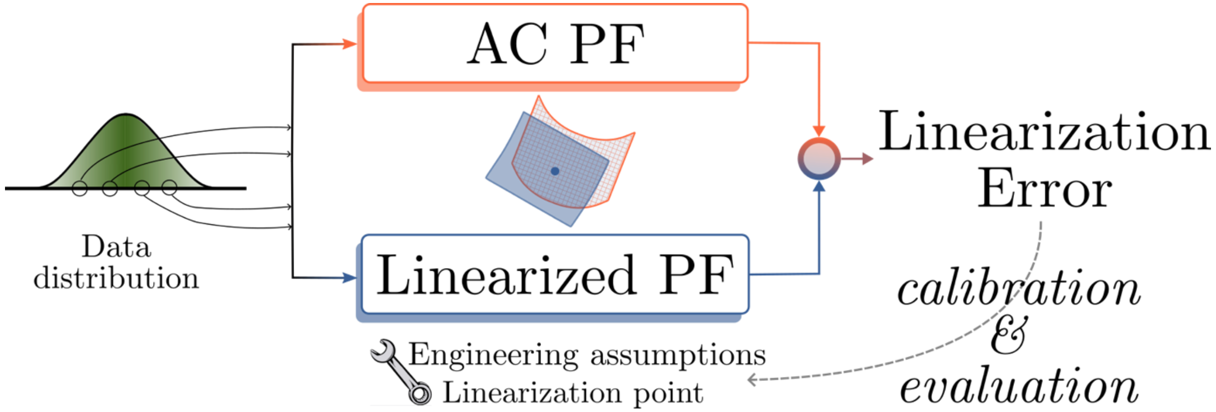
Power flow linearization

"Life can only be understood backwards, but it must be lived forwards"

Søren Kierkegaard

Today's talk

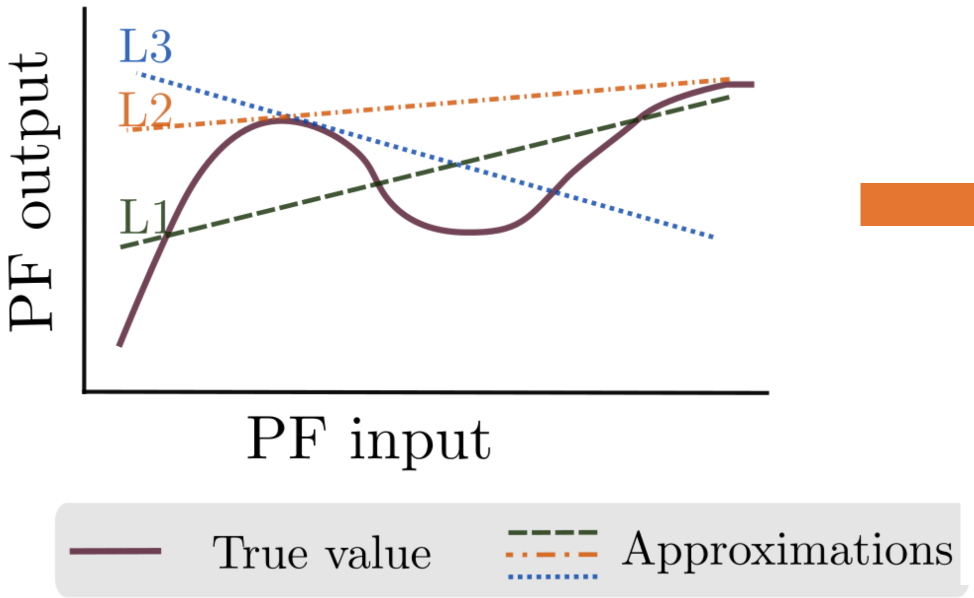
Power flow linearization

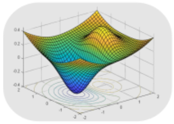


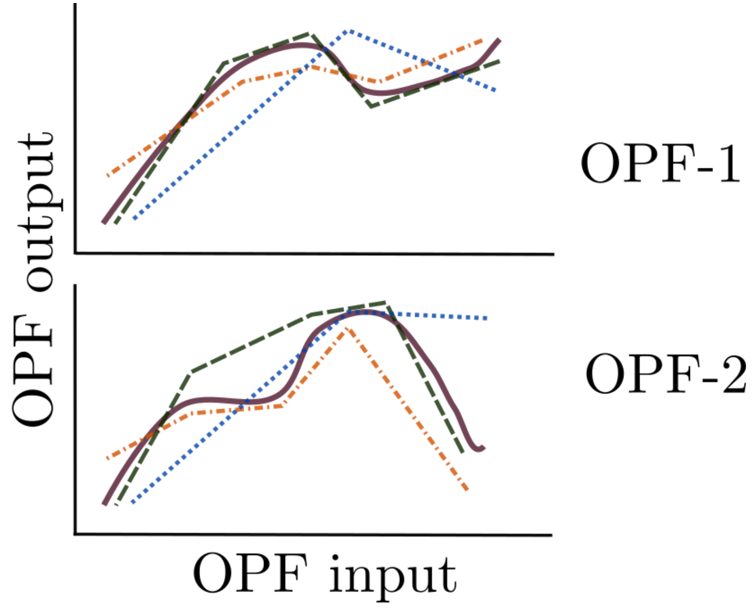
Outstanding contributions from Saverio, Andrey, Kyri, and several others…

A Survey of Relaxations and Approximations of the Power Flow Equations

Daniel K. Molzahn[1] and Ian A. Hiskens[2]







Optimal linearization for optimization

$$\begin{aligned} \mathbf{x}^\star(\boldsymbol{\theta}) = \arg\min_{\mathbf{x}} \ & \mathbf{x}^\top \mathbf{C}\mathbf{x} && \text{(OPF1)} \\ \text{s.to} \ & \mathbf{x}^\top \mathbf{P}_j \mathbf{x} + \mathbf{q}_j^\top \mathbf{x} = \mathbf{r}_j^\top \boldsymbol{\theta}, \quad j = 1:J \\ & \mathbf{S}\mathbf{x} + \mathbf{T}\boldsymbol{\theta} \leq \mathbf{u} \end{aligned}$$

$$\begin{aligned} \hat{\mathbf{x}}^\star(\boldsymbol{\theta}) = \arg\min_{\mathbf{x}} \ & \mathbf{x}^\top \mathbf{C}\mathbf{x} && \text{(OPF2)} \\ \text{s.to} \ & \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{R}\boldsymbol{\theta} \\ & \mathbf{S}\mathbf{x} + \mathbf{T}\boldsymbol{\theta} \leq \mathbf{u} \end{aligned}$$

- Example of an application-informed metric

$$\ell(\mathbf{A}, \mathbf{b}) = \mathbb{E}_{\boldsymbol{\theta}\sim p_{\boldsymbol{\theta}}} \|\mathbf{x}^\star(\boldsymbol{\theta}) - \hat{\mathbf{x}}^\star(\boldsymbol{\theta})\|_2^2 \approx \frac{1}{S}\sum_{s=1}^{S} \|\mathbf{x}^\star(\boldsymbol{\theta}_s) - \hat{\mathbf{x}}^\star(\boldsymbol{\theta}_s)\|_2^2$$

- Obtain optimal linearization as

$$(\mathbf{A}^\star, \mathbf{b}^\star) = \arg \min_{(\mathbf{A},\mathbf{b})\in\mathcal{A}} \ell(\mathbf{A}, \mathbf{b}; \boldsymbol{\theta}_s, \mathbf{x}^\star(\boldsymbol{\theta}_s))$$ bilevel optimization

- Gradient descent-based optimization

$\nabla_{\hat{\mathbf{x}}^\star} \ell$ — easy

$\nabla_{\mathbf{A},\mathbf{b}} \hat{\mathbf{x}}^\star$ — sensitivity analysis presented for $\nabla_{\boldsymbol{\theta}} \hat{\mathbf{x}}^\star$

Optimal power flow sensitivities

- Consider a parametric OPF setting

For QPs, see MPP

$$\mathbf{x}(\boldsymbol{\theta}) = \arg\min_{\mathbf{x}} \quad f(\mathbf{x};\boldsymbol{\theta}) \qquad \text{s.to} \quad \mathbf{g}(\mathbf{x};\boldsymbol{\theta}) \leq \mathbf{0} : \quad \boldsymbol{\lambda}$$

KKT →

$$\nabla_{\mathbf{x}} f + (\nabla_{\mathbf{x}}\mathbf{g})^\top \boldsymbol{\lambda} = \mathbf{0}$$
$$\mathbf{g} \odot \boldsymbol{\lambda} = \mathbf{0}$$
$$\mathbf{g} \leq \mathbf{0}, \quad \boldsymbol{\lambda} \geq \mathbf{0}$$

- Perturb KKT to find point $(\mathbf{x}+\mathrm{d}\mathbf{x}, \boldsymbol{\lambda}+\mathrm{d}\boldsymbol{\lambda})$ optimal for $\boldsymbol{\theta}+\mathrm{d}\boldsymbol{\theta}$ [Castillo-Conejo+'06]

- Unless $(\mathbf{x}, \boldsymbol{\lambda}; \boldsymbol{\theta})$ is degenerate, perturbed point satisfies KKT inequalities too

- Compute total differential of two equality constraints $\boldsymbol{\lambda} \odot (\nabla_{\mathbf{x}}\mathbf{g} \cdot \mathrm{d}\mathbf{x} + \nabla_{\boldsymbol{\theta}}\mathbf{g} \cdot \mathrm{d}\boldsymbol{\theta}) + \mathbf{g} \odot \mathrm{d}\boldsymbol{\lambda} = \mathbf{0}$

- Obtain linear system $\mathbf{S}\begin{bmatrix}\mathrm{d}\mathbf{x}\\ \mathrm{d}\boldsymbol{\lambda}\end{bmatrix} = \mathbf{U}\,\mathrm{d}\boldsymbol{\theta}$ → find $(\nabla_{\boldsymbol{\theta}}\mathbf{x}, \nabla_{\boldsymbol{\theta}}\boldsymbol{\lambda})$ if S^{-1} exists

Q) When is $\mathbf{S}$ invertible? What if it is not?

Existence of sensitivities

- *When?* certain second-order optimality conditions and LICQ holds [Fiacco'76]

> Linear independence constraint qualification (LICQ) requires $\{\nabla_{\mathbf{x}} g_i \mid \lambda_i > 0\}$ to be linearly independent

- LICQ often fails for inverter dispatch posed as a quadratic program [Singh et al'20]
- *What then?* For QPs, multiparametric programming can still provide $\nabla_{\boldsymbol{\theta}} \mathbf{x}$
- For general non-convex OPF, the system $\mathbf{S} \begin{bmatrix} \mathrm{d}\mathbf{x} \\ \mathrm{d}\boldsymbol{\lambda} \end{bmatrix} = \mathbf{U}\, \mathrm{d}\boldsymbol{\theta}$, if consistent, has a unique solution in $\mathrm{d}\mathbf{x}$
- *Happy news:* When S is not invertible, Jacobian $\nabla_{\boldsymbol{\theta}} \mathbf{x}$ may still exist

Singh, Kekatos, Giannakis, "Learning to Solve the AC-OPF using Sensitivity-Informed Deep Neural Networks," *IEEE Trans. Power Sys.*, 2021

Sensitivity analysis takeaways

- Sensitivities of optimal primal/dual variable w.r.t. any parameter can be computed for *(non) convex OPF*

Given a minimizer, one can compute sensitivities for optimization problems just by solving a linear system

Can we do sensitivity analysis when using convex relaxations?

- Sensitivity computation requires optimal primal and dual variables, and cost/constraint functions
- When exact, relaxations yield true optimal primal variables, but different dual variables! ☹
- *More happy news:* We can use SDP or SOCP relaxations (when exact) to compute sensitivities for non-convex problems

Singh, Kekatos, Giannakis, "Learning to Solve the AC-OPF using Sensitivity-Informed Deep Neural Networks," *IEEE Trans. Power Sys.*, 2021

Jalali, Singh, Kekatos, Giannakis, Liu, "Fast Inverter Control by Learning the OPF mapping using Sensitivity Informed Gaussian Processes," *IEEE Trans. Smart Grids*, 2022

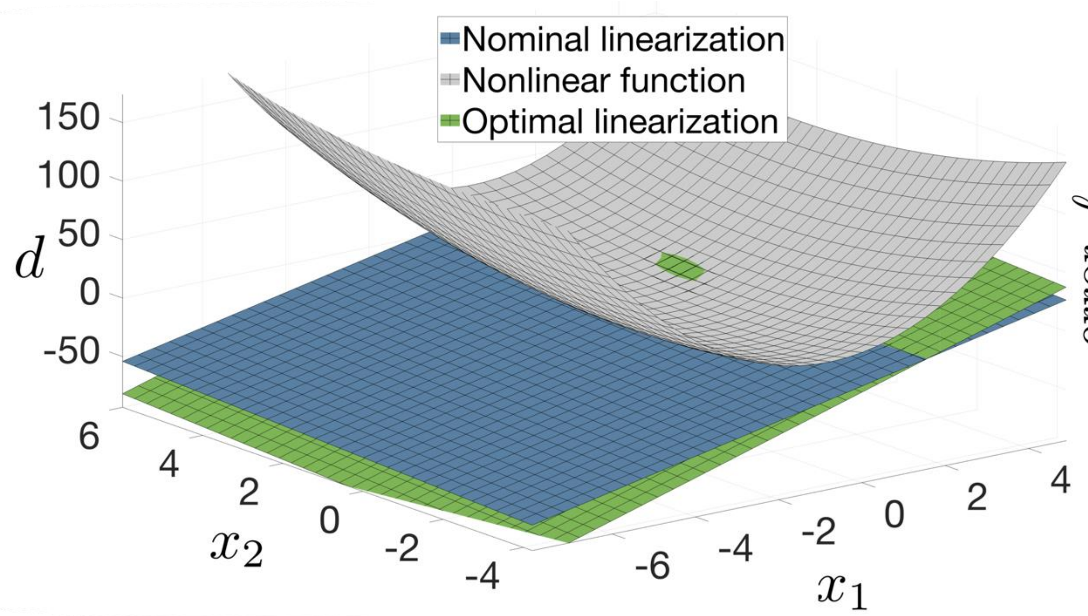
Preliminary tests

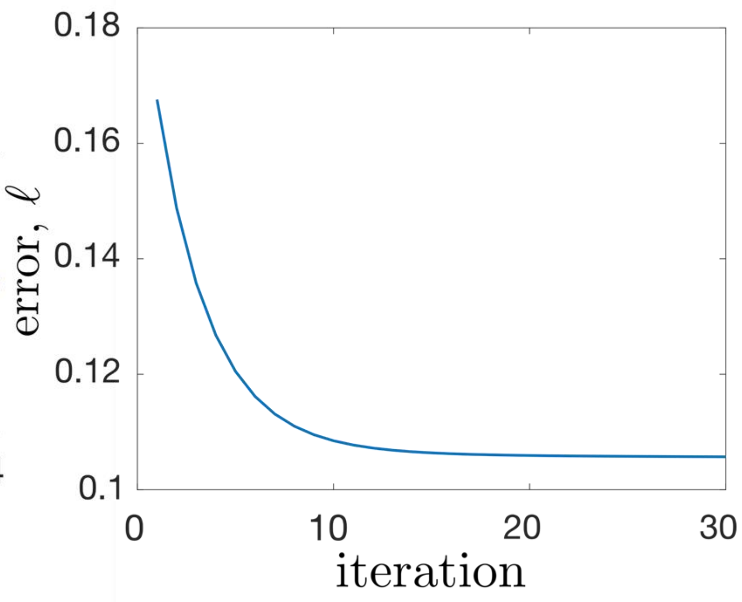
$$\mathbf{x}^\star(\theta) = \begin{bmatrix} x_1^\star \\ x_2^\star \end{bmatrix} = \arg\min_{\mathbf{x}} \; (x_1 - 4)^2 + (x_2 - 2)^2 \; \text{(P1)}$$
$$\text{s.to} \quad 4x_1^2 + 2x_1 + x_2^2 - x_1 x_2 = \theta$$

$$\longrightarrow$$

$$\hat{\mathbf{x}}^\star(\theta) = \arg\min_{\mathbf{x}} \; (x_1 - 4)^2 + (x_2 - 2)^2 \; \text{(P2)}$$
$$\text{s.to} \quad a x_1 + b x_2 + c = \theta$$

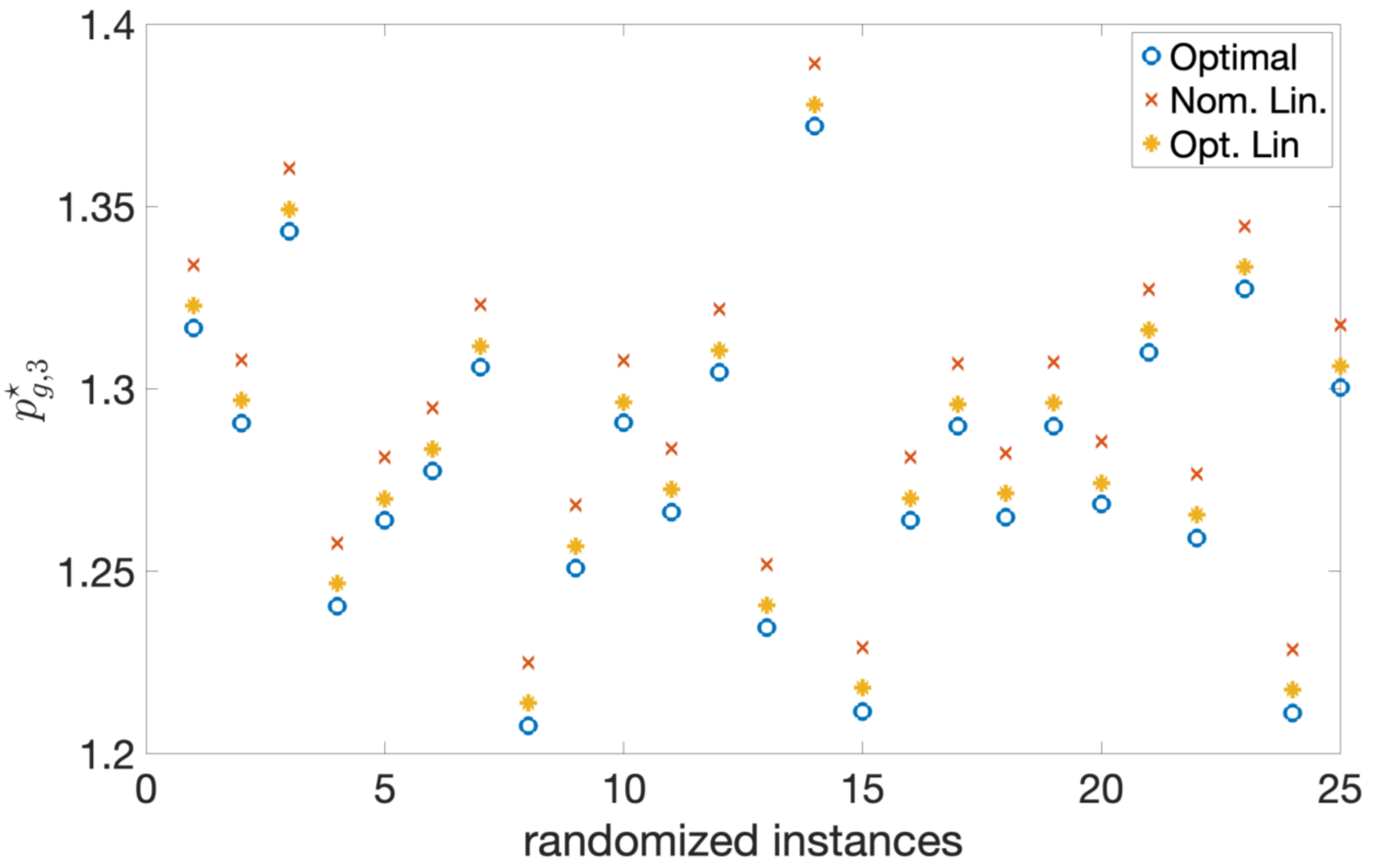
$$(a^\star, b^\star, c^\star) = \arg\min_{a,b,c} \sum_{s=1}^{10} [\mathbf{x}^\star(\theta_s) - \hat{\mathbf{x}}^\star(\theta_s)]^2 \qquad \theta \sim \mathcal{U}[5, 15]$$





IEEE 9 bus case

- 3 generators
- Nominal linearization is at the PF solution at base case
- 25 training samples, 200 gradient steps
- 25 test samples, error reduction 70%

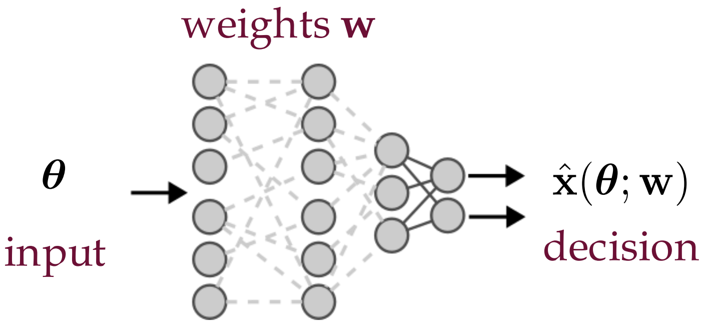


Next steps

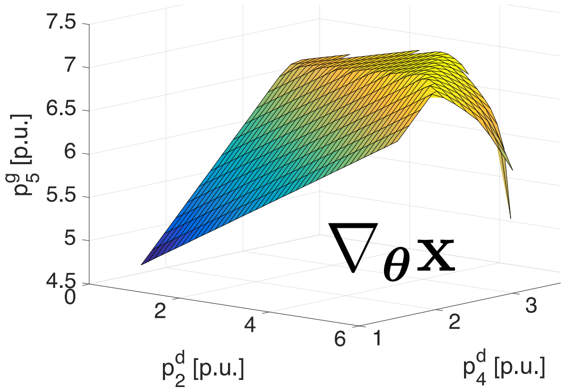
- ☐ Other meaningful error metrics
- ☐ Better optimization techniques
- ☐ Convergence analysis

Outline

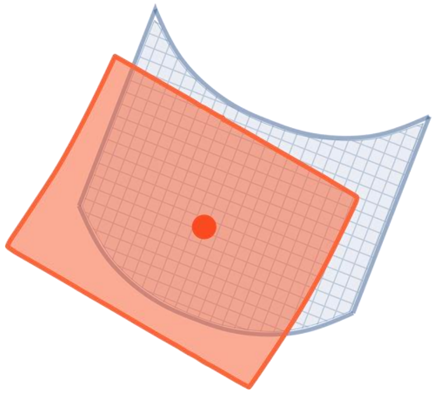
$$\mathbf{x}(\boldsymbol{\theta}) = \arg\min_{\mathbf{x}} \; f(\mathbf{x};\boldsymbol{\theta}) \quad \text{s.to} \quad \mathbf{g}(\mathbf{x};\boldsymbol{\theta}) \leq \mathbf{0}$$



Learning to optimize

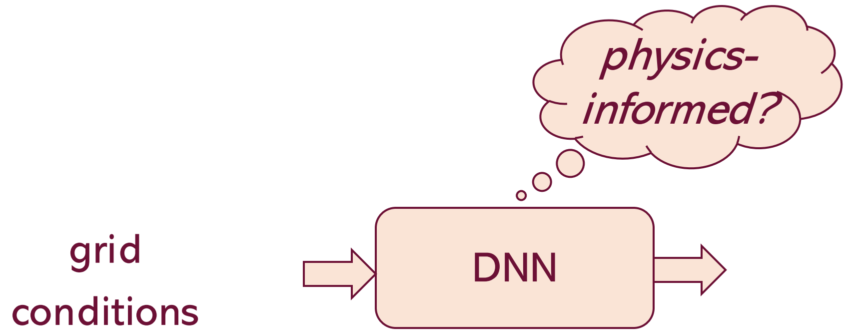


Sensitivity analysis



Power flow linearization

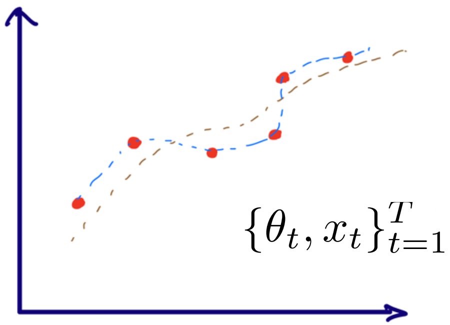
Learning for OPF



- DER setpoints
- warm-starts
- binding constraints

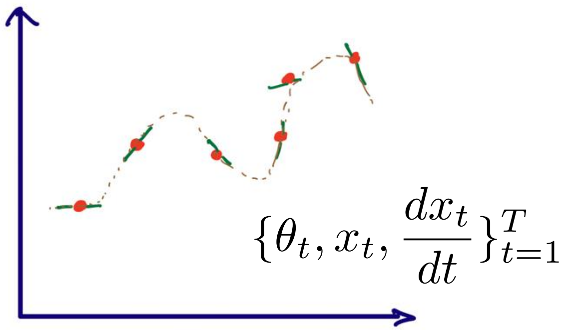
$\boldsymbol{\theta}$

$\mathbf{x}(\boldsymbol{\theta})$



- Polynomial regression
- DNN, GP,...
- *High data requirement*

KKT conditions, dual variables, function forms



- (Physics) informed regularizer
- Can be done for DNN & GP
- *Reduced data requirement*

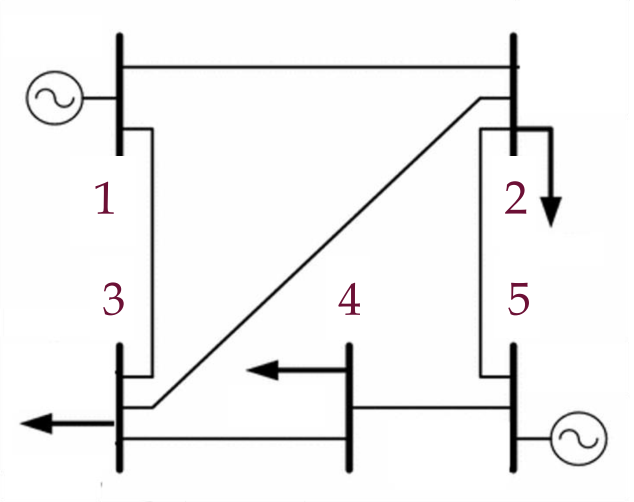
Singh+, "Learning to Optimize Power Distribution Grids using Sensitivity-Informed Deep Neural Networks," *IEEE SmartGridComm '20*

Sensitivity-informed learning

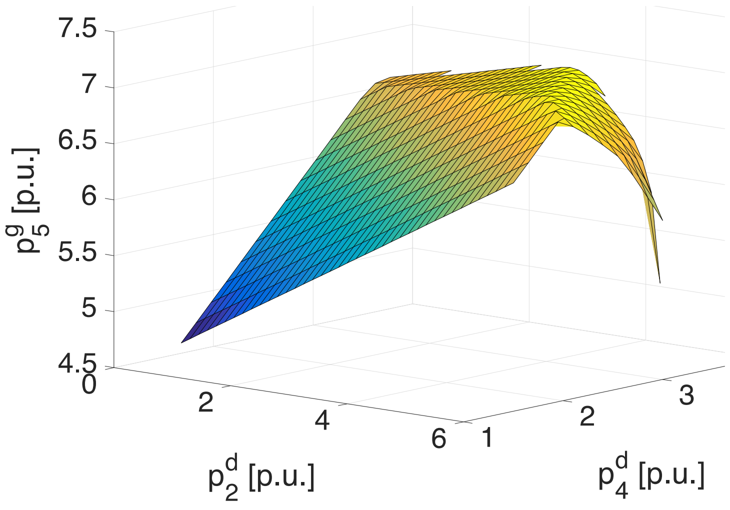
- Augmented supervised-learning cost

$$\min_{\mathbf{w}} \sum_{t=1}^{T} \left[\|\hat{\mathbf{x}}(\boldsymbol{\theta}_t) - \mathbf{x}(\boldsymbol{\theta}_t)\|_2^2 + \rho\|\nabla_{\boldsymbol{\theta}}\hat{\mathbf{x}}(\boldsymbol{\theta}_t) - \nabla_{\boldsymbol{\theta}}\mathbf{x}(\boldsymbol{\theta}_t)\|_F^2\right]$$

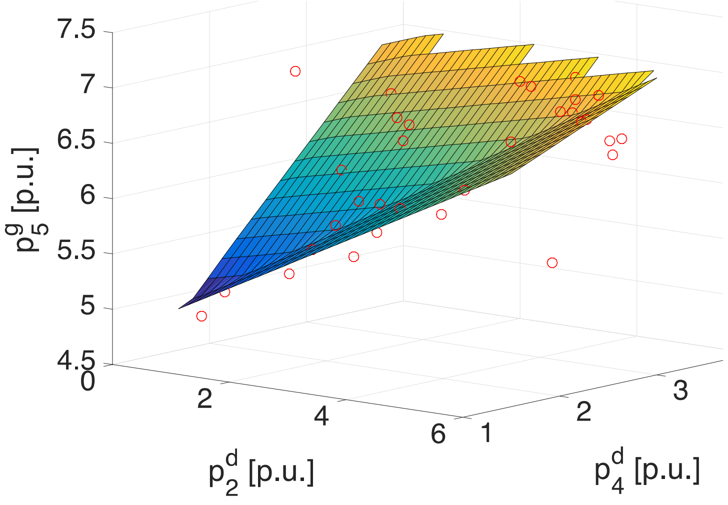
P-DNN SI-DNN



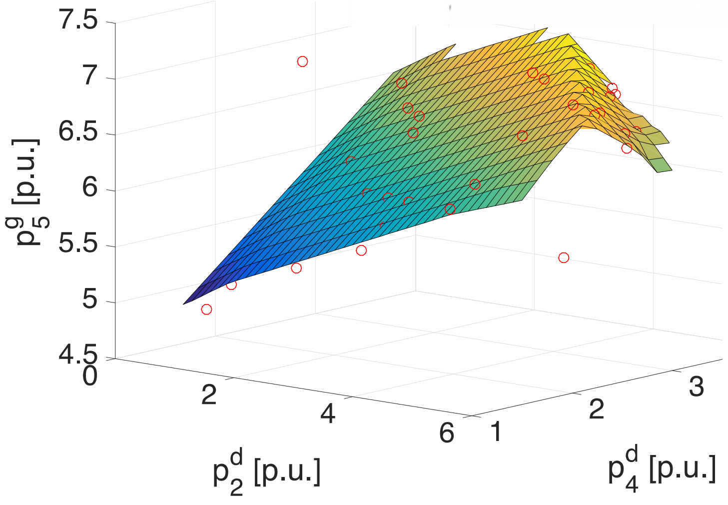
- PJM 5-bus example: Solve AC-OPF over varying loads 2+4 to dispatch GENs 1+5



actual OPF solution



P-DNN prediction



SI-DNN prediction

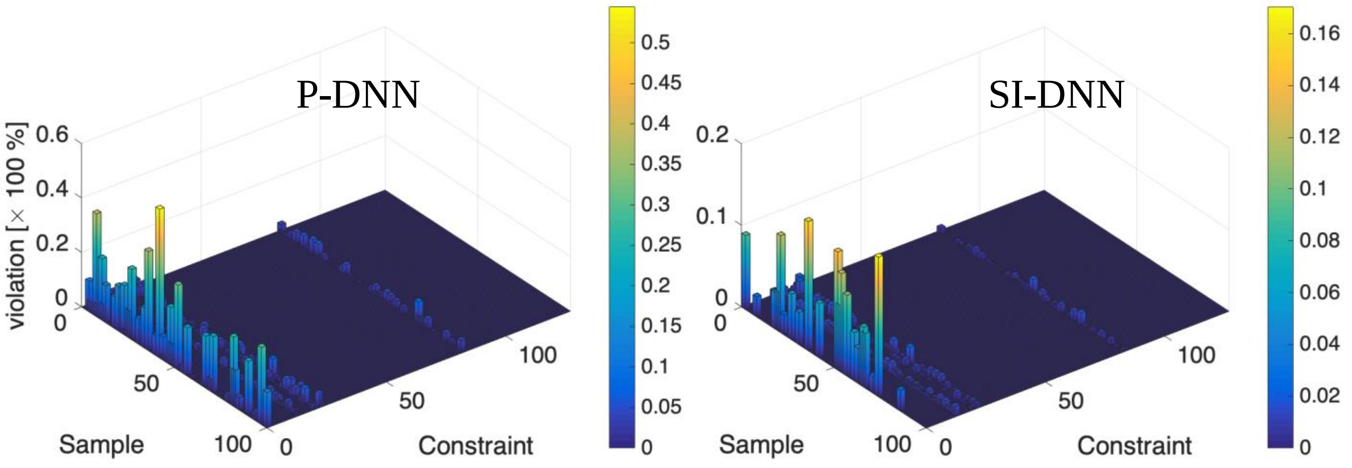
Quantifying accuracy improvements

AVERAGE TEST MSE [$\times 10^{-3}$] AND TRAINING TIME [IN SEC] FOR PREDICTING MATPOWER SOLUTION ON IEEE 39-BUS SYSTEM

Training Size	**P-DNN**		**SI-DNN**	
	MSE	Time	MSE	Time
10	8.6	738	3.3	746
50	4.3	739	2.1	756
100	3.2	747	2.0	776
250	1.9	302	2.0	332

Train. Size	**IEEE 118-bus**				**Illinois 200-bus**			
	P-DNN		**SI-DNN**		**P-DNN**		**SI-DNN**	
	MSE	Time	MSE	Time	MSE	Time	MSE	Time
25	1.8	447	1.1	483	0.19	452	0.04	491
50	1.7	458	1.1	527	0.15	456	0.04	524
100	1.6	463	0.9	610	0.09	471	0.06	608

Quantifying feasibility improvements



AVERAGE TEST MSE [$\times 10^{-3}$] FOR PREDICTING SDP SOLUTIONS, AND CONSTRAINT VIOLATION STATISTICS ON THE IEEE 39-BUS SYSTEM

Train. Size	P-DNN				SI-DNN			
	MSE	(a)	(b)	(c)	MSE	(a)	(b)	(c)
10	6.3	2.61	0.50	9.78	0.91	2.52	0.37	3.35
50	3.6	2.45	0.55	7.38	0.62	2.58	0.27	2.06
100	2.5	2.59	0.53	6.87	0.67	2.52	0.27	1.96

(a) #violations /instance; (b) max. violation; (c) mean violation [$\times 10^{-2}$]

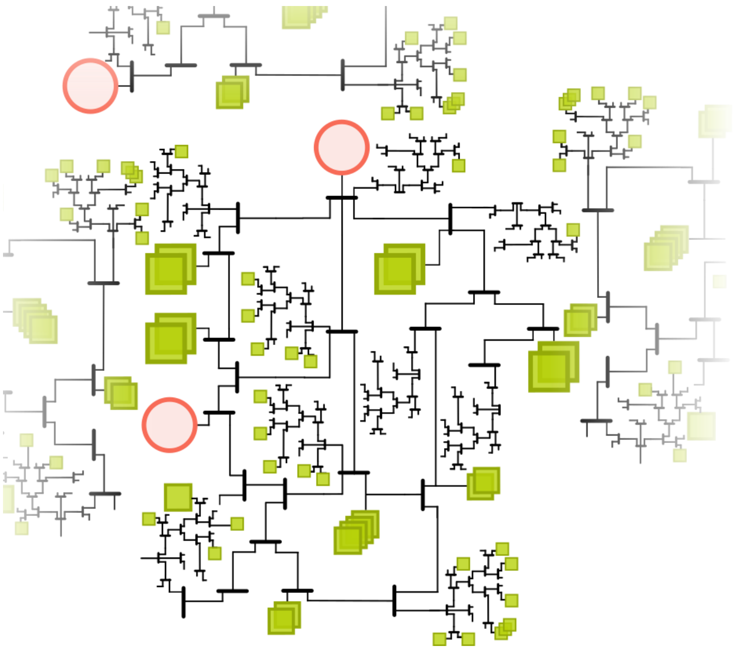
Sensitivity regularization is tantamount to local-approximation Vs point-approximation

- ☑ Sensitivities information improves learning
- ☑ Physics-based model regularization
- ☑ Approach applies to DC and AC OPF across solvers and relaxations
- ☑ Extensions possible to non-parametric learning models

Jalali, Singh, Kekatos, Giannakis, Liu, "Fast Inverter Control by Learning the OPF mapping using Sensitivity Informed Gaussian Processes," *IEEE Trans. Smart Grids,* 2022

Parting thoughts

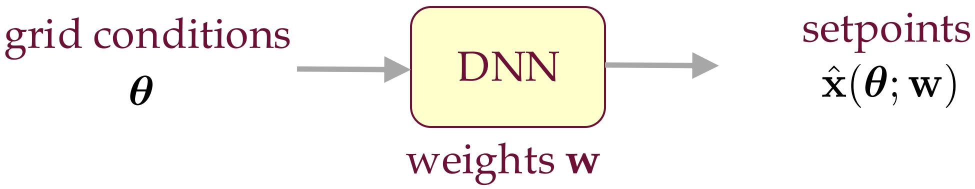
- Sensitivity analysis of OPF is possible and "scalable"
- Application-informed PF linearization could be a treasure box
  - ☑ Use-inspired research
  - ☑ Bilevel optimization
  - ☑ Sensitivity analysis
- Scalable solutions combining physics & data will drive the future
- Sensitivities can help build the bridge
- *I am looking for PhD students ☺*



Thank You!

Rationale for sensitivity regularization

- Consider local approximations



True mapping $\quad x(\theta_t + \epsilon) \approx x(\theta_t) + \epsilon\, x'(\theta_t)$

DNN mapping $\quad \hat{x}_{\mathbf{w}}(\theta_t + \epsilon) \approx \hat{x}_{\mathbf{w}}(\theta_t) + \epsilon\, \hat{x}'_{\mathbf{w}}(\theta_t)$

- Suppose ϵ is a zero-mean random variable with $\mathbb{E}[\epsilon^2] = \sigma^2$

$$\mathbb{E}_\epsilon \left[\left(\hat{x}_{\mathbf{w}}(\theta_t + \epsilon) - x(\theta_t + \epsilon) \right)^2 \right]$$
$$\approx (\hat{x}_{\mathbf{w}}(\theta_t) - x(\theta_t))^2 + \sigma^2 (\hat{x}'_{\mathbf{w}}(\theta_t) - x'(\theta_t))^2$$

$$\min_{\mathbf{w}} \sum_{t=1}^{T} \left[\|\hat{\mathbf{x}}(\boldsymbol{\theta}_t) - \mathbf{x}(\boldsymbol{\theta}_t)\|_2^2 + \rho \|\nabla_{\boldsymbol{\theta}} \hat{\mathbf{x}}(\boldsymbol{\theta}_t) - \nabla_{\boldsymbol{\theta}} \mathbf{x}(\boldsymbol{\theta}_t)\|_F^2 \right]$$

Sensitivity regularization is tantamount to local-approximation Vs point-approximation

Alternatively, viewed as (almost free) data augmentation