Optimizing Power Distribution Grids on a Data Budget

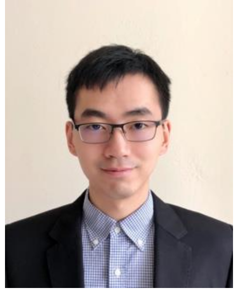
Vassilis Kekatos (kekatos@purdue.edu)

Seventh Workshop on Autonomous Energy Systems

Friday, September 6, 2024

Ridley Annin

Manish Singh (UW-Madison)



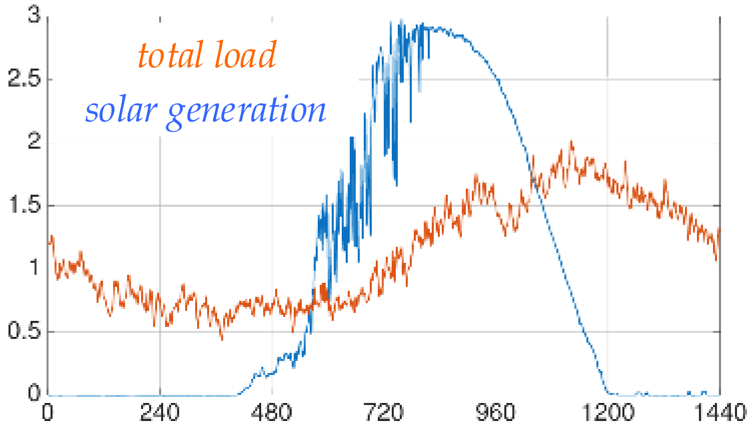
Junjie Qin

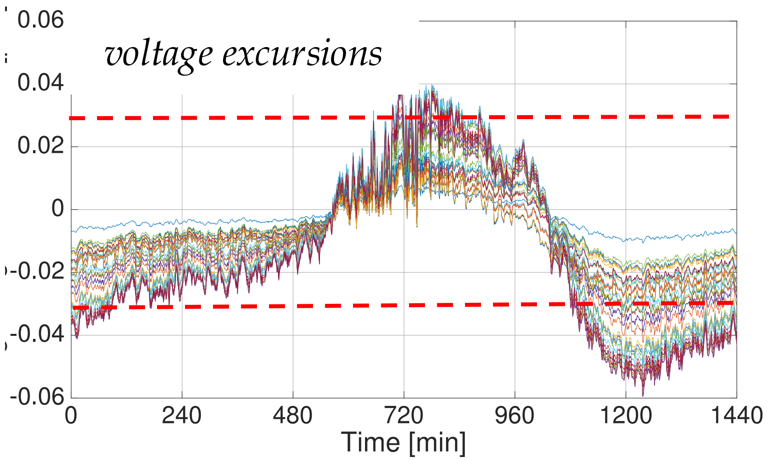
Acknowledgements

NSF-2150596

Motivation

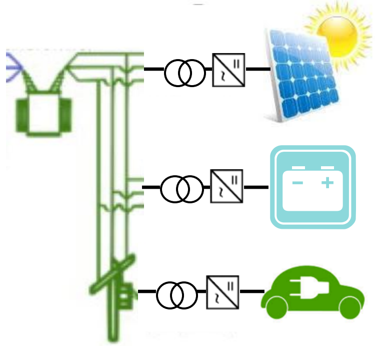
- Voltage fluctuations due to solar and other DERs





- Inefficiency of voltage control devices

- Reactive power control using smart inverters



- Optimal scheduling of DERs via OPF

OPF data

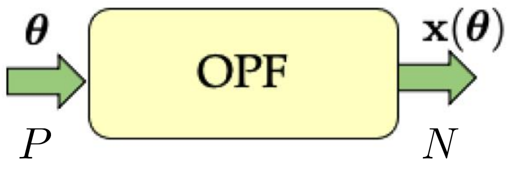
- DSO schedules DERs using OPF for varying *grid data* $\boldsymbol{\theta} = [\mathbf{p}_g;\ \mathbf{p}_\ell;\ \mathbf{q}_\ell]$

$$\begin{aligned} \mathbf{x}(\boldsymbol{\theta}_t) = \arg\min_{\mathbf{q}} \quad & \ell(\mathbf{q};\boldsymbol{\theta}_t) \\ \text{s.to} \quad & \mathbf{v}(\mathbf{q};\boldsymbol{\theta}_t) \in \mathcal{V} \\ & -\bar{\mathbf{q}} \leq \mathbf{q} \leq \bar{\mathbf{q}} \end{aligned}$$

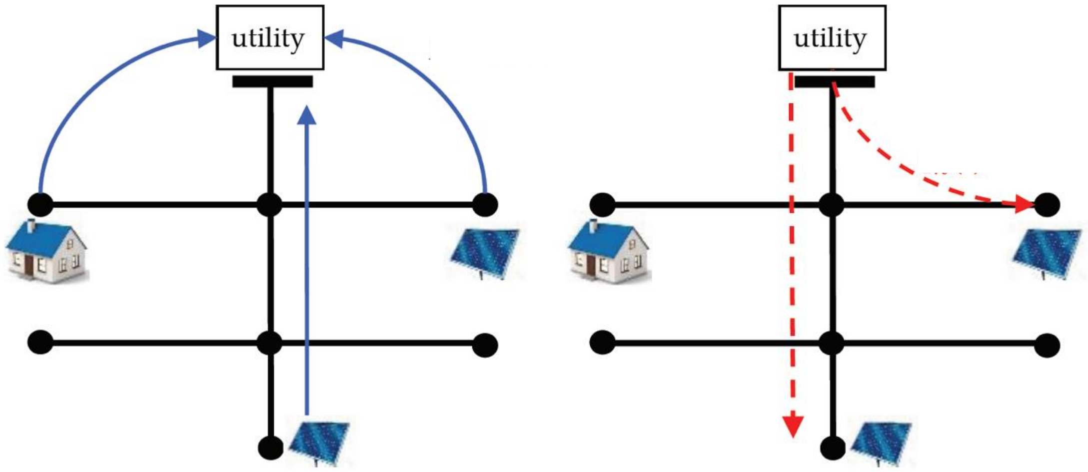
← ohmic losses

← voltage/network constraints

← DER limits

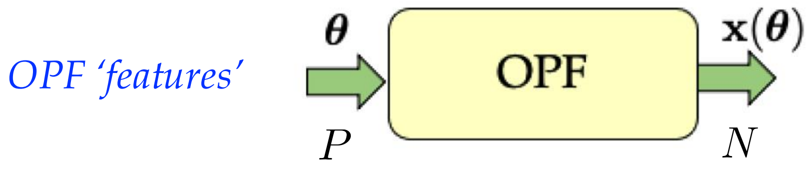


- Data collected in decentralized fashion
  - numerous buses sampled frequently
  - data privacy/cyber-security
  - incomplete observability
  - where to install meters

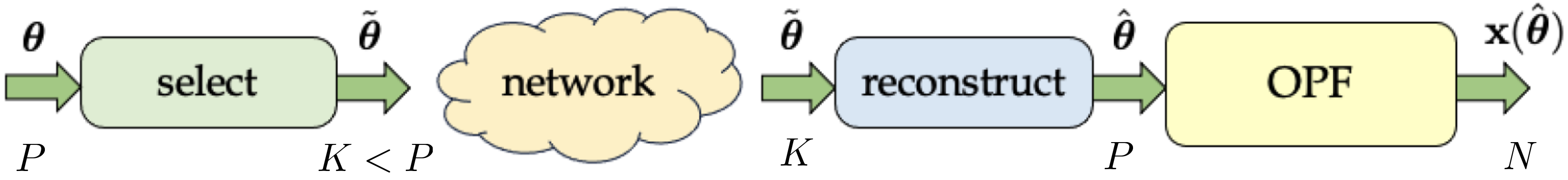


OPF data distillation

- ***Fact:*** DSO must collect much data to solve the OPF

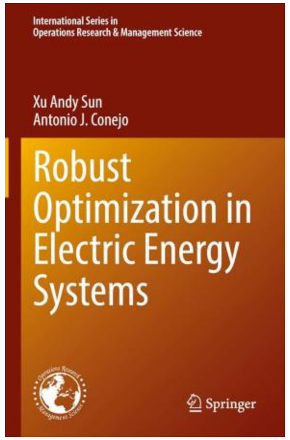


- ***Need:*** How to solve OPF on a *data budget*?



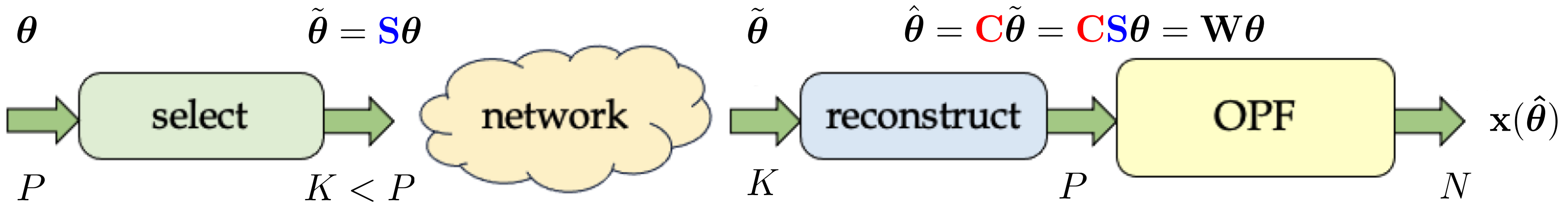
- ***Question:*** How to select and reconstruct OPF data?
  - redundancies in OPF data and problem structure

Related works



- Handle uncertain parameters via stochastic/robust OPF [Roald+'23]
- Compress OPF datasets via PCA to train DNNs [Park+'23]
- Optimal placement of sensors for observability [Bhela+Kekatos'17]
- Optimal placement of sensors for controllability [Lin+'13, Dorfler+'14, Summers+'16]
- Optimal network reduction [Chevalier-Almassalkhi'22, Caliskan-Tabuada'12, Nikolakakos+'18]

General methodology



- How to design $\mathbf{W} = \mathbf{CS}$?

- Collect OPF data from representative loading scenarios $\boldsymbol{\Theta} = [\theta_1 \ \theta_2 \ \cdots \ \theta_T]$
  - normalized data (zero-mean and unit-variance)

- *Type-1: Data-centric data distillation*
  - design W to improve data fidelity
  - PCA, DEIM, Group Lasso (GL)

- *Type-2: Minimizer-centric data distillation*
  - design W to improve minimizer fidelity
  - Bilevel Group Lasso (BGL)

Principal component analysis (PCA)

- Finds best rank-K approximation of matrix $\boldsymbol{\Theta}$

$$\min_{\mathbf{W}} \sum_{t=1}^{T} \|\boldsymbol{\theta}_t - \mathbf{W}\boldsymbol{\theta}_t\|_2^2 = \|\boldsymbol{\Theta} - \mathbf{W}\boldsymbol{\Theta}\|_F^2$$
$$\text{s.to} \quad \text{rank}(\mathbf{W}) = K$$

$$\mathbf{W}_{\text{PCA}} = \mathbf{U}_K \mathbf{U}_K^\top$$

$\mathbf{U}_K$ top-K eigenvectors of $\mathbf{C}_\theta = \frac{1}{T}\boldsymbol{\Theta}\boldsymbol{\Theta}^\top$

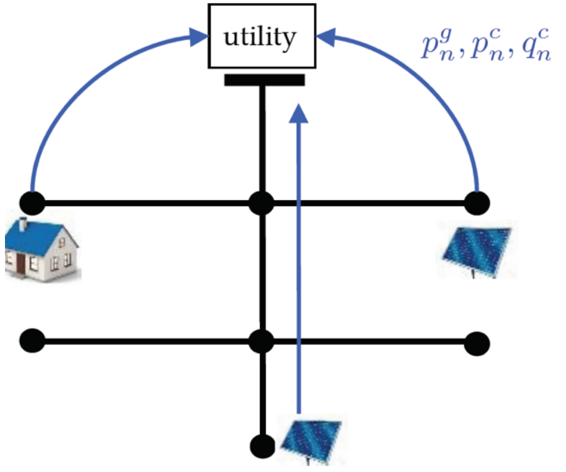
- Easy to find through EVD of covariance matrix

- Achieves minimum rank-K reconstruction error $E_{\text{PCA}} = \|\boldsymbol{\Theta} - \mathbf{W}_{\text{PCA}}\boldsymbol{\Theta}\|_F^2$

- Compresses but does not select! → … at least serves as a benchmark

Discrete empirical interpolation method (DEIM)

- Suppose selection matrix $\mathbf{S}$ has been found
- DEIM reconstructs data as $\hat{\mathbf{\Theta}} = \mathbf{U}_K(\mathbf{S}\mathbf{U}_K)^{-1}\mathbf{S}\mathbf{\Theta}$
- Neat interpolation property $\mathbf{S}\hat{\mathbf{\Theta}} = \mathbf{S}\mathbf{\Theta}$!
- How to find selection matrix $\mathbf{S}$?
  - minimize upper bound $E_{\text{PCA}} \leq E_{\text{DEIM}}(\mathbf{S}) \leq \| (\mathbf{S}\mathbf{U}_K)^{-1} \|_2^2 \cdot E_{\text{PCA}}$
- Hard problem; low-complexity greedy algorithm 🙂
- DEIM selects data but cannot select groups of data



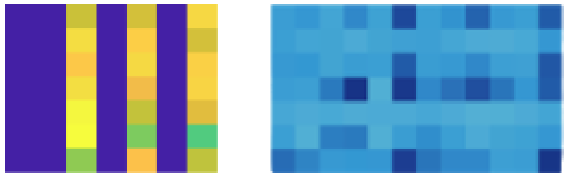
Barrault, Maday, Nguyen, and Patera, "An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations," *Comptes Rendus Mathematique*, vol. 339, no. 9, pp. 667–672, 2004.

Group Lasso (GL)

- Finds a column-sparse **W** through convex *group lasso* (GL) problem

$$\text{(GL):} \quad \min_{\mathbf{W}} \; \frac{1}{2T}\|\mathbf{\Theta} - \mathbf{W}\mathbf{\Theta}\|_F^2 + \lambda_1 \sum_{p=1}^{P} \|\mathbf{w}_p\|_2$$

$\hat{\mathbf{\Theta}} =$



$\mathbf{W}$ $\mathbf{\Theta}$

- Effects simultaneous data selection and compression $\hat{\boldsymbol{\theta}} = \mathbf{W}\boldsymbol{\theta} = \sum_{p=1}^{P} \theta_p \mathbf{w}_p$

- Columns of **W** can be stacked together in groups

- Parameter λ_1 controls block sparsity; use bisection to select exactly K columns

- Two-stage GL: use GL to find the support (**S**) and least-squares fitting to finalize **C**

Yuan, Lin, *"Model selection and estimation in regression with grouped variables,"* 2007.

Proximal gradient descent for GL

- GL can be solved using SOCP, ADMM, PGD
- Generalizes gradient descent for non-differ. objectives

$$\min_{\mathbf{W}} \underbrace{\frac{1}{2T}\|\mathbf{\Theta} - \mathbf{W}\mathbf{\Theta}\|_F^2}_{f_1(\mathbf{W})} + \underbrace{\lambda_1 \sum_p \|\mathbf{w}_p\|_2}_{g(\mathbf{W})}$$

- Gradient descent step for differentiable part

$$\mathbf{Y}^i = \mathbf{W}^i - \mu_i \nabla_{\mathbf{W}} f_1(\mathbf{W}^i)$$

where $\nabla_{\mathbf{W}} f_1(\mathbf{W}) = (\mathbf{W} - \mathbf{I})\mathbf{C}_\theta$

- Proximal step for non-differentiable part

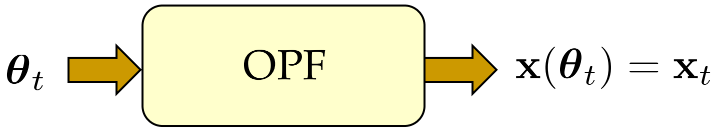
$$\mathbf{w}_p^{i+1} = \begin{cases} \left(1 - \frac{\lambda_1 \mu_i}{\|\mathbf{y}_p^i\|}\right) \cdot \mathbf{y}_p^i & , \|\mathbf{y}_p^i\| \geq \lambda_1 \mu_i \\ \mathbf{0} & , \text{otherwise.} \end{cases}$$

- Accelerated PGD takes gradient step on extrapolated point (memory)

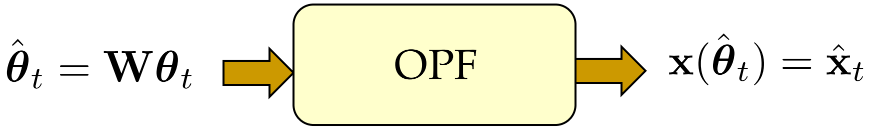
$$\bar{\mathbf{W}}^i = \mathbf{W}^i + \frac{\alpha_{i-1} - 1}{\alpha_i}\left(\mathbf{W}^i - \mathbf{W}^{i-1}\right)$$

Minimizer-centric data distillation

- Type-1 methods are agnostic to end use of OPF data



- Type-2 methods aim at fidelity of OPF solution



- Given labeled dataset $\mathcal{D} = \{(\boldsymbol{\theta}_t, \mathbf{x}_t)\}_{t=1}^{T}$, design $\boldsymbol{W}$ via *bilevel group lasso* (BGL)

$$\text{(BGL):} \quad \min_{\mathbf{W}} \quad \frac{1}{2T}\sum_{t=1}^{T}\|\mathbf{x}_t - \hat{\mathbf{x}}_t\|_2^2 + \lambda_2 \sum_{p=1}^{P}\|\mathbf{w}_p\|_2$$

$$\text{s.to} \quad \hat{\mathbf{x}}_t \text{ is the OPF solution for } \hat{\boldsymbol{\theta}}_t = \mathbf{W}\boldsymbol{\theta}_t \; \forall t$$

- Nonconvex bilevel program
  - replace inner problems by KKT to solve BGL as MINLP
  - PGD iterative algorithm

Proximal gradient descent for BGL

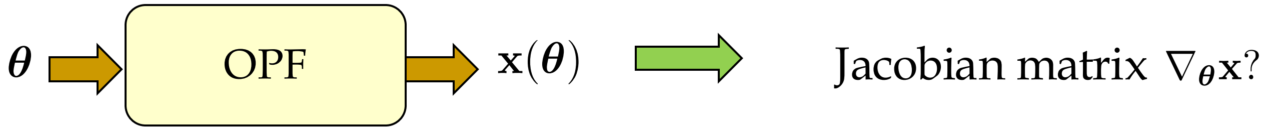
$$\min_{\mathbf{W}} \underbrace{\frac{1}{2T}\sum_{t=1}^{T}\|\mathbf{x}_t - \hat{\mathbf{x}}_t(\mathbf{W})\|_2^2}_{f_2(\mathbf{W})} + \lambda_2 \underbrace{\sum_{p=1}^{P}\|\mathbf{w}_p\|_2}_{g(\mathbf{W})}$$

- Gradient descent for 'differentiable' part: $\mathbf{Y}^i = \mathbf{W}^i - \mu_i \nabla_{\mathbf{W}} f_2(\mathbf{W}^i)$
- Proximal step like before
- More complicated extrapolation due to non-convexity
  - take proximal gradient step at 3 candidate points and select the best
  - enjoys convergence to stationary (critical) point
- Gradient computation $\nabla_{\mathbf{w}} f_2 = \frac{1}{T}\sum_{t=1}^{T}(\nabla_{\mathbf{w}}\hat{\mathbf{x}}_t)^\top(\hat{\mathbf{x}}_t - \mathbf{x}_t)$ where $\nabla_{\mathbf{w}}\hat{\mathbf{x}}_t = \nabla_{\hat{\boldsymbol{\theta}}_t}\hat{\mathbf{x}}_t \cdot \nabla_{\mathbf{w}}\hat{\boldsymbol{\theta}}_t$

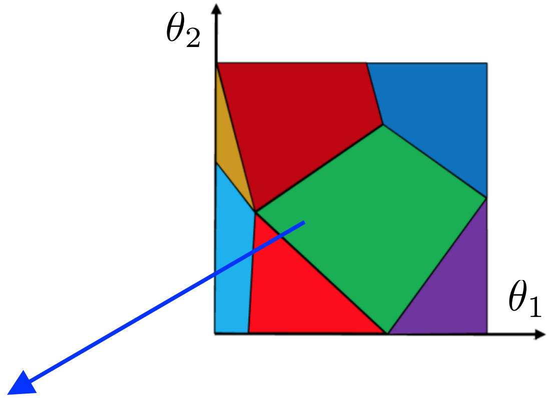
sensitivity analysis of OPF $\qquad \hat{\boldsymbol{\theta}}_t = \mathbf{W}\boldsymbol{\theta}_t$

Li and Lin, "Accelerated proximal gradient methods for nonconvex programming," in *Neural Information Processing Systems*, 2015.

Sensitivity analysis of the OPF



- If optimal primal/dual solutions are known, find Jacobian by solving system of linear equations!
- Need to solve T *(or 3T)* OPFs and find their Jacobians per PGD iteration…
  - stochastic PGD (one OPF instance per iteration)
- If OPF is LP/QP, use *multiparametric programming (MPP)*
  - to expedite *batch* OPF computations by an order of magnitude

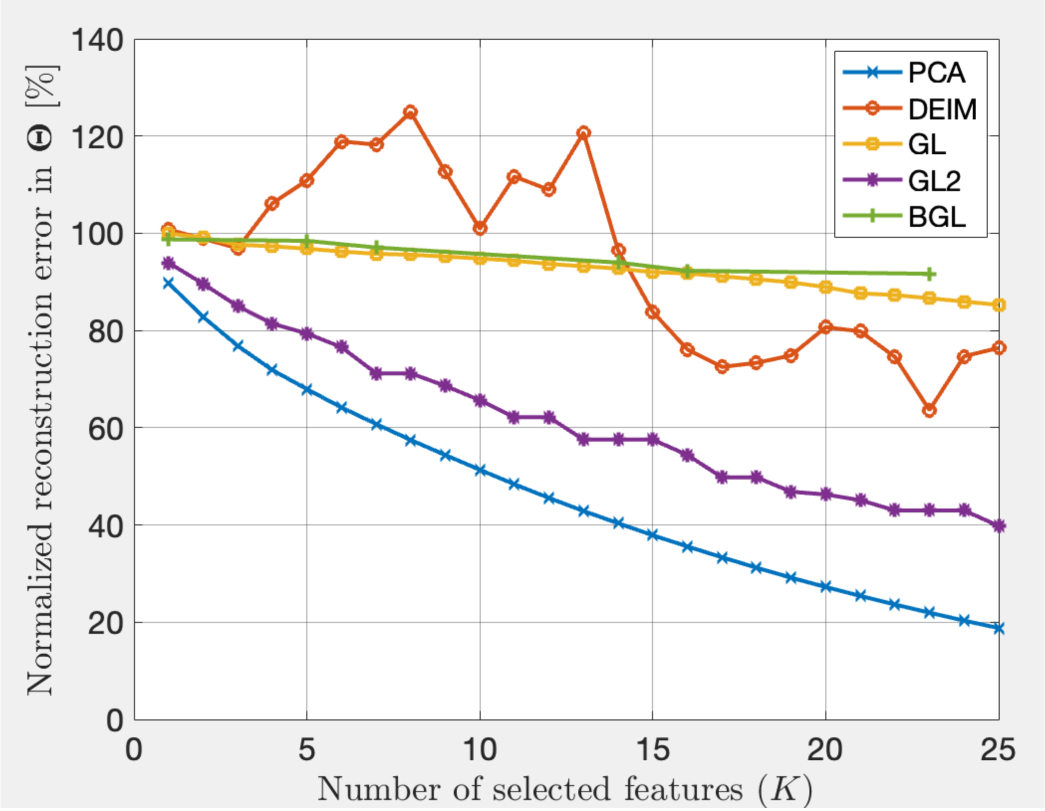


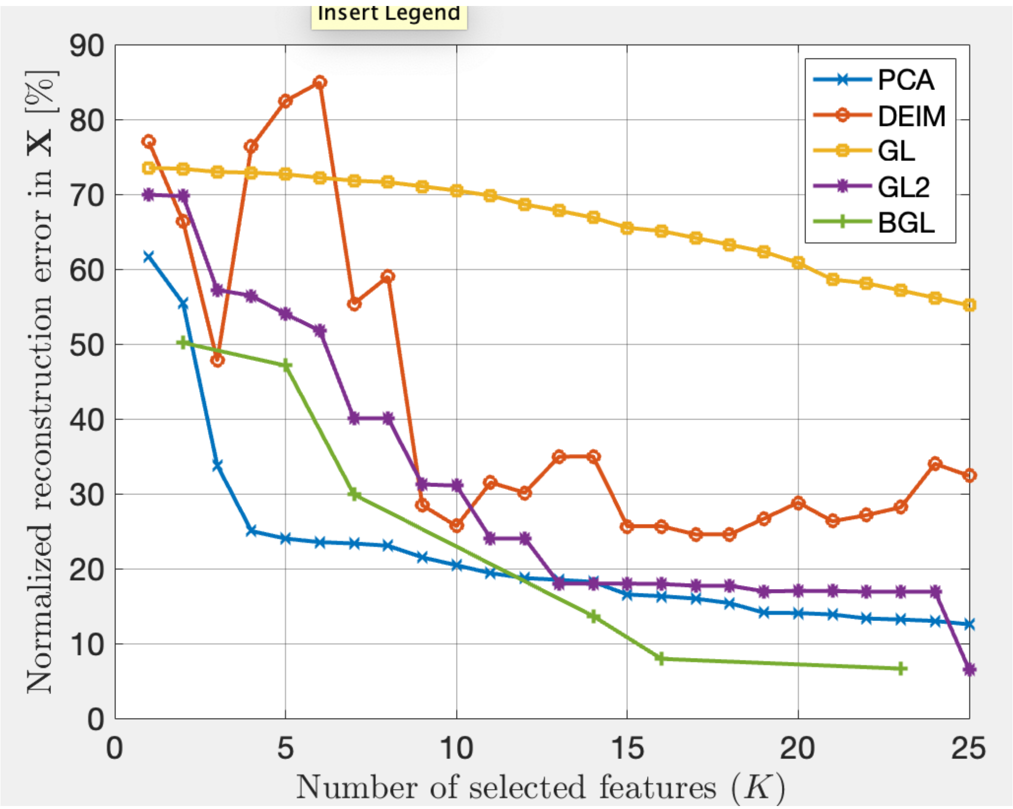
$$\mathbf{x}(\boldsymbol{\theta}) = \mathbf{A}_k\boldsymbol{\theta} + \mathbf{b}_k \quad \text{for all} \quad \boldsymbol{\theta} : \mathbf{C}_k\boldsymbol{\theta} + \mathbf{d}_k \leq \mathbf{0}$$

Singh, Kekatos, and Giannakis, "Learning to Solve the AC-OPF using Sensitivity-Informed DNNs," *IEEE TPWRS*, 2022.
Taheri, Jalali, Kekatos, and Tong, "Fast Probabilistic Hosting Capacity Analysis for Active Distribution Systems," *IEEE TSG*, 2021.

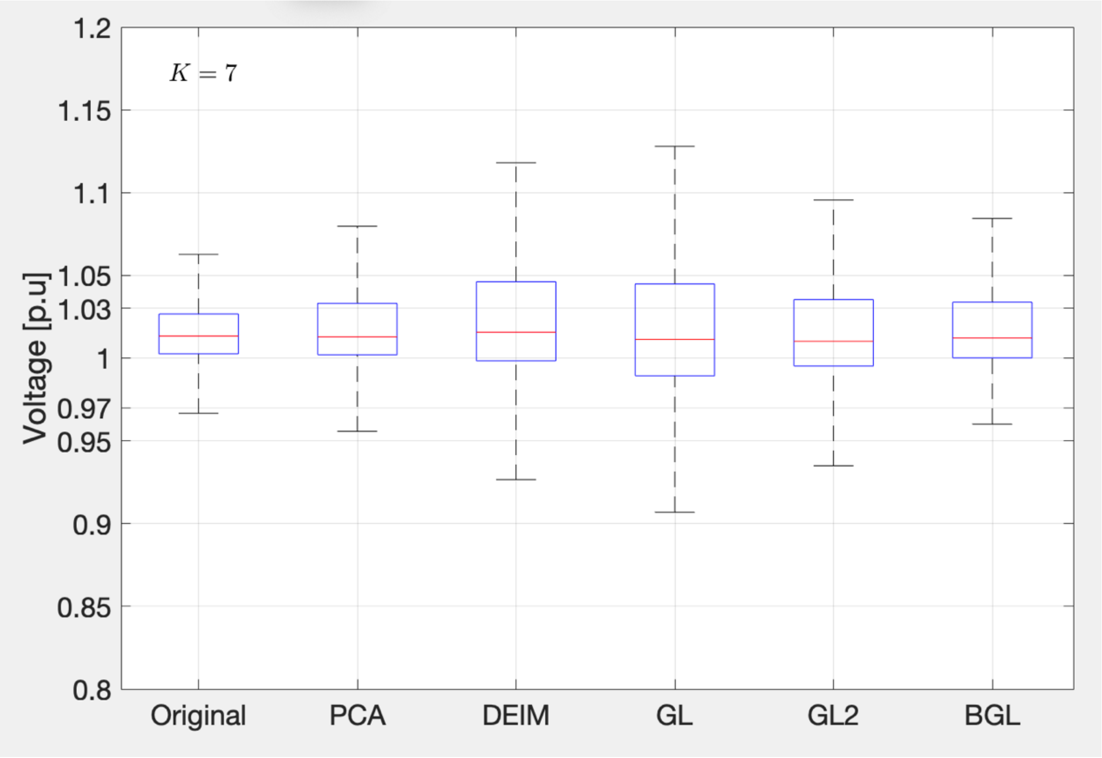
Numerical tests

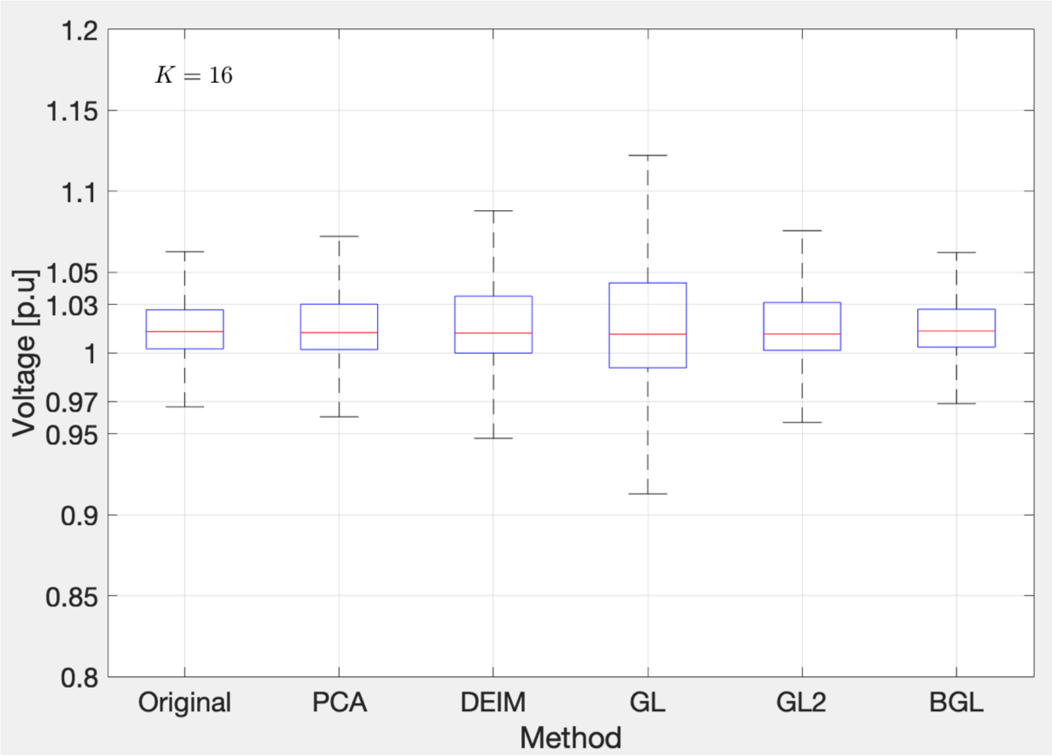
- Load/solar from Pecan Street; EV data from NREL
- IEEE 37-bus feeder; *P=50; N=10; T=800*





Voltage deviations





Conclusions

OPF data distillation

- ☑ three methods
- ☑ PGD algorithms
- ☑ differentiation through OPF

Ongoing work

- ❑ stochastic PGD
- ❑ comparison to MINLP
- ❑ multiphase/AC-OPF
- ❑ meter placement for OPF
- ❑ OPF data security
- ❑ unsupervised learning
- ❑ nonlinear reconstruction

Thank You!

Preprint to appear on arxiv soon

Related references

1) Jalali, Singh, Kekatos, Giannakis, and Liu, "Fast Inverter Control by Learning the OPF Mapping using Sensitivity-Informed Gaussian Processes," *TSG'23*.
2) Singh, Kekatos, Giannakis, "Learning to Solve the AC-OPF using Sensitivity-Informed Deep Neural Networks," *TPWRS'22*.
3) Taheri, Jalali, Kekatos, Tong, "Fast Probabilistic Hosting Capacity Analysis for Active Distribution Systems," *TSG'21*.
4) Singh, Gupta, Kekatos, Cavraro, and Bernstein, "Learning to Optimize Power Distribution Grids using Sensitivity-Informed Deep Neural Networks," *IEEE Smart Grid Comm*, 2020.