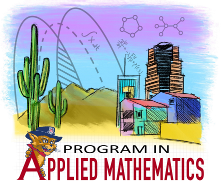
THE UNIVERSITY OF ARIZONA



Towards Model Reduction for Power System Transients with Physics-Informed Neural PDE

Misha Chertkov Applied Math @ UArizona

The Fifth Autonomous Energy Systems Workshop

July 15, 2022 – NREL/Golden

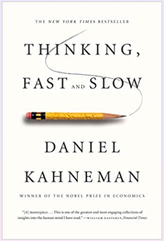
- L. Pagnier, J. Fritzsch, P. Jacquod and M. Chertkov, "Toward Model Reduction for Power System Transients With Physics-Informed PDE", in IEEE Access, vol. 10, pp. 65118-65125, 2022, doi: 10.1109/ACCESS.2022.3183336.

Laurent Pagnier

Julian Fritzsch

Philippe Jacquod

Outline

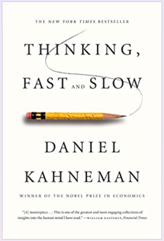


System 1 & 2 in Deep Learning & AI

- "From System 1 Deep Learning to System 2 Deep Learning" – Yoshua Bengio, NeurIPS 2019
- "Combining Fast and Slow Thinking for Human-like and Efficient Navigation in Constrained Environments" – M. Ganappini, et al, arXiv:2201.07050

- System 1 – operates automatically & quickly

- System 2 – allocates attention to effortfull mental activities

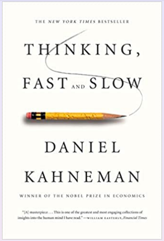


System 1 & 2 in Deep Learning & AI

- "From System 1 Deep Learning to System 2 Deep Learning" – Yoshua Bengio, NeurIPS 2019
- "Combining Fast and Slow Thinking for Human-like and Efficient Navigation in Constrained Environments" – M. Ganappini, et al, arXiv:2201.07050

- System 1 – operates automatically & quickly
  - **Deep Learning** empowered by **Automatic Differentiation**

- System 2 – allocates attention to effortfull mental activities
  - **Physics Informed Machine Learning** – more generally **Explainable Heuristics in Quantitative Sciences**



System 1 & 2 in Deep Learning & AI

- "From System 1 Deep Learning to System 2 Deep Learning" – Yoshua Bengio, NeurIPS 2019
- "Combining Fast and Slow Thinking for Human-like and Efficient Navigation in Constrained Environments" – M. Ganappini, et al, arXiv:2201.07050

- System 1 – operates automatically & quickly
  - **Deep Learning** empowered by **Automatic Differentiation**

- System 1.5
  - 20th century Applied Math – ODE, PDE, **Sensitivity Analysis**

- System 2 – allocates attention to effortfull mental activities
  - **Physics Informed Machine Learning** – more generally **Explainable Heuristics in Quantitative Sciences**

Physics Informed Machine Learning for Power Systems

Machine Learning (e.g. Neural Network, Graph Models, etc)

- will make Power System Computations
  - faster (efficient)
  - possible even when data/measurements incomplete
- requires ground-truth data
  - actual measurements (Phasor Measurement Units, etc)
  - power flow solvers (microscopic simulations) – reliable, possibly heavy
- can be power-system "informed" (System 2) vs "agnostic" (System 1)
  - What is System 1 today may become System 2 tomorrow (with proper theory & enough of experiments)
- methods/options are many
  - should be gauged to available data, level of uncertainty, etc

Incomplete Review: Brief, Recent, Biased

AI/ML in Power Systems (System 1, System 2 & juxtaposition)

- Structure Learning, Sparse Measurements, Graphical Models, Focus on Power Distribution: Deka, et al [2016-2019]
- Learning ODE: Power Transmission, Dynamic Coefficients in Swing Equations, Deterministic and Stochastic, Lokhov, et al [2017]
- Real-time Faulted Line Localization and PMU Placement in Power Transmission through CNN: Li, et al [2018]
- Collocation Point Neural ODE for Power Systems: Misuris, et al [2018]
- Learning a Generator Model from Terminal Bus Data: many ML schemes, tradeoffs, ranking models according to regimes, Stulov et al [2019]
- Learning from power system data stream, phasor-detective approach, Escobar et al [2019]

Incomplete Review: Brief, Recent, Biased

AI/ML in Power Systems (System 1, System 2 & juxtaposition)

- **Physics-Informed** Graphical Neural Network for **Parameter & State Estimations** in Power Systems https://arxiv.org/abs/2102.06349 (Pagnier & MC))
- **Embedding Power Flow** into Machine Learning for Parameter and State Estimation https://arxiv.org/abs/2103.14251 (Pagnier & MC)
- **Which Neural Network to Choose** for Post-Fault Localization, Dynamic State Estimation and Optimal Measurement Placement in Power Systems? https://arxiv.org/abs/2104.03115 (Afonin & MC))

Machine Learning (Neural Networks) Setting

NN models: General

- $\mathrm{NN}_{\vec{\phi}}(\vec{x}) = \vec{y}$
  - Vector, $\vec{\phi}$, of Not-Interpretable Parameters
  - Input vector: $\vec{x}$
  - Output vector: $\vec{y}$

NN models: Loss Functions

- L2 norm $\|\cdots\|$
- Probabilistic (Cross Entropy or Kullback-Leibler)
- Regularizations, e.g. L1 (sparsity, physical, etc)

NN models: Architectures

- Convolutional NN (LeCun 1989 –)
- Graph NN (Scarcelli. et al 2009 –)
- Neural ODE (Chen et al 2008 –)
- Collocation Point NN (Lagaris et al 1998, Raissi et al 2019 –)
- Hamiltonian NN (Greydanus et al 2018 –)

Power Flow Equations

- grid-graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- complex-valued powers: $\forall a \in \mathcal{V}: \;\; S_a \equiv p_a + \boldsymbol{i} q_a$
- complex-valued (electric) potentials, $\forall a \in \mathcal{V}: \;\; V_a \equiv v_a \exp(\boldsymbol{i}\theta_a)$,
- Power Flow (PF) equations:

$$p_a = \sum_{b;\{a,b\}\in\mathcal{E}} v_a v_b \Big[g_{ab} \cos\left(\theta_a - \theta_b\right) + \beta_{ab} \sin\left(\theta_a - \theta_b\right) \Big],$$

$$q_a = \sum_{b;\{a,b\}\in\mathcal{E}} v_a v_b \Big[g_{ab} \sin\left(\theta_a - \theta_b\right) - \beta_{ab} \cos\left(\theta_a - \theta_b\right) \Big],$$

- Direct PF Map: $\boldsymbol{\Pi_Y} : \boldsymbol{S} \equiv (S_a | a \in \mathcal{V}) \mapsto \boldsymbol{V} \equiv (V_a | a \in \mathcal{V})$ - implicit (need to solve eqs. - System 1 & System 2 ML may be useful `https://arxiv.org/abs/2103.14251` L. Pagnier & MC)

Task: State & Parameter Estimation

- Inverse PF Map: $\boldsymbol{S} = \boldsymbol{\Pi}_{\boldsymbol{Y}}^{-1}(\boldsymbol{V})$ – explicit (do not need to solve eqs. – System 1 and System 2 ML may be useful https://arxiv.org/abs/2102.06349 L. Pagnier and MC)

- State Estimation
  - Full Observability: given $\mathcal{G}$ and $\boldsymbol{Y}$ to estimate injected/consumed active and reactive powers = application of the inverse PF map, $\boldsymbol{\Pi}^{-1}$
  - Limited Observability:
    - Complement Missing power injections/consumptions at the nodes where voltages and phases are measured
    - Challenging Version: to reconstruct injected/consumed powers and also voltages and phases at all nodes of the system. (super-resolution – will not discuss)
- Parameter Estimation
  - Reconstruct Graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, and line characteristics, $\boldsymbol{Y}$

Outline

How does model reduction work?

- Ground Truth – reliable but computations "heavy"
- $\Rightarrow$
- Reduced Model – lighter computations-wise, loosing some accuracy (but hopefully not too much)

Transient (seconds) Dynamics of the grid

- Swing Equation: $m_i\ddot{\theta}_i + d_i\dot{\theta}_i = p_i - \sum_j v_i v_j b_{ij}(\theta_i - \theta_j)$
- Reduced Model Options?

How does model reduction work?

- **Ground Truth** – reliable but computations "heavy"
- $\Rightarrow$
- Reduced Model – lighter computations-wise, loosing some accuracy (but hopefully not too much)

Transient (seconds) Dynamics of the grid

- Swing Equation: $m_i\ddot{\theta}_i + d_i\dot{\theta}_i = p_i - \sum_j v_i v_j b_{ij}(\theta_i - \theta_j)$
- Reduced Model Options?

PDE as the Reduced Model

- $m(\boldsymbol{r})\frac{\partial^2}{\partial t^2}\theta(t;\boldsymbol{r}) + d(\boldsymbol{r})\frac{\partial}{\partial t}\theta(t;\boldsymbol{r}) = p(t;\boldsymbol{r}) + \sum_{\alpha,\beta=1,2} \partial_{r_\alpha} b_{\alpha\beta}(\boldsymbol{r})\partial_{r_\beta}\theta(t;\boldsymbol{r})$

How does model reduction work?

- **Ground Truth** – reliable but computations "heavy"
- $\Rightarrow$
- Reduced Model – lighter computations-wise, loosing some accuracy (but hopefully not too much)

Transient (seconds) Dynamics of the grid

- Swing Equation: $m_i\ddot{\theta}_i + d_i\dot{\theta}_i = p_i - \sum_j v_i v_j b_{ij}(\theta_i - \theta_j)$
- Reduced Model Options?

PDE as the Reduced Model

- $m(\mathbf{r})\frac{\partial^2}{\partial t^2}\theta(t;\mathbf{r}) + d(\mathbf{r})\frac{\partial}{\partial t}\theta(t;\mathbf{r}) = p(t;\mathbf{r}) + \sum_{\alpha,\beta=1,2} \partial_{r_\alpha} b_{\alpha\beta}(\mathbf{r})\partial_{r_\beta}\theta(t;\mathbf{r})$
- Why is Partial Differential Equation modeling a sound option for model reduction?

Why is PDE a sound option for model reduction?

Approximating the swing ODEs by a PDE? Really?

- Naively: increases # degrees of freedom

Why is PDE a sound option for model reduction?

Approximating the swing ODEs by a PDE? Really?

- Naively: increases # degrees of freedom

... but thinking a bit more (system 2):

- It has a sense because
  - Solutions of linear 2+1 dimensional PDE assume spatial regularization via a 2d grid with fewer # grid points
  - Operations are much more efficient over a regular grid
  - # physical parameters can be reduced dramatically via coarsening – fewer & large-scale harmonics

Why is PDE a sound option for model reduction?

Approximating the swing ODEs by a PDE? Really?

- Naively: increases # degrees of freedom

... but thinking a bit more (system 2):

- It has a sense because
  - Solutions of linear 2+1 dimensional PDE assume spatial regularization via a 2d grid with fewer # grid points
  - Operations are much more efficient over a regular grid
  - # physical parameters can be reduced dramatically via coarsening – fewer & large-scale harmonics

Inspired by – 1+1 PDE modeling of PS:

- A. Semlyen, 1974.
- J. S. Thorp, C. E. Seyler, and A. G. Phadke, 1998.
- M. Parashar, J. S. Thorp, and C. E. Seyler, 2004.

Why is PDE a sound option for model reduction?

Approximating the swing ODEs by a PDE? Really?

- Naively: increases # degrees of freedom

... but thinking a bit more (system 2):

- It has a sense because
  - Solutions of linear 2+1 dimensional PDE assume spatial regularization via a 2d grid with fewer # grid points
  - Operations are much more efficient over a regular grid
  - # physical parameters can be reduced dramatically via coarsening – fewer & large-scale harmonics

How can we make it work?

- In the Core of **This Talk** !

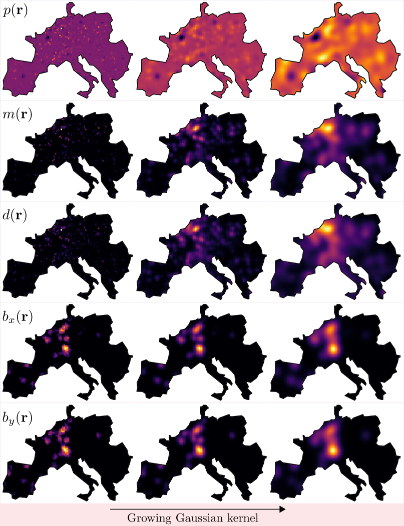
From Swing Model to PDE Model

- From Swing Equation: $m_i\ddot{\theta}_i + d_i\dot{\theta}_i = p_i - \sum_j v_i v_j b_{ij}(\theta_i - \theta_j)$
- To PDE as the Reduced Model
  $m(\mathbf{r})\frac{\partial^2}{\partial t^2}\theta(t;\mathbf{r}) + d(\mathbf{r})\frac{\partial}{\partial t}\theta(t;\mathbf{r}) = p(t;\mathbf{r}) + \sum_{\alpha,\beta=1,2} \partial_{r_\alpha} b_{\alpha\beta}(\mathbf{r})\partial_{r_\beta}\theta(t;\mathbf{r})$
- $\forall i: \quad \theta_i(t) \to \theta(t;\mathbf{r}),\ m_i \to m(\mathbf{r}),\ d_i \to d(\mathbf{r}),\ p_i(t) \to p(t;\mathbf{r}),\ b_{ij} \to b_{\alpha\beta}(\mathbf{r}),\ \forall \alpha,\beta = 1,2.$

Neumann Boundary Conditions:

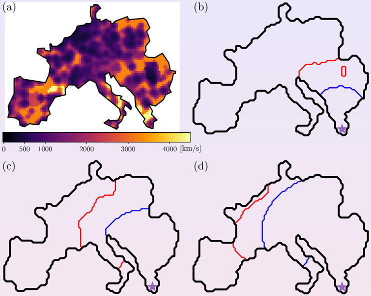
- Vanishing normal derivative of the angle field on the domain boundary $\partial\Omega$:
  $\forall t,\ \forall \mathbf{r} \in \partial\Omega: \quad \sum_{\alpha,\beta=1,2} n_\alpha(\mathbf{r}) b_{\alpha\beta}(\mathbf{r})\partial_{r_\beta}\theta(t;\mathbf{r}) = 0$
- e.g. guaranteeing equilibration to the same frequency

Initialization Parameters: Artificial Diffusion & Calibration



- Artificial Diffusion (AD) – "diffusive growth" of non-uniform distribution of parameters
- Generalization of the metodology developed in M. Parashar, J. S. Thorp, and C. E. Seyler (2004) for 1+1 PDEs
- AD is stopped when parameters satisfy some smoothness criterion – advantageous because it allows the optimal width of the Gaussian kernel to be self-determined

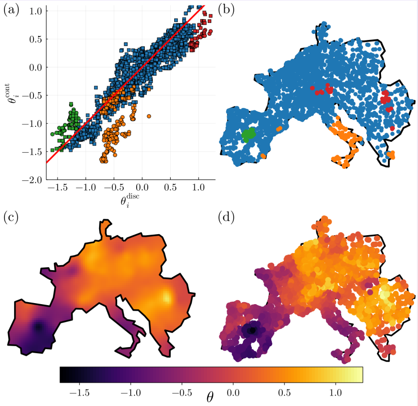
Speed of EM waves: Inhomogeneous Map



- PanTaGruEl model: 3809 buses, 618 generators and 4944 lines. (3221 nodes in the "full" discretization of our PDE model.)

- (a) Assessment of the local propagation speed as $c(\mathbf{r}) = \sqrt{b(\mathbf{r})/m(\mathbf{r})}$.
- (b)-(d) Fronts of the perturbation at incremental time intervals of $\Delta t = 0.6$s, after a fault in Greece (violet star), for inhomogeneous (red) and average parameters (blue) – slower.

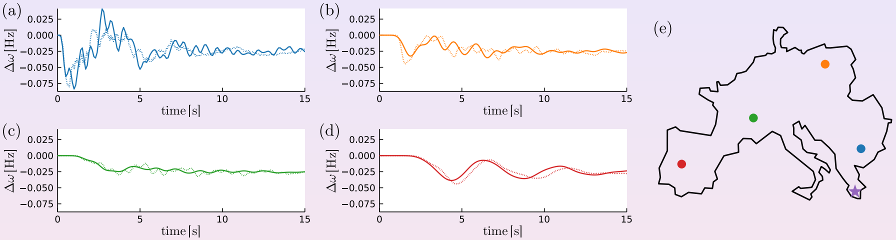
Steady State Test & Adjustments



- (a) One-to-one comparison of local voltage angle: for each bus in the discrete model the nearest node in the continuous mesh is selected. The red line indicates a perfect match.
- (b) The outliers marked in orange, red and green correspond to the points marked on the map in (b). The square markers correspond the solution after adjusting the susceptances.
- (c) PDE solution $\theta(\boldsymbol{r})$ after adjustment.
- (d) GT (ODEs) solution $\theta^{\rm disc}$.

Steady State: PDE vs Ground Truth (ODE)

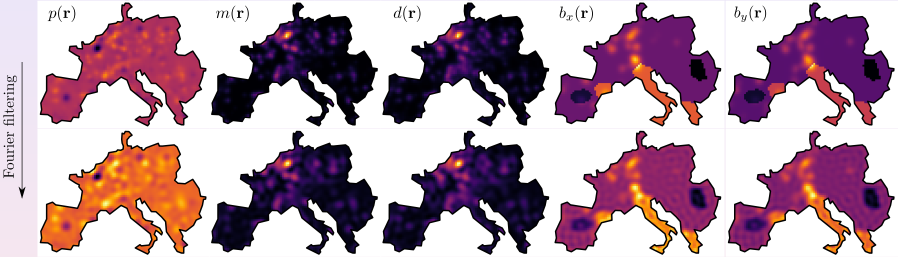
Frequency Response of Generators



PDE vs Ground Truth (ODE)

- Response in (a) Bulgaria, (b) Poland, (c) France, and (d) Spain to a 900 MW loss of power in Greece.
- dotted – PDE, solid – Ground Truth (ODEs)

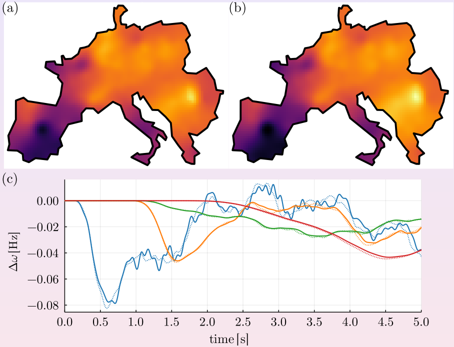
Coarse-Graining/Filtering



Distribution of the grid parameters: Low-Pass Filtering

- Original vs a Fourier Low-Pass Filter with a cut-off frequency 30% (of the maximum).

Coarse-Graining/Filtering



- (a) – original vs (b) – filtered: Comparison of Steady state solution
- (c) Frequency response original (solid) vs fitlered (dotted)

30% of filtering – almost no loss accuracy

What did we achieve so far?

- **Construction** of the **reduced** PDE model (of the ODE/swing equations). Included:
  1. accurate resolution of the **boundary conditions**
  2. efficient and flexible identification of parameters based on **Artificial Diffusion** and **Fourrier Filtering**
- **Validation** via **Static** and **Dynamic** Tests – reduced PDE vs Ground Truth (ODEs)
- **Observation:** — PDE offers significant gain in efficiency:
  1. Evaluation of PDE is faster at least factor of ten (for the same number of discretization points)
  2. Can run PDE at much lower resolution
  3. Can use much fewer degrees of freedom ⇒ **to learn**

Work in Progress: Towards Physics (System 2) Informed ML

- Functional maps for $m(\boldsymbol{r}), d(\boldsymbol{r})$ and $b_{\alpha\beta}(\boldsymbol{r})$ will be modeled as Neural Networks (System 1)
- Artifical Difussion (AD) for the warm start

Work in Progress: Towards Physics (System 2) Informed ML

- Functional maps for $m(\boldsymbol{r}), d(\boldsymbol{r})$ and $b_{\alpha\beta}(\boldsymbol{r})$ will be modeled as Neural Networks (System 1)
- Artifical Difussion (AD) for the warm start

Goal — efficient & accurate evaluation of multiple scenarios

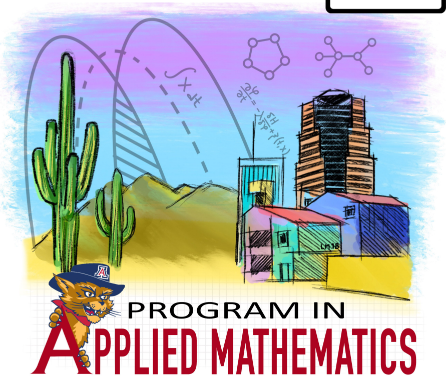
- Auotmatic & much faster than an individual Dynamic Simultation of today (also faster than real dynamics)

- Research focused, since 1976, one of the US first [dynamical systems, integrability, turbulence …]
- Interdisciplinary: 100+ professors/ 26 departments/ 8 colleges across UA campus (CoS & CoE & Optics – top 3)
- Mixing traditional @ contemporary Applied Math
- Graduate, Ph.D. focused, no terminal M.Sc.
- 60 Ph.D students (**13/**16/10 enrolled in **2021**/20/19)
- **3 Core Courses** (1st year -- Methods, Analysis, Algorithms) https://appliedmath.arizona.edu/students/new-core-courses
- Strong collaborations with Industry (e.g. Raytheon, Uber, Intel, Critical Path, etc) and National Labs (e.g. LANL, LLNL, NREL, NNSS, etc)
- 5 seminar/colloquium series – recorded and posted online
- Participation in many UA & National Edu Projects

http://appliedmath.arizona.edu/

chertkov@arizona.edu

SCAN ME



Support is Appreciated !!

- **Energy Systems**: UArizona start up + DOE/ARPA-E

Thanks for your attention !

Outline

Task SE & PE. Reduced Modeling.

- Setting of Partial Observability
- Find Equivalent (Reduced) Model of Power System
- "Inspired" by Kron Reduction
  - $\boldsymbol{I}^{(o)} = \boldsymbol{Y}^{(r)}\boldsymbol{V}^{(o)}$
  - "o" - observed; "r" - reduced
  - $\mathcal{G}^{(r)} \equiv (\mathcal{V}^{(o)}, \mathcal{E}^{(r)})$
  - $\boldsymbol{Y}^{(r)} \doteq (\{a,b\}|Y_{ab}^{(r)} \neq 0)$ – associated with the effective (not necessarily real) power lines, $\{a,b\} \in \mathcal{E}^{(r)}$. $\boldsymbol{Y}^{(r)}$
- Reduced Model
  - $\boldsymbol{S}^{(o)} = \Pi^{-1}_{\boldsymbol{Y}^{(r)}}(\boldsymbol{V}^{(o)})$
  - Learn it !?

Task: SE & PE. PIML of Reduced Model

- Power Graphical NN (System 2):

$$\min_{\varphi, \boldsymbol{Y}^{(r)}} L_{\text{Power-GNN}}\left(\varphi, \boldsymbol{Y}^{(r)}\right),$$

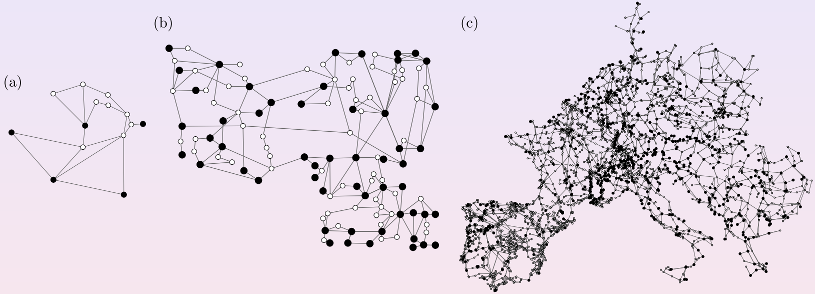
$$L_{\text{Power-GNN}}\left(\varphi, \boldsymbol{Y}^{(r)}\right) \equiv \frac{1}{N|\mathcal{V}^{(o)}|} \sum_{n=1}^{N} \Bigg\| \boldsymbol{S}_n^{(o)} - \underbrace{\Pi^{-1}_{\boldsymbol{Y}^{(r)}}\left(\boldsymbol{V}_n^{(o)}\right)}_{\text{physics = interpretable}} - \underbrace{\Sigma_{\varphi}\left(\mathcal{V}_n^{(o)}, S_n^{(o)}\right)}_{\text{NN = "sub-scale"}} \Bigg\|^2 + \underbrace{\mathcal{R}(\varphi)}_{\text{regularization}}$$

  - SIMULTANEOUSLY physics-informed and physics-blind parts
  - Compare with Vanilla-NN (System 1)

$$L_{\text{NN}} \doteq \frac{1}{N|\mathcal{V}^{(\mathbf{o})}|} \sum_{n=1}^{N} \left\| \boldsymbol{S}_n^{(o)} - \text{NN}_{\varphi}(\boldsymbol{V}_n^{(o)}) \right\|^2$$

Task: SE & PE. Power GNN vs Vanilla NN. Experiments.

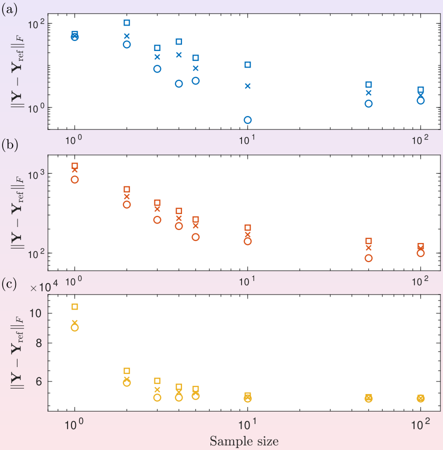
IEEE 14-bus [panel (a)], IEEE 118-bus [panel (b)] and PanTaGruEl [panel (c)] models



State Estimation Test: Six set of samples were generated for each network. Average mismatch of predicted power injections (on the training set in parentesis)

	case #1	case #2	case #3	case #4	case #5	case #6
Vanilla NN	4.9E-6	7.2E-5	6.3E-3	5.2E-2	6.3E-2	1.4E0
	(4.2E-6)	(6.6E-5)	(5.0E-5)	(3.7E-5)	(1.2E-4)	(4.2E-6)
Power-GNN	3.0E-6	5.8E-7	6.9E-7	1.3E-6	2.9E-7	3.0E-6

Task: SE & PE. Power GNN vs Vanilla NN. Experiments.



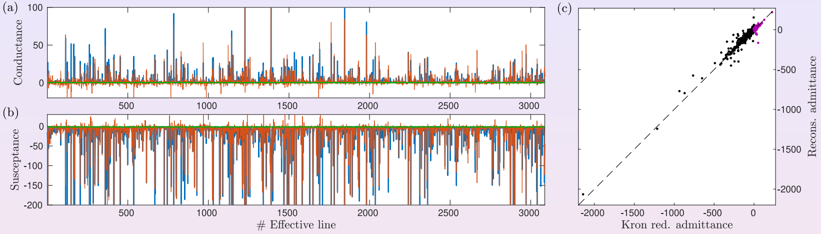
Full Observability. Parameter Estimation.

- Reconstruction of the admittance matrix $\boldsymbol{Y}$ for IEEE 14-bus (a), IEEE 118-bus (b) and PanTaGruEl (c) models
- The min, mean and max values are displayed as circles, crosses and squares respectively (for 10 realizations.)

Notice !!

- Quality of the reconstruction by Power-GNN – especially for large network

Task: SE & PE. Power GNN vs Vanilla NN. Experiments.



Partial Observability. Parameter Estimation. PanTaGruEl model

- Initial (pre-training) values – in green.
- Trained values and their Kron-reduction counterparts – red and blue respectively.
- (c) shows alternative visualization of the reference-vs-predicted values of the line conductances (purple) and susceptances (black)

Notice !!

- Quality of the reconstruction by Power-GNN – especially for large network