Machine Learning for Power Systems: From Pure Data-Driven to Physics-Informed Methods

Prof. Nanpeng Yu
Department of Electrical and Computer Engineering
Department of Computer Science
(cooperating faculty)
Director of Energy, Economics, and Environment Research Center
nyu@ece.ucr.edu
951.827.3688

UNIVERSITY OF CALIFORNIA, RIVERSIDE
Team Members

- Energy, Economics, and Environment (E3) Research Center
  - Director: Dr. Nanpeng Yu.
  - Over 40 affiliated faculty in CSE, ECE, STAT and SPP

- Project Scientists and Postdoctoral Researchers
  - Dr. Koji Yamashita, Dr. Mikhail Bragin, Dr. Yinglun Li, and Dr. Zuzhao Ye

- PhD and MS Students
  - Yuanbin Cheng, Osten Anderson, Jingtao Qin, Joseph Brown, Maojie Tang, Zhixuan Tan, Shaorong Zhang, and Jeffrey Chan

- Alumni PhD Students

Dr. Brandon Foggo
Machine Learning Scientist at Habitat Energy

Dr. Wei Wang
Senior Data Scientist at Walmart Labs

Dr. Yuanqi Gao
Machine Learning Engineer at Lucid Motors

Dr. Jie Shi
Postdoctoral Researcher at Cornell

Dr. Wenyu Wang
Senior Engineer at Quanta Technology

Dr. Farzana Kabir
EV Grid Infrastructure Specialist at California Energy Commission

Dr. Yinglun Li
Senior ML Scientist at Alibaba International Digital Commercial Group

Dr. Zuzhao Ye
Founder of AmpTrans
Research Projects and Sponsors

- Over $16 Million of R&D Funding in the past 10 years
  - Federal government, state agencies, private companies
- Research Tracks
  - Physics-informed Machine Learning for Power Systems
  - Scalable Optimization in Critical Infrastructure Systems
  - Transportation and Building Electrification
  - Decarbonization Planning
  - Energy Efficient Data Center
Outline

- Volume, Variety, Velocity & Value of Big Data in Power Systems
- Applications of Machine Learning in Power Systems
  - Transmission system, distribution system, end-use customers
- Motivation for Physics-informed Methods in Power Systems
- Leverage Unique Properties of Data from Power Systems
  - Low rank and sparsity, high and low entropy
- Integrate Machine Learning Models with Physical Models of Power Systems
  - Renewable resource model, power flow model
  - Topology of power network, system control model
  - Power system dynamic model
- Summary and Discussions
Volume and Variety

▶ Variety of Big Data in Power Systems
  ▶ Supervisory control and data acquisition (SCADA) system
  ▶ Phasor Measurement Unit (PMU), Micro-PMU
  ▶ Digital Fault Recorder, Equipment Monitors
  ▶ Census Data, Advanced Metering Infrastructure (AMI)
  ▶ Weather Station, Electricity Market
  ▶ Geographical Information System

▶ Volume of Big Data in Power Systems
  ▶ PMUs
    ▶ Over 2,500 PMU (10,000 measurement channels) in the U.S.
    ▶ Roughly 450 TB of data generated annually (60Hz)
  ▶ AMI
    ▶ Over 1.1 billion smart meter installations worldwide as of 2022.
    ▶ Over 2 petabytes of smart meter data.
Velocity and Value

- **Velocity of Data in Power Systems**
  - **Sampling Frequency**
    - AMI Data: Sampling frequency increases from once a month to 1 reading every 15 mins
    - PMU: Mature technology (30 – 240 Hz), continuous point on wave (1920 – 61,400 Hz)
  - **Bottleneck in Communication Systems of Distribution Network**
    - Limited bandwidth for ZigBee network
    - Most of the utilities in the US receives smart meter data with ~24 hour delay
  - **Edge Computing Trend**
    - Itron and Landis+Gyr extend edge computing capability of smart meters
    - Centralized → distributed / decentralized monitoring, computing, control and decision making

- **Value of Big Data Analytics in Power Systems**
  - According to GTM Research, electric utilities around the world spent over $3.8 billion on data analytics solutions in 2020.
  - Analysis from Indigo Advisory group indicates that the market for AI in the energy sector could be worth $13 billion as of the end of 2023.
Applications of Machine Learning in Transmission Systems and Electricity Market

Electricity Market Applications
Price & Load Forecasting, Algorithmic Trading

Learn Power System Dynamics
Accelerate UC & OPF

Event Detection & Classification
Generator Tripping

Equipment Monitoring
Identify Faulty Equipment in Substations

Model Validation & Parameter Estimation

State Estimation
Linear State Estimation
Applications of Machine Learning for Power Distribution Systems and End Use Customers

Spatio-temporal Forecasting
Electric Load / DERs – Short-Term / Long-Term

Anomaly Detection
Electricity Theft, Unauthorized Solar Interconnection

System Monitoring
State Estimation & Visualization

Distribution System Controls
Deep Reinforcement Learning

Network Topology and Parameter Identification
Transformer-to-customer, Phase connectivity, Impedance estimation

Equipment Monitoring
Predictive Maintenance Online Diagnosis

Customer Behavior Analysis
Customer segmentation, nonintrusive load monitoring, demand response
Physics-Informed ML for Power Systems: Motivation

- Purely Data-driven ML algorithms: widely adopted by researchers and practitioners to solve a myriad of problems in power systems.
  - Big data: e.g. load & price forecasting, predictive maintenance of transformers
  - Struggle to deal with system monitoring, sequential decision-making, large-scale optimization, and control problems in power systems.

- Technical Challenges
  - Accuracy, generalization capability, sample efficiency, safety and interpretability

- Physics-informed ML Algorithms
  - Synergistic combination of machine learning & physical model or information
  - Embed domain knowledge, unique data property, system model in ML algorithms
  - Introduce inductive bias, improve explainability, and generalize to unforeseen scenarios from a finite training dataset
Low Rank and Sparsity Properties of Power System Measurement Data

Rank of the matrix: the number of linearly independent rows or columns in the matrix
Low Rank and Sparsity: Voltage Event Detection Using Optimization with Structured Sparsity Inducing Norms

Key Observations

- Voltage related events trigged by system faults are often regional events
- The $X - L$ during voltage event periods have row-sparse structure
- Rows of residual matrix correspond to PMUs highly impacted by the event

Main Idea

- Decompose the streaming PMU data matrix $X$ into
  - A low-rank matrix $L$, a row-sparse event-pattern matrix $S$, and a noise matrix $G$
- Extract anomaly features from $L$ & $S$
- Use clustering algorithm to identify power system voltage events

Fig. 2. The heatmap of “$X - L$” (left) and “$X - L - G$” (right) for normalized active power data (scaled from 0 to 1). The event happens approximately at the red line.

Decompose PMU Data Matrix with Proximal Bilateral Random Projection (PBRP) to Detect Events

\[
\min_{L,S} \frac{1}{2} \|X - L - S\|_F^2 \\
\text{s.t.} \quad \text{rank}(L) = r, \\
\quad S \text{ is row-sparse.}
\]

⇒

\[
\min_{L,S} \frac{1}{2} \|X - L - S\|_F^2 + \lambda \|S\|_{21} \\
\text{s.t.} \quad \text{rank}(L) = r, \\
\quad L^{(k)} = \arg \min_{\text{rank}(L) = r} \frac{1}{2} \|X - L^{(k-1)} - S^{(k-1)}\|_F^2 \\
\quad S^{(k)} = \arg \min_S \frac{1}{2} \|X - L^{(k)} - S\|_F^2 + \lambda \|S\|_{21}
\]

Solution Approach: Coordinate Descent

Residual PMU data matrices during voltage events have distinctive sparsity structure

Computationally efficient PBRP algorithm is proposed to decompose PMU data matrices

Online voltage event detection algo. shows better accuracy & scalability on PMU data (Eastern Interconnection)
Low and High Entropy Power System Measurements

Information entropy of data: the average amount of information conveyed by an event, when considering all possible outcomes

$$H(x) = - \sum_{x \in X} p(x) \log p(x)$$
High & Low Entropy of Dataset: Information Loading-based Regularization

Background
- Abstract Representation of Deep Neural Network based Classifier

Main Idea
- Control the amount of information compression between the input layer and the last hidden layer of a deep neural network
- Balance memorization and generalization

Algorithm
- Augment the typical cross-entropy loss function with estimated mutual information between the input layer and the hidden representation

\[ L_T = L_{CE} - \beta \hat{I}(X;Z) \]

Phase Connectivity Identification

- Very few electric utility companies have completely accurate phase connectivity information in GIS!
- Validated using real-world distribution circuits data from SCE and PG&E.

---

**System Event Classification with PMU Data**

- Formulated as a classification problem
- Normal operation condition, line event, generator event, oscillation event

![PMU Data Diagram]

**Learned Representation**

- (a) Baseline
- (b) Baseline+info
- (c) Baseline+GSP
- (d) Baseline+GSP+info

**F1 Score on Testing Dataset from Eastern Interconnection**

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Non-event</th>
<th>Line-event</th>
<th>Generator event</th>
<th>Oscillation event</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td>0.554</td>
<td>0.879</td>
<td>0.881</td>
<td>0.208</td>
</tr>
<tr>
<td><strong>Baseline+info</strong></td>
<td>0.596</td>
<td>0.928</td>
<td>0.924</td>
<td>0.205</td>
</tr>
<tr>
<td><strong>Baseline+GSP</strong></td>
<td>0.894</td>
<td>0.937</td>
<td>0.907</td>
<td>0.922</td>
</tr>
<tr>
<td><strong>Baseline+GSP+info</strong></td>
<td>0.973</td>
<td>0.971</td>
<td>0.962</td>
<td>0.986</td>
</tr>
</tbody>
</table>

Renewable Energy Resource Model

Solar PV System Performance Model: used to understand and predict energy or power output from PV systems under a wide range of environmental, design, and site.
Physical Solar PV Performance Model

Estimation of Behind-the-Meter Solar Generation

Net Metering: \( net\ load = load - solar\ generation \)

Mixture Hidden Markov Model for Load Modeling

\[
L_{nt} = a_{zt} + b_n + c_{n,zt}x_t + \epsilon_{nt}, \quad n = 1 \ldots N,
\]

\[
t = 1 \ldots T, \quad b_n \sim N(0, \sigma^2), \quad \epsilon_{nt} \sim N(0, \lambda_{zt}^2)
\]

Disaggregation method

\[
\arg\min_{\theta_S} \sum_{t=1}^{T} (S_t - g_t(\theta_S))^2
\]

subject to \( \theta_{S,min} \leq \theta_S \leq \theta_{S,max} \)

\( g_t(\theta_S) \): solar PV generation at time \( t \) based on the physical solar PV system performance model with the technical parameters \( \theta_S \).

\[
P_{ac} = g(\theta_S) = \eta(\eta_{nom}, P_{dc})P_{dc}
\]

\[
P_{dc} = g'(P_{dc0}, \theta_t, \theta_{az}, l) = (1 - l) \times \frac{E_{tr}(\theta_t, \theta_{az})}{E_0} P_{dc0}[1 + \gamma(T_c(\theta_t, \theta_{az}) - T_0)]
\]
Testing Results

<table>
<thead>
<tr>
<th>Error Metric</th>
<th>Variable</th>
<th>MHMM (solar panel scenario 1)</th>
<th>HMM reg. (solar panel scenario 1)</th>
<th>MHMM (solar panel scenario 2)</th>
<th>HMM reg. (solar panel scenario 2)</th>
<th>Consumer Mixture Model</th>
<th>SunDance Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>Solar</td>
<td>0.13</td>
<td>0.19</td>
<td>0.12</td>
<td>0.18</td>
<td>0.37</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Load</td>
<td>0.13</td>
<td>0.19</td>
<td>0.12</td>
<td>0.18</td>
<td>0.37</td>
<td>0.49</td>
</tr>
<tr>
<td>MASE</td>
<td>Solar</td>
<td>2.13</td>
<td>2.61</td>
<td>2.11</td>
<td>2.58</td>
<td>3.85</td>
<td>3.74</td>
</tr>
<tr>
<td></td>
<td>Load</td>
<td>0.43</td>
<td>0.48</td>
<td>0.42</td>
<td>0.48</td>
<td>0.74</td>
<td>0.81</td>
</tr>
<tr>
<td>CV</td>
<td>Solar</td>
<td>0.47</td>
<td>0.58</td>
<td>0.45</td>
<td>0.57</td>
<td>0.77</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Load</td>
<td>0.29</td>
<td>0.33</td>
<td>0.28</td>
<td>0.32</td>
<td>0.46</td>
<td>0.57</td>
</tr>
</tbody>
</table>

- Validated using real-world smart meter and solar PV generation data from Austin, Texas.
- The MHMM follows the actual load much more closely than the benchmark algorithms during the low load periods.
- The synergistic combination of physical solar PV system performance model and the statistical community load model yields more accurate solar PV generation estimation.
- The MHMM allows sharing of info. across individual customers, which leads to more accurate load and solar PV generation estimates.

Steady State Model of Power Systems

Power Flow Model: Linearized or Nonlinear Model of Power Flow
Key Idea

There exists approximate linear models between power and voltage magnitudes to distribution secondaries.

A large discrepancy between estimated & measured power consumption data indicates potential theft.

Let $y_i(t) = x(t)^T \sum_{j=1}^{n_c} y_j(t) [\beta_i^x \beta_i^y] + \epsilon_i(t)$

Where $x(t) = [1, v_1(t), v_2(t), \cdots, v_{n_c}(t)]^T$, $y_i(t) = p_i(t)$

Let $\tilde{y}_i^{(e)}$ and $\tilde{y}_i$ denote the out-of-sample residual time series for the energy thief.

Lemma 1. $\tilde{y}_i^{(e)} - \tilde{y}_i = - \sum_{j \neq i} \beta_j^y y_i^s$

Lemma 1 & Lemma 3 combine to show that a thief’s residuals will become negative once he or she begins to steal power.

Lemma 2. $\sum_j \tilde{y}_j^{(e)} = \sum_j \tilde{y}_j = 0$

Lemma 2 shows that the residuals of the other customers will raise to balance their sum.

Lemma 3. $\forall \delta > 0$, there exists a training data window length $T > 0$ such that for each $j$: $\mathbb{P} \left( \beta_j^y \geq -\delta \right) > 1 - \delta$

Testing Results with Real-world Smart Meter Data

- 12 KV circuit from Southern California Edison, 6 months of smart meter data from 980 customers and 190 transformers.
- The average electric load consumed by the customer is 1.6 kWh.
- The mean of the estimation residual is -0.01 kWh and its standard deviation is 0.1 kWh.
- Anomaly score for customer \( k \) is much higher than that of all other customers in the feeder.

<table>
<thead>
<tr>
<th>( d_k )</th>
<th>Ranking</th>
<th>( \sum_t d_i/N )</th>
<th>PR(( d_i, 95 ))</th>
<th>( \max_{i \neq k} d_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>79.4</td>
<td>1 (0.1%)</td>
<td>8.1</td>
<td>14.6</td>
<td>42.2</td>
</tr>
</tbody>
</table>

- Anomaly score increases monotonically with the amount of stolen electricity.
- In all cases, customer \( k \)’s anomaly score will surpass the 95th percentile of all customers if it steals more than 32 kWh in two weeks or 0.19 kWh per hour.
- In cases 1-3, if customer \( k \) steals more than 0.38 kWh of power per hour, then its anomaly score will the largest of all customers.
Physics-informed Phase Connectivity Identification in Power Distribution System

The time difference version of the physical model of distribution network

\[ \tilde{v}(t) = X\tilde{v}^{ref}(t) + X\hat{K}X^T\tilde{p}(t) + X\hat{L}X^T\tilde{q}(t) + n(t) \]

The likelihood of observing \( \{\tilde{v}(t)\}_{t=1}^T \), given \( x \), \( \{\tilde{p}(t)\}_{t=1}^T \) and \( \{\tilde{q}(t)\}_{t=1}^T \) is

\[ 
Prob(\{\tilde{v}(t)\}_{t=1}^T \mid \{\tilde{p}(t)\}_{t=1}^T, \{\tilde{q}(t)\}_{t=1}^T; x) 
= \frac{|\Sigma_N|^{-\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T} [\tilde{v}(t) - \tilde{v}(t, x)]^T \Sigma_N^{-1} [\tilde{v}(t) - \tilde{v}(t, x)] \right\} 
\]

Lemma 1. Let \( x^* \) be the correct phase connection. If the following two conditions are satisfied, then as \( T \to \infty \), \( x^* \) is a global optimizer of \( f(x) \).

1. \( n(t_k) \) is i.i.d. and independent of \( \tilde{v}^{ref}(t_l), \tilde{p}(t_l), and \tilde{q}(t_l) \), for \( \forall t_k, t_l \in Z^+ \).
2. \( \tilde{v}^{ref}(t_k), \tilde{p}(t_k), and \tilde{q}(t_k) \) are independent of \( \tilde{v}^{ref}(t_l), \tilde{p}(t_l), and \tilde{q}(t_l) \), for \( \forall t_k, t_l \in Z^+, t_k \neq t_l \).

Directly minimizing \( f(x) \) is difficult.

Key Idea: phase identification problem \( \rightarrow \) maximum marginal likelihood estimation (MMLE) problem.

Numerical Results with Smart Meter Data

Number of Loads per Phase in the IEEE Test Circuits

<table>
<thead>
<tr>
<th>Feeder</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>BC</th>
<th>CA</th>
<th>ABC</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>37-bus</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>123-bus</td>
<td>18</td>
<td>17</td>
<td>17</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>5</td>
<td>85</td>
</tr>
<tr>
<td>342-bus</td>
<td>30</td>
<td>38</td>
<td>31</td>
<td>35</td>
<td>31</td>
<td>33</td>
<td>10</td>
<td>208</td>
</tr>
</tbody>
</table>

Smart Meter Data

- Length: 90 days of hourly average real power consumption data
- Provided by FortisBC
- Measurement noise ~ 0.1 and 0.2 accuracy class smart meters established in ANSI.
- Voltage measurements rounded to the nearest 1 V for the primary loads & 0.1 V for the secondary loads.

Phase Identification Accuracy of Different Methods with 90 days of Meter Data

<table>
<thead>
<tr>
<th>Method</th>
<th>Meter Class</th>
<th>37-Bus Feeder</th>
<th>123-Bus Feeder</th>
<th>342-Bus Feeder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation-based Approach</td>
<td>0.1%</td>
<td>100%</td>
<td>98.75%</td>
<td>81.82%</td>
</tr>
<tr>
<td></td>
<td>0.2%</td>
<td>100%</td>
<td>97.5%</td>
<td>81.31%</td>
</tr>
<tr>
<td>Clustering-based Approach</td>
<td>0.1%</td>
<td>100%</td>
<td>100%</td>
<td>93.43%</td>
</tr>
<tr>
<td></td>
<td>0.2%</td>
<td>100%</td>
<td>98.75%</td>
<td>91.41%</td>
</tr>
<tr>
<td>MMLE-based Algorithm</td>
<td>0.1%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>0.2%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

- The MMLE-based algorithm outperforms the correlation and clustering-based approaches.
- The improvement in accuracy increases as the complexity of the distribution feeder increases.
Graphical Model of Power Network

Graph Model for Power Systems: Nodes → Vertices, Power Lines → Branches
Graph Model: Physics-informed Graphical Learning for Distribution Line Parameter Estimation

- Key Idea: Embed physical equations of power flow in the graphical learning model
- Inspired by graphical neural network (GNN)
- Difference between physics-informed GL and GNN
  - Leverage 3Φ power flow-based physical transition fcn. to replace the deep neural networks in GNN.
- Key Step: Derive the gradient of voltage magnitude loss function w.r.t. line segment’s resistance and reactance parameters with an iterative method.

- Estimate distribution line parameters with SGD considering prior estimates of line parameters and physical constraints.
- Improve computation efficiency with grid partition scheme and fast forward/backward function.

Fast GL and Numerical Study Results

- 15 days of smart meter data from 13.2 kV distribution circuit in National Grid
- 177 line sections, 491 loads, 23 solar PV systems
- Feeder partitioned into 10-subnetworks (parallel estimation)
- Fast Graphical Learning has significantly higher MADR improvement (30% improvement in MADR)
- Approximately 10 times reduction in computation time

MADR Improvement of Parameter Estimation Methods in the Test Feeder (Avg. / Choose Optimal Value)

<table>
<thead>
<tr>
<th>Network</th>
<th>LMLE</th>
<th>FGL</th>
<th>FGL+CON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole Network</td>
<td>10.8% / 13.5%</td>
<td>20.7% / 25.5%</td>
<td>29.4% / 30.9%</td>
</tr>
<tr>
<td>Sub-Net 1</td>
<td>10.3% / 9.1%</td>
<td>20.1% / 21.6%</td>
<td>23.1% / 27.1%</td>
</tr>
<tr>
<td>Sub-Net 2</td>
<td>7.3% / 9.2%</td>
<td>13.6% / 20.9%</td>
<td>26.7% / 29.3%</td>
</tr>
<tr>
<td>Sub-Net 3</td>
<td>9.9% / 12.2%</td>
<td>34.5% / 40.8%</td>
<td>41.6% / 43.7%</td>
</tr>
<tr>
<td>Sub-Net 4</td>
<td>4.5% / 4.7%</td>
<td>5.1% / 5.0%</td>
<td>12.4% / 13.2%</td>
</tr>
<tr>
<td>Sub-Net 5</td>
<td>11.6% / 14.8%</td>
<td>20.8% / 22.1%</td>
<td>21.0% / 22.3%</td>
</tr>
<tr>
<td>Sub-Net 6</td>
<td>12.0% / 24.3%</td>
<td>37.0% / 62.4%</td>
<td>61.8% / 63.5%</td>
</tr>
<tr>
<td>Sub-Net 7</td>
<td>9.3% / 9.9%</td>
<td>16.2% / 18.0%</td>
<td>31.6% / 32.9%</td>
</tr>
<tr>
<td>Sub-Net 8</td>
<td>13.9% / 16.5%</td>
<td>22.3% / 23.0%</td>
<td>24.9% / 25.5%</td>
</tr>
<tr>
<td>Sub-Net 9</td>
<td>17.3% / 21.3%</td>
<td>30.8% / 34.5%</td>
<td>37.9% / 38.2%</td>
</tr>
<tr>
<td>Sub-Net 10</td>
<td>7.6% / 9.6%</td>
<td>2.2% / 8.9%</td>
<td>19.0% / 20.4%</td>
</tr>
</tbody>
</table>

Avg. Runtime (Seconds) of Main Functions of Parameter Estimation Methods in Sub-networks of different sizes

<table>
<thead>
<tr>
<th>Method</th>
<th>Function</th>
<th>7-Bus</th>
<th>14-Bus</th>
<th>22-Bus</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGL</td>
<td>Fast-FORWARD</td>
<td>0.0142</td>
<td>0.0313</td>
<td>0.0519</td>
</tr>
<tr>
<td>FGL</td>
<td>Fast-BACKWARD</td>
<td>0.068</td>
<td>0.1046</td>
<td>0.1761</td>
</tr>
<tr>
<td>GL</td>
<td>FORWARD</td>
<td>0.3028</td>
<td>1.1603</td>
<td>3.2839</td>
</tr>
<tr>
<td>GL</td>
<td>BACKWARD</td>
<td>0.1057</td>
<td>0.3325</td>
<td>1.0065</td>
</tr>
<tr>
<td>LMLE</td>
<td>Gradient Calculation</td>
<td>0.0079</td>
<td>0.1866</td>
<td>0.8531</td>
</tr>
</tbody>
</table>

Solve Large-Scale MIP Problem with Generic GL

- Solving unit commitment (UC) or security constrained UC (SCUC) problems is crucial to market operations.
- The UC or SCUC problem boils down to solving large-scale mixed integer programing (MIP) problems.
- State-of-the-art commercial solvers (e.g. Gurobi, CPLEX) use sophisticated branch and bound algorithms.

\[ \min_x \ c_1^T x_1 + \cdots + c_n^T x_n \]
\[ a_{11} x_1 + \cdots + a_{1n} x_n \leq b_1 \]
\[ a_{21} x_1 + \cdots + a_{2n} x_n \leq b_2 \]
\[ \vdots \]
\[ a_{m1} x_1 + \cdots + a_{mn} x_n \leq b_m \]

- Used to solve large-scale MIP problems (e.g. UC)
- Drawbacks
  - Scalability: the number of constraints in the bipartite graph increases exponentially with the number of nodes in a power system


Google Deep Mind
Accelerate UC Solution with Physics-informed GL

- Learn to branch and dive for optimization problems on large-scale networks
- Instead of adopting bi-partite graph, leveraging power network to perform physics-informed GL

Outperforms commercial MIP solver (e.g. Gurobi and CPLEX) on large-scale UCR problems (6515-bus)

Better scalability, optimality and interpretability

System Control Model

System Control Model: Model-based Controller or Heuristic Control Policy from Human Operator
Control Model: Batch Constrained Reinforcement Learning-based Distribution System Control

- **Motivation**
  - Costly & time consuming to learn optimal control strategy by directly interacting with physical network.
  - Learning from finite historical dataset lead to large extrapolation errors.

- **Solution:** Batch-constrained soft actor-critic algorithm

- **Key Idea:** train a control policy
  - Maximize the total discounted return
  - Minimize dissimilarity between learned control policy & behavior policy of the batch data

---

Dynamic Model of Power Systems

Dynamic Model of Power System: Differential and Algebraic Equations (DAEs), Hamiltonian and Nearly Hamiltonian System
Dynamic Model: Generator Dynamic Parameter Estimation with Physics-based Neural Ordinary Differential Equations

Key Ideas

- Convert the forward solver of ODEs for power system dynamics into physics-informed neural networks.
- Calculate the loss function based on the difference between dynamic simulation results from the neural networks and pseudo PMU measurements.

- Calculate the gradient of the loss function w.r.t. dynamic parameters based on the neural ODE technique and the adjoint method.
- Update the dynamic parameters iteratively with a quasi-Newton Method.

Numerical Study Results

WECC 3-machine 9-bus system

- Generate noisy PMU measurements from dynamic simulation data.
- A single transmission line is disconnected at 5s. The simulation time is 10s.
- Two disturbance scenarios (between nodes 5 and 7, between nodes 8 and 9)
- When the PMU data length is 3s, our proposed algorithm achieves the lowest relative estimation error.
- The physics-based neural ODE algorithm outperforms the baseline algorithm in terms of estimation accuracy for most of the unknown parameters.
- Physics-based neural ODE algorithm has much shorter computation time than the baseline algorithm.
- When the data length is 3s, the running time of our model is just 4.82 seconds, which is nearly 8 times faster than the baseline model.
- Mini-batch scheme of neural network training shortens model running time.
Learning Dynamic System with Hamiltonian Neural Networks

- **Hamiltonian mechanics**: can predict the motion of an energy-conserved system.
- State variables: generalized position $\mathbf{q} = [q_1, q_2, \cdots, q_n]^T$ momentum $\mathbf{p} = [p_1, p_2, \cdots, p_n]^T$
- $\mathbf{q}$ and $\mathbf{p}$ correspond to voltage angle $\delta$ and angular speed $\omega$ in power system dynamic model
- Model single machine infinite bus system as a Hamiltonian: $m_1 \ddot{\delta} + d_1 \dot{\delta} + B_{12} V_1 V_2 \sin(\delta) - P_1 = 0$
- Can we learn the Hamiltonian or Lagrangian function (energy conservation law) directly?

The Hamiltonian can be regarded as the energy function

---

Can we formulate the SMIB system as a Hamiltonian function?

\[ m_1 \ddot{\delta} + d_1 \dot{\delta} + B_{12} V_1 V_2 \sin(\delta) - P_1 = 0 \]

Unfortunately, the answer is No. If the damping coefficient \(d_1\) is positive, then the SMIB system is a dissipative system instead of a conservative system.

---

**Hamiltonian System**

Hamiltonian differential equations

\[
\frac{dq}{dt} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q}
\]

Total energy is conserved:

\[
\frac{dH}{dt} = \frac{\partial^T H}{\partial q} \frac{dq}{dt} + \frac{\partial^T H}{\partial p} \frac{dp}{dt} = 0
\]

**Nearly Hamiltonian System**

Hamiltonian differential equations

\[
\frac{dq}{dt} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q} - D \frac{\partial H}{\partial p}
\]

Total energy decreases:

\[
\frac{dH}{dt} = \frac{\partial^T H}{\partial q} \frac{dq}{dt} + \frac{\partial^T H}{\partial p} \frac{dp}{dt} = -\frac{\partial^T H}{\partial p} D \frac{\partial H}{\partial p} \leq 0
\]

---


---

Fig. 2. Architecture of baseline neural network, HNN and NHNN
Summary

- Key: synergistic combine ML algorithms with physical domain knowledge
- Drastic improvement in accuracy, generalization capability, sample efficiency, safety and interpretability
- Types of power system domain knowledge or physical model
  - Unique data properties: low rank, high/low entropy
  - Steady state models: renewable resource model, topology, power flow model
  - Dynamic models: dynamic simulation and control models, Hamiltonian function
- How to combine physical model with machine learning model?
  - Iterative fitting the physical and data-driven models
  - Embed physics model in machine learning algorithms
    - Data preprocessing, loss function, neural architecture
Discussions

- Importance of Collaborating with Electric Utilities
  - Learn about current practice and realistic challenges
  - Work on solutions and datasets that could address real-world problems
    - Smart meter data for phase connectivity identification
    - PMU data from all three U.S. interconnections for system monitoring
- Real-world data is the best source for validating your ML algorithms
  - Simulation data does not possess the same property as real-world data
  - Algorithm trained with simulation data will struggle to solve real-world problems
- Reproducibility: Dataset and Open-Source Software
  - Learning Power System Dynamics with Neural ODE and NHNN
  - RL-based Control in Distribution System
  - Repository of synthetic PMU data generated from U.S. power grid data
    - pmuBAGE: https://github.com/NanpengYu/pmuBAGE
Recent Publications

Theoretical Machine Learning


Machine Learning for Power Transmission Systems

Recent Publications

Machine Learning for Power Distribution Systems and DERs

Contact Information:
Nanpeng Yu, nyu@ece.ucr.edu
Lab Website: https://intra.ece.ucr.edu/~nyu/

Thank You

Questions?

Collaborating Companies: SCE, PG&E, National Grid, Fortis BC, Con Edison, NYPA, ComEd, Exelon, EPRI, RPU, ISO-NE and CAISO.

Funding Support: DOE, NSF, CEC, NYSERDA, EPRI, DEED, UCOP, NREL & Utilities.