High-dimensional Data Analytics Using Low-dimensional Models in Power Systems

Meng Wang

Assistant Professor
Department of Electrical, Computer & Systems Engineering
Rensselaer Polytechnic Institute (RPI)
Troy, NY, USA
IEEE SGSMA 2019
May 20, 2019
Acknowledgment

Dr. Yingshuai Hao
Dr. Pengzhi Gao
Shuai Zhang
Ren Wang

Prof. Joe H. Chow
Big Data and Low-Dimensional Models

• Despite the ambient dimension, many high-dimensional datasets have intrinsic low-dimensional structures such as sparsity, low-rankness, and low-dimensional manifolds.

• These low-dimensional models enable the development of fast, model-free methods with provable performance guarantees for data recovery and information extraction.
Big Data in Power Systems

- Phasor Measurement Units (PMUs)
  - PMUs provide synchronized phasor measurements at a sampling rate of 30 or 60 samples per second.
  - Multi-channel PMUs can measure bus voltage phasors, line current phasors, and frequency. 2000+ PMUs in the North America.
  - Data availability and quality issues, e.g., data losses due to communication congestions.
  - Limited incorporation into the real-time operations.
Low Dimensionality of PMU data

- 6 PMUs measure 37 voltage/current phasors. 30 samples/second for 20 seconds.
- Singular values decay significantly. Mostly close to zero. Singular values can be approximated by a sparse vector.
- Low-dimensionality also used in Chen, Xie, Kumar 2013, Dahal, King, Madani 2012 for dimensionality reduction.
Convert Data to Information

Objective: Develop computationally efficient data-driven methods for power system situational awareness.

- PMU data quality improvement: missing data recovery, bad data correction, and detection of cyber data attacks.
- Real-time event identification through machine learning.
Outline

1. Motivation

2. Data Recovery and Error Correction

3. Event Identification

4. Conclusions
PMU Data Quality Issues

- Data losses and errors resulting from communication congestions and device malfunction.
- California Independent System Operator reported that 10%-17% of data in 2011 had availability and quality issues.
- Reliable data needed for real-time situational awareness and control.
Simultaneous and Consecutive Data Losses

A recorded PMU dataset: consecutive data losses on three phases of line for an hour.

Figure: Measured voltage phasor magnitudes

Figure: Measured current phasor magnitudes
Low-rank Matrix Completion

\[
\begin{bmatrix}
? & ? & ? \\
? & ? & ? \\
? & ? & ? \\
? & ? & ? \\
\end{bmatrix}
\]

- The problem includes theoretical analysis (e.g., Candes, Recht 2012) and recovery methods (e.g., nuclear norm minimization (Fazel 2002)).

- Applications in collaborative filtering, computer vision, remote sensing, load forecasting, electricity market inference.
Low-rank Matrix Completion

Low-rank Matrix Completion Problem!
Low-rank Matrix Completion

Low-rank Matrix Completion Problem!

- A literature includes theoretical analysis (e.g., Candes, Recht 2012) and recovery methods (e.g., nuclear norm minimization (Fazel 2002)).
  \[
  \min_X \|X\|_* = \text{sum of singular values of } X
  \]
  s.t. \( X \) is consistent with the observed entries,

- Applications in collaborative filtering, computer vision, remote sensing, load forecasting, electricity market inference
Low-rank Matrix Completion for PMU Data Recovery

Advantages:

- No modeling of the power system.
- Analytical performance guarantee.
- Tolerate a significant percentage of missing data/bad measurements at random locations.
Low-rank Matrix Completion for PMU Data Recovery

Advantages:

- No modeling of the power system.
- Analytical performance guarantee.
- Tolerate a significant percentage of missing data/bad measurements at random locations.

Limitations:

- Do not model temporal dynamics sufficiently.
- Low-rank matrix completion methods fail to recover a column/row if the complete column/row is lost. Simultaneous and consecutive data losses are frequent in PMU data.
- Convex optimization problems are computationally expensive for large datasets.
Our Contribution

Our developed model-free data recovery and error correction methods

- First-order algorithms to solve nonconvex optimization problems with provable global optimality.
- Recover/correct simultaneous and consecutive data losses/errors.
- Differentiate bad data from system events.

Low-rank Hankel Structure of PMU Data

Observation matrix:

\[ \mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n] \in \mathbb{C}^{n_c \times n} \]

Hankel structure:

\[ \mathcal{H}_\kappa(\mathbf{Y}) = \begin{bmatrix}
\mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_{n-\kappa+1} \\
\mathbf{y}_2 & \mathbf{y}_3 & \cdots & \mathbf{y}_{n-\kappa+2} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{y}_\kappa & \mathbf{y}_{\kappa+1} & \cdots & \mathbf{y}_n
\end{bmatrix} \]

\[ \mathcal{H}_\kappa(\mathbf{Y}) \in \mathbb{C}^{\kappa n_c \times (n-\kappa+1)} \]
can still be approximated by a low-rank matrix.

The low-rank Hankel property results from the reduced-order dynamical system.
Low-rank Hankel Structure of PMU Data

**Figure:** Measurements that contain a disturbance

**Figure:** The low-rank approximation errors to $\mathcal{H}_κ(\mathbf{Y})$

**Figure:** The low-rank approximation errors to $\mathcal{H}_κ(\mathbf{Y})$, where $\mathbf{Y}$ is a column permutation of $\mathbf{Y}$. 
Robust Data Recovery

Let $\mathbf{M} = \mathbf{Y} + \mathbf{S}$ denote the partially corrupted measurements, where $\mathbf{S}$ denotes the sparse errors. The robust data recovery problem is formulated as

$$
\min_{\mathbf{X}, \mathbf{S} \in \mathbb{C}^{nc \times n}} \| \mathcal{P}_\Omega (\mathbf{X} + \mathbf{S} - \mathbf{M}) \|_F^2
$$

subject to $\text{rank}(\mathcal{H}_\kappa (\mathbf{X})) = r$, $\| \mathbf{S} \|_0 \leq s$.

(1)
Our proposed alternating projection algorithm

Initialization: $X_0 = 0$, thresholding $\varepsilon_0$;

Two stages of iterations:

- In the $k$-th outer iteration:
  - Increase the desired rank $k$ from 1 to $r$ gradually;

- In the $l$-th inner iteration:
  - Update $S_l$ based on the current estimated thresholding $\xi_l$;
  - Update $X_l$ along the gradient descent direction $\mathcal{P}_\Omega (X_l + S_l - M)$;
  - Project the Hankel matrix $H_k X_l$ into the rank-$k$ matrix set;
  - Obtain $X_{l+1}$ from the matrix after projection;
  - Update $\xi_{l+1}$ based on $X_{l+1}$. 
Theoretical results

Theorem

Suppose the number of observed data exceeds $O(r^3 \log^2(n))$ and each row of $S$ has at most $O\left(\frac{1}{r}\right)$ fraction of nonzeros, the algorithm converges to the original data matrix linearly as

$$\|X_l - Y\|_F \leq \varepsilon \quad \text{after} \quad l = O\left(\log\left(\frac{1}{\varepsilon}\right)\right) \text{ iterations.}$$

- Required number of observations: $O(r^3 \log^2(n))$, less than the bound $O(nr \log^2(n))$ of recovery with convex relaxation approach;
- Fraction of corruptions it can correct: $O\left(\frac{1}{r}\right)$ in each row;
- Low computational complexity per iteration: $O(rn_c n \log n)$;
- Recovery guarantees on simultaneous data losses and corruptions across all channels.
Numerical experiments

**Figure:** One case of 8% random bad data and 40% random missing data

**Figure:** Consecutive bad data, 3% random bad data and 20% missing data
Outline

1. Motivation
2. Data Recovery and Error Correction
3. Event Identification
4. Conclusions
Existing Approaches

Recent development of data driven approaches of event identification, see sample work [Wang W et al., 2014, Rafferty et al., 2016, Valtierra R et al., 2014, Jiang H et al., 2014]:

- Advantage: model-free, robust to model errors.
- Limitations
  - **Single events** or multiple events with long time intervals and minor overlapping
  - A large number of training datasets
  - No clear physical interpretations
  - High training complexity.
Challenges and Goals

• Challenges
  • Events happen close in time or location have overlapping impacts on the measurements.
  • **Not enough training datasets** for all possible conditions and successive events.
  • The **rapid occurrence** of cascading failures requires online algorithms.

• Goals
  • Train on the **a small number of single events**
  • Identify **overlapping successive** events in real time
Our Approach: Identification using Subspace Representation

- Identify an event by comparing the row subspace of the real-time PMU data matrix with a dictionary of subspaces obtained from recorded PMU data.

\[
\begin{align*}
M_1 &= \ldots \\
M_{n1} &= \ldots \\
M_1 &= \ldots \\
M_{n2} &= \ldots \\
M_1 &= \ldots \\
M_{n3} &= \ldots \\
M &= \\
\end{align*}
\]

**Figure:** Dictionary construction from historical datasets and real-time data identification through subspace comparison


Offline Training

- **Extract** dominant eigenvalues $\lambda$ of the state matrix and the dominant singular values $\delta$ of the data matrix of recorded single events as features;
- Input $\lambda$ and $\sigma$ to a 2-layer convolutional neural network (CNN) and **combine two paths** in the fully connected layer;
- Train this **2-layer CNN** classifier to identify the type of an event.

*Figure: Offline Training*
Online Testing

Given the two successive events occurring at $T_1$ and $T_2$ respectively, there are **three steps** to identify the type of second event:

- **Step 1:** predict the impact of the first event after time $T_2$ and subtract it from the obtained measurements

**Figure:** Online Testing
Online Testing

Given the two successive events occurring at $T_1$ and $T_2$ respectively, there are three steps to identify the type of second event:

- **Step 1:** predict the impact of the first event after time $T_2$ and subtract it from the obtained measurements.
- **Step 2:** various methods can be employed to predict the measurements, such as time series analysis and Hankel matrix-based methods [Hao et al., 2018].

Figure: Online Testing
Online Testing

Given the two successive events occurring at $T_1$ and $T_2$ respectively, there are three steps to identify the type of second event:

- **Step 2**: The features $(\lambda, \sigma)$ are extracted from the residual measurements;

**Figure**: Online Testing
Online Testing

Given the two successive events occurring at $T_1$ and $T_2$ respectively, there are three steps to identify the type of second event:

1. **Prediction of the impacts**
2. **Subtraction**
3. **Classification**

**Prediction-Subtraction Process**
- Compute eigenvalues and singular values
- The trained CNN classifier

**Classification**
- Classify the dominant features

**Step 2:** The features $(\lambda, \sigma)$ are extracted from the residual measurements;

**Step 3:** The trained CNN outputs the type of the second event in real time.

**Figure:** Online Testing
Numerical Results

Table: Total Number of the Testing Datasets

<table>
<thead>
<tr>
<th>Types</th>
<th>LT+GT</th>
<th>LT+TP</th>
<th>LT+LT</th>
<th>GT+GT</th>
<th>GT+TP</th>
<th>TP+GT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>169</td>
<td>404</td>
<td>378</td>
<td>107</td>
<td>58</td>
<td>146</td>
</tr>
</tbody>
</table>

- Line trips (LTs), generator trips (GTs), and three-phase short circuit (TPs) in the 68-bus power system generated using PSS/E;
- The training set includes 967 single events at different locations with different topologies and various pre-event conditions;
- The testing set includes 1262 two-event cases, where “LT+GT” means a line trip event is followed by a generator trip event.
- The time interval between any two successive events varies from 0.5 to 2 seconds.
Performance with a Small Training Dataset

- CNN-F achieves a higher IAR when given a small training data;
- The IAR of **CNN-T** is sensitive to the size of the training data.
The Impact of Subtracting the First Event

Table: IAR of CNN-F and CNN-T with and without subtracting the first event ($\Delta T = 1$ second)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Process</th>
<th>LT %</th>
<th>GT %</th>
<th>TP %</th>
<th>Overall %</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN-F</td>
<td>NS</td>
<td>79.4</td>
<td>68.3</td>
<td>96.4</td>
<td>81.9</td>
</tr>
<tr>
<td>CNN-F</td>
<td>SP</td>
<td>95.9</td>
<td>89.1</td>
<td>97.3</td>
<td>94.2</td>
</tr>
<tr>
<td>CNN-T</td>
<td>NS</td>
<td>94.8</td>
<td>83.2</td>
<td>78.4</td>
<td>85.1</td>
</tr>
<tr>
<td>CNN-T</td>
<td>SP</td>
<td>73.2</td>
<td>62.4</td>
<td>78.4</td>
<td>71.5</td>
</tr>
</tbody>
</table>

- “Not Subtract (NS)” means using the measurements after the second event directly;
- “Subtract the Prediction (SP)” means using the residual after subtracting the first event;
- Subtracting the impact of the first event can enhance CNN-F’s performance significantly.
Robustness to Noise

Table: IAR of identifying the second event by CNN-F with different signal-noise-ratio (SNR) noisy measurements

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>IAR of LT (%)</td>
<td>77.3</td>
<td>81.4</td>
<td>82.5</td>
<td>85.6</td>
<td>82.5</td>
<td>90.7</td>
<td>91.7</td>
</tr>
<tr>
<td>IAR of GT (%)</td>
<td>75.2</td>
<td>82.2</td>
<td>86.1</td>
<td>86.1</td>
<td>85.1</td>
<td>86.1</td>
<td>82.2</td>
</tr>
<tr>
<td>IAR of TP (%)</td>
<td>93.7</td>
<td>97.3</td>
<td>97.3</td>
<td>98.2</td>
<td>98.2</td>
<td>99.1</td>
<td>99.1</td>
</tr>
<tr>
<td>Overall IAR (%)</td>
<td>82.5</td>
<td>87.4</td>
<td>88.9</td>
<td>90.3</td>
<td>90.0</td>
<td>92.2</td>
<td>91.3</td>
</tr>
</tbody>
</table>

- The overall IAR of CNN-F is more than 80% for different noise levels;
- IAR is more than 90% when the SNR is higher than 70 dB;
- Significant events like TP events are less sensitive to noise.
Conclusions

• A framework of power system data analytics by exploiting the low-dimensional structure of spatial-temporal data blocks.

• Data quality improvement with analytical guarantees. (Missing data recovery, detection of cyber data attacks.)

• Real-time event identification approach using a small number of recorded single events for training.
Q & A
References

- Candes, Emmanuel, Benjamin Recht (2012)
  Exact Matrix Completion via Convex Optimization
  *Communications of the ACM*, 111 – 119.

- Fazel M (2002)
  Matrix Rank Minimization with Applications
  *PhD thesis, Stanford University*

  Load Curve Data Cleansing and Imputation via Sparsity and Low Rank
  *IEEE Transactions on Smart Grid*, 2347 – 2355.

  Electricity Market Forecasting via Low-rank Multi-kernel Learning

  False Data Injection Attacks Against State Estimation in Electric Power Grids
Xie L, Mo Y, Sinopoli B (2010)
False Data Injection Attacks in Electricity Markets
*Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, 226 – 231.

Malicious Data Attacks on Smart Grid State Estimation: Attack Strategies and Countermeasures
*Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, 220 – 225.

Sedghi H, Jonckheere E (2013)


1-Bit Matrix Completion *Information and Inference*, 189 – 223.
Pengzhi Gao, Ren Wang, Meng Wang, and Joe H. Chow (2016)
Low-rank Matrix Recovery from Quantized and Erroneous Measurements: Accuracy-preserved Data Privatization in Power Grids
_Asilomar Conference on Signals, Systems and Computers_, 374 – 378.

Pengzhi Gao, Meng Wang, Scott G. Ghiocel, Joe H. Chow, Bruce Fardanesh, and George Stefopoulos (2016)
Missing Data Recovery by Exploiting Low-dimensionality in Power System Synchrophasor Measurements

Pengzhi Gao, Meng Wang, Joe H. Chow, Scott G. Ghiocel, Bruce Fardanesh, George Stefopoulos, and Michael P. Razanousky (2016)
Identification of Successive “Unobservable” Cyber Data Attacks in Power Systems
Le Xie, Yang Chen, and P. R. Kumar (2014)
Dimensionality Reduction of Synchrophasor Data for Early Event Detection: Linearized Analysis

Mark A. Davenport, Yaniv Plan, Ewout van den Berg, and Mary Wootters (2014)
1-bit Matrix Completion
*Information and Inference* 3(3), 189 – 223.

Sonia A. Bhaskar (2016)
Probabilistic Low-Rank Matrix Completion from Quantized Measurements
*Journal of Machine Learning Research* 17(60), 1 – 34.

Akshay Soni, Swayambhoo Jain, Jarvis Haupt, and Stefano Gonella (2016)
Noisy Matrix Completion under Sparse Factor Models
Tony Cai, and Wen-Xin Zhou (2013)  
A Max-Norm Constrained Minimization Approach to 1-Bit Matrix Completion  

Olga Klopp, Jean Lafond, Eric Moulines, and Joseph Salmon (2015)  
Adaptive Multinominal Matrix Completion  

Andrew S. Lan, Christoph Studer, and Richard G. Baraniuk (2014)  
Matrix Recovery from Quantized and Corrupted Measurements  

Pengzhi Gao, Meng Wang, Joe H. Chow, Scott G. Ghiocel, Bruce Fardanesh, George Stefopoulos, and Michael P. Razanousky (2016)  
H. Farhangi (2010)
The Path of the Smart Grid

Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities
*IEEE Transactions on Smart Grid*, 6(2), 911 – 918.

Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data
*IEEE Transactions on Smart Grid*, 7(1), 136 – 144.

G. W. Hart (1992)
Nonintrusive appliance load monitoring
*Proceedings of the IEEE*, 80(12), 1870 – 1891.
Inferring Personal Information from Demand-Response Systems

Li Fengjun, Luo Bo, and Liu Peng (2011)
Secure and Privacy-Preserving Information Aggregation for Smart Grids

Barbosa Pedro, Brito Andrey, Almeida Hyggo, and Clau Sebastian (2014)
Lightweight Privacy for Smart Metering Data by Adding Noise

McLaughlin Stephen, McDaniel Patrick, and Aiello William (2011)
Protecting Consumer Privacy from Electric Load Monitoring

Variability and Trend-Based Generalized Rule Induction Model to NTL Detection in Power Companies


Saeed Aghabozorgi, Seyed Shirkhorshidi Ali, and Ying Wah Teh (2015)
Piao, Minghao and et al (2014)  
Subspace Projection Method Based Clustering Analysis in Load Profiling  

Chen Yi, Nasrabadi Nasser M, and Tran Trac D (2011)  
Hyperspectral Image Classification Using Dictionary-Based Sparse Representation  

Missing Data Recovery for High-dimensional Signals with Nonlinear Low-dimensional Structures  
*IEEE Transactions on Signal Processing*, 65(20), 5421 – 5436.

Sonia A. Bhaskar (2016)  
Probabilistic Low-Rank Matrix Completion from Quantized Measurements  
*Journal of Machine Learning Research* 17(60), 1 – 34.
Ng Andrew, Jordan Michael, and Weiss Yair (2002)
On Spectral Clustering: Analysis and an Algorithm
Advances in neural information processing systems, 2, 849 – 856.

Elhamifar Ehsan, and Vidal Rene (2013)
Sparse Subspace Clustering: Algorithm, Theory, and Applications
IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(11), 2765 – 2781.

Liu Guangcan, Lin Zhouchen, Yan Shuicheng, Sun Ju, Yu Yong, and Ma Yi (2013)
Robust Recovery of Subspace Structures by Low-Rank Representation

Rahmani Mostafa, and George K. Atia (2017)
Innovation Pursuit: A New Approach to Subspace Clustering
Tseng Paul (2000)
Nearest Q-Flat to m Points

Pengzhi Gao, Ren Wang, Meng Wang, and Joe H. Chow (2018)
Low-rank Matrix Recovery from Noisy, Quantized and Erroneous Measurements
*IEEE Transactions on Signal Processing*, 60(11), 2918 – 2932.

Andrew S. Lan, Christoph Studer, and Richard G. Baraniuk (2014)
Matrix Recovery from Quantized and Corrupted Measurements

Mark A. Davenport, Yaniv Plan, Ewout van den Berg, and Mary Wootters (2014)
1-bit Matrix Completion
*Information and Inference* 3(3), 189 – 223.
Wenting Li and Meng Wang and Joe H. Chow (2018)
Real-time Event Identification through Low-dimensional Subspace Characterization of High-dimensional Synchrophasor Data
*IEEE Transactions on Power Systems*, 4937-4947.

Wei Wang, Li He, Penn Markham, Hairong Qi, Yilu Liu, Qing Charles Cao, and Leon M Tolbert (2014)
Multiple event detection and recognition through sparse unmixing for high-resolution situational awareness in power grid

Robust subspace clustering
Valtierra Rodriguez, Martin and de Jesus Romero-Troncoso, Rene and Osornio-Rios, Roque Alfredo and Garcia-Perez, Arturo (2015)  
Detection and classification of single and combined power quality disturbances using neural networks  

Huaiguang Jiang, Jun Jason Zhang, Wenzhong Gao, and Ziping Wu. (2014)  
Fault detection, identification, and location in smart grid based on data-driven computational methods  

Rafferty, Mark and Liu, Xueqin and Laverty, David M and McLoone, Sean. (2016)  
Real-time multiple event detection and classification using moving window PCA  
Model-less Data Quality Improvement of Streaming Synchrophasor Measurements by Exploiting the Low-Rank Hankel Structure

Elsner, James B. (2002)
Analysis of time series structure: SSA and related techniques
*Taylor & Francis, 2002.*