High-dimensional Data Analysis in Power System Monitoring

Meng Wang
Rensselaer Polytechnic Institute

CURENT Course
October 8, 2014
Big Data Issues

- Large in quantity or complex in structure.
- The curse of dimensionality. Challenges in acquisition, storage, and processing.
- Applications: Medical imaging, genomic data analysis, social network analysis, power system monitoring, Internet monitoring, etc.
SCADA (supervisory control and data acquisition) System provides power measurements every five seconds.

PMUs (Phasor Measurement Units) can provide synchronized phasor measurements of the power system at a sampling rate of 30 samples per second or more.
After the DOE smart grid investment program starting in 2009, the installation of PMUs in the North American power system will be close to 2,000 PMUs soon.

Multi-channel PMUs can measure bus voltage phasors, line current phasors, and frequency.
Applications:

- State estimation,
- Oscillation detection and electromechanical mode identification,
- Stability analysis and post-event analysis,
- Disturbance detection and location, dynamic security assessment,

Task specific methods, developed when the coverage of PMUs on a power network was quite sparse.

Mostly exploit the spatial and the temporal correlations in the measurements separately. Lack a common framework.
As the coverage of PMUs become denser, one can analyze PMU data collectively from PMUs located in electrically close regions and distinct control regions.

Processing *spatial-temporal blocks* of PMU data for tasks such as missing data recovery, data compression and storage, disturbance triggering, detection of cyber data attacks.

How to process the high-dimensional PMU data blocks?
Outline

1. Introduction
2. High-dimensional Data Analysis Using Low-dimensional Models
3. Missing PMU Data Recovery
4. Detection of Cyber Data attacks
Outline

1. Introduction

2. High-dimensional Data Analysis Using Low-dimensional Models

3. Missing PMU Data Recovery

4. Detection of Cyber Data attacks
Exploiting Low-dimensional Structures in High-dimensional Datasets

- It is relatively easy to acquire, store and process high-dimensional datasets if they have *low-dimensional* structures.
- Examples: Sparse signals, low rank matrices.
- Wide existence of low-dimensional structures: images, rating matrices, network measurements, ...
- Recent arts: compressed sensing theory, low rank matrix theory.
Wide Existence of Low-dimensional Structures

Sparse signals

- a vector in $\mathbb{R}^n$
- $k$ nonzero coefficients under certain basis. ($k << n$)
- Imaging, Communication, Biology

Low-rank matrices

- rank $<<$ dimension.
- Collaborative filtering (Netflix problem)
- System identification
- Internet traffic analysis

http://www.inkfarm.com/Image-File-Extensions
http://www.netflix.com
Compressed Sensing for Sparse Signals

Unknown sparse signal $x$, observation $y$, measurement matrix $A$.

$$y = Ax, \ m << n.$$  

Given $A$, sparse $x$ can be recovered from $y$!

- Significant reduction in the required number of measurements. $k$-sparse $n$-dimensional signals, $m = O(k \log \frac{n}{k})$.
- Simple compression methods. $A$ could be random matrices. Wide existence of good $A$'s.
- Efficient recovery methods. e.g., $\ell_1$ minimization.
Compressed Sensing for Sparse Signals

\[ y = A x \]

\[ m = n \]

\[ \ell_0\text{-minimization} \]

\[ \min_z \|z\|_0 \quad \text{s.t.} \quad Az = y, \]

where \( \|z\|_0 \) = number of non-zero entries of \( z \).

- bad news: computationally hard.
Compressed Sensing for Sparse Signals

\[ \begin{align*}
  y & = A x \\
  m & = n
\end{align*} \]

\[ \begin{align*}
  \ell_1\text{-minimization} \\
  \min_{z} \|z\|_1 = \sum_{i} |z_i| \quad \text{s.t.} \quad Az = y,
\end{align*} \]

where \( \|z\|_1 = \sum_{i} |z_i| \).

The sparse signal \( x \) is the solution to \( \ell_1 \) problem!

- computationally efficient.
- theoretical guarantee. (Candès and Tao 2006, Donoho 2006)
Low-rank Matrix Completion

Netflix Problem

Low rank matrix completion problem.
How shall we recover the missing entries of a low-rank matrix?

\[
\min_X \text{Rank}(X) \\
\text{s.t. } X_{ij} = M_{ij}, \forall i \in \Omega.
\]

\(\Omega\): locations of the observed entries.

- NP hard.
Nuclear norm minimization (Fazel 2002, Candés & Recht 2009), recover the missing data by solving a convex program.

\[
\min_{X} \|X\|_* = \text{sum of the singular values of } X \\
\text{s.t. } X_{ij} = M_{ij}, \forall i \in \Omega.
\]

- Convex relaxation of Rank minimization problem.
- Computationally efficient. Can be solved by Semidefinite Programming.
- Theoretical guarantee. (Candés & Recht 2009)
Low-rank Matrix Completion

Theorem (Candés & Recht 09, Gross 11, Recht 11)

All entries of a rank-$r$ matrix $L \in \mathbb{C}^{n_1 \times n_2}$ can be correctly recovered, as long as $O(rn \log^2 n)(n = \max(n_1, n_2))$ randomly selected entries of $L$ are observed.

Significant saving in the number of observations when $r$ is small.

Theorem (Candés & Tao 09)

If each entry is observed with prob. $p = \frac{m}{n_1 n_2}$, no method whatsoever can succeed with

$$m \leq Cnr \log n,$$

for some fixed constant $C$.

Near-optimal recovery via Nuclear norm minimization.
Connection to Compressed Sensing

General setup

Rank minimization

\[
\min_X \text{Rank}(X) \quad \text{s.t.} \quad \langle A^i, X \rangle = \text{Trace}\left((A^i)^T X\right) = b^i, \quad \forall i = 1, \ldots, m
\]

Convex relaxation.

\[
\min_X \|X\|_* \quad \text{s.t.} \quad \langle A^i, X \rangle = \text{Trace}\left((A^i)^T X\right) = b^i, \quad \forall i = 1, \ldots, m
\]

Consider the special case \( X = \text{diag}(x), x \in \mathbb{R}^n \), then

\[
\text{Rank}(X) = \|x\|_0, \\
\|X\|_* = \|x\|_1.
\]

Reduce to compressed sensing!
**Principle Component Analysis: Classical Method**

- widely used for dimension reduction.
- convert observations to a set of values of linearly uncorrelated variables, called *principle components*.

Stack all the data points as column vectors of a matrix $M$, PCA seeks the best rank-$k$ approximation by solving

\[
\min_L \|M - L\|_2 \\
\text{s.t. } \text{Rank}(L) \leq k
\]

Not robust to even a few outliers in $M$. 
Robust PCA

- The observation matrix is the sum of a low-rank matrix $L_0$ (actual data) and a sparse matrix (corruptions).

$$M = L_0 + S_0$$

- Decomposition of a low-rank matrix and a sparse matrix.
- Applications in computer vision, image processing, network traffic analysis, etc.
(Hard) Optimization problem

\[
\min_{L,C} \text{Rank}(L) + \lambda \|C\|_0
\]

s.t. \(L + C = M\)

- \(\|C\|_0\): the number of nonzero entries in \(C\).
- not tractable.

Convex relaxation

\[
\min_{L,C} \|L\|_* + \lambda \|C\|_1
\]

s.t. \(L + C = M\)

- \(\|C\|_1 = \sum_{ij} |C_{ij}|\)
- Theoretical guarantee: Candés & Li & Ma & Wright 11, Chandrasekaran & Sanghavi & Parrilo & Willsky 11, Recht & Fazel & Parrilo 10.
Does low-dimensionality exist in PMU measurements?
Low-rank Property of PMU data

PMUs in Central NY Power Current magnitudes of Systems

- 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 observed in Chen & Xie & Kumar 2013, Dahal & King & Madani 2012)
How could we use the low dimensionality in PMU data management for power system monitoring?
Publications

- **Missing data recovery**

- **Cyber data attacks**
  Wang, Gao, Ghiocel, Chow, Fardanesh, Stefopoulos, Razanousky, IEEE SmartGridComm 2014

- **Low rank approach to data analysis**
  Wang, Chow, Gao, Jiang, Xia, Ghiocel, Fardanesh, Stefopoulos, Kokai, Saito, Razanousky, HICSS 2015
Outline

1. Introduction

2. High-dimensional Data Analysis Using Low-dimensional Models

3. Missing PMU Data Recovery

4. Detection of Cyber Data attacks
We observe partial entries of the complex matrix (rectangular form of voltage and current phasors). How shall we recover the missing points for offline applications like system identification and past event analysis?

<table>
<thead>
<tr>
<th>time</th>
<th>channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

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

- Wide applications in collaborative filtering, computer vision, machine learning, remote sensing, and system identification.
- Problem formulation: given part of the entries of a matrix, recover the remaining entries.
- the rank of the matrix is much less than its dimension.
Low-rank Matrix Completion

Nuclear norm minimization

$$\min_{X} \|X\|_* = \text{sum of singular values of } X$$

s.t. $X$ is consistent with the observed entries,

Quite a few recovery algorithms exist, e.g., singular value thresholding (SVT) (Cai et al. 2010), information cascading matrix completion (ICMC) (Meka et al. 2009).
Theorem (Candés & Recht 09, Gross 11, Recht 11)

All entries of a rank-$r$ matrix $L \in \mathbb{C}^{n_1 \times n_2}$ can be corrected recovered, as long as $O(rn\log^2 n)(n = \max(n_1, n_2))$ randomly selected entries of $L$ are observed.

- Significant saving in the number of observations when $r$ is small.
- Existing analysis assumes that the locations of missing points are selected randomly.
Our Results

- The locations of missing PMU data are usually correlated.
  - temporal correlation: loss of consecutive measurements in one PMU channel.
  - channel correlation: loss of measurements in multiple PMU channels simultaneously.

Theorem (Gao & Wang & Ghiocel & Chow 14)

*Although the locations of the missing entries of a rank-$r$ matrix are temporally or spatially correlated, all missing entries can be correctly recovered as long as $O(n^{2-\frac{1}{r+1}} r^{\frac{1}{r+1}} \log^{\frac{1}{r+1}} n)$ entries are observed.*

- The first theoretical guarantee of low-rank matrix completion when the locations of missing entries are correlated.
Simulation

- 6 PMUs measuring 37 bus voltage phasors and line current phasors.
- 30 samples per second.

Current magnitudes in four datasets.
Simulation Results

- Recover missing data points that are temporally correlated.
  - the first 5-second PMU data of dataset #1.

Relative recovery error of the SVT algorithm

Relative recovery error of the ICMC algorithm
Simulation Results

- Recover missing data points that are spatially correlated.
  - the first second PMU data of dataset #1.

![Relative recovery error of the SVT algorithm](image1)

![Relative recovery error of the ICMC algorithm](image2)

Relative recovery error of the SVT algorithm

Relative recovery error of the ICMC algorithm
We also develop an online missing data recovery algorithm that could fill in the missing data instantaneously and detect system disturbance.

**Table:** Relative recovery error of OLAP on four datasets

<table>
<thead>
<tr>
<th>$p_{avg}$</th>
<th>Dataset #1</th>
<th>Dataset #2</th>
<th>Dataset #3</th>
<th>Dataset #4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.0082</td>
<td>0.0026</td>
<td>0.0006</td>
<td>0.0186</td>
</tr>
<tr>
<td>0.10</td>
<td>0.0089</td>
<td>0.0039</td>
<td>0.0009</td>
<td>0.0235</td>
</tr>
<tr>
<td>0.15</td>
<td>0.0119</td>
<td>0.0053</td>
<td>0.0011</td>
<td>0.0442</td>
</tr>
<tr>
<td>0.20</td>
<td>0.0145</td>
<td>0.0060</td>
<td>0.0013</td>
<td>0.0494</td>
</tr>
<tr>
<td>0.25</td>
<td>0.0153</td>
<td>0.0066</td>
<td>0.0014</td>
<td>0.0726</td>
</tr>
<tr>
<td>0.30</td>
<td>0.0191</td>
<td>0.0075</td>
<td>0.0017</td>
<td>0.0779</td>
</tr>
<tr>
<td>0.35</td>
<td>0.0204</td>
<td>0.0082</td>
<td>0.0018</td>
<td>0.1137</td>
</tr>
<tr>
<td>0.40</td>
<td>0.0227</td>
<td>0.0091</td>
<td>0.0019</td>
<td>0.1189</td>
</tr>
</tbody>
</table>

Missing data recovery by OLAP algorithm
1 Introduction

2 High-dimensional Data Analysis Using Low-dimensional Models

3 Missing PMU Data Recovery

4 Detection of Cyber Data attacks
Example of Cyber Data Attacks

Measurements $V_1$, $V_2$, $I_{12}$, and $I_{13}$. Estimate $V_3$.

Redundancy in measurements can be used to detect bad data.

Cyber data attack: manipulate $I_{12}$ and $I_{13}$ simultaneously.

Can result in significant error in $V_3$.

The removal of attacked measurements will make $V_3$ unobservable.
Cyber Data Attacks

- The worst interacting bad data. (Liu & Ning & Reiter 11).
- An intruder with the system topology information can simultaneously manipulate multiple measurements so that these attacks cannot be detected by any bad data detector.
- Cyber data attacks can potentially lead to significant errors to the outcome of state estimation.
- Also termed as “unobservable attacks”, because the removal of affected measurements would make the system unobservable. (Kosut & Jia & Thomas & Tong 10).
Existing Approaches

Cyber attacks in SCADA system. A protection perspective.

- Usually protect key PMUs to avoid these attacks. (Kosut & Jia & Thomas & Tong 10, Kim & Poor 11, Bobba et al. 10, Dán & Sandberg 10)

- Sedghi & Jonckheere 13: Detection of cyber data attacks in SCADA system. Assume the measurements at different time instants are i.i.d. distributed.
Our Contributions

A new detection method of cyber data attacks in PMU measurements.

- no other assumptions expect for the low-rankness of the PMU data.
- The detection method can identify the attacks even when the system is under disturbance.
- The PMU channels under attack can be identified by solving a convex optimization problem.
System Model

- \( \pi \) model of a transmission line.

- Current \( I_{ij} \) from bus \( i \) to bus \( j \) is related to bus voltage \( V^i \) and \( V^j \) by

\[
I_{ij} = \frac{V^i - V^j}{Z_{ij}} + V^i \frac{Y_{ij}}{2}.
\]

- Voltage and current phasor measurements can be represented by linear functions of state variables.
The intruder can only attack a small number of PMUs continuously.
The intruder injects an unobservable attack at each time instant.
An unobservable attack is a linear combination of errors in state variables.
Mathematical Formulation

- $t$: total number of sampling time instants.
- $p$: total number of PMU channels.
- $n$: total number of buses. $n < p$
- $\bar{L} \in \mathbb{C}^{t \times p}$: the actual voltage and current phasors in $t$ instants.
- $\bar{D} \in \mathbb{C}^{t \times n}$: the additive error in state variables.
- $N \in \mathbb{C}^{t \times p}$: the noise.
- $M \in \mathbb{C}^{t \times p}$: the obtained measurements that are under attack.
- $W \in \mathbb{C}^{t \times p}$: relates state variables with phasors.

$$M = \bar{L} + \bar{C}W^T + N$$
Measurements under attack

- $\bar{L}$: low-rank. From correlations in measurements.
- $\bar{C}$: column sparse. The intruder has limited access to the system.
- $N$: $\|N\|_F \leq \varepsilon$.

Given $M$ and $W$, how could we identify $\bar{L}$ and $\bar{C}$?

Decomposition of a low-rank matrix and a transformed column-sparse matrix.
Matrix Decomposition

- Decomposition of a low-rank matrix and a sparse matrix.
- Through convex optimization.

\[
\min_{L \in \mathbb{C}^{t \times p}, C \in \mathbb{C}^{t \times n}} \|L\|_* + \lambda \sum_{i,j} |C_{i,j}| \quad \text{s.t.} \quad \|L + C - M\|_F \leq \varepsilon \tag{1}
\]

- Theoretical guarantee:
  Candés & Li & Ma & Wright 11, Chandrasekaran & Sanghavi & Parrilo & Willsky 11, Recht & Fazel & Parrilo 10.

- Applications: Internet traffic analysis, image processing, medical imaging, etc.
Connection to Related Work

- Xu & Caramanis & Sanghavi 12: Decomposition of a low-rank matrix and a column-sparse matrix.

\[
\text{Measurements} = \text{Low Rank} + \text{Column Sparse} + \text{Noise}
\]

Our methods and proofs are built upon those in Xu & Caramanis & Sanghavi 12. Extension to general cases.

\[
\text{Measurements} = \text{Actual phasors} + \text{Errors in bus voltages} \times \text{column sparse} + \text{Noise}
\]
Mardani & Mateos & Giannakis 13: Decomposition of a low-rank matrix plus a compressed sparse matrix. Internet traffic anomaly detection.

Our focus: column-sparse matrices, $W$ is arbitrary.
Our Approach

- Find \((L^*, C^*)\), the optimum solution to the following optimization problem

\[
\min_{L\in\mathbb{C}^{t\times p}, C\in\mathbb{C}^{t\times n}} \|L\|_* + \lambda \|C\|_{1,2} \quad \text{s.t.} \quad \|L + CW^T - M\|_F \leq \epsilon \quad (2)
\]

- Compute the SVD of \(L^* = U^*\Sigma^*V^*\).
- Find column support of \(D^* = C^*W^T\), denoted by \(\mathcal{I}^*\).
- Return \(L^*_{\mathcal{I}^*c}, U^*\) and \(\mathcal{I}^*\).

(2) is convex and can be solved efficiently.
Theoretical Guarantee

Theorem (Noiseless measurements, $N = 0$)

With a properly chosen $\lambda$, the solution returned by our method

1. identifies the PMU channels under attack.
2. identifies the measurements that are not attacked.
3. recovers the correct subspace spanned by actual phasors.

Theorem (Noisy measurements, $N \neq 0$)

With a properly chosen $\lambda$, the solution returned by our method is sufficiently close (with distance depending on the noise level) to a solution such that 1-3 hold.
Numerical Results

- Simulate the case that the intruder alters the PMU channels that measure $I_{12}^1$, $I_{52}^1$, $I_{13}^1$ and $I_{43}^1$.
- The voltage phasor estimates of Buses 2 and 3 are corrupted.
Conclusion

- Leverage the low-dimensional structures in high-dimensional data to address the challenges in data acquisition, storage, and information extraction.
- Connection of low-rank methods with power system monitoring. Analysis of spatial-temporal blocks of PMU measurements.
- Missing data recovery: theoretical guarantee of successful recovery when the locations of the missing points are correlated.
- Detection of cyber data attacks: decomposition of a low-rank and transferred column-sparse matrix.