Missing Data Recovery for High-dimensional Signals with Nonlinear Low-dimensional Structures

Pengzhi Gao

Department of Electrical, Computer & Systems Engineering

Rensselaer Polytechnic Institute

GlobalSIP 2015, Orlando FL

December 15, 2015

Acknowledgement

- Prof. Meng Wang (My advisor), Ms. Genevieve de Mijolla, Dr. Scott G. Ghiocel, Prof. Joe H. Chow. (**Rensselaer Polytechnic Institute**)
- Dr. Bruce Fardanesh, Dr. George Stefopoulos. (**New York Power Authority**)
- Dr. Matthew Berger, Dr. Lee M. Seversky. (**Air Force Research Laboratory**)

NSF

CURENT
CENTER FOR ULTRA-WIDE-AREA RESILIENT ELECTRIC ENERGY TRANSMISSION NETWORKS

nyserda
Energy. Innovation. Solutions.

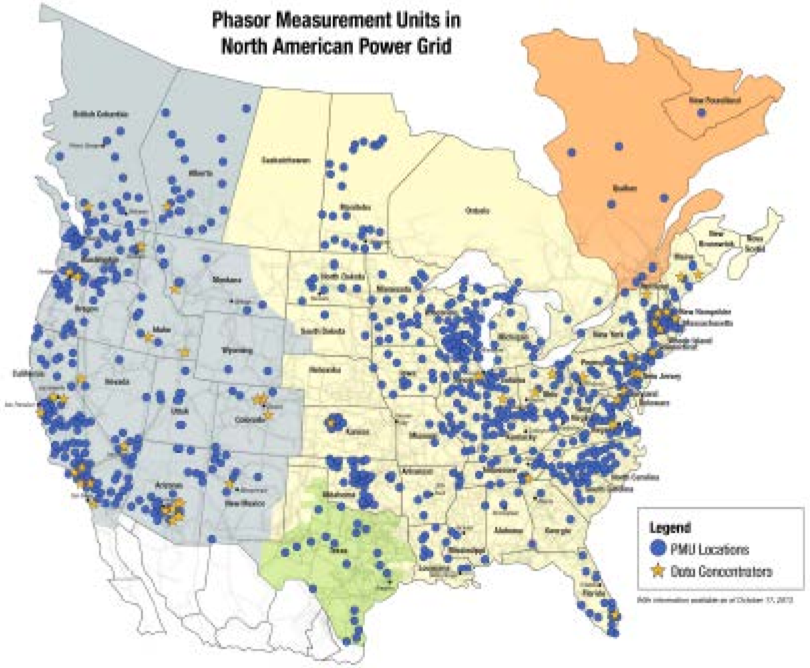
Outline

- Background and challenges of PMU data processing
- Low-dimensional structures of PMU data
- Our new model of low-dimensional structures
- Missing data recovery: theoretical guarantee and numerical results

Phasor Measurement Units

- Multi-channel PMUs can measure bus voltage phasors, line current phasors, and frequency.
- 1500+ PMUs in the North America. Will increase to 2,000 soon.

Example of PMU
http://www.macrodyneusa.com/model_1690.html



Current Installation of PMUs
https://www.naspi.org/documents

Big Data in Power Systems

PMU data is envisioned to provide the following capabilities:

- Improved accuracy of power system state estimation
- Disturbance location and recognition

PMU data is considered to be a source of Big Data in power systems.

- The fast sampling rate (30 Hz or more) and the increasing deployment of PMUs introduce a challenge to data storage. For example, the Tennessee Valley Authority with 120 PMUs installed throughout the eastern half of North America needs to manage 36 GB data per day.

PMU Data Analysis

State estimation

- (Zhou & Centeno & Thorp & Phadke 06, Vanfretti & Chow & Sarawgi & Fardanesh 11, Rice & Heydt 06.)

Stability analysis and post-event analysis

- (Messina & Vittal 07, Vanfretti & Chow 10, Zhou & Huang & Tuffner & Pierre & Jin 10.)

Disturbance detection and location, dynamic security assessment

- (Lin & Liu & Chen 04, Ma & Makarov & Miller & Nguyen 08, Fan & Kavasseri & Miao & Osborn & Bilke 08)

Lack a common framework. Usually exploit the spatial and the temporal correlations in the measurements separately. We consider the missing data recovery problem in PMU data.

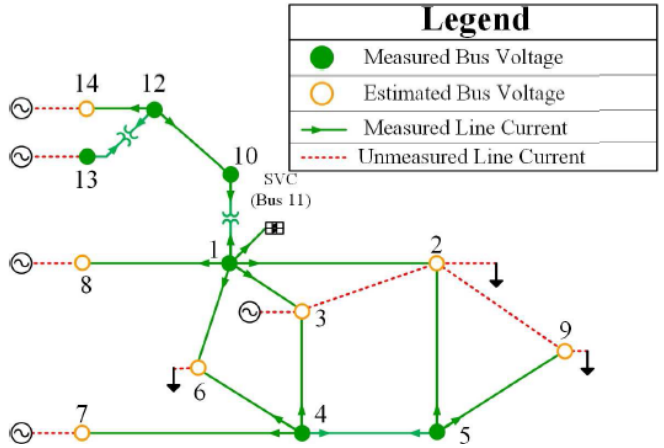
Collective Processing on Spatial-temporal PMU Data Blocks

- Challenge: High dimensionality of the PMU data blocks.

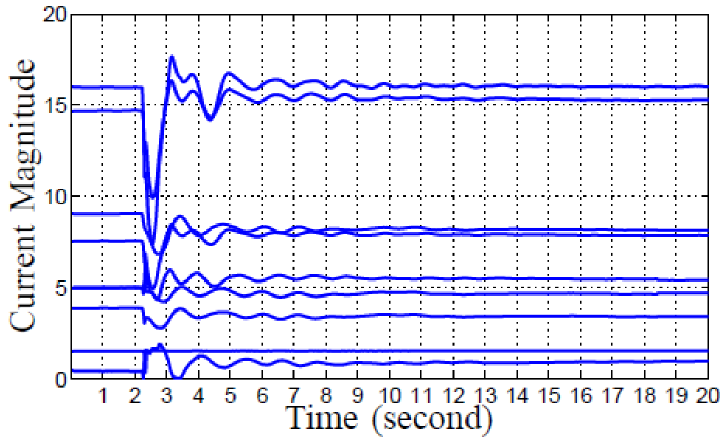
  High sampling rate, process several PMUs at the same time.

- Technical approach: Exploiting the low-dimensional structures in the PMU data blocks.

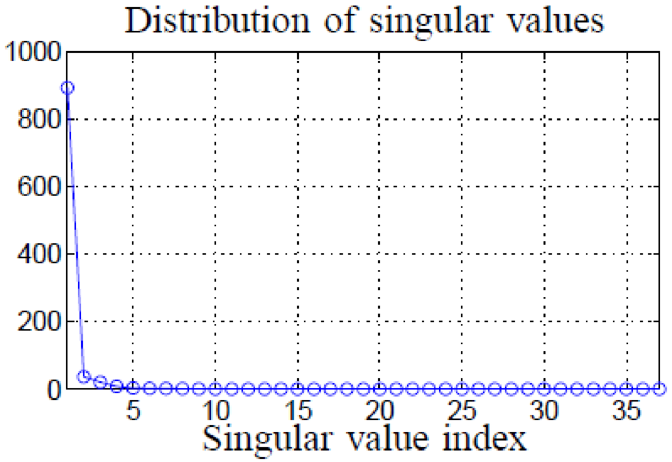
A Simple Low-dimensional Structure: Low-rankness



PMUs in Central NY Power System



Current magnitudes of PMU data

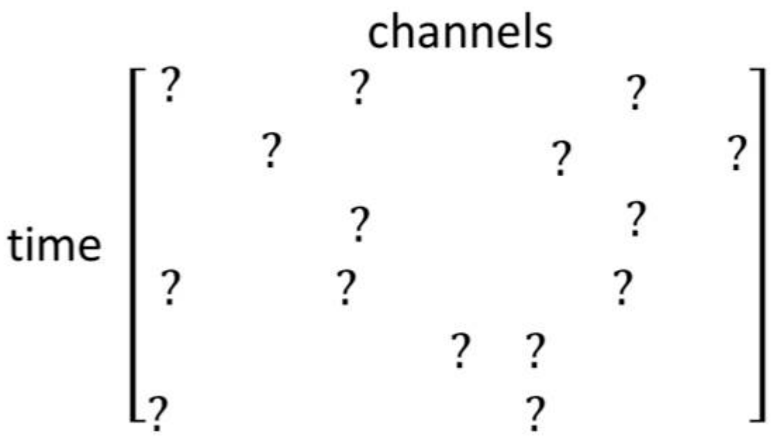


Singular values of the PMU data matrix

- 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-rankness also used in Chen & Xie & Kumar 2013, Dahal & King & Madani 2012 for dimensionality reduction.

Missing Data Recovery in Low-rank PMU Data Matrix

- Data Losses happen due to PMU malfunction or communication congestion between PMUs and Phasor Data Concentrator (PDC).
- Existing missing data recovery: interpolation from measurements in the same channel.
- Our approach: leverage low-rankness of PMU data blocks.



Low-rank matrix completion problem!

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).

Our Previous Results on Low-rank Matrix Completion

- **Theoretical guarantee of low-rank matrix completion when the locations of missing points are correlated.**

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.

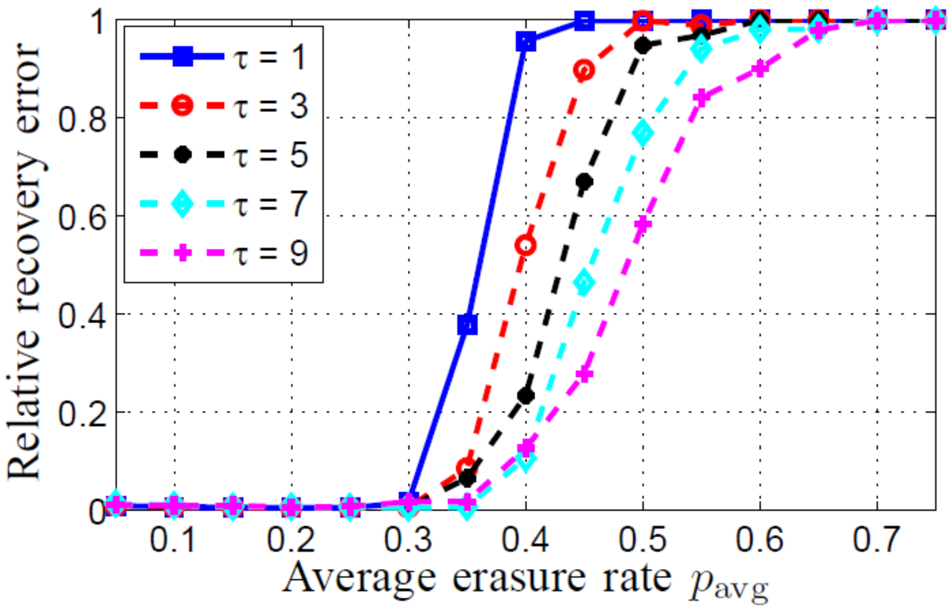
- **Online Algorithm for PMU Data Processing (OLAP)**

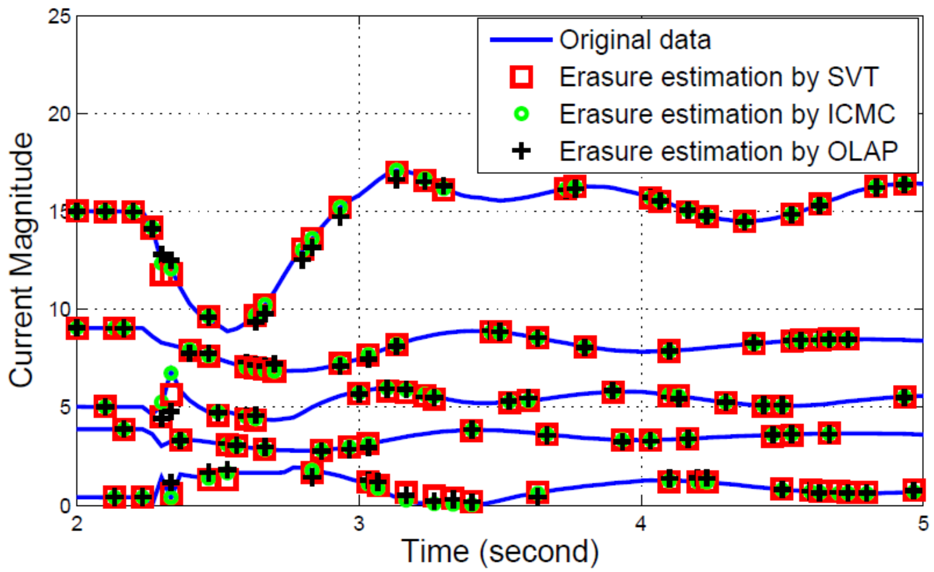
OLAP can fill in the missing data for real-time applications like state estimation and disturbance detection.

Gao, Wang, Ghiocel, and Chow, IEEE Power & Energy Society General Meeting 2014, accepted at IEEE Trans on Power Systems 2014.

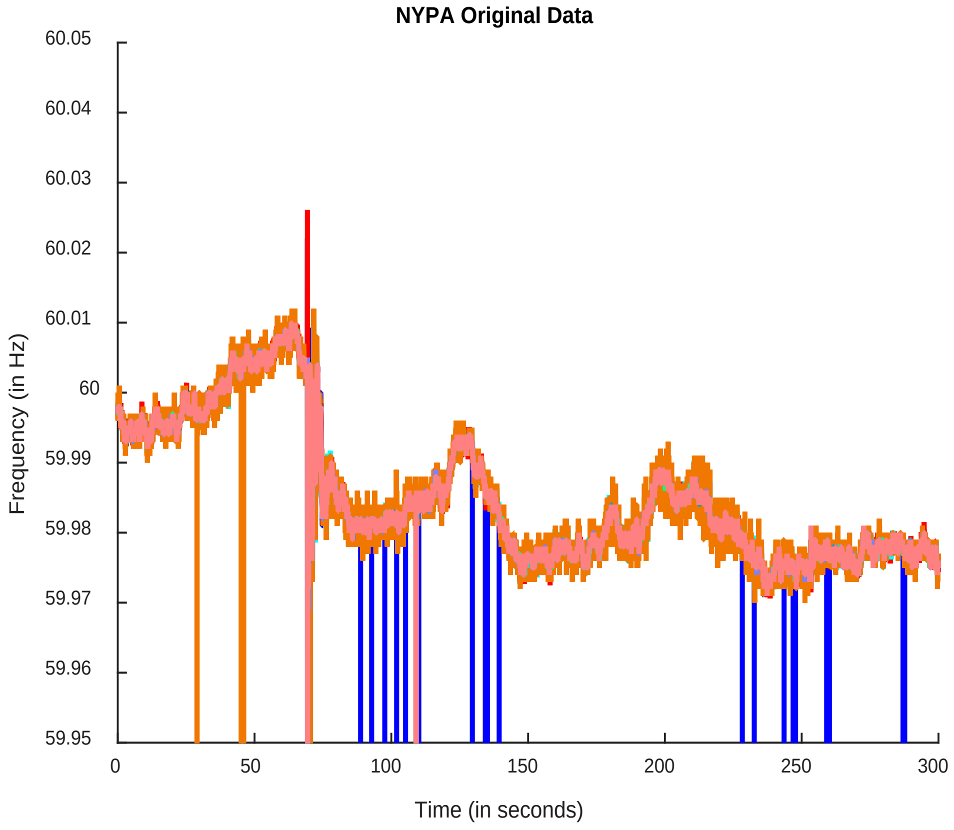
Missing Data Recovery by Low-rank Matrix Completion

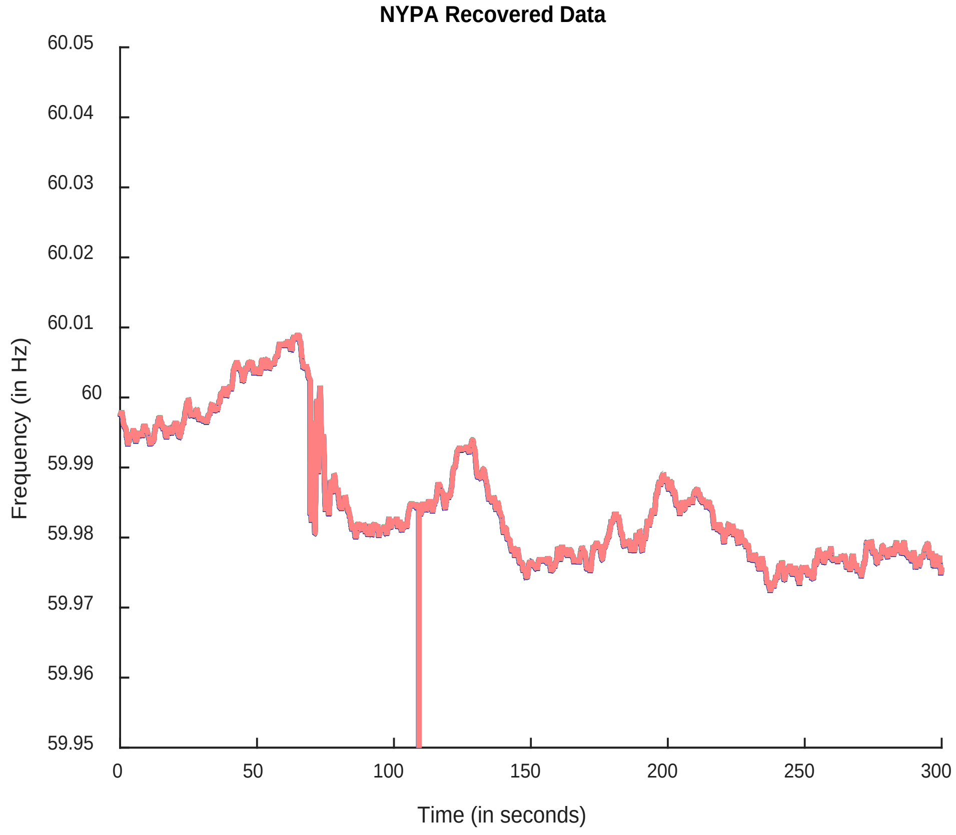
- Low-rank matrix completion method can tolerate up to 30% erasure rate.
- OLAP performances very close to block processing methods.





Missing Data Recovery by Low-rank Matrix Completion



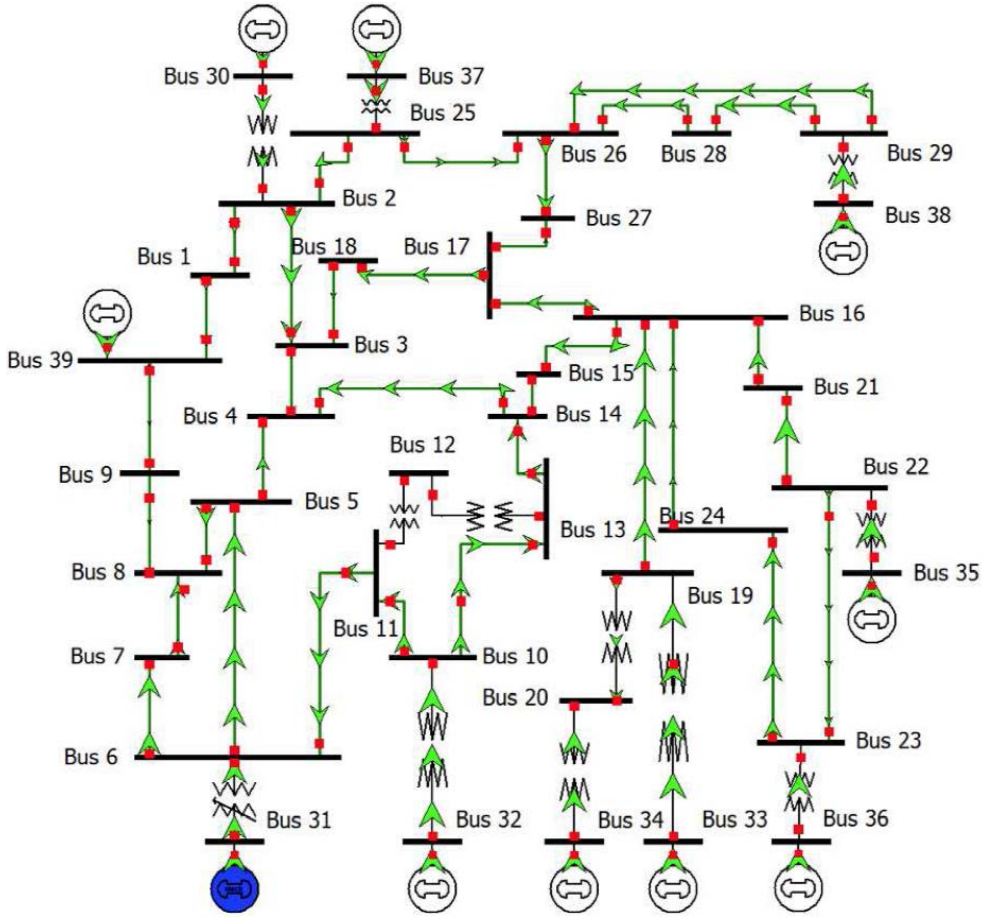


From Low-rankness to General Low-dimensionality

- Low-rankness: When voltage and current phasors are affected by one disturbance, the corresponding columns belong to the same low-dimensional subspace.

- Consider the power system in a larger area. The rank of the PMU data matrix increases if the system experiences multiple disturbances.

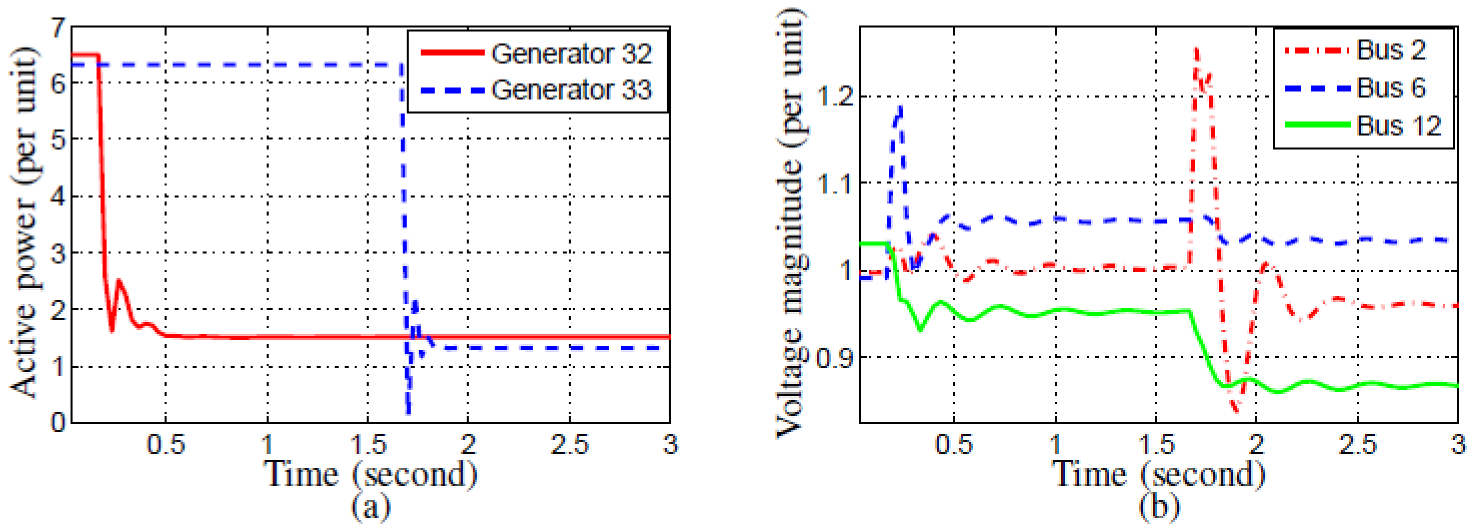
From Low-rankness to General Low-dimensionality



- We use Power System Toolbox (PST) to simulate the PMU data based on the linear model of IEEE 39 New England Power System

- We assume that sixteen PMUs are installed at bus 2, 4, 6, 8, 10, 12, 16, 18, 20, 22, 26, 33, 36, 37, 38 and 39.

From Low-rankness to General Low-dimensionality



Consider a 57×90 PMU data matrix.

When there is one disturbance in the power system, the approximate rank is 7.

When there are two disturbances in the power system, the approximate rank is 16.

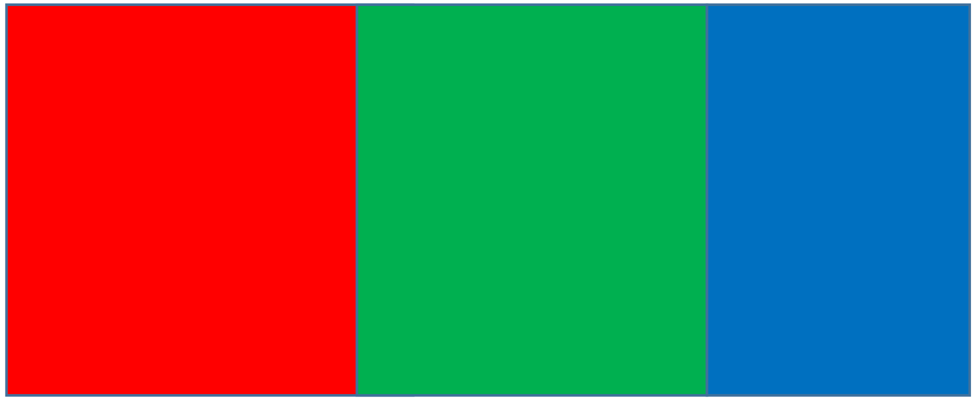
Each event corresponds to a low-dimensional subspace. Low-rankness is no longer the best model to describe the datasets containing multiple events.

One Example of the General Low-dimensional Models

Eriksson & Balzano & Nowark 11

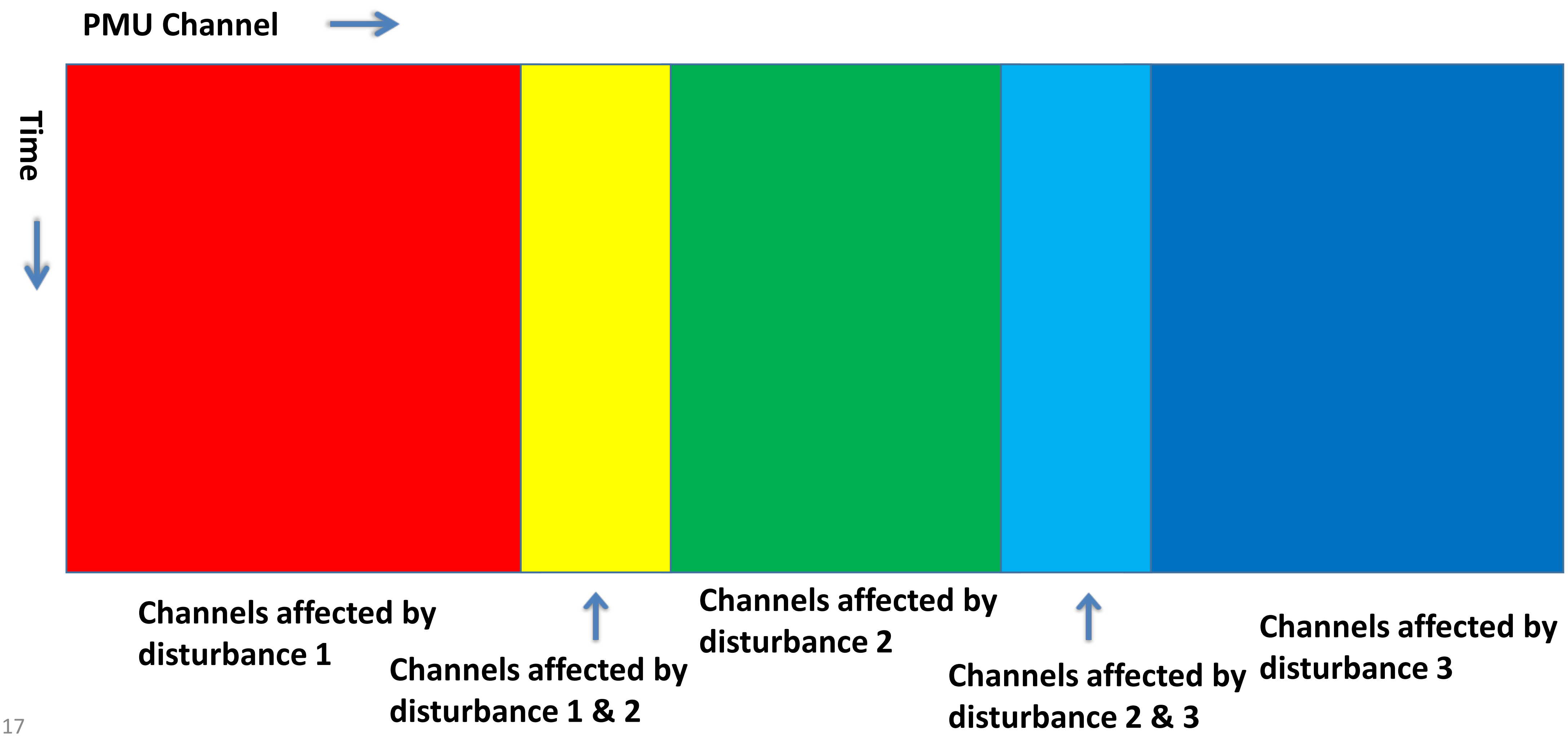
(High-Rank Matrix Completion and Subspace Clustering with Missing Data)

- Consider the matrix completion problem that the columns belong to the union of subspaces (UoS).



Columns belong to subspace 1 **Columns belong to subspace 2** **Columns belong to subspace 3**

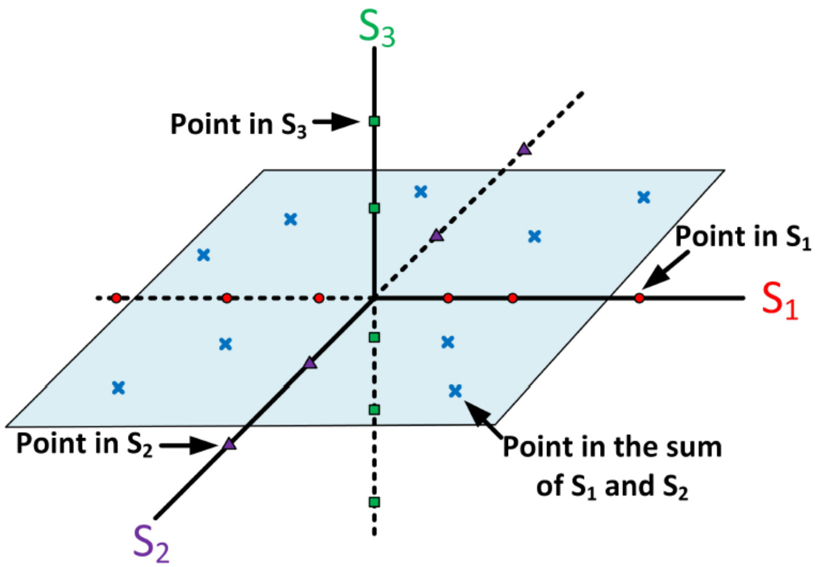
Limitation of Using UoS Model in PMU Data



Our Proposed General Model

- A vector belongs to the sum of n subspaces if and only if it can be decomposed into n vectors that belong to the n subspaces respectively.

- Each subspace corresponds to one disturbance. The sum of subspaces corresponds to the overlapping area of multiple disturbances.



Lines S_1, S_2 and S_3 are three one-dimensional subspace in $\mathbb{R}^3$.

Missing Data Recovery under the Model of Union and Sums of Subspaces

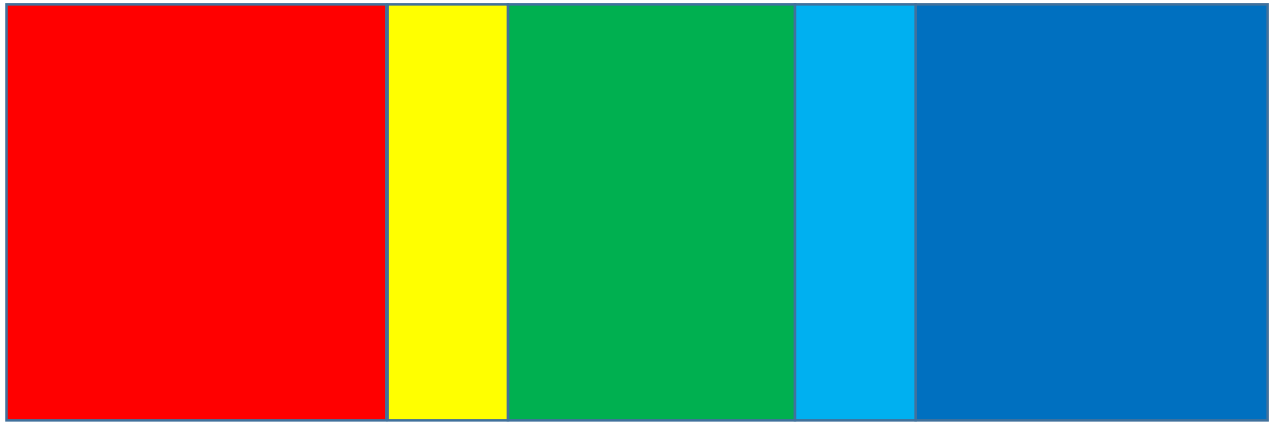
- Borrow the idea from **Eriksson & Balzano & Nowark 11**
- Key technical difference: Address the overlapping effects modeled by the sum of subspaces.

Our Method

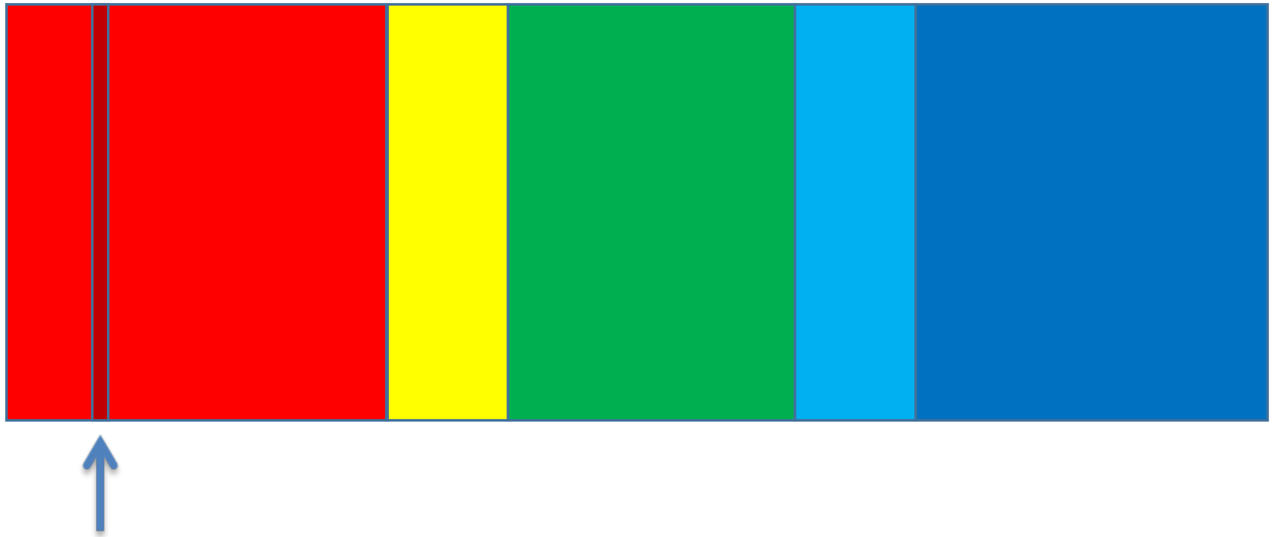
STEP 1: Find Local Neighborhood & Estimate Local Subspace

- select a column uniformly at random and find its near-neighbors
- apply robust low-rank matrix completion to columns in neighborhood to estimate local subspace

Find Local Neighborhood

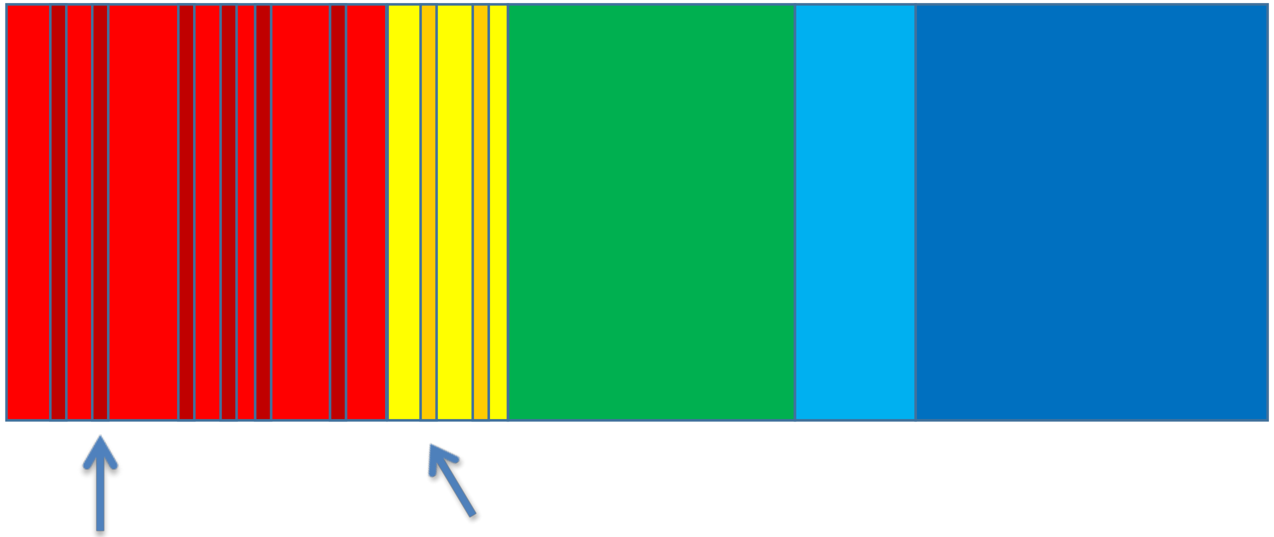


Find Local Neighborhood



Randomly select a column (seed)

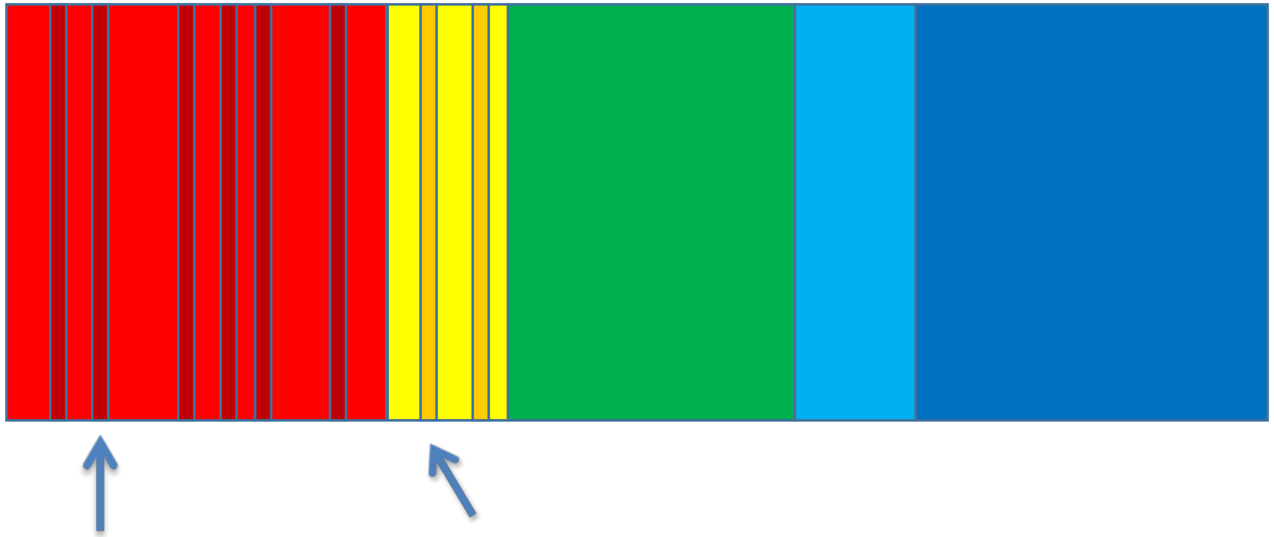
Find Local Neighborhood



Randomly select a column (seed)

Find its near-neighbors based on (partial) distance measurements

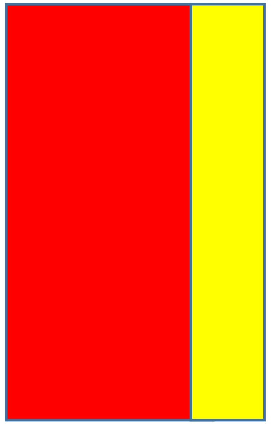
Find Local Neighborhood



Randomly select a column (seed)

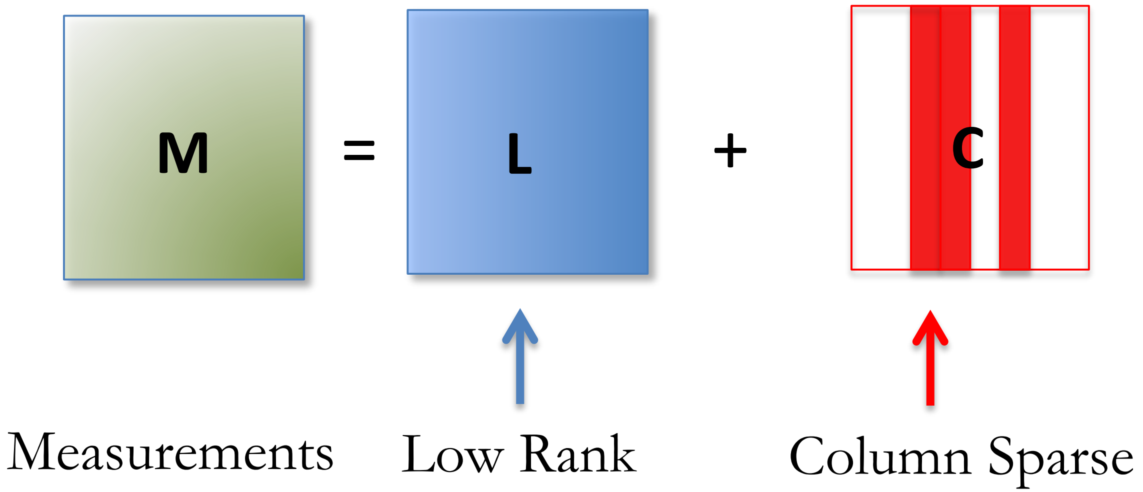
Find its near-neighbors based on (partial) distance measurements

Local neighbor matrix



The columns do not belong to one single subspace!

Robust Matrix Completion with Corrupted Columns



Low-rank matrix completion problem when some number of the columns are arbitrarily corrupted.

Chen & Xu & Caramanis 11 shows that the true subspace can be recovered by

$$
\begin{aligned}
&min_{L,C}\ \ ||L||_* + \lambda ||C||_{1,2} \\
&subject\ to\ \ M_\Omega = L_\Omega + C_\Omega
\end{aligned}
$$

Our Method

STEP 1: Find Local Neighborhood & Estimate Local Subspace

- select a column uniformly at random and find its near-neighbors
- apply robust low-rank matrix completion to columns in neighborhood to estimate local subspace

Repeat to generate a number of candidate subspaces

STEP 2: Subspace Pruning & Full Matrix Completion

- discard all subspaces spanned by union of two or more others
- project each remaining column on to all possible sum of subspaces to determine which it belongs to and impute missing data

Theoretical Results

When the measurements do not contain noise

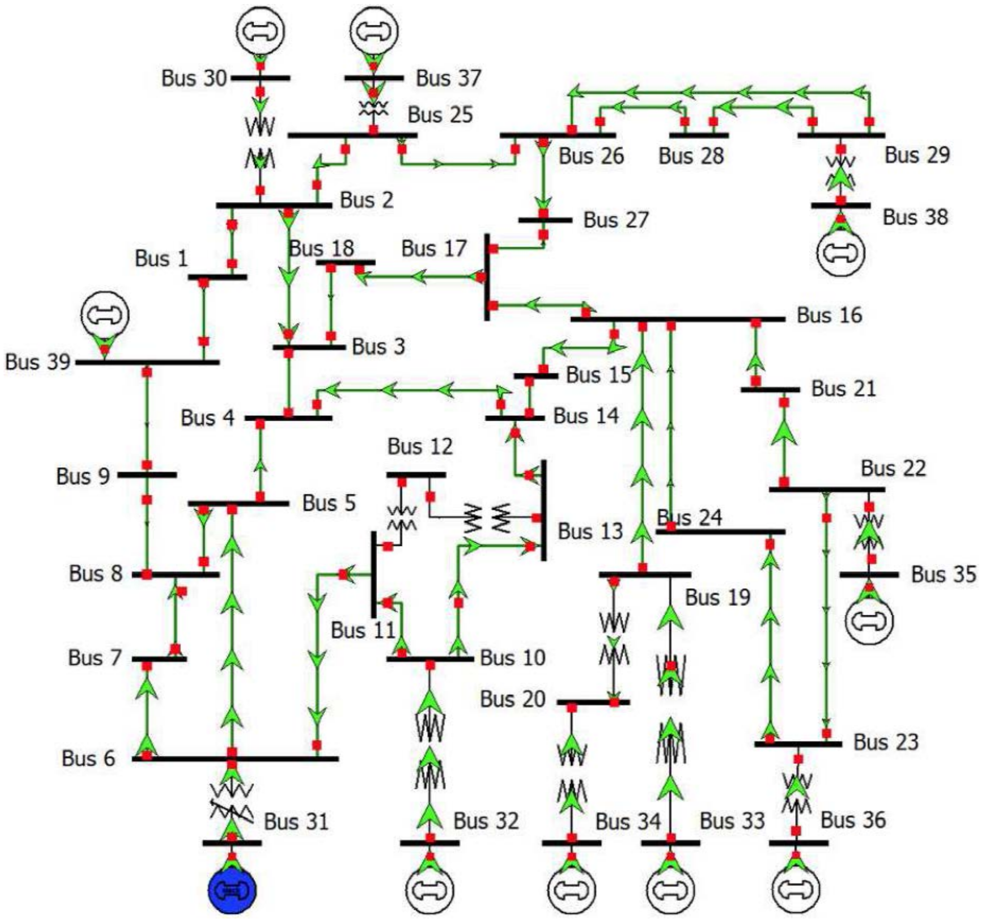
- **We prove that $\mathbf{O}\left(\boldsymbol{r}^{2}\mathbf{N}\boldsymbol{log}^{3}\boldsymbol{n}\right)$ random observation are sufficient to determine the missing points of each column of an $\mathbf{n} \times \mathbf{N}$ matrix with columns belonging to the "union and sums" of $\mathbf{k}$ $\mathbf{r}$-dimensional subspaces.**

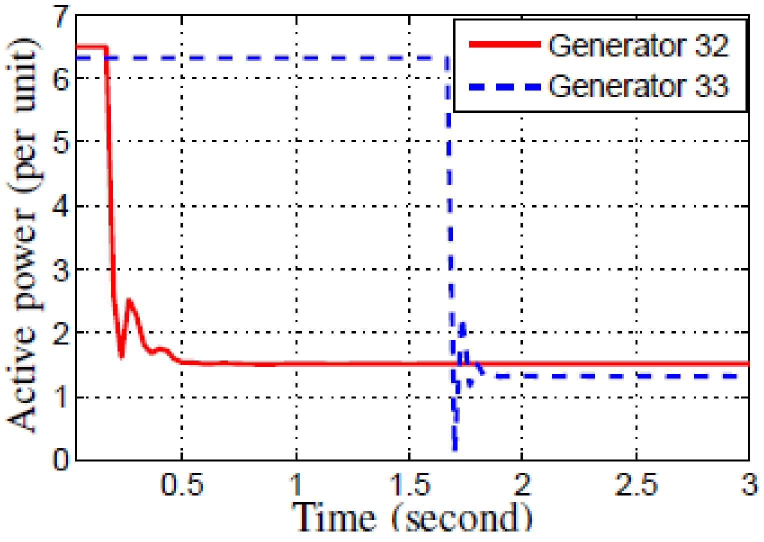
When the measurements contains noise

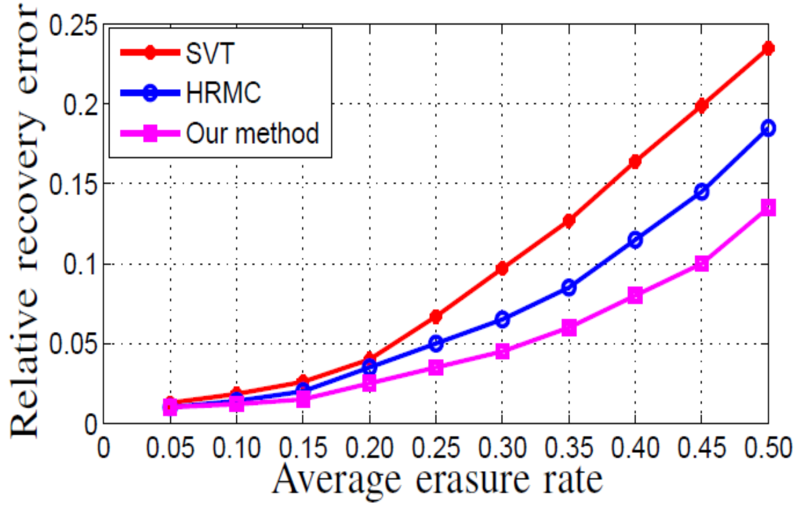
- **We provide upper bounds of recovery errors with respect to the square of the noise level.**

Gao, Wang, Chow, Berger and Seversky, IEEE GlobalSIP 2015, submission to IEEE Journal of selected topics in signal processing 2015.

Simulation







HRMC: Method proposed in Eriksson's paper.

Conclusions

- Low-dimensional structures exist in high-dimensional PMU data.
- Propose a new model to characterize the low-dimensional structures in PMU data.
- Show the theoretical guarantee and numerical results of our proposed method.


Q&A