Model-less Data Quality Improvement of Streaming Synchrophasor Measurements by Exploiting the Low-Rank Hankel Structure

Yingshuai Hao, *Student member, IEEE,* Meng Wang, *Member, IEEE*, Joe H. Chow, *Fellow, IEEE*, Evangelos Farantatos, *Member, IEEE*, Mahendra Patel, *Fellow, IEEE*

***Abstract*—This paper presents a new framework to improve the quality of streaming synchrophasor measurements with the existence of missing data and bad data. The method exploits the low-rank property of the Hankel structure to identify and correct bad data, as well as to estimate and fill in missing data. The method is advantageous compared to existing methods in the literature that only estimate missing data by leveraging the low-rank property of the synchrophasor data observation matrix. The proposed algorithm can efficiently differentiate event data from bad data, even in the existence of simultaneous and consecutive bad data. The algorithm has been verified through numerical experiments on recorded synchrophasor datasets.**

***Index Terms*—missing data recovery, bad data detection, phasor measurement unit, Hankel matrix, low dimensionality**

I. Introduction

Phasor measurement units (PMU) provide time-synchronized high-resolution measurements of voltage and current phasors at rates from 30 up to 120 samples per second. Due to various reasons including communication congestions and PMU malfunctions, data losses and data quality degradation are common in PMU measurements [1]. California Independent System Operator (ISO) reported that 10%-17% of data it received in 2011 had availability and quality issues [2]. Data quality issues limit the use of synchrophasor data and synchrophasor applications like state estimation [3] and disturbance identification [4] in the control room. Thus, missing points and bad measurements should be recovered and corrected in a timely fashion, respectively.

Traditionally, accurate data recovery requires measurement redundancy and power system models. The data recovery performance relies heavily on the accuracy of the system model. The recent abundance of synchronized PMU data enables model-less methods for data quality improvements. The spatial-temporal blocks of synchrophasor data exhibit intrinsic low-dimensional structure despite the high ambient dimension. This low-dimensionality has been recently exploited for dimensionality reduction [4], [5], missing data recovery [6], [7], and bad data detection [8]. Reference [6] for the first time applies low-rank matrix completion methods to recover location-correlated data losses that result from communication congestions. Reference [7] proposes a new low-dimensional model to characterize the measurements that are affected by multiple events in power systems and develops a data recovery method for this model. Reference [8] develops an identification method for bad data and cyber data attacks assuming the ground-truth measurements are low-rank. These methods are block-processing methods that fill in the data or correct the errors in a fixed time window simultaneously.

Online missing data recovery methods that exploit the low-dimensionality have been developed. The methods in [9], [10] assume that the data points belong to a time-varying low-dimensional subspace and fill in the missing data by continuously updating the subspace and projecting the partial observations to the subspace. These methods assume the dimension of the subspace does not change. Reference [6] also proposes an online missing data recovery method that adapts to the dimension change of the PMU data.

For online identification of bad synchrophasor measurements, reference [11] employs a phasor state estimator to identify angle biases and current scaling errors. Reference [12] develops a model-less approach to identify sparse bad measurements through exploiting the data similarities across measuring channels. Sparse bad data means that at most a small fraction of the measurements are erroneous at a given time. The sparsity assumption is also employed in [13], which further exploits the low-dimensionality of the data to correct bad measurements in network applications.

All the above block-processing and online methods require at least a few accurate measurements at each time instant to recover the missing points and correct bad data. If the measurements at one time instant are completely lost or corrupted, none of these method will succeed in data recovery. Since communication congestion could lead to data losses, simultaneous data losses across all channels for one instant or even consecutively are not uncommon in PMU data.

This paper develops an *online* algorithm that estimates the missing entries and corrects the bad measurements in streaming synchrophasor data. Central to the model-less method is the low-dimensionality of the Hankel matrix, which naturally characterizes the power system dynamics without modeling the dynamical system. Reference [14] exploits this low-rank Hankel structure to develop a block-processing method to recover missing data. It does not handle streaming data and does not consider bad data. The online algorithm developed in this paper has two distinctive features that have not been

Y. Hao, M. Wang, and J.H. Chow are with the Dept. of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY. Email: {haoy2, wangm7, chowj}@rpi.edu. E. Farantatos and M. Patel are with Electric Power Research Institute, Palo Alto, CA. Email: {efarantatos, mpatel}@epri.com

addressed in all the above references. 1) It can recover the data points accurately even when simultaneous data losses and bad measurements happen across all PMU channels consecutively. 2) It can also distinguish consecutive simultaneous bad data from system disturbances. Our algorithm has low memory cost and computational complexity and can pre-process streaming synchrophasor data for subsequent real-time applications.

The rest of the paper is organized as follows. Section II demonstrates and analyzes the low-rank property of the Hankel synchrophasor data block. Section III presents the proposed approach to correct bad data and fill in the missing data. Section IV provides the theoretical performance analysis. Sections V documents the results of the numerical experiments on recorded PMU datasets. Section VI concludes the paper.

Notation: Vectors are bold lowercase, matrices are bold uppercase, and scalars are in normal font. For example, $\mathbf{Z}$ is a matrix and $\mathbf{z}$ is vector. $\mathbf{Z}^T$ and $\mathbf{Z}^*$ denote the transpose and conjugate transpose of $\mathbf{Z}$. $\overline{\mathbf{Z}}$ denotes a matrix with a random column permutation of $\mathbf{Z}$. $\mathbf{z}(i)$ denotes the ith entry of vector $\mathbf{z}$. $\|\mathbf{z}\|$ represents the ℓ_2-norm of $\mathbf{z}$. $\|\mathbf{Z}\|$ and $\|\mathbf{Z}\|_F$ denote the spectral norm and the Frobenius norm of $\mathbf{Z}$, respectively. For one index set Ψ, $|\Psi|$ represents its cardinality, $\mathbf{z}_\Psi$ and $\mathbf{Z}_\Psi$ represent the subvector of $\mathbf{z}$ and submatrix of $\mathbf{Z}$ with only the rows labeled by Ψ, respectively, while $\mathbf{Z}^\Psi$ denotes the submatrix of $\mathbf{Z}$ with only the columns labeled by Ψ.

II. Low-rank Hankel structure of synchrophasor streams

A. Low-rank property of synchrophasor data blocks

To illustrate the low-rank property of the Hankel matrix constructed by synchrophasor data, we use a recorded synchrophasor dataset from six multi-channel PMUs in Central New York Power System. Eleven voltage phasors at voltage levels of 230 kV, 345 kV and 765 kV are measured at a rate of 30 samples per second. Fig. 1 shows a recorded dataset of 20 seconds, where a disturbance occurs around 2.3 seconds.

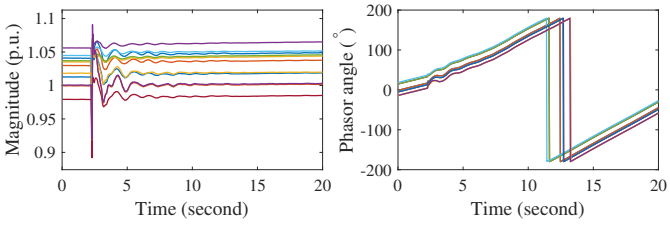

Fig. 1: The measured voltage phasors

Consider a matrix $\mathbf{Y}$ containing the measured phasors from m channels during τ consecutive time instants,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_\tau \end{bmatrix} \in \mathbb{C}^{m\times\tau}, \tag{1}$$

where $\mathbf{y}_i$ denotes the phasors of m channels at instant i. With the singular value decomposition (SVD) of $\mathbf{Y}$ as $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$, if we approximate $\mathbf{Y}$ by the rank-r matrix $\mathbf{U}\mathbf{\Sigma}^r\mathbf{V}^*$, where $\mathbf{\Sigma}^r$ is the diagonal matrix with only the first r largest singular values in $\mathbf{\Sigma}$, then the rank-r approximation error is defined as

$$e^r(\mathbf{Y}) = \|\mathbf{U}\mathbf{\Sigma}^r\mathbf{V}^* - \mathbf{Y}\|_F / \|\mathbf{Y}\|_F = \|\mathbf{\Sigma} - \mathbf{\Sigma}^r\|_F / \|\mathbf{\Sigma}\|_F. \tag{2}$$

For example, the data block $\mathbf{Y} \in \mathbb{C}^{11\times 600}$ in Fig. 1 has a rank-one approximation error of 0.92% and a rank-two approximation error of 0.37%.

The low-rank property of PMU data blocks has been exploited to conduct disturbance identification [4], missing data recovery [6], [7], cyber data attack detection [8], etc. One limitation of the low-rank model is that it does not capture the temporal correlations in time series explicitly. A column permutation of $\mathbf{Y}$, denoted as $\overline{\mathbf{Y}}$, would result in different time series, while the ranks of $\overline{\mathbf{Y}}$ and $\mathbf{Y}$ are the same.

B. Hankel structure and its low-rank property

Hankel matrices are widely used in the analysis of linear time-invariant (LTI) systems, like the system order identification and realization [15]. Let κ denote the number of vectors in each column of the Hankel matrix. The constructed Hankel matrix $\mathcal{H}_\kappa(\mathbf{Y})$ from $\mathbf{Y}$ with parameter κ is

$$\begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_{\tau-\kappa+1} \\ \mathbf{y}_2 & \mathbf{y}_3 & \cdots & \mathbf{y}_{\tau-\kappa+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_\kappa & \mathbf{y}_{\kappa+1} & \cdots & \mathbf{y}_\tau \end{bmatrix} \in \mathbb{C}^{\kappa m\times(\tau-\kappa+1)}. \tag{3}$$

We observe that for synchrophasor data, the Hankel matrix $\mathcal{H}_\kappa(\mathbf{Y})$ is still low-rank. Note that the low-rank Hankel property does not hold in general for a low-rank matrix. Fig. 2 shows the approximation errors of the constructed Hankel matrices using the data in Fig. 1 with different approximate rank r and κ. All the matrices $\mathcal{H}_\kappa(\mathbf{Y})$ can be approximated by a rank-2 or rank-3 matrix with a negligible error.

The low-rankness of the Hankel matrix can be verified from the linearized dynamic system model [14]. The linear state-space model around an equilibrium point is

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t, \;\; \mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t, \tag{4}$$

with state vector $\mathbf{x}_t \in \mathbb{C}^n$, output vector $\mathbf{y}_t \in \mathbb{C}^m$, input vector $\mathbf{u}_t \in \mathbb{C}^p$, state matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, input matrix $\mathbf{B} \in \mathbb{C}^{n\times m}$, output matrix $\mathbf{C} \in \mathbb{C}^{m\times n}$ and feedthrough matrix $\mathbf{D} \in \mathbb{C}^{m\times p}$.

The stability of power systems can be studied from the eigen-decomposition of $\mathbf{A}$ [16]. Suppose $\mathbf{A}$ could be diagonalized as $\mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$, where $\mathbf{P} = [\mathbf{p}_1, \cdots, \mathbf{p}_n]$, $\mathbf{P}^{-1} = [\mathbf{q}_1, \cdots, \mathbf{q}_n]^*$, $\mathbf{p}_i \in \mathbb{C}^n$, $\mathbf{q}_i \in \mathbb{C}^n, i \in \{1, \cdots, n\}$, and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_\mathbf{1}, \cdots, \lambda_\mathbf{n})$. Then with an impulse input, we have

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t = \mathbf{C}\mathbf{A}^{t-1}\mathbf{x}_1 = \sum_{i=1}^{n} \lambda_i^{t-1}\mathbf{q}_i^*\mathbf{x}_1\mathbf{C}\mathbf{p}_i. \tag{5}$$

Here λ_i corresponds to the ith mode of the system. In practice, a mode might be highly damped ($|\lambda_i| \approx 0$), or not excited by the input ($\mathbf{q}_i^*\Delta\mathbf{x}_1 \approx 0$), or not directly measured ($\|\mathbf{C}\mathbf{p}_i\| \approx 0$). If $\mathbf{y}_t$ only includes r dominant modes, denoted by $\lambda_1, \cdots, \lambda_r$ for notational simplicity, we have

$$\mathbf{y}_t \simeq \sum_{i=1}^{r} \lambda_i^{t-1}\mathbf{q}_i^*\mathbf{x}_1\mathbf{C}\mathbf{p}_i. \tag{6}$$

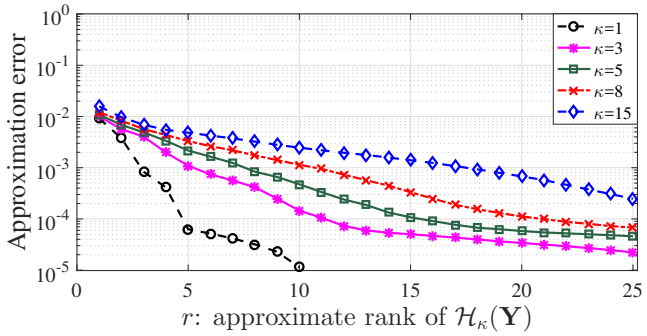

Fig. 2: The approximation errors of Hankel matrices from $\mathbf{Y}$

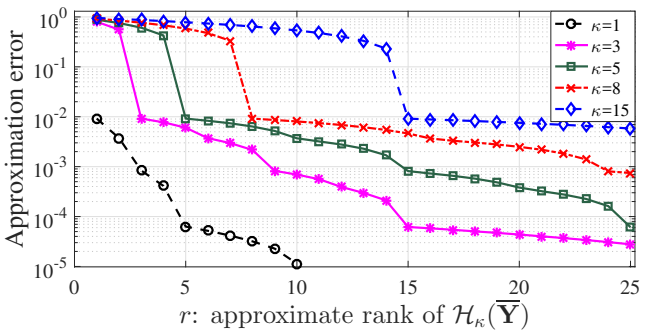

Fig. 3: The approximation errors of Hankel matrices from $\overline{\mathbf{Y}}$

The constructed Hankel matrix $\mathcal{H}_\kappa(\mathbf{Y}) = \mathbf{E}_L\mathbf{E}_R$, where

$$\mathbf{E}_L = \begin{bmatrix} \mathbf{C}\mathbf{p}_1 & \mathbf{C}\mathbf{p}_2 & \cdots & \mathbf{C}\mathbf{p}_r \\ \lambda_1\mathbf{C}\mathbf{p}_1 & \lambda_2\mathbf{C}\mathbf{p}_2 & \cdots & \lambda_r\mathbf{C}\mathbf{p}_r \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{\kappa-1}\mathbf{C}\mathbf{p}_1 & \lambda_2^{\kappa-1}\mathbf{C}\mathbf{p}_2 & \cdots & \lambda_r^{\kappa-1}\mathbf{C}\mathbf{p}_r \end{bmatrix} \in \mathbb{C}^{\kappa m\times r}, \tag{7}$$

$$\mathbf{E}_R = \begin{bmatrix} \mathbf{q}_1^*\mathbf{x}_1 & \lambda_1\mathbf{q}_1^*\mathbf{x}_1 & \cdots & \lambda_1^{\tau-\kappa}\mathbf{q}_1^*\mathbf{x}_1 \\ \mathbf{q}_2^*\mathbf{x}_1 & \lambda_2\mathbf{q}_2^*\mathbf{x}_1 & \cdots & \lambda_2^{\tau-\kappa}\mathbf{q}_2^*\mathbf{x}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_r^*\mathbf{x}_1 & \lambda_r\mathbf{q}_r^*\mathbf{x}_1 & \vdots & \lambda_r^{\tau-\kappa}\mathbf{q}_r^*\mathbf{x}_1 \end{bmatrix} \in \mathbb{C}^{r\times(\tau-\kappa+1)}. \tag{8}$$

Then we can conclude $\text{rank}(\mathcal{H}_\kappa(\mathbf{Y})) \le r, \forall\kappa$. Since $\mathbf{Y} = \mathcal{H}_1(\mathbf{Y})$, $\text{rank}(\mathbf{Y}) \le r$.

The low-rank Hankel property also holds beyond impulse inputs. If a time-varying input $\mathbf{u}_t$ can be decomposed as $\mathbf{u}_t = \sum_{i=1}^d \beta_i^t\mathbf{w}_i$, for some $\mathbf{w}_i \in \mathbb{C}^p$ and constant β_i, $i = 1, \cdots, d$, then with $\mathbf{D} = \mathbf{0}$ in (4), the rank of $\mathcal{H}_\kappa(\mathbf{Y})$ is bounded by $r + d$. The proof is quite simple and we skip the details.

We remark that the above analysis on dynamical systems is only introduced to illustrate the low-rank Hankel property. Our method exploits this low-rank property and does not require any information or estimation of the dynamical model.

C. Differentiation of event data from consecutive bad data using the Hankel structure

Based on the above discussion, when an event happens, the corresponding Hankel matrix of the synchrophasor data would exhibit low-rank property. With a random column permutation, the data matrix will lose the temporal correlation in the time series, and the corresponding Hankel matrix after permutation will not be low-rank. For consecutive bad data, which may result from malfunction of PMUs or malicious data manipulation, if there is weak or no dynamic correlation in the bad data, then the corresponding Hankel data matrix is not low-rank. In this case, a random permutation would not affect the rank much. Thus one can exploit the low-rank Hankel structure to differentiate consecutive bad data from event data. The idea is to randomly permute the columns of the data matrix. If the rank of the Hankel matrix of the original data matrix is much less than the rank of the Hankel matrix of the permuted matrix, then it corresponds to an event. If the two ranks are similar, it corresponds to bad data.

To illustrate this idea, consider the data block $\mathbf{Y}$ in Fig. 1. Let $\overline{\mathbf{Y}}$ denote the resulting matrix of one realization after applying a random column permutation to $\mathbf{Y}$. Fig. 3 demonstrates the approximation errors of $\mathcal{H}_\kappa(\overline{\mathbf{Y}})$. Compared to the approximation errors $e^r(\mathcal{H}(\mathbf{Y}))$ shown in Fig. 2, we can find $e^r(\mathcal{H}_\kappa(\overline{\mathbf{Y}})) \gg e^r(\mathcal{H}_\kappa(\mathbf{Y}))$ for any fixed $\kappa > 1$ and r. Moreover, in this example, $\mathbf{Y}$ can be approximated by a rank-one matrix with a relative error of 10^{-2}, then $\overline{\mathbf{Y}}$ can also be approximated by a rank-one matrix, because the column permutation does not change the rank. Then the Hankel matrix $\mathcal{H}_\kappa(\overline{\mathbf{Y}})$ can be approximated by a rank-κ matrix with a relative error around 10^{-2}. This explains the sharp drop of the approximation error at $r = \kappa$ in Fig. 3.

We further provide a simplified example to illustrate the difference between $\mathcal{H}_\kappa(\mathbf{Y})$ and $\mathcal{H}_\kappa(\overline{\mathbf{Y}})$ when $\mathbf{Y}$ contains event data. Suppose in (6), let $m = 1$ and $r = 1$, i.e., we only have one channel of measurements and one dominant mode. Assume $\mathbf{q}_1^*\mathbf{x}_1\mathbf{C}\mathbf{p}_1 = 1$ for notational simplicity and let $\alpha = \lambda_1$. Assume $|\lambda_1| < 1$. Let $\tau = 6$, then $\mathbf{Y} = [\alpha^0, \alpha^1, \cdots, \alpha^5]$. Let $\kappa = 3$, then

$$\mathcal{H}_3(\mathbf{Y}) = \begin{bmatrix} \alpha^0 & \alpha^1 & \alpha^2 & \alpha^3 \\ \alpha^1 & \alpha^2 & \alpha^3 & \alpha^4 \\ \alpha^2 & \alpha^3 & \alpha^4 & \alpha^5 \end{bmatrix}.$$

The rank of $\mathcal{H}_3(\mathbf{Y})$ is 1. Suppose after a random permutation, $\overline{\mathbf{Y}} = [\alpha^0, \alpha^3, \alpha^2, \alpha^4, \alpha^1, \alpha^5]$, then the rank of $\mathcal{H}_3(\overline{\mathbf{Y}})$ is 3. Thus, we have $e^1(\mathcal{H}_3(\mathbf{Y})) = 0$, while $e^1(\mathcal{H}_3(\overline{\mathbf{Y}})) > 0$.

There is no evident difference between $e^r(\mathcal{H}_\kappa(\mathbf{Y}))$ and $e^r(\mathcal{H}_\kappa(\overline{\mathbf{Y}}))$, if there is weak or no temporal correlation in $\mathbf{Y}$. Existing bad data detectors such as [12], [13] can succeed only if there exists a small fraction of bad measurements at one time instant and fail to identify simultaneous bad data across channels. In the proposed method, using the low-rank Hankel structure, we can identify bad data even if bad data happen across all channels simultaneously and consecutively.

III. Online Algorithm for Data Estimation of Bad and Missing Synchrophasor Data

Let $\mathbf{y}_t$ denote the obtained observations (possibly with bad data and missing points) at time instant t and $\widehat{\mathbf{y}}_t$ denote the estimated measurements after removing bad data and filling in missing points. Given any channel i, if $\mathbf{y}_t(i) = \widehat{\mathbf{y}}_t(i)$, $\mathbf{y}_t(i)$ and $\widehat{\mathbf{y}}_t(i)$ are considered as trusted data. Otherwise, $\mathbf{y}_t(i)$ is either bad data or missing data, and $\widehat{\mathbf{y}}_t(i)$ is an estimated data.

A. Main idea of the algorithm

The algorithm initializes with trusted $\mathbf{y}_1 \sim \mathbf{y}_L$ in a window of length L. If $\mathbf{y}_1 \sim \mathbf{y}_L$ contain missing data or bad data, one

could apply block-processing methods such as [17], [18] to fill in missing points and correct bad data to obtain $\widehat{\mathbf{y}}_1 \sim \widehat{\mathbf{y}}_L$.

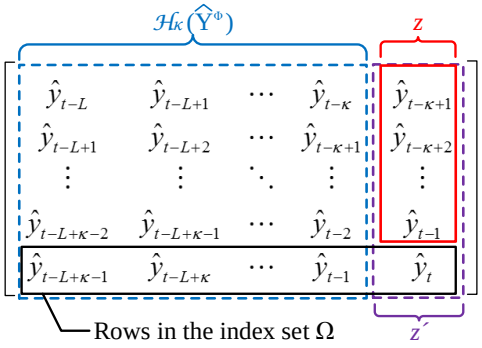

Fig. 4: Illustration of the Hankel matrix constructed at time t

We first explain the main idea of the algorithm assuming all the data points are trusted. The key is to exploit the fact that the Hankel matrix constructed by $\widehat{\mathbf{y}}_{t-L} \sim \widehat{\mathbf{y}}_t$, as shown in Fig. 4, has rank r. Let Φ denote the set $\{t-L, ..., t-1\}$, and let $\widehat{\mathbf{Y}}^\Phi$ denote the data block $\widehat{\mathbf{y}}_{t-L} \sim \widehat{\mathbf{y}}_{t-1}$, i.e., $\widehat{\mathbf{Y}}^\Phi = [\widehat{\mathbf{y}}_{t-L}, ..., \widehat{\mathbf{y}}_{t-1}]$. The Hankel matrix in Fig. 4 can be decomposed by two parts: the first $L-\kappa+1$ columns, which is $\mathcal{H}_\kappa(\widehat{\mathbf{Y}}^\Phi)$, and the last column, denoted by $\mathbf{z}'$.

The algorithm first computes the SVD of $\mathcal{H}_\kappa(\widehat{\mathbf{Y}}^\Phi)$ to estimate its r-dimensional column basis, denoted by $\widehat{\mathbf{U}}^r$ (Step 1). Since the last column $\mathbf{z}'$ belongs to the span of $\widehat{\mathbf{U}}^r$, it can be written as a linear combination of the columns in $\widehat{\mathbf{U}}^r$. Steps 2 and 3 compute vector $\widehat{\mathbf{d}} \in \mathbb{C}^r$ such that $\mathbf{z}'$ equals to $\widehat{\mathbf{U}}^r\widehat{\mathbf{d}}$. Since in the worst case, the measurements at time t can be all lost or all corrupted, we only use the top $(\kappa-1)m$ entries in $\mathbf{z}'$, denoted by $\mathbf{z}$ in Fig. 4 and defined in (9), to compute $\widehat{\mathbf{d}}$. Mathematically, let $\Psi = \{1, \cdots, (\kappa-1)m\}$, $\widehat{\mathbf{U}}^r_\Psi$ denotes the submatrix of $\widehat{\mathbf{U}}^r$ by only keeping the top $(\kappa-1)m$ rows. Then $\mathbf{z}$ equals to $\widehat{\mathbf{U}}^r_\Psi\widehat{\mathbf{d}}$. $\widehat{\mathbf{d}}$ is computed by solving a least-squares problem in (10) in Step 3.

With the computed $\widehat{\mathbf{d}}$, let $\Omega = \{(\kappa-1)m+1, \cdots, \kappa m\}$, we use $\mathbf{h} = \widehat{\mathbf{U}}^r_\Omega\widehat{\mathbf{d}}$ as the prediction of $\mathbf{y}_t$. If $\mathbf{y}_t(i)$ is close to the corresponding prediction $\mathbf{h}(i)$, we let $\widehat{\mathbf{y}}_t(i) = \mathbf{y}_t(i)$, and both are labeled as trusted (Step 4). Otherwise, $\widehat{\mathbf{y}}_t(i)$ is untrusted, and $\widehat{\mathbf{y}}_t(i) = \mathbf{h}(i)$.

With the new observations $\widehat{\mathbf{y}}_t$, the algorithm recomputes $\widehat{\mathbf{d}}$ by $\mathbf{z}'$. Since $\widehat{\mathbf{U}}^r\widehat{\mathbf{d}}$ equals $\mathbf{z}'$, $\widehat{\mathbf{d}}$ is computed by solving a least-squares problems in (11). After updating $\widehat{\mathbf{d}}$, every untrusted values in $\widehat{\mathbf{y}}_t$ is replaced with the corresponding entry in $\widehat{\mathbf{U}}^r_\Omega\widehat{\mathbf{d}}$.

To handle the general case that the observations contain untrusted entries, only trusted observations in $\mathbf{z}$ and $\mathbf{z}'$ are used in Step 3 and Step 5 to compute $\widehat{\mathbf{d}}$, respectively.

B. Distinguishing an event from consecutive bad data

One key component of our algorithm is to distinguish an event from bad data. When an event happens in the system at time t, the measurements usually change significantly such that some obtained measurements in $\mathbf{y}_t$ are much different from the corresponding predictions in $\widehat{\mathbf{y}}_t$ obtained in step 4. Then these measurements will be considered as untrusted data. Thus, one challenge for bad data detection is to distinguish a system event from bad data. The differentiation usually requires state-estimation-based bad data detectors or incorporating other prior information such as features of events.

Algorithm 1 Online robust bad and missing data estimation from Hankel matrix

Require: κ, L, approximation error threshold e_a, bad data threshold $\mathbf{s} \in \mathbb{R}^m$, threshold of the number of the affected channels n_s, threshold of the ratio of the approximation errors $\eta > 1$, counter C. (n_s, η, and C are used in Subroutine 2)

Initialization: obtain $\widehat{\mathbf{y}}_1 \sim \widehat{\mathbf{y}}_L$ from $\mathbf{y}_1 \sim \mathbf{y}_L$, set $C = 0$

For $t = L+1, L+2, \cdots$ **do**

1. Let $\Phi = \{t-L, \cdots, t-1\}$, conduct SVD on $\mathcal{H}_\kappa(\widehat{\mathbf{Y}}^\Phi) = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}^*$. Find the smallest r satisfying $e^r\left(\mathcal{H}_\kappa(\widehat{\mathbf{Y}}^\Phi)\right) \le e_a$, where e^r is defined in (2). Keep the first r dominant left singular vectors $\widehat{\mathbf{U}}^r \in \mathbb{C}^{\kappa m \times r}$.
2. Construct $\mathbf{z}$ as
$$\mathbf{z} = [\widehat{\mathbf{y}}^T_{t-\kappa+1}, \widehat{\mathbf{y}}^T_{t-\kappa+2}, \cdots, \widehat{\mathbf{y}}^T_{t-1}]^T \in \mathbb{C}^{(\kappa-1)m}. \quad (9)$$
$\Psi \subseteq \{1, ..., (\kappa-1)m\}$ is the index set of trusted entries in $\mathbf{z}$.
3. Compute $\widehat{\mathbf{d}}$ as
$$\widehat{\mathbf{d}} = \arg\min_{\mathbf{d}\in\mathbb{C}^r} \|\mathbf{z}_\Psi - \widehat{\mathbf{U}}^r_\Psi\mathbf{d}\| = ((\widehat{\mathbf{U}}^r_\Psi)^*\widehat{\mathbf{U}}^r_\Psi)^{-1}(\widehat{\mathbf{U}}^r_\Psi)^*\mathbf{z}_\Psi, \quad (10)$$
where $\widehat{\mathbf{U}}^r_\Psi \in \mathbb{C}^{|\Psi|\times r}$ and $\mathbf{z}_\Psi \in \mathbb{C}^{|\Psi|}$.
4. Compute $\mathbf{h} = \widehat{\mathbf{U}}^r_\Omega\widehat{\mathbf{d}}$, where the index set $\Omega = \{(\kappa-1)m+1, \cdots, \kappa m\}$. For every channel i where an observation $\mathbf{y}_t(i)$ is available, if $|\mathbf{y}_t(i) - \mathbf{h}(i)| \le \mathbf{s}(i)$, set $\widehat{\mathbf{y}}_t(i) = \mathbf{y}_t(i)$, and label $\mathbf{y}_t(i)$ and $\widehat{\mathbf{y}}_t(i)$ as trusted.
5. If not all entries in $\mathbf{y}_t$ are trusted, then update $\widehat{\mathbf{d}}$ as
$$\widehat{\mathbf{d}} = ((\widehat{\mathbf{U}}^r_{\Psi'})^*\widehat{\mathbf{U}}^r_{\Psi'})^{-1}(\widehat{\mathbf{U}}^r_{\Psi'})^*\mathbf{z}'_{\Psi'}, \quad (11)$$
where
$$\mathbf{z}' = [\widehat{\mathbf{y}}^T_{t-\kappa+1}, \widehat{\mathbf{y}}^T_{t-\kappa+2}, \cdots, \widehat{\mathbf{y}}^T_t]^T \in \mathbb{C}^{\kappa m}, \quad (12)$$
$\Psi' \subseteq \{1, ..., \kappa m\}$ is the index set of trusted entries in $\mathbf{z}'$. Compute $\mathbf{h}' = \widehat{\mathbf{U}}^r_\Omega\widehat{\mathbf{d}}$. For every i that $\mathbf{y}_t(i)$ is untrusted, set $\widehat{\mathbf{y}}_t(i)$ as $\mathbf{h}'(i)$.
6. Employ subroutine 2 to determine whether $\mathbf{y}_t$ is event data.

End for

We propose a data-driven approach without prior information to differentiate an event from bad data. Subroutine 2 exploits the low-rank Hankel property discussed in Section II-C to identify an event. The event identification is the last step of our algorithm at each time instant. Note that existing bad data detection methods such as [12], [13] work only when few channels contain bad measurements. Subroutine 2 can further identify simultaneous and consecutive bad data.

Step 1 of subroutine 2 identifies bad data if untrusted measurements exist in only few channels at one instant. Γ_1 contains the indices of channels with untrusted data. If $|\Gamma_1|$ is less than n_s, where n_s is a predetermined threshold, it is not an event. Step 2 examines channels in Γ_1. Set Γ_2 records the indices of channels in Γ_1 of which the measurements from $t-L/2+1$ to t are all identified as untrusted. If there are less than n_s channels in Γ_2, then these untrusted measurements are determined as bad data. Otherwise, set the counter C as $L/2$, the algorithm further collects the data in the next

Subroutine 2 Event data differentiation from bad data

If $C = 0$:

1. Compare $\widehat{\mathbf{y}}_t \in \mathbb{C}^m$ with $\mathbf{y}_t$. Let $\Gamma_1 = \{i : \widehat{\mathbf{y}}_t(i) \neq \mathbf{y}_t(i)\}$. If $|\Gamma_1| < n_s$, then $\mathbf{y}_t$ is not event data, return; otherwise, proceed to step 2.
2. Let $\Gamma_2 = \emptyset$. For each channel $i \in \Gamma_1$, if the observations $\mathbf{y}_\tau(i), \tau = t-L/2+1, \cdots, t$, are all identified as untrusted data, then $\Gamma_2 = \Gamma_2 \cup \{i\}$.
   If $|\Gamma_2| \geq n_s$, set $C = L/2$; otherwise, return.

Elseif $C = 1$:

3. Reset $C = 0$. Let $\Phi = \{t-L+1, \cdots, t\}$, let $\mathbf{Z} \in \mathbb{C}^{|\Gamma_2| \times L}$ contain the rows of $\mathbf{Y}^\Phi$ with row indices in Γ_2. For the Hankel matrices $\mathcal{H}_\kappa(\mathbf{Z})$ and $\mathcal{H}_\kappa(\overline{\mathbf{Z}})$, if
$$e^1(\mathcal{H}_\kappa(\overline{\mathbf{Z}})) \geq \eta e^1(\mathcal{H}_\kappa(\mathbf{Z})), \tag{13}$$
   we identify the observations in $\mathbf{Z}$ as event data, declare the detection of an event, record the instant as t_d, update the recovered measurements $\widehat{\mathbf{y}}_\tau(i) = \mathbf{y}_\tau(i), \forall \tau \in \Phi, \forall i \in \Gamma_2$ and store them as trusted entries, return; otherwise, keep $\widehat{\mathbf{y}}_t$, return.

Else: $C \leftarrow C - 1$.

$L/2$ time steps and obtains time series $\mathbf{Y}^\Phi$ in a window size of L. Step 3 first constructs $\mathbf{Z} \in \mathbb{C}^{|\Gamma_2| \times L}$ with only the measurements from the channels labeled by Γ_2 of $\mathbf{Y}^\Phi$. $\overline{\mathbf{Z}}$ is a random column permutation of $\mathbf{Z}$. As discussed in Section II-C, if the data result from an event, the rank of $\mathcal{H}_\kappa(\mathbf{Z})$ would be much smaller than the rank of $\mathcal{H}_\kappa(\overline{\mathbf{Z}})$, because the temporal correlations are lost after the permutation. If the data result from consecutive bad data, the rank of these two Hankel matrices would not differ much. (13) exploits this property to identify an event. If an event is identified, all the measurements of these channels in this window are labeled as trusted data.

In order to obtain $\widehat{\mathbf{y}}_t$, the memory requirement of the algorithm is $\mathcal{O}(\kappa m L)$, and the computational complexity is $\mathcal{O}(\kappa m L \cdot \min(\kappa m, L))$. Both are determined by L and κ. When L is less than κm, the computational complexity is $\mathcal{O}(\kappa m L^2)$.

C. Parameter selection of the algorithm

Parameters L and κ are related since the number of columns $L-\kappa+1$ in the Hankel matrix should be positive from Fig. 4. Moreover, to reduce the impact of the noise on subspace basis estimation, $L - \kappa + 1$ should not be too small. A relatively small L is suggested because it reduces the computational complexity. Moreover, if L is too large, the long time window may contain some obsolete modes, degrading the estimation accuracy of the subspace. Theoretical analysis in Section IV shows that the estimation accuracy can be improved with more trusted entries in steps 3 and 5 of Algorithm 1. Thus we prefer a large κ. In practice, it is reasonable to set $L \approx 2\kappa$.

Since the matrix is not exactly low-rank due to the noise, e_a affects the approximated rank of the Hankel matrix. We think it is reasonable to set e_a around $1\% \sim 5\%$.

When a disturbance happens, the measurements change significantly and deviate much from the prediction. The algorithm gradually updates the subspace and thus improves the prediction accuracy over time. The difference between the actual data and the predictions reduces gradually. Thus we suggest $\mathbf{s}$ as a decaying function of the interval between the current instant and the instant of the latest disturbance. Moreover, $\mathbf{s}$ can be different for different channels as some channels may be noisier than others. With a high noise level, we can set a larger $\mathbf{s}$ to reduce the false alarm rate.

The parameter n_s can be set as a fixed ratio (which is less than one) multiplying the total number of PMU channels. A disturbance may affect only a few buses in a weakly connected system, while it could impact the whole region for a strongly connected area. Thus, we suggest using a small ratio for a weakly connected system and a large ratio for a strongly connected system.

The parameter η is related to the measurement type (phasor, frequency, etc), L, κ, and the noise level. In the case of a high noise level, we could decrease η slightly to maintain the detection accuracy.

In practice, the proper values of $\mathbf{s}$, n_s and η could be learned from historical data, with the aim to strike a balance between the false alarm rate and the detection accuracy.

D. Other comments

If one only needs to estimate missing data (no bad data exist), Algorithm 1 can be simplified as follows.

- The algorithm is triggered only when the measurement $\mathbf{y}_t$ at time t contains missing points.
- Steps 2-4 and step 6 are unnecessary.
- In step 5, Ψ' is the index set of the observed entries in $\mathbf{y}_{t-\kappa+1} \sim \mathbf{y}_t$. One estimates the missing points in $\mathbf{y}_t$ by the corresponding entries in $\mathbf{h}'$.

The simplified version is an extension of algorithm OLAP [6] to $\mathcal{H}_\kappa(\widehat{\mathbf{Y}}^\Phi)$. OLAP only uses the observed entries in $\mathbf{y}_t$ to estimate the missing points, while Algorithm 1 uses all the observed entries in $\mathbf{y}_{t-\kappa+1} \sim \mathbf{y}_t$. Intuitively, the estimation accuracy is improved by exploiting more observations. Theorem 2 in Section V justifies this intuition analytically.

We remark that since the information of the system model is not required, our method cannot detect scaling errors. Due to problems in the instruments such as calibration errors, some measurement channels may continuously produce bad data, and the bad measurements differ from the true data by a constant scaling factor. Since the low-rank Hankel structure still holds in this case, our method cannot detect these errors. The detection of these scaling errors requires system model and topology information. One can identify these errors by first computing system states and then checking the relations among these states. Our method can be employed as a pre-processing step to fill in missing data and correct some forms of bad data. It can be followed by conventional bad data detectors and state estimators to correct scaling errors [11].

IV. Theoretical analysis of the algorithm

This section presents the theoretical analysis of our algorithm. We first show that the prediction $\mathbf{h}$ in step 4 of Algorithm 1 is entry-wise close to the actual data (Theorem 1). We then show that after removing the bad data, our estimation of the missing points in step 5 is sufficiently close to the ground-truth data (Theorem 2). We lastly compare using

Hankel matrix ($\kappa > 1$) with using the original data matrix ($\kappa = 1$) in data estimation and show the former performs better than the latter in (20). The proofs are deferred to Appendix.

To simplify the presentation of the theoretical analysis, we first introduce the notations used in this section. The notations are defined in a general form to analyze various steps in Algorithm 1. We drop the superscript r for the rank, and let $\mathbf{U} \in \mathbb{C}^{q\times r}$ denote the ground-truth orthonormal column basis of the r-dimensional subspace in $\mathbb{C}^q$. Let $\widehat{\mathbf{U}} \in \mathbb{C}^{q\times r}$ denote the orthonormal column basis estimated from historical measurements. $\widehat{\mathbf{U}}$ may be different from $\mathbf{U}$. The measurements $\mathbf{y} \in \mathbb{C}^q$ in the span of $\mathbf{U}$ can be written as $\mathbf{y} = \mathbf{U}\mathbf{d}+\mathbf{w}$, where $\mathbf{d} \in \mathbb{C}^r$ contains the coefficients, and $\mathbf{w} \in \mathbb{C}^q$ denotes the measurement noise. When partial observations are available, let set Ψ contain the indices of available entries in $\mathbf{y}$, and the available observations are represented by $\mathbf{y}_\Psi \in \mathbb{C}^{|\Psi|}$. Then the remaining part of $\mathbf{y}$ can be estimated from

$$\widehat{\mathbf{d}} = \underset{\mathbf{d}\in\mathbb{C}^r}{\arg\min} \|\mathbf{y}_\Psi - \widehat{\mathbf{U}}_\Psi \mathbf{d}\|_2 = (\widehat{\mathbf{U}}_\Psi^* \widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^* \mathbf{y}_\Psi, \quad (14)$$

$$\widehat{\mathbf{y}} = \widehat{\mathbf{U}}\widehat{\mathbf{d}} = \widehat{\mathbf{U}}(\widehat{\mathbf{U}}_\Psi^* \widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^* \mathbf{y}_\Psi. \quad (15)$$

In Algorithm 1, $q = \kappa m$. $\widehat{\mathbf{U}}$ corresponds to the column basis of $\mathcal{H}_\kappa(\widehat{\mathbf{Y}}_\Phi)$ shown in Fig. 4, i.e., $\widehat{\mathbf{U}}^r$ obtained in Step 1. If we let $\mathbf{y}$ and Ψ in (14) and (15) represent $\mathbf{z}$ in (9) and Ψ in Step 2, respectively, then (14) models (10) in Step 3. The last m entries of $\widehat{\mathbf{y}}$ in (15) correspond to $\mathbf{h}$ in Step 4. If we let $\mathbf{y}$ and Ψ in (14) and (15) represent $\mathbf{z}'$ in (12) and Ψ' in Step 5, respectively, then (14) models (11) in Step 5. The last m entries of $\widehat{\mathbf{y}}$ in (15) correspond to $\mathbf{h}'$ in Step 5.

Before presenting our theoretical analysis, we first introduce the conventional definition of subspace incoherence and vector incoherence that are used in the study of data recovery [19].

Definition 1. *Let $\mathbf{W} \in \mathbb{C}^{q\times r}$ denote one basis of an r-dimensional subspace $S(\mathbf{W})$, then the incoherence of $S(\mathbf{W})$ is defined as*

$$\mu(S(\mathbf{W})) = \mu(\mathbf{W}) = \max_{i\in\{1,\cdots,q\}} \frac{q}{r} \left\|\mathbf{e}_i^* \mathbf{W}(\mathbf{W}^*\mathbf{W})^{-1}\mathbf{W}^*\mathbf{e}_i\right\|,$$

where $\mathbf{e}_i \in \mathbb{C}^q$ has its i-th entry as one and others as zeros.

For the special case of a vector $\mathbf{w} \in \mathbb{C}^q$, let $\mu(\mathbf{w})$ denote the incoherence of the 1-dimensional subspace spanned by $\mathbf{w}$, then we have $\mu(\mathbf{w}) = q\|\mathbf{w}\|_\infty^2/\|\mathbf{w}\|^2$. It is established in the literature that if the column and row subspaces of a low-rank matrix have a small incoherence, one can recover the matrix from a small number of observations. Intuitively, if a row of the matrix is not related with other rows at all, then one cannot use other rows to fill in the missing entries of this row and needs to observe almost all the entries in this row to recover the matrix. This bad scenario corresponds to the property of a large incoherence.

For example, consider the $n \times n$ rank-2 matrix M

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & \cdots & 2 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}. \quad (16)$$

M has one entry M_{1n} being 2 and all other entries being 1. One needs to sample M_{1n} directly to be able to correctly recover M. If not, the low-rank estimation from partial samples would be a rank-1 all-one matrix. Without any prior information about where the value 2 is, one essentially needs to sample almost all entries. The basis vectors of the column subspace of M are $[1, 0, \cdots, 0]^T$ and $[0, 1, \cdots, 1]^T$. Following Definition 1, we have the incoherence $\mu = n/2$, which is very large. Please see [19] for more details about the incoherence.

Let $\sigma_{\min}(\widehat{\mathbf{U}}^*\mathbf{U})$ denote the smallest singular value of $\widehat{\mathbf{U}}^*\mathbf{U}$. If the estimated subspace is very close to the ground truth, then $\sigma_{\min}(\widehat{\mathbf{U}}^*\mathbf{U})$ is close to one. We characterize the bad data detection performance through the following theorem.

Theorem 1. *Let $\delta > 0$, assume $\sigma_{\min}(\widehat{\mathbf{U}}^*\mathbf{U}) \geq \sqrt{1-\epsilon^2}$ for some $\epsilon \in (0, 1)$, and $|\Psi| \geq \frac{8}{3} r\mu(\widehat{\mathbf{U}}) \log(\frac{2r}{\delta})$, then with a probability at least $1-\delta$,*

$$\|\widehat{\mathbf{y}} - \mathbf{y}\|_\infty \leq \left(\sqrt{\frac{q}{1-\gamma}} + 1\right)\left(\epsilon\sqrt{r}\|\mathbf{d}\|_\infty + \|\mathbf{w}\|_\infty\right), \quad (17)$$

where $\gamma = \sqrt{\frac{8r\mu(\widehat{\mathbf{U}})}{3|\Psi|}\log(\frac{2r}{\delta})}$.

Theorem 1 indicates that if the subspace estimation is accurate (ϵ is small), and there are enough trusted observations in time $t-L$ to $t-1$, then the difference between $\mathbf{h}(i)$ and the noisy measurement of channel i at time t is upper bounded by the right-hand side of (17), which is small if ϵ is small, and the noise level is low. If $\mathbf{y}_t(i)$ is corrupted and significantly different from the ground-truth data, then $|\mathbf{h}(i) - \mathbf{y}_t(i)|$ is large, and the bad data can be detected in step 4 for a properly selected threshold $\mathbf{s}(i)$. The required number of trusted measurements $|\Psi|$ is only $\frac{8}{3} r\mu(\widehat{\mathbf{U}}) \log(\frac{2r}{\delta})$, which is much less than the dimension q of $\mathbf{y}$. If the subspace has a small incoherence and $\widehat{\mathbf{U}}$ is close to $\mathbf{U}$, then our method can tolerate a significant amount of missing data.

We next show that after deleting bad data, the recovered data in step 5 is close to the ground truth.

Theorem 2. *Let $\delta > 0$, with the same assumptions as Theorem 1, then with a probability at least $1-2\delta$, $\|\widehat{\mathbf{y}} - \mathbf{U}\mathbf{d}\|$ is no greater than*

$$E := \frac{1+\beta_{\mathbf{w}}}{1-\gamma}\sqrt{\frac{r\mu(\widehat{\mathbf{U}})}{|\Psi|}}\|\mathbf{w}\| + \left(1 + \frac{1+\beta_{\mathbf{u}^\perp}}{1-\gamma}\sqrt{\frac{r\mu(\widehat{\mathbf{U}})}{|\Psi|}}\right)\epsilon\|\mathbf{d}\|, \quad (18)$$

where $\beta_{\mathbf{w}} = \sqrt{2\mu(\mathbf{w})\log(\frac{1}{\delta})}$, $\gamma = \sqrt{\frac{8r\mu(\widehat{\mathbf{U}})}{3|\Psi|}\log(\frac{2r}{\delta})}$, $\mathbf{u}^\perp = (\mathbf{I} - \widehat{\mathbf{U}}\widehat{\mathbf{U}}^*)\mathbf{U}\mathbf{d}$, $\beta_{\mathbf{u}^\perp} = \sqrt{2\mu(\mathbf{u}^\perp)\log(\frac{1}{\delta})}$.

$\beta_{\mathbf{w}}$ is usually upper bounded by a constant if the noise levels in all channels are similar. If the number of trusted observations $|\Psi|$ is large, then γ is small, $\sqrt{\frac{r\mu(\widehat{\mathbf{U}})}{|\Psi|}}$ is small, and the first term of (18) is small. The second term of (18) is small if ϵ is small, i.e., the subspace estimation is accurate.

We now show that the data recovery accuracy is improved using the Hankel structure compared with using the original data matrix. To simplify the analysis, we assume that the subspace estimation is accurate, i.e., $\widehat{\mathbf{U}} = \mathbf{U}$, and $\epsilon = 0$. We use $\mathbf{U}_\kappa$ to denote the actual subspace of the Hankel matrix in Fig. 4, and $\mathbf{U}_1$ corresponds to the case that $\kappa = 1$. The following lemma is important for our analysis.

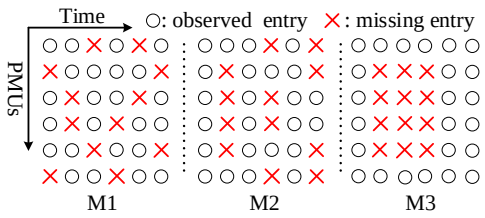

Fig. 5: Three modes of missing data

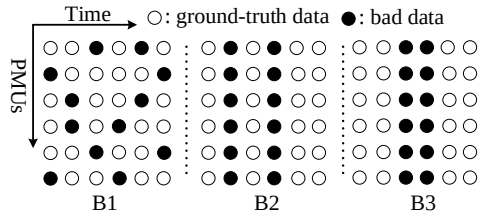

Fig. 6: Three modes of bad data

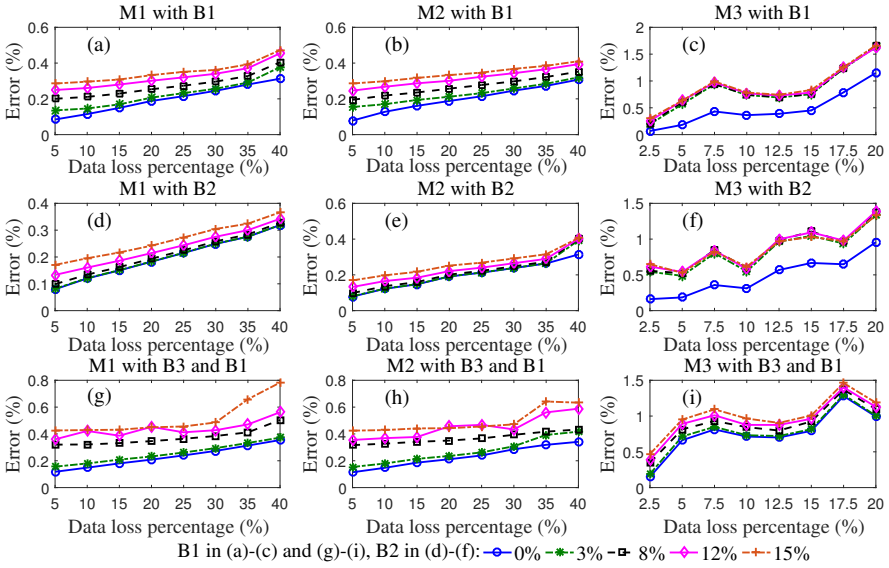

Fig. 7: Estimation errors of three missing data modes and three bad data modes

Lemma 1. *If the measurements follow the model in* (6)*, then*

$$\mu(\mathbf{U}_\kappa) < \kappa\mu(\mathbf{U}_1),\ \forall \kappa > 1. \tag{19}$$

If the missing data and bad data happen randomly across time and channels, $|\Psi_\kappa| = \kappa|\Psi_1|$ holds approximately, where Ψ_κ is the set of trusted entries in $\mathbf{z}'$ in (12) when the number Hankel block is κ. Moreover, if the noise distribution does not change much, $\mu(\mathbf{w}_\kappa) = \mu(\mathbf{w}_1)$ holds approximately, where $\mathbf{w}_\kappa \in \mathbb{C}^{\kappa m}$ contains the noise in κ time steps. Moreover, $\|\mathbf{w}_\kappa\| \approx \sqrt{\kappa}\|\mathbf{w}_1\|$. Let E_κ denote the error bound in (18) when the number of Hankel block is κ. Then combining the above discussion, (18), and Lemma 1, we have

$$E_\kappa/\sqrt{\kappa} < E_1. \tag{20}$$

Since $\mathbf{y}_t$ is the last m entries of $\widehat{\mathbf{y}}$ in $\mathbb{C}^{\kappa m}$, and Theorem 2 provides an upper bound E_κ of the ℓ_2-norm of the error vector $\widehat{\mathbf{y}} - \mathbf{U}\mathbf{d}$, (20) indicates that recovery error bound is smaller if $\kappa > 1$. Thus, one can enhance the recovery accuracy using the Hankel structure.

V. Numerical Experiments

To test the accuracy of Algorithm 1 in Section III, numerical experiments are conducted on practical recorded synchrophasor datasets, where a phasor is represented by its rectangular form. The simulation is implemented on a desktop with 3.4GHz Intel Core i7 and 16 GB memory. Some data points are deleted to simulate missing data. Bad data are also added to the datasets. To study the performance under extreme conditions (like simultaneous and consecutive data losses), three modes of data losses are used:

- M1: Data losses occur randomly and independently across PMU channels and time instants.
- M2: Data losses occur at random time instants independently. When data losses occur, randomly selected k out of m channels lose data.
- M3: Data losses occur on fixed k PMU channels simultaneously and consecutively.

The three missing data modes are illustrated in Fig. 5, where $m = 6$ and $k = 4$ in the diagram. We also simulate three modes of bad data, of which a diagram is shown in Fig. 6:

- B1: Bad data occur randomly and independently across channels and time instants. The additive errors in voltage magnitudes and angles follow Gaussian $\mathcal{N}(0, 0.05^2)$ and $\mathcal{N}(0, 10^2)$, respectively.
- B2: Bad data occur in all channels simultaneously at random time instants. The additive errors are independent across time but are the same for all channels at one time instant. The errors are generated from $\mathcal{N}(0, 0.05^2)$ for magnitudes and $\mathcal{N}(0, 10^2)$ for angles.
- B3: Bad data occur in all channels simultaneously and consecutively for some time. The additive errors are independent across time but are the same for all channels at one instant. The errors are drawn from $\mathcal{N}(-0.1, 0.05^2)$ for magnitudes and $\mathcal{N}(-20, 10^2)$ for angles.

Note that Gaussian distributions are used to generate significant additive errors as bad data, while the algorithm makes no assumption on the distribution of bad data. We choose the mean of -0.1 and -20 in mode B3 to simulate significant measurement changes, which could also result from an event. We will show our method can distinguish bad data in mode B3 from system events. The estimation error is defined as $\|\widehat{\mathbf{Y}} - \mathbf{Y}\|_F/\|\mathbf{Y}\|_F$, where $\mathbf{Y}$ is the ground-truth data. Every result is averaged over 50 independent runs.

As pointed out in [20], the data noises produced by PMUs are random and their statistics depend on the specific signal processing methods used in various PMUs. To further evaluate the robustness of our method against the measurement noise, we use three different datasets recorded at different locations. The datasets are provided by New York Power Authority (NYPA), New York ISO (NYISO), and ISO New England (ISO-NE), respectively. The parameters shared in common by the simulations are $L = 10$, $\kappa = 6$, $e_a = 0.02$, $\eta = 2.4$.

A. Voltage phasors of a generator trip event by NYPA

The measured voltage phasors are shown in Fig. 1, where $m = 11$, k is set as 5 in missing data modes M2 and M3. Since

the system is strongly connected, n_s is set as $m \times 60\% \approx 7$.

To identify bad data, we set a threshold for the magnitude change and a threshold for the angle change. Since the change of the voltage magnitude is usually minor, the threshold is set as $0.0015f(t)$, where $f(t) = \max(2, 30e^{-3(t-t_d)/5})$, t_d is the instant of the latest detected disturbance in subroutine 2. Thus $f(t)$ starts with 30 at t_d and drops to 2 after about 5 seconds. For the phasor at channel i at time t, the threshold of the angle change is estimated from the first L measurements and set as $f(t) \cdot \sum_{\tau=2}^{L} \frac{|\theta(i,\tau)-\theta(i,\tau-1)|}{L-1}, \forall i \in \{1, \cdots, m\}$, where $\theta(i,\tau)$ is the angle of phasor i at time τ.

Fig. 7 shows the estimation errors of different modes and percentages of missing data and bad data. We first study random bad data with different patterns of missing data, as shown in Figs. 7(a)-7(c). The estimation error generally increases when the percentage of missing data or bad data increases. In mode M3, the error is mostly dominated by simultaneous and consecutive data losses and does not change much when the number of bad data increases. Fig. 8 shows the observed data and the estimated data of one case with 8% bad data in mode B1 and 40% missing data in mode M1.

Figs. 7(d)-7(f) show the estimation errors of bad data in mode B2 with different missing data patterns. Fig. 9 illustrates one example of 8% bad data in B2. The estimated data are similar to the bottom figures in Fig. 8, so we skip the details.

We next test a mixture of simultaneous and consecutive bad data in mode B3 and random bad data in mode B1. For mode B3, additive bad data are injected to all channels during data columns of $210 \sim 219, 420 \sim 429, 520 \sim 529$, and the length of duration is 10 instants. Fig. 10(a) shows one example of the observations with bad data in mode B3, 3% bad data in mode B1, and 40% missing data in mode M2. The injected bad data in mode B3 are highlighted in red circles. Fig. 10(b) shows the estimated data. One can see that bad data in mode B3 are correctly identified, while the actual system event data around 2.3 seconds remain unaffected. Thus, our method can differentiate simultaneous and consecutive bad data from event data accurately. Figs. 7(g)-7(i) show the estimation errors when we maintain the same bad data in mode B3 and increase the percentages of bad data in mode B1 and missing data in different modes.

We also vary parameters κ and L in our algorithm and test the estimation performance. The results are shown in Fig. 11, where we generate 8% bad data in mode B1 and vary the missing data in mode M1. We first fix the window length $L = 10$ and change κ from 4 to 7. A larger κ helps improve the estimation accuracy slightly, since more trusted entries are employed to obtain $\widehat{\mathbf{y}}_t$. We then fix $\kappa = 6$ and change L from 10 to 40. A smaller L improves the estimation accuracy. The results are consistent with our analysis in Section III-C.

To obtain $\widehat{\mathbf{y}}_t$ from the partially observed $\mathbf{y}_t$ with corruptions, the average running time of our experiments is 1.5 ms, far less than 1/30 s. Thus our algorithm can be implemented as the preprocessing to improve the quality of streaming synchrophasor measurements in real time.

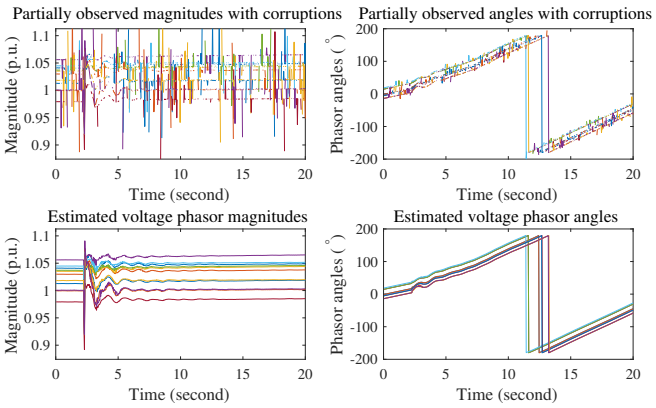

Fig. 8: The observed and estimated voltage phasors with 8% bad data in mode B1 and 40% missing entries in mode M1

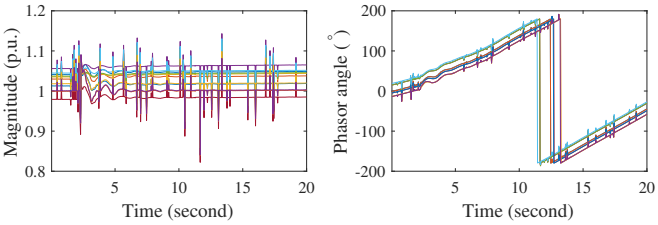

Fig. 9: Voltage phasors with 8% bad data in mode B2

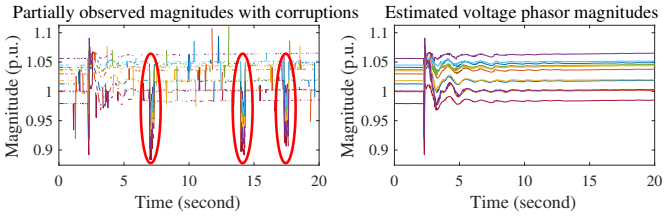

Fig. 10: The observed and estimated voltage phasor magnitudes with bad data in mode B3 (plus 3% bad data in mode B1 and 40% missing data in mode M2)

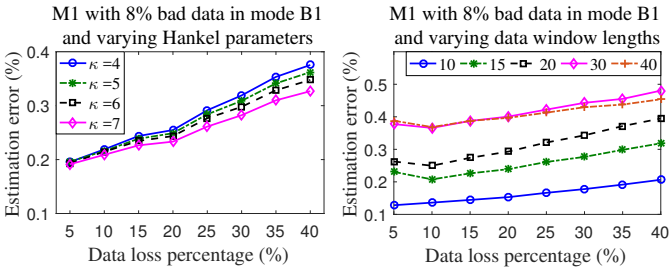

Fig. 11: The estimation errors with varying parameters

B. Current phasors of a disturbance by ISO-NE

There are 124 current phasors in the dataset with 30 samples per second. With the unit value of each channel estimated from the first L measurements, Fig. 12 shows the magnitudes of 10 phasors in the dataset. There is one event occurring around 10 seconds, and the number of affected channels is around 15.

We still test the modes of missing and bad data M1$\sim$ M3 and B1 $\sim$ B3 with $m = 124$ and $k = 62$ in M2 and M3. Since the system is not strongly connected, $n_s = m \times 10\% \approx 12$. Due to the distinct noise levels of different channels, for the phasor of channel i at instant t, the threshold of magnitude change is $f(t) \cdot \max(\sigma_i, 0.008)$, where σ_i denotes the noise

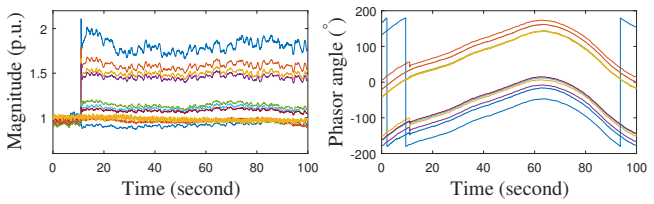

Fig. 12: Current phasors of 10 channels of ISO-NE dataset

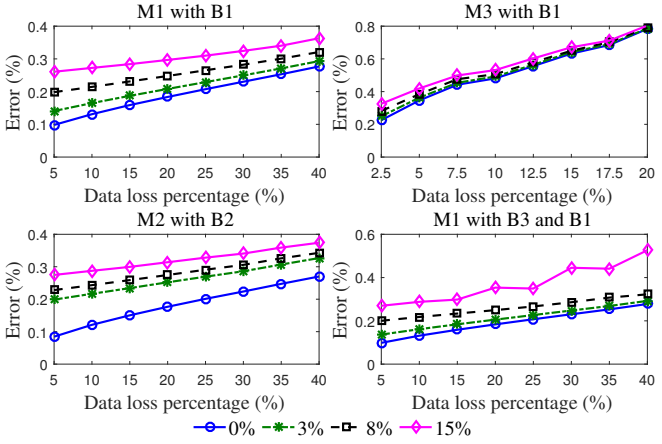

Fig. 13: Estimation errors of ISO-NE dataset

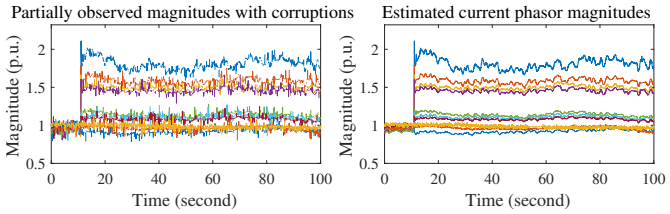

Fig. 14: The observed and estimated current phasor magnitudes with 8% bad data in B1 and 20% missing data in M1

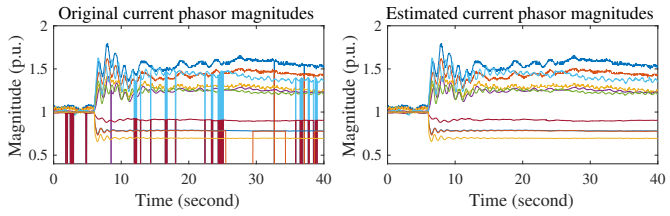

Fig. 15: The actual phasor magnitudes and corresponding estimated magnitudes of 10 channels of NYISO dataset

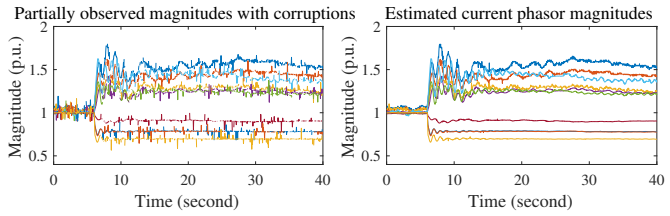

Fig. 16: The observed and estimated current phasor magnitudes with 8% bad data in B1 and 15% missing data in M1

level of channel i, and we estimate it as the standard deviation of the first L measurements. $f(t) = \min(8, 24e^{-3(t-t_d^i)/5})$, where t_d^i denotes the instant of the latest disturbance affecting channel i. Thus $f(t)$ is always 8, if one channel is not affected by events. The threshold of angle change is still estimated from the first L measurements as $f(t) \cdot \sum_{\tau=2}^{L} \frac{|\theta(i,\tau)-\theta(i,\tau-1)|}{L-1}$. For mode B3, additive bad data are injected to all channels during data columns of $1500 \sim 1509,\ 2500 \sim 2509$.

Part of the results is shown in Fig. 13. The estimation errors are similar to the results in Fig. 7. Fig. 14 shows one example of the estimated data with 8% bad data in B1 and 20% missing data in M1, where the estimation error is 0.4%.

C. Current phasors of a generator trip event by NYISO

There are 44 current phasors with 30 samples per second in the dataset. Part of the measurements is missing. Fig. 15(a) shows the original magnitudes of 10 phasors. The event occurring at around 6 seconds affects all the channels. The estimated data from the incomplete dataset is shown in Fig. 15(b). With the incomplete dataset in Fig. 15(a), we further inject 8% additive errors following mode B1, and erase 15% measurements randomly to test the performance of our algorithm. With n_s as $m \times 60\% \approx 26$, due to the strong connectivity of the system, and other parameters same as presented in Section V-B, Fig. 16(a) shows one example of the magnitudes of partially observed phasors with corruptions. Fig. 16(b) shows the magnitudes of the estimated data. The estimation error is 0.7%. We can find the missing data are filled in and almost all the bad data are corrected.

All the three datasets contain an event. Although the measurements change significantly when an event starts, as shown in Figs. 1, 12, and 15, our method does not mistake them for bad data. Our method can correctly detect events and only correct bad data.

VI. Conclusion and Discussions

An online algorithm is developed in this paper to identify and correct bad data, as well as to estimate missing data of streaming synchrophasor measurements. The central idea is to exploit the low-rank property of the Hankel matrix of the measurements. The advantage of this method is that it can detect and correct bad data even when they exist bad data on all channels and for consecutive instants, in contrast to existing methods that can only detect sparse bad measurements. The algorithm is evaluated both theoretically and numerically on synchrophasor datasets recording different events.

Acknowledgment

This research is supported by NSF Grant #1508875, EPRI #1007316, and the ERC Program of NSF and DoE under the supplement to NSF Award EEC-1041877 and the CURENT Industry Partnership Program. We thank NYPA, NYISO and ISO-NE for providing recorded synchrophasor datasets.

References

[1] A. Silverstein and J. E. Dagle, "Successes and challenges for synchrophasor technology: An update from the north american synchrophasor initiative," in *2012 45th Hawaii Int. Conf. System Science*, 2012, pp. 2091–2095.

[2] California ISO, "Five year synchrophasor plan," Folsom, CA, Tech. Rep., 2011.

[3] J. De La Ree, V. Centeno, J. S. Thorp, and A. G. Phadke, "Synchronized phasor measurement applications in power systems," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 20–27, 2010.

[4] L. Xie, Y. Chen, and P. R. Kumar, "Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2784–2794, 2014.

[5] N. Dahal, R. L. King, and V. Madani, "Online dimension reduction of synchrophasor data," in *Proc. IEEE PES Transmission and Distribution Conf. Expo.*, 2012, pp. 1–7.
[6] P. Gao, M. Wang, S. G. Ghiocel, J. H. Chow, B. Fardanesh, and G. Stefopoulos, "Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1006–1013, 2016.
[7] P. Gao, M. Wang, J. H. Chow, M. Berger, and L. M. Seversky, "Missing data recovery for high-dimensional signals with nonlinear low-dimensional structures," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5421–5436, 2017.
[8] P. Gao, M. Wang, J. H. Chow, S. G. Ghiocel, B. Fardanesh, G. Stefopoulos, and M. P. Razanousky, "Identification of successive "unobservable" cyber data attacks in power systems." *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5557–5570, 2016.
[9] L. Balzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," in *Proc. Allerton Conf. Communication, Control, and Computing*, 2010, pp. 704–711.
[10] Y. Chi, Y. C. Eldar, and R. Calderbank, "PETRELS: parallel subspace estimation and tracking by recursive least squares from partial observations," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5947–5959, 2013.
[11] S. G. Ghiocel, J. H. Chow, G. Stefopoulos, B. Fardanesh, D. Maragal, B. Blanchard, M. Razanousky, and D. B. Bertagnolli, "Phasor-measurement-based state estimation for synchrophasor data quality improvement and power transfer interface monitoring," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 881–888, 2014.
[12] M. Wu and L. Xie, "Online detection of low-quality synchrophasor measurements: A data-driven approach," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2817–2827, 2017.
[13] M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low rank," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 50–66, 2013.
[14] S. Zhang, Y. Hao, M. Wang, and J. H. Chow, "Multi-channel missing data recovery by exploiting the low-rank hankel structures," in *Proc. Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process. (CAMSAP)*, 2017, pp. 1–5.
[15] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 946–977, 2013.
[16] P. Kundur, *Power system stability and control*. New York: McGraw-Hill, 1994.
[17] P. Netrapalli, U. Niranjan, S. Sanghavi, A. Anandkumar, and P. Jain, "Non-convex robust pca," in *Adv. Neural Inf. Process. Syst.*, 2014, pp. 1107–1115.
[18] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Adv. Neural Inf. Process. Syst.*, 2009, pp. 2080–2088.
[19] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
[20] Y. Seyedi, H. Karimi, and J. M. Guerrero, "Centralized disturbance detection in smart microgrids with noisy and intermittent synchrophasor data," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2775–2783, 2017.
[21] L. Balzano, B. Recht, and R. Nowak, "High-dimensional matched subspace detection when data are missing," in *Proc. Int. Symp. Inf. Theory*, 2010, pp. 1638–1642.

APPENDIX

A. Proof of Theorem 1

The proof of theorem 1 relies on the following lemma.

Lemma 2 ([21], Lemma 3)**.** *With the same assumptions as Theorem 1,*

$$\|(\mathbf{U}_\Psi^*\mathbf{U}_\Psi)^{-1}\| \le \frac{q}{(1-\gamma)|\Psi|}$$

holds with probability at least $1-\delta$.

Proof of Theorem 1. For any matrix $\mathbf{A}\in\mathbb{C}^{q\times r}$, we have

$$\|\mathbf{A}\|_\infty = \max_{\mathbf{x}\neq\mathbf{0}}\frac{\|\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} \le \max_{\mathbf{x}\neq\mathbf{0}}\frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|/\sqrt{r}} = \sqrt{r}\|\mathbf{A}\|.$$

Let $\mathbf{U}\mathbf{d} = \widehat{\mathbf{U}}\widehat{\mathbf{U}}^*\mathbf{U}\mathbf{d} + \mathbf{u}^\perp$, where $\mathbf{u}^\perp = (\mathbf{I}-\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*)\mathbf{U}\mathbf{d}$, then

$$\begin{aligned}
&\widehat{\mathbf{U}}\widehat{\mathbf{d}} - \mathbf{U}\mathbf{d} = \widehat{\mathbf{U}}(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^*\mathbf{y}_\Psi - \mathbf{U}\mathbf{d}\\
=&\widehat{\mathbf{U}}(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^*(\widehat{\mathbf{U}}_\Psi\widehat{\mathbf{U}}^*\mathbf{U}\mathbf{d} + \mathbf{u}_\Psi^\perp + \mathbf{w}_\Psi) - (\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*\mathbf{U}\mathbf{d} + \mathbf{u}^\perp)\\
=&\widehat{\mathbf{U}}(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^*(\mathbf{u}_\Psi^\perp + \mathbf{w}_\Psi) - \mathbf{u}^\perp. \qquad (21)
\end{aligned}$$

Then, following (21), we have

$$\begin{aligned}
&\|\widehat{\mathbf{y}} - \mathbf{y}\|_\infty = \|\widehat{\mathbf{U}}\widehat{\mathbf{d}} - (\mathbf{U}\mathbf{d}+\mathbf{w})\|_\infty\\
\le&\sqrt{|\Psi|}\|\widehat{\mathbf{U}}(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^*\|(\|\mathbf{u}_\Psi^\perp\|_\infty + \|\mathbf{w}_\Psi\|_\infty)\\
&+\|\mathbf{u}^\perp\|_\infty + \|\mathbf{w}\|_\infty\\
\le&(\sqrt{|\Psi|}\|(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^*\| + 1)(\|\mathbf{u}^\perp\|_\infty + \|\mathbf{w}\|_\infty),
\end{aligned}$$

where we use the fact that $\|\widehat{\mathbf{U}}\| = 1$ in deriving the last inequality. To bound $\|(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^*\|$, we have

$$\begin{aligned}
&\|(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^*\|^2\\
=&\max_{\|\mathbf{x}\|=1}\mathbf{x}^*\left((\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^*\right)^*(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^*\mathbf{x}\\
=&\max_{\|\mathbf{x}\|=1}\mathbf{x}^*\widehat{\mathbf{U}}_\Psi(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1/2}(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1/2}\widehat{\mathbf{U}}_\Psi^*\mathbf{x}\\
\le&\|(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\|\max_{\|\mathbf{x}\|=1}\mathbf{x}^*\widehat{\mathbf{U}}_\Psi(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^*\mathbf{x}\\
=&\|(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\| \le \frac{q}{(1-\gamma)|\Psi|}.
\end{aligned}$$

The last equality holds, since $\widehat{\mathbf{U}}_\Psi(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\widehat{\mathbf{U}}_\Psi^*$ is a projection matrix, and its spectral norm is 1. The last inequality follows from Lemma 2.

Since $\mathbf{u}^\perp = (\mathbf{I}-\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*)\mathbf{U}\mathbf{d}$, we have

$$\begin{aligned}
&\|(\mathbf{I}-\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*)\mathbf{U}\|^2\\
=&\max_{\|\mathbf{x}\|_2=1}\mathbf{x}^*(\mathbf{U}-\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*\mathbf{U})^*(\mathbf{U}-\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*\mathbf{U})\mathbf{x}\\
=&\max_{\|\mathbf{x}\|_2=1}\mathbf{x}^*\mathbf{x} - \mathbf{x}^*\mathbf{U}^*\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*\mathbf{U}\mathbf{x}\\
=&1-\min_{\|\mathbf{x}\|_2=1}\mathbf{x}^*\mathbf{U}^*\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*\mathbf{U}\mathbf{x}\\
=&1-\sigma_{\min}^2(\widehat{\mathbf{U}}^*\mathbf{U}) \le \epsilon^2.
\end{aligned}$$

$$\begin{aligned}
\|\mathbf{u}^\perp\|_\infty &= \|(\mathbf{I}-\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*)\mathbf{U}\mathbf{d}\|_\infty\\
&\le \sqrt{r}\|(\mathbf{I}-\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*)\mathbf{U}\|\|\mathbf{d}\|_\infty \le \epsilon\sqrt{r}\|\mathbf{d}\|_\infty.
\end{aligned}$$

Thus, we have

$$\|\widehat{\mathbf{y}} - \mathbf{y}\|_\infty \le \left(\sqrt{q/(1-\gamma)}+1\right)\left(\epsilon\sqrt{r}\|\mathbf{d}\|_\infty + \|\mathbf{w}\|_\infty\right),$$

which completes the proof of Theorem 1. □

B. Proof of Theorem 2

The proof of Theorem 2 needs the following lemma [21].

Lemma 3 ([21], Lemma 2)**.** *With the same assumptions as Theorem 2, the following inequality holds for any* $\mathbf{y}\in\mathbb{C}^q$ *with probability at least* $1-\delta$*:*

$$\|\mathbf{U}_\Psi^*\mathbf{y}_\Psi\|^2 \le (1+\beta_\mathbf{y})^2\frac{|\Psi|}{q}\frac{r\mu(\mathbf{U})}{q}\|\mathbf{y}\|^2.$$

Proof of Theorem 2. From (21), we have

$$\begin{aligned}
&\|\widehat{\mathbf{y}}-\mathbf{U}\mathbf{d}\| \\
\leq&\|\widehat{\mathbf{U}}\|\|(\widehat{\mathbf{U}}_\Psi^*\widehat{\mathbf{U}}_\Psi)^{-1}\|\|\widehat{\mathbf{U}}_\Psi^*(\mathbf{u}_\Psi^\perp+\mathbf{w}_\Psi)\|+\|\mathbf{u}^\perp\| \\
\leq&\frac{q}{(1-\gamma)|\Psi|}(\|\widehat{\mathbf{U}}_\Psi^*\mathbf{u}_\Psi^\perp\|+\|\widehat{\mathbf{U}}_\Psi^*\mathbf{w}_\Psi\|)+\|\mathbf{u}^\perp\| \\
\leq&\frac{q}{(1-\gamma)|\Psi|}\frac{\sqrt{|\Psi|r\mu(\widehat{\mathbf{U}})}}{q}(1+\beta_{\mathbf{w}})\|\mathbf{w}\| \\
&+\left(\frac{q}{(1-\gamma)|\Psi|}\frac{\sqrt{|\Psi|r\mu(\widehat{\mathbf{U}})}}{q}(1+\beta_{\mathbf{u}^\perp})+1\right)\|\mathbf{u}^\perp\| \\
=&\frac{1+\beta_{\mathbf{w}}}{1-\gamma}\sqrt{\frac{r\mu(\widehat{\mathbf{U}})}{|\Psi|}}\|\mathbf{w}\|+\left(1+\frac{1+\beta_{\mathbf{u}^\perp}}{1-\gamma}\sqrt{\frac{r\mu(\widehat{\mathbf{U}})}{|\Psi|}}\right)\|\mathbf{u}^\perp\|.
\end{aligned}$$

With $\|\mathbf{u}^\perp\|\leq\|(\mathbf{I}-\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*)\mathbf{U}\|\|\mathbf{d}\|\leq\epsilon\|\mathbf{d}\|$, Theorem 2 holds. □

C. Proof of Lemma 1

Proof. If the measurements follow the model in (6), then $\mathbf{E}_L\in\mathbb{C}^{\kappa m\times r}$ in (7) spans the column space of Hankel matrix. When $\kappa=1$, the column space is spanned by $\mathbf{D}_L\in\mathbb{C}^{m\times r}$, where $\mathbf{D}_L=[\mathbf{C}\mathbf{p}_1,\ \mathbf{C}\mathbf{p}_2,\ \cdots,\ \mathbf{C}\mathbf{p}_r]$. Let $\mathbf{\Lambda}=\text{diag}(\lambda_1,\cdots,\lambda_r)$, then we have

$$\begin{aligned}
\mathbf{E}_L^*\mathbf{E}_L=&\mathbf{D}_L^*\mathbf{D}_L+\mathbf{\Lambda}^*\mathbf{D}_L^*\mathbf{D}_L\mathbf{\Lambda}+\cdots+(\mathbf{\Lambda}^*)^{\kappa-1}\mathbf{D}_L^*\mathbf{D}_L\mathbf{\Lambda}^{\kappa-1}\\
&\succ(\mathbf{\Lambda}^*)^i\mathbf{D}_L^*\mathbf{D}_L\mathbf{\Lambda}^i\succ\mathbf{0},\quad\forall i\in\{0,\cdots,\kappa-1\},
\end{aligned}$$

where $A\succ B$ means that matrix $A-B$ is positive definite.

$$(\mathbf{E}_L^*\mathbf{E}_L)^{-1}=\left(\sum_{i=0}^{\kappa-1}(\mathbf{\Lambda}^*)^i\mathbf{D}_L^*\mathbf{D}_L\mathbf{\Lambda}^i\right)^{-1}\prec\left((\mathbf{\Lambda}^*)^i\mathbf{D}_L^*\mathbf{D}_L\mathbf{\Lambda}^i\right)^{-1}.$$

Thus, $\forall k\{1,...,m\}$, $\forall i\in\{0,\cdots,\kappa-1\}$, we have

$$\begin{aligned}
&\mathbf{e}_{k+mi}^*\mathbf{E}_L(\mathbf{E}_L^*\mathbf{E}_L)^{-1}\mathbf{E}_L^*\mathbf{e}_{k+mi}\\
=&\mathbf{e}_k^*\mathbf{D}_L\mathbf{\Lambda}^i(\mathbf{E}_L^*\mathbf{E}_L)^{-1}(\mathbf{\Lambda}^*)^i\mathbf{D}_L^*\mathbf{e}_k\\
<&\mathbf{e}_k^*\mathbf{D}_L\mathbf{\Lambda}^i((\mathbf{\Lambda}^*)^i\mathbf{D}_L^*\mathbf{D}_L\mathbf{\Lambda}^i)^{-1}(\mathbf{\Lambda}^*)^i\mathbf{D}_L^*\mathbf{e}_k\\
=&\mathbf{e}_k^*\mathbf{D}_L(\mathbf{D}_L^*\mathbf{D}_L)^{-1}\mathbf{D}_L^*\mathbf{e}_k.
\end{aligned}\tag{22}$$

Then,

$$\begin{aligned}
&\mu(\mathbf{U}_\kappa)=\mu(\mathbf{E}_L)\\
=&\max_{\substack{k\in\{1,\cdots,m\}\\ i\in\{0,\cdots,\kappa-1\}}}\frac{\kappa m}{r}\|\mathbf{e}_{k+mi}^*\mathbf{E}_L(\mathbf{E}_L^*\mathbf{E}_L)^{-1}\mathbf{E}_L^*\mathbf{e}_{k+mi}\|\\
<&\max_{k\in\{1,\cdots,m\}}\frac{\kappa m}{r}\|\mathbf{e}_k^*\mathbf{D}_L(\mathbf{D}_L^*\mathbf{D}_L)^{-1}\mathbf{D}_L^*\mathbf{e}_k\|\\
=&\kappa\mu(\mathbf{D}_L)=\kappa\mu(\mathbf{U}_1),
\end{aligned}$$

where the inequality follows from (22). □

Yingshuai Hao (S'14) received the B.E. degree from Shandong University, Jinan, China, in 2011 and the M.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014.

He is pursuing the Ph.D. degree in electrical engineering at Rensselaer Polytechnic Institute, Troy, NY. His research interests include cyber security of power systems, and PMU data quality improvement.

Meng Wang (M'12) received B.S. and M.S. degrees from Tsinghua University, China, in 2005 and 2007, respectively. She received the Ph.D. degree from Cornell University, Ithaca, NY, USA, in 2012.

She is an Assistant Professor in the department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute, Troy, NY, USA. Her research interests include high-dimensional data analytics, machine learning, power systems monitoring, and synchrophasor technologies.

Joe H. Chow (F'92) received the M.S. and Ph.D. degrees from the University of Illinois, Urbana-Champaign, Urbana, IL, USA.

After working in the General Electric power system business in Schenectady, NY, USA, he joined Rensselaer Polytechnic Institute, Troy, NY, USA, in 1987, where he is Institute Professor of Electrical, Computer, and Systems Engineering. His research interests include power system dynamics and control, FACTS controllers, and synchronized phasor data. He is a member of the US National Academy of Engineering. He is a past recipient of the IEEE PES Charles Concordia Power Engineering Award.

Evangelos Farantatos (S'06-M'13) received the Diploma in Electrical and Computer Engineering from the National Technical University of Athens, Greece, in 2006 and the M.S. and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, GA, USA, in 2009 and 2012, respectively.

He is a Technical Leader with the Grid Operations and Planning R&D Group at EPRI, Palo Alto, CA. He is managing and leading the technical work of various projects related to synchrophasor technology, power systems monitoring and control, power systems stability and dynamics, system protection, state estimation. renewable energy resources modeling, DER integration, and grid operation with high levels of inverter based resources. In summer 2009, he was an intern in MISO.

Mahendra Patel (M'74-SM'81-F'16) received the B.E. from Sardar Patel University, Vallabh Vidyanagar, India, the M.S.E.E. degree from West Virginia University, Morgantown, WV, USA, and M.B.A. degree from the University of Pittsburgh, Pittsburgh, PA, USA.

He is a Technical Executive with the Grid Operations and Planning Department, Electric Power Research Institute (EPRI), Knoxville, TN, USA. He worked at PJM before joining EPRI. He has more than 40 years of experience in the Electric Power Industry in transmission planning and system reliability, transmission technologies, system dynamics, synchrophasor technology, switching surges and insulation coordination, voltage stability, system protection and power quality.