Modeless Streaming Synchrophasor Data Recovery in Nonlinear Systems

Yingshuai Hao, *Student member, IEEE,* Meng Wang, *Member, IEEE*, Joe H. Chow, *Fellow, IEEE*

***Abstract*—This paper develops a model-free approach to recover the missing points in streaming synchrophasor measurements obtained in nonlinear dynamical systems. It can accurately recover simultaneous and consecutive data losses across all channels for some time consecutively without modeling the nonlinear dynamics at all. The idea is to lift the nonlinear system to an infinite-dimensional linear dynamical system and exploit the low-rank Hankel in the lifted dimension to characterize the system dynamics. The kernel technique is employed to handle the implicit lifting function. Compared with existing model-free synchrophasor data recovery methods, our approach drops the assumption of linear systems and applies to general nonlinear systems. The algorithm has low computational complexity and can be implemented in real time. The method is validated through numerical experiments on recorded synchrophasor datasets.**

***Index Terms*—missing data recovery, phasor measurement unit, Hankel matrix, low dimensionality, kernel technique**

I. Introduction

Phasor measurement units (PMU) have been increasingly deployed in North America and provided time-synchronized high-resolution phasor measurements for the wide-area monitoring and control of power systems [1]. Due to reasons like PMU malfunction or communication congestions, the synchrophasor measurements that arrive at the operator contain data losses and bad measurements [2]. The percentage of measurements with data quality issues can be up to 17% [3]. These data issues limit the incorporation of synchrophasor measurements in real-time system monitoring and control.

Conventional data recovery methods [4], [5] require solving state estimation, and the recovery accuracy relies heavily on the accuracy of the system model. Moreover, the underlying assumption is that enough PMUs are deployed such that the system is observable. This does not always hold in practice. The recent abundance of synchrophasor data enables the development of model-free methods for data recovery. Some clustering-based algorithms, like [6], can be applied to fill in the missing measurements. The idea is first to partition the training dataset with fully-observed data points into several clusters. Then any data sample with missing entries is assigned to one cluster, and the missing entries are estimated with the samples in the same cluster. The performance heavily depends on the completeness of the training dataset and the similarities of the data samples to the training dataset. In addition, it cannot handle the simultaneous data losses across all sampling channels. Reference [7] combined the Kalman filter and one measurement updating rule, which was first studied in [8] to represent the temporal correlation among consecutive synchrophasor measurements, to improve the data quality. Though these methods have low complexity and are easy to implement, the spatial correlation among synchrophasor measurements across different channels is not employed, resulting in the greatly degraded performance when the missing data ratio is high. Reference [9] studied the intra-PMU and inter-PMU measurement correlations and leveraged them to differentiate the spoofed data from the measurements under the normal operation. The method, however, does not consider the existence of missing data and does not take the event data into account. References [10]–[13] studied the so-called low-rank property of the PMU data matrix, i.e., the spatial-temporal blocks of PMU data exhibit intrinsic low-dimensionality despite the high ambient dimension. This property of the PMU data matrix has been exploited to develop model-free methods with analytical guarantees for PMU data recovery [11], [12] and bad data correction [13].

The low-rank matrix assumption characterizes the spatial correlations among different PMUs but does not fully capture the temporal dynamics in the measurements. One resulting limitation in low-rank-based data recovery approaches, including PMU data recovery methods [11]–[13] and general low-rank matrix completion methods such as [14], [15], is that they cannot recover the missing data if all the measurements across all channels are lost simultaneously. However, simultaneous data losses across all channels is not uncommon in practice. To address this fundamental limitation, references [16], [17] prove that the Hankel matrix of the PMU data matrix is low-rank, as long as the measurements can be viewed as the output of a reduced-order linear dynamical system. Exploiting the low-rank Hankel property, model-free methods for missing data recovery have been developed for both block processing [16], [17] and streaming data processing [18], and these methods can recover simultaneous data losses across all channels consecutively for some time without modeling system dynamics. Moreover, the percentage of data losses that can be recovered by the approaches in [16]–[18] is significantly higher than that can be handled by the existing low-rank approaches. The low-rank Hankel property has also been exploited to correct simultaneous bad data across all channels [19].

Despite the superior analytical and numerical performance, the low-rank Hankel-based approaches in [16]–[19] are all based on the assumption that the system can be approximated well by a linear dynamical system. When the system is highly nonlinear, especially immediately after a significant event, the linear approximation is no longer accurate, and the missing

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.

data recovery performance degrades. This motivates us to develop an algorithm that can recover the missing entries in streaming synchrophasor data with a high accuracy even when the underlying system is highly nonlinear.

The central idea of this paper is to lift the data in a nonlinear system to a much higher dimension (potentially infinite dimension) such that the lifted data follow a linear dynamical system in the lifted space. This distinguishes the work from our previous work in [16]–[19], where no data lifting exists. We then exploit the low-rank Hankel property in the lifted space to develop model-free approach to recover the missing data. Our proposed low-rank Hankel property in the lifted space characterizes the temporal dynamics in a nonlinear system without any prior assumption on the system model. Our algorithm can recover the missing data with a higher accuracy even when data losses occur across all PMU channels simultaneously and consecutively during a nonlinear event. The algorithm has a low computational complexity and can be implemented in real time.

In practice, the proposed method could be applied with a combination of the method in [18]. Reference [18] presents a method to check the occurrence of an event or not. So if an event is determined, we resort to the method proposed in this paper to fill in the missing measurements. The difference between the estimated data and the initial estimation could be recorded. When the difference is below a specified threshold level, we think the system can be approximated well by a linear model and switch to the method in [18] to recover the missing measurements.

The rest of the paper is organized as follows. Section II describes the low-rank Hankel property in the lifted space for nonlinear systems. Section III presents the proposed approach to fill in the missing data. Sections IV documents the results of the numerical experiments on recorded PMU datasets. Section V 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. $\mathbf{z}_{\mathbb{R}}$ and $\mathbf{z}_{\mathbb{I}}$ denote the real part and imaginary part of $\mathbf{z}$, respectively. $\mathbf{Z}^*$ and $\mathbf{Z}^\dagger$ denote the conjugate transpose and pseudo-inverse of $\mathbf{Z}$. $\mathbf{z}(i)$ denotes the i-th entry of vector $\mathbf{z}$. $\mathbf{Z}(i,j)$ denotes the (i,j)-th entry of matrix $\mathbf{Z}$. The inner product between two vectors is $\langle \mathbf{z}_1, \mathbf{z}_2 \rangle = \mathbf{z}_1^* \mathbf{z}_2$, and corresponding l_2 norm of vector $\mathbf{z}$ is given as $\|\mathbf{z}\| = \langle \mathbf{z}, \mathbf{z} \rangle^{1/2}$. For one index set Ψ, $\mathbf{Z}^\Psi$ denotes the submatrix of $\mathbf{Z}$ with only the columns labeled by Ψ, while $\mathbf{Z}_\Psi$ represents the submatrix of $\mathbf{Z}$ with only the rows labeled by Ψ.

II. Low-rank Lifted Hankel Property

A. Data recovery exploiting the low-rank Hankel structure

Let $\mathbf{y}_i \in \mathbf{C}^m$ denote the measured phasors of m channels at instant i, then the phasors from m channels during T consecutive time instants can be denoted as follows:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_T \end{bmatrix} \in \mathbb{C}^{m \times T}. \tag{1}$$

Let κ denote the number of vectors in each column of the Hankel matrix. Then the constructed Hankel matrix $\mathcal{H}_\kappa(\mathbf{Y})$ from $\mathbf{Y}$ with parameter κ is

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

Given a general matrix $\mathbf{X}$ with its singular value decomposition (SVD) as $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$, $\mathbf{X}$ can be approximated by a rank-r matrix $\mathbf{U}\mathbf{\Sigma}^r\mathbf{V}^*$, where $\mathbf{\Sigma}^r$ is the diagonal matrix obtained by setting all but the largest r singular values in $\mathbf{\Sigma}$ to zero. The rank-r approximation error ratio is defined as

$$e^r(\mathbf{X}) = \|\mathbf{U}\mathbf{\Sigma}^r\mathbf{V}^* - \mathbf{X}\|_F / \|\mathbf{X}\|_F = \|\mathbf{\Sigma} - \mathbf{\Sigma}^r\|_F / \|\mathbf{\Sigma}\|_F. \tag{3}$$

The low-rank Hankel property of the synchrophasor data has been demonstrated on a recorded dataset in Central New York Power System with 11 voltage phasors measured at a rate of 30 samples per second [18]. Fig. 1 (Fig. 1 of [18]) shows the recorded dataset of 20 seconds, where a disturbance occurs around 2.3 seconds. Fig. 2 contains a time window of 2.7 seconds immediately after the start of the disturbance, as indicated with two dotted lines in Fig. 1.

Fig. 3 shows the approximation errors of the constructed Hankel matrices using the synchrophasor measurements in Fig. 2 with different approximate rank r and κ. For example, when $\kappa = 5$, a rank-5 approximation to $\mathcal{H}_5(\mathbf{Y})$ has an error ratio of 0.002. When $\kappa = 1$, $\mathcal{H}_1(\mathbf{Y})$ is reduced to $\mathbf{Y}$. Thus, one can approximate Hankel matrices with low-rank matrices with very small error ratios.

References [16], [18] also provide the analytical justification of the low-rank Hankel property. One can view the measurements in the time window of Fig. 2 as the output of a linear dynamical system after an impulse input, i.e.,

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

with state vector $\mathbf{x}_t \in \mathbb{C}^n$, and output vector $\mathbf{y}_t \in \mathbb{C}^m$. Then the Hankel matrix $\mathcal{H}_\kappa(\mathbf{Y})$ is approximately rank-r if the system can be approximated by an order-r linear system.

The low-rank Hankel property captures the spatial and temporal correlations among the time series of a linear dynamical system without directly modeling the system. Exploiting the low-rank Hankel property, references [16], [17], [19] develop fast block-processing algorithms to recover missing data and correct bad measurements. Reference [18] develops an online algorithm to recover the missing points and correct bad data in streaming PMU data in real time. The method can differentiate event data from consecutive bad data without any prior assumption of system events and models.

The above data-driven approaches exploiting the low-rank Hankel property are based on the critical assumption that the system can be approximated by a linear dynamical system in the observation window. Although the linear dynamic model is an accurate approximation when the power system is under normal operations or non-major events, the power system after a significant event can be highly nonlinear, and the linear approximation is no longer accurate. If one still approximates

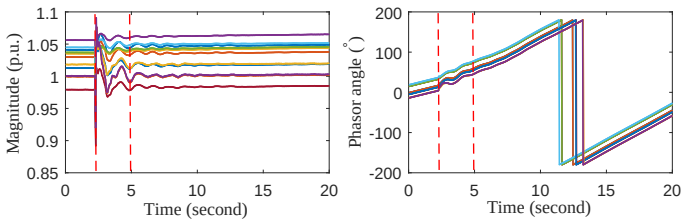


Fig. 1: The measured voltage phasors

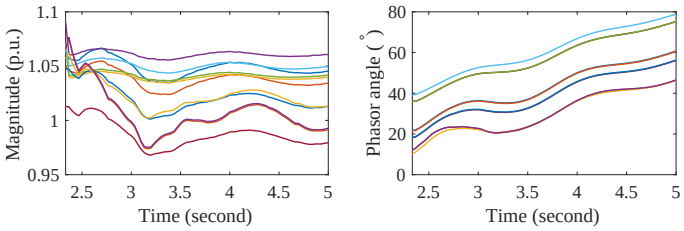


Fig. 2: Voltage phasors under system disturbances

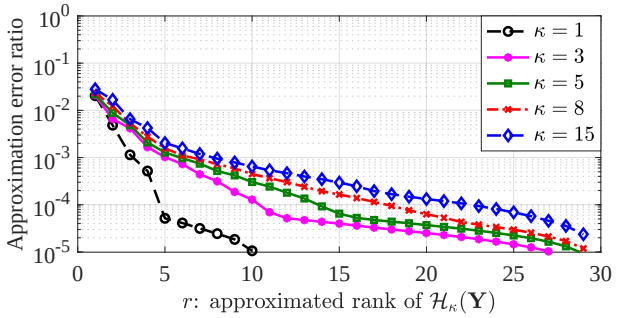


Fig. 3: Approximation error ratios of Hankel matrices from observation data $\mathbf{Y}$ in Fig. 2

the Hankel matrix of the measurements by a rank-r matrix when the system is highly nonlinear, r needs to be large to capture the nonlinear features. A very large r usually results in the over-fitting problem, which causes significant errors in missing data recovery. We will show an example of over-fitting when applying the approach in [18] to recover missing measurements in Fig. 12 of Section IV.

B. Low-rank lifted Hankel matrix

This paper focuses on missing data recovery when the system is highly nonlinear. The central idea is to characterize the nonlinear dynamics without directly modeling the system. Consider a non-linear dynamical system after an impulse input,

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t), \quad \mathbf{y}_t = g(\mathbf{x}_t), \tag{5}$$

with the state variable $\mathbf{x} \in \mathbb{C}^n$, and the observation $\mathbf{y} \in \mathbb{C}^m$. From Koopman theory, for every nonlinear dynamical system, there exists a Koopman operator that lifts the system to an infinite-dimensional but linear system with the same observations [20], [21]. In other words, there exists a mapping $\widehat{\Phi}(\cdot)$ from $\mathbb{C}^n$ to $\mathbb{C}^N$ with $\mathbf{z}_t := \widehat{\Phi}(\mathbf{x}_t) \in \mathbb{C}^N$, where N can be infinite, $\mathbf{A} \in \mathbb{C}^{N\times N}$, and $\mathbf{C} \in \mathbb{C}^{m\times N}$ such that

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t, \quad \mathbf{y}_t = \mathbf{C}\mathbf{z}_t. \tag{6}$$

Then the oscillation modes and the stability of the nonlinear system can be analyzed in the equivalent lifted linear system. Data-driven approaches have been developed to estimate the dominant modes of the lifted linear system [20], [21].

Let

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_T \end{bmatrix} \in \mathbb{C}^{N\times T}, \tag{7}$$

$$\mathcal{H}_\kappa(\mathbf{Z}) = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_{T-\kappa+1} \\ \mathbf{z}_2 & \mathbf{z}_3 & \cdots & \mathbf{z}_{T-\kappa+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{z}_\kappa & \mathbf{z}_{\kappa+1} & \cdots & \mathbf{z}_T \end{bmatrix} \in \mathbb{C}^{\kappa N\times(T-\kappa+1)}. \tag{8}$$

If the lifted linear dynamical system can be approximated by a reduced-order system in the lifted space $\mathbb{C}^N$, then the rank of $\mathcal{H}_\kappa(\mathbf{Z})$ is much smaller than its ambient dimension. The low-rank property of the lifted Hankel matrix $\mathcal{H}_\kappa(\mathbf{Z})$ characterizes the dominant nonlinear dynamics of the system without directly computing the dominant modes.

Since the lifting function is often unknown or difficult to compute, $\mathbf{z}_t$ cannot be directly computed. One standard trick in machine learning is to apply the kernel method [22]. The kernel function $\widehat{\mathcal{K}}$ is defined as

$$\widehat{\mathcal{K}}(\mathbf{x}_i, \mathbf{x}_j) = \langle \widehat{\Phi}(\mathbf{x}_i), \widehat{\Phi}(\mathbf{x}_j) \rangle = \langle \mathbf{z}_i, \mathbf{z}_j \rangle. \tag{9}$$

Given $\widehat{\mathcal{K}}$, one can compute the inner product of $\mathbf{z}_i$ and $\mathbf{z}_j$ without knowing $\mathbf{z}_i$ and $\mathbf{z}_j$ explicitly. For example, reference [23] lifts a nonlinear dynamical system to a linear system in a higher dimension and employs the kernel trick to predict the future measurements. Reference [23] needs to estimate the operator $\mathbf{A}$ in (6), which is computationally complex, and the estimation error of $\mathbf{A}$ leads to the prediction inaccuracy of $\mathbf{z}_t$'s. We will show here that by employing the low-rank Hankel structure, the estimation of $\mathbf{A}$ is no longer needed.

We assume that one can estimate $\mathbf{x}_t$'s from $\mathbf{y}_t$'s so that we define a kernel function $\mathcal{K}$ on $\mathbf{y}_t$'s directly, i.e.,

$$\mathcal{K}(\mathbf{y}_i, \mathbf{y}_j) = \langle \mathbf{z}_i, \mathbf{z}_j \rangle. \tag{10}$$

Popular choices of kernel functions include Gaussian kernel, polynomial kernel, and sigmoid kernel [22]. Reference [23] shows that the Gaussian kernel outperforms the polynomial kernel in their nonlinear system. Moreover, different from the polynomial kernel, the Gaussian kernel is defined in the infinite dimensional lifted space. Since the measurements are represented by complex numbers, we employ the complex Gaussian kernel [24] $\mathcal{K}_{\mathbb{C}\mathcal{G}}(\mathbf{y}_i, \mathbf{y}_j)$, which is defined as

$$\mathcal{K}_{\mathcal{G}}(\mathbf{y}_{i\mathbb{R}}, \mathbf{y}_{j\mathbb{R}}) + \mathcal{K}_{\mathcal{G}}(\mathbf{y}_{i\mathbb{I}}, \mathbf{y}_{j\mathbb{I}}) + \mathrm{j}(\mathcal{K}_{\mathcal{G}}(\mathbf{y}_{i\mathbb{R}}, \mathbf{y}_{j\mathbb{I}}) - \mathcal{K}_{\mathcal{G}}(\mathbf{y}_{i\mathbb{I}}, \mathbf{y}_{j\mathbb{R}})), \tag{11}$$

where $\mathbf{y}_{i\mathbb{R}}, \mathbf{y}_{i\mathbb{I}}$ denote the real part and imaginary part of $\mathbf{y}_i$, respectively. j is the imaginary unit. $\mathcal{K}_{\mathcal{G}}(\mathbf{y}_i, \mathbf{y}_j)$ denotes the real Gaussian kernel, which is defined as

$$\mathcal{K}_{\mathcal{G}}(\mathbf{y}_i, \mathbf{y}_j) = \exp(-\gamma \|\mathbf{y}_i - \mathbf{y}_j\|_2^2), \tag{12}$$

where γ is a parameter selected by users. Please see the Appendix for a detailed discussion of the Gaussian kernel.

One can compute a rank-r approximation to $\mathcal{H}_\kappa(\mathbf{Z})$ according to (3). Let $\mathcal{H}_\kappa(\mathbf{Z}) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ denote the SVD of $\mathcal{H}_\kappa(\mathbf{Z})$, where the diagonal matrix $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \cdots, \sigma_{T-\kappa+1})$ contains the singular values. Since $\mathcal{H}_\kappa(\mathbf{Z})$ is not explicitly known, one cannot compute its SVD directly. Instead, we employ

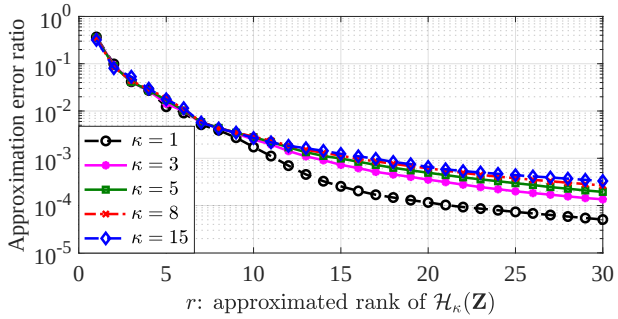


Fig. 4: Approximation error ratios of Hankel matrices from the lifted data $\mathbf{Z}$

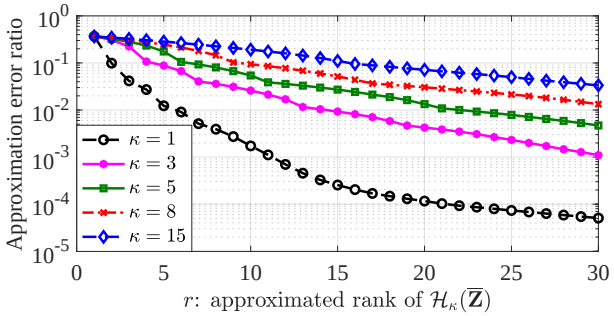


Fig. 5: Approximation error ratios of Hankel matrices from a column-wise permuted lifted data $\overline{\mathbf{Z}}$

the kernel trick and compute the SVD of $\mathcal{H}_\kappa(\mathbf{Z})^*\mathcal{H}_\kappa(\mathbf{Z}) = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^*$. The singular value of $\mathcal{H}_\kappa(\mathbf{Z})^*\mathcal{H}_\kappa(\mathbf{Z})$ is the square of the corresponding singular value of $\mathcal{H}_\kappa(\mathbf{Z})$. The (i,j)-entry of $\mathcal{H}_\kappa(\mathbf{Z})^*\mathcal{H}_\kappa(\mathbf{Z})$ is

$$\begin{bmatrix}\mathbf{z}_i^* & \cdots & \mathbf{z}_{\kappa+i-1}^*\end{bmatrix}\begin{bmatrix}\mathbf{z}_j\\ \vdots \\ \mathbf{z}_{\kappa+j-1}\end{bmatrix} = \sum_{l=0}^{\kappa-1}\mathcal{KCG}(\mathbf{y}_{i+l},\mathbf{y}_{j+l}) \quad (13)$$

and can be computed according to (11) and (12). After obtaining $\sigma_i^2, i = 1,\cdots,T-\kappa+1$ through the SVD of $\mathcal{H}_\kappa(\mathbf{Z})^*\mathcal{H}_\kappa(\mathbf{Z})$, the rank-$r$ approximation error ratio of $\mathcal{H}_\kappa(\mathbf{Z})$ can be computed as

$$e^r(\mathcal{H}_\kappa(\mathbf{Z})) = \sqrt{\sum_{i=r+1}^{T-\kappa+1}\sigma_i^2 \Big/ \sum_{i=1}^{T-\kappa+1}\sigma_i^2}. \quad (14)$$

Here we still use the data shown in Fig. 2 to validate the low-rankness of $\mathcal{H}_\kappa(\mathbf{Z})$. The rank of $\mathcal{H}_\kappa(\mathbf{Z})$ is bounded by $\min(\kappa N, T-\kappa+1) = T-\kappa+1$. Here $T = 81$. As shown in Fig. 4, $\mathcal{H}_\kappa(\mathbf{Z})$ can be approximated by lower-rank matrices with relatively small error ratios. Moreover, the low-rank Hankel property characterizes the temporal dynamics in the dynamical system, and this does not hold for a general low-rank matrix. To see this, let $\overline{\mathbf{Z}}$ denote the obtained matrix after a random column-wise permutation of $\mathbf{Z}$, Fig. 5 shows the approximation errors of $\mathcal{H}_\kappa(\overline{\mathbf{Z}})$. For any $\kappa \geq 2$ and any $r \geq 2$, $e^r(\mathcal{H}_\kappa(\overline{\mathbf{Z}}))$ is much larger than $e^r(\mathcal{H}_\kappa(\mathbf{Z}))$. That is because after a permutation, the temporal dynamics is no longer preserved in $\mathcal{H}_\kappa(\overline{\mathbf{Z}})$.

We notice that for the same κ and r, $e^r(\mathcal{H}_\kappa(\mathbf{Z}))$ is larger than $e^r(\mathcal{H}_\kappa(\mathbf{Y}))$, as one can see from Figs. 3 and 4. However, note that the ambient dimension of $\mathbf{z}_t$ is also much larger than that of $\mathbf{y}_t$, as $N \gg m$. Analyzing the data in the lifted dimension $\mathbb{C}^N$ resolves the problem of over-fitting using the linear approximation.

III. Online Algorithm for Missing Data Estimation

We next present our data recovery method by exploiting the low-rank Hankel property in the lifted space. Our approach exploits the nonlinear dynamics in the data without modeling the dynamical system directly.

Let $\mathbf{y}_t \in \mathbb{C}^m$ denote the obtained observations (possibly with missing points) at time instant t. Let $\widehat{\mathbf{y}}_t \in \mathbb{C}^m$ denote the recovered measurements of $\mathbf{y}_t$. $\widetilde{\mathbf{y}}_t$ represents the initial estimation of $\mathbf{y}$ in our algorithm. The corresponding lifted data of $\widehat{\mathbf{y}}_t$ and $\widetilde{\mathbf{y}}_t$ are denoted as $\widehat{\mathbf{z}}_t$ and $\widetilde{\mathbf{z}}_t$, respectively. Let Ω represent the index set of observed entries in $\mathbf{y}_t$, and we set $\widehat{\mathbf{y}}_t(i) = \widetilde{\mathbf{y}}_t(i) = \mathbf{y}_t(i), \forall i \in \Omega$.

Similar to the algorithm in [18], our algorithm initializes with $\mathbf{y}_1 \sim \mathbf{y}_L$ in a window of length L and recovers missing data at each time instant t for $t > L$ in a streaming fashion. If there exist unobserved entries in $\mathbf{y}_1 \sim \mathbf{y}_L$, these missing points can be estimated from block-processing methods such as those in [12], [17] to obtain $\widehat{\mathbf{y}}_1 \sim \widehat{\mathbf{y}}_L$.

Let Θ denote the set of instants $\{t-L, ..., t-1\}$, and let $\widehat{\mathbf{Z}}^\Theta = [\widehat{\mathbf{z}}_{t-L}, ..., \widehat{\mathbf{z}}_{t-1}]$. As shown in Fig. 6, the Hankel matrix constructed by $\widehat{\mathbf{z}}_{t-L}$ to $\widehat{\mathbf{z}}_t$ can be divided into two parts: the first $L-\kappa+1$ columns, represented as $\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta)$, and the last column, denoted by $\mathbf{w}'$. Reference [18] similarly constructs a Hankel matrix, which can be regarded as a special case of our current approach with $\Phi(\mathbf{y}) = \mathbf{y}$. In this case, $\widehat{\mathbf{z}}_t = \widehat{\mathbf{y}}_t$, and $\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta) = \mathcal{H}_\kappa(\widehat{\mathbf{Y}}^\Theta)$ in Fig. 6.

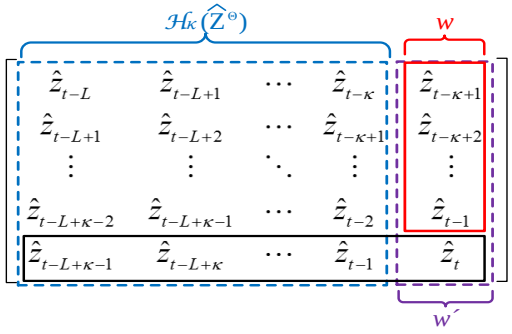


Fig. 6: Illustration of the Hankel matrix constructed at time t

Our algorithm includes two major steps: 1) estimate $\widehat{\mathbf{z}}_t$ by leveraging the low-rank Hankel structure $\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta)$; 2) determine $\widehat{\mathbf{y}}_t$ as the pre-image of $\widehat{\mathbf{z}}_t$. The low-rank Hankel method in [18] can be viewed as a special case of Step 1 with $\Phi(\mathbf{y}) = \mathbf{y}$. Step 2 does not exist in [18].

A. Step 1: $\widehat{\mathbf{z}}_t$ estimation

Since the lifted Hankel structure is low-rank in the lifted space, the last column $\mathbf{w}'$ lies in the subspace spanned by the column vectors of $\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta)$. In other words, $\mathbf{w}'$ can be written as $\mathbf{w}' = \mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta)\widehat{\mathbf{d}}$ for some $\widehat{\mathbf{d}} \in \mathbb{C}^{L-\kappa+1}$. If $\widehat{\mathbf{d}}$ is known,

then from the bottom block in Fig. 6, $\widehat{\mathbf{z}}_t$ can be estimated by a linear combination of $\widehat{\mathbf{z}}_{t-L+\kappa-1}$ to $\widehat{\mathbf{z}}_{t-1}$, i.e.,

$$\widehat{\mathbf{z}}_t = \sum_{i=1}^{L-\kappa+1} \widehat{\mathbf{z}}_{t-L+\kappa-2+i}\widehat{\mathbf{d}}(i) := \sum_{i=1}^{L-\kappa+1} \widehat{\mathbf{z}}_{\tau(i)}\widehat{\mathbf{d}}(i), \quad (15)$$

where $\tau(i)$ is defined as $t-L+\kappa-2+i$ to simplify notations. Then the problem is to estimate $\widehat{\mathbf{d}}$.

Our algorithm needs to compute $\langle \widehat{\mathbf{z}}_j, \mathbf{z}_t \rangle = \mathcal{K}_{\mathbb{C}\mathcal{G}}(\widehat{\mathbf{y}}_j, \mathbf{y}_t)$ for some $j < t$. It cannot be computed directly because $\mathbf{y}_t$ contains missing data. Therefore, we first obtain an initial estimation, denoted as $\widetilde{\mathbf{y}}_t$, from the observed entries of $\mathbf{y}_t$ and compute $\mathcal{K}_{\mathbb{C}\mathcal{G}}(\widehat{\mathbf{y}}_j, \widetilde{\mathbf{y}}_t)$ instead.

We employ the algorithm in [18] to compute $\widetilde{\mathbf{y}}_t$. The approach in [18] first determines the dominant subspace basis $\mathbf{U}^r \in \mathbb{C}^{\kappa m \times r}$ from the SVD of $\mathcal{H}_\kappa(\widehat{\mathbf{Y}}^\Theta)$. Since $\mathbf{w}'$ can be represented by $\mathbf{U}^r\mathbf{d}$ for some $\mathbf{d} \in \mathbb{C}^{L-\kappa+1}$, the method in [18] computes vector $\mathbf{d}$ using linear regression with the observed entries in $\mathbf{w}'$. Finally, $\widetilde{\mathbf{y}}_t$ is computed as:

$$\widetilde{\mathbf{y}}_t(i) = \begin{cases} (\mathbf{U}^r\mathbf{d})(i), & \forall i \notin \Omega; \\ \mathbf{y}_t(i), & \forall i \in \Omega. \end{cases} \quad (16)$$

The approximate rank r is determined from a fixed approximation ratio in [18]. When the linear model is no longer accurate for a nonlinear event, r might be large, and that leads to over-fitting. Since our algorithm here only needs an initial estimation $\widetilde{\mathbf{y}}_t$, we fix a small r like 1 or 2 to avoid over-fitting.

After computing $\widetilde{\mathbf{y}}_t$, with the historically estimated data $\widehat{\mathbf{z}}_{t-\kappa+1} \sim \widehat{\mathbf{z}}_{t-1}$, we estimate $\widehat{\mathbf{d}}$ by solving from the following linear regression problem:

$$\begin{aligned} \widehat{\mathbf{d}} &= \underset{\mathbf{d}\in\mathbb{C}^{L-\kappa+1}}{\arg\min} \|\mathbf{w}' - \mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta)\mathbf{d}\|_2^2 + \lambda\|\mathbf{d}\|_2^2 \\ &= \left(\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta)^*\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta) + \lambda\mathbf{I}\right)^{-1}\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta)^*\mathbf{w}' \\ &:= (\mathbf{H} + \lambda\mathbf{I})^{-1}\mathbf{h}, \end{aligned} \quad (17)$$

where $\mathbf{H} = \mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta)^*\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta) \in \mathbb{C}^{(L-\kappa+1)\times(L-\kappa+1)}, \mathbf{h} = \mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Theta)^*\mathbf{w}' \in \mathbb{C}^{L-\kappa+1}$. Here a penalty term $\|\mathbf{d}\|_2^2$ with weight λ is added to reduce over-fitting. The (i,j)-th entry in $\mathbf{H}$ can be computed by

$$\begin{aligned} \mathbf{H}(i,j) &= \begin{bmatrix} (\widehat{\mathbf{z}}_{\tau(i)-\kappa+1})^* & \cdots & (\widehat{\mathbf{z}}_{\tau(i)})^* \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{z}}_{\tau(j)-\kappa+1} \\ \vdots \\ \widehat{\mathbf{z}}_{\tau(j)} \end{bmatrix} \\ &= \sum_{l=0}^{\kappa-1} \mathcal{K}_{\mathbb{C}\mathcal{G}}(\widehat{\mathbf{y}}_{\tau(i)-l}, \widehat{\mathbf{y}}_{\tau(j)-l}). \end{aligned} \quad (18)$$

The i-th entry of $\mathbf{h}$ can be computed as

$$\begin{aligned} \mathbf{h}(i) &= \begin{bmatrix} (\widehat{\mathbf{z}}_{\tau(i)-\kappa+1})^* & \cdots & (\widehat{\mathbf{z}}_{\tau(i)})^* \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{z}}_{t-\kappa+1} \\ \vdots \\ \widetilde{\mathbf{z}}_t \end{bmatrix} \\ &= \sum_{l=1}^{\kappa-1} \mathcal{K}_{\mathbb{C}\mathcal{G}}(\widehat{\mathbf{y}}_{\tau(i)-l}, \widehat{\mathbf{y}}_{t-l}) + \mathcal{K}_{\mathbb{C}\mathcal{G}}(\widehat{\mathbf{y}}_{\tau(i)}, \widetilde{\mathbf{y}}_t). \end{aligned} \quad (19)$$

We estimate $\widehat{\mathbf{d}}$ by combining (17), (18) and (19). Then $\mathbf{z}_t$ can be represented by (15). Note that we do not need to compute $\widehat{\mathbf{z}}_t$ explicitly. (15) together with the kernel function is sufficient for the subsequent computation.

Algorithm 1 Online missing data estimation exploiting Hankel matrix and kernels

Require: κ, L, penalty weight λ, stopping criteria δ
Initialization: obtain $\widehat{\mathbf{y}}_1 \sim \widehat{\mathbf{y}}_L$ from $\mathbf{y}_1 \sim \mathbf{y}_L$
For $t = L+1, L+2, \cdots$
If $\mathbf{y}_t$ is not fully observed:

1. Let $\Theta = \{t-L, \cdots, t-1\}$, with the data block $\widehat{\mathbf{Y}}^\Theta = [\widehat{\mathbf{y}}_{t-L}, \cdots, \widehat{\mathbf{y}}_{t-1}]$, compute $\widetilde{\mathbf{y}}_t$ from the algorithm in [18] as an initial estimate of $\widehat{\mathbf{y}}_t$.
2. Construct matrix $\mathbf{H}$ and vector $\mathbf{h}$ following (18) and (19), respectively.
3. Compute coefficient vector $\widehat{\mathbf{d}}$ following (17).
4. Follow (25) to iteratively update the estimation until $\|\widehat{\mathbf{y}}_t^{(l)} - \widehat{\mathbf{y}}_t^{(l-1)}\| \le \delta$. Return $\widehat{\mathbf{y}}_t^{(l)}$ as the estimation $\widehat{\mathbf{y}}_t$.

End if
End for

B. Step 2: $\widehat{\mathbf{y}}_t$ estimation

$\widehat{\mathbf{y}}_t$ can be estimated by solving

$$\min_{\mathbf{y}\in\mathbb{C}^m} \|\Phi(\mathbf{y}) - \widehat{\mathbf{z}}_t\| \quad \text{s.t.} \quad \mathbf{y}(j) = \mathbf{y}_t(j),\ \forall j \in \Omega. \quad (20)$$

Although $\widehat{\mathbf{z}}_t$ is not explicitly known, we have

$$\begin{aligned} &\|\Phi(\mathbf{y}) - \widehat{\mathbf{z}}_t\|_2^2 \\ =&\langle\Phi(\mathbf{y}), \Phi(\mathbf{y})\rangle - \langle\Phi(\mathbf{y}), \widehat{\mathbf{z}}_t\rangle - \langle\widehat{\mathbf{z}}_t, \Phi(\mathbf{y})\rangle + \langle\widehat{\mathbf{z}}_t, \widehat{\mathbf{z}}_t\rangle \\ \overset{(a)}{=}&2 - 2\langle\Phi(\mathbf{y}), \textstyle\sum_{i=1}^{L-\kappa+1} \Phi(\widehat{\mathbf{y}}_{\tau(i)})\widehat{\mathbf{d}}(i)\rangle_{\mathbb{R}} + 2 \\ \overset{(b)}{=}&4 - 2\sum_{i=1}^{L-\kappa+1} \left(\mathcal{K}_{\mathbb{C}\mathcal{G}}(\mathbf{y}, \widehat{\mathbf{y}}_{\tau(i)})\widehat{\mathbf{d}}(i)\right)_{\mathbb{R}}. \end{aligned} \quad (21)$$

(a) holds since $\mathcal{K}_{\mathbb{C}\mathcal{G}}(\mathbf{y},\mathbf{y}) = \langle\Phi(\mathbf{y}), \Phi(\mathbf{y})\rangle = 2, \forall \mathbf{y}$, $\langle\mathbf{y}_1, \mathbf{y}_2\rangle = \langle\mathbf{y}_2, \mathbf{y}_1\rangle^*$, and (15). (b) follows from (10).

$\mathbf{y}$ can be decomposed as $[\mathbf{y}_\Omega; \mathbf{y}_{\Omega^c}]$, where Ω^c denotes the complementary set of Ω. Similarly, we decompose $\widehat{\mathbf{y}}_{\tau(i)}$ as $[(\widehat{\mathbf{y}}_{\tau(i)})_\Omega; (\widehat{\mathbf{y}}_{\tau(i)})_{\Omega^c}]$. Given $\mathbf{y}_\Omega$, we combine (20) and (21) and rewrite the constrained optimization problem in (20) as an unconstrained optimization problem:

$$\min_{\mathbf{y}_{\Omega^c}} -\sum_{i=1}^{L-\kappa+1} \left(\mathcal{K}_{\mathbb{C}\mathcal{G}}\left([\mathbf{y}_\Omega; \mathbf{y}_{\Omega^c}], [(\widehat{\mathbf{y}}_{\tau(i)})_\Omega; (\widehat{\mathbf{y}}_{\tau(i)})_{\Omega^c}]\right)\widehat{\mathbf{d}}(i)\right)_{\mathbb{R}} \quad (22)$$

The objective function in (22) is a function of $\mathbf{y}_{\Omega^c}$, and we denote it by $F(\mathbf{y}_{\Omega^c})$. We exploit the idea in [25] to develop a heuristic to solve (22). Note that

$$\nabla F(\mathbf{y}_{\Omega^c}) = 0 \quad (23)$$

is a necessary condition of the optimality of (22). We can rewrite (23) as (24), where $\mathbf{y}$ exists in both sides of the equations. Based on (24), we propose an iterative algorithm to solve (22). With the computed $\widehat{\mathbf{y}}_t^{(l)}$ in the l-th iteration, we can update $\widehat{\mathbf{y}}_t^{(l+1)}$ in the $(l+1)$-th iteration following (25). In (24) and (25), the summation of i is from 1 to $L-\kappa+1$.

Our proposed missing data recovery algorithm is summarized in Algorithm 1. It can be easily extended to handle both missing data and bad measurements simultaneously. Similar to the idea in [18], we can first predict the measurements at time

t from past observations using the algorithm in [18]. After obtaining observation $\mathbf{y}_t$, for each observed entry $\mathbf{y}_t(i)$, $\mathbf{y}_t(i)$ is determined as trusted if it deviates from the prediction by no more than a pre-defined threshold. The untrusted entries in $\mathbf{y}_t$ will be removed and treated as missing data. We then run Algorithm 1 to recover the missing data.

C. Computational complexity analysis

The computational complexity of estimating $\widetilde{\mathbf{y}}_t \in \mathbb{C}^m$ in Line 1 using the approach in [18] is $\mathcal{O}(\kappa mL \cdot \min(\kappa m, L))$ [18]. In Line 2, the complexity of computing $\mathbf{H}(i,j)$ following (18) is $\mathcal{O}(\kappa m)$. Since the dimensionality of $\mathbf{H}$ is $(L-\kappa+1)\times(L-\kappa+1)$, totally it costs $\mathcal{O}(\kappa mL^2)$ to compute $\mathbf{H}$. Similarly, the complexity to compute $\mathbf{h}$ is $\mathcal{O}(\kappa mL)$. The computation of $\widehat{\mathbf{d}}$ in Line 3 needs $\mathcal{O}(L^3)$ flops. In Line 4, each iteration requires $\mathcal{O}(Lm)$ flops. Let n_s denote the number of iterations, then the complexity of Line 4 is $\mathcal{O}(Lmn_s)$.

The total computational complexity at time t to estimate $\widehat{\mathbf{y}}_t$ is $\mathcal{O}(\kappa mL \cdot \min(\kappa m, L) + \kappa mL^2 + L^3 + Lmn_s)$. When $L < \kappa m$, the complexity is simplified to $\mathcal{O}(\kappa mL^2 + Lmn_s)$. If n_s is further regraded as a constant, the computational cost is $\mathcal{O}(\kappa mL^2)$, in the same order as the algorithm in [18].

D. A Simplified algorithm without estimation of $\widetilde{\mathbf{y}}_t$

To further reduce the computational complexity, we can skip the initial estimation of $\widetilde{\mathbf{y}}_t$ and only use the data in $\mathbf{w}$, as shown in Fig. 6, to estimate vector $\widehat{\mathbf{d}}$. Define $\mathbf{P} \in \mathbb{R}^{(\kappa-1)N\times\kappa N}$ with $\mathbf{P}(i,i)=1$ when $i=1,\cdots,(\kappa-1)N$ and 0 otherwise. Then $\widehat{\mathbf{d}}$ can be computed from

$$\begin{aligned}\widehat{\mathbf{d}} &= \underset{\mathbf{d}\in\mathbb{C}^{L-\kappa+1}}{\arg\min} \|\mathbf{P}(\mathbf{w}' - \mathcal{H}_\kappa(\widehat{\mathbf{Z}}^{\Theta})\mathbf{d})\|_2^2 + \lambda\|\mathbf{d}\|_2^2 \\ &= \left(\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^{\Theta})^*\mathbf{P}^*\mathbf{P}\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^{\mathbf{\Theta}}) + \lambda\mathbf{I}\right)^{-1}\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^{\Theta})^*\mathbf{P}^*\mathbf{P}\mathbf{w}' \\ &:= (\mathbf{H_P} + \lambda\mathbf{I})^{-1}\mathbf{h_P}.\end{aligned} \tag{26}$$

where $\mathbf{H_P} = \mathcal{H}_\kappa(\widehat{\mathbf{Z}}^{\Theta})^*\mathbf{P}^*\mathbf{P}\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^{\Theta}) \in \mathbb{C}^{(L-\kappa+1)\times(L-\kappa+1)}$, $\mathbf{h_P} = \mathcal{H}_\kappa(\widehat{\mathbf{Z}}^{\Theta})^*\mathbf{P}^*\mathbf{P}\mathbf{w}' \in \mathbb{C}^{L-\kappa+1}$. Similar to (18) and (19), the (i,j)-th entry of $\mathbf{H_P}$ and the i-th entry of $\mathbf{h_P}$ can be computed as follows:

$$\mathbf{H_P}(i,j) = \sum_{l=1}^{\kappa-1}\mathcal{K}_{\mathbb{C}\mathcal{G}}(\widehat{\mathbf{y}}_{\tau(i)-l}, \widehat{\mathbf{y}}_{\tau(j)-l}), \tag{27}$$

$$\mathbf{h_P}(i) = \sum_{l=1}^{\kappa-1}\mathcal{K}_{\mathbb{C}\mathcal{G}}(\widehat{\mathbf{y}}_{\tau(i)-l}, \widehat{\mathbf{y}}_{t-l}). \tag{28}$$

Algorithm 1 can be simplified as follows. We skip Line 1 and compute $\mathbf{H_P}$ and $\mathbf{h_P}$ following (27) and (28) instead in Line 2. Then we follow (26) instead to compute $\widehat{\mathbf{d}}$ in Line 3 and finally follow Line 4 to obtain $\widehat{\mathbf{y}}_t$.

The total complexity without initial estimation of $\widetilde{\mathbf{y}}_t$ is $\mathcal{O}(Lmn_s + \kappa mL^2 + L^3)$. In Section IV, we compare the data recovery performance of with and without $\widetilde{\mathbf{y}}_t$.

IV. Numerical Experiments

Here we use the recorded practical synchrophasor dataset to test the performance of the proposed algorithm in Section III, where a phasor is represented in the rectangular form. All the simulations are conducted on a desktop with Window 7 system, 3.4GHz Intel Core i7 and 16GB memory. Four modes of data loss patterns are tested:

- Mode 1: Missing data points exist randomly across all m PMU channels and all time instants.
- Mode 2: Missing data points exist in all m channels simultaneously at random time instants.
- Mode 3: Missing data points exist across all m channels simultaneously at consecutive time instants.
- Mode 4: Block-wise data losses occur across k randomly selected channels and d consecutive instants repeatedly for fixed k and d.

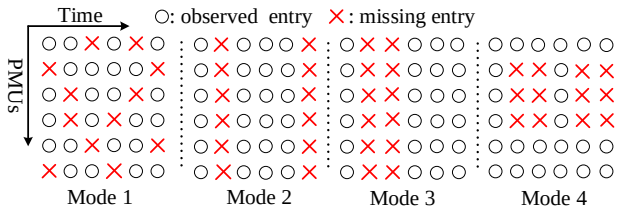


Fig. 7: Diagram of four missing data modes

$$\begin{aligned}(\mathbf{y}_{\Omega^c})_{\mathbb{R}} &= \frac{\sum_i((\widehat{\mathbf{y}}_{\tau(i)})_{\Omega^c})_{\mathbb{R}} \times \mathcal{K}_{\mathcal{G}}(\mathbf{y}_{\mathbb{R}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{R}}) \times \widehat{\mathbf{d}}_{\mathbb{R}}(i) - \sum_i((\widehat{\mathbf{y}}_{\tau(i)})_{\Omega^c})_{\mathbb{I}} \times \mathcal{K}_{\mathcal{G}}(\mathbf{y}_{\mathbb{R}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{I}}) \times \widehat{\mathbf{d}}_{\mathbb{I}}(i)}{\sum_i \mathcal{K}_{\mathcal{G}}(\mathbf{y}_{\mathbb{R}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{R}}) \times \widehat{\mathbf{d}}_{\mathbb{R}}(i) - \sum_i \mathcal{K}_{\mathcal{G}}(\mathbf{y}_{\mathbb{R}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{I}}) \times \widehat{\mathbf{d}}_{\mathbb{I}}(i)}, \\ (\mathbf{y}_{\Omega^c})_{\mathbb{I}} &= \frac{\sum_i((\widehat{\mathbf{y}}_{\tau(i)})_{\Omega^c})_{\mathbb{I}} \times \mathcal{K}_{\mathcal{G}}(\mathbf{y}_{\mathbb{I}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{I}}) \times \widehat{\mathbf{d}}_{\mathbb{R}}(i) + \sum_i((\widehat{\mathbf{y}}_{\tau(i)})_{\Omega^c})_{\mathbb{R}} \times \mathcal{K}_{\mathcal{G}}(\mathbf{y}_{\mathbb{I}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{R}}) \times \widehat{\mathbf{d}}_{\mathbb{I}}(i)}{\sum_i \mathcal{K}_{\mathcal{G}}(\mathbf{y}_{\mathbb{I}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{I}}) \times \widehat{\mathbf{d}}_{\mathbb{R}}(i) + \sum_i \mathcal{K}_{\mathcal{G}}(\mathbf{y}_{\mathbb{I}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{R}}) \times \widehat{\mathbf{d}}_{\mathbb{I}}(i)}.\end{aligned} \tag{24}$$

With the computed $\widehat{\mathbf{y}}_t^{(l)}$ at the l-th itearation, compute

$$\begin{aligned}(\mathbf{y}_{\Omega^c})_{\mathbb{R}} &= \frac{\sum_i((\widehat{\mathbf{y}}_{\tau(i)})_{\Omega^c})_{\mathbb{R}} \times \mathcal{K}_{\mathcal{G}}\left((\widehat{\mathbf{y}}_t^{(l)})_{\mathbb{R}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{R}}\right) \times \widehat{\mathbf{d}}_{\mathbb{R}}(i) - \sum_i((\widehat{\mathbf{y}}_{\tau(i)})_{\Omega^c})_{\mathbb{I}} \times \mathcal{K}_{\mathcal{G}}\left((\widehat{\mathbf{y}}_t^{(l)})_{\mathbb{R}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{I}}\right) \times \widehat{\mathbf{d}}_{\mathbb{I}}(i)}{\sum_i \mathcal{K}_{\mathcal{G}}\left((\widehat{\mathbf{y}}_t^{(l)})_{\mathbb{R}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{R}}\right) \times \widehat{\mathbf{d}}_{\mathbb{R}}(i) - \sum_i \mathcal{K}_{\mathcal{G}}\left((\widehat{\mathbf{y}}_t^{(l)})_{\mathbb{R}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{I}}\right) \times \widehat{\mathbf{d}}_{\mathbb{I}}(i)}, \\ (\mathbf{y}_{\Omega^c})_{\mathbb{I}} &= \frac{\sum_i((\widehat{\mathbf{y}}_{\tau(i)})_{\Omega^c})_{\mathbb{I}} \times \mathcal{K}_{\mathcal{G}}\left((\widehat{\mathbf{y}}_t^{(l)})_{\mathbb{I}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{I}}\right) \times \widehat{\mathbf{d}}_{\mathbb{R}}(i) + \sum_i((\widehat{\mathbf{y}}_{\tau(i)})_{\Omega^c})_{\mathbb{R}} \times \mathcal{K}_{\mathcal{G}}\left((\widehat{\mathbf{y}}_t^{(l)})_{\mathbb{I}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{R}}\right) \times \widehat{\mathbf{d}}_{\mathbb{I}}(i)}{\sum_i \mathcal{K}_{\mathcal{G}}\left((\widehat{\mathbf{y}}_t^{(l)})_{\mathbb{I}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{I}}\right) \times \widehat{\mathbf{d}}_{\mathbb{R}}(i) + \sum_i \mathcal{K}_{\mathcal{G}}\left((\widehat{\mathbf{y}}_t^{(l)})_{\mathbb{I}}, (\widehat{\mathbf{y}}_{\tau(i)})_{\mathbb{R}}\right) \times \widehat{\mathbf{d}}_{\mathbb{I}}(i)};\end{aligned} \tag{25}$$

update $\widehat{\mathbf{y}}_t^{(l+1)}$ as: $(\widehat{\mathbf{y}}_t^{(l+1)})_{\Omega^c} = (\mathbf{y}_{\Omega^c})_{\mathbb{R}} + \mathrm{j}\cdot(\mathbf{y}_{\Omega^c})_{\mathbb{I}}$; $\quad(\widehat{\mathbf{y}}_t^{(l+1)})_{\Omega} = (\widehat{\mathbf{y}}_t)_{\Omega}$.

Fig. 7 illustrates the four data loss modes, where $m = 6$, $k = 3$, and $d = 2$ in the diagram. Given a data loss percentage ϵ, $mT\epsilon$ points in the ground-truth measurement block $\mathbf{Y} \in \mathbb{C}^{m\times T}$ are erased following the given data loss mode to simulate the missing data. Modes 2-4 are introduced to simulate the extreme scenarios of simultaneous and consecutive data losses. The estimation error is evaluated by $\|\widehat{\mathbf{Y}} - \mathbf{Y}\|_F / \|\mathbf{Y}\|_F$. Each result is averaged over 50 independent runs.

In the following simulations, if not otherwise specified, $L = 10, \kappa = 6, \gamma$ in (11) is set as 5, λ in (17) is set as 10^{-3}, and $\delta = 10^{-3}$. We do not delete data in the first L instants. We compute the average norm of $\mathbf{y}_1$ to $\mathbf{y}_L$ and use it to normalize the measurement vectors at the remaining instants. Since no modeling of the system is needed in the paper and most of the methods mentioned in Section I cannot handle the cases of modes $2 \sim 4$, here the missing data are estimated with the following five methods:

1) **Duplication**: the missing points in $\mathbf{y}_t$ are filled in with the corresponding entries in $\widehat{\mathbf{y}}_{t-1}$;
2) **Low-rank Hankel structure**: the algorithm in [18] that exploits the low-rank Hankel structure of $\mathbf{Y}$. The approximation error threshold to determine the rank of the constructed Hankel matrix is set as 0.01;
3) **Low-rank lifted Hankel structure**: our proposed Algorithm without an initial estimation of $\widetilde{\mathbf{y}}_t$;
4) **Low-rank lifted Hankel structure**: our proposed Algorithm with an initial estimation of $\widetilde{\mathbf{y}}_t$. We fix $r = 1$ in estimating $\widetilde{\mathbf{y}}_t$ from (16).
5) **Kalman filter**: the method proposed in [7], where the measurements of each channel are estimated individually.

A. Missing data recovery of voltage phasors

We use the recorded data shown in Fig. 2. In Mode 4, $k = 11$ and $d = 5$ in the experiment. The results are shown in Fig. 8. Sorted by the increasing order of average estimation errors, these first four methods are, from top to the bottom, our low-rank lifted Hankel algorithm with $\widetilde{\mathbf{y}}_t$ estimation, our low-rank lifted Hankel algorithm without $\widetilde{\mathbf{y}}_t$ estimation, the low-rank Hankel method [18], and duplication. The Kalman filter method performs well only when the data loss percentage is not over 20%, and this is consistent with the conclusion in [7]. In contrast, our proposed method has a more stable performance and can estimate the missing entries with a high accuracy even when data losses occur more severely.

Fig. 9 shows one example of partially observed data with 5% consecutive data erasures in Mode 3. We compare the estimated data using the low-rank Hankel approach in [18] and our Algorithm 1 with $\widetilde{\mathbf{y}}_t$ estimation. The estimated data by [18] follow the dominant linear trend in the data before the data losses, causing an increasing gap between the estimation and the actual values. The average estimation error is 0.58%. In contrast, no evident gap is observed by using Algorithm 1, and the estimation error is 0.15%. If we focus on the last time instant of consecutive data losses when the error is most significant, the error at that time instant t^*, denoted by $\|\widehat{\mathbf{y}}_{t^*} - \mathbf{y}_{t^*}\|/\|\mathbf{y}_{t^*}\|$, is 3.53% for the approach in [18] and 1.04% for our Algorithm 1, respectively.

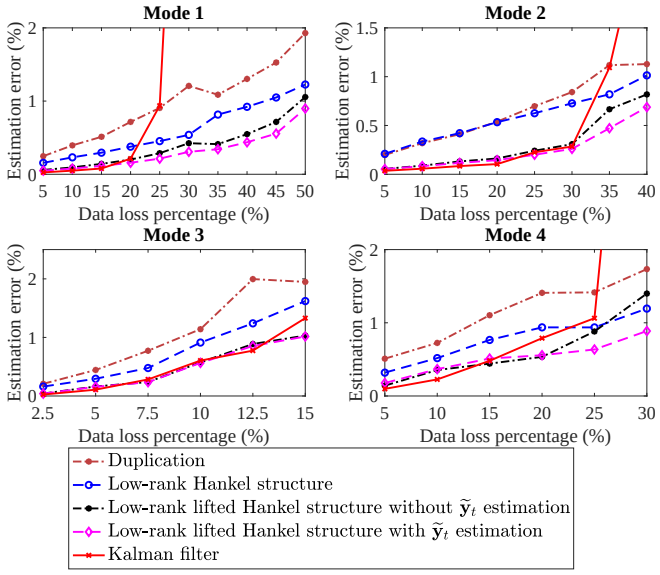


Fig. 8: Estimation errors of voltage phasors

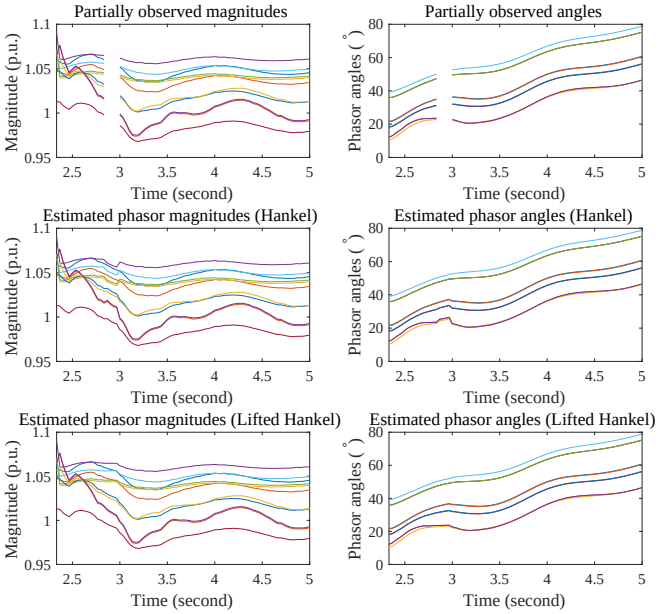


Fig. 9: Partially observed and estimated voltage phasors with 5% data losses in Mode 3

B. Missing data recovery of current phasors

We also test on the current phasors corresponding to the event in Fig. 1. As shown in Fig. 10, 22 current phasors are measured at 30 samples per second. In Mode 4, $k = 22$ and $d = 5$. The results are shown in Fig. 11. Algorithm 1 with $\widetilde{\mathbf{y}}_t$ estimation generally achieves the best performance.

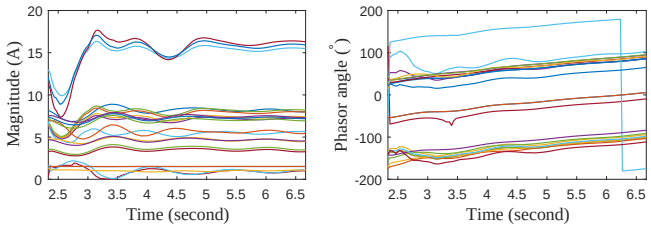


Fig. 10: Current phasors under system disturbances

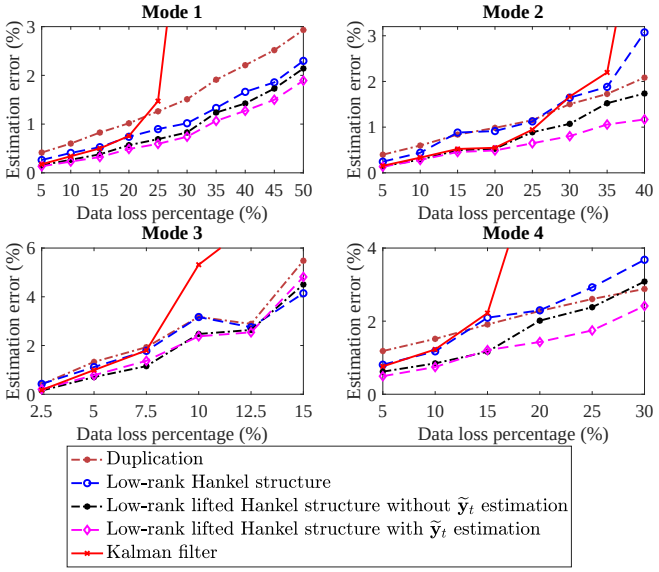


Fig. 11: Estimation errors of current phasors

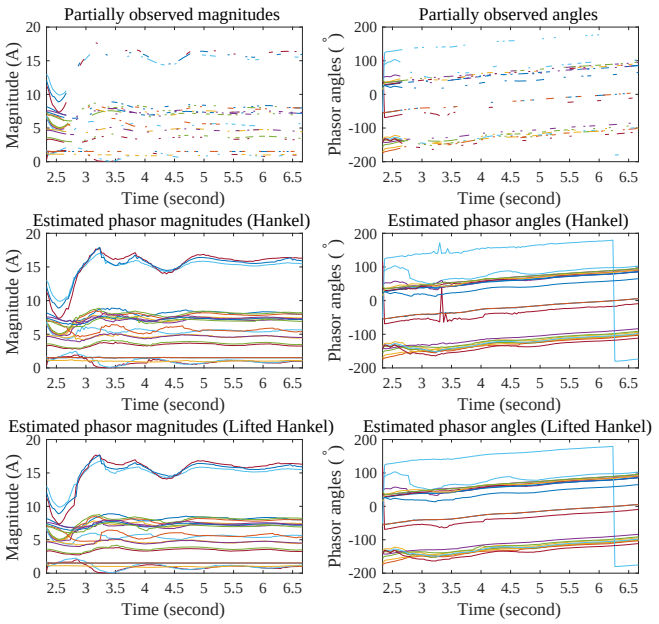


Fig. 12: Partially observed and estimated current phasors

Fig. 12 shows one example of the observed data with 50% data losses in Mode 1 and the estimated data from the approach in [18] and our Algorithm 1. The average estimation errors are 1.93% and 1.51%, respectively. The low-rank Hankel approach in [18] results in significant errors occasionally due to over-fitting, as shown in Fig. 12. More specifically, for a fixed approximation error threshold, the resulting rank r under the linear model can be larger than the number of observations at a given time instant when the data loss ratio is high. One can address the over-fitting issue by increasing the approximation error threshold, but that leads to an increase in the recovery error. In contrast, our methods using the lifted Hankel model do not have the over-fitting issue.

C. Performance evaluation on the frequency f and the rate of change of frequency (ROCOF)

Considering the frequency f and ROCOF are import indicators about the operation status of power systems, we further compute them from the estimated phasor angels and compare them with the ground-truth data. The computation of f and ROCOF follows the formulas provided in [26], i.e.,

$$\begin{aligned}\Delta f_t &= \frac{6(\theta_t-\theta_{t-1})+3(\theta_{t-1}-\theta_{t-2})+(\theta_{t-2}-\theta_{t-3})}{20\pi\times\Delta t},\\ f_t &= 60+\Delta f_t,\\ \text{ROCOF} &= \frac{\Delta f_t-\Delta f_{t-1}}{\Delta t},\end{aligned} \tag{29}$$

where θ_t, f_t denotes the phasor angle and frequency at instant t, respectively. Due to the sampling rate of 30 samples per channel per second, $\Delta t = 1/30$ second.

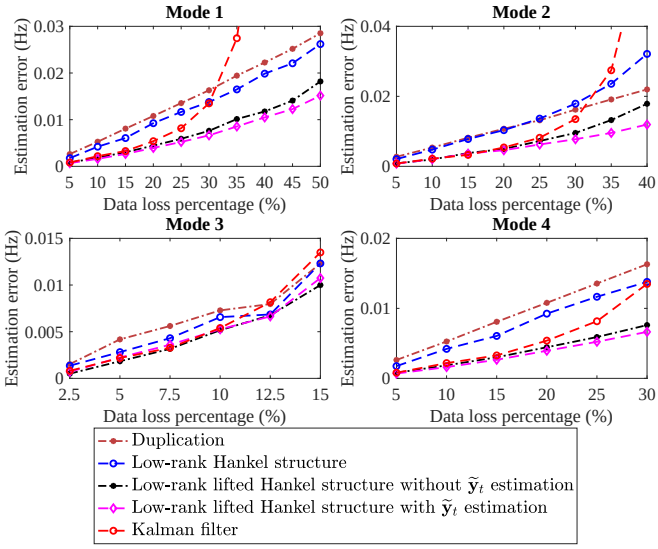


Fig. 13: Performance comparison of frequency estimation from current phasors

To evaluate the performance, we compute the average error of $\sum_{t=1}^{T}\left(|f_t-\widehat{f}_t|\right)/T$ over all the channels and 50 trials. Similarly, the estimation performance of ROCOF is also evaluated in terms of average error. The results based on the estimated current phasors are shown in Figs. 13 and 14, where the curves in the two figures are much similar, and in most cases, the data recovery from our proposed method with $\widetilde{\mathbf{y}}_t$ estimation have the best performance. Thus, though the data recovery performance in Figs. 11, 13 and 14 are evaluated in different aspects, they all support the conclusion that our proposed algorithm achieves a better performance.

D. Influence of parameters on the data recovery performance

As presented in Sections IV-A and IV-B, several key parameters, like κ, L, λ and δ, are determined before the implementation of this proposed method. Here we discuss the influence of these parameters on the missing data recovery performance. The results are compared over the same missing

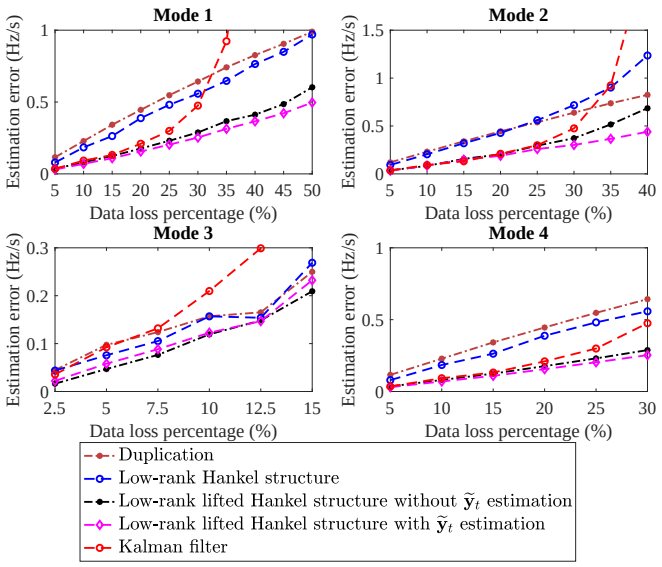


Fig. 14: Performance comparison of ROCOF estimation from current phasors

voltage phasors in Mode 1. Other parameters in Algorithm 1 remain the same.

1) Parameter κ: With L, λ and δ fixed as 10, 10^{-3} and 10^{-3}, the recovery performance with different κ is demonstrated in Fig. 15. We can find when κ is around $L/2$, the estimation achieves the best performance, which is consistent with the previous suggestions in [17], [18].

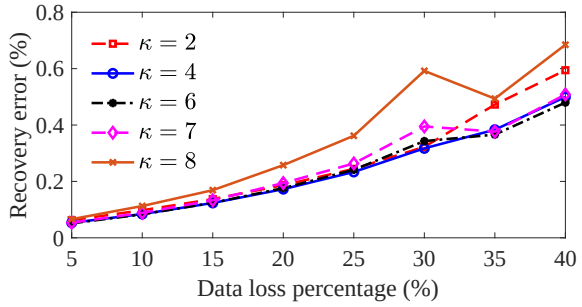


Fig. 15: Estimation errors of voltage phasors with varying κ

2) Parameter window length L: Here we fix κ, λ as 6, 10^{-3} and 10^{-3}, respectively. The estimation errors with the varying window length L are shown in Fig. 16, where the algorithm achieves the best performance when $L \approx 2\kappa$. The reason is that Hankel matrix $\mathcal{H}_\kappa(\widehat{\mathbf{Z}}^\Phi)$ in Fig. 6 has $L-\kappa+1$ columns, though a larger L helps reduce the effect of noise on the subspace estimation, it tends to have more obsolete transient components in the past measurements, which affects the data recovery accuracy.

3) Parameter λ: Here we keep κ as 6, L as 10, and δ as 10^{-3}. Fig. 17 demonstrates how the performance is affected by regularization weight λ. The data estimation performance does not change much in a wide range of λ from 10^{-6} to 10^{-4}. Then as λ increases, the error goes up, which means λ is larger than necessary. The reason is that the synchrophasor measurements have a high signal-to-noise ratio, thus a small value of λ, like 10^{-4}, is appropriate in practical applications.

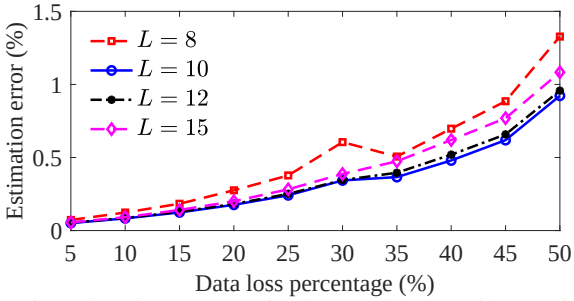


Fig. 16: Estimation errors of voltage phasors with varying L

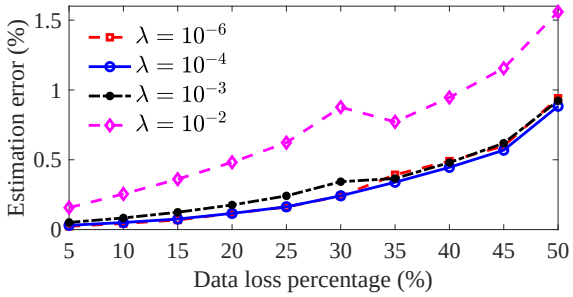


Fig. 17: Estimation errors of voltage phasors with varying λ

4) Parameter δ: Parameters κ, L and λ are kept as $6, 10$ and 10^{-3}, respectively. Fig. 18 shows the result with δ varying from 10^{-1} to 10^{-8}, where no evident difference in the estimation error is observed.

Fig. 19 demonstrates the average number of iterations needed when δ and missing data ratio changes in Mode 1. Here we consider voltage phasors, and the result is averaged over 50 trials. When the data loss ratio changes, the number of iterations is always 3 with $\delta = 10^{-8}$. Moreover, considering the stopping criteria is $\|\widehat{\mathbf{y}}_t^{(l)} - \widehat{\mathbf{y}}_t^{(l-1)}\| \le \delta$, we find $\|\widehat{\mathbf{y}}_t^{(3)} - \widehat{\mathbf{y}}_t^{(2)}\| \ll \|\widehat{\mathbf{y}}_t^{(2)} - \widehat{\mathbf{y}}_t^{(1)}\| \ll \|\widehat{\mathbf{y}}_t^{(1)} - \widehat{\mathbf{y}}_t^{(0)}\|$ holds in our numerical simulations. Thus, though a smaller δ corresponds to more iterations in theory, the accuracy is still close to the case of $\delta = 10^{-1}$ in our simulation. The same phenomena hold in voltage phasors of other modes and current phasors as well, i.e., the iterates always converge to stationary points in a few iterations, and the performance does not change much with varying δ.

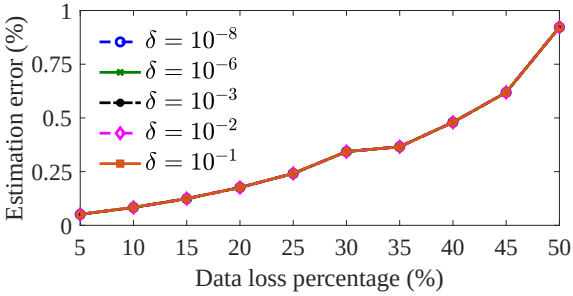


Fig. 18: Estimation errors of voltage phasors with varying δ

E. Algorithm robustness and running time

1) The number of channels: Here we test the performance of our algorithm with the incomplete voltage phasors in Mode 1. When we change the number of channels processed simultaneously, all the parameters remain as the original values.

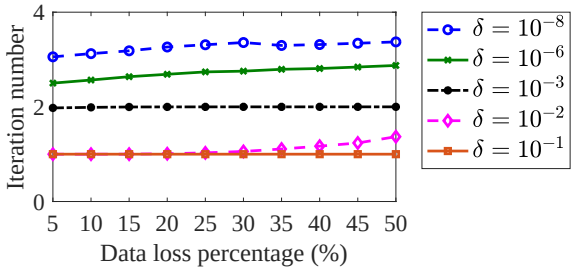


Fig. 19: Average number of iterations with varying δ

We pick channel 1 and randomly select $m-1$ other channels. We estimate the missing data on these m channels together. Fig. 20 demonstrates the average estimation error of channel 1 over 50 trials.

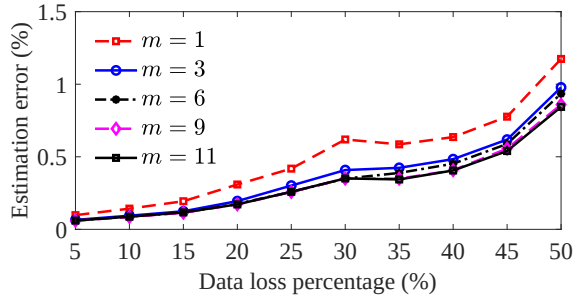


Fig. 20: Estimation errors of voltage phasors in channel 1 with a varying total number of channels

We can find that recovering more channels simultaneously helps improve the estimation accuracy, if they are deployed close to each other such that a disturbance affects the measurements of all channels. The correlations among channels help improve the estimation performance. Even if we estimate the missing entries with only the measurements on one channel by itself, the performance is still acceptable.

2) Average running time: Section III-C analyzes the complexity of the proposed algorithm. Here we record the running time of the simulation over 50 trials and compute the average time filling in the missing entries on each instant, as shown in Fig. 21. Based on the theoretical analysis in Section III-C, the average time should be proportional to the number of measurement channels and not related to the data loss percentages. Though the running time of 22 current phasors is not exactly twice of the time of 11 voltage phasors, the ratio always keeps around 1.5. As data loss percentage increases, the slight increase of the running time in Fig. 21 results from the increasing number of instants at which the measurements are not fully observed. Note that the processing time is below 1 ms, far less than 1/30 s, which means the algorithm can estimate the missing data in the current samples well before the arrival of the measurements in the next instant. Thus, the algorithm can be implemented to improve the quality of streaming synchrophasor measurements in real time.

V. Conclusion and Discussions

This paper presents an online algorithm to estimate missing data points of streaming synchrophasor measurements, especially under severe disturbances. The key idea is to characterize the dynamics of a non-linear system through the low-rank

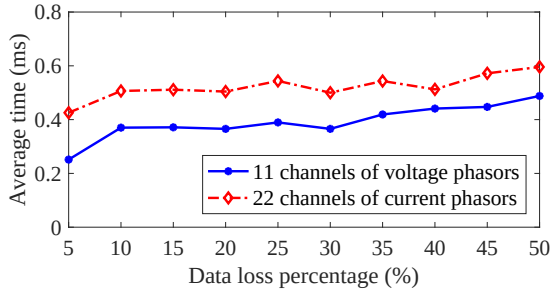


Fig. 21: Average running time in mode 1

Hankel property in the lifted space. The resulting data-driven method does not require the modeling the nonlinear system and can be implemented efficiently in real time. Compared to existing algorithms, the algorithm can handle the extreme cases of data losses with superior performance.

Acknowledgment

This research is supported by NSF Grants #1508875, #1736326, and #1932196, 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 for providing recorded synchrophasor datasets.

References

[1] V. V. Terzija *et al.*, "Wide-area monitoring, protection, and control of future electric power networks." *Proc. IEEE*, vol. 99, no. 1, pp. 80–93, 2011.

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

[3] California ISO, "Five year synchrophasor plan," https://www.caiso.com/Documents/FiveYearSynchrophasorPlan.pdf, 2011.

[4] S. G. Ghiocel *et al.*, "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.

[5] J. Zhao, M. Netto, and L. Mili, "A robust iterated extended kalman filter for power system dynamic state estimation," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3205–3216, 2017.

[6] M. Rahman and M. Islam, "Missing value imputation using decision trees and decision forests by splitting and merging records: Two novel techniques," *Knowledge-Based System*, vol. 53, pp. 51–65, 2013.

[7] K. D. Jones, A. Pal, and J. S. Thorp, "Methodology for performing synchrophasor data conditioning and validation," *IEEE Trans. Power Syst.*, vol. 30, no. 3, pp. 1121–1130, 2015.

[8] F. Gao, J. Thorp, A. Pal, and S. Gao, "Dynamic state prediction based on Auto-Regressive (AR) model using PMU data," in *Proc. IEEE Power and Energy Conference at Illinois (PECI)*, 2012, pp. 1–5.

[9] J. Landford *et al.*, "Fast sequence component analysis for attack detection in smart grid," in *2016 5th International Conference on Smart Cities and Green ICT Systems (SMARTGREENS)*, April 2016, pp. 1–8.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] E. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053 –2080, 2010.
[15] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proc. 45th Annu. ACM Symp. Theory of Comput.*, 2013, pp. 665–674.
[16] 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.
[17] ——, "Multi-channel hankel matrix completion through nonconvex optimization," *IEEE J. Sel. Topics Signal Process., Special Issue on Signal and Information Processing for Critical Infrastructures*, vol. 12, no. 4, pp. 617–632, 2018.
[18] Y. Hao, M. Wang, J. H. Chow, E. Farantatos, and M. Patel, "Modelless data quality improvement of streaming synchrophasor measurements by exploiting the low-rank hankel structure," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6966–6977, 2018.
[19] S. Zhang and M. Wang, "Correction of simultaneous bad measurements by exploiting the low-rank hankel structure," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2018, pp. 646–650.
[20] Y. Susuki and I. Mezić, "Nonlinear koopman modes and power system stability assessment without models," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 899–907, 2014.
[21] E. Barocio, B. C. Pal, N. F. Thornhill, and A. R. Messina, "A dynamic mode decomposition framework for global power system oscillation analysis," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 2902–2912, 2015.
[22] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *Ann. Stat.*, pp. 1171–1220, 2008.
[23] L. Ralaivola *et al.*, "Dynamical modeling with kernels for nonlinear time series prediction," in *Adv.Neural Inf. Process. Syst.*, 2004, pp. 129–136.
[24] F. A. Tobar, A. Kuh, and D. P. Mandic, "A novel augmented complex valued kernel LMS," in *Proc. 7th Sensor Array and Multichannel Signal Process. Workshop (SAM)*, 2012, pp. 473–476.
[25] S. Mika, B. Schölkopf, A. J. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel pca and de-noising in feature spaces," in *Adv. Neural Inf. Process. Syst.*, 1999, pp. 536–542.
[26] IEEE Std C37.118.1-2011 (Revision of IEEE Std C37.118-2005), "IEEE standard for synchrophasor measurements for power systems," pp. 1–61, Dec 2011.

APPENDIX: MOTIVATION OF USING GAUSSIAN KERNEL FOR NONLINEAR SYSTEMS

To simplify our analysis, here we assume $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^m$ and consider the real Gaussian kernel in (12) to simplify the analysis. Without loss of generality, let $\gamma = 1/2$, then based on Taylor expansion,

$$\begin{aligned}
&\mathcal{K}_{\mathcal{G}}(\mathbf{y}_1, \mathbf{y}_2) = \exp(-\frac{1}{2}\|\mathbf{y}_1 - \mathbf{y}_2\|_2^2) \\
&= \exp(-\frac{1}{2}\|\mathbf{y}_1\|_2^2) \cdot \exp(\mathbf{y}_1^T \mathbf{y}_2) \cdot \exp(-\frac{1}{2}\|\mathbf{y}_2\|_2^2) \\
&= \exp(-\frac{1}{2}\|\mathbf{y}_1\|_2^2) \cdot \sum_{k=0}^{\infty} \frac{(\sum_{i=1}^m \mathbf{y}_1(i)\mathbf{y}_2(i))^k}{k!} \cdot \exp(-\frac{1}{2}\|\mathbf{y}_2\|_2^2).
\end{aligned}$$

Note that

$$\begin{aligned}
&\left(\sum_{i=1}^m \mathbf{y}_1(i)\mathbf{y}_2(i)\right)^k = \sum_{\sum n_i = k} \prod_{i=1}^m \binom{k}{n_i} (\mathbf{y}_1(i)\mathbf{y}_2(i))^{n_i} \\
&= \sum_{\sum n_i = k} \binom{k}{n_1, \cdots, n_m} \mathbf{y}_1(1)^{n_1} \cdots \mathbf{y}_1(m)^{n_m} \mathbf{y}_2(1)^{n_1} \cdots \mathbf{y}_2(m)^{n_m},
\end{aligned}$$

If $\Phi(\mathbf{y}) \in \mathbb{R}^\infty$ consists of all the entries in the form of

$$\exp(-\frac{1}{2}\|\mathbf{y}\|_2^2) \binom{k}{n_1, \cdots, n_m}^{1/2} (k!)^{-1/2} \mathbf{y}(1)^{n_1} \cdots \mathbf{y}(m)^{n_m}$$

with $\sum_{i=1}^m n_i = k$ and $k = 0, \cdots, \infty$, then we have

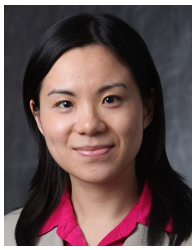
$$\langle \Phi(\mathbf{y}_1), \Phi(\mathbf{y}_2) \rangle = \exp(-\frac{1}{2}\|\mathbf{y}_1 - \mathbf{y}_2\|_2^2) = \mathcal{K}_{\mathcal{G}}(\mathbf{y}_1, \mathbf{y}_2).$$

Then $\Phi(\mathbf{y})$ is a vector with infinity dimension and consists of all the possible monomials that can be constructed with entries in $\mathbf{y}$. Each monomial is scaled by $\exp(-\frac{1}{2}\|\mathbf{y}\|_2^2)\binom{k}{n_1, \cdots, n_m}^{1/2}(k!)^{-1/2}$, where $\sum_{i=1}^m n_i = k$.

Since we assume $\mathbf{x}_t$ can be determined from $\mathbf{y}_t$, then from (5), we can find a function $h(\cdot)$ such that $\mathbf{y}_{t+1} = h(\mathbf{y}_t)$. Applying the Taylor expansion to $h(\cdot)$, one can express it as $\mathbf{y}_{t+1} = \widehat{\mathbf{D}}\Phi(\mathbf{y}_t)$ for some matrix $\widehat{\mathbf{D}}$, because $\Phi(\mathbf{y}_t)$ includes all the possible monomials with entries in $\mathbf{y}_t$. Since $\Phi(\mathbf{y}_{t+1})$ includes all the monomials with entries in $\mathbf{y}_{t+1}$, we finally obtain that $\Phi(\mathbf{y}_{t+1}) = \mathbf{D}\Phi(\mathbf{y}_t)$, where $\mathbf{D}$ is in $\mathbb{R}^{\infty \times \infty}$.

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 received 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 Associate 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.