

Online Voltage Event Detection Using Synchronphasor Data with Structured Sparsity-Inducing Norms

Xianghao Kong, *Student Member, IEEE*, Brandon Foggo, *Member, IEEE*, Koji Yamashita, *Member, IEEE*, Nanpeng Yu, *Senior Member, IEEE*

Abstract—This paper develops an accurate and computationally efficient data-driven framework to detect voltage events from PMU data streams. It develops an innovative Proximal Bilateral Random Projection (PBRP) algorithm to quickly decompose the PMU data matrix into a low-rank matrix, a row-sparse event-pattern matrix and a noise matrix. The row-sparse pattern matrix significantly distinguishes events from normal behavior. These matrices are then fed into a clustering algorithm to separate voltage events from normal operating conditions. Large-scale numerical study results on real-world PMU data show that the proposed algorithm is computationally more efficient and achieves higher F scores than state-of-the-art benchmarks.

Index Terms—Phasor measurement unit (PMU), event detection, low-rank and sparse matrix decomposition, bilateral random projection.

I. INTRODUCTION

PHASOR Measurement Units (PMUs) are capable of recording both amplitudes and angles of voltage and current phasors. PMUs have been widely deployed in transmission networks worldwide to improve the situational awareness of power system operators [1]. The rapid adoption of PMUs in transmission grids has led to tremendous growth in the amount of synchronphasor data. For example, with a 60 Hz sampling rate, each PMU can generate more than one gigabyte of data per day [2]. The wealth of PMU data enables the development of various data-driven applications to better monitor, protect, and control power systems. In particular, data-driven event detection algorithms are critical to making system operators aware of abnormal system conditions [3]. The early detection of power system events enables the operators to take corrective control actions in response to disturbance events.

The existing literature on the data-driven power system event detection using synchronphasor data can be clustered

X. Kong, B. Foggo, K. Yamashita, and N. Yu are with Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521, USA. Email: {xkong016@, bfogg001@, nyu@ece, kyamashi@}ucr.edu.

Disclaimer: this report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

into three groups. The first group of literature leverages signal processing techniques and statistical analysis to detect system events. Signal processing techniques such as discrete wavelet decomposition [3], dissipating energy flow [4], empirical mode decomposition [5], self-coherence spectrum [6], Teager-Kaiser energy operator [7], dynamic programming-based swinging door trending [8], graph signal processing [9] [10], and parallel detrended fluctuation analysis [11] are adopted to detect oscillation events [6], voltage events [3], [5], [9], frequency events [3], [5], [9], [11] and sequence of voltage and frequency events [7]. A multi-hypothesis statistical testing framework is developed in [12] to detect power line outages. This group of techniques has achieved great success at various types of power system events. However, most signal processing techniques analyze PMU data streams collected from different locations separately and did not fully exploit underlying spatial-temporal correlations in the PMU dataset.

The second group of papers adopts deep learning techniques to detect and classify abnormal events. Convolution neural networks (CNN)-based classifiers [13] are built using the rate of change of frequency and relative angle shift signals to detect and classify generator trip and load disconnection events. A novel information loading based regularization and a graph signal processing-based PMU sorting algorithm were developed to improve the parameter sharing scheme of the CNN framework [14]. An ensemble-based learning algorithm, combining multiple machine learning algorithms, is proposed in [15]. Despite the high accuracy achieved by deep learning techniques, their success heavily depends on the availability of numerous high-quality event labels rarely available in practice.

The third research group leverages the low-dimensionality and approximates low-rank properties of PMU data to detect power system events. Recognizing that high dimensional PMU data lie close to a low dimensional manifold, principle component analysis (PCA)-based event detection algorithms have been developed [16]. Two statistics derived from a moving window PCA on PMU data matrices are used to detect frequency and islanding events [17]. It has been shown that pilot PMUs identified by the PCA can be used to approximate non-pilot PMUs' data streams. An event alert will be issued when the normalized approximation error is larger than a pre-specified threshold [18]. In [19], a PCA-based method is developed to measure the similarity of operation states between a pair of buses, and the k -reachability is adopted to detect power system events. By exploiting the approximate low-rank

property of PMU data matrices, subspace characterization [20] and matrix completion-based [21], [22] approaches have been proposed to detect power system events. The PCA and matrix completion-based event detection algorithms model PMU data as the sum of a low-rank matrix and a noise matrix. They overcome the shortcomings of algorithms in the first two groups and do not require a large number of event labels for training. However, the PMU data decomposition methods in the third group are oversimplified and ignore the unique structural pattern of events.

Meanwhile, sparsity-inducing norms have been widely adopted in other power grid fields. Examples include malicious cyber attack detection [23]–[25], imbalance identification [26] and line outage detection [27]. Routtenberg et al. [26] successfully leveraged the sparse structure in voltage measurements to localize imbalances in the power grid. Zhu et al. [27] leverage sparse overcomplete representations with l_1 norms to identify sparse power line outages. Liu et al. [24] propose a novel matrix decomposition method based on l_1 norm to detect the malicious attacks in the power grid. Hao et al. [23] assume the measurement matrix can be recovered by a low-rank matrix and a sparse-attack matrix and apply l_1 norm to induce sparsity. The l_1 norm only constraints a matrix to be element-wise sparse, i.e., non-zero elements appear randomly in the matrix. This sparsity-inducing constraint is too weak to cater a structured sparsity, like row-sparse or column-sparse. To break this limitation, Gao et al. [25] proposed an attack identification algorithm with the l_{21} norm, assuming that the measurement data matrix is a low-rank matrix plus a transformed column-sparse matrix. This time, the sparse matrix is not directly stripped from the original measurement matrix, but needs to be multiplied by a known transform matrix. To construct the transform matrix, we need to know the structure of the power grid in advance, as well as the impedance and admittance between different buses, which is hard to obtain in practice. Most of the algorithms in this group rely on singular value decomposition (SVD), whose computation time drastically increases with the number of PMUs and the analysis window length. This drawback dramatically limits the scalability of the event detection algorithm. Our research develops a novel method that decomposes the PMU measurement matrix into a low-rank matrix, a sparse matrix, and a noise matrix without a transform matrix in a computationally efficient manner.

Mining the real-world data gathered by hundreds of PMUs, covering thousands of events across the U.S., discovers that regional events mostly show a unique sparsity in the noise matrix (Fig. 2). As the row of the noise matrix corresponds to individual PMUs, the sparsity emerges in its row space, depending on how sensitive a PMU is to events. Among a wide variety of power system events, voltage-related events mainly triggered by system faults are recognized as a regional event. Therefore, the row-sparse property in the noise matrix can exert an effect, especially in an event with a significant voltage dip. This paper proposes further decomposing the noise matrix into a row-sparse event-pattern matrix and a pure noise matrix, in light of the above. This innovative low-rank and sparse matrix decomposition framework, extracting anomaly features from both the low-rank matrix and the row-sparse event-

pattern matrix, enables event alerts. Finally, an unsupervised clustering technique is adopted to distinguish normal system operation data from that of the power system voltage events.

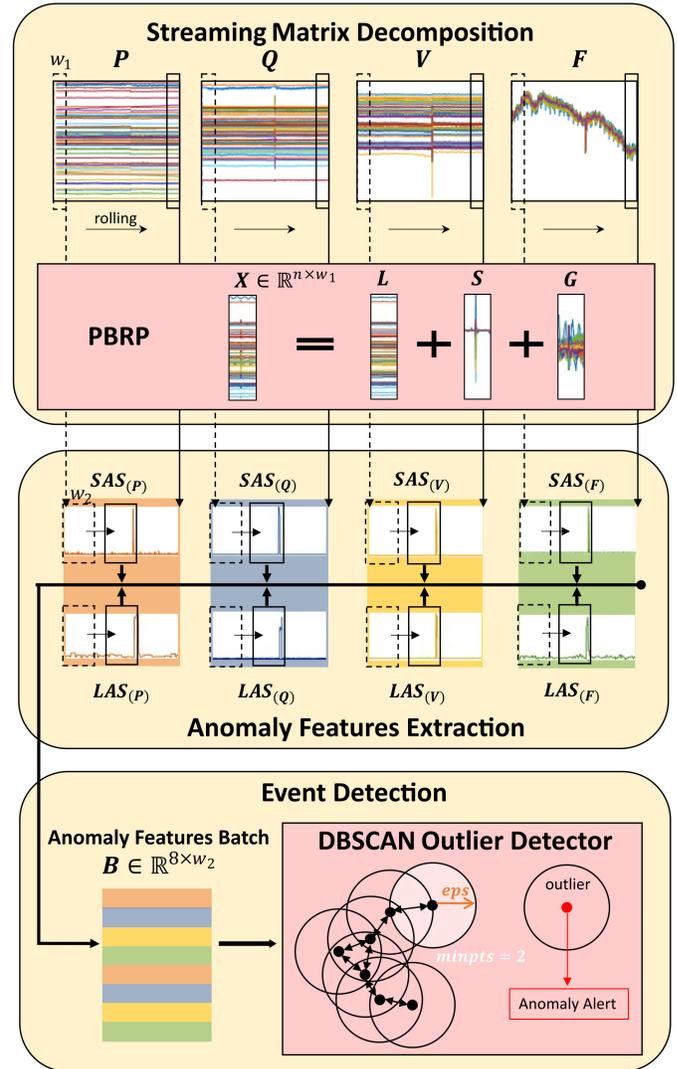


Fig. 1. Overview of online event detection framework based on low-rank and sparse matrix decomposition model.

Unique contributions of this work include:

- 1) The proposed iterative matrix decomposition approach, PBRP, which greatly accelerates the solution of a general low-rank and sparse matrix decomposition problem where the residual matrix has a row-sparse structure.
- 2) The data-driven event detection framework based on PBRP yields better theoretical and empirical computation efficiency than existing SVD-based subspace characterization approaches
- 3) Unlike the model-based or supervised deep learning methods, our proposed voltage event detection algorithm does not rely on detailed physical model or a large amount of event labels for training.
- 4) The numerical study on a large-scale real-world PMU dataset with hundreds of PMUs and voltage events shows that the PBRP-based event detection framework provides higher F1 and F2 scores than state-of-the-art

algorithms. The proposed algorithm can also estimate event area/location along with PMUs that are sensitive to an event.

II. TECHNICAL METHODS

A. Overall Framework

The overall framework of the proposed online voltage event detection algorithm is summarized in Fig. 1. The proposed algorithm has three modules: a streaming matrix decomposition module, an anomaly feature extraction module, and an event detection module based on cluster analysis. The first module separately decomposes four types of streaming PMU data matrices in w_1 -length windows. Each type of data matrix (X) would be decomposed into a low-rank matrix (L), a sparse event-pattern matrix (S), and a noise matrix (G) via PBRP algorithm. The second module extracts useful features, two anomaly scores, LAS and SAS from decomposed matrices, L and S . The third module performs cluster analysis on extracted features within w_2 -length windows to identify anomalies.

The technical details of these three modules are presented in the next few subsections. Subsection II-B provides the formulation and iterative solution approach to the low-rank and sparse event-pattern matrix decomposition problem. Subsection II-C presents the bilateral random projection technique to reconstruct the low-rank matrix. Subsection II-D develops the proximal method to solve the sparse matrix. The anomaly feature extraction method and the clustering algorithm are described in subsection II-F.

B. Low-Rank and Sparse Event-Pattern Matrix Decomposition

Let n denote the number of PMUs under consideration. We collect the streaming PMU data into a matrix time series $X_t \in \mathbb{R}^{n \times w}$ by placing new instances of data in the rightmost column while removing the leftmost column. These matrices are decomposed in a specific way based on prior knowledge of their properties.

The first property of note has to do with the rank of these matrices. As shown by small-scale empirical studies and theoretical derivations, voltage and current phasor data under normal conditions exhibits a low-rank structure [21]. Using large-scale PMU data from the Eastern Interconnection of the continental U.S. power transmission grid, the active power (P), reactive power (Q), voltage magnitude (V), and frequency (F) data does possess the low-rank property, which means the low-rank property holds up well during steady-state operation periods. In Table I, the analysis on a representative event shows that the largest singular value of the reactive power data matrix accounts for 99.988% of the variance, while this percentage drops to 59.743% during event periods. Suppose that this normal behavior from the streaming PMU data matrices is decomposed as:

$$X = L + (X - L), \quad (1)$$

where L is an approximation of X with rank r . Then the matrix, $X - L$, contains information from the data that is residual from normal behavior, which is a promising first step towards event detection.

The second property of note has to do with the structure of $X - L$ during voltage event periods. It turns out that these matrices have specific patterns of sparsity that we can take advantage of. The main component of this structure comes directly from the fact that voltage events, when they occur, often significantly affect limited area/zones. As such, the number of PMUs interacting with a voltage event is prone to be limited (Fig. 2). The right subfigure shows the heatmap of an event pattern matrix representing one-second min-max normalized active power data. Thus, we propose to decompose the PMU data matrix as follows:

$$X = L + S + G, \quad (2)$$

where $L \in \mathbb{R}^{n \times w}$ is a low-rank matrix, $S \in \mathbb{R}^{n \times w}$ is a row-sparse event-pattern matrix representing the impact of voltage events, and $G \in \mathbb{R}^{n \times w}$ denotes a noise matrix. Then, the problem can be formulated as:

$$\begin{aligned} \min_{L,S} \quad & \frac{1}{2} \|X - L - S\|_F^2 \\ \text{s.t.} \quad & \begin{cases} \text{rank}(L) = r, \\ S \text{ is row-sparse.} \end{cases} \end{aligned} \quad (3)$$

It turns out that this ‘‘row-sparse’’ event pattern can be captured by using l_{21} regularization, where the l_{21} norm of S is defined as $\|S\|_{21} = \sum_i \sqrt{\sum_j s_{ij}^2}$. In other words, by adding the l_{21} norm on event-related matrix, S , to the objective function as a penalty term, the solution will yield the desired row-sparse structure. The problem can then be relaxed into a new one:

$$\begin{aligned} \min_{L,S} \quad & \frac{1}{2} \|X - L - S\|_F^2 + \lambda \|S\|_{21} \\ \text{s.t.} \quad & \text{rank}(L) = r, \end{aligned} \quad (4)$$

where the λ is a penalty coefficient of the l_{21} norm. This optimization problem (4) can be solved with Coordinate Descent.

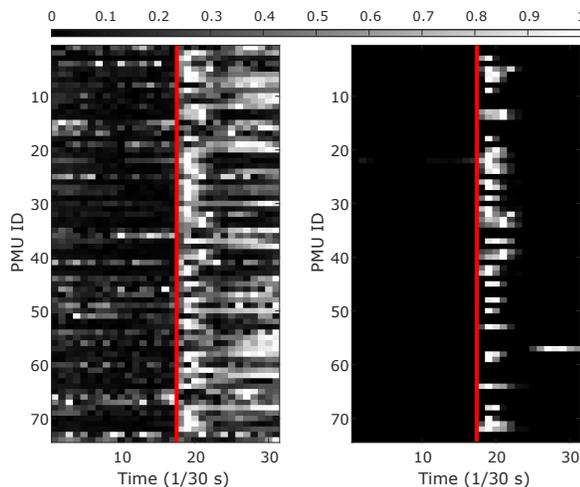


Fig. 2. The heatmap of ‘‘ $X - L$ ’’ (left) and ‘‘ $X - L - G$ ’’ (right) for normalized active power data (scaled from 0 to 1). The event happens approximately at the red line.

TABLE I
SINGULAR VALUE DECOMPOSITION OF P, Q, V, AND F DATA MATRICES OVER 1 SECOND (30 SAMPLES)

Data Type	Electrical Quantity	Singular Value Percentage Variance ($\frac{\sigma_i^2}{\sum_i \sigma_i^2}$)			Singular Value Proportion ($\frac{\sigma_i}{\sum_i \sigma_i}$)		
		1st	2nd	3rd	1st	2nd	3rd
Non-event Data	P (Active Power)	99.999261%	0.000536%	0.000109%	99.242040%	0.229736%	0.103609%
	Q (Reactive Power)	99.988472%	0.008789%	0.001427%	97.134683%	0.910695%	0.366895%
	V (Voltage Magnitude)	99.999995%	0.000005%	0.000000%	99.963393%	0.022302%	0.003274%
	F (Frequency)	99.999999%	0.000000%	0.000000%	99.996304%	0.000824%	0.000654%
Event Data	P (Active Power)	95.003242%	4.933310%	0.045058%	78.391273%	17.863547%	1.707207%
	Q (Reactive Power)	59.743182%	40.185730%	0.068034%	53.278058%	43.695845%	1.797904%
	V (Voltage Magnitude)	99.545736%	0.447828%	0.006371%	92.892350%	6.230519%	0.743139%
	F (Frequency)	99.999994%	0.000006%	0.000000%	99.971389%	0.023498%	0.001269%

This means that we alternate between solving the following two sub-problems until $\frac{\|X - L^{(k)} - S^{(k)}\|_F^2}{\|X\|_F^2}$ converges:

$$\begin{cases} L^{(k)} = \arg \min_{\text{rank}(L)=r} \frac{1}{2} \|X - L - S^{(k-1)}\|_F^2 \\ S^{(k)} = \arg \min_S \frac{1}{2} \|X - L^{(k)} - S\|_F^2 + \lambda \|S\|_{21} \end{cases} \quad (5)$$

where $L^{(k)}$ and $S^{(k)}$ denote an estimate of L and S , respectively, in the k -th iteration.

To solve the first sub-problem, we forgo time-consuming exact optimization and instead choose to approximate its solution via an enhanced version of Bilateral Random Projections (BRPs).

C. Bilateral Random Projections

Bilateral Random Projections (BRPs) are a fast and accurate method of low-rank matrix approximation. We showcase a new approximation approach here. First, consider the following column-row-echelon decomposition of a matrix $X \in \mathbb{R}^{n \times w}$ that has a rank r .

$$X = \underbrace{[\mathbf{c}_1 \quad \mathbf{c}_2 \quad \cdots \quad \mathbf{c}_r]}_C \underbrace{[I_r \mid E_{w-r}]}_E P, \quad (6)$$

where $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_r$ are any choice of r linearly independent columns of X , $I_r \in \mathbb{R}^{r \times r}$, $E_{w-r} \in \mathbb{R}^{r \times (w-r)}$ and $P \in \mathbb{R}^{w \times w}$ acts on E as a column permutation matrix (EP is the reduced row-echelon form of X , and P just moves the pivot columns of that echelon form back to their original positions). This decomposition expresses a matrix as a selection of r linearly independent columns from X and uses the matrix E_{w-r} to generate the remaining columns from the selected ones. We can also express this same decomposition as $X = (XA_1)EP$ where A_1 acts on X to select those r independent columns. If we then use a matrix A_2^T to select r linearly independent rows of X and collect them into a matrix H^T (via left-application, i.e., $H^T = A_2^T X$), then we have that

$$A_2^T X = A_2^T C E P. \quad (7)$$

The left hand side of (7) has rank r , and EP has rank r as well due to its I_r submatrix, so it must be the case that $A_2^T C$ has rank r as well. But since $A_2^T C$ has dimension $r \times r$, this

means that $A_2^T C$ is invertible. Thus $EP = (A_2^T C)^{-1} A_2^T X$. In other words:

$$X = CEP = C(A_2^T C)^{-1}(A_2^T X) = C(A_2^T C)^{-1} H^T. \quad (8)$$

To get from here to BRPs, all we need is relaxing the requirement that the matrices A_1 and A_2 choose r linearly independent columns/rows of X , and instead let them randomly choose subspaces via random linear combinations of the columns/rows of X . This can be done by simply drawing the elements of A_1 and A_2 randomly, according to some distribution (in this case, the standard normal distribution). The Johnson–Lindenstrauss lemma [28] ensures that such random selection well approximates the process of selecting independent rows and columns of X . In accordance with [29], we enhance this approximation algorithm with a *power scheme* technique. The idea of this enhancement is simple — instead of approximating X itself, $\tilde{X} = (X X^T)^q X$ for some integer $q \geq 1$ are approximated. This leads to a higher likelihood of randomly selecting the most important low-rank approximations because the matrix \tilde{X} , while having the same row and column spaces as X , has its singular values exponentiated to the power of $2q+1$, which leads to a much higher discrepancy between the larger and smaller singular values. Once we have an approximation of \tilde{X} given by $\tilde{X} \approx \tilde{L} = C(A_2^T C)^{-1} H^T$, we can use the same approximated subspaces of \tilde{X} to get an approximation of X via $X \approx L = \tilde{L}^{\frac{1}{2q+1}}$. This can be done efficiently by taking the QR decomposition of C and H , denoted $Q_{col} R_{col}$ and $Q_{row} R_{row}$ respectively, and computing

$$L = \tilde{L}^{\frac{1}{2q+1}} = Q_{col} [R_{col} (A_2^T C)^{-1} R_{row}^T]^{\frac{1}{2q+1}} Q_{row}^T \quad (9)$$

Finally, the aforementioned algorithm is iteratively performed using the previous iterations' C matrix as the projection matrix for the next iterations' H matrix and vice versa. We couple these iterations with an adaptive rank reduction scheme. This is done so that if a too high rank, r , as input is chosen, an even lower rank approximation can still be found if a good one exists. Specifically, at the end of each iteration, the enhanced algorithm checks if the rank of the combined row-space and column-space containing matrix, $A_2^T \tilde{X} A_1$, is below our rank parameter. If so, the algorithm reduces the rank parameter and continues the loop. However, if the rank of this matrix fail to reduce this iteration, the iteration is terminated. The entire algorithm can be found in Algorithm 1.

Algorithm 1 Closed-Form BRPs with Power Scheme

Input: $X \in \mathbb{R}^{n \times w_1}$, rank r , power Q
Output: $BRP(X) \triangleq L$

- 1: Initialize: $\tilde{X} = (XX^T)^q X$;
- 2: Initialize: $\forall i, j, A_{1,(i,j)} \sim \mathcal{N}(0, 1), A_{2,(i,j)} \sim \mathcal{N}(0, 1)$;
- 3: **while** true **do**
- 4: $C = \tilde{X}A_1 = Q_{col}R_{col}$;
- 5: $H = \tilde{X}^T A_2 = Q_{row}R_{row}$;
- 6: **if** $rank(A_2^T C) < r$ **then**
- 7: $r = rank(A_2^T C)$;
- 8: **else**
- 9: **break**;
- 10: **end if**
- 11: $A_1 = H$;
- 12: $A_2 = C$;
- 13: **end while**
- 14: $L = Q_{col}[R_{col}(A_2^T C)^{-1}R_{row}^T]^{\frac{1}{2q+1}}Q_{row}^T$;
- 15: **return** L ;

In Algorithm 1, it is worth mentioning that the power scheme is strong enough when $q \geq 3$ according to [29]. It costs $(2q+1)nwr$ floating-point operations (flops) to perform two projections and $r^2(n+w)$ flops to perform each QR decomposition. The matrix division in (9) requires an additional $nwr+2wr^2+4r^3$ flops. In general, the rank r is much smaller than n or w . Thus, the computational complexity of the BRP-based decomposition method is $O(nwr)$. This is far faster than the traditional matrix decomposition based on the SVD, whose computational complexity is $O(\min(nw^2, n^2w))$ [30].

D. Proximal Methods

The second sub-problem in (5) has the form of

$$\min_{S \in \mathbb{R}^{n \times w}} f(S) + \lambda\Omega(S), \quad (10)$$

where $f: \mathbb{R}^{n \times w} \rightarrow \mathbb{R}$ is a convex function, and $\Omega: \mathbb{R}^{n \times w} \rightarrow \mathbb{R}$ is a sparsity-inducing norm. In this problem, F is half the square of the Frobenius norm of the difference between $X-L$ and S , and Ω is the square of the l_{21} norm. This is exactly the form of minimization that defines the *proximal operator* [31] given by:

$$Prox_{\lambda\Omega}(\mathbf{u}) = \arg \min_{\mathbf{v} \in \mathbb{R}^w} \frac{1}{2} \|\mathbf{u} - \mathbf{v}\|_2^2 + \lambda\Omega(\mathbf{v}). \quad (11)$$

In this case, the \mathbf{u} variable found in the generic proximal operator definition can be replaced with $X-L$, and \mathbf{v} with S . In the specific case where $\Omega(\cdot)$ is the l_{21} norm, (11) is called ‘‘group Lasso’’ [31], and the proximal operator for l_{21} is

$$Prox_{\lambda\|\cdot\|_{21}}(X[i, :]) = (1 - \lambda\|X[i, :]\|_2)_+ X[i, :], \quad (12)$$

where $X[i, :] \in \mathbb{R}^w$ is the i -th row of X , $\|X[i, :]\|_2 = \sqrt{\sum_j x_{ij}^2}$, and $(\cdot)_+ \triangleq \max(\cdot, 0)$. From (12), we can see that the proximal operator maps a row of the target matrix at once.

E. Proximal BRP Algorithm

We call the proposed iterative approach to solve the matrix decomposition problem (5), *Proximal BRP (PBRP)*. It is summarized in the Algorithm 2.

Algorithm 2 Proximal BRP (PBRP)

Input: $X \in \mathbb{R}^{n \times w_1}$, rank r , power factor Q , λ , ϵ
Output: L, S

- 1: Initialization: $L = S = \mathbf{0}$
- 2: **while** $\frac{\|X-L-S\|_F^2}{\|X\|_F^2} \geq \epsilon$ **do**
- 3: $L = BRP(X - S)$
- 4: $S = Prox_{\lambda\|\cdot\|_{21}}(X - L)$;
- 5: **end while**
- 6: **return** L, S

F. Feature Engineering and Anomaly Detection with DBSCAN

After individually obtaining the low-rank matrix L and the sparse matrices associated with P, Q, V , and F , we take useful measurements from them - which we will call *anomaly scores*. Since the sparse matrix S represents the event patterns, its l_{21} norm is selected as the first anomaly score. We will call this score ‘*SAS*’ (*S*-Anomaly Score) for short. Furthermore, the sparsity of S intuitively comes from the regional variance of events, and thus, *SAS* summarizes the spatial features of the data. Drawing on the idea in [21], the maximum temporal difference of L is treated as the second anomaly score:

$$LAS \triangleq \max_{i,j} \left(\left| \frac{L[i, j] - L[i, j-1]}{L[i, j-1]} \right| \right), \quad (13)$$

which summarizes the temporal features of the data.

Algorithm 3 Voltage Event Detection Framework

Input: $X_{(l)} \in \mathbb{R}^{n \times T}$ ($l \in \{P, Q, V, F\}$); Analysis window lengths: w_1, w_2 ; DBSCAN: $minpts, eps$; Adaptive Scheme: $\theta = 100, uprate = 1.1, downrate = 0.9$.

Output: Anomaly Alert

- 1: **for** $t = w_1 + 1 : T$ **do**
- 2: **for** data type $l = \{P, Q, V, F\}$ **do**
- 3: $AS_{(l)}[t] \stackrel{L, S}{\leftarrow} PBRP(X_{(l)}[:, t - w_1 : t], \lambda_{(l)})$;
- 4: **if** $\max_{i=1, \dots, n} \left| \frac{std_i}{\mu_i} \right| > \theta$ **then**
- 5: $\lambda_{(l)} = \lambda_{(l)} \times uprate$;
- 6: **end if**
- 7: **if** $\|S\|_{21} == 0$ **then**
- 8: $\lambda_{(l)} = \lambda_{(l)} \times downrate$;
- 9: **end if**
- 10: **end for**
- 11: Construct the high-level feature batch:

$$B[t - w_2 + 1 : t] = \begin{pmatrix} AS_{(P)}[t - w_2 + 1 : t] \\ AS_{(Q)}[t - w_2 + 1 : t] \\ AS_{(V)}[t - w_2 + 1 : t] \\ AS_{(F)}[t - w_2 + 1 : t] \end{pmatrix};$$
- 12: **if** $t \geq w_2$ **then**
- 13: DBSCAN(normalize(B), $eps, minpts$) \rightarrow Alert;
- 14: **end if**
- 15: **end for**

After accumulating the two anomaly scores of P, Q, V , and F for w_2 time steps, an anomaly feature batch, $B \in \mathbb{R}^{8 \times w_2}$ is obtained. This feature batch can be regarded as a time series of length w_2 containing 8-dimensional feature vectors. Since we intend to build an unsupervised framework, we need an

unsupervised algorithm to do classification on B . There are two widely used unsupervised clustering algorithms: k-means and DBSCAN. Compared with k-means, DBSCAN is more conducive to the detection of outliers [32]. DBSCAN is a clustering algorithm based on density, so it is more suitable for various complex-shaped datasets, while k-means is mainly intended for convex datasets with spherical distributions. In DBSCAN, there are three types of points: *core* points, *density-reachable* points, and *outliers*. They are defined as follows: If a point has at least ‘*minpts*’ other points in its ‘*eps*’ neighborhood, it is a *core* point. The points that are in the ‘*eps*’ neighborhood of a *core* point are called a *density-reachable* point. Points that are neither core nor density reachable are identified as *outliers*.

In our proposed framework, we apply DBSCAN on B every time it is updated. If the 8-dimensional feature vector corresponding to time t is identified as the first outlier by the DBSCAN in the time series, the voltage event is deemed to occur at time t .

G. Summary of the Overall Event Detection Framework

The proposed voltage event detection framework is summarized in Algorithm 3, and illustrated in Fig. 1. For each time window w_1 and each data type l , the anomaly score (AS) that includes SAS and LAS is obtained. These anomaly scores are calculated based on L and S that are the outputs of the PBRP algorithm. Since P , Q , V , and F data streams have different scales, the corresponding penalty coefficients λ_l are adjusted separately and adaptively in Lines 4-9. When the maximum relative standard deviation (RSD) $\frac{std_i}{\mu_i}$ for PMUs in S exceeds a threshold θ , λ_l increases. The larger the RSD, the greater the degree of signal dispersion - thus requiring the weight of S to be larger to alert an event. Correspondingly, λ_l increases. However, if the l_{21} norm of S becomes zero, which means that no event pattern exists in X , λ_l decreases. After all anomaly scores within a time window of w_2 are calculated, we construct the anomaly feature batch B and apply the density-based cluster analysis on it to identify potential voltage events.

III. NUMERICAL STUDY WITH REAL-WORLD PMU DATA

The proposed event detection algorithm is validated using real-world PMU datasets that were recorded following voltage events. The online algorithm for PMU data processing and disturbance detection [21] and [22] are selected as benchmark algorithms.

A. Dataset Description

The PMU dataset is collected from the Eastern Interconnection of the U.S. power transmission grid. The dataset includes P , Q , V , and F readings from 187 PMUs with a sampling frequency of 30 Hz covering 668 labeled voltage events. Each event contains 3 minutes or 5400 samples of data (Fig. 3). Event datasets are divided into a validation dataset with 80% of event samples and a testing dataset with 20% of event samples. Our algorithm does not require training. The validation dataset is only used to determine the hyper-parameters of the entire framework, and the testing dataset helps evaluate the final performance.

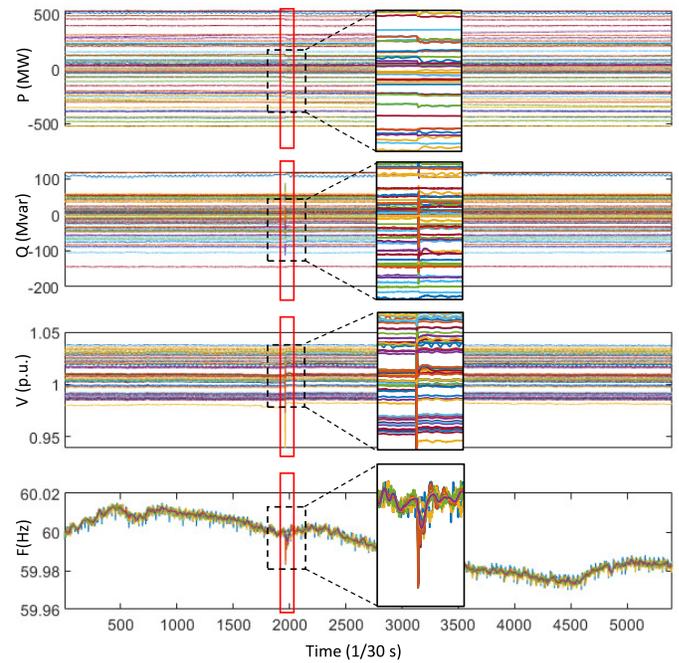


Fig. 3. Voltage event example with four electric quantities. The event occurs around the middle of the red box. Each colored line represents the readings from a PMU.

B. Benchmark Method

The online algorithm for PMU data processing (OLAP) [21] and OLAP with a Hankel matrix (HOLAP) [22] are two state-of-the-art power system event detection methods. The Hankel matrix is more sensitive to the change in temporal correlation of the time series, but its size is larger. Both methods leverage the low-rank property of their target matrices (the original data matrix in OLAP and the data matrix constructed with a Hankel structure in HOLAP). OLAP computes the ratio of the first two singular values, denoted ζ , in a short time window and detects system disturbances by using changes in this ratio. HOLAP computes the rank-1 approximation error of the original Hankel matrix as well as that of a permuted version of itself, and identifies an event by using the change in the ratio of these two errors. This later ratio is denoted as η . To make a fair comparison, OLAP and HOLAP are embedded into our proposed framework to replace PBRP by using ζ and η as anomaly scores. In particular, Lines 3-9 in Algorithm 3 are replaced with the two comparison algorithms.

C. Hyper-parameter Settings

Important hyper-parameters of three algorithms are summarized in TABLE II. The OLAP and HOLAP adopt the optimal parameters identified in [21] and [22] respectively. The DBSCAN module calculates Euclidean distance between individual samples. Only the *eps* is fine-tuned because all the algorithms are only sensitive to this hyper-parameter.

D. Numerical Results

We rely on the validation dataset to determine the optimal value of the hyper-parameter ‘*eps*’ in DBSCAN. At the initial

TABLE II
HYPER-PARAMETER SETTINGS

Algorithm Components	Window Length	PBRP	OLAP	HOLAP	DBSCAN
Parameter Values	$w_1 = 30$	$\epsilon = 0.001$	$\bar{w} = 5$	$\kappa = 5$	$minpts = 2$
	$w_2 = 300$	$\lambda = 10$			$eps : 2 \sim 13$
		$r = 5$			
		$q = 5$			

time t , the P, Q, V, F data in w_1 -length window are respectively passed through the PBRP algorithm (Section II-E) to obtain their corresponding L and S . Then, we extract anomaly features SAS and LAS from the L and S matrices (Section II-F), generating an 8-dimensional feature vector at time t . Packing it with $w_2 - 1$ feature vectors before time t into a batch B , and send B to DBSCAN for clustering analysis. If there is an outlier point, an alert will be issued. The result of the above process is shown in Fig. 4. After getting the best ‘ eps ’, we use the best hyper-parameters to evaluate the final performance of our voltage event detection framework on the testing dataset.

1) Performance of Voltage Event Detection Framework:

Two commonly used evaluation metrics in classification problems, F1 and F2 scores, are used to evaluate the performance of our voltage event detection framework. The two F-scores are calculated based on *precision* and *recall* that are further derived based on True Positive (TP), False Positive (FP), and False Negative (FN). The TP means events are detected within 1 second of the labeled event time. The FP denotes the scenarios where the algorithm reports an event outside the above 1 second time window. The FN comprises the cases where no event is detected within the aforementioned 1 second time window. *Recall* and *precision* are calculated as: $Recall = \frac{TP}{TP+FN}$, $Precision = \frac{TP}{TP+FP}$. F1 and F2 scores are derived based on *precision* and *recall* as follows: $F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$, $F2 = \frac{5 \times Precision \times Recall}{4 \times Precision + Recall}$. The F1 score is the harmonic mean of the *precision* and *recall*, while the F2 score weighs *recall* higher than *precision*.

The F1 and F2 scores of the proposed PBRP and the two benchmark on the validation dataset are evaluated to select the appropriate hyper-parameter, eps (Fig. 4). Note that two eps are selected for each algorithm, one that optimizes F1 score and the other optimizes F2 score. After the hyper-parameters are selected, we apply the three algorithms to the testing dataset. The PBRP algorithm achieves significantly higher F1 and F2 scores than the benchmark algorithms on the testing dataset, mainly due to a substantial improvement in Recall (Table III). Note that in Table III the top (bottom) three rows correspond to the hyperparameters optimized for F1 (F2) score.

The improvements over the benchmark can mostly be attributed to the ability to capture the spatial properties with the anomaly score SAS . Both OLAP and HOLAP are capable of capturing temporal anomalies. However, as seen in Fig. 5, these temporal anomalies, which are also captured by the LAS indicator, does not become pronounced right away. This leads to delay in the detection of events and significantly decreases

TABLE III
F SCORES OF THREE ALGORITHMS ON THE TESTING DATASET

Statistics	OLAP	HOLAP	PBRP
Precision	0.8889	0.8824	0.8881
Recall	0.8955	0.8955	0.9478
F1 Score	0.8922	0.8889	0.9170
Precision	0.8089	0.8571	0.8000
Recall	0.9478	0.9403	0.9851
F2 Score	0.9163	0.9224	0.9415

the recall. In contrast, the spatial anomaly indicators reach their peaks very quickly as soon as the event begins.

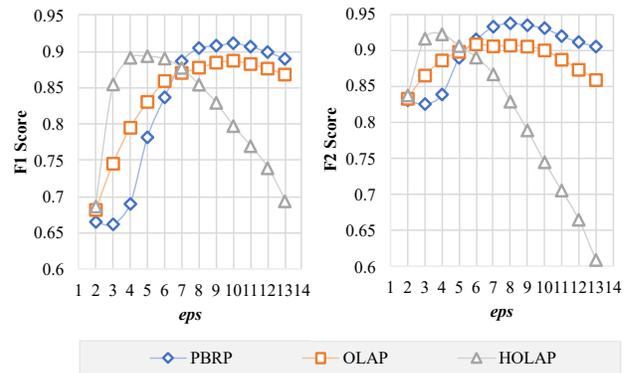


Fig. 4. F scores of PBRP and benchmark algorithms on validation dataset.

2) Computational Efficiency: The computation complexity of BRP-based and SVD-based matrix decomposition approaches were discussed at the end of Section II-C. This computational efficiency is crucial to apply the detection algorithms online. Faster event detection allows greater flexibility in the design of any submodules that follow.

To showcase that our method is faster than its competitors, we varied the number of PMUs while fixing $w_1 = 30$ and $w_2 = 300$. Therefore, the size of the matrix to be decomposed is only proportional to the number of PMUs. All algorithm’s computation times are reported in Table IV after averaging over the detection of 100 randomly selected voltage events. The partial computation time excludes the time of the cluster analysis (Line 13 in Algorithm 3). The PBRP algorithm uses 50% less computation time compared to the benchmark algorithms. Furthermore, as the number of PMUs increases, the increase in the computation time of the PBRP algorithm is slower than that of the benchmark algorithms. Thus, the proposed algorithm shows better applicability to larger grids and larger numbers of sensors compared to benchmark algorithms.

TABLE IV
AVERAGE COMPUTATION TIME OF EVENT DETECTION ALGORITHMS OVER THREE-MINUTE TIME PERIOD

Number of PMUs	50			100			150		
	Algorithm	Partial (s)	Total (s)	Algorithm	Partial (s)	Total (s)	Algorithm	Partial (s)	Total (s)
Computation Time (s)	HOLAP	61.78	68.46	HOLAP	181.50	189.25	HOLAP	336.27	344.58
	OLAP	7.53	15.01	OLAP	9.58	17.33	OLAP	16.99	24.79
	PBRP	2.18	8.46	PBRP	3.13	9.40	PBRP	4.29	10.53

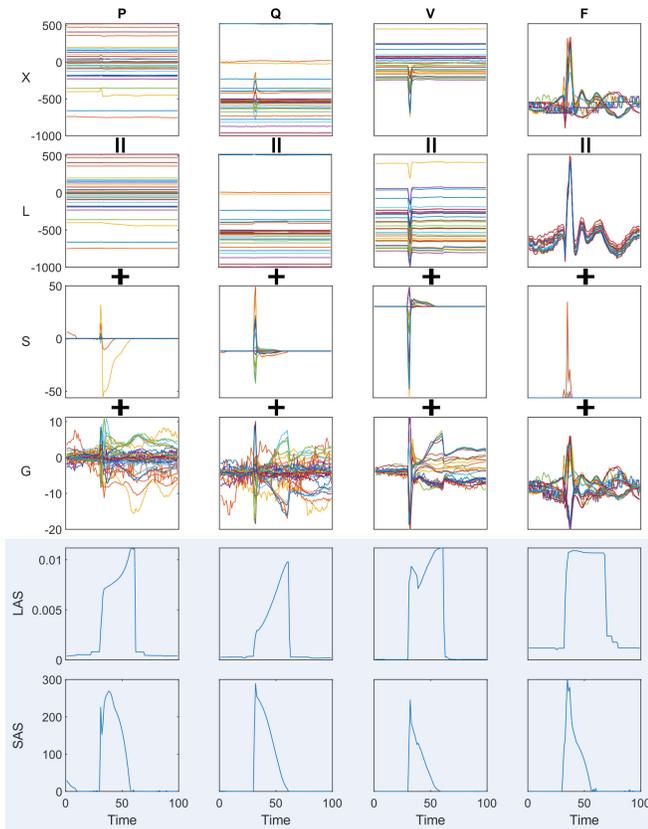


Fig. 5. An example of decomposition of streaming PMU data matrix X with corresponding anomaly scores for a voltage event.

3) *Identification of PMUs That Can Capture Voltage Events:* As shown in Figs. 2 and 3, significant influence on voltage events is often observed in a limited number of PMUs. The sparsity structure in the S matrix could help identify which PMUs are closely related to a particular voltage event. For example, we could first identify the non-zero elements in matrix S at the start of the events. These row indices of these non-zero elements correspond to distinctive PMUs that firmly grasp an event. Fig. 6 illustrates the identification of the distinctive PMUs on the voltage event. The left column figures depict the original P, Q, V, and F data obtained from PMUs. The middle column is the heat map of sparse matrix S . We remove the PMUs in the original X matrix corresponding to the rows with zero elements at timestamp 24 identified as the start of the event. The right column shows the filtered X , containing highly sensitive PMUs to the event. The algorithm effectively identifies all the PMUs with sizable dynamic behavior in P, Q, V, and F data streams.

IV. CONCLUSION

This paper reveals the distinctive sparsity structure of residual PMU data matrices during regional voltage events. This distinctive characteristic of voltage events motivates us to decompose the PMU data matrix into a low-rank matrix, a row-sparse event-pattern matrix, and a noise matrix. The key features extracted from the low-rank and row-sparse event-pattern matrices are leveraged in a clustering module to

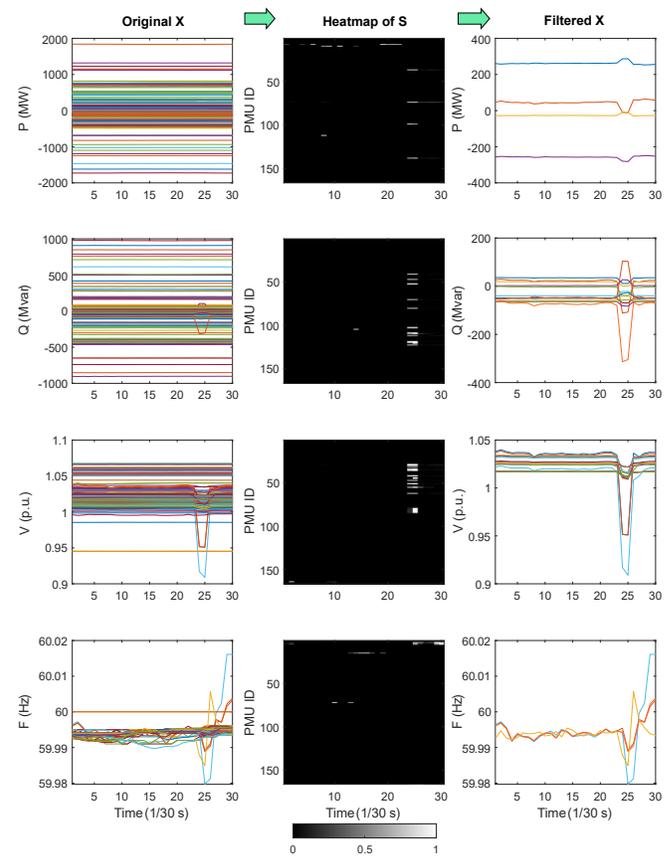


Fig. 6. An example of identifying PMUs that are sensitive to voltage events. Elements in the heatmap of S are scaled from 0 to 1.

differentiate voltage events from normal operating conditions. A computationally efficient proximal bilateral random project-based algorithm, PBRP, is proposed to perform the matrix decomposition with structured sparsity-inducing norms. The feature extracted from the row-sparse event-pattern matrix significantly enhances the voltage event detection performance. A large-scale numerical study with real-world PMU data shows that our proposed *online* voltage event detection algorithm yields lower computation time, higher accuracy, and scalability than state-of-the-art benchmark. The proposed algorithm can also specify event area/zones by identifying the PMUs most sensitive to the detected event.

Although the proposed algorithm displays outstanding performance for detecting voltage-related events, it only shows marginal improvement in detecting frequency related events due to less pronounced row-sparsity structure for the residual PMU data matrix. In the future, we plan to further examine the unique structures of the PMU data matrices, which correspond to the frequency events.

ACKNOWLEDGMENT

This material is based upon work supported by the Department of Energy under Award Number DE-OE0000916.

REFERENCES

- [1] A. G. Phadke and T. Bi, "Phasor measurement units, WAMS, and their applications in protection and control of power systems," *Journal of*

- Modern Power Systems and Clean Energy*, vol. 6, no. 4, pp. 619–629, 2018.
- [2] C. Tu, X. He, Z. Shuai, and F. Jiang, “Big data issues in smart grid—a review,” *Renewable and Sustainable Energy Reviews*, vol. 79, pp. 1099–1107, 2017.
 - [3] D.-I. Kim, T. Y. Chun, S.-H. Yoon, G. Lee, and Y.-J. Shin, “Wavelet-based event detection method using pmu data,” *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1154–1162, 2015.
 - [4] K. Kirihara, J. Yamazaki, P. Chongfuangprinya, S. Konstantinopoulos, C. Lackner, J. H. Chow, S. Maslennikov, and Y. Liu, “Speeding up the dissipating energy flow based oscillation source detection,” in *2019 International Conference on Smart Grid Synchronized Measurements and Analytics*, College Station, TX, USA, May 2019, pp. 1–8.
 - [5] S. S. Negi, N. Kishor, K. Uhlen, and R. Negi, “Event detection and its signal characterization in pmu data stream,” *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 3108–3118, 2017.
 - [6] N. Zhou and J. Dagle, “Initial results in using a self-coherence method for detecting sustained oscillations,” *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 522–530, 2014.
 - [7] R. Yadav, A. K. Pradhan, and I. Kamwa, “Real-time multiple event detection and classification in power system using signal energy transformations,” *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1521–1531, 2018.
 - [8] M. Cui, J. Wang, J. Tan, A. R. Florita, and Y. Zhang, “A novel event detection method using PMU data with high precision,” *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 454–466, 2018.
 - [9] J. Shi, B. Foggo, X. Kong, Y. Cheng, N. Yu, and K. Yamashita, “Online event detection in synchrophasor data with graph signal processing,” in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, Tempe, AZ, USA, Nov. 2020, pp. 1–7.
 - [10] E. Drayer and T. Routtenberg, “Detection of false data injection attacks in smart grids based on graph signal processing,” *IEEE Systems Journal*, vol. 14, no. 2, pp. 1886–1896, 2019.
 - [11] M. Khan, P. M. Ashton, M. Li, G. A. Taylor, I. Pisica, and J. Liu, “Parallel detrended fluctuation analysis for fast event detection on massive pmu data,” *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 360–368, 2014.
 - [12] C. Chen, J. Wang, and H. Zhu, “Effects of phasor measurement uncertainty on power line outage detection,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 6, pp. 1127–1139, 2014.
 - [13] W. Wang, H. Yin, C. Chen, A. Till, W. Yao, X. Deng, and Y. Liu, “Frequency disturbance event detection based on synchrophasors and deep learning,” *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3593–3605, Jul. 2020.
 - [14] J. Shi, B. Foggo, and N. Yu, “Power system event identification based on deep neural network with information loading,” *arXiv preprint arXiv:2011.06718*, 2020.
 - [15] M. Zhou, Y. Wang, A. K. Srivastava, Y. Wu, and P. Banerjee, “Ensemble-based algorithm for synchrophasor data anomaly detection,” *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2979–2988, 2018.
 - [16] Y. Ge, A. J. Flueck, D.-K. Kim, J.-B. Ahn, J.-D. Lee, and D.-Y. Kwon, “Power system real-time event detection and associated data archival reduction based on synchrophasors,” *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2088–2097, 2015.
 - [17] M. Rafferty, X. Liu, D. M. Laverty, and S. McLoone, “Real-time multiple event detection and classification using moving window pca,” *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2537–2548, 2016.
 - [18] L. Xie, Y. Chen, and P. Kumar, “Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis,” *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2784–2794, 2014.
 - [19] S. Liu, Y. Zhao, Z. Lin, Y. Liu, Y. Ding, L. Yang, and S. Yi, “Data-driven event detection of power systems based on unequal-interval reduction of pmu data and local outlier factor,” *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1630–1643, 2019.
 - [20] W. Li, M. Wang, and J. H. Chow, “Fast event identification through subspace characterization of PMU data in power systems,” in *2017 IEEE Power & Energy Society General Meeting*, Chicago, IL, USA, Jul. 2017, pp. 1–5.
 - [21] 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, 2015.
 - [22] 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 Transactions on Power Systems*, vol. 33, no. 6, pp. 6966–6977, 2018.
 - [23] J. Hao, R. J. Piechocki, D. Kaleshi, W. H. Chin, and Z. Fan, “Sparse malicious false data injection attacks and defense mechanisms in smart grids,” *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1–12, 2015.
 - [24] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, “Detecting false data injection attacks on power grid by sparse optimization,” *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 612–621, 2014.
 - [25] P. Gao, M. Wang, J. H. Chow, S. G. Ghiocel, B. Fardanesh, G. Stefopoulos, and M. P. Razaousky, “Identification of successive “unobservable” cyber data attacks in power systems through matrix decomposition,” *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5557–5570, 2016.
 - [26] T. Routtenberg and Y. C. Eldar, “Centralized identification of imbalances in power networks with synchrophasor data,” *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1981–1992, 2017.
 - [27] H. Zhu and G. B. Giannakis, “Sparse overcomplete representations for efficient identification of power line outages,” *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2215–2224, 2012.
 - [28] S. Dasgupta and A. Gupta, “An elementary proof of the johnson-lindenstrauss lemma,” *Random Structure & Algorithm*, vol. 22, no. 1, pp. 60–65, Jan. 2003.
 - [29] T. Zhou and D. Tao, “Bilateral random projections,” in *2012 IEEE International Symposium on Information Theory Proceedings*, Cambridge, MA, USA, Jul. 2012, pp. 1286–1290.
 - [30] L. Knockaert, B. De Backer, and D. De Zutter, “Svd compression, unitary transforms, and computational complexity,” *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2724–2729, 1999.
 - [31] F. Bach, R. Jenatton, J. Mairal, and G. Obozinsk, *Optimization with sparsity-inducing penalties*, 1st ed. Now Foundations and Trends, 2012.
 - [32] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, “DBSCAN revisited, revisited: why and how you should (still) use DBSCAN,” *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.



Xianghao Kong received the B.S. degree in Computer Science from Hangzhou Dianzi University, Hangzhou, China, in 2019. He is now pursuing his Ph.D. degree in Computer Science from University of California, Riverside. His research interests lie in time-efficient machine learning and big data analysis in dynamical systems.



Brandon Foggo (M’19) received the B.S. degree in Electrical Engineering from the University of California, Los Angeles, in 2015, and the Ph.D. degree in Electrical and Computer Engineering from the University of California, Riverside, in 2019. His research interests lie in statistical learning theory and information theory, particularly in their merging, as well as applications to cyber physical systems with an emphasis on power distribution systems.



Koji Yamashita (M’04) received the B.S. and M.S. degrees in Electrical Engineering from Waseda University, Tokyo, Japan, in 1993 and 1995, respectively. He also received his Ph.D. degree in Electrical and Computer Engineering from Michigan Technological University, Houghton, MI, USA, in 2020. He was a research scientist with the power system division of CRIEPI, Japan in 1995–2018. He is currently a postdoctoral scholar at University of California Riverside, cultivating machine learning techniques for power system dynamics.



Nanpeng Yu (M'11-SM'16) received his B.S. in Electrical Engineering from Tsinghua University, Beijing, China, in 2006. Dr. Yu also received his M.S. and Ph.D. degree in Electrical Engineering from Iowa State University, Ames, IA, USA in 2007 and 2010 respectively. He is currently an Associate Professor in the department of Electrical and Computer Engineering at University of California, Riverside, CA, USA. His current research interests include machine learning in smart grid, electricity market design and optimization, and smart energy

communities. Dr. Yu is an Editor of IEEE Transactions on Smart Grid and IEEE Transactions on Sustainable Energy.