Real-time Event Identification through Low-dimensional Subspace Characterization of High-dimensional Synchrophasor Data

Wenting Li, *Student Member, IEEE,* Meng Wang, *Member, IEEE*, Joe H. Chow, *Fellow, IEEE*

***Abstract*—This paper proposes a data-driven real-time event identification method based on the measurements of Phasor Measurement Units (PMUs). The central idea is to characterize an event by the low-dimensional subspace spanned by the dominant singular vectors of a data matrix that contains spatial-temporal blocks of PMU data. The subspace representation is robust to system initial conditions and characterizes the system dynamics. A dictionary of subspaces that correspond to different events are established off-line, and an event is identified online with the most similar event in the dictionary through subspace comparison. The compact subspace representation reduces the dictionary size and the computational time of the event identification method. Numerical experiments on both simulated events in an IEEE 68-bus power system and the recorded data in New England to validate the proposed method.**

***Index Terms*—disturbance identification, Phasor Measurement Unit (PMU), low-dimensional approximation, singular value decomposition, subspace.**

I. Introduction

POWER system faults and events happen irregularly and potentially lead to wide-area oscillations and even cascading failures. The Northeast blackout of 2003 in the U.S. started with a local fault but escalated to a widespread blackout partially because the operator was not aware of the fault and took no action at the early stage. Therefore, fast event detection, identification, and location are important to enhance the wide-area situational awareness of power systems and prevent cascading failures.

The current control room operators can identify the tripped elements within seconds based on the status of circuit breakers, provided in the Supervisory Control and Data Acquisition (SCADA) system. With the increasing amounts of data obtained by Phasor Measurement Units [2], many efforts have been devoted to exploit the obtained measurements to locate events such as generator trips [3], line outages [4]–[6], and oscillations [7]–[10]. Model-based identification methods (see e.g., [4], [5], [11]) usually require the modeling of the power system, and the identification performance depends critically on the accuracy of estimated parameters. Data-driven methods [12]–[21] are receiving increasing attention partially due to the increasingly denser coverage of Phasor Measurement Units [2], which measure voltage and current phasors directly at a rate of thirty samples per second or above. Data-driven methods extract features (including direct features like a frequency [15] or its derivative [14], as well as indirect features like wavelet coefficients [18]) from measurements and classify those with similar features as resulting from the same event type. One limitation of data-driven methods is that the computational complexity generally increases when the dataset size increases. Moreover, the classifiers trained from data do not have clear physical interpretations.

Some event identification methods can only identify an event with a delay of tens of seconds after an event initiates. For instance, most line outage localization methods require the new steady state condition [5], [22]–[24], and it takes seconds or beyond to reach a new steady state after an event starts. Only a few work such as [25]–[27] exploit the transient response of the system to locate the outage. The advantage of identification methods using the transient dynamics is that they require a shorter time window, like one second or even less, and therefore can identify an event in a more timely fashion. These methods, however, require an accurate system dynamic model to generate synthetic datasets for comparison or training and thus, suffer from model inaccuracies.

This paper proposes a novel real-time data-driven method for event identification by exploiting the transient dynamics after an event happens. The method can quickly identify an event, as only one second of data is needed in our numerical experiments. Moreover, it does not require any modeling of the power system. The critical innovation is to characterize an event by a low-dimensional row subspace spanned by the dominant singular vectors of the data matrix that contains spatial temporal blocks of measurements from multiple PMUs. This subspace characterization is robust to system initial conditions and captures the system dynamics. Then an event is identified by comparing the obtained data with a pre-computed event dictionary with each dictionary atom corresponding to a row subspace of an event.

This proposed method has the following distinctive features: 1. The low-dimensional subspace can be computed through Singular Value Decomposition (SVD) [28] or its faster variants like [29]. 2. The dictionary size is much smaller, compared with the existing dictionary of time series [20] or explicit features in complicated model [30]. The reduction of the dictionary size reduces the computational complexity of both the offline training and the online event identification. 3. The method identifies events shortly after the event starts (e.g., within 1 second) and can be implemented in real time, while existing methods many require a time window of 10-

W. Li, M. Wang, and J.H. Chow are with the Dept. of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY. Email: {liw14, wangm7, chowj}@rpi.edu.

Preliminary and partial results will appear in [1].

30 seconds of data [20], [31]. 4. The approach can identify both the internal and the external events only based on local measurements in one control region, respectively.

The rest of the paper is organized as follows. The physical interpretation and numerical validation of a subspace representation are described in Section II. The event identification method is introduced in Section III. Sections IV and V record the numerical experiments in both the simulated data in the IEEE 68-bus test system and the recorded data from New England. Section VI concludes the paper.

II. Subspace characterization of system events

A. Low-rank Property of PMU Data Matrices

PMU measurements at multiple buses across time are collected into an $m \times T$ matrix M, where m is the number of measurements at one time instant, and T is the number of time steps. The matrix can often be approximated by a low-rank matrix with rank r much less than m and T. This approximate low-rank property of PMU data matrices has been observed and exploited to recover missing PMU data [32], detect system events [32], [33], and identify cyber data attacks [34].

Let σ_i denote its ith largest singular value of M. Throughout this paper, given predetermined $\tau \in (0,1)$ and $\Delta \gg 1$, the approximate rank r of M is the largest i such that

$$\sigma_i/\sigma_{\bar{r}+1} > \Delta \tag{1}$$

where $\bar{r}$ is the smallest integer such that

$$(\Sigma_{i=1}^{\bar{r}}\sigma_i)/(\Sigma_{j=1}^{m}\sigma_j) \geq \tau \tag{2}$$

holds. Δ in (1) guarantees that only dominate singular values remain in the approximation. τ in (2) measures the approximation ratio. The best rank-r approximation to M in terms of minimizing $\frac{\|M-M'\|_F}{\|M\|_F}$ over rank-r matrices M') is

$$M_r = U_r \Sigma_r V_r^\dagger, \tag{3}$$

where $\Sigma_r \in \mathbb{R}^{r\times r}$ contains the r largest singular values as its diagonal entries, and $U_r \in \mathbb{C}^{m\times r}$, $V_r \in \mathbb{C}^{T\times r}$ contain the corresponding r left and right singular vectors, respectively. $V_r^\dagger$ denotes the conjugate transpose. U_r characterizes the correlations of measurements in different PMUs. V_r characterizes the dominant dynamics. Columns of V_r form a unitary basis of the row subspace of M_r, denoted by span(V_r).

B. Subspace Characterization of Events

We motivate the physical interpretation of span(V_r) through linear system analysis. If the power system is linearized around one equilibrium point, the resulting discrete-time linear time invariant (LTI) system model is

$$\mathbf{x(t+1)} = \mathbf{Ax(t)} + \mathbf{Bu(t)}, \tag{4}$$

$$\mathbf{y(t)} = \mathbf{Cx(t)} + \boldsymbol{\epsilon}\mathbf{(t)}, \tag{5}$$

where vectors $\mathbf{x(t)} \in \mathbb{C}^{\mathbf{n}}$ and $\mathbf{y(t)} \in \mathbb{C}^{\mathbf{m}}$ represent deviations at time t in state variables and measurements from the equilibrium point. $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are the state, input, and output matrix respectively. $\mathbf{u(t)}$ is the control input. $\boldsymbol{\epsilon}(t)$ is the measurement noise. Let λ_k, $\mathbf{l}_k$, and $\mathbf{r}_k$ denote the kth eigenvalue of $\mathbf{A}$ and the corresponding left and right eigenvectors.

We ignore measurement noise ($\boldsymbol{\epsilon}(t)$) and consider the system dynamics after an impulse input. $\mathbf{x_0} \in \mathbb{C}^{\mathbf{n}}$ denotes the system initial state after the impulse input. M contains the measurements from time 1 to T after the impulse input. M can be written as the sum of n rank-1 matrices, i.e.,

$$M = [\mathbf{y(1)}, \mathbf{y(2)}, \cdots, \mathbf{y(T)}] = \sum_{\mathbf{k=1}}^{\mathbf{n}} \mathbf{l_k^\dagger x_0} \cdot (\mathbf{Cr_k}) \cdot \boldsymbol{\beta}_\mathbf{k}^\dagger, \tag{6}$$

where

$$\boldsymbol{\beta}_k^\dagger = [\lambda_k, \lambda_k^2, \cdots, \lambda_k^T]. \tag{7}$$

We assume without loss of generality that rank-1 matrices in (6) are sorted in the decreasing order of the matrix 2-norm, which is $|\mathbf{l}_k^\dagger \mathbf{x_0}|\|\mathbf{Cr_k}\|\|\boldsymbol{\beta_k}\|$ for the kth rank-1 matrix. The k-th matrix for $k > r$ is relatively small, measured by its 2-norm, if the corresponding mode is not excited by the input ($|\mathbf{l}_k^\dagger \mathbf{x_0}| \approx \mathbf{0}$); or not directly measured ($\|C\mathbf{r}_k\| \approx 0$); or highly damped ($\|\boldsymbol{\beta}_k\| \approx 0$, i.e., $|\lambda| \approx 0$). Then if we only keep the first r rank-1 matrices in the summation of (6) and project M to the row subspace of the resulting rank-r matrix, we obtain another rank-r approximation to M, denoted by M_r'. We have

$$M_r' = M\Psi_r(\Psi_r^\dagger \Psi_r)^{-1}\Psi_r^\dagger, \tag{8}$$

where

$$\Psi_r := \begin{bmatrix}\boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \dots & \boldsymbol{\beta}_r\end{bmatrix} \in \mathbb{C}^{T\times r}. \tag{9}$$

Note that Ψ_r only depends on the first r singular values λ_k's that are sorted in the decreasing order of $|\mathbf{l}_k^\dagger \mathbf{x_0}|\|\mathbf{Cr_k}\|\|\boldsymbol{\beta_k}\|$. Changes in system conditions, including topology changes and power injection variations, could change the initial condition $\mathbf{x}_0$, singular vectors $\mathbf{l}_k$'s, $\mathbf{r}_k$'s, and singular values λ_k's. As long as the dominate r singular values remain the same, Ψ_r does not change. Thus, the subspace spanned by Ψ_r can represent the dominant dynamics of M.

Moreover, if all but the first r matrices in (6) are indeed small, the rank-r approximation M_r' in (8) should be close to the best rank-r approximation M_r in (3). Notice that Ψ_r is a complex matrix and V_r is a real matrix if a real-valued M is chosen, then one can check that in the T-dimensional real space $\mathbb{R}^T$, the span of V_r is the same as the span of $[\Psi_{real}, \Psi_{image}]$, where $\Psi_{real}, \Psi_{image} \in \mathbb{R}^{T\times r}$ are real and imaginary parts of Ψ_r, respectively. We focus on $\mathbb{R}^T$ here for notational simplicity. We then have

$$\text{span}(\Psi_r) = \text{span}(V_r), \tag{10}$$

where the spans are considered in $\mathbb{R}^T$ and have the same dimension.

Note that (10) holds under the linear system model with an impulse input and no noise. (10) is derived for a physical interpretation of span(V_r), while our proposed event identification method only uses the row subspace span(V_r) and does not depend on (10). This paper studies span(V_r) instead of (4)-(7), which are often used in mode analysis [7]–[10]. As shown in Section II, span(V_r) is closely related to the system modes without solving the system parameters like λ_i's.

C. Comparison of Different Events through Subspace Angles

The proposed event identification method is based on the following intuition. If two events are of the same type, at nearby locations, and have close initial states, their respective λ_i's ($i = 1, ..., r$) that correspond to the largest r terms of $|\mathbf{l}_k^\dagger \mathbf{x_0}| \|\mathbf{C r_k}\| \|\boldsymbol{\beta_k}\|$ are similar. That in turn means that Ψ_r's of these events are similar. Then from (10), we expect that span(V_r)'s might be similar. In contrast, span(V_r)'s of two events of different types and at remote locations might be very different. Thus, roughly speaking, the similarity of span(V_r) could be employed to study the similarity of events.

To quantify this intuition, we employ the average subspace angle [35] to measure the similarity of two subspaces in $\mathbb{R}^T$. The angle between an l-dimensional subspace $\mathcal{S}_l$ and a k-dimension subspace $\mathcal{S}_k$ in $\mathbb{R}^T$, spanned by orthogonal bases B_l and B_k, respectively, can be computed as

$$\theta(\mathcal{S}_l, \mathcal{S}_k) = \arccos(\sqrt{\|B_k^\dagger B_l\|_F^2 / \min\{k, l\}}). \quad (11)$$

θ equals 90° if two subspaces are orthogonal to each other. θ equals 0° if two subspaces are the same ($l = k$) or one is embedded in the other ($l \neq k$). Thus, a smaller θ indicates higher affinity of two subspaces.

We remark that the subspace angle has the limitation of the Zero Distance Problem [36]. If S_l is a low-dimensional subspace embeded in S_k with $l < k$, $\theta(\mathcal{S}_l, \mathcal{S}_k)$ would be 0° although these two subspaces are different. To address this issue, when comparing two events, we use the corresponding low-dimensional subspaces span(V_r)'s with the same dimension. When the individual approximate ranks determined by (1)-(2) of two events are different, we set r to be the larger value of the two. See Section IV-B for details.

We use two simple examples to illustrate this intuition.

Example 1. Consider two first-order systems with real-valued eigenvalues λ and $\hat{\lambda}$, receptively. Suppose $\lambda \leq \hat{\lambda} \leq (1+\varepsilon)\lambda$ for some positive ε. The smaller ε means the closer the eigenvalues are. Then

$$\Psi = [\lambda, \lambda^2, ..., \lambda^T]^\dagger, \text{ and } \hat{\Psi} = [\hat{\lambda}, \hat{\lambda}^2, ..., \hat{\lambda}^T]^\dagger. \quad (12)$$

From (11), the angle between span(Ψ) and span($\hat{\Psi}$) is

$$\theta(\text{span}(\Psi), \text{span}(\hat{\Psi})) = \arccos\left(\frac{\sqrt{|\Sigma_{k=1}^t \lambda^k \hat{\lambda}^k|^2}}{\sqrt{\Sigma_{k=1}^t \lambda^{2k}}\sqrt{\Sigma_{k=1}^t \hat{\lambda}^{2k}}}\right) \in [0, \arccos(\frac{1}{(1+\varepsilon)^t})]. \quad (13)$$

Then, a smaller ε leads to a smaller θ.

We next study the impact of different events on λ_i's in a system of multiple machines connecting to an infinite bus. The analysis is largely simplified for an illustration purpose.

Example 2. Fig. 1 (a) shows a multi-machine system connected to an infinite bus and (b) shows its equivalent model. The infinite bus is set to be the reference with its voltage denoted by $E_b\angle 0$. The generators are modeled by an equivalent generator with the internal voltage $E'\angle\delta$, rotor angle δ, angular speed of the rotor ω, inertial constant H, and the transient reactance X'_d. T_m is the equivalent mechanical torque. X_E denotes the equivalent line reactance of two parallel lines, and the line resistance is ignored. Let $X_T = X'_d + X_E$. Then the state equations of the continuous-time linear dynamical system can be written as

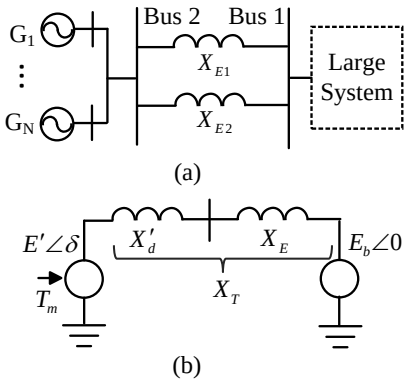

Fig. 1: (a) A multi-machine infinite-bus system (b) Its equivalent model

$$\begin{bmatrix} \Delta\dot{\delta} \\ \Delta\dot{\omega} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & \omega_0 \\ \frac{-K_s}{2H} & \frac{-K_D}{2H} \end{bmatrix}}_{A'} \begin{bmatrix} \Delta\delta \\ \Delta\omega \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{\Delta T_m}{2H} \end{bmatrix} \quad (14)$$

where $K_s = \dfrac{E' E_b \cos\delta_0}{X_T}$ and K_D are the synchronizing torque coefficient and the damping torque coefficient of the generator, respectively [37]. δ_0 is the initial rotor angle; ω_0 is rated angular speed. Since the damping torque is often neglected when computing the eigenvalues of A [38], [39], we also assume $K_D = 0$ to simplify the computation in this example. We compute the eigenvalues of A' of the continuous model by

$$\lambda'_{1,2} = \pm j\sqrt{\frac{2\omega_0 E' E_b \cos\delta_0}{H X_T}}. \quad (15)$$

We assume that the voltage E' does not change much in a very short time window, then λ under different events can be analyzed as follows.

(1) Line trip (LT) event. X_E increases if one of the two parallel lines is tripped. Then X_T increases, and λ' decreases.

(2) Three-phase (TP) short circuit event at bus 1. It can be equivalently modeled as that E_b becomes zero, then λ' decreases significantly.

(3) Load change (LC) event. The steady-state output real power of the generator can be computed from $P = (E' E_b \sin\delta)/X_T$. If the increase of the load consumption results from the increase of E_b, then from (15), $|\lambda'|$ increases correspondingly.

D. Numerical validation of the subspace property

We next validate numerically on the IEEE 16-generator 68-bus test system (Fig. 2) that (10) holds and span(V_r) of the data matrix can characterize the system event.

We simulate PMU data through the nonlinear model with a data rate of 30 samples per second by Power System Toolbox (PST) [40]. Then we calculate the span(V_r) of the three types simulated data: a 0.5 p.u. load change event at bus 5, a line trip event at line 5-6, and a three-phase short circuit event after

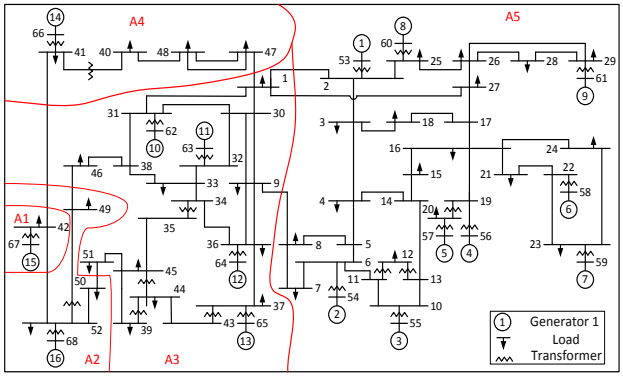

Fig. 2: IEEE 16-machine 68 bus power system with 5 coherent groups

clearance at bus 5. The voltage magnitudes in one second of all buses are collected into a matrix M. Then the r dominant right singular vectors of each M are collected in V_r, where r is determined by (1) - (2) with $\Delta = 10, \tau = 0.99$.

We next compute Ψ_r of line trip events based on (6) through a linear model of the same system. Two line trip events are considered. One is between bus 5 and bus 6, denoted by line 5-6, and the other is at line 26-29. The steps include

1) Compute system matrices A and C corresponding to voltage magnitudes in (4)-(5). The function svm_mgen in PST is employed to calculate A' and C of the continuous-time system based on the perturbation theory, where A' is the state matrix of the continuous-time model. The discrete-time system matrix A is computed by $A = e^{A'T_S}$, where the step size $T_S = 0.03$.
2) Calculate all the eigenvalues λ_k's, left and right eigenvectors $\mathbf{l}_k$'s and $\mathbf{r}_k$'s of A by QR decomposition [28], where $\lambda_i, l_i, r_i, (i = 1, 2, \ldots, 32)$;
3) Compute Ψ_r from Ψ_n. From (8), the dominant r vectors of β_k's in the rank-r approximation[1] is selected based on coefficients $|\mathbf{l}_k^\dagger \mathbf{x_0}| \|\mathbf{C r_k}\| \|\beta_\mathbf{k}\|$, where x_0 is the simulation data of the first step after the event happens.

Table I: The subspace angles between span(Ψ_r) of line trip events and span(V_r) of the data matrices of different events at bus 5

	span(V_r) of line trip	span(V_r) of load change	span(V_r) of short circuit
span(Ψ_r) of line trip 5-6	3.22°	8.85°	6.79°
span(Ψ_r) of line trip 26-29	5.78°	7.82°	26.79°

When span(Ψ_r) and span(V_r) correspond to line trip 5-6, the subspace angle is 3.22°, which is the smallest among all six cases. The subspace angle between span(Ψ_r) of the line trip 26-29 and span(V_r) of line trip 5-6 is also small (5.78°), even though the line trips happen at different locations. Thus, the subspaces of the same type of events are similar as long as the events are closely related. In contrast, the angle between span(Ψ_r) of a line trip event and span(V_r) of other events are much larger.

III. Subspace-based Event Identification

The proposed method is outlined in Fig. 3. It is centered on subspace-based event identification, including event detection and location for a practical implementation. A dictionary of events are constructed offline from sample PMU historical datasets during a window size T. Once an event is detected (through either the detection method here or any other event detector), all the PMU data in the next T time steps are collected to estimate the row subspace and column subspace. The event is identified through comparing the row subspace with the dictionary computed offline. The most influenced buses can be located by comparing the significance of changes at different locations.

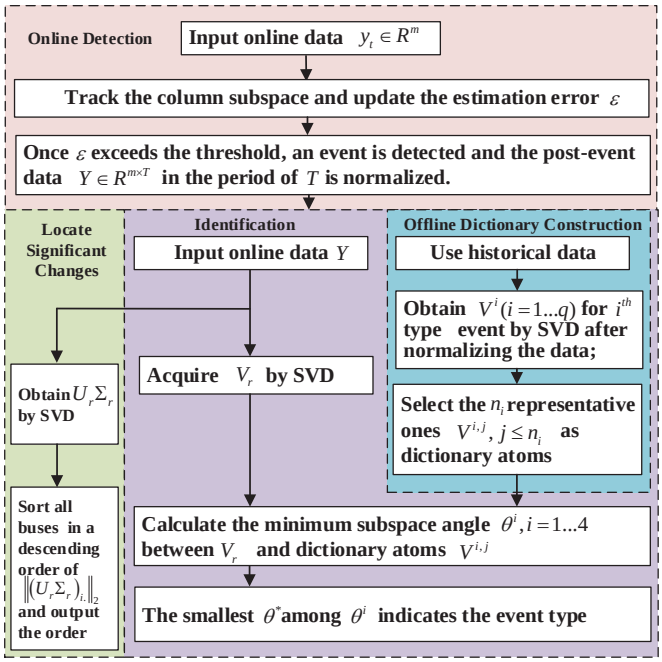

Fig. 3: The flow diagram of the online detection, identification and location method

The proposed method includes the following components.

1. Offline construction of an event dictionary

The dictionary is constructed from historical event datasets, as shown in Fig. 4. Given each dataset and parameters τ and Δ, we compute a low-dimensional row subspace from PMU data M within a window size T (e.g. thirty samples for one second of PMU data). These row subspaces are candidate dictionary atoms. Since a row subspace is robust to system initial conditions as described in Section II-B, row subspaces computed from events with the same type, at nearby location may have a small subspace angle. This property is exploited to reduce the dictionary size.

We refine the dictionary as follows. After selecting a subspace as an atom, all subspaces with angles less than a predefined threshold ϑ are removed from the dictionary. Let $\bar{V}^{i,j}$ denote the ith dictionary atom that corresponds to event type j. The total number of atoms in a dictionary is $\Sigma_{i=1}^{q} n_i$ for q types of events like line trip, load change, and three-phase short circuit events. The refinement significantly reduces the dictionary size. We will demonstrate that in Section IV.

2. Online event detection

We detect an event when the column subspace changes, similar to [32]. At time step t, the obtained measurement

[1]The rank of Ψ_r is always an even number due to the complex property and is not necessary to be the same with that of V_r.

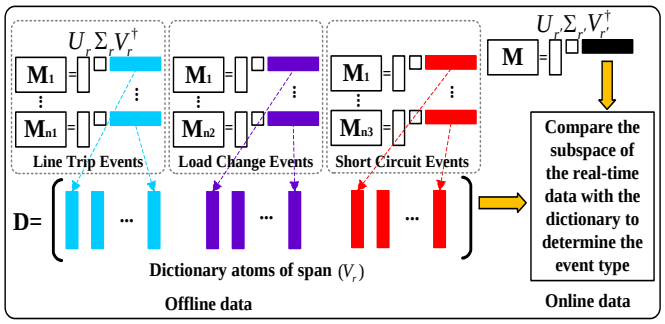

Fig. 4: Dictionary construction from historical datasets and real-time data identification through subspace comparison

vector $\mathbf{y}(t)$ is projected onto the current estimate of the column subspace with a basis $\hat{U}_r$. If the projection error, defined as $\varepsilon = \|\mathbf{y}_t - \hat{U}_r\hat{U}_r^\dagger \mathbf{y}(t)\|_2$, exceeds a predetermined threshold, then $\mathbf{y}(t)$ does not belong to span($\hat{U}_r$), and an event is declared to have happened. The column subspace can be estimated by SVD of past data in a fixed window or computed by subspace tracking methods like SPIRIT [41]. The event detection functionality is independent of others and can be achieved with other event detectors.

3. Online event identification based on the row subspace

Once an event is detected, all the PMU data in T time steps after an event happens are collected into a matrix. The row subspace spanned by V is computed by SVD [28] with r determined by (1)-(2), and when the two subspaces have different ranks the larger one is selected to maximize their difference. Define the minimum subspace angle θ^* with the dictionary as

$$\theta^* = \min_{j=1}^{q} \min_{i=1}^{n_j} \theta(V, \bar{V}^{i,j}). \tag{16}$$

If the minimum is achieved with some dictionary atom $\bar{V}^{i^*,j^*}$, the event is identified to be type j^*.

4. Location of most influenced buses

Here we use bus voltage magnitude measurements to identify the buses that are most influenced by the event, such that the operator can take proper measures in that area if needed.

Given the rank-r approximation $M_r = U_r\Sigma_r V_r^\dagger$ of a data matrix, the ith row of $U_r\Sigma_r$, denoted by $(U_r\Sigma_r)_{i\cdot}$, contains the coordinates of $(M_r)_{i\cdot}$ with respect to the basis V_r. Then $\|(U_r\Sigma_r)_{i\cdot}\|_2 = \|(M_r)_{i\cdot}\|_2$. Note that $(M_r)_{i\cdot}$ is the projection of the voltage magnitudes of bus i to the low-dimensional space. With the intuition that the bus most influenced by the event has the largest $\|(M_r)_{i\cdot}\|_2$, we sort the buses in a descending order of $\|(U_r\Sigma_r)_{i\cdot}\|_2$ and locate the most influenced k buses for a predetermined integer k.

5.External Event Identification with local data

A regional operator only has real-time data within its own control region and usually relies on manual-entry notification from other control regions about significant external events. Fast identification of significant external events helps an operator to make timely decisions. We can extend our method to identify external events in real time using only local data, provided that these events affect the local measurements.

To implement our method, one needs historical local measurements when an external event happens, as well as the information of event types. The operator builds a dictionary based on these historical events and compare the real-time data with the dictionary. All the steps remain the same. The size of the dictionary is generally larger than that needed to identify local events that are directly measured, because the impact on local measurements by external events is usually less significant than local events. In this case, the top k buses with the highest $\|(U_r\Sigma_r)_{i\cdot}\|_2$ values indicate the local buses that are most affected by the external event.

We remark that our method requires full data in T time steps to apply SVD. If some data points are lost during transmission or do not arrive at the operator within a specified waiting period, these measurements are considered as data losses. One could apply real-time missing data recovery methods such as [32] before executing the event identification method.

IV. Numerical validation on synthetic datasets

The method is tested on the 68-bus power system and 100MVA system base as shown in Fig. 2. Five types of events described are simulated: load change events (A 0.5 p.u. or 1.5 p.u. step change of active load input at certain bus is used to simulate the abrupt load change by ml_sig in PST.), line trip events, three-phase short circuit events (cleared after 0.2 second, and the corresponding line is tripped), and double line-to-ground short circuits (with zero sequence impedance 0.001 p.u. and negative sequence impedance 0.5 p.u.) and generator trip events. The data rate is set to be 30 samples per second. Results on voltage magnitude measurements are reported here, while we obtained similar results for phasor angle measurements. One second of data after an event starts are selected for analysis, i.e., $T = 30$, such that the fast dynamics are still maintained after the low-rank approximation. The observation window for short circuit events starts after the short is cleared.

Event datasets are generated with different pre-event system conditions. The difference between pre-event conditions are measured by the average relative difference of active power flow over all lines, denoted by $\bar{\eta}$. Specifically,

$$\bar{\eta}_a^\mu = (\Sigma_{j=1}^L |P_j^\mu - P_j^a|/|P_j^a|)/L, \tag{17}$$

where P_j^a and P_j^μ are active power flows (directed) at line j under condition "a" and μ, respectively. L is the total number of lines.

Table II: Different pre-event conditions

Labels of pre-event condition	Meanings	$\bar{\eta}_a^\mu$%
μ = "a"	the original condition	0
μ = "b"	different power injections	93.4
μ = "c"	different power injections	48.7
μ = "d"	lose line 8-9	28.9
μ = "e"	lose line 3-18	20.7
μ = "f"	lose line 8-9 and line 2-3	71.3
μ = "g"	lose line 4-14	15.4

A. Properties of the subspace representation

We first verify that the row subspace is robust to pre-event system conditions. Fig. 5 shows the voltage magnitudes of mainly influenced buses when the line connecting bus 1 and bus 2 is tripped at $t = 0.5$ second under two different system

conditions a and b. These differences in pre-event conditions lead to different voltage time series, thus directly comparing the time series measurements may miss the similarities of the same type of events.

We pick one second of data in both datasets starting from $t_1 = 0.55$ second. With $\tau = 0.99$ and $\Delta = 10$, the approximate rank defined in (1)-(2) is 6 for both data matrices. Fig. 6 shows the corresponding six principal right singular vectors, which form a basis of the row subspace. The angle of these two subspaces (defined in (11)) is 1.13°, indicating that these two subspaces are very close to each other. Thus, the row subspace can tolerate some variations of initial conditions.

Since the singular vectors can be viewed as linear combinations of the sinusoids at the dominate modes, their frequencies are closely related to the system modes, even though not exactly the same. To demonstrate this, we first estimate the oscillation modes employing the Prony method [42], which fits the time series $y(t)$ with a sum of sinusoids, i.e.,

$$y(t) = \Sigma_{i=1}^{n} S_i e^{(\lambda_i \Delta t)t}, \tag{18}$$

where S_i and $f_i = \lambda_i/(2\pi\Delta t)$ are the magnitude and frequency of each sinusoid to be estimated, and Δt is the step size. Here we employ one second of the bus voltage magnitude of bus 2 after the line 1-2 is tripped and fit the data with $n = 6$ and $\Delta t = 0.03$. S_i's and f_i's are estimated, and three dominant oscillation frequencies corresponding to the largest S_i's are 0.46 Hz, 1.28 Hz, and 2.11 Hz. In comparison, the approximated frequencies of singular vectors V_r in Fig. 6 are also measured. The frequencies that correspond to the dominate six singular vectors are 0.31 Hz, 0.49 Hz, 1.11 Hz, 1.23 Hz, 1.85 Hz, and 2.22 Hz. They are not exactly the same as the system modes because each singular vector could be a linear combination of a few modes.

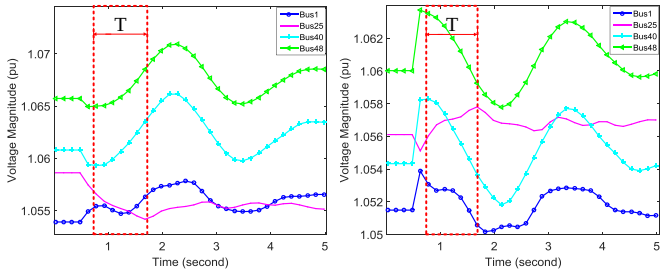

Fig. 5: The voltage magnitude of bus 1,25,40,48 when line 1-2 is tripped under pre-event condition "a" (left) ; under pre-event condition "b" (right)

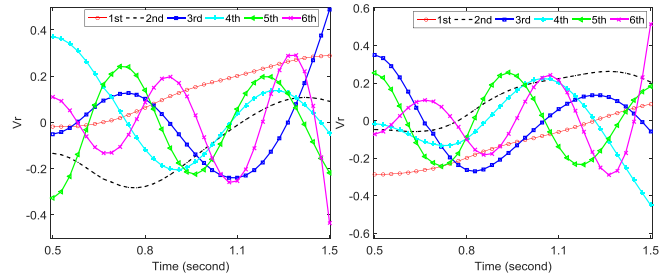

Fig. 6: The six principal singular vectors of V_r when line 1-2 is tripped under pre-event condition "a" (left); under pre-event condition "b" with $\eta_a^b = 93.45\%$ (right)

B. Dictionary construction

We simulate 250 events, including 64 line trip, 50 load change, 60 three-phase short circuit, 16 generator trip, and 60 double line-to-ground events at different locations with the same pre-event system condition "a". An event dictionary is constructed following Step 1 in Section III. $\tau = 0.99$, $\Delta = 10$, and $\vartheta = 5^\circ$. After the refinement, the dictionary has 70 atoms, listed in Table III. For LT (or DL) events, LT1-2 means the event at the line between bus 1 and 2. For TP (or LC), TP2 means a three phase short circuit event at bus 2. For GT, GT1 means tripping generator G1 in Fig. 2. Each atom corresponds to a row subspace, represented by a $30 \times r$ matrix $\bar{V}$ (r is usually $3 \sim 6$ here).

To avoid the Zero Distance problem of the subspace angle, when comparing two data matrices with approximate ranks r_1 and r_2 ($r_1 > r_2$), we compute the subspace angle of two rank-r_1 subspaces of these events. Thus, we store a $30 \times r_{\max}$ matrix for each event even though its approximate rank is less than $r_{\max}$. In the following experiments, $r_{\max}$ is set to 6.

Table III: The labels of different types of dictionary atoms

Dictionary Types (DT)	The Label of Dictionary Atoms
Line Trip (LT)	1-2, 4-14, 5-8, 12-11, 15-16, 30-32, 32-33, 33-34, 35-45, 42-41, 44-45
Three Phase (TP)	2, 3, 8, 12, 16, 25, 28, 31, 32, 33, 34, 35, 36, 37, 39, 43, 44, 50, 53, 56, 59, 62
Load Change (LC)	15, 16, 24, 32, 41, 42, 47, 48
Generator Trip (GT)	1, 2, 10, 11, 13, 14, 15
Double Line-to-ground (DL)	1-2,1-30, 1-31,2-3,9-36,10-11,10-13,13-14, 17-27,19-20,26-27,26-29,30-31, 30-32,31-38, 33-38,34-36,35-34,35-45,36-37,39-44,52-42

C. Event identification and location

To test the identification performance under various pre-event conditions resulting from difference power injections and topologies, we generate 500 events under pre-event conditions "b" and "c" due to different power injections and 750 events under conditions "d","e","f" due to different topologies. The conditions are described in Table II. Table IV records the minimum subspace angle between some sample testing case under condition "b" and the dictionary atoms of different types of events. The corresponding dictionaries closest to the events are listed in the brackets[2]. As highlighted in bold, the minimum subspace angle is achieved between a dictionary atom of the same event type and the test case. Subspace angles between a test case and dictionary atoms of different event types are generally much larger, with an exception that the subspace angle between a generator trip and a line trip can

Table IV: Minimum subspace angles between a test case with a pre-event condition "b" and dictionary atoms of one event type with pre-event condition "a". $\bar{\eta}_a^b = 93.45\%$.

Event \ DT	LC	LT	TP	GT	DL
LC_{17}	$\mathbf{2.6}^\circ(LC_{15})$	10.8°	9.5°	9.8°	14.6°
LT_{7-8}	12.2°	$\mathbf{0.7}^\circ(LT_{5-8})$	12.0°	3.0°	3.8°
TP_{17}	17.1°	7.9°	$\mathbf{4.4}^\circ(TP_3)$	9.4°	4.3°
GT_{12}	10.4°	$1^\circ(LT_{44-45})$	5.1°	$1.6^\circ(G_{10})$	2.0°
DL_{6-11}	13.6°	17.4°	8.8°	17.6°	$\mathbf{2.6}^\circ(DL_{4-5})$

[2]LC_{15} denotes the dictionary atom of load change event at bus 15; LT_{5-8} denotes that of the line trip event between line 5-8.

be small. For example, the subspace angle between GT_{12} and line trip 44-45 is $1.0°$, while the angle between GT_{12} and GT_{10} is $1.57°$. That is because the dominant features of a generator trip and a nearby line trip event are very similar when only one second of data are considered. In fact, generator and line trip events were previously differentiated based on measurements in a longer period, e.g., $15 \sim 30$ seconds of frequency measurements in [14], [20]. Although the subspaces are similar, a generator trip event is usually more significant than a line trip event. We can use the sum of singular values to further separate them.

If an event is (pre-)identified as a line or generator trip event after comparing with the dictionary, we compute the sum of its dominate singular values, denoted by $\xi = \sum_{i=1}^{r} \sigma_r$. If ξ is larger than a predetermined threshold ξ_0, this event is identified as a generator trip event. Otherwise, it is identified as a line trip event. ξ_0 is set to be $\xi_0 = c\bar{\xi}_{LT}$, where $\bar{\xi}_{LT}$ is the average ξ of all the line trip events in the training set and c is a margin constant. In this setup, the average sum of singular values of the training generator trip events is about eight times that of line trip events, so we set c to be 1.5. We also utilized $\xi_1 = 3\bar{\xi}_{DL}$ to further distinguish the three phase short circuit and double line-to-ground events, where $\bar{\xi}_{DL}$ is the average ξ of all the double line-to-ground events.

Table V records the identification and location results under the criteria as follows (Notice that the location of line trip events are considered as successful if one of the two related buses are correctly identified):

Identification Accuracy Rate (IAR): the ratio of the number of accurately identified events to the total number of events;

Average Rank (AR): the average rank of the correct bus location in the order we obtain in Step 4. The correct bus denotes the bus that is closest to the location of the event. For a line trip event that relates to two buses, we choose the bus with a smaller rank.

Table V: The IAR of different events on different conditions with the original dictionary

IAR of Events % \ Conditions	b	c	d	e	f	AR
Line Trip	95.3	95.3	100	95.7	100	1.7
Three Phase	96.7	100	86.8	95.6	88.2	1.25
Load Change	100	98	100	98.5	98.5	1.5
Generator Trip	93.8	93.8	93.8	93.8	87.5	2.38
Double Line-to-ground	95.0	90.0	96.7	95.2	96.7	1.4

Condition "b" and "c" denote the different initial conditions of diverse power injections, and most IAR on these conditions are higher than 90%; condition "d"-"f" have different initial conditions due to topology change at different locations, and their IAR are higher than 85%. AR is around 1 to 2, indicating that the location of the event usually corresponds to significant changes in the measurements and thus, is on the top of the sorted buses in Step 4.

We next analyze the reasons for wrong diagnosis. Take the testing events under condition "f" as an example. The incorrect identification mainly results from differences in the pre-event condition with that of the dictionary, as well as the similarities between some generator and line trip events. First, compared with condition "a", two lines (8-9 and 2-3) are open under condition "f", and the resulting η_a^f is as high as 71.3% as shown in Table II. The large differences in the initial condition will lead to non-trivial subspace angles between the same event under both conditions, which in turn lead to misidentification for some events. Second, tripping a heavily loaded line and a generator in the same coherent group (see [39], [43] for the definition of power system coherence) can result in similar dominant dynamics. For example, generator 10 provides 5 p.u. active power, which is lower than the average generator output of 11.5 p.u. Line 44-45 is a heavily loaded line carrying 2.6 p.u. active power. Tripping generator 10 and line 44-45 produce similar dynamics. That leads to misclassification of tripping generator 10 as a line-trip event.

D. Dictionary Enlargement

We then demonstrate the improvement of our method by enlarging the training set. We generate 250 events under condition "g" and build a dictionary using all 500 events under conditions "a" and "g". The resulting enlarged dictionary has 111 atoms, compared with the original 70 atoms. Table VI records the identification results with the enlarged dictionary. Compared with Table V, the identification performance improves when we have more training data. We also emphasize that our method does not require a large training set. A small training set such as our original training set of 250 events under condition "a" can lead to desirable results. We still employ the original training set in the following experiments.

Table VI: Event identification with an enlarged dictionary

IAR of Events % \ Conditions	b	c	d	e	f
Line Trip	95.3	100	100	95.7	100
Three Phase	96.7	100	97.1	100	95.6
Load Change	100	100	100	100	100
Generator Trip	93.8	93.8	93.8	93.8	93.8
Double Line-to-ground	95.0	90.0	96.7	95.2	96.7

E. Parameter sensitivity to Δ and τ

The rank r depends on Δ and τ in (1)-(2). We study the impact of Δ and τ on the identification performance of 250 events under condition "b". We fix $\tau = 0.99$ and change Δ from 10 to 1. As shown in Table VII, the change of IAR is generally less than 5%. We then fix $\Delta = 10$ and vary τ. When τ decreases, IAR slightly decreases because some dominant dynamics are ignored in the approximation. On the other hand, when the measurements are highly noisy, τ cannot be too high, since otherwise a large amount of noise would be maintained in the approximation. τ in $[0.97, 0.99]$ is appropriate for our synthetic and recorded datasets.

Table VII: IAR of events under condition "b" with various Δ's and τ's

IAR % of	$\tau = 0.99$			$\Delta = 10$		
	Δ =10	5	1	τ =0.99	0.98	0.97
Line Trip	95.3	92.2	90.6	95.3	86	80
Three Phase	96.7	93.3	91.7	96.7	91.7	93.3
Load Change	100	98	100	100	92.0	90.0
Generator Trip	93.8	93.8	93.8	93.8	87.5	87.5
Double Line-to-ground	95.0	96.7	96.7	95.0	96.7	96.7

F. Identification peformance of noisy data

We next test our identification method when the measurements contain noise. IEEE Standard C37.118 [44] does not specify the signal-noise-ratio(SNR) of PMU measurements. A SNR of 92 dB in voltage and frequency is simulated in [33]. The SNRs of current, voltage, and frequency measurements of PMUs in Public Service Company in New Mexico are measured to be about 45 dB [45]. We employ the method in [45] to compute the SNR of the recorded PMU data from ISO-NE (datasets in Section V), and the average SNR of voltage magnitudes is 86.6 dB.

Here, we test different SNRs from 40 dB to 100 dB. We add random Gaussian noise with zero mean to both the training set of 250 events under condition "a" and the testing set of 250 events under condition "b". We set $\tau = 0.97, \Delta = 10$. Tables VIII records the identification performance when SNR changes. When the SNR is relatively high, between 80 to 100 dB, we directly build a dictionary based on the noisy training set and test the performance on the noisy testing set. When the SNR is 100 dB, the resulting dictionary has 74 atoms. When the SNR is 80 dB, the dictionary has 128 atoms.

When the SNR is low, we first apply filters to reduce the noise in the training and testing sets. We then build a dictionary based on the filtered training set and test the performance on the filtered testing set. When the SNR is 60 dB, we apply an exponential filter [46] to the measurements of each PMU channel. Let $z(n)$ denote the input data and $\hat{z}(n)$ denote the output filtered data. The exponential filter averages the input with the previous output, i.e., $\hat{z}(n) = w\hat{z}(n-1)+(1-w)z(n)$, where $w \in (0,1)$ is the weight factor. Here we set $w = 0.1$. The exponential filter only needs to store the previous input, and its performance can be readily adjusted by tuning w. The constructed dictionary contains 138 atoms. When the SNR is 40 dB, the signal is so noisy that the exponential filter provide unsatisfactory denoising performance. We instead employ a moving average filter of order 20 [47]. The input $z(n)$ and output $\hat{z}(n)$ of the moving average filter is related by $\hat{z}(n) = \frac{1}{M}\Sigma_{j=0}^{M-1} z(n+j)$. M is 20 here. The moving average filter stores a longer history than the exponential filter and has a better denoising performance. The resulting dictionary contains 85 atoms.

In practice, if the measurements in a power system generally have high SNR, one can directly apply our method. If the SNR in a system is relatively low, one only needs to first employ filters to reduce the noise and then apply our method.

Table VIII: IAR of events under condition "b" with noisy measurements

SNR (dB)	40	60	80	100
Line trip IAR %	79.7	85.9	92.2	90.6
Three Phase IAR %	91.7	95.0	98.3	91.7
Load change IAR %	100	85.0	100	100
Generator trip IAR %	75.0	93.8	93.8	93.8
Double Line-to-ground IAR %	86.7	91.7	91.7	93.3
AR	11	5.0	1.9	1.7

G. Identification performance with other types of PMU measurements

Although we employ voltage magnitudes in the study, our method applies to other types of measurements like frequency, current magnitudes, and angles. We test both frequency and current magnitude and include the results on frequency here due to the page limit.

Table IX: Identification performance with frequency measurements

IAR of Events % \ Conditions	b	c	d	e	f
Line Trip	90.0	85.1	78.4	80.4	88.1
Three Phase	84.3	85.8	89.7	90.9	94.1
Load Change	100	100	100	100	100
Generator Trip	81.2	81.2	68.7	75	62.5
Double Line-to-ground	97.0	97.0	98.4	96.8	98.4

We set $\tau = 0.99$, $\Delta = 10$ as before and construct a dictionary of 94 atoms from 250 events under condition "a". If an event is (pre-)identified as a line trip and a generator trip, we further employ a threshold $\xi_2 = 3\bar{\xi}_{LT}$ to distinguish line trip (below ξ_2) and generator trip events (above ξ_2), where $\bar{\xi}_{LT}$ is the average sum of singular values of the frequency measurements in the training sets of line trip events. Similarly, since some three-phase short circuit events and double line-to-ground events have similar dynamics but different magnitudes, we further set a threshold $\xi_3 = 1.5\bar{\xi}_{DL}$ to distinguish them, where DL stands for double line-to-ground. The performance on voltage magnitude is slighter better than that of frequency measurements.

H. External event identification with local data

The 68-bus power system can be divided into five coherent groups (A1-A5) [43]. We treat the combination of A1, A2 and A3 as a local area of 26 buses and treat A4 and A5 as external areas, as shown in Fig. 2. With the local measurements of different external events, a dictionary of 35 atoms is constructed, including 20 atoms for line trip events and 15 for three phase short circuit events. Then we test the identification performance of 40 three-phase short circuit events and 40 line trip events occurring outside the local area. The event types of all test cases are correctly identified with more than IAR=97%. Some sample test cases are shown in Table X. LT_{2-3} and TP_4 represent a line trip event and a short circuit event between line 2-3 and bus 4, respectively. One can see that the similarities among the events of the same type are still preserved for external events.

The buses with large $\|(U_r\Sigma_r)_{i.}\|_2$ can provide information about the location of the external event. When the external line 6-11 (in area A5) is tripped (see Fig. 2), the top 5 buses are buses 9, 30, 1, 31, and 38. They are close to the boundary between the local area and A5. When line 47-48 (in area A4) is tripped, the top 5 bus are buses 49, 46, 38, 31 and 1. They are close to the boundary between the local area and A4.

Table X: Event identification results with local data

Dictionary Types	External LT Events		External TP Events	
	LT_{2-3}	LT_{11-12}	TP_4	TP_{18}
Line Trip	$\mathbf{1.85}^\circ$	$\mathbf{2.71}^\circ$	21.66°	15.17°
Three Phase	10.44°	9.02°	$\mathbf{3.54}^\circ$	$\mathbf{1.83}^\circ$

V. Experiment Results on Recorded PMU datasets

We also test our method on recorded PMU datasets in ISO-New England. That includes 32 events (8 line trips, 11 faults

and 13 load change), which happened on different days and at different times of the day. The line trip events are scheduled line trips without faults. The load change events correspond to load (pump storage) shedding. The faults are short circuit events.

The average SNR of the recorded PMU data estimated by the method in [45] is 86.6 dB and we employ the exponential filter to reduce the noise with $w = 0.7$.

We randomly select 22 events as training datasets and 10 as testing datasets. For each event, we select the voltage magnitudes of 136 buses in one second after the event occurs. $\tau = 0.97$ and $\Delta = 10$ in (1)-(2). A dictionary of six atoms (two for each type) is constructed based on the training datasets. The dimension of each dictionary atoms is $30 \times r_{max}$ with $r_{max} = 4$.

Fig. 7 shows the magnitudes of some bus voltages under different events. Note that (a) and (b) are both line trip events at different locations. Although the shapes of their curves look different, the corresponding low-dimensional subspace of Line trip B has a subspace angle of $6.05°$ with that of Line trip A. In contrast, the minimum angle between the subspace of line trip B and the dictionary of load change events (or the dictionary of short circuit events) is $11.2°$ (or $30.21°$).

Table XI: Minimum subspace angles between a test case and the dictionary atoms of three types of events in recorded PMU data

Events \ DT	Load Change	Fault	Line Trip
Load Change 1	**4.08°**	16.91°	18.29°
Load Change 2	**3.12°**	20.81°	14.39°
Fault 1	24.95°	**6.33°**	23.86°
Fault 2	8.93°	**3.73°**	15.76°
Line Trip 1	7.25°	5.85°	**3.93°**
Line Trip 2	11.20°	30.21°	**4.27°**

Table XII: Performance of the testing recorded PMU datasets

Type of events	Line Trip	Fault	Load Change
IAR %	100	100	100
AR	3	2.8	2.5

All the 10 testing datasets are successfully identified and the sample results are shown in Table XI. Load change 1 and 2 denote pump storages shedding at different locations and time instants. The minimum subspace angle is reached between an event and the dictionary atoms of the same event type. The minimum subspace angle between Load Change 1 and the dictionary atoms of Load Change is $4.08°$, while the angles with the dictionary atoms of short circuit and line trip events are $16.91°$ and $18.29°$, respectively. Thus, the event type can be correctly determined by the minimum subspace angle. The exact locations of the events are confidential and not included here, but AR is shown in Table XII. It shows that the location of an event is within top three buses in the order output by Step 4.

VI. Conclusions

This paper develops a data-driven real-time event identification method that identifies an event within seconds. The new idea here is to characterize an event by a low-dimensional row subspace of a high-dimensional spatial-temporal PMU data matrix. The row subspace largely depends on the eigenvalues

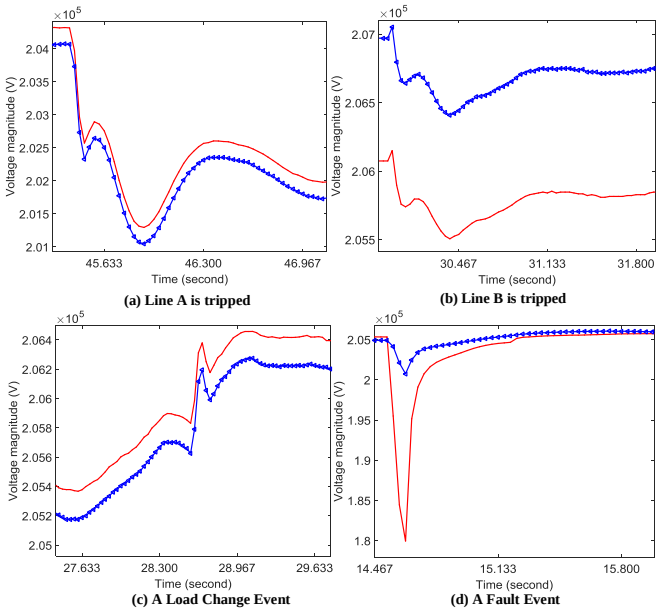

Fig. 7: Voltage magnitudes of three types of events in New England.

of the system state matrix and is a compact characteristic of system dynamics without power system modeling. The proposed method constructs a dictionary of row subspaces of different events offline and identifies an event online by comparing the subspace of the obtained data with the dictionary. This subspace representation reduces the dictionary size while maintaining the identification accuracy. The proposed method is further extended to identify events outside a control region only based on measurements within a control region. The method is verified on simulated datasets in the IEEE 68-bus test system and recorded PMU datasets in New England. Numerical studies of our method on the sensitivity to system power injections and topology changes, dictionary size, parameter selections, measurement noise are reported.

Future work includes identification of successive events including cascading failures and the extension of the method to microgrids in non-interconnected and remote zones.

Acknowledgement

We thank Frankie Qiang Zhang, Xiaochuan Luo, and Slava Maslennikov at ISO-NE for valuable discussions and suggestions throughout this project and for providing us the recorded PMU data in New England. This research is supported in part by the ERC Program of NSF and DoE under the supplement to NSF Award EEC-1041877 and the CURENT Industry Partnership Program, and in part by NSF Grant 1508875.

References

[1] W. Li, M. Wang, and J. H. Chow, "Fast event identification through subspace characterization of PMU data in power systems," *accepted to IEEE Power and Energy Society (PES) General Meeting*, 2017.

[2] A. Phadke and J. Thorp, *Synchronized phasor measurements and their applications*. Springer, 2008.

[3] J. N. Bank, R. M. Gardner, J. K. Wang, A. J. Arana, and Y. Liu, "Generator trip identification using wide-area measurements and historical data analysis," in *Proc. IEEE PES Power Systems Conference and Exposition*, 2006, pp. 1677–1681.

[4] J. E. Tate and T. J. Overbye, "Line outage detection using phasor angle measurements," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1644–1652, 2008.

[5] H. Zhu and G. B. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2215–2224, 2012.

[6] P. Bhui and N. Senroy, "Online identification of tripped line for transient stability assessment," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2214–2224, 2016.

[7] 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.

[8] M. L. Crow and A. Singh, "The matrix pencil for power system modal extraction," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 501–502, 2005.

[9] G. Liu, J. Quintero, and V. M. Venkatasubramanian, "Oscillation monitoring system based on wide area synchrophasors in power systems," in *Proc. iREP Symposium-Bulk Power System Dynamics and Control-VII. Revitalizing Operational Reliability*. IEEE, 2007, pp. 1–13.

[10] D. Trudnowski, J. Johnson, and J. Hauer, "Making prony analysis more accurate using multiple signals," *IEEE Trans. Power Syst.*, vol. 14, no. 1, pp. 226–231, 1999.

[11] M. J. Smith and K. Wedeward, "Event detection and location in electric power systems using constrained optimization," in *Proc. IEEE Power Energy Society General Meeting*, 2009, pp. 1–6.

[12] P. G. Axelberg, I. Y.-H. Gu, and M. H. Bollen, "Support vector machine for classification of voltage disturbances," *IEEE Trans. Power Del.*, vol. 22, no. 3, pp. 1297–1303, 2007.

[13] T. S. Bi, X. N. Song, J. T. Wu, and Q. X. Yang, "Novel method for disturbance identification in power systems," in *Proc. IEEE Power Engineering Society General Meeting*, 2006.

[14] A. Bykhovsky and J. H. Chow, "Power system disturbance identification from recorded dynamic data at the northfield substation," *International Journal of Electrical Power & Energy Systems*, vol. 25, no. 10, pp. 787 – 795, 2003.

[15] O. P. Dahal, S. M. Brahma, and H. Cao, "Comprehensive clustering of disturbance events recorded by phasor measurement units," *IEEE Trans. Power Del.*, vol. 29, no. 3, pp. 1390–1397, 2014.

[16] A. K. Ghosh and D. L. Lubkeman, "The classification of power system disturbance waveforms using a neural network approach," *IEEE Trans. Power Del.*, vol. 10, no. 1, pp. 109–115, 1995.

[17] T. Y. Ji, Q. H. Wu, L. Jiang, and W. H. Tang, "Disturbance detection, location and classification in phase space," *IET Generation, Transmission Distribution*, vol. 5, no. 2, pp. 257–265, 2011.

[18] H. Jiang, J. J. Zhang, and D. W. Gao, "Fault localization in smart grid using wavelet analysis and unsupervised learning," in *Proc. Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2012, pp. 386–390.

[19] Y. Song, W. Wang, Z. Zhang, H. Qi, and Y. Liu, "Multiple event detection and recognition for large-scale power systems through cluster-based sparse coding," *IEEE Trans. Power Syst.*, 2017.

[20] W. Wang, L. He, P. Markham, H. Qi, Y. Liu, Q. C. Cao, and L. M. Tolbert, "Multiple event detection and recognition through sparse unmixing for high-resolution situational awareness in power grid," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1654–1664, 2014.

[21] T. Yin, S. S. Wulff, J. W. Pierre, D. Duan, D. J. Trudnowski, and M. Donnelly, "Initial investigation of data mining applications in event classification and location identification using simulated data from minniwecc," in *Proc. North American Power Symposium (NAPS)*. IEEE, 2016, pp. 1–6.

[22] J. E. Tate and T. J. Overbye, "Double line outage detection using phasor angle measurements," in *Proc. IEEE Power Energy Society General Meeting*, 2009, pp. 1–5.

[23] J. Wu, J. Xiong, and Y. Shi, "Efficient location identification of multiple line outages with limited pmus in smart grids," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1659–1668, 2015.

[24] Y. Zhao, J. Chen, A. Goldsmith, and H. V. Poor, "Identification of outages in power systems with uncertain states and optimal sensor locations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 6, pp. 1140–1153, 2014.

[25] P. Bhui and N. Senroy, "Online identification of tripped line for transient stability assessment," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2214–2224, 2016.

[26] M. Garcia, T. Catanach, S. Vander Wiel, R. Bent, and E. Lawrence, "Line outage localization using phasor measurement data in transient state," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3019–3027, 2016.

[27] G. Rovatsos, X. Jiang, A. D. Domínguez-García, and V. V. Veeravalli, "Statistical power system line outage detection under transient dynamics," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2787–2797, 2017.

[28] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[29] C. G. Baker, K. A. Gallivan, and P. Van Dooren, "Low-rank incremental methods for computing dominant singular subspaces," *Linear Algebra and its Applications*, vol. 436, no. 8, pp. 2866–2888, 2012.

[30] H. Jiang, J. J. Zhang, W. Gao, and Z. Wu, "Fault detection, identification, and location in smart grid based on data-driven computational methods," *IEEE Trans. Smart Grid*, vol. 5, no. 6, pp. 2947–2956, 2014.

[31] H. Jiang, X. Dai, W. Gao, J. Zhang, Y. Zhang, and E. Muljadi, "Spatial-temporal synchrophasor data characterization and analytics in smart grid fault detection, identification and impact causal analysis," *Smart Grid*.

[32] 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.

[33] 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.

[34] 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.

[35] M. Soltanolkotabi, E. Elhamifar, E. J. Candes *et al.*, "Robust subspace clustering," *The Annals of Statistics*, vol. 42, no. 2, pp. 669–699, 2014.

[36] G. Zuccon, L. A. Azzopardi, and C. Van Rijsbergen, "Semantic spaces: Measuring the distance between different subspaces," in *International Symposium on Quantum Interaction*. Springer, 2009, pp. 225–236.

[37] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*. McGraw-hill New York, 1994, vol. 7.

[38] P. M. P. M. Anderson, A. A. A.-A. A. Fouad, I. of Electrical, and E. Engineers, *Power system control and stability*, 2nd ed. Piscataway, N.J. : IEEE Press : Wiley-Interscience, 2003, previous ed.: Ames : Iowa State University Press, 1977. [Online]. Available: http://swbplus.bsz-bw.de/bsz105088617cov.htm

[39] J. H. Chow, *Power system coherency and model reduction*. Springer, 2013.

[40] J. H. Chow and K. W. Cheung, "A toolbox for power system dynamics and control engineering education and research," *IEEE Trans. Power Syst.*, vol. 7, no. 4, pp. 1559–1564, 1992.

[41] S. Papadimitriou, J. Sun, and C. Faloutsos, "Streaming pattern discovery in multiple time-series," in *Proceedings of the 31st international conference on Very large data bases*. VLDB Endowment, 2005, pp. 697–708.

[42] J. F. Hauer, C. Demeure, and L. Scharf, "Initial results in prony analysis of power system response signals," *IEEE Trans. Power Syst.*, vol. 5, no. 1.

[43] G. Rogers, *Power system oscillations*. Springer Science & Business Media, 2012.

[44] K. E. Martin, "Synchrophasor measurements under the ieee standard c37. 118.1-2011 with amendment c37. 118.1 a," *IEEE Transactions on Power Delivery*, vol. 30, no. 3, pp. 1514–1522, 2015.

[45] M. Brown, M. Biswal, S. Brahma, S. J. Ranade, and H. Cao, "Characterizing and quantifying noise in pmu data," in *Power and Energy Society General Meeting (PESGM), 2016*. IEEE, 2016, pp. 1–5.

[46] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2014.

[47] S. W. Smith *et al.*, "The scientist and engineer's guide to digital signal processing," 1997.

Wenting Li (S'16) received the B.E. degree from Harbin Institute of Technology, Heilongjiang, China, in 2013.

She is pursuing the Ph.D. degree in electrical engineering at Rensselaer Polytechnic Institute, Troy, NY. Her research interests include high dimensional data analysis, event detection, identification and location in power system with PMU data.

Meng Wang (M'12) received the Ph.D. degree from Cornell University, Ithaca, NY, USA, in 2012.

She is an Assistant Professor in the department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute. Her research interests include high dimensional data analysis and their applications in power systems monitoring and network inference.

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 a Professor of Electrical, Computer, and Systems Engineering. His research interests include multivariable control, power system dynamics and control, FACTS controllers, and synchronized phasor data.