Robust Recovery of PMU Signals with Outlier Characterization and Stochastic Subspace Selection

Kaustav Chatterjee, *Student Member, IEEE,* Kaveri Mahapatra, *Student Member, IEEE,* and Nilanjan Ray Chaudhuri, *Senior Member, IEEE*

***Abstract*—This paper proposes an improvement on the standalone robust principal component analysis (R-PCA)-based approach for recovering clean signals from corrupted synchrophasor measurements. The contributions of this paper are twofold. First, a kernel principal component analysis (K-PCA)-based metric is proposed for detecting and differentiating event-induced outliers from spurious outliers in data, which is then used as an indicator to suspend R-PCA in the event window to minimize the overall error in signal recovery. Second, a formal approach based on the recursive Bayesian framework is proposed for selecting the most appropriate subspace from a library of subspaces to be used by R-PCA. The paper combines the ideas of robust signal recovery, corruption-resilient event outlier detection, and stochastic subspace selection into a composite approach for correcting anomalies in synchrophasor data. The effectiveness of the proposed methodology is validated on simulated data from IEEE $16-$machine, $5-$area test system.**

***Index Terms*—PMU Data Quality, Anomaly Correction, Robust PCA, Kernel PCA, Recursive Bayesian Estimation.**

NOMENCLATURE

r	Sample number (and window number)
p	Number of PMU channels (measurements)
$\mathbf{x}^{(r)}$	Vector of PMU measurements$-$ observed signal values at the r^{th} sample
$\boldsymbol{\ell}^{(r)}$	True measurements$-$ uncorrupted signal values at the r^{th} sample
$\boldsymbol{\delta}^{(r)}$	Vector of additive signal corruptions
$\hat{\boldsymbol{\delta}}^{(r)}$	Estimated value of the signal corruptions
$\hat{\boldsymbol{\ell}}^{(r)}$	Reconstructed signal$-$ clean measurements recovered from $\mathbf{x}^{(r)}$
$\phi(\mathbf{x}^{(r)})$	Mapping of $\mathbf{x}^{(r)}$ in feature space
N	Number of samples in a data window
$\boldsymbol{\Phi}^{(r)}$	r^{th} data window mapped in feature space
$\mathbf{C}^{(r)}$	Covariance matrix corresponding to $\boldsymbol{\Phi}^{(r)}$
$\lambda_j^{(r)}, \mathbf{v}_j^{(r)}$	j^{th} eigenpair associated with $\mathbf{C}^{(r)}$
$\mathbf{K}^{(r)}$	Kernel matrix corresponding to $\boldsymbol{\Phi}^{(r)}$
$N\lambda_j^{(r)}, \boldsymbol{\alpha}_j^{(r)}$	j^{th} eigenpair associated with $\mathbf{K}^{(r)}$
$\zeta_j^{(r)}$	Norm of principal component (PC) score of the window $\boldsymbol{\Phi}^{(r)}$ along direction $\mathbf{v}_j$
q	Degree of the polynomial kernel
S_i	i^{th} low-dimensional subspace in the library
N_s	Total number of subspaces in the library
ρ	Dimension of each subspace in the library
$\mathbf{U}_i$	Matrix of ρ orthonormal bases spanning S_i
$\mathcal{Z}^{(r)}$	Set of observations $\{\mathbf{x}^{(r_0)} \dots \mathbf{x}^{(r-1)}, \mathbf{x}^{(r)}\}$ from end of an event till any r^{th} sample
$\mathbf{m}$	Random variable (r.v.) indicating the choice of a subspace
m_i	Value of the r.v. for choice of S_i
$\mathbb{P}(m_i\|\mathcal{Z}^{(r)})$	Probability that S_i is the true subspace given the observation set $\mathcal{Z}^{(r)}$
$\mathbb{P}(\mathbf{x}^{(r)}\|m_i)$	Likelihood that $\mathbf{x}^{(r)}$ lies in the span of S_i

Authors are with the School of Electrical Engineering and Computer Science, The Pennsylvania State University, State College, PA 16801, USA (e-mail: kuc760@psu.edu, kzm221@psu.edu, nuc88@psu.edu)

Financial support from the NSF grant under award CNS 1739206 is gratefully acknowledged.

I. INTRODUCTION

HIGH precision time-synchronized data from Phasor Measurement Units (PMUs) have enhanced situational awareness and decision-making in modern power systems [1]. However, reliability of these applications is dependent on the quality of the sensor data and the security of the cyber-infrastructure relaying them. Often data packets are lost or delayed in transit, and measurements are corrupted with colored noise, spurious outliers, and non-zero biases [2]. Moreover, as outlined in [3] and [4], a dedicated intranet-based communication network in NASPInet architecture is not immune to cyber attacks. Attackers with malicious intent can breach the security in the cyber infrastructure and corrupt the data streams with erroneous inputs [5].

There exists extensive literature on detection of anomalies in PMU data including [6]–[11] and references therein. Detection methods include Bayesian approximation-based approaches [6], Kalman filter-based approaches [7], maximum likelihood estimation methods [8], principal component analysis (PCA)-based approaches [9], and ensemble learning methods [11] among others. However, challenge remains in recovering clean measurements from corruptions with strong spatio-temporal correlation.

To that end, the low-rank property of a time-window of correlated measurements has been exploited in [12]–[14] for estimating missing data samples. A density-based spatial clustering approach has been proposed in [15] for online attack detection, classification, and recovery of clean data from corrupted measurements. In [16] recovery of missing data samples is posed as a low-rank matrix completion problem, and is solved using alternating direction method of multipliers (ADMM). Extending the nature of anomalies beyond missing samples and spurious outliers, authors in [17] have used a principal component pursuit (PCP)-based approach for correcting corruptions involving parameter manipulation, noise injection, and fault replay attacks. However, these block-processing

techniques are less suitable for real-time applications since they require finite waiting time in filling a data window.

In our recent work [10], a robust principal component analysis (R-PCA)-based vector-processing approach is proposed for recovering clean signals from online streaming of corrupted measurements. Building on the theory of compressed sensing [18]–[20], the proposed algorithm in [10] identifies the corrupted channels and reconstructs clean samples by solving a sparse recovery problem using a library of low-rank subspaces derived offline from simulated data.

Although the proposed approach in [10] is successful in correcting anomalies in PMU data, few challenges remain. It fails to distinguish between outliers due to bad data and those induced by power system events. As a consequence, it is prone to false positives in an event window. In this context, we define an *event* as the onset of a disturbance, and the time-window capturing the event-induced outlier as an *event window*, as shown in Fig. 1. Also, reliable signal recovery in the post-event oscillatory period and estimation of system modes following an event is dependent on the right choice of subspaces. It is possible for the post-event subspace to be radically different from the pre-event subspace. Therefore, there remains a challenge in accurately updating the low-rank subspace following the occurrence of an event, which induces a change in network topology or operating condition. Next, we describe these challenges in more detail, identify the gaps in literature, and highlight the contributions of the paper.

II. Challenges, Gaps, and Contributions

1) Separation of event outliers from bad data

Onset of a disturbance like a fault or tripping of a line or a generator might induce outliers in the system response (henceforth, referred to as 'event outliers') that do not belong to the low-dimensional pre-event subspace. Therefore, a standalone R-PCA-based signal recovery algorithm [10] using the pre-event subspace may wrongly classify the event-induced outliers as bad data and attempt to correct them. It will be shown in Section IV that this can result in incorrect reconstruction of voltage sag during fault, leading to potential errors in event localization. In this regard, we suggest to integrate an event outlier detector in the signal recovery framework. Following the detection of an event outlier, we propose to suspend signal reconstruction and use raw PMU data until the end of the event window. However, it is also important to ensure that the event outlier detection is itself immune to bad data and should not wrongly stop the signal recovery algorithm at spurious outliers.

Extensive literature exists on use of multivariate statistical methods for event detection and localization in power systems including [21]–[24] and references therein. In [21], [22] authors have combined the classical PCA with T^2 and Q statistics to capture the occurrence of an event and to separate it from normal operating conditions. A major limitation of PCA-based methods is in the assumption of linear relationship between the measured variables. With the inherent nonlinearity in the system, such assumptions may lead to inaccurate results [25]. Improvements on this have been suggested in [23] and [24] using a kernel PCA (K-PCA)-based nonlinear technique.

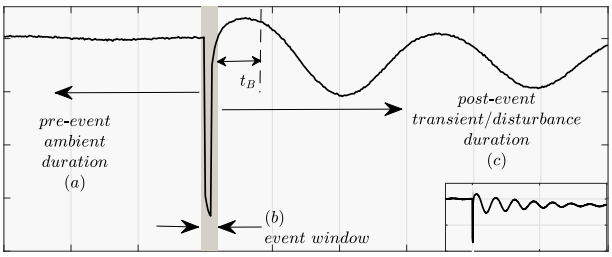

Fig. 1: Pre-event ambient duration, event window, and post-event transient duration. Inset: signal with different pre-event and post-event equilibriums.

However, these use large data windows in detection, which increases the computational requirements. When adapted to smaller windows, the metrics used in detection are sensitive to noise and data anomalies, and therefore cannot be used here.

Contribution: In this paper, we extend the K-PCA framework to formulate a corruption-resilient event detection metric. We propose that with an appropriate choice of a kernel function, the difference between the largest eigenvalues of the kernel matrices derived from two successive data windows can be used in detecting event-induced outliers and the onset of an event window. Unlike the T^2 and Q statistics presented in literature [22], [23], the proposed method is insensitive to noise, corruption, and spurious outliers. Since eigenvalue calculation is computationally intensive, our focus has also been on reducing the computational burden without compromising the detector selectivity. We derive an upper bound on the difference of the eigenvalues as a closed form expression of the data samples without performing eigen-decomposition on the kernel matrices. Further, it is shown that the bound itself can act as a detector, thereby significantly reducing computation in real-time. In addition, a rank-based measure has been proposed for detecting the end of an event window (see, Fig. 1) and the transition to post-event transients. These detection metrics are then used to suspend signal recovery in the event window and eliminate the error due to inaccurate reconstruction of event outliers. □

In [26], a PMU data recovery approach is proposed which exploits the Hankel structure derived from a window of measurements in detection and correction of spurious anomalies. The algorithm in [26] has a subroutine designed for detection of events, which is based on the count of channels with steady mismatch between the reported and reconstructed signal values over a time window. The detection, therefore, is not instantaneous and has to wait for a window of samples to decide whether the mismatch is due to an event or bad data. In contrast, the K-PCA-based metric proposed in this paper detects the event at the instant of its onset, and thus is more suited for online applications. Moreover, the event detection method in [26] places complete trust on the reported samples and ignores the results of reconstruction in the period of mismatch, including the duration of post-event oscillations following the event-outlier. In this context we would like to highlight that most wide-area monitoring applications rely on accurate recovery of these transients for inferring the stability of the post-event system. Therefore, in this paper, we limit our trust on the data beyond the window containing the event outlier. The recovery of post-event transients using the

proposed approach is discussed next.

2) Subspace selection and recovery of post-event transients

As mentioned previously, the post-event operating condition could be different from the condition preceding the event, and therefore, reliable signal recovery during the post-event transient (see, Fig. 1) is contingent upon fast and accurate estimation of the new subspace. In [26] the subspace is formed from the incoming data samples. However, in the initial duration following an event with only a few samples observed, the subspace estimation from a short window of data can be challenging due to the sensitivity of singular vectors to noise and nonlinearity. As we shall show later in Section IV (see, *ST* approach) this error can propagate during reconstruction of samples over time. To resolve this, in [10] it was assumed that the network connectivity information from the topology processor of SCADA EMS is available and can be used in selecting a suitable subspace from a library of pre-computed subspaces. In this paper, we would like to limit our reliance on the data from the Network Topology Processor as SCADA is more prone to attacks and has a much slower polling rate compared to PMUs.

Contributions: We adopt a data-driven recursive Bayesian framework for selecting the most appropriate subspace from the library for use in post-event signal recovery. Following the occurrence of an event, the proposed method derives the likelihoods of all subspaces in the library from the PMU measurements at each instant, and recursively computes their probabilities of being the true subspace. The subspace with probability approaching 1 is chosen for signal recovery. We use an efficient termination criteria for accelerating the selection process. It is also shown that the proposed selection methodology is robust to random missing samples and uncorrelated spurious outliers. However, to avoid error due to choice of incorrect subspace, signal reconstruction is suspended during the selection process. □

Note that as opposed to ignoring the anomaly correction for a significant duration in the post-event transient window, in our proposed approach we suspend signal recovery only in the window of the event outlier and in the duration of subspace selection t_B, a much lesser time span, as shown in Fig. 1.

The rest of this paper is organized as follows. In Section III, the proposed approach for robust anomaly correction is presented, which combines the attributes of corruption-resilient event detection, stochastic subspace selection, and robust signal recovery. To that end, the K-PCA-based approach for detecting event outliers and the Bayesian approach for subspace selection are described in detail. Finally in Section IV, case studies are presented to highlight the efficacy of the proposed approach, followed by conclusions in Section V.

Notations: Matrices are denoted by bold uppercase, vectors by bold lowercase, and scalars by normal font. Exponent (r) denotes the the window number when used on matrix, and the sample number when used on a vector or scalar. For example, $\mathbf{X}^{(r)}$ is a matrix derived from the r^{th} window, $\mathbf{x}^{(r)}$ is a vector denoting the r^{th} sample, and $x^{(r)}$ is a scalar computed at the r^{th} instant. $\phi(\cdot)$ is a vector function, and $\phi^T(\mathbf{x}^{(r)})$ denotes transpose of the function value.

III. Proposed Robust Anomaly Correction Approach

Addressing the challenges highlighted in the previous section, we propose a coordinated approach for correcting anomalies in PMU data. As discussed, it has three main functionalities – (A) *Corruption-resilient detection of event-induced outliers* – identifying the onset of an event and suspending signal recovery in event window, (B) *Stochastic subspace selection* – choosing a subspace from library that best approximates the post-event transient condition, and (C) *Robust signal recovery* – anomaly detection and signal reconstruction in the post-event duration using the selected subspace. A brief overview of the proposed approach described in Fig. 2 is discussed next.

At any instant, the incoming data sample is first passed through the K-PCA-based event detection algorithm to check for event outliers. If an event outlier is detected, the event sign flag ES and the subspace estimation flag $\mathcal{M}$ are set to 1, signalling the onset of event window. In this duration (region (b) of Fig. 1) signal recovery is suspended and the incoming data samples are passed without processing. At the end of event window (detected using rank-based measure, discussed later), the flag ES toggles back to 0 with $\mathcal{M}$ still at 1. This marks the transition to post-event event transients and initiates the subspace selection process. In this duration, the incoming data samples are used to update the posterior probabilities of the library subspaces using the recursive Bayesian approach. Eventually one of these subspaces is selected at the end of duration t_B (see, Fig. 1), and flag $\mathcal{M}$ returns to 0. For every data sample thereafter, this subspace is used in signal recovery.

If an event outlier is not detected in the incoming data sample (implying that the data is from the regions (a) or (c) of Fig. 1) and if a subspace has already been selected at a prior instant, signal recovery is performed using that subspace. Note that in Fig. 2 we initialize flags as $ES = 0$, $\mathcal{M} = 1$, this is to ensure that at the very beginning the algorithm has a subspace selected from the library before it proceeds with signal recovery. The action of the algorithm on different signal regions with respective flag statuses have been outlined in Tables I and II.

TABLE I: FLAG STATUS AND SIGNAL RECOVERY

Flag Status	**Duration** (see, Fig. 1)	**Signal Recovery**
$ES = 0$, $\mathcal{M} = 0$	Pre-event or post-event duration beyond t_B	YES
$ES = 1$, $\mathcal{M} = 1$	Event window	NO
$ES = 0$, $\mathcal{M} = 1$	Subspace selection duration t_B	NO

TABLE II: FLAG STATUS AND ACTIONS IN THE ALGORITHM

Flag Status	**Actions**
$ES = 0$, $\mathcal{M} = 0$	Perform K-PCA-based event outlier detection Perform R-PCA-based signal recovery
$ES = 1$, $\mathcal{M} = 1$	Perform incremental rank computation to detect end of event window
$ES = 0$, $\mathcal{M} = 1$	Perform recursive Bayesian estimation Perform K-PCA-based event outlier detection

Assumptions and justifications: We make the following assumptions in the paper regarding the nature of data anomalies and availability of subspaces in library.

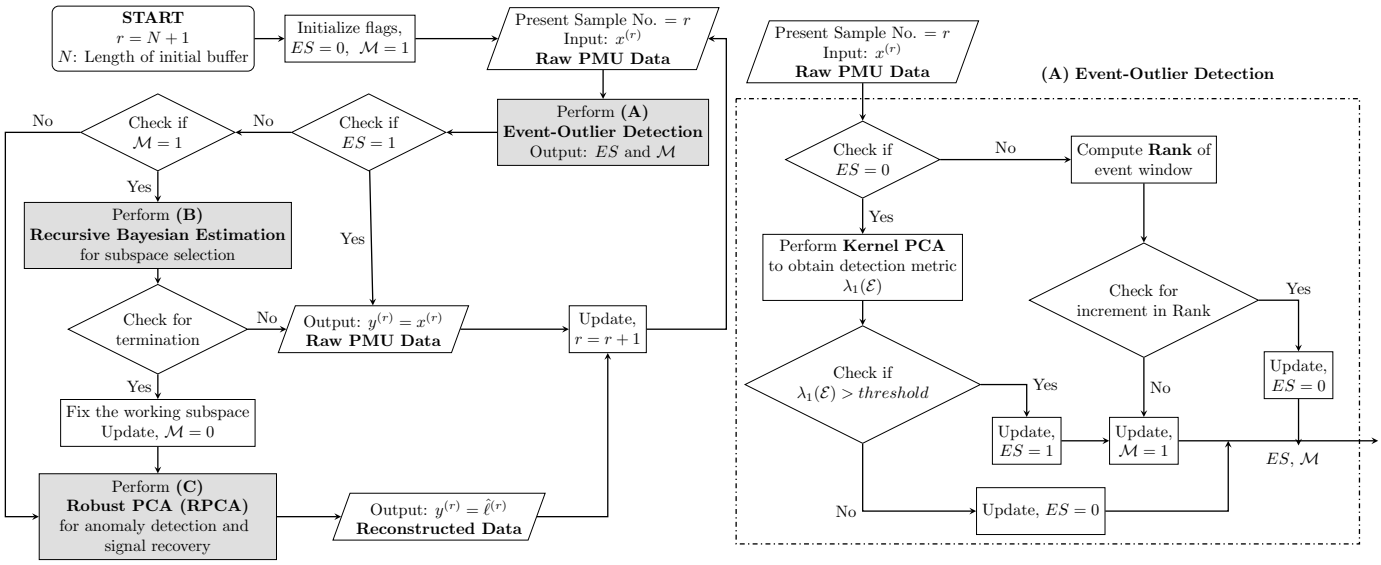

Fig. 2: (left) Flowchart describing the proposed robust anomaly correction approach. (right) Expanded view of the event outlier detection approach.

(1) We assume that at any given instant of time the corruption in data is limited to only a small fraction of the total number of available PMU channels. We believe this to be a realistic assumption for wide-area monitoring applications, since PMUs are spread across wide geographies and therefore, coordinated corruption and correlated anomaly in a large fraction of channels is improbable.

(2) We assume the availability of close-by subspaces (from $\mathcal{N}-1$ contingency studies). Ideally, this requires an exhaustive library with subspaces corresponding to each $\mathcal{N}-1$ condition. Our assertion is if subspaces in the library are exhaustive enough in capturing possible post-contingency modal signatures – that should be adequate. Fortunately, for a realistic meshed bulk power system, such a set of subspaces should be a reasonably small subset of subspaces corresponding to all $\mathcal{N}-1$ contingencies. Thus, it is realistic to assume that the library will have subspaces close to all post-event conditions.

The detailed approaches for event detection, subspace selection, and signal recovery outlined as (A), (B), and (C) in the flowchart of Fig. 2 are discussed next.

A. Corruption-resilient Detection of Event-induced Outliers: Kernel PCA-based Approach

Kernel principal component analysis [27] is a generalization of PCA for nonlinear dimensionality reduction. It is a two-step process: 1) Mapping the input data to a higher dimensional feature space (where data points are linearly separable), and 2) Using PCA in feature space for dimensionality reduction. K-PCA does not explicitly compute the mapping to the higher dimensional feature space, but uses a function of the measured data to encode the mapping information. The choice of this function (called 'kernel') thus decides the efficacy of the mapping and the analysis that follows in the feature space.

Overview of the kernel method: Let $\mathbf{x}^{(r)}$ be the vector of p PMU measurements at the r^{th} sampling instant. Consider a moving window of N such vectors. At each sampling instant the vector of the latest observations is added to the window as the last column and the oldest observations from the first column are discarded. Detrending is performed to ensure that each PMU signal has zero mean and unit variance in consecutive overlapping windows, pairwise taken at a time. Let $\tilde{\mathbf{x}}^{(r)}$ be obtained after detrending $\mathbf{x}^{(r)}$ in the windows r and $r+1$ taken together, and $\phi(\tilde{\mathbf{x}}^{(r)})$ be the mapping of $\tilde{\mathbf{x}}^{(r)}$ in a high-dimensional feature space. Without loss of generality, let us assume that the data in feature space is centered.

Let $(\lambda_j^{(r)}, \mathbf{v}_j^{(r)})$ be any j^{th} eigenpair of the covariance matrix $\mathbf{C}^{(r)}$, corresponding to the r^{th} window of data mapped in feature space. Therefore, for this window we can write,

$$\mathbf{C}^{(r)}\mathbf{v}_j^{(r)} = \frac{1}{N}\sum_{i=0}^{N-1}\phi(\tilde{\mathbf{x}}^{(r+i)})\phi^T(\tilde{\mathbf{x}}^{(r+i)})\mathbf{v}_j^{(r)} = \lambda_j^{(r)}\mathbf{v}_j^{(r)} \quad (1)$$

$$\mathbf{v}_j^{(r)} = \sum_{i=0}^{N-1}\phi(\tilde{\mathbf{x}}^{(r+i)})\left\{\frac{\phi^T(\tilde{\mathbf{x}}^{(r+i)})\mathbf{v}_j^{(r)}}{N\lambda_j^{(r)}}\right\} \triangleq \sum_{i=0}^{N-1}\phi(\tilde{\mathbf{x}}^{(r+i)})\alpha_{j_i}^{(r)} \quad (2)$$

Eigenvectors $\mathbf{v}_j^{(r)}$ are the principal directions in feature space. Substituting (2) in (1) and pre-multiplying by $\phi^T(\mathbf{x}^{(r+l)})$,

$$\sum_{i=0}^{N-1}\left\{\phi^T(\tilde{\mathbf{x}}^{(r+l)})\phi(\tilde{\mathbf{x}}^{(r+i)})\sum_{n=0}^{N-1}\phi^T(\tilde{\mathbf{x}}^{(r+i)})\phi(\tilde{\mathbf{x}}^{(r+n)})\alpha_{j_n}^{(r)}\right\} = N\lambda_j^{(r)}\sum_{n=0}^{N-1}\phi^T(\tilde{\mathbf{x}}^{(r+l)})\phi(\tilde{\mathbf{x}}^{(r+n)})\alpha_{j_n}^{(r)} \quad (3)$$

Let $\mathbf{\Phi}^{(r)} \triangleq [\ \phi(\tilde{\mathbf{x}}^{(r)}) \quad \dots \quad \phi(\tilde{\mathbf{x}}^{(r+N-1)})\ ]$, and $\mathbf{K}^{(r)} \triangleq \mathbf{\Phi}^{(r)^T}\mathbf{\Phi}^{(r)}$ be the inner product matrix in feature space with $\mathbf{k}_{i,j}$ as the $(i,j)^{\text{th}}$ element. Then (3) can be rewritten as,

$$\sum_{i=0}^{N-1}\mathbf{k}_{l+1,i+1}^{(r)}\sum_{n=0}^{N-1}\mathbf{k}_{i+1,n+1}^{(r)}\alpha_{j_{n+1}}^{(r)} = N\lambda_j^{(r)}\sum_{n=0}^{N-1}\mathbf{k}_{l+1,n+1}^{(r)}\alpha_{j_{n+1}}^{(r)} \quad (4)$$

Concatenating the expression in (4) for $l = 0, 1, \dots N-1$,

$$(\mathbf{K}^{(r)})^2\boldsymbol{\alpha}_j^{(r)} = N\lambda_j^{(r)}\mathbf{K}^{(r)}\boldsymbol{\alpha}_j^{(r)} \Longrightarrow \mathbf{K}^{(r)}\boldsymbol{\alpha}_j^{(r)} = N\lambda_j^{(r)}\boldsymbol{\alpha}_j^{(r)} \quad (5)$$

where, $\boldsymbol{\alpha}_j^{(r)} \triangleq [\ \alpha_{j_1}^{(r)} \quad \dots \quad \alpha_{j_N}^{(r)}\ ]^T$.

$$\mathbf{K}^{(r)} \triangleq \left[\begin{array}{c|c} a & \mathbf{c}^T \\ \hline \mathbf{c} & \mathbf{A} \end{array}\right] \left\{ \begin{array}{|c|c|c|} \hline (\tilde{\mathbf{x}}^{(r)^T}\tilde{\mathbf{x}}^{(r)})^q & (\tilde{\mathbf{x}}^{(r)^T}\tilde{\mathbf{x}}^{(r+1)})^q & \dots \quad (\tilde{\mathbf{x}}^{(r)^T}\tilde{\mathbf{x}}^{(r+N-1)})^q \\ \hline (\tilde{\mathbf{x}}^{(r+1)^T}\tilde{\mathbf{x}}^{(r)})^q & (\tilde{\mathbf{x}}^{(r+1)^T}\tilde{\mathbf{x}}^{(r+1)})^q & \dots \quad (\tilde{\mathbf{x}}^{(r+1)^T}\tilde{\mathbf{x}}^{(r+N-1)})^q \quad (\tilde{\mathbf{x}}^{(r+1)^T}\tilde{\mathbf{x}}^{(r+N)})^q \\ \vdots & \vdots & \ddots \quad \vdots \quad \vdots \\ (\tilde{\mathbf{x}}^{(r+N-1)^T}\tilde{\mathbf{x}}^{(r)})^q & (\tilde{\mathbf{x}}^{(r+N-1)^T}\tilde{\mathbf{x}}^{(r+1)})^q & \dots \quad (\tilde{\mathbf{x}}^{(r+N-1)^T}\tilde{\mathbf{x}}^{(r+N-1)})^q \quad (\tilde{\mathbf{x}}^{(r+N-1)^T}\tilde{\mathbf{x}}^{(r+N)})^q \\ \hline & (\tilde{\mathbf{x}}^{(r+N)^T}\tilde{\mathbf{x}}^{(r+1)^T})^q & \dots \quad (\tilde{\mathbf{x}}^{(r+N)^T}\tilde{\mathbf{x}}^{(r+N-1)})^q \quad (\tilde{\mathbf{x}}^{(r+N)^T}\tilde{\mathbf{x}}^{(r+N)})^q \\ \hline \end{array} \right\} \left[\begin{array}{c|c} \mathbf{A} & \mathbf{b} \\ \hline \mathbf{b}^T & d \end{array}\right] \triangleq \mathbf{K}^{(r+1)}$$

Fig. 3: Kernel matrices for r^{th} and $(r+1)^{\text{th}}$ windows of detrended PMU data

From (5), $(N\lambda_j^{(r)}, \boldsymbol{\alpha}_j^{(r)})$ is an eigenpair of $\mathbf{K}^{(r)}$ in the window of interest. Further, since $\mathbf{C}^{(r)}$ is real symmetric, the eigenvectors can be ensured to be orthonormal if,

$$\mathbf{v}_j^{(r)^T}\mathbf{v}_j^{(r)} = (\boldsymbol{\alpha}_j^{(r)})^T\mathbf{K}^{(r)}\boldsymbol{\alpha}_j^{(r)} = 1 \implies \left\|\boldsymbol{\alpha}_j^{(r)}\right\|_2 = \frac{1}{\sqrt{N\lambda_j^{(r)}}}$$

The reproducing kernel Hilbert space property allows the inner product $\phi^T(\tilde{\mathbf{x}}^{(i)})\phi(\tilde{\mathbf{x}}^{(j)})$ in feature space to be expressed as a positive definite symmetric function of pre-images in input space. This is commonly referred to as the '*kernel trick*', and the positive definite symmetric function $\mathcal{K}$ is called a *kernel*. Some commonly used kernels include Gaussian kernel, linear kernel, polynomial kernel, etc [27]. In this paper, we use the polynomial kernel described in (6). For a polynomial kernel the entries of the inner product matrix $\mathbf{K}^{(r)}$, also called the kernel matrix, can be expressed as,

$$\begin{aligned} \mathbf{k}_{i+1,j+1}^{(r)} &= \phi^T(\tilde{\mathbf{x}}^{(r+i)})\phi(\tilde{\mathbf{x}}^{(r+j)}) = \mathcal{K}(\tilde{\mathbf{x}}^{(r+i)}, \tilde{\mathbf{x}}^{(r+j)}) \\ &= (\tilde{\mathbf{x}}^{(r+i)^T}\tilde{\mathbf{x}}^{(r+j)})^q \quad \forall q \geq 1, \;\; \forall i,j \in \{0,1,\dots N-1\} \end{aligned} \tag{6}$$

Hence, the kernel matrix corresponding to the r^{th} window $\mathbf{K}^{(r)}$, and the $(r+1)^{\text{th}}$ window $\mathbf{K}^{(r+1)}$ can be written as shown in Fig. 3. If the mapped data in feature space is not centered, one can alternatively center the kernel matrix as described in [27]. The analysis hereafter shall consider centered kernels.

Onset of an event window: We detect events, which induce outliers at its onset. As described previously, these outliers can challenge the accuracy of R-PCA-based signal recovery. We propose that the occurrence of an event is captured in the projection of $\boldsymbol{\Phi}^{(r)}$ along the direction of maximum variance in feature space. The magnitude of this quantity is same as the norm of the PC score along the first principal direction $\mathbf{v}_1$. The PC score norm $\zeta_1^{(r)}$ of a window of mapped data along $\mathbf{v}_1^{(r)}$ can be computed as,

$$\zeta_1^{(r)} = ||\mathbf{v}_1^{(r)^T}\boldsymbol{\Phi}^{(r)}||_2 = N\lambda_1^{(r)}||\boldsymbol{\alpha}_1^{(r)^T}||_2 = \sqrt{N\lambda_1^{(r)}} \tag{7}$$

We define a metric, that uses the difference in PC score norms between two consecutive time-windows to detect the onset of an event. The norm of the projection, as in (7), is also the square root of the largest eigenvalue of the kernel matrix $\mathbf{K}^{(r)}$ corresponding to the r^{th} window. With a suitable choice of kernel function it can be ensured that the largest eigenvalue corresponding to a window with an incoming event outlier is significantly higher compared to the windows containing ambient data, and thus can be used in detecting the onset of an event. However, this requires an eigen-decomposition to be performed on the matrix $\mathbf{K}^{(r)}$ of size $N \times N$. To circumvent this, we derive an upper bound on the change in $(\zeta_1^{(r)})^2$ between successive time-windows in terms of inner products on data samples, and show that the bound can itself act as a detector, thereby reducing computation in real-time.

Referring to Fig. 3, let $\mathbf{P_r}\mathbf{K}^{(r)}\mathbf{P_c} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{c} \\ \hline \mathbf{c}^T & a \end{array}\right] \triangleq \mathbf{B}$ where, $\mathbf{P_r} = \left[\begin{array}{c|c} \mathbf{0}_{(N-1)\times 1} & \mathbf{I}_{(N-1)\times(N-1)} \\ \hline 1 & \mathbf{0}_{1\times(N-1)} \end{array}\right]$ and $\mathbf{P_c} = \mathbf{P_r}^{-1}$

From (5), the eigen-decomposition of $\mathbf{K}^{(\mathbf{r})}$ can be written as,

$$\begin{aligned} &\mathbf{K}^{(r)}\boldsymbol{\alpha}_j = N\lambda_j^{(r)}\boldsymbol{\alpha}_j^{(r)} \\ \implies &\mathbf{P_r}\mathbf{K}^{(r)}\mathbf{P_c}\hat{\boldsymbol{\alpha}}_j^{(r)} = N\lambda_j^{(r)}\mathbf{P_r}\mathbf{P_c}\hat{\boldsymbol{\alpha}}_j^{(r)} = N\lambda_j^{(r)}\hat{\boldsymbol{\alpha}}_j^{(r)} \\ \implies &\mathbf{K}^{(r)} \text{ and } \mathbf{P_r}\mathbf{K}^{(r)}\mathbf{P_c} \text{ have same eigenvalues.} \end{aligned} \tag{8}$$

From Fig. 3, $\mathbf{K}^{(r+1)} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{b} \\ \hline \mathbf{b}^T & d \end{array}\right] \triangleq \mathbf{B} + \mathcal{E}$, where $\mathcal{E}$ is the perturbation in the kernel with respect to the previous window.

$$\mathcal{E} = \left[\begin{array}{c|c} \mathbf{0}_{(N-1)\times(N-1)} & \mathbf{b}-\mathbf{c} \\ \hline (\mathbf{b}-\mathbf{c})^T & d-a \end{array}\right] \tag{9}$$

Using the classical perturbation bound [28] between two successive windows, the absolute value of the difference in the square of PC score norms along $\mathbf{v}_1$ can be expressed as,

$$\begin{aligned} \Delta(\zeta_1)^2 &= |(\zeta_1^{(r+1)})^2 - (\zeta_1^{(r)})^2| = |N\lambda_1^{(r+1)} - N\lambda_1^{(r)}| \\ &= |\lambda_1(\mathbf{K}^{(r+1)}) - \lambda_1(\mathbf{K}^{(r)})| = |\lambda_1(\mathbf{B}+\mathcal{E}) - \lambda_1(\mathbf{B})| \\ &\leq \|\mathcal{E}\|_2 = \lambda_1(\mathcal{E}) \end{aligned} \tag{10}$$

where, $\lambda_1(\mathcal{E})$ denotes the largest eigenvalue of $\mathcal{E}$ given by,

$$\lambda_1(\mathcal{E}) = \frac{1}{2}\left\{(d-a) + \sqrt{(d-a)^2 + 4(\mathbf{b}-\mathbf{c})^T(\mathbf{b}-\mathbf{c})}\right\} \tag{11}$$

$\Delta(\zeta_1)^2$, and its bound $\lambda_1(\mathcal{E})$ are our proposed metrics for detecting event outliers. Between two successive windows r and $r+1$, the term $d-a$ equals to $||\tilde{\mathbf{x}}^{(r+N)}||_2^{2q} - ||\tilde{\mathbf{x}}^{(r)}||_2^{2q}$ and each entry in the vector $\mathbf{b}-\mathbf{c}$ equals to $(\tilde{\mathbf{x}}^{(l+1)^T}\tilde{\mathbf{x}}^{(r+N)})^q - (\tilde{\mathbf{x}}^{(l+1)^T}\tilde{\mathbf{x}}^{(r)})^q$, for $l = r, \dots (r+N-1)$. If both the windows are from ambient condition, then the difference between the outgoing data vector $\tilde{\mathbf{x}}^{(r)}$ and the incoming vector $\tilde{\mathbf{x}}^{(r+N)}$ is small – only source of difference being the ambient noise and minor variations about equilibrium. However, if $\tilde{\mathbf{x}}^{(r+N)}$ corresponds to an event outlier, the difference would be significantly larger compared to the previous case.

Detection threshold: Let us consider a simple case with 1 p.u. pre-fault voltage magnitude at all p channels and that each voltage dips to 0 at the fault instant. For any i^{th} signal (channel), the measurement row in the overlapping windows r and $r+1$ (taken together) can then be

expressed as, $x_i = [\mathbf{1}_{1\times N} \quad 0]$. The sample mean and standard deviation of the signal in these combined windows are $\frac{N}{N+1}$ and $\frac{1}{\sqrt{N+1}}$, respectively. Therefore, on detrending, $\tilde{x}_i = [\frac{1}{\sqrt{N+1}}\mathbf{1}_{1\times N} \quad \frac{-N}{\sqrt{N+1}}]$. Since the fault is assumed to be observable in all p signals in same extent, $\tilde{\mathbf{x}}^{(r+N)} = \frac{-N}{\sqrt{N+1}}\mathbf{1}_{p\times 1}$ and $\tilde{\mathbf{x}}^{(r+i)} = \frac{1}{\sqrt{N+1}}\mathbf{1}_{p\times 1}, \forall i \in \{0, 1, \dots N-1\}$. Therefore, from (7), $a = (\frac{p}{N+1})^q$, $d = (\frac{N^2 p}{N+1})^q = aN^{2q}$, $\mathbf{b} = a(-N)^q\mathbf{1}_{(N-1)\times 1}$, and $\mathbf{c} = a\mathbf{1}_{(N-1)\times 1}$. Substituting these in (11), $\lambda_1(\mathcal{E}) = \{a(N^{2q} - 1) + \sqrt{a^2(N^{2q}-1)^2 + 4a^2(N-1)(-1+(-N)^q)^2}\}/2$. For sufficiently large q, $\lambda_1(\mathcal{E}) \to \mathcal{O}(aN^{2q}) = \mathcal{O}\big((pN)^q\big)$. This gives the order of magnitude of $\lambda_1(\mathcal{E})$ for an event outlier.

In most realistic cases, the extent of the fault would not be same in all signals. We express this empirically by introducing a correction factor $\beta_1 \in (0, 1)$ to p in the expression of $\lambda_1(\mathcal{E})$. The choice of β_1 shall decide the threshold $(\beta_1 pN)^q$, in detecting the event outlier. $\lambda_1(\mathcal{E})$ exceeding this predefined threshold, indicates the onset of an event window.

If $\tilde{\mathbf{x}}^{(r+N)}$ corresponds to a bad data or malicious corruption, assuming only a small fraction of PMUs can be corrupted simultaneously, we can write $\lambda_1(\mathcal{E}) \to \mathcal{O}\big((\beta_2 pN)^q\big)$. Here, β_2 ($< \beta_1$) captures the effect of number of affected channels and degree of corruption. Since $\beta_2 < \beta_1 < 1$, for large values of q, $(\beta_2 pN)^q << (\beta_1 pN)^q$.

Comments on choice of polynomial kernel: For understanding why a polynomial kernel works on our data set, let us refer to the discussions in Appendix B.1 of [27]. It explains that for a map resulting from a polynomial kernel of degree q, the terms corresponding to q^{th} order correlations are weighed with an extra factor $\sqrt{q!}$ compared to the self-terms. Therefore, K-PCA with polynomial kernels should mainly pick up variance for q^{th} order correlations. In our case, this is of merit because PMU measurements from multiple channels are correlated for event-induced outliers as opposed to anomalies and corruptions appearing only in a small fraction of channels. Therefore, event outliers would get separated from other data classes in the direction of the q^{th} order correlation in feature space. Moreover, higher the degree of the polynomial, higher is the weighing factor and therefore, greater is the separation in feature space. Hence, we recommend using higher order polynomial kernels in event outlier detection.

End of an event window: In this approach, we prefer voltage magnitude and frequency signals for detection of onset and end of event window. Following the event outlier detection, we initiate a progressively increasing event window that accumulates new data samples at each time-step. We compute the rank of this matrix at each instant. As long as the event persists, the numerical rank of this window would be 1. The instant when the event ends (e.g. fault is cleared) and a measurement vector from the post-event system enters the window, the numerical rank would increase to 2. The increment in rank indicates the end of an event window. At this point, the stochastic subspace selection is initiated, which is described next.

B. Stochastic Subspace Selection: Bayesian Approach

As mentioned previously, the R-PCA-based algorithm needs a subspace representing the present operating condition, which is selected from a subspace library. Such a library consists of a large number of low-rank subspaces computed from offline planning simulations and archived data, considering plausible nominal and off-nominal scenarios in the operating envelop.

Let us assume that there are N_s subspaces in the library. We propose a probabilistic framework based on recursive Bayesian approach for selecting the most appropriate subspace.

Let $\mathcal{Z}^{(r)} \triangleq \{\mathbf{x}^{(r_0)} \dots \mathbf{x}^{(r-1)}, \mathbf{x}^{(r)}\}$ be the set of PMU observations, starting at the r_0^{th} sample at the end of event window till any r^{th} instant ($r > r_0$). Let $\mathbb{P}(m_i|\mathcal{Z}^{(r)})$ be the posterior probability that subspace S_i is the true subspace, given the observation set $\mathcal{Z}^{(r)}$. Then, following the framework of recursive Bayesian estimation [29], the posterior probability of subspace S_i at the r^{th} instant can be written as,

$$\mathbb{P}(m_i|\mathcal{Z}^{(r)}) = \frac{\mathbb{P}(\mathbf{x}^{(r)}|m_i)\mathbb{P}(m_i|\mathcal{Z}^{(r-1)})}{\sum_{j=1}^{N_s}\mathbb{P}(\mathbf{x}^{(r)}|m_j)\mathbb{P}(m_j|\mathcal{Z}^{(r-1)})} \tag{12}$$

where, $\mathbb{P}(\mathbf{x}^{(r)}|m_i)$ is the likelihood that $\mathbf{x}^{(r)}$ lies in the span of S_i. The likelihood can be assumed to be a normal distribution in $e_i^{(r)}$, where $e_i^{(r)}$ is defined as the distance [20] between the observation span containing $\mathbf{x}^{(r)}$ and the subspace S_i. Let $\mathbf{U}_i$ be the matrix of $\rho-$orthonormal bases spanning S_i. Then, $e_i^{(r)} = ||(\mathbf{I} - \mathbf{U_i}\mathbf{U_i}^T)\mathbf{Z}^{(r)}||_2$, with $\mathbf{Z}^{(r)} = [\mathbf{x}^{(r_0)} \dots \mathbf{x}^{(r)}]$. The closer the span to the subspace, the higher is the likelihood of $\mathbf{x}^{(r)}$ belonging to S_i. Eqn (12) can then be expressed as,

$$\mathbb{P}(m_i|\mathcal{Z}^{(r)}) = \frac{exp\{-\frac{1}{2}C_f(e_i^{(r)})^2\}\mathbb{P}(m_i|\mathcal{Z}^{(r-1)})}{\sum_{j=1}^{N_s}exp\{-\frac{1}{2}C_f(e_j^{(r)})^2\}\mathbb{P}(m_j|\mathcal{Z}^{(r-1)})} \tag{13}$$

At $r = r_0$, all subspaces are initialized with equal probabilities, $\mathbb{P}(m_i|\mathcal{Z}^{(r_0)}) = \frac{1}{N_s}, \forall i \in \{1, 2, .., N_s\}$. The likelihoods are computed at each time-step and the posterior probabilities are updated recursively. Since the likelihood $\mathbb{P}(\mathbf{x}^{(r)}|m_i)$ for the true subspace is expected to take high values consistently at every time-step, the corresponding posterior probability would eventually approach 1. All other subspaces would be rejected as their probabilities converge to 0. The rate of convergence depends on the value of C_f.

In case the subspace exactly matching the operating condition is not present in the library, the proposed approach selects the subspace that best approximates the operating condition. However, the process might take finitely many iterations for the corresponding probability to approach 1. Any attack on PMU data in this duration might jeopardize the selection process. Hence, the subspace selection mechanism must be accelerated and the desired subspace should be chosen in the least possible time-steps. The termination criteria as derived in [30] is used to stop the Bayesian estimation and identify the desired subspace much before the probability reaches 1.

Termination criteria: As in [30], if we assume that the likelihoods vary little about their expected values, then, for simplifying the analytical deductions we may replace $\mathbb{P}(\mathbf{x}^{(r)}|m_i)$ by its expected value L_i. The difference in probabilities between

two successive time-steps can then be expressed as,

$$\mathbb{P}(m_i|\mathcal{Z}^{(r+1)}) - \mathbb{P}(m_i|\mathcal{Z}^{(r)}) = \frac{\mathbb{P}(m_i|\mathcal{Z}^{(r)})\Big\{L_i - \sum_{j=1}^{N_s} L_j \mathbb{P}(m_j|\mathcal{Z}^{(r)})\Big\}}{\sum_{j=1}^{N_s} L_j \mathbb{P}(m_j|\mathcal{Z}^{(r)})} \quad (14)$$

Further, we define the denominator in (14) as the weighted likelihood, $L_w^{(r)}$. With assumptions of minor variations in likelihoods, it is derived in [30] that $L_w^{(r)}$ is monotonically non-decreasing. Thus, subspaces for which $L_i < L_w^{(r)}$, it is implied that $\mathbb{P}(m_i|\mathcal{Z}^{(r+1)}) < \mathbb{P}(m_i|\mathcal{Z}^{(r)})$, and hence, their posterior probabilities would monotonically decrease to 0 [30]. Although nothing can be said about subspaces for which $L_i > L_w^{(r)}$. However, from the theory of recursive Bayesian estimation we know that there can be only one subspace for which the posterior probability would approach one. Therefore, if at any point in time there is only one subspace for which $L_i > L_w^{(r)}$, it can be identified as the true subspace and the process can be terminated [30]. The respective probabilities for all other subspaces would eventually converge to 0.

Remark: The proposed approach is robust to single or multiple uncorrelated outliers and missing data values. However, coordinated attacks like fault injections with strong spatio-temporal correlation can jeopardize the selection process. In this paper, we shall assume that no attack happens while the posterior probabilities are estimated. This is a weak assumption, since high values of C_f can ensure that termination is achieved in less than a second of end of event window. However, it is not desirable to keep C_f so high that termination is achieved before the likelihoods have settled around their expected values. That would violate the assumptions made in deriving the termination criteria. □

Once, the Bayesian estimation is terminated, flag $\mathcal{M}$ toggles back to 0 (see, Fig. 2), and the subspace selected from the library is used in the robust signal recovery discussed next.

C. Robust Signal Recovery: Robust PCA-based Approach

As shown in Fig. 2, robust recovery is performed on the data vector $\mathbf{x}^{(r)}$ when flags ES and $\mathcal{M}$ are both 0. The idea is to decompose $\mathbf{x}^{(r)}$ into vectors $\boldsymbol{\ell}^{(r)}$ and $\boldsymbol{\delta}^{(r)}$, such that $\boldsymbol{\ell}^{(r)}$ is in the low-rank subspace spanned by true measurements and $\boldsymbol{\delta}^{(r)}$is the vector of additive signal corruptions, $\mathbf{x}^{(r)} = \boldsymbol{\ell}^{(r)} + \boldsymbol{\delta}^{(r)}$. Since, we have assumed at any instant the corruption is limited to a small fraction of channels, $\boldsymbol{\delta}^{(r)}$ is sparse.

Corresponding to the present operating condition, let S_i be the subspace selected from library by the Bayesian approach (see, Section III-B). Next, $\mathbf{x}^{(r)}$ is projected onto the subspace orthogonal to S_i as in (15)

$$\boldsymbol{\gamma}^{(r)} = \mathbf{Q}\mathbf{x}^{(r)} = \mathbf{Q}(\boldsymbol{\ell}^{(r)} + \boldsymbol{\delta}^{(r)}) = \mathbf{Q}\boldsymbol{\delta}^{(r)} + \boldsymbol{\nu}^{(r)} \quad (15)$$

where, $\mathbf{Q} = \mathbf{I} - \mathbf{U}_i\mathbf{U}_i^T$ and $\boldsymbol{\gamma}^{(r)}$ is the projection of $\mathbf{x}^{(r)}$ in the span normal to S_i. Since $\boldsymbol{\ell}^{(r)} \in S_i$, the projection ensures that the contribution of $\boldsymbol{\ell}^{(r)}$ is nullified while preserving that of $\boldsymbol{\delta}^{(r)}$. However, because of noise and due to approximations in limiting the dimension of the subspace to ρ, the term $\mathbf{Q}\boldsymbol{\ell}^{(r)}$ is not exactly zero, but negligibly small as captured in $\boldsymbol{\nu}^{(r)}$.

The recovery of the sparse corruption vector $\boldsymbol{\delta}^{(r)}$ from $\mathbf{x}^{(r)}$ can be posed as an optimization problem [19] shown in (16).

$$\min_{\boldsymbol{w}^{(r)}} \left\|\boldsymbol{w}^{(r)}\right\|_1 \quad \text{s.t.} \quad \left\|\boldsymbol{\gamma}^{(r)} - \mathbf{Q}\boldsymbol{w}^{(r)}\right\|_2 \leq \eta^{(r)} \quad (16)$$

The optimal solution to (16), $\boldsymbol{w}^{*(r)}$ is an estimate of corruption in the measurement vector, $\hat{\boldsymbol{\delta}}^{(r)} = \boldsymbol{w}^{*(r)}$. The estimated clean signal can then be obtained as, $\hat{\boldsymbol{\ell}}^{(r)} = \mathbf{x}^{(r)} - \hat{\boldsymbol{\delta}}^{(r)}$. The thresholding term $\eta^{(r)}$ in (16) is updated in every time-step as, $\eta^{(r)} = ||\mathbf{Q}\hat{\boldsymbol{\ell}}^{(r-1)}||_2$. Henceforth, in this paper we shall refer to $\hat{\boldsymbol{\ell}}^{(r)}$ as the reconstructed PMU data at the r^{th} instant, which is the output of this algorithm.

Algorithm 1 Robust Anomaly Correction

Input: $\mathbf{x}^{(r)}$, p, N, q, β_1, C_f
Output: $\mathbf{y}^{(r)}$
Initialization : $ES \leftarrow 0$, $\mathcal{M} \leftarrow 1$, $r_0 \leftarrow 1$
for $r = N+1,\ N+2,\ \ldots$ **do**
  if $ES = 0$ **then**
    Calculate $\lambda_1(\mathcal{E})$ from eqn. (11)
    if $\lambda_1(\mathcal{E}) > (\beta_1 pN)^q$ **then**
      $ES \leftarrow 1$, $\mathcal{M} \leftarrow 1$, $r_e \leftarrow r$, $\quad \mathbf{y}^{(r)} \leftarrow \mathbf{x}^{(r)}$
    end if
    if $\mathcal{M} = 1$ and $ES = 0$ **then**
      $\mathbf{Z}^{(r)} \leftarrow [\mathbf{x}^{(r_0)}\ \ldots\ \mathbf{x}^{(r)}]$
      for $j = 1, 2, \ldots N_s$ **do**
        $e_j^{(r)} \leftarrow ||(\mathbf{I} - \mathbf{U_j}\mathbf{U_j}^T)\mathbf{Z}^{(r)}||_2$
        $\mathbb{P}(\mathbf{x}^{(r)}|m_j) \leftarrow exp\{-\frac{1}{2}C_f(e_j^{(r)})^2\}$
      end for
      for $i = 1, 2, \ldots N_s$ **do**
        $\mathbb{P}(m_i|\mathcal{Z}^{(r)}) \leftarrow \frac{\mathbb{P}(\mathbf{x}^{(r)}|m_i)\mathbb{P}(m_i|\mathcal{Z}^{(r-1)})}{\sum_{j=1}^{N_s}\mathbb{P}(\mathbf{x}^{(r)}|m_j)\mathbb{P}(m_j|\mathcal{Z}^{(r-1)})}$
      end for
      $L_w^{(r)} \leftarrow \sum_{j=1}^{N_s}\mathbb{P}(\mathbf{x}^{(r)}|m_j)\mathbb{P}(m_j|\mathcal{Z}^{(r)})$
      $\mathcal{T} \leftarrow \emptyset$
      for $j = 1, 2, \ldots N_s$ **do**
        if $\mathbb{P}(\mathbf{x}^{(r)}|m_j) > L_w^{(r)}$ **then**
          $\mathcal{T} \leftarrow \mathcal{T} \cup \{j\}$
        end if
      end for
      if $|\mathcal{T}| = 1$ **then**
        $\mathcal{M} \leftarrow 0$, $U \leftarrow U_j : j \in \mathcal{T}$
      else
        $\mathbf{y}^{(r)} \leftarrow \mathbf{x}^{(r)}$
      end if
    else if $\mathcal{M} = 0$ and $ES = 0$ **then**
      $\mathbf{Q} \leftarrow \mathbf{I} - \mathbf{U}\mathbf{U}^T$, $\boldsymbol{\gamma}^{(r)} \leftarrow \mathbf{Q}\mathbf{x}^{(r)}$, $\eta^{(r)} \leftarrow ||\mathbf{Q}y^{(r-1)}||_2$
      $\hat{\boldsymbol{\delta}}^{(r)} \leftarrow \underset{\boldsymbol{w}^{(r)}}{\arg\min}||\boldsymbol{w}^{(r)}||_1$ s.t. $||\boldsymbol{\gamma}^{(r)} - \mathbf{Q}\boldsymbol{w}^{(r)}||_2 \leq \eta^{(r)}$
      $\hat{\boldsymbol{\ell}}^{(r)} \leftarrow \mathbf{x}^{(r)} - \hat{\boldsymbol{\delta}}^{(r)}$, $\quad \mathbf{y}^{(r)} \leftarrow \hat{\boldsymbol{\ell}}^{(r)}$
    end if
  else if $ES = 1$ **then**
    $\mathbf{y}^{(r)} \leftarrow \mathbf{x}^{(r)}$
    Compute numerical rank of $[\mathbf{x}^{(r_e)} \ldots \mathbf{x}^{(r)}]$ with tol.
    if rank > 1 **then**
      $ES \leftarrow 0$, $r_0 \leftarrow r$, $\mathbb{P}(m_i|\mathcal{Z}^{(r_0)}) \leftarrow \frac{1}{N_s}, \forall 1 \leq i \leq N_s$
    end if
  end if
end for

The proposed anomaly correction approach combining the attributes of event outlier detection, subspace selection, and R-PCA-based signal recovery is summarized in Algorithm 1.

IV. Case Studies

Test system: The positive-sequence fundamental-frequency phasor model of the IEEE 16-machine, 5-area New England - New York system [31] is considered. Voltage magnitude signals from PMUs installed at 10 major inter-tie buses as highlighted in red, in Fig. 4 are used in the analysis. The reporting rate of the PMUs is assumed to be 60 Hz.

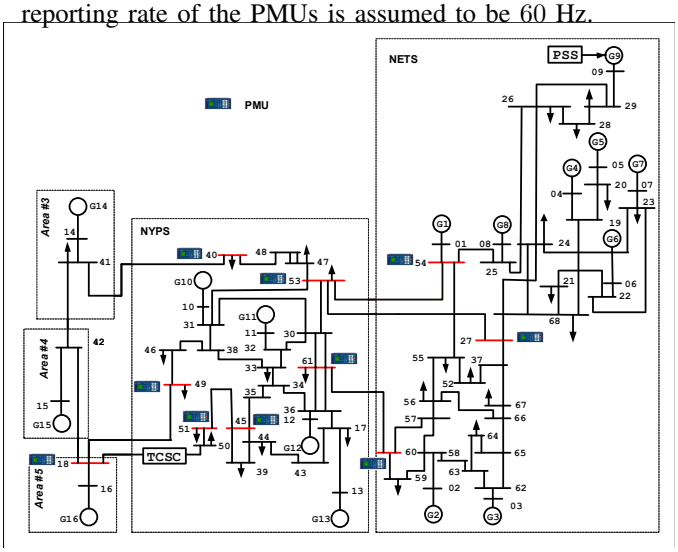

Fig. 4: Single-line diagram of 16-machine, 5−area New England - New York system with PMUs installed at major inter-tie buses highlighted in red.

Subspace library: The library shown in Table I contains subspace information corresponding to 5 different network configurations. SVD [28] was performed on a 40 s window following a self-clearing fault on each of these network configurations. For each subspace S_i, singular vectors corresponding to the three largest singular values are stored in the library.

TABLE III: SUBSPACE LIBRARY

S_1	S_2	S_3	S_4	S_5
$18-42$ out[1]	$18-49$ out[1]	$40-41$ out[1]	$53-54$ out[1]	Nominal Condition

[1]one of the double circuit lines in Fig. 4 is out

Corruption model: We consider following types of anomalies in PMU data−

(1) *Fault resembling attack:* We simulate bus fault at the desired location, and then record and replay that time series data including the fault outlier and the post-fault oscillations at multiple PMU channels to fool the operator into believing that it is an actual fault,

(2) *Noise injection attack:* We replace successive samples of measurement data from multiple locations using white Gaussian noise with a constant mean. In simulation, the mean value is set equal to the magnitude of the last correctly received data sample, and

(3) *Missing data attack:* We assume no data is received from multiple PMUs for consecutive instants. For blocks of missing samples, we use the last received sample and define corruption as the difference between the original data and the latest recorded sample. At any instant, let Ω be the index set of channels with missing data, and $\mathbf{I}_\Omega$ be the submatrix of the identity matrix $\mathbf{I}$ with columns indexed by Ω. Therefore, for a missing sample,

$$\mathbf{x}^{(r)} = (\mathbf{I} - \mathbf{I}_\Omega\ \mathbf{I}_\Omega^T)\ \boldsymbol{\ell}^{(r)} + \mathbf{I}_\Omega\ \mathbf{I}_\Omega^T\ \mathbf{x}^{(r-1)} = \boldsymbol{\ell}^{(r)} + \boldsymbol{\delta}^{(r)} \tag{17}$$

where, signal corruption, $\boldsymbol{\delta}^{(r)} = \mathbf{I}_\Omega\ \mathbf{I}_\Omega^T\ (\mathbf{x}^{(r-1)} - \boldsymbol{\ell}^{(r)})$.

Note that the case of signal recovery from random spurious outliers is the same as signal recovery from the fault outlier in the fault replay attack described above, and therefore is not studied separately. Also, at any instant we assume that data from only a fraction of PMUs is corrupted.

A. Case I: Signal recovery without event detection

First, we consider the standalone R-PCA algorithm working with the nominal subspace S_5 in absence of the proposed event detection method. A three-phase self-clearing fault is simulated at bus 18 in Fig. 4. The outliers in the voltage magnitudes induced by the fault do not belong to S_5, and when projected onto the orthogonal subspace (see, eqn. (15)) produce sufficiently large $\boldsymbol{\gamma}^{(r)}$ to get mistaken as data anomaly. As evident from Fig. 5, the standalone signal recovery algorithm is unable to trace the voltage sags in the selected signals.

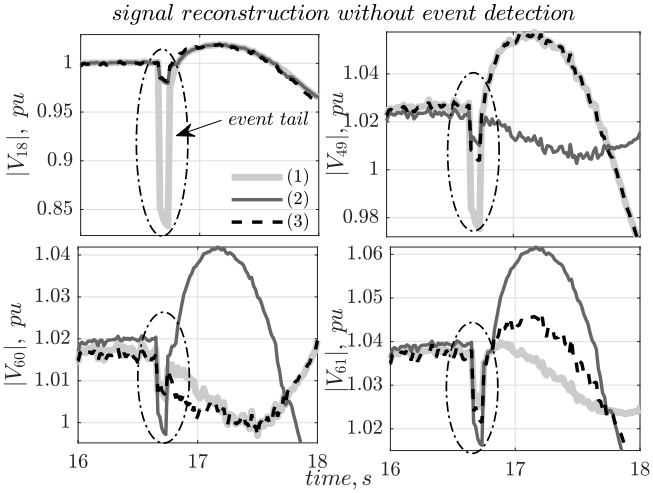

Fig. 5: Case I: Original voltage magnitudes during a fault − (1), and its reconstruction with a standalone R-PCA algorithm [10] using ST − (2), and using a fixed pre-fault subspace − (3).

We have also studied an approach where the subspace is derived online and updated dynamically from the reconstruction of a moving window of incoming data samples. We refer to this as the *subspace tracking (ST)* approach. Fig. 5 shows that ST produces similar error in reconstruction compared to the case with fixed subspace S_5. This is because in absence of an event detector, the window of data forming the subspace has measurements from both the fault and the pre-fault conditions.

B. Event detection and comparison with PCA-based approach

In this study, we demonstrate the ability of the proposed K-PCA-based metric (see, Section III-A) to detect event outliers and the onset of an event window using the voltage signals from 10 PMU locations shown in Fig. 4. We simulate two instances of three-phase fault near bus 18 and a fault resembling

attack as shown in Fig 6. The fault at $t = 5$ s near bus 18 is cleared by opening line 18-42, while fault at $t = 25$ s is self-clearing in nature. The recorded fault data from $t = 25$ s is replayed at $t = 45$ s at buses 18 and 42. The analysis was performed on a moving window of 25 samples for both PCA and K-PCA. In addition for K-PCA, q was set to 10. It can be seen from the Fig. 6 that the PCA-based T^2 event detection metric not only fails to differentiate between events (fault at 5 s and 25 s) and attack (at 45 s). However, when $\lambda(\mathcal{E})$ is used as a metric for detection, it clearly identifies the event and does not have a false triggering during the attack. Henceforth, we use this K-PCA-based metric to identify event outliers and suspend signal recovery in event window.

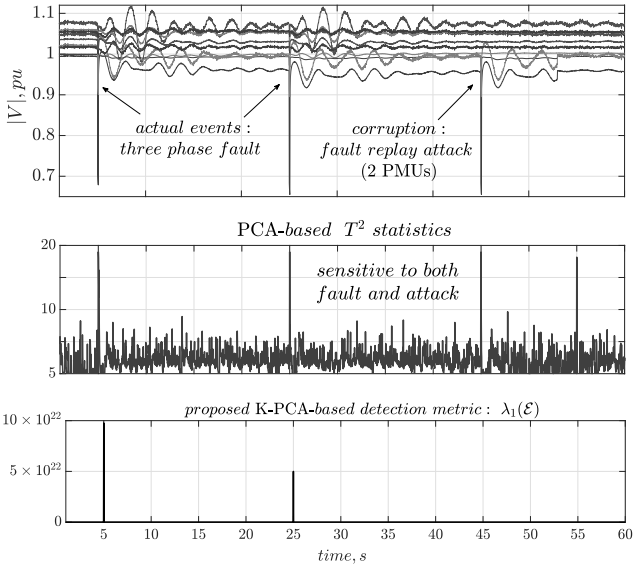

Fig. 6: Corruption-resilient detection of faults using the proposed K-PCA-based metric and comparison with PCA-based T^2 metric, using a window of 25 samples and with $q = 10$.

C. Case II: Signal recovery using the proposed approach with the exact subspace in library

Here, we study the effectiveness of the anomaly correction approach proposed in Section III, to recover clean signals from measurements containing both event-induced outliers and malicious corruption. A three-phase fault is simulated near bus 18 at $t = 16.67$ s and cleared at $t = 16.75$ s, by opening one of the double circuit lines connecting buses $18 - 42$. It is assumed that there is an intelligent attacker in the system with the capability of recording PMU outputs at selected locations and injecting that in the data stream at a later time. In this simulation, 400 samples of voltage magnitude data recorded following the fault at $t = 16.67$ s is replayed at bus locations 18 and 49 at $t = 21.67$ s, as shown in Figs. 7 (a)-(b). The challenge here is to identify the fault replay attack as a data anomaly and correct it while preserving the signature of the actual fault.

The K-PCA-based approach is applied on the incoming voltage data from the 10 PMU locations in Fig. 4, for detection of event-induced outliers. In each time-step K-PCA was performed on a moving window of 20 samples. The degree of the polynomial kernel q was taken as 10, and the detection

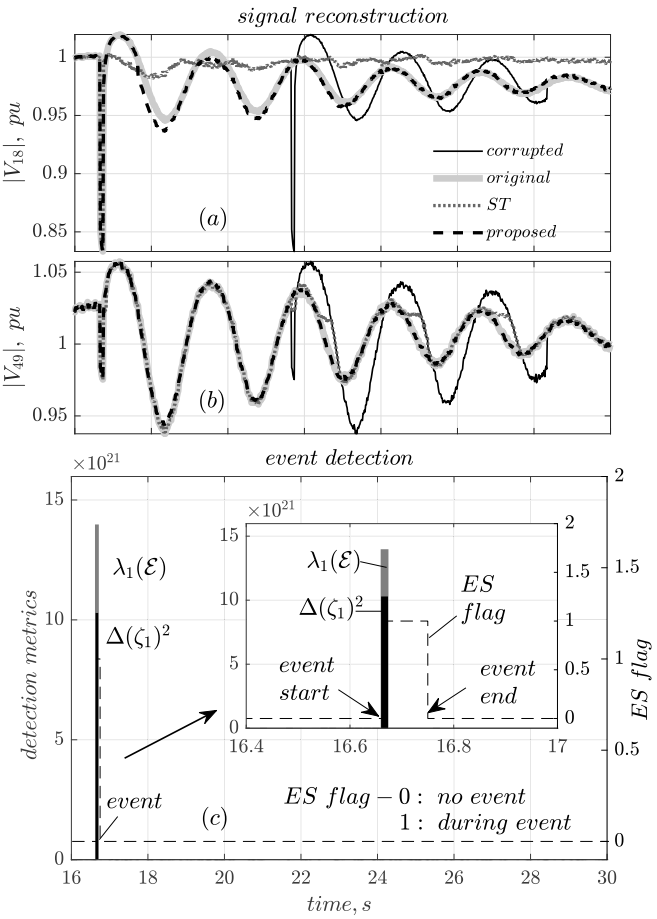

Fig. 7: (a)-(b): Case II: Fault resembling attack on signals $|V_{18}|$ and $|V_{49}|$. Signal recovery using the proposed approach is compared with ST approach. (c): K-PCA-based metrics for detection of event-induced outliers.

threshold was set at 10^{20} (assumed, in the worst case at any instant, not more than 50% of the channels are corrupted, therefore, $\beta_1 = 0.5 \implies$ threshold $= (0.5 \cdot 10 \cdot 20)^{10} = 10^{20}$, see Section III-A). The metrics $\Delta(\zeta_1)^2$ and $\lambda_1(\mathcal{E})$ as plotted in Fig. 7 (c), are able to detect and distinguish the onset of the fault from the attack. Further, the instant of fault clearing was identified at $t = 16.75$ s, by progressively computing the numerical rank of the event window at each time-step. The flag ES is 1 in this duration, and recovery is suspended.

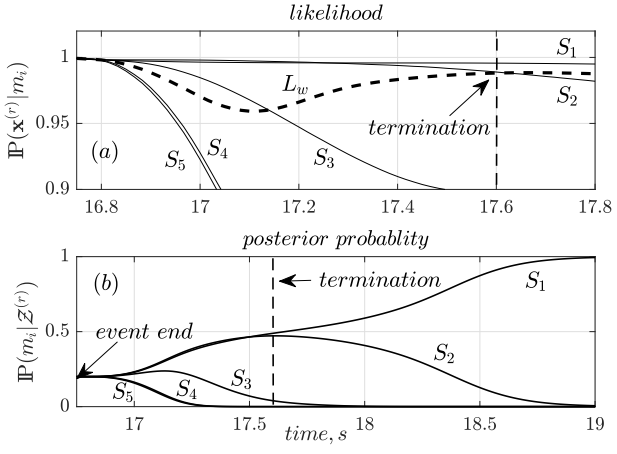

Fig. 8: Case II: (a) Likelihoods and (b) posterior probabilities of all 5 subspaces as obtained from recursive Bayesian estimation.

The Bayesian estimation for subspace selection is initiated

at the instant of fault clearing, with $C_f = 10$. The likelihoods at each iteration, and the corresponding posterior probabilities for all 5 subspaces from the library are plotted in Fig. 8. As seen from Fig. 8 (a), beyond $t = 17.6$ s there is only one subspace S_1 with likelihood above the weighted value L_w. And hence, following the arguments in Section III-B, termination is achieved at $t = 17.6$ s, and S_1 is selected.

The effectiveness of the termination criteria can be validated from the probability plot in Fig. 8 (b). Had the estimation not been terminated, the posterior probability for S_1 would have settled at 1 after finitely many iterations, with all others converging to 0. Choice of S_1 is also consistent with our knowledge that the fault in the simulation was cleared by opening a double circuit line connecting $18 - 42$.

The recovery of signals $|V_{18}|$ and $|V_{49}|$, using the proposed approach is shown in Figs. 7 (a)-(b).

Robustness of subspace selection to missing data: In the case above, it took 51 samples ($t = 16.75$ s $- 17.6$ s) to converge to the desired subspace S_1. Now, we study the impact of missing data on subspace selection and termination time. Two scenarios of data loss are considered in which a block of consecutive data samples are missed in signals $|V_{18}|$, $|V_{49}|$, and $|V_{60}| -$ (1) towards the beginning of subspace estimation $-$ after 5 samples from initiation, and (2) after accumulation of few samples $-$ after 15 samples from initiation. The termination time for these are presented in Table IV.

TABLE IV: IMPACT OF MISSING SAMPLES ON TERMINATION TIME

Missing data scenario	t_B (in samples)
missed 10 samples starting at 5^{th} sample ($t = 16.83$ s)	57
missed 10 samples starting at 15^{th} sample ($t = 17$ s)	52
missed 20 samples starting at 15^{th} sample ($t = 17$ s)	53

As demonstrated, the approach is fairly robust to missing samples. The robustness can be attributed to the following. First, in calculating the likelihoods in eqn. (13) the span of $\mathbf{x}^{(r)}$ is considered, which includes previously observed samples, and second, upon missing a sample we use the latest available data value.

D. Case III: Signal recovery using the proposed approach in absence of the exact subspace

1) *Fault resembling attack:* The study here is similar to Case II, except that S_1 is removed and the library now contains subspaces $S_2 - S_5$. The subspace selection process is illustrated in the likelihood and the probability plots of Fig. 9. The Bayesian estimation is terminated at $t = 17.2$ s, and S_2 is chosen for signal recovery. This can be justified from the argument that in absence of the exact post-fault subspace S_1, the selection approach chooses the subspace that is nearest to it. The angle θ_{ij}, as computed from (18), can be used as a measure of proximity between two subspaces S_i and S_j [32].

$$\theta_{ij} = sin^{-1}\Big\{ \Big\| (\mathbf{I} - \mathbf{U}_i\mathbf{U}_i^T)\mathbf{U}_j \Big\|_2 \Big/ \|\mathbf{U}_j\|_2 \Big\} \qquad (18)$$

TABLE V:
ANGLE BETWEEN S_1 AND OTHER SUBSPACES, in rads.

θ_{21}	θ_{31}	θ_{41}	θ_{51}
0.17	0.29	1.26	1.24

Using the measure in (18), the angle between S_1 and all other subspaces are computed, see Table V. It is evident that S_2 is closest to S_1, and thus best approximates the post-fault condition in absence of S_1. The recovery of signals $|V_{18}|$ and $|V_{49}|$ using S_2 as the reference subspace is shown in Fig. 10.

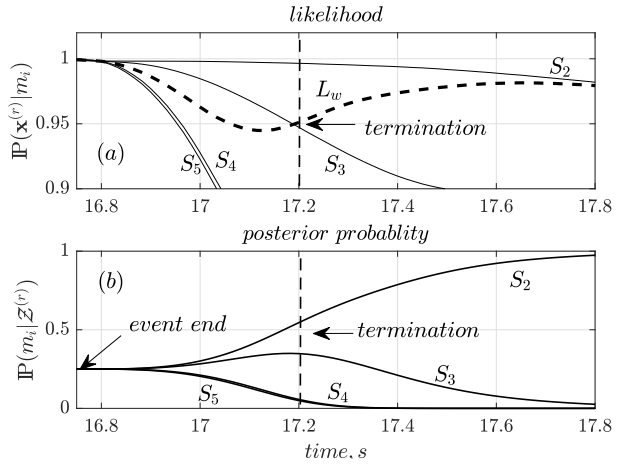

Fig. 9: Case III (1): (a) Likelihoods and (b) posterior probabilities of remaining 4 subspaces as obtained from recursive Bayesian estimation.

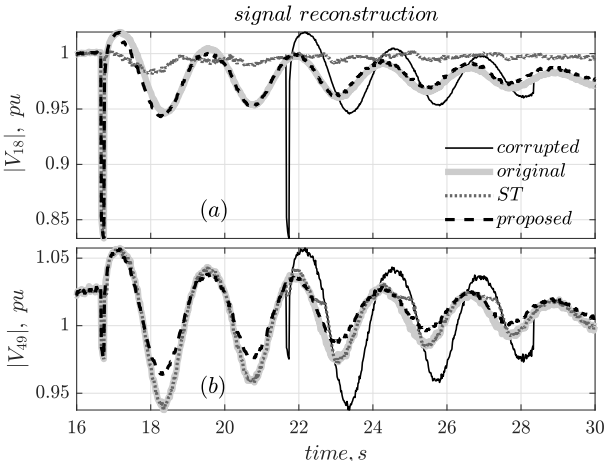

Fig. 10: Case III (1): (a)-(b): Fault resembling attack on signals $|V_{18}|$ and $|V_{49}|$. Signal recovery using the proposed approach in absence of exact subspace compared with ST approach.

TABLE VI: COMPARISON OF RECONSTRUCTION ERRORS

Fault Resembling Attack (20% Corruption: 18, 49)	**Average MSE**	**Average Std. Dev.**	**Maximum MSE**
Case I (3): *using pre-fault subspace*	$4.6e-3$	$7.8e-4$	1.14
Case I (2): *using ST approach*	$1.55e-4$	$2.28e-4$	$6.1e-3$
Case II: **Proposed** *using exact subspace* S_1	$2.39e-6$	$4.54e-6$	$2.38e-4$
Case III: **Proposed** *using subspace* S_2	$1.24e-5$	$2.18e-5$	$6.69e-4$

The statistical dispersion of the error in reconstruction (as defined in [10]) of all 10 signals, during $t = 16 - 30$ s for the

three cases discussed above, are presented in Table VI. The effectiveness of the proposed method is evident.

2) *Noise injection attack:* We have also studied the effectiveness of the proposed signal recovery approach under noise injection attack. Following the fault near bus 18 at $t = 16.67$ s, 400 samples of constant mean white Gaussian noise is injected in signals $|V_{45}|$ and $|V_{60}|$ at $t = 21.67$ s replacing the actual measurements. The signal-to-noise ratio (SNR) in the attack duration is 40 dB. In defining SNR, the constant mean represents the signal amplitude. It is assumed that the exact post-fault subspace S_1 is not available, and thus, the Bayesian framework selects S_2 for signal recovery. The corrupted and reconstructed signals are shown in Fig. 11.

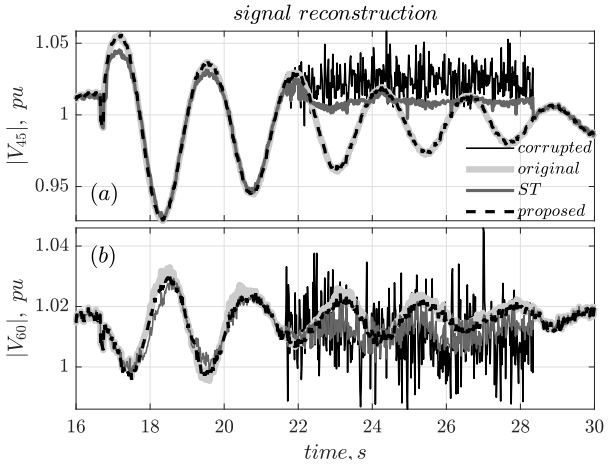

Fig. 11: Case III (2): (a)-(b): Noise injection attack on signals $|V_{45}|$ and $|V_{60}|$. Signal recovery using the proposed approach in absence of exact subspace compared with ST approach.

3) *Missing data attack:* We demonstrate the ability of the proposed approach in recovering measurements from missing data samples. In this study, following a fault near bus 18 at $t = 16.67$ s, we perform missing data attack on signals $|V_{18}|$, $|V_{49}|$, and $|V_{60}|$ for 400 consecutive samples. In absense of the exact post-fault subspace S_1, the Bayesian framework selects S_2 for signal recovery. The corrupted and reconstructed signals are shown in Fig. 12.

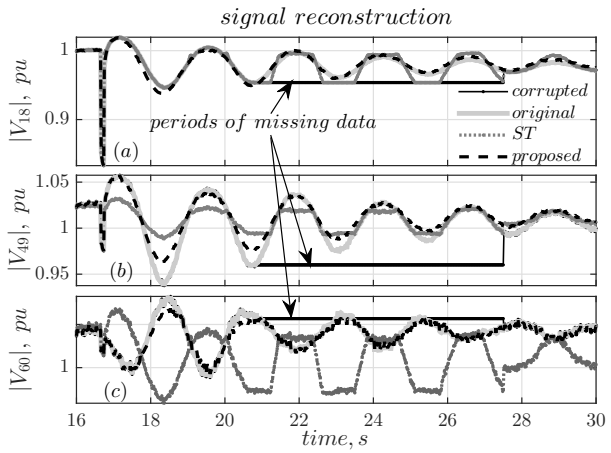

Fig. 12: Case III (3): (a)-(c): Missing data attack on signals $|V_{18}|$, $|V_{45}|$, and $|V_{60}|$. Signal recovery using the proposed approach in absence of exact subspace compared with ST approach.

4) *Robustness to corruption in multiple channels*: Finally, we demonstrate the robustness of the proposed approach by varying the fraction of corrupted channels. As described before, we assume that the exact subspace S_1 is not present in the library, and thus, signal reconstruction is performed with a close by subspace S_2. For each of two attack types − fault replay and missing data, we study 5 cases with number of corrupted signals at each instant ranging from 20% to 50%. The statistical dispersion of the errors in signal recovery for these cases are presented in Tables VII and VIII.

TABLE VII: RECONSTRUCTION ERRORS IN ABSENCE OF S_1 WITH VARYING PERCENTAGE OF CORRUPTION: FAULT REPLAY ATTACK

% Channels Corrupted	Signals (Bus No.s)	Average MSE	Average Std. Dev.	Maximum MSE
20%	18, 49	$1.24e-5$	$2.18e-5$	$6.69e-4$
30%	18, 27, 61	$1.25e-5$	$2.19e-5$	$6.69e-4$
30%	18, 27, 54	$1.27e-5$	$2.22e-5$	$6.69e-4$
40%	18, 27, 54, 60	$1.69e-5$	$2.55e-5$	$6.69e-4$
50%	18, 27, 54 60, 61	$1.75e-5$	$2.61e-5$	$6.69e-4$
50%	18, 27, 53 54, 60	$1.78e-5$	$2.64e-5$	$6.69e-4$

TABLE VIII: RECONSTRUCTION ERRORS IN ABSENCE OF S_1 WITH VARYING PERCENTAGE OF CORRUPTION: MISSING DATA ATTACK

% Channels Corrupted	Signals (Bus No.s)	Average MSE	Average Std. Dev.	Maximum MSE
20%	18, 40	$1.24e-5$	$2.17e-5$	$6.69e-4$
30%	18, 27, 61	$1.25e-5$	$2.18e-5$	$6.69e-4$
30%	18, 27, 54	$1.30e-5$	$2.22e-5$	$6.69e-4$
40%	18, 27, 54, 60	$1.78e-5$	$2.55e-5$	$6.69e-4$
50%	18, 27, 54 60, 61	$1.78e-5$	$2.59e-5$	$6.69e-4$
50%	18, 27, 53 54, 60	$1.78e-5$	$2.60e-5$	$6.69e-4$

V. Conclusions

A standalone R-PCA-based signal recovery algorithm might mistake an event-induced outlier as a data anomaly, thereby producing erroneous signal reconstruction. To solve this, a corruption-resilient K-PCA-based metric was proposed, which suspends the R-PCA algorithm upon event detection. A recursive Bayesian framework was proposed for autonomous selection of appropriate subspace from a library, to be used in signal reconstruction. A termination criterion for accelerating the subspace search was also discussed. The ideas of corruption-resilient event outlier detection and stochastic subspace selection following a change in topology augments the traditional R-PCA to robustify PMU data recovery.

References

[1] S. Brahma, R. Kavasseri, H. Cao, N. R. Chaudhuri, T. Alexopoulos, and Y. Cui, "Real-Time Identification of Dynamic Events in Power Systems Using PMU Data, and Potential Applications− Models, Promises, and Challenges," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 294–301, Feb. 2017.

[2] C. Huang, F. Li, D. Zhou, J. Guo, Z. Pan, Y. Liu, and Y. Liu, "Data Quality Issues for Synchrophasor Applications Part I: A Review," *Journal of Modern Power Systems and Clean Energy*, vol. 4, no. 3, pp. 342–352, Jul. 2016.

[3] R. Bobba, E. Heine, H. Khurana, and T. Yardley, "Exploring a tiered architecture for NASPInet," in *2010 Innovative Smart Grid Technologies (ISGT)*, Jan. 2010, pp. 1–8.
[4] A. Ashok, M. Govindarasu, and J. Wang, "Cyber-Physical Attack-Resilient Wide-Area Monitoring, Protection, and Control for the Power Grid," *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1389–1407, Jul. 2017.
[5] K. Chatterjee, V. Padmini, and S. A. Khaparde, "Review of Cyber Attacks on Power System Operations," in *2017 IEEE Region 10 Symposium*, Jul. 2017, pp. 1–6.
[6] H. M. Khalid and J. C. Peng, "A Bayesian Algorithm to Enhance the Resilience of WAMS Applications Against Cyber Attacks," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2026–2037, Jul. 2016.
[7] K. Manandhar, X. Cao, F. Hu, and Y. Liu, "Detection of Faults and Attacks Including False Data Injection Attack in Smart Grid Using Kalman Filter," *IEEE Trans. Control of Net. Syst.*, vol. 1, no. 4, pp. 370–379, Dec. 2014.
[8] J. Zhao, S. Wang, L. Mili, B. Amidan, R. Huang, and Z. Huang, "A Robust State Estimation Framework Considering Measurement Correlations and Imperfect Synchronization," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4604–4613, Jul. 2018.
[9] K. Mahapatra, N. R. Chaudhuri, R. G. Kavasseri, and S. M. Brahma, "Online Analytical Characterization of Outliers in Synchrophasor Measurements: A Singular Value Perturbation Viewpoint," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 3863–3874, Jul. 2018.
[10] K. Mahapatra and N. R. Chaudhuri, "Online Robust PCA for Malicious Attack-Resilience in Wide-Area Mode Metering Application," *IEEE Trans. Power Syst.*, pp. 1–1, 2019.
[11] M. Zhou, Y. Wang, A. K. Srivastava, Y. Wu, and P. Banerjee, "Ensemble based Algorithm for Synchrophasor Data Anomaly Detection," *IEEE Trans. Smart Grid*, pp. 1–1, 2018.
[12] P. Gao, M. Wang, J. H. Chow, M. Berger, and L. M. Seversky, "Missing Data Recovery for High-Dimensional Signals With Nonlinear Low-Dimensional Structures," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5421–5436, Oct. 2017.
[13] P. Gao, R. Wang, M. Wang, and J. H. Chow, "Low-Rank Matrix Recovery From Noisy, Quantized, and Erroneous Measurements," *IEEE Trans. Signal Process.*, vol. 66, pp. 2918–2932, Jun. 2018.
[14] P. Gao, M. Wang, S. G. Ghiocel, and J. H. Chow, "Modeless reconstruction of missing synchrophasor measurements," in *2014 IEEE PES General Meeting | Conference Exposition*, Jul. 2014, pp. 1–5.
[15] X. Wang, D. Shi, J. Wang, Z. Yu, and Z. Wang, "Online Identification and Data Recovery for PMU Data Manipulation Attack," *IEEE Trans. Smart Grid*, pp. 1–1, 2019.
[16] M. Liao, D. Shi, Z. Yu, Z. Yi, Z. Wang, and Y. Xiang, "An Alternating Direction Method of Multipliers Based Approach for PMU Data Recovery," *IEEE Trans. Smart Grid*, pp. 1–1, 2018.
[17] K. Mahapatra and N. R. Chaudhuri, "Malicious Corruption-Resilient Wide-Area Oscillation Monitoring using Principal Component Pursuit," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1813–1825, Mar. 2019.
[18] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[19] H. Guo, C. Qiu, and N. Vaswani, "An Online Algorithm for Separating Sparse and Low-Dimensional Signal Sequences From Their Sum," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4284–4297, Aug. 2014.
[20] P. Narayanamurthy, V. Daneshpajooh, and N. Vaswani, "Subspace Tracking from Missing and Outlier Corrupted Data," *arXiv:1810.03051 [cs.LG]*, Oct. 2018.
[21] E. Barocio, B. C. Pal, D. Fabozzi, and N. F. Thornhill, "Detection and Visualization of Power System Disturbances using Principal Component Analysis," in *2013 IREP Symposium Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid*, Aug. 2013, pp. 1–10.
[22] X. Liu, D. M. Laverty, R. J. Best, K. Li, D. J. Morrow, and S. McLoone, "Principal Component Analysis of Wide-Area Phasor Measurements for Islanding Detection- A Geometric View," *IEEE Trans. Power Del.*, vol. 30, no. 2, pp. 976–985, Apr. 2015.
[23] X. Liu, J. M. Kennedy, D. M. Laverty, D. J. Morrow, and S. McLoone, "Wide-Area Phase-Angle Measurements for Islanding Detection- An Adaptive Nonlinear Approach," *IEEE Trans. Power Del.*, vol. 31, no. 4, pp. 1901–1911, Aug. 2016.
[24] G. Liu, H. Chen, X. Sun, N. Quan, and L. Wan, "Low-Complexity Nonlinear Analysis of Synchrophasor Measurements for Events Detection and Localization," *IEEE Access*, vol. 6, pp. 4982–4993, 2018.
[25] K. Chatterjee and N. R. Chaudhuri, "Corruption-resilient Detection of Event-Induced Outliers in PMU Data: A Kernel PCA Approach," in *2019 IEEE PES General Meeting | Conference Exposition*, Aug. 2019 (Accepted), pp. 1–5.
[26] Y. Hao, M. Wang, J. H. Chow, E. Farantatos, and M. Patel, "Modelless Data Quality Improvement of Streaming Synchrophasor Measurements by Exploiting the Low-Rank Hankel Structure," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6966–6977, Nov. 2018.
[27] B. Scholkopf, A. Smola, and K. Muller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Max-Planck-Institut fur biologische Kybernetik, Technical Report*, no. 44, 1996.
[28] G. H. Golub and C. F. V. Loan, *Matrix Computations*. The John Hopkins University Press, Baltimore, MD, 1989.
[29] A. J. Huang, *Bayesian Estimation and Tracking: Practical Guide*. John Wiley & Sons, Inc., 2012.
[30] Y. Chen, F. Liu, S. Mei, G. He, and Q. Lu, "An Improved Recursive Bayesian Approach for Transformer Tap Position Estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2830–2841, Aug. 2013.
[31] N. R. Chaudhuri, "Wide-area Monitoring and Control of Future Smart Grids," PhD thesis, Imperial College, London, U.K., 2011. [Online]. Available: http://hdl.handle.net/10044/1/7026.
[32] A. Björck and G. H. Golub, "Numerical Methods for Computing Angles between Linear Subspaces," *Mathematics of Computation*, vol. 27, no. 123, pp. 579–594, 1973.

Kaustav Chatterjee (S'19) received his B.E. and M.Tech. degrees in Electrical Engineering from Jadavpur University, India and Indian Institute of Technology Bombay, India in 2015 and 2018 respectively. Currently, he is pursuing Ph.D. in Electrical Engineering at The Pennsylvania State University, USA. His research interests lie in the intersection of signal processing and system theory for real-time monitoring of large power systems and online estimation of system dynamics.

Kaveri Mahapatra (S'16) received the M. Tech. degree from Siksha 'O' Anusandhan University, India in 2013. She is currently pursuing her Ph.D. degree in the School of Electrical Engineering and Computer Science at the Pennsylvania State University, USA. Her current research interests include wide area monitoring, protection and control, cyber security, soft computing and optimization, and power system dynamics.

Nilanjan Ray Chaudhuri (S'08-M'09-SM'16) received his Ph.D. degree from Imperial College London, London, UK in 2011 in Power Systems. From 2005-2007, he worked in General Electric (GE) John F. Welch Technology Center. He came back to GE and worked in GE Global Research Center, NY, USA as a Lead Engineer during 2011-2014. Presently, he is an Assistant Professor with the School of Electrical Engineering and Computer Science at Penn State, University Park, PA. He was an Assistant Professor with North Dakota State University, Fargo, ND, USA during 2014-2016. He is a member of the IEEE and IEEE PES. Dr. Ray Chaudhuri is the lead author of the book Multi-terminal Direct Current Grids: Modeling, Analysis, and Control (Wiley/IEEE Press, 2014). He served as an Associate Editor of the IEEE TRANSACTIONS ON POWER DELIVERY (2013 – 2019) and IEEE PES LETTERS (2016 - present). Dr. Ray Chaudhuri was the recipient of the National Science Foundation Early Faculty CAREER Award in 2016 and Joel and Ruth Spira Excellence in Teaching Award in 2019.