UNIVERSITY OF CALIFORNIA
RIVERSIDE

Tackling the Challenges in Power Distribution System State Estimation With Low-Observability

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Alireza Akrami

September 2022

Dissertation Committee:

Dr. Hamed Mohsenian-Rad, Chairperson
Dr. Salman Asif
Dr. Konstantinos Karydis

The Dissertation of Alireza Akrami is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to thank my advisor, Prof. Hamed Mohsenian-Rad, for his help and support during my journey towards earning my PhD degree. His trust on me, provided me with an opportunity to learn and grow.

I would like to thank Dr. Salman Asif and Dr. Konstantinos Karydis for their support and being a part of my dissertation defense committee.

Finally, I wish to thank my beloved wife, Fatemeh, who has stood by me through all my travails during this journey. Her love, support, and encouragement are the true reasons of my success. Also, I would like to thank my parents and my dear brother for their endless support.

The content of this thesis is a reprint of the material that are appeared in the following publications:

1. **A. Akrami**, M. S. Asif, and H. Mohsenian-Rad, "Sparse Distribution System State Estimation: An Approximate Solution Against Low Observability," 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2020, pp. 1-5.

2. **A. Akrami**, M. S. Asif, and H. Mohsenian-Rad, "Sparse Tracking State Estimation for Low-Observable Power Distribution Systems Using D-PMUs," in IEEE Transactions on Power Systems, vol. 37, no. 1, pp. 551-564, Jan. 2022.

3. **A. Akrami** and H. Mohsenian-Rad, "Event-Triggered Distribution System State Estimation: Sparse Kalman Filtering with Reinforced Coupling," *under review* in IEEE Transactions on Smart Grid.

4. **A. Akrami** and H. Mohsenian-Rad, "Joint Sparse Estimation of Sensitivity Distribution Factors and Power Flows in Low-Observable Power Distribution Systems," *submitted to* 2023 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT).

To my beloved Fatemeh, Maman, and Baba.

ABSTRACT OF THE DISSERTATION

Tackling the Challenges in Power Distribution System State Estimation With Low-Observability

by

Alireza Akrami

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, September 2022
Dr. Hamed Mohsenian-Rad, Chairperson

State estimation is a fundamental task in power system monitoring. The focus in this thesis is on Distribution System State Estimation (DSSE). One of the main challenges in DSSE is the lack of observability due to the small number of sensor installations in practical power distribution circuits, where the number of measurements is far fewer than the number of state variables. In this thesis, our goal is to develop DSSE methods which address the low-observability challenges.

First, we leverage the high reporting rate of a small number of distribution-level phasor measurement units (D-PMUs), a.k.a., micro-PMUs, to unmask and characterize sparsity patterns among the state variables in radial power distribution systems. Accordingly, the DSSE problem is formulated over the differential synchrophasors as an adaptive group sparse recovery problem to track the changes that are made in the states of the system and captured by D-PMU measurements. To enhance the performance of the proposed method, the formulated DSSE is further augmented by the side information on the support of the vector of unknowns that is obtained from the outcome of an event-zone identification

analysis prior to solving the DSSE problem.

Second, to capture the dynamics of the power distribution system, we model the DSSE problem under an event-triggered setting, where we use the estimations of the state variables during the previous events as priori information to predict the state variables at the current event. Accordingly, a novel data-driven method based on elastic net regression is proposed to learn the event-triggered state transition matrix; despite the low-observability in the system. Here, in the absence of direct power measurements, we enhance our ability in sparse recovery by developing a new reinforced physics-based coupling method among the state variables, in which we add a novel set of linear differential power flow equations to the DSSE problem formulation in forms of virtual measurements.

Third, we study the joint estimation of sensitivity distribution factors and power flows in low-observable power distribution systems by developing a novel physics-aware measurement-based approach that takes into account the sparsity features of the problem extracted for radial power distribution systems.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Power systems consist of power generation units, power transmission lines, and power distribution feeders as shown in Fig. 1.1. Typically, the generated power by bulk power plants is stepped up by transformers to higher voltage level; and then it transmitted through the power transmission network to the consumers, i.e., loads that are located within the power distribution feeders, after the voltage level is stepped down.

Based on the operating points of the power system, its state can be classified into four categories, namely normal, alert, emergency, and restorative [108]. Reliable operation of power systems, requires the grid operator to make sure power system is continuously operated within the normal state; and if it is transited to alert or emergency state, proper controlling actions are taken. Knowledge on the current state of the system requires the grid operator to know about the values of operating points of the system are at every location.

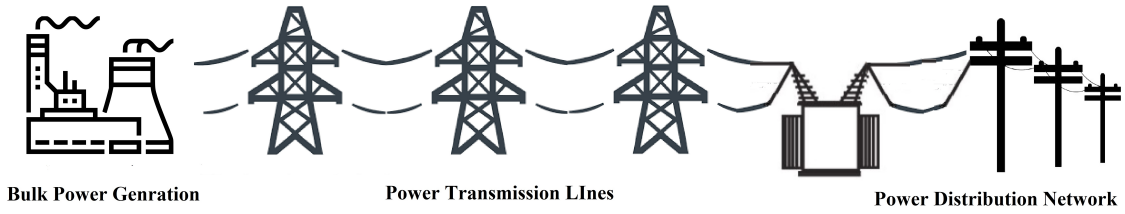

Figure 1.1: Schematic of a power system.

Power system state estimation (PSSE) is the process of obtaining state variables, typically nodal voltage phasors or line current phasors, from the available measurements in the power grid [3, 109]. PSSE can be classified into two categories: 1) transmission system state estimation (TSSE), and 2) distribution system state estimation (DSSE). TSSE is a very well-studied topic, as measurement redundancy is sufficient in power transmission networks to maintain full-observability and process bad data to solve the system of equations in the state estimation problem [83]. Formerly, DSSE did not receive many attentions, as traditional power distribution systems were passive and static, the load profiles changed very slowly over time, and there were no power generation units at the demand side. However, growing penetration of distributed energy resources (DER) with uncertain and variable generation at the power distribution side, as well as development of smart grid technologies has changed the nature of power distribution systems into an active type, where the need for continuous and real-time monitoring of the power distribution system is highlighted more than ever [75, 93, 93]. Accordingly, developing DSSE as the core module in monitoring of power distribution systems has turned into an interesting research topic.

1.2 Challenges in DSSE Problem

Due to the major differences that exists between the power transmission level and power distribution level, one cannot simply generalize a TSSE method to be used as a DSSE method [72]. These differences are as follows.

First, power distribution liens do have higher ratio of r/x, which means that DC power flow equations that may be used in a TSSE method are inaccurate in the DSSE method. Therefore, power flow equations should be in form of AC [90].

Secondly, power distribution systems are highly unbalanced [62], while TSSE methods are typically formulated for the balanced operation of power transmission networks. Unbalanced operation causes more complexity to the power flow equations in the state estimation problem, such that single phase equivalent of the system cannot be used anymore and the equations should be formulated based on the three-phase model [114, 119, 28].

Thirdly, high penetration of renewable energy resources introduces uncertainty to the DSSE problem. Additionally, there exists an uncertainty in the network model such as network topology and line parameters [97, 12, 44, 92]. TSSE methods cannot generally deal with these uncertainties and they need to be addressed.

Finally, power distribution systems suffer from lack of observability, which is due the lower number of installed sensor, expanded size of the distribution grid in comparison to the transmission network, and limited communication infrastructure [20, 4, 101].

Therefore, developing a DSSE method that is capable to address these challenges is of prominence. In this thesis, our focus is on the low-observability issue in the DSSE problem which is the most important and challenging aspect of the aforementioned issues.

1.3 State-of-the-Art in DSSE Methods

1.3.1 Conventional WLS-DSSE Methods

The state estimation problem can be formulated as the following system of equations:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}, \tag{1.1}$$

where $\mathbf{z}$ is the vector of measurements, $\mathbf{h}(.)$ is the measurement function, $\mathbf{x}$ is the vector of state variables, and $\mathbf{e}$ denotes the measurements error. The most common approach to formulate the state estimation problem is to write it in form of weighted least square (WLS) optimization as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} (\mathbf{z} - \mathbf{h}(\mathbf{x}))^{\top} \mathbf{W} (\mathbf{z} - \mathbf{h}(\mathbf{x})), \tag{1.2}$$

where $\hat{\mathbf{x}}$ is the estimated vector of state variables and $\mathbf{W}$ is the weight matrix which is associated with the measurement error. Based on the type of measurements and the selection of state variables, the measurement function $\mathbf{h}(.)$ can be either linear or non-linear.

The DSSE methods which are formulated as WLS optimization, mainly differ from each other in problem setting, such as type of measurements, type of state variables, and linearization of measurement function. Based on the choice of state variables, conventional DSSE methods can be classified into two categories: 1) nodal voltage based DSSE methods, and 2) line current based DSSE methods. State variables can be in rectangular coordinates or polar coordinates.

In general, presence of power measurements causes the measurement function to become non-linear. The common approach to deal with the non-linearity of the measure-

ment function is to use the Gauss-Newton method or Newton-Raphson method, take derivative of objective function with respect to the state variables, and iteratively updating the estimation until convergence [50]. Sometimes, the power injections are first converted to current injections, and then current injections are mapped to the voltage state variables in a linear form [73].

Deployment of direct voltage and current measurements in the DSSE problem can further help with making the measurement functions linear, without the need to linearize it. In [57] , a DSSE method based on the measurements of distribution level phasor measurement units (D-PMUs) is proposed which is computationally efficient, as it does not need to iteratively calculate the gain matrix for the DSSE problem. In [69], a DSSE method is developed which uses fast decouple power flow measurements to avoid high computation cost. However, it suffers from lack of robustness against bad data.

Line current based state estimation methods are more common in the power distribution feeders with radial topology. For current based approaches, power measurements can easily be represented in form of current provided that parameters of the distribution feeder are known. This alone removes the non-linearity of the measurement function [17]. Moreover, for unbalanced networks, the current based approaches can be written in form of phase decoupled equations.

In next section, we review the alternative DSSE approaches that are developed to address the drawbacks of conventional methods. We discuss the properties of these methods from various aspects.

1.3.2 Alternatives to WLS-DSSE Methods

While conventional WLS based methods are widely proposed in the literature to address the DSSE problem, there have been alternative approaches to address the challenges that WLS DSSE methods yet cannot tackle. In general, WLS DSSE is not robust against bad data which is a common issue in the power distribution monitoring system. In [80], a method based on least median of squares (LMS) is proposed which shows more robustness against the bad data. However, the computation cost for this method is higher than WLS and it needs measurement redundancy which is not usually the case in power distribution systems. In [49], a DSSE method based on least absolute value (LAV) is proposed which can identify and reject bad data due to the use of ℓ_1-norm minimization. An special type of DSSE approach which has the properties of both WLS and LAV estimator, is M-estimator, also known as Scheweppes-Hubber generalized M-estimator (SHGM) [34].

1.3.3 Centralized vs. Distributed DSSE Methods

Conventional DSSE methods are centralized, i.e., all measurements are sent into a central module to run and solve the DSSE problem. This requires large computations and well-established communication infrastructures. There are another class of DSSE methods which perform this task in a distributed fashion [51, 28]. These approaches are also known in the literature as multi-area DSSE methods [86, 117, 103]. In distributed approaches, the network is split into multiple areas, each area performs the estimation locally, and then, exchanges the estimations with other areas such that the estimation for the whole network is obtained. The exchange of estimations can be done in parallel [48] or in sequence [89].

1.3.4 Static vs. Dynamic DSSE Methods

The conventional DSSE approaches are mainly static, which means that they only consider a snapshot of the system at a time to estimate the state variables. Another class of estimators which takes into account the evolution of state variables over a period of time are dynamic DSSE approaches, which are also known as forecasting-aided state estimation (FASE) [61, 106, 74]. The advantge of FASE methods is that they can track the changes that frequently happen in the state of the system due to the addition of DERs and respondent loads in the system, while static DSSE approaches are not capable to do so.

Dynamic DSSE methods consist of two steps. In the first step, state variables for the current time instant are predicted from the estimations in previous time instances by means of a state-space model whose accuracy has a critical role in the performance of the DSSE method. In general, power distribution systems are assumed to operate under quasi-steady-state condition, which means the state-space model is in form of linear mapping with zero mean noise [95]. In the second step, the predictions are refined based on the measurements that are available in the current time instant by use of any types of a Kalman filte. A survey on different types of FASE methods is provided in [40].

1.4 Measurements and Distribution System Observability

As mentioned earlier, one of the main challenges in developing a DSSE method is the observability issue. In state estimation problems, observability means whether or not enough measurements are available such that the system of equations in the DSSE problem can be solved. Observability of the system is tied with the number, type, and location

of sensors. Observability analysis is usually performed in two forms of numerical analysis and topological analysis. In [23], an observability index is defined for power distribution systems which takes into account the uncertainties of the system. In [20], observability analysis is done based on graph theory. Different types of measurements might be used in the DSSE problem with respect to the problem setting and assumption to maintain the required observability for the system.

1.4.1 Real-Time Measurements

The main type of measurements in DSSE methods are the real-time measurements which can include nodal voltages, line currents, nodal injection powers, line power flows, and switch statuses that are collected by D-PMUs, supervisory control and data acquisition (SCADA), remote terminal units (RTUs), and smart meters. The reporting rate for different sensors can vary as high as once every 60 milliseconds for D-PMUs to as low as once every 15-30 minutes for smart meters. Therefore, the DSSE method should be capable to integrate heterogeneous types of measurements [58]. Another issue with integration of various types of measurements is unsynchronized measurements which needs to be addressed [60].

Developing a DSSE method for modern power distribution systems with high penetration of DERs, mainly needs to deploy those types measurements which can help it with capturing the dynamics of the system. Those measurements exclusively come from D-PMUs which can measure the magnitude and angle of voltage and current phasors up to 120 readings per second. D-PMUs use the global positioning system (GPS) as the time reference to time stamp all the measurements that are collected from different locations such that they can be used together without any concerns about lack of synchronization among

the measurements. D-PMus have been used in the literature for different applications such as state estimation [8, 9, 39], power system reconfiguration [11], and enhancing situational awareness [45, 118, 102, 82, 63]

1.4.2 Pseudo-Measurements

Due to the lack of measurement redundancy, other types of inputs in form of measurements are required for the DSSE problem, otherwise the undetermined system of equations cannot be solved through a conventional method. A common approach at the distribution level is to use an artificial stream of data that is called pseudo-measurements to maintain full-observability. Pseudo-measurements are not real measurement data, but are generated through statistical or machine learning approaches, so that their integration in the DSSE problem can make the system of DSSE equations overdetermined, and accordingly, solvable. Pseudo-measurements are typically nodal power injections. In statistical approaches, historical load data such as customer billing data and load profiles are used to generate data. Meanwhile, different factors such as demand response incentives, seasonal changes, and behind the meter generation can impact the accuracy of the pseudo-measurements which should be taken into account [98, 78]. There are also methods which consider the correlation between different variables in the system to improve the accuracy of pseudo-measurements [99, 85].

Machine learning approaches have also received attention recently to generate the pseudo-measurements [54, 47]. The advantage of machine learning based approaches is that accuracy of generated data might be higher provided that sufficient training data is available. Meanwhile, larger training data means that computation time would be higher.

1.5 Primary Research Question

While conventional DSSE methods in the literature address the observability issue in the DSSE problem mainly by making the network fully-observable, there are another class of the DSSE methods, which aim to solve the DSSE problem under the low-observability condition, and without making the network fully-observable. These emerging types of DSSE methods, leverage the underlying sparsity features that can be extracted in power distribution systems under different settings and use compressed sensing tools from signal theory to perform the state estimation task [36]. In continuations of this emerging approach, in this thesis our goal is to develop a novel class of DSSE approaches which can address the DSSE problem under low-observability condition, and without making the network fully-observable. Accordingly, in this thesis, we try to answer the following question:

"*How we can use the available limited and partially unsynchronized/missing measurements in a low-observable distribution feeder to develop a DSSE method such that we can properly monitor the distribution feeder in the real-time?*"

1.6 Summary of Contributions

The summary of contributions for each Chapter in this thesis are listed as follows:

1) Developing a novel sparse tracking DSSE method to leverage the measurements from only a small number of D-PMUs in a low-observable three-phase distribution feeder which is formulated as an adaptive group-sparse signal recovery problem, to estimate the differential voltage phasors and track the changes on the state variables of the system.

2) Establishing the grounds for sparsity in the DSSE problem in differential mode based on the engineering characteristics of power distribution systems as well as analyzing the

events that occur on the distribution feeders and are captured by D-PMUs. We introduce four types of state variables in the context of *group sparsity* with respect to their tendency towards being zero or non-zero.

3) Developing an event zone identification method that is used before running the sparse tracking DSSE. It provides the DSSE problem formulation with side information on the support of the sparse vector of the four types of state variables,

4) Deriving sufficient conditions for the *uniqueness* of the DSSE solution that is obtained from the sparse recovery problem with respect to the constructed side information,

5) Developing a mechanism to identify potential drifting, caused by accumulative estimation error of differential synchrophasors; thus to calibrate the estimated voltage phasors.

6) Developing an event-triggered DSSE method that uses the voltage and current phasor measurements from only a few D-PMUs. In this new method, the power distribution system is modeled as an event-triggered dynamic system; as opposed to a time-triggered dynamic system, which is commonly used in the literature. In this setting, state variables are defined in differential mode under low-observability conditions, where all the changes in the state variables are due to the occurrences of physical events such as load switching, capacitor bank switching, a sudden change in the generation level of renewable energy resources, etc.

7) A data-driven method is developed to learn the event-triggered state transition matrix. It takes into account the inherent sparsity due to the radial network topology and spatial-temporal correlations of the state variables in power distribution systems and works as a elastic net regression method. To reduce computational complexity, we propose a zonal extension of the basic nodal form for the event-triggered state transition matrix.

8) Two different approaches are proposed to solve the formulated sparse event-triggered DSSE problem. The first approach is exact and works based on the batch ADMM. The second method is approximate to further lower computational complexity. It works in three steps: Kalman filtering, incorporating sparsity, and backward smoothing.

9) Developing a novel joint sparse estimation method for the sensitivity distribution factors (SDF) matrices and power flows in a low-observable power distribution system. Unlike the existing methods in the literature, which are either model-based or measurement-based, the proposed method is a hybrid method, where a model-based approach is considered with respect to the physics-aware component of our design; while every other aspect of our approach is measurement-based.

10) Defining three scenarios based on the availability of measurements, and accordingly, developing an optimization model for each scenario, to obtain the joint solution. Also, developing a novel iterative solution to deal with the non-linearity of the optimization problem that is caused because of the missing measurements in the regression model.

Chapter 2

Sparse Tracking State Estimation for Low-Observable Power Distribution Systems Using D-PMUs

2.1 Introduction

2.1.1 Background and Motivation

Continuous monitoring and control of distribution systems has not been common across the whole network, but at few critical nodes such as substations, due to the passive slowly-changing nature of loads, as well as lack of instrumentation. The conventional monitoring usually has been performed through SCADA in terms of distribution system au-

tomation at a local level. However, penetration of distributed energy resources, emergence of new types of load, and development of demand side management programs have caused an evolution in the operation of power distribution grids [15, 10, 53, 5]. Distribution system state estimation (DSSE) is an important monitoring tool in power distribution system operation to help utilities with real-time grid monitoring capabilities in order to maintain reliable and secure operation of power distribution systems [37]. DSSE methods may use different types of available measurements with various time resolution in order to recover the state variables of the system, i.e., the nodal voltage phasors and branch current phasors [58].

The performance of DSSE is directly affected by the extent of observability in the power distribution system, which depends on the type, number, and location of sensors [3]. In practice, power distribution feeders often suffer from low-observability; because the number of sensors is much less than the number of state variables in many distribution feeders. The installation of smart meters and advanced metering infrastructure (AMI) have improved observability in power distribution networks. However, smart meters report measurements only once every 15 minutes to 1 hour [75]. These low reporting rates are not sufficient to capture the high dynamics of power distribution systems [112, 100].

The above issue can be resolved with the advent of the state-of-the-art D-PMUs, a.k.a., micro-PMUs[1], which report phasor measurements once every 8 to 100 milliseconds [82]. D-PMUs have been used in recent years at distribution feeders for various tasks, such as to achieve situational awareness [82] and to support optimal grid reconfiguration [11].

[1]Micro-PMU is a trademark of PSL [1]. Therefore, we use the term D-PMUs; since our study is not specific to any particular commercial technology.

D-PMUs can also be used in DSSE algorithms [39]. However, to resolve the low-observability issue to run a conventional DSSE, we need to install hundreds of D-PMUs on each feeder. This is cost prohibitive due to, not only the high cost of sensors, but also the cost of the communications infrastructure.

Therefore, finding a suitable solution to mitigate the low-observability in DSSE is an important practical challenge.

2.1.2 Related Works

The common approach to compensate for low-observability in DSSE is to use an artificial stream of data, so-called pseudo-measurements. Pseudo-measurements are often constructed by using historical load data or real-time AMI data [20]. The impact of seasonal changes in customers behavior may also be considered, e.g., see [38]. However, due to the uncertainty and the variability in distribution systems, pseudo-measurements are typically not accurate [35]. Lack of time synchronization is another factor that can negatively impact the accuracy of pseudo-measurements [14]. Inaccurate pseudo-measurements can create ill-conditioned mathematical optimization in the DSSE problem formulation; which may prevent it from converging to a reliable solution [35].

There have been efforts to make pseudo-measurements more robust against system uncertainties, through either statistical approaches or machine learning methods. In [98], a Gaussian mixture model is used for the load probability density function for inclusion in a conventional WLS-DSSE. In [47], a two-stage data clustering method is used to construct the pseudo-measurements. In [79], artificial neural networks are trained to generate pseudo-measurements from limited measurements. However, statistical methods still re-

quire reliable statistical models to be accurate. As for the machine learning methods, they require large sets of reliable training data to generate accurate pseudo-measurements.

The authors in [91] proposed a new data-driven DSSE method based on training a deep neural network model to solve the DSSE problem without adding pseudo-measurements; instead they added physical information of the underlying power distribution feeder, such as the parameters of the distribution lines to further increase the accuracy. While this method is very promising, as a data-driven method it naturally requires access a considerably large data set in order to train the machine learning model. Furthermore, as the size of network grows, it may become difficult to cover all the possible scenarios in the underlying physical system in the training process.

Numerically, observability is evaluated by checking the rank of DSSE gain matrix. If the gain matrix is full-column-rank and invertible, the network is said to be fully-observable. Otherwise, it means that the system of equations for DSSE problem is undetermined and more measurements are needed. According to compressed sensing theorem, for an under-determined system of equations, if the vector of unknowns is k-sparse, i.e. it has only k nonzero entries, then it can be recovered through sparse signal recovery [41].

Accordingly, a different approach that has emerged recently is to address low-observability through compressed sensing. In this approach, instead of trying to make the low-observable DSSE problem fully-observable through the above aforementioned methods, the goal is to rather extract and leverage the potential sparsity features in the DSSE problem so as to solve the original DSSE problem while it remains low-observable. In this regard, the authors in [13] used the correlation among nodal voltages to compress the measurements;

and then they used techniques from compressed sensing to solve the DSSE problem. In [77], the authors developed a current-based sparse DSSE method that is built on the assumption that the load currents are negligible in comparison with the injected current from the substation. This consideration results in making the problem sparse. In [43], a method based on matrix completion is developed to estimate the missing values in the DSSE problem, i.e., to estimate those state variables that do not have direct measurements. In [76], a block tensor completion is proposed to estimate the nodal voltages. In this method, the correlations among the state variables are used in a tensor norm minimization problem; which is a generalization of the matrix completion method. Finally, in [36], the authors used the correlation among the nodal voltages, and they accordingly presented a comparison among the performance of four different compressed sensing techniques, where the robustness of sparse recovery against bad data is also discussed.

A brief comparison between the proposed method versus the comparable methods that we discussed above is provided in Table 2.1.

2.2 Sparse Linear DSSE Problem Formulation in Differential Mode

Consider a multi-phase power distribution network that is represented by a graph $\mathcal{G} := (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ denotes the set of nodes and $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$ denotes the set of distribution lines. Let $\boldsymbol{\varphi} = \{A, B, C\}$ denote the set of phases. Suppose $\mathbf{V}^t$ denotes the vector of all

Table 2.1: Comparison among DSSE methods which use compressed sensing

Reference	Grounds for Sparsity	State Variables	Solution Approach	Tracking	Analysis on Optimality
[13]	Correlated structure of nodal voltages	Voltage	ℓ_1 norm minimization solved by Newton-Raphson	No	Yes
[77]	Neglecting load currents vs. the substation current	Current	Lasso minimization	No	No
[43]	Missing measurements	Voltage	Rank minimization	No	No
[36]	Temporal correlation of state variables	Voltage	Trace norm minimization	No	No
[23]	Correlated structure of nodal voltage	Voltages	Survey which includes several methods	No	No
Proposed Work	Grouped differential synchrophasors	Voltage + Current	Group Lasso minimization augmented by partial side information solved by ADMM	Yes	Yes

voltage phasors at time slot t, assuming that they are represented in the rectangular form:

$$\mathbf{V}^t := \left[\begin{array}{cc} \Re(v^t_{n,\phi}) & \Im(v^t_{n,\phi}) \end{array} \right]^\top, \quad \forall n \in \mathcal{N}, \quad \forall \phi \in \varphi, \tag{2.1}$$

where $\Re(.)$ denotes the real part of the phasor, $\Im(.)$ denotes the imaginary part of the phasor, and $v^t_{n,\phi}$ denotes the voltage phasor at bus n on phase ϕ at time slot t. The length of $\mathbf{V}^t$ is $|\mathcal{N}|$. In a similar setting, $\mathbf{I}^t$ denotes the vector of all the current phasors at time slot t:

$$\mathbf{I}^t := \left[\begin{array}{cc} \Re(i^t_{kl,\phi}) & \Im(i^t_{kl,\phi}) \end{array} \right]^\top, \quad \forall kl \in \mathcal{L}, \quad \forall \phi \in \varphi, \tag{2.2}$$

where $i^t_{kl,\phi}$ denotes the current phasor at line kl on phase ϕ at time slot t. The length of $\mathbf{I}^t$ is $|\mathcal{L}|$. Since we study radial distribution feeders, we have $|\mathcal{L}| = |\mathcal{N}| - 1$.

2.2.1 DSSE Problem Under Low-Observability Conditions

Let $\mathbf{x}^t$ be a $N \times 1$ vector which contains all state variables of the power distribution system at time interval t. Note that any independent combination of nodal voltage phasors and line current phasors that can identify the state of the system can be selected as the vector of state variables. Importantly, Throughout this work, we assume that the state variables include the vector of voltage phasors for all nodes, i.e., $\mathbf{V}^t$, and the vector of current phasors for all lines, i.e., $\mathbf{I}^t$. That is,

$$\mathbf{x}^t := \left[\begin{array}{cc} (\mathbf{V}^t)^\top & (\mathbf{I}^t)^\top \end{array} \right]^\top, \tag{2.3}$$

where we have $N = |\mathcal{N}| + |\mathcal{L}|$.

Also, let $\mathbf{y}^t$ denote the vector of all available measurements. At each time slot t, we assume that the measurements come from only D-PMUs. Since our focus is on a

challenging scenario where the distribution network is not fully-observable, we assume that D-PMUs are installed at only a few nodes. For example, one D-PMU can be installed at the substation; and one or at most two D-PMUs can be installed on each lateral, as we will see in the case studies in Section 2.4. D-PMUs measure voltage phasors on three-phases at nodes $\mathcal{N}_m \subseteq \mathcal{N}$ and current phasors on three-phases at lines $\mathcal{L}_m \subseteq \mathcal{L}$.

D-PMUs provide two types of measurements, voltage phasors and current phasors. Accordingly, there are two types of equations that we need to include in the state estimation problem. First, there are equations that map the state variables to the voltage phasor measurements through identity mapping for each node $n \in \mathcal{N}_m$:

$$\Re(v_{n,\phi}^t) = \frac{1}{2} e_{n,\phi}^\top (v_n^t + \overline{v_n^t}), \tag{2.4}$$

$$\Im(v_{n,\phi}^t) = \frac{1}{2j} e_{n,\phi}^\top (v_n^t - \overline{v_n^t}), \tag{2.5}$$

where $e_{n,\phi}$ is the ϕ-th canonical basis vector and v_n^t is the vector of voltage phasors at all phases at bus n at time slot t in complex form. Second, there are equations which map the state variables to the current measurements for each line segment $kl \in \mathcal{L}_m$ through the KCL equations:

$$\Re(i_{kl,\phi}^t) = \frac{1}{2} e_{kl,\phi}^\top (Y_{kl}(v_k^t - v_l^t) + \overline{Y}_{kl}(\overline{v_k^t} - \overline{v_l^t})), \tag{2.6}$$

$$\Im(i_{kl,\phi}^t) = \frac{1}{2j} e_{kl,\phi}^\top (Y_{kl}(v_k^t - v_l^t) - \overline{Y}_{kl}(\overline{v_k^t} - \overline{v_l^t})), \tag{2.7}$$

where Y_{kl} is the admittance of line segment kl.

For the rest of this work, we represent the equations in (2.4)-(2.7) in a compact form through measurement matrix $\mathbf{\Psi}$. Thus, we have:

$$\mathbf{y}^t = \mathbf{\Psi} \mathbf{x}^t. \tag{2.8}$$

Matrix $\mathbf{\Psi}$ represents $|\mathcal{N}_m| + |\mathcal{L}_m|$ equations. Therefore, its size is $(|\mathcal{N}_m| + |\mathcal{L}_m|) \times N$. The DSSE problem in its standard form is the problem of solving the system of equations in (2.8) to obtain the state variables $\mathbf{x}^t$ from the available measurements $\mathbf{y}^t$.

When we have measurement redundancy, it means that we have many more measurements than necessary to solve the equations in (2.8). In this work, our focus is on a challenging scenario where the distribution network is not fully-observable, i.e., we do not have enough measurements to solve the system of linear equations in (2.8).

Without loss of generality, we assume that $\text{rank}(\mathbf{\Psi}) = m_1$; that is, the measurement matrix is full row-rank. This means that the D-PMU measurements associated with each row of matrix $\mathbf{\Psi}$ are linearly independent.

Under the low-observability circumstances, we do not have enough measurements to solve the system of linear equations in (2.8). Thus, one can use the circuit laws to introduce additional equations, based on the so-called virtual measurements, to capture the relationships among the state variables at the line segments that are not equipped with sensors, i.e., $\forall kl \notin \mathcal{L}_m$:

$$0 = \frac{1}{2} e_{kl,\phi}^{\top} (Y_{kl}(v_k^t - v_l^t) + \overline{Y}_{kl}(\overline{v_k^t} - \overline{v_l^t})) - \Re(i_{kl,\phi}^t), \tag{2.9}$$

$$0 = \frac{1}{2j} e_{kl,\phi}^{\top} (Y_{kl}(v_k^t - v_l^t) - \overline{Y}_{kl}(\overline{v_k^t} - \overline{v_l^t})) - \Im(i_{kl,\phi}^t). \tag{2.10}$$

Again, let us represent the equations in (2.9) and (2.10) in a compact form as:

$$\mathbf{0} = \mathbf{\Phi} \mathbf{x}^t. \tag{2.11}$$

Matrix $\mathbf{\Phi}$ represents $|\mathcal{L}| - |\mathcal{L}_m|$ equations. Therefore, its size is $(|\mathcal{L}| - |\mathcal{L}_m|) \times N$. While the system of equations in (2.8) maps the unknown phasors to the known phasors, the system of equations in (2.11) relates the unknown phasors to other unknown phasors.

In a standard DSSE problem, if a bus or a line segment is already equipped with a sensor, then we do not need to consider the corresponding virtual measurement anymore. Because it does not provide any new independent equation. However, in a low-observable network, virtual measurements can sometimes provide new independent equations.

Now, let us concatenate the system of equations in (2.8) and (2.11) to obtain:

$$\mathbf{z}^t = \mathbf{H}\mathbf{x}^t \tag{2.12}$$

where

$$\mathbf{H} := \begin{bmatrix} \boldsymbol{\Psi} \ \boldsymbol{\Phi} \end{bmatrix}^\top, \quad \mathbf{z}^t := \begin{bmatrix} \mathbf{y}^t \ \mathbf{0} \end{bmatrix}^\top. \tag{2.13}$$

Matrix $\mathbf{H}$ is $M \times N$, where M denote the rank of matrix $\mathbf{H}$, i.e., the number of independent rows in stacking of matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$. For the rest of this work, we assume that matrix $\mathbf{H}$ is full-row-rank, i.e., it includes only the independent rows in $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$. The DSSE problem under the low-observability condition is the problem of solving (2.12) when

$$M < N. \tag{2.14}$$

In that case, the system of linear equations in (2.12) is undetermined and its solution is unspecified; i.e., many choices of $\mathbf{x}^t$ can lead to the same $\mathbf{z}^t$. The set of solutions are

$$\mathbf{H}^\top(\mathbf{H}\mathbf{H}^\top)^{-1}\mathbf{z}^t + \text{null}(\mathbf{H}), \tag{2.15}$$

where null$(\cdot)$ denotes the null space of matrix $\mathbf{H}$. Therefore, the DSSE problem cannot be solved in its standard form.

2.2.2 DSSE Problem in Differential Mode

Let us consider the same overall setting as in Section 2.2.1, but let us represent all the D-PMU measurements and all the state variables in a differential mode as

$$\Delta \mathbf{z}^t := \mathbf{z}^t - \mathbf{z}^{t-1}, \tag{2.16}$$

$$\Delta \mathbf{x}^t := \mathbf{x}^t - \mathbf{x}^{t-1}. \tag{2.17}$$

Here, we represent the measurements and state variables in time slot t in terms of their difference compared to their values in time slot $t-1$. We can write the relationship in (2.12) in differential mode as

$$\Delta \mathbf{z}^t = \mathbf{H} \Delta \mathbf{x}^t. \tag{2.18}$$

If (2.14) holds, then the system is still not fully-observable even in differential mode. Thus, in principle, we still cannot find a unique solution for the equations in (2.18); just like in Section 2.2.1, where we could not find a unique solution for the equations in (2.12). However, unlike in (2.12), we can solve the differential form of the DSSE problem in (2.18) by posing it as a sparse signal recovery problem [41]. The key is to show that the vector of state variables in the DSSE problem in differential mode, i.e., $\Delta \mathbf{x}^t$, is a sparse vector. Next, we explain the grounds for sparsity.

2.2.3 Grounds for Sparsity

In this section, we examine the engineering characteristics in power distribution systems and make the case that the DSSE problem in differential mode can be formulated as a sparse signal recovery problem. The fundamental observation throughout this section

is as follows; given the short time interval in D-PMU measurements, the changes in the states of the system from time $t-1$ to time t are caused by only one major change in the components of the power distribution system. We refer to these major changes as events, such as load switching, a sudden change in a DER output, etc. [45].

When an event takes place on a distribution feeder, it may impact a subset (or all) of nodal voltages and branch currents. The impact of an event on the network can be studied by considering compensation theorem from circuit theory [67]; based on which we can replace the circuit element that is the source cause of an event with a current source that injects the same level of current at the location of the event. The advantage of using compensation theorem is that it provides us with a direct relationship and also an equivalent circuit model to study how an event may change the state variables.

1) Partitioned Network Representation at an Event

Before we explain the specific grounds for sparsity, let us first define a few new notations. Consider the power distribution network in Fig. 2.1(a). Suppose a major load switching event occurs at bus 29 during the short time-period between time $t-1$ and time t. Consider the path between the substation and the event bus; as marked in red. Let us represent this path with $\mathcal{T} := (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of buses and $\mathcal{E}$ is the set of line segments on this path.

Next, let us partition the rest of the distribution system into four subgraphs as $(\mathcal{V}_i, \mathcal{E}_i),\ \forall i \in \{1, \cdots, 4\}$. These partitions are marked in Fig. 2.1(a). Note that, together

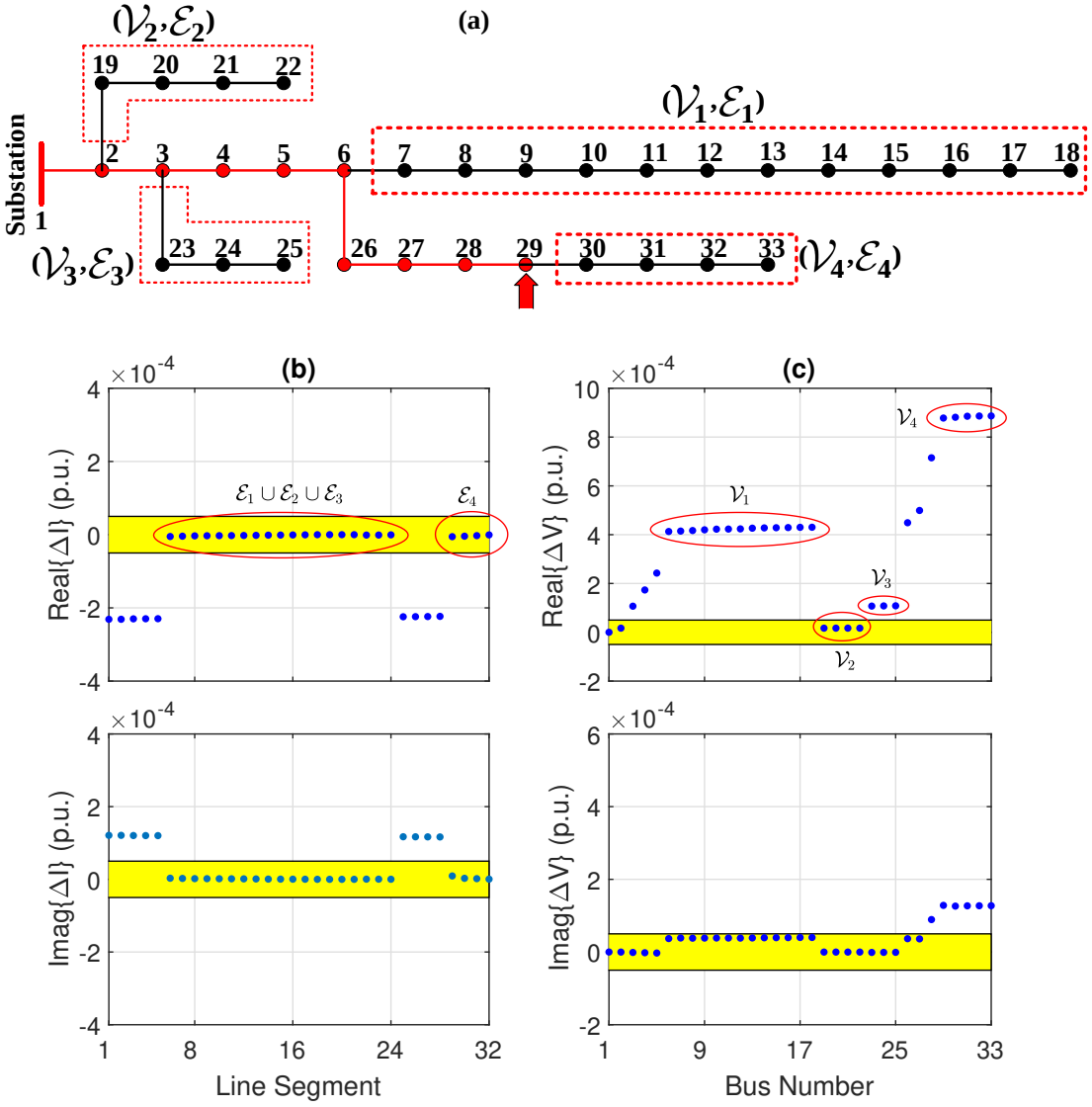

Figure 2.1: An example to illustrate sparsity: (a) An instance of the IEEE 33-node test system right after a major load change at node 29; (b) The changes in line currents, i.e., ΔI; (c) The changes in bus voltages, i.e., ΔV. The values within the yellow ribbon are considered approximately zero.

with path $(\mathcal{V}, \mathcal{E})$, these partitions cover the entire distribution system. That is, we have:

$$\bigcup_{i=1}^{4} \mathcal{V}_i = \mathcal{N} \backslash \mathcal{V}, \qquad \bigcup_{i=1}^{4} \mathcal{E}_i = \mathcal{L} \backslash \mathcal{E}. \tag{2.19}$$

While the above notations are defined based on the example in Fig. 2.1(a), the basic idea can be similarly applied to any power distribution system with any given bus for the event.

In our analysis, it is critical to figure out under what circumstances the changes in voltage or current are non-zero or (approximately) zero, i.e., they lay outside or inside the zero approximation region; as marked with the yellow ribbon in Figs. 2.1(b) and (c). Note that, a key feature of sparse signal recovery is its capability to manage the width of the zero approximation region as a trade-off against estimation error.

2) Impact of an Event on Line Current

Following the compensation theorem [67], suppose we replace the event on node 29 in Fig. 2.1(a) by an equivalent current source. If we write the Kirchhoff's current law at all nodes in $\mathcal{V}$, we can show that the change in the current for all the line segments in set $\mathcal{E}$ is equal to the amount of current that is injected by a current source, in the equivalent circuit of the power distribution system that is constructed by applying the compensation theorem; because the change in the injected current to all the nodes is zero; except for the node where the event occurs. Moreover, if we write the same system of equations for all the nodes in $\mathcal{N} \backslash \mathcal{V}$, we can show that the amount of change in the current for all the line segments in set $\mathcal{L} \backslash \mathcal{E}$ is zero; otherwise it means that there is a non-zero change in the injected current to one of the nodes outside the red path, i.e., there is another event somewhere else on the distribution feeder which is in contradiction to our initial assumption.

In summary, once an event occurs somewhere on a power distribution feeder, the line segments on the red path may experience non-zero changes; while the line segments outside the red path do not experience any change in their current. This alone creates a major sparsity in vector $\Delta \mathbf{x}^t$.

3) Impact of an Event on Nodal Voltage

When an event takes place, the nodal voltages on the nodes along the red path are directly affected. This is because of the changes in the current of the line segments that are along the red path. However, the voltages on the nodes outside the red path may or may not change drastically depending on several factors, such as line parameters, network topology, and most importantly, how the event changes voltage across the nodes that are located on the boundaries of the red path and other partitions. To elaborate, suppose s is a node on the red path that has a neighboring node r outside the red path. For example s could be bus 2, 3, 6, or 29 in Fig. 2.1 (a). By writing down the Eq. (2.6) for line segment sr that lays outside the red path, one can show that the following equality holds:

$$\frac{R_{sr}}{X_{sr}} = \frac{\Re(\Delta v_s^t) - \Re(\Delta v_r^t)}{\Im(\Delta v_s^t) - \Im(\Delta v_r^t)}, \tag{2.20}$$

where R_{sr} and X_{sr} are the resistance and reactance of line sr, respectively. From (2.20), whether or not the differential voltage phasors for the nodes outside the red path are zero, depends on the value of other parameters in (2.20). In either case, the same zero/non-zero condition would hold for any other node in the same partition.

As a result, there exists a group sparsity among the nodal voltages in differential mode for the buses that are in the same partition $\mathcal{V}_i,\ \forall i \in \{1, \cdots, 4\}$ outside the red path.

For example, in Fig. 2.1(c), the parameters in (2.20) are such that the differential voltage phasors for the nodes in partition $\mathcal{V}_1$ are all non-zero and the differential voltage phasors for the nodes in partition $\mathcal{V}_2$ are all approximately zero.

4) Fundamental Conclusions

We can combine the analysis to conclude the following corollary; which builds a foundation for our analysis:

Corollary 1: At each time slot t, we can divide the state variables in differential mode into four types:

1. **Differential current phasors known to be zero**: For all the line segments in set $\mathcal{L}\backslash\mathcal{E}$, ΔI would be zero.

2. **Differential current phasors known to be non-zero**: For all the line segments in set $\mathcal{E}$, ΔI would be non-zero.

3. **Grouped differential voltage phasors**: For all buses in each of the sets $\mathcal{V}_i$, $\forall i \in \{1, \cdots, 4\}$, either ΔV would be (approximately) zero or non-zero for all such buses.

4. **Differential voltage phasors known to be non-zero**: For all buses in set $\mathcal{V}$, ΔV would be non-zero.

The proof of Corollary 1 is as follows:

Consider the network in Fig. 2.2, which is the equivalent circuit of a power distribution feeder at the time of event that is obtained by applying the Compensation Theorem [67]. The event at bus m is replaced by a current source which injects the same level of

current due to the event to event bus. By applying the Kirchhoff's Current Law (KCL) at the event bus, i.e., the last bus on the red path, we have:

$$\Delta i_m - \Delta i_{lm} + \sum_{j=1}^{N} \Delta i_{ml'_j} = 0. \tag{2.21}$$

We can write a similar equation at node l as

$$\Delta i_l + \Delta i_{nl} - \Delta i_{lm} = 0. \tag{2.22}$$

Based on the Compensation Theorem, since no event occurs at bus l, there is no change in the injected current to this bus, i.e., we have: $\Delta i_l = 0$. This is also true for all other buses in set $\mathcal{V}$. Thus, $\Delta i_{nl} = \Delta i_{lm}$. If we continue writing the same type of equations for the buses on the red path until we reach the substation, we can conclude that the change in the line current for all the line segments in set $\mathcal{E}$ is equal to the change in the injected current that is drawn from the substation:

$$\Delta i_{sub} = \Delta i_{lm} \quad \forall lm \in \mathcal{E}. \tag{2.23}$$

Next, let us write the KCL at all nodes in set $\mathcal{V}_1$. For the same reason, the change in current for all the line segments in set $\mathcal{E}_1$ is equal to $\Delta i_{ml'_1}$. Consider the whole set of $(\mathcal{V}_1, \mathcal{E}_1)$ as a super node. If the value of injected current to this supper node, i.e., $\Delta i_{ml'_1}$ is non-zero, then it means that this current is being injected to one of the inner nodes of the super node, which means that there is an event somewhere inside the super node that would be against our initial assumption; where only one event happens across the feeder during each time slot t. Therefore, $\Delta i_{ml'_1} = 0$. Accordingly, we have:

$$\sum_{j=1}^{N} \Delta i_{ml'_j} = 0. \tag{2.24}$$

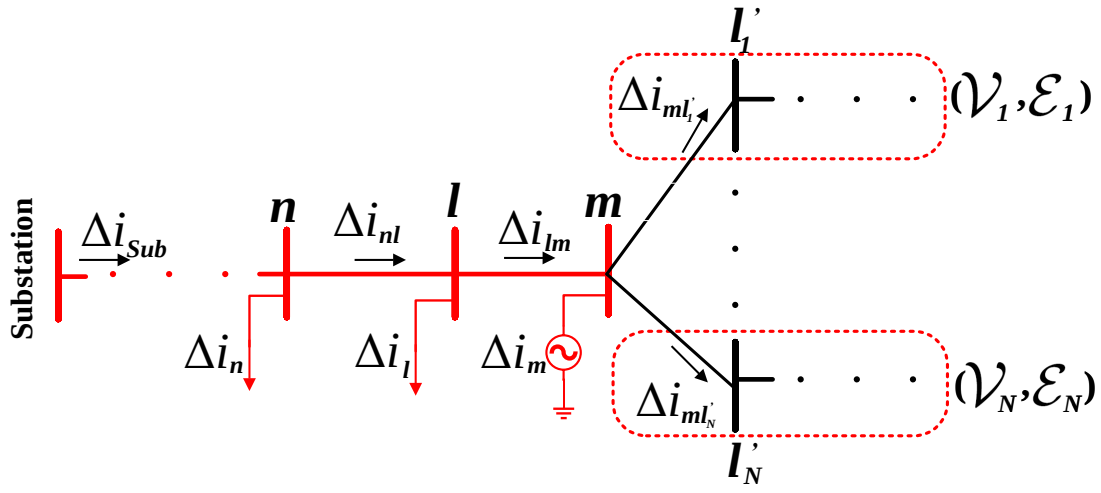

Figure 2.2: Equivalent circuit of distribution feeder when an event happens at node m.

By replacing (2.24) in (2.21), we obtain:

$$\Delta i_m = \Delta i_{sub}. \tag{2.25}$$

Therefore, when an even happens, the change in current for all the line segments in set $\mathcal{E}$ is non-zero, and equal to the value of injected current from the substation node, or the value of the equivalent current source. Also, the change in current for all the line segments in set $\mathcal{L}\backslash\mathcal{E}$ is zero. This is exactly in accordance to the illustrative example that was shown in Fig. 2.1(b). Next, let us write the Ohm's law for line segment ml_1' as in (2.6). Since $\Delta i_{ml_1'} = 0$, we can obtain

$$\frac{R_{ml_1'}}{X_{ml_1'}} = \frac{\Re(\Delta v_m) - \Re(\Delta v_{l_1'})}{\Im(\Delta v_m) - \Im(\Delta v_{l_1'})}. \tag{2.26}$$

Thus, based on the values of the parameters, there are two scenarios to consider. First, if the impedance of line ml_1' and the differential voltage at node m are such that

$$\frac{R_{ml_1'}}{X_{ml_1'}} \approx \frac{\Re(\Delta v_m)}{\Im(\Delta v_m)}; \tag{2.27}$$

then, we can (approximately) replace $\Delta v_{l'_1}$ by zero. This is the case that holds for the nodes in set $\mathcal{V}_2$ in the illustrative example shown in Fig. 2.1(c). Second, if the corresponding parameters are not such that (2.27) holds, then $\Delta v_{l'_1} \neq 0$.

The discussion on the grounds for sparsity in this work are made based on the assumption that only one major event happens across the distribution feeder during each time slot of the proposed DSSE process. This is a reasonable assumption in practice; as it is observed in the previous studies that have looked at real-world D-PMU data. However, for the occasional scenarios, where two or more events occur simultaneously during the same 100 msec time slot, one may need to modify the grounds for sparsity that we discussed in Section 2.2.3. Addressing such cases could be the subject of an extension of this work in the future.

The grounds for sparsity that we discussed above are developed based on the assumption that the distribution feeder has a radial topology; which is very common in practice. The above discussions are not applicable to transmission systems, where the network has a meshed topology. Therefore, the analysis in this work is specific to the DSSE; because it takes advantage of the features of power distribution systems.

Note that, in principle, we do not know the actual type for any of the above state variables in advance; unless the location (i.e., the bus number) of the event is known to us.

2.2.4 Event Zone Identification

Based on the analysis in Section 2.2.3, the key to unlock the sparsity characteristics in the DSSE problem in differential mode is to focus on the major events that occur across the low-observable power distribution system.

When an event(s) happens whose type and location is unknown to us, all our information on the unknown vector would be limited to the fact that $\Delta \mathbf{x}^t$ might be a sufficiently sparse vector whose true value can be obtained by means of sparse signal recovery. Additional information on the type and location of the event could be translated into a partial knowledge on the support of unknown vector which is beneficial for the sparse recovery to improve the solution. This information might be such precise that tells us the exact location of the event, and ultimately the location of zero and non-zero entries on the unknown vector. Or, it might be less accurate even compromised with error such that only leads into a brief knowledge about a subset of unknowns. Sparse signal recovery has got the capability to leverage any information within this spectrum.

There are recent methods that use D-PMU measurements to identify the location of events on power distribution systems, e.g., see [45]. If such methods are truly accurate, then we can use them to identify exactly which state variables belong to each of the four groups that we introduced in Section 2.2.3.

However, in practice, event location identification is not exact. This is particularly the case in low-observable networks, where the number of D-PMUs may not be enough to identify the exact location of all events. Interestingly, our approach here does not require knowing the exact location of the event.

Suppose we do not know the exact location of the event at bus 29 in Fig. 2.1(a), but we do know that the event occurs somewhere on the lateral that includes bus 26 to bus 33. In other words, suppose we only know the zone of the event. In that case, all the discussions in Section 2.2.3 on the features of voltage and current differential phasors for all

nodes and line segments in $(\mathcal{V}_1, \mathcal{E}_1)$ and $(\mathcal{V}_2, \mathcal{E}_2)$ and $(\mathcal{V}_3, \mathcal{E}_3)$ would still be valid. Moreover, all the previous analysis for the nodes and line segments on the path between the substation and the event zone up to node 29 would also still be valid.

The difference between knowing the exact event bus versus only knowing the zone of the event would affect our analysis only with respect to the differential voltage and current phasors for the nodes and line segments in $(\mathcal{V}_4, \mathcal{E}_4)$. Given that we do not know the exact location of the event at bus 29, we can simply merge set $(\mathcal{V}_4, \mathcal{E}_4)$ to set $(\mathcal{V}, \mathcal{E})$. That is, we simply do not include set $(\mathcal{V}_4, \mathcal{E}_4)$ in our sparsity assumptions; everything else will remain the same as in Section 2.2.3. In this condition, without loss of generality, we assume that $(\mathcal{V}_4, \mathcal{E}_4)$ belongs to the path of the event and we treat their corresponding state variables as we do for the ones on the red path in Fig.2.1(a). Note that the predefined four types of state variables will be deployed in the sparse recovery in terms of partial knowledge on the support of unknown vector as we will discuss in next Section. Thus, as far as we obtain a valid knowledge on a part of unknown vector and the error is small, we can expect improvements. Accordingly, we should make sure that number of nodes in each zone are such that the caused error on the side information is kept small. Meanwhile, the number of nodes within each zone should be such that the event zone is detected correctly from the available sensors.

As for how to obtain the event zone, we propose an extension of the event location identification method in [45].

The method starts from two D-PMUs, one at the substation and one at the end of a lateral of interest. It then goes through a forward nodal differential voltage calculation

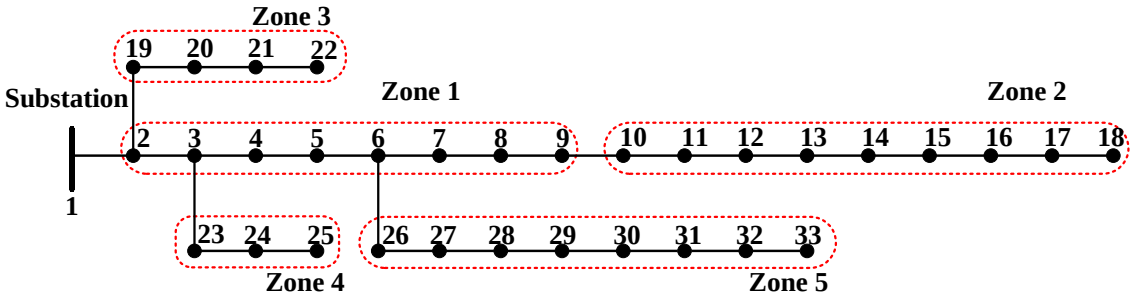

Figure 2.3: A typical zoning of the radial distribution system.

that starts from the D-PMU at the substation and ends at the D-PMU at the end of the latter; and also a backward nodal differential voltage calculation that starts from the D-PMU at the end of the lateral and ends at the D-PMU at the substation. The bus number for the event is then obtained by comparing the two nodal differential voltage calculations as taking the minimum of their discrepancy, as in Eq. (12) in [45].

To find the zone of event, first, we split the distribution feeder into several mutually exclusive zones and represent each zone by an index c. One simple way to form the zones is to consider all the nodes on the same lateral as one zone, provided that we have installed at least one D-PMU on each lateral. (see Fig. 2.3).

Then, we can simply replace Eq. (12) in [45] with the following:

$$c^\star = \arg\min_c \frac{1}{n_c} \left\| \Delta \mathbf{V}^{c,f} - \Delta \mathbf{V}^{c,b} \right\|_2, \tag{2.28}$$

where n_c denotes the number of buses in zone c; $\Delta \mathbf{V}^{c,f}$ is the vector of estimated differential voltage phasors from the forward sweep in zone c; and $\Delta \mathbf{V}^{c,b}$ is the same vector from the backward sweep. For more details please refer to [45].

2.2.5 DSSE as a Sparse Signal Recovery Problem

We now go back to the DSSE problem in differential mode based on the system of linear equations in (2.18). Recall that this system of equations is undetermined. Based on grounds for sparsity that we established in Section 2.2.3, we can now formulate this problem in a form that can be solved using techniques in sparse signal recovery. Accordingly, we can track the changes in the state variables across the power distribution network whenever an event happens; despite the fact that the system is low-observable due to (2.14).

To recover a sparse solution of the system of equations in (2.18), we can solve the following basis pursuit problem which is a well-known convex relaxation of the original intractable ℓ_0-norm minimization problem in sparse recovery [41]:

$$\min_{\Delta \mathbf{x}^t} ||\Delta \mathbf{x}^t||_1 \quad \text{s.t.} \quad \Delta \mathbf{z^t} = \mathbf{H} \Delta \mathbf{x}^t, \tag{2.29}$$

where $\| \cdot \|_1$ denotes the ℓ_1-norm.

Suppose, we do not know where the zone of event is and all of our information about the vector of state variables is limited to the fact that it is a sparse vector. In that case, we cannot distinguish the state variables from each other and all we can do is to solve the optimization problem in (2.29).

The optimization (2.29) is convex but non-smooth, c.f. [22]. A computationally convenient way to solve (2.29) is to use ADMM algorithm [22]. In this regard, we first rewrite problem (2.29) as:

$$\begin{aligned} &\min_{\Delta \mathbf{x}^t, \Delta \mathbf{y}} \frac{1}{2} ||\Delta \mathbf{z}^t - \mathbf{H} \Delta \mathbf{x}^t||_2^2 + \lambda ||\Delta \mathbf{y}||_1 \\ &\text{s.t.} \qquad \Delta \mathbf{x}^t = \Delta \mathbf{y}. \end{aligned} \tag{2.30}$$

Next, we use the following system of iterative equations to solve the above reformulated optimization problem [22]:

$$\begin{aligned}
\Delta \mathbf{x}_{k+1}^{t} &:= (\mathbf{H}^{\top}\mathbf{H} + \rho \mathbf{I})^{-1}(\mathbf{H}^{\top}\Delta \mathbf{z}^{t} + \rho(\Delta \mathbf{y}_{k} - \mathbf{u}_{k})) \\
\Delta \mathbf{y}_{k+1} &:= S_{\lambda/\rho}(\Delta \mathbf{x}_{k+1}^{t} + \mathbf{u}_{k}) \\
\mathbf{u}_{k+1} &:= \mathbf{u}_{k} + \Delta \mathbf{x}_{k+1}^{t} - \Delta \mathbf{y}_{k+1}
\end{aligned} \tag{2.31}$$

where $\rho > 0$ is the penalty parameter for the augmented Lagrangian function, subscript k denotes the k-th iteration, u is the dual variable corresponding to the constraint in (2.30), and operator $S_{\lambda/\rho}$ is the proximal operator.

It should be noted that the iterations in (2.31) are internal to each time slot. That is, every time slot t involves its own iterations as in (2.31) in order to obtain the state estimation solutions at that time slot.

The ADMM iterations in (2.31) determine the values of the very small entries, i.e. those that can be approximated by zero in the sparse recovery process, based on the values of λ and ρ, and the pre-determined maximum number of iterations. This provides an advantage in implementing the DSSE algorithm because it provides a knob to control the extent of the zero-approximation for the states that seem small.

2.2.6 Dynamic Reweighting

Given the differential nature of the state variables in this study, the performance of the proposed DSSE problem can be further enhanced through learning its sparse characteristics, i.e., learning which state variables can be zero-approximated. This can be done in each time slot by using a dynamic reweighting based on the largest obtained values from the previous time slot. The changes in the states of the system are due to changes in loads or

grid components. The buses that experience the largest changes in their voltage are often those that are close to such loads or grid components. In this regard, we propose to rewrite (2.30) as:

$$\begin{aligned} &\min_{\Delta\mathbf{x}^t, \Delta\mathbf{y}} \frac{1}{2}||\Delta\mathbf{z}^t - \mathbf{H}\Delta\mathbf{x}^t||_2^2 + \lambda||\Delta\mathbf{y}||_1 \\ &\text{s.t.} \qquad \mathbf{F}^t\Delta\mathbf{x}^t = \Delta\mathbf{y}. \end{aligned} \tag{2.32}$$

where $\mathbf{F}^t$ is a diagonal $N \times N$ reweighting matrix. Suppose we are at time slot t. Let Γ_{t-1} denote the set of half of largest differential states that were non-zero during time slot $t-1$, i.e., during the previous time slot. Updating rule for entries of $\mathbf{F}^t$ is:

$$f_{ii}^t = \begin{cases} 1 \qquad ;\text{if} \qquad i \in \Gamma_{t-1} \\ 1/|_i^{t-1} + \epsilon| \quad ; \quad \text{otherwise} \end{cases} \tag{2.33}$$

where ϵ is a very small positive value. From (2.33), if the previous differential state was within the half of largest non-zero, i.e., the voltage or current phasor corresponding to that state variable had a considerable change, then the weight corresponding to that differential state remains 1, which is the default value as in (2.30). However, if the previous differential state was zero, then the weight is set in an inverse proportional relationship with its value during the previous time slot. Devoting a small weight (here 1) to large non-zero elements of state variables vectors, relaxes the constraint of sparsity on these entries. Regarding the simulation results, large non-zero state variables are generally: 1) real parts of voltage difference for the nodes that are closer to the node with the load change, and 2) current difference for the line segments close to the substation node or node with the load change. Inverse proportional relationship assigns large penalty coefficients to the approximately zero element, which helps the algorithm to improve its accuracy in the next time slots by ignoring the corresponding state variables.

Given the updated matrix $\mathbf{F}^t$ in each time slot t, the iterative solutions in (2.31) are updated as:

$$\begin{aligned}
\Delta\mathbf{x}_{k+1}^t &:= (\mathbf{H}^\top\mathbf{H} + \rho(\mathbf{F}^t)^\top\mathbf{F}^t)^{-1}(\mathbf{H}^\top\Delta\mathbf{z}^t + \rho\mathbf{F}^t(\Delta\mathbf{y}_k - \mathbf{u}_k)) \\
\Delta\mathbf{y}_{k+1} &:= S_{\lambda/\rho}(\mathbf{F}^t\Delta\mathbf{x}_{k+1}^t + \mathbf{u}_k) \\
\mathbf{u}_{k+1} &:= \mathbf{u}_k + \mathbf{F}^t\Delta\mathbf{x}_{k+1}^t - \Delta\mathbf{y}_{k+1}
\end{aligned} \tag{2.34}$$

Thus far, we have not incorporated any information about the zero/non-zero being of state variables that we extract from the identification of events. In the next step, we will modify (2.29) to incorporate the following "side information" so that we can solve the above optimization problem:

1) the four groups of state variables that we introduced in Section 2.2.3; and 2) the zone of the event that we identified in Section 2.2.4.

2.3 Solution Method, Uniqueness, and Calibration

2.3.1 Incorporating Side Information

The sparse DSSE problem as written in (2.29) treats all the entries in the unknown vector $\Delta\mathbf{x}^t$ in a uniform manner and does not include any prior information on the support of unknown vector. Since the side information provides us with an estimate of the support of the unknown vector, we can exploit that knowledge to improve the performance of the sparse recovery method. To incorporate the side information that we listed in Section 2.2.4, we modify (2.29) into an adaptive group-sparse recovery problem [26]. Recall from Section

2.2.3 that we divided all the state variables in differential mode into four types. Let us refer to them as Type 1, Type 2, Type 3, and Type 4. Accordingly, suppose vector $\Delta\mathbf{x}^t$ is partitioned into P groups. Three partitions are formed based on the state variables of Type 1, Type 2, and Type 4, respectively. For the state variables of Type 3, the number of partitions is equal to the number of sets $\mathcal{V}_i$, as we defined in Section 2.2.3. As we discussed in Section 2.2.3, the state variables in each partition have similar tendency towards being (approximately) zero or non-zero. If instead of the event location, we use the event zone, then the definition of partitions are adjusted accordingly, as we discussed in Section 2.2.4.

Let us represent the state variables in each partition p by $\Delta\mathbf{x}_p^t$, where $p = 1, \ldots, P$. Let w_p denote the adaptive weight for partition p. Set $w_p = 1$ for Type 1 state variables because they can be replaced by zero. Set $w_p = 0$ for Type 2 and Type 4 state variables; since they are expected to be non-zero. Set w_p for Type 3 state variables to be a number between 0 and 1 based on their distribution of historical zero/non-zero values.

To incorporate the weights, we reformulate the basis pursuit problem (2.29) as the following weighted ℓ_1-norm minimization:

$$\min_{\Delta\mathbf{x}^t} \sum_{p=1}^{P} w_p \|\Delta\mathbf{x}_p^t\|_1 \quad \text{s.t.} \quad \Delta\mathbf{z}^t = \sum_{p=1}^{P} \mathbf{H}_p \Delta\mathbf{x}_p^t, \tag{2.35}$$

where $\mathbf{H}_p$ denotes a submatrix of $\mathbf{H}$ with only the columns associated with the state variables in partition p.

It is more common to solve the related unconstrained relaxation of (2.35), a.k.a., weighted Lasso problem [104] as

$$\min_{\Delta\mathbf{x}^t} \frac{1}{2} \|\Delta\mathbf{z}^t - \sum_{p=1}^{P} \mathbf{H}_p \Delta\mathbf{x}_p^t\|_2^2 + \lambda \sum_{p=1}^{P} w_p \|\Delta\mathbf{x}_p^t\|_1, \tag{2.36}$$

where $\lambda > 0$ is a regularization parameter which controls the trade-off between the estimation error and the sparsity level of the unknown vector $\Delta\mathbf{x}^t$. By changing λ we change the width of the yellow ribbon in Figs. 2.1(b) and (c). The first term in (2.36) is the squared error loss and the second term is the weighted ℓ_1-norm penalty. If λ is such that $\lambda w_p > \|\mathbf{H}_p^\top \Delta\mathbf{z}^t\|_\infty$, then we force the corresponding estimated $\Delta\mathbf{x}_p^t$ to be zero.

2.3.2 ADMM Solution

Similar to previous Section, we use ADMM to solve problem (2.36). We obtain the augmented Lagrangian for (2.36), under constraints $\Delta\mathbf{x}_p^t - \boldsymbol{\mu}_p = 0$, as

$$\begin{aligned}\mathcal{L}_\rho(\Delta\mathbf{x}^t, \boldsymbol{\mu}, \mathbf{u}) = \frac{1}{2}\|\Delta\mathbf{z}^t - \sum_{p=1}^{P} \mathbf{H}_p \Delta\mathbf{x}_p^t\|_2^2 + \lambda \sum_{p=1}^{P} w_p \|\boldsymbol{\mu}_p\|_1 \\ + \rho\langle \mathbf{u}, \Delta\mathbf{x}^t - \boldsymbol{\mu}\rangle + \frac{\rho}{2}\|\sum_{p=1}^{P}(\Delta\mathbf{x}_p^t - \boldsymbol{\mu}_p)\|_2^2,\end{aligned} \tag{2.37}$$

where $\Delta\mathbf{x}^t$ and $\boldsymbol{\mu}$ are the primal variables, $\mathbf{u}$ is the dual variable, and $\rho > 0$ is a Lagrangian parameter.

The ADMM method alternately updates all the primal and dual variables at every iteration as follows

$$\Delta\mathbf{x}^t = \arg\min_{\Delta\mathbf{x}^t} \mathcal{L}_\rho(\Delta\mathbf{x}^t, \tilde{\boldsymbol{\mu}}, \tilde{\mathbf{u}}) \tag{2.38}$$

$$\boldsymbol{\mu} = \arg\min_{\boldsymbol{\mu}} \mathcal{L}_\rho(\Delta\mathbf{x}^t, \boldsymbol{\mu}, \tilde{\mathbf{u}}) \tag{2.39}$$

$$\mathbf{u} = \tilde{\mathbf{u}} + (\Delta\mathbf{x}^t - \boldsymbol{\mu}), \tag{2.40}$$

where $(\Delta\tilde{\mathbf{x}}, \tilde{\boldsymbol{\mu}}, \tilde{\mathbf{u}})$ and $(\Delta\mathbf{x}, \boldsymbol{\mu}, \mathbf{u})$ denote the estimates from the previous and the current iteration, respectively. The solutions to (2.38) and (2.39) can be obtained by setting the

derivatives of (3.29) w.r.t. $\Delta\mathbf{x}^t$ and $\boldsymbol{\mu}^t$, respectively. The resulting closed form solutions for the updates can be described as

$$\Delta\mathbf{x}^t = (\mathbf{H}^\top\mathbf{H} + \rho\mathbf{I})^{-1}(\mathbf{H}^\top\Delta\mathbf{z}^t + \rho(\tilde{\boldsymbol{\mu}} - \tilde{\mathbf{u}})) \tag{2.41}$$

$$\boldsymbol{\mu}_p = \mathcal{S}_{w_p\lambda/\rho}(\Delta\mathbf{x}_p^t + \tilde{\mathbf{u}}_p) \qquad \forall p \in P \tag{2.42}$$

$$\mathbf{u} = \tilde{\mathbf{u}} + (\Delta\mathbf{x}^t - \boldsymbol{\mu}) \tag{2.43}$$

where $\mathcal{S}_{w_p\lambda/\rho}(\cdot)$ is a soft-thresholding operator; and $\boldsymbol{\mu}_p$, $\Delta\mathbf{x}_p^t$, and $\tilde{\mathbf{u}}_p$ forms the p-th partition of the respective vectors.

Importantly, the convergence of the ADMM is guaranteed for the above sparse DSSE problem; because the objective function in the proposed problem formulation is the sum of several convex functions with linear constraints, c.f [22].

2.3.3 Uniqueness of the Sparse DSSE Solution

The conditions for the uniqueness of the solution in a sparse signal recovery problem typically depend on the structure of the measurement matrix $\mathbf{H}$, the number of non-zero entries, and the support and the sign sequence of the non-zero entries. A number of criteria have been developed to analyze the uniqueness of the solution which leads to extremely tight sufficiency condition [24, 25, 41, 42, 104, 46]. Mutual coherence and restricted isometry-based analysis address the worst-case scenarios and provide pessimistic results for matrices with highly correlated columns.

Thus, to check the uniqueness of the sparse DSSE solution, we extend the dual certificate-based method in [113], where we incorporate the side information by assigning different values to w_p for different partitions $p = 1, \ldots, P$.

Suppose the true sparse signal $\Delta\mathbf{x}$ in (2.18) has non-zero entries on an index set Γ. Let us denote $\Delta\mathbf{x}_\Gamma$ as the non-zero entries, $\mathbf{s}_\Gamma = \text{sign}(\Delta\mathbf{x}_\Gamma)$, and $\mathbf{H}_\Gamma$ as a matrix that consists of the columns in $\mathbf{H}$ that correspond to the indices in Γ. We can show that $\Delta\mathbf{x}$ is also the unique solution of the ℓ_1-norm minimization problem in (2.36) using the following theorem.

Theorem 1: A vector $\Delta\mathbf{x}$ that satisfies the equations in (2.18) with support Γ and sign sequence $\mathbf{s}_\Gamma$ can be recovered from the unique optimal point of the optimization problem in (2.36) if the following conditions are satisfied:

1. Submatrix $\mathbf{H}_\Gamma$ is full-rank.

2. For all $\gamma \notin \Gamma$, the following inequality holds:

$$|\mathbf{H}_\gamma^\top \mathbf{H}_\Gamma (\mathbf{H}_\Gamma^\top \mathbf{H}_\Gamma)^{-1} \mathbf{W}_\Gamma \mathbf{s}_\Gamma| < w_\gamma. \tag{2.44}$$

3. λ and $\mathbf{W}$ are selected so that

$$\text{sign}(\Delta\mathbf{x}_\Gamma - \lambda(\mathbf{H}_\Gamma^\top \mathbf{H}_\Gamma)^{-1} \mathbf{W}_\Gamma \mathbf{s}_\Gamma) = \mathbf{s}_\Gamma. \tag{2.45}$$

$\mathbf{H}_\gamma$ denotes the column of $\mathbf{H}$ with index γ, $\mathbf{W}_\Gamma$ is a diagonal matrix whose diagonal entries are w_p for the partitions associated with the indices in Γ, and w_γ is the corresponding weight of the partition that γ belongs to.

The proof of Theorem 1 is as follows:

We largely follow the procedure outlined in [46]. We will first discuss the necessary and sufficient conditions that any optimal point of the convex optimization problem in (2.36) must satisfy. Let us rewrite the problem in (2.36) in a compact form as

$$\min_{\Delta\mathbf{x}} \frac{1}{2} \|\Delta\mathbf{z} - \mathbf{H}\Delta\mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\Delta\mathbf{x}\|_1, \tag{2.46}$$

where $\mathbf{W}$ is a diagonal matrix with weights w_γ for $\gamma \in \{1, \ldots, N\}$. The optimality conditions can be obtained by using the first-order necessary KKT conditions that are also sufficient; the optimization problem in (2.36) is convex. A necessary condition for vector $\Delta\mathbf{x}^*$ to be an optimal solution of (2.46) is that zero vector belongs to the subgradient of its objective. We can write this condition as

$$\mathbf{H}^\top(\mathbf{H}\Delta\mathbf{x}^* - \Delta\mathbf{z}) + \lambda\mathbf{W}\partial\|\Delta\mathbf{x}^*\|_1 = 0. \tag{2.47}$$

$\partial\|\Delta\mathbf{x}^*\|_1$ denotes the subgradient which can be defined as

$$\partial\|\Delta\mathbf{x}^*\|_1 \in \begin{cases} \text{sign}(\Delta\mathbf{x}^*) & \Delta\mathbf{x}^* \neq 0 \\ [-1, 1] & \Delta\mathbf{x}^* = 0. \end{cases}$$

We can write (2.47) in a compact form as

$$|\mathbf{H}_\gamma^\top(\mathbf{H}\Delta\mathbf{x}^* - \Delta\mathbf{z})| \leq \lambda w_\gamma.$$

One can show that the solution is a unique minimum if the following sufficient conditions are satisfied [22]:

$$\mathbf{H}_\gamma^\top(\mathbf{H}\Delta\mathbf{x}^* - \Delta\mathbf{z}) = -\lambda w_\gamma \text{sign}(\Delta\mathbf{x}_\gamma^*), \text{ if } \Delta\mathbf{x}_\gamma^* \neq 0 \tag{2.48}$$

$$|\mathbf{H}_\gamma^\top(\mathbf{H}\Delta\mathbf{x}^* - \Delta\mathbf{z})| < \lambda w_\gamma, \text{ if } \Delta\mathbf{x}_\gamma^* = 0. \tag{2.49}$$

We can now prove that if a vector $\Delta\mathbf{x}$ with support Γ and sign sequence $\mathbf{s}_\Gamma$ satisfies the set of equations $\mathbf{H}\Delta\mathbf{x} = \Delta\mathbf{z}$ and the conditions in (2.44) and (2.45), then we can explicitly write the unique solution of (2.46) supported on Γ as

$$\Delta\mathbf{x}_\Gamma^* = \Delta\mathbf{x}_\Gamma - \lambda(\mathbf{H}_\Gamma^\top\mathbf{H}_\Gamma)^{-1}\mathbf{W}_\Gamma\mathbf{s}_\Gamma. \tag{2.50}$$

The sign condition in (2.45) implies that $\text{sign}(\Delta \mathbf{x}^*_\Gamma) = \mathbf{s}_\Gamma$. To prove the optimality of $\Delta \mathbf{x}^*$ in (2.50), we need to show that $\Delta \mathbf{x}^*$ satisfies the necessary and sufficient conditions in (2.48) and (2.49). Let us first look at the condition in (2.48) as

$$\begin{aligned}
&\mathbf{H}_\Gamma^\top (\mathbf{H}_\Gamma \Delta \mathbf{x}^*_\Gamma - \Delta \mathbf{z}) \\
&= \mathbf{H}_\Gamma^\top (\mathbf{H}_\Gamma [\Delta \mathbf{x} - \lambda (\mathbf{H}_\Gamma^\top \mathbf{H}_\Gamma)^{-1} \mathbf{W}_{\mathbf{\Gamma}} \mathbf{s}_\Gamma] - \Delta \mathbf{z}) \\
&= \mathbf{H}_\Gamma^\top (\underbrace{\mathbf{H}_\Gamma \Delta \mathbf{x} - \Delta \mathbf{z}}_{=0}) - \lambda \mathbf{W}_{\mathbf{\Gamma}} \mathbf{s}_\Gamma \\
&= -\lambda \mathbf{W}_{\mathbf{\Gamma}} \mathbf{s}_\Gamma = -\lambda \mathbf{W}_{\mathbf{\Gamma}} \text{sign}(\mathbf{x}^*_\Gamma).
\end{aligned} \tag{2.51}$$

For the condition in (2.49), we can show that for any $\gamma \notin \Gamma$

$$\begin{aligned}
&|\mathbf{H}_\gamma^\top (\mathbf{H}_\Gamma \Delta \mathbf{x}^*_\Gamma - \Delta \mathbf{z})| \\
&= |\mathbf{H}_\gamma^\top (\mathbf{H}_\Gamma [\Delta \mathbf{x} - \lambda (\mathbf{H}_\Gamma^\top \mathbf{H}_\Gamma)^{-1} \mathbf{W}_{\mathbf{\Gamma}} \mathbf{s}_\Gamma] - \Delta \mathbf{z})| \\
&= |\lambda \mathbf{H}_\gamma^\top \mathbf{H}_\Gamma (\mathbf{H}_\Gamma^\top \mathbf{H}_\Gamma)^{-1} \mathbf{W}_{\mathbf{\Gamma}} \mathbf{s}_\Gamma| \\
&< \lambda w_\gamma.
\end{aligned} \tag{2.52}$$

The last inequality is due to the main condition in (2.44). The above inequality completes the proof; since $\Delta \mathbf{x}^*_\Gamma$, as defined in (2.50), satisfies the necessary and sufficient conditions to be the strict optimal solution of the problem in (2.46).

The conditions in (2.44) relax the tightness of similar conditions discussed in [113] for standard Lasso problem in which all weights w_γ are equal. From (2.44), if a w_γ is very large, then the corresponding condition becomes irrelevant and does not affect the uniqueness of the solution. This condition supports our claims that incorporating the side information can help us recover the correct solution with provable guarantees.

2.3.4 Computational Complexity

The main computational complexity for solving the Lasso problem in (2.36) comes from the complexity of solving the least-squares optimization problem in (2.38). This optimization problem is solved once at every iteration. The solution for this problem is provided in (2.41). We can pre-calculate and store the inverse matrix $(\mathbf{H}^\top\mathbf{H}+\rho\mathbf{I})^{-1}$ and use it in every iteration, for as long as the topological configuration of our network remains unchanged. The typical computational cost for solving such least-square optimization problem for a dense matrix $\mathbf{H}$ with size $M \times N$ is $\mathcal{O}(MN + N^2)$. However, since $\mathbf{H}$ in our problem is sparse, the cost of computing is reduced to $\mathcal{O}(kM + N^2)$, where k is the number of nonzero entries in each row of matrix $\mathbf{H}$. For the base case scenario in our case studies where we have only six D-PMUs in the system, we have: $k \leq 8$. As for the number of iterations in our case studies, in the worst case scenario, we approximately use a maximum of 1000 iterations in the process of solving the Lasso problem.

As for the WLS-DSSE method with linear equations, the cost of computation per iteration is $\mathcal{O}(MN^2 + N^3)$. Therefore, our proposed method is computationally less expensive than the conventional WLS-DSSE method. We will further discuss the computation time in the case studies in Section 2.4.1.

2.3.5 Sparse Tracking Linear DSSE

The solution for $\Delta\mathbf{x}^t$ tells us whether and how the states of the power distribution system change during each measurement interval. This by itself is an important result. However, to make the DSSE results more useful to the utility operator, we need to convert

the results from the differential mode back to the standard mode to obtain the standard state variables. This can be done in the context of tracking state estimation.

Similar to the real-world power distribution networks, we assume that there are other legacy meters available in addition to the D-PMUs, such as through a SCADA system and/or an AMI. Regardless of how the legacy meters are installed at the primary side and/or the secondary side of the network, they are assumed to be sufficient to achieve full-observability at the moment when such legacy measurements become available. However, importantly, as we discussed in Section I-A, the low reporting rate of the legacy meters requires the power system operator to operate the power distribution system without updating the state variables for 30 to 60 minutes, i.e., until the next readings of the legacy meters become available [91]. Accordingly, our goal in this work is to take advantage of the high reporting rate of the D-PMUs to continue updating the state variables; yet addressing the low-observability in D-PMU measurements, i.e., the fact that in practice only a few D-PMUs can be available on each power distribution feeder.

In this setting, the measurements from legacy meters are assumed to be used in order to update the initial values for the tracking state estimation problem at time $t = 0$. The details on how such initial values are obtained by using the measurements from the legacy meters are not of concern in this study; because they are beyond the scope of this work.

Once the initial values are provided by the measurements from the legacy meters, we start the use of the proposed DSSE method to continue updating the state variables without access to any new measurement from the legacy meters. In this regard, we keep

solving the low-observable DSSE problem in (2.18) once every 100 milliseconds solely based on the measurements from only a few D-PMUs for the next 30 to 60 minutes. In this 30 to 60 minutes period, we keep adding the estimated values that are obtained in differential mode to the initial values in order to obtain the nodal voltage phasors for the whole network at each iteration:

$$\mathbf{V}^t = \mathbf{V}^0 + \sum_{\tau=1}^{t-1} \Delta \mathbf{V}^\tau \tag{2.53}$$

where $\mathbf{V}^0$ denotes the vector of the initial voltage phasors that is obtained at time $t = 0$ by using the measurements from the legacy low-reporting-rate meters in the system. Note that, if needed, the time-step for the proposed tracking DSSE can be adjusted with respect to the time frame in which the dynamics in the power distribution system may change due to the events.

2.3.6 Drift Identification and Calibration

Drift Identification

By solving the adaptive group Lasso in (2.36), we minimize the error in estimating the differential voltage phasors at each time slot. However, as time goes by, we inevitably see accumulative error in estimating the voltage phasors in (2.53). If the time interval between the two consecutive AMI readings is too long, then such accumulative error can ultimately result in a considerable drift in the DSSE results from the true values. Therefore, we need a mechanism to detect drifting and calibrate the estimated values. In order to identify a possible drifting in the state estimation results, we check the following condition:

$$\|\mathbf{z}^t - \mathbf{H}\mathbf{x}^t\|_2^2 \geq \beta, \tag{2.54}$$

where β is a threshold parameter. If condition (2.54) holds at any time t, then we trigger a calibration mechanism.

Calibration

Our calibration method works based on the assumption that D-PMU data is reliable and the bad data is already removed. At each time interval t, D-PMUs can serve as reference to obtain voltage at the few specific locations where D-PMUs are installed. Accordingly, whenever calibration is needed, we use the available D-PMU measurements in each partition p as the reference point to calibrate the voltage estimation for all the nodes within the same partition through:

$$\mathbf{V}_{p,cal}^{t} := \mathbf{V}_{p}^{t} \times \zeta_p, \tag{2.55}$$

where ζ_p is the ratio of the voltage phasor measured by the D-PMU to the voltage phasor estimated at the same bus in partition p. The ratio is calculated separately for the real part and the imaginary part of the voltage phasor.

2.4 Case Studies

The case studies are done by simulating the IEEE 33-bus distribution test network [18]. Tracking DSSE is performed upon receiving the D-PMU measurements at 100 milliseconds intervals during the one hour period between two consecutive AMI readings. The legacy meters' data provide the initial states for the proposed tracking DSSE. At most, one event is assumed to occur within each time slot between the two readings of D-PMUs. Each event can be a change in active/reactive power load, distributed generation, or capacitor

bank switching; and it is assumed to be within ±50% of its value at the previous time slot. The loads are assumed to be constant-power loads.

Unless we state otherwise, we consider five zones in the network. The longer lateral is divided into two zones. Each shorter lateral is one zone. One D-PMU is placed at the substation; and five D-PMUs are placed at the end of each zone, i.e. at buses 9, 18, 22, 25, and 33. Thus, the ratio of the available measurements to the unknowns is $12/65 = 0.185$. Error in D-PMU measurements have zero mean and standard deviation of $\sigma_V = 0.1\%$ for voltage measurements and $\sigma_I = 1\%$ for current measurements. A total of 120,000 random scenarios are generated in MATPOWER in MATLAB R2018b.

2.4.1 Performance Comparison

Performance comparison is done with the conventional WLS-DSSE method [57]. For the WLS-DSSE method, we inevitably need to add pseudo-measurements to make the network fully-observable; otherwise the WLS-DSSE method cannot solve the undetermined system of equations for the DSSE problem. To have a consistent comparison with our proposed method, we used the initial measurements from the legacy meters at time $t = 0$ as pseudo-measurement for the rest of time slots. Of course, the pseudo-measurements that come from the legacy meters will be updated at the next interval when new measurements become available, i.e., after 30 to 60 minutes. Since differential synchrophasors are small, we use mean absolute percentage error (MAPE) to assess the estimation of differential phasors:

$$\text{MAPE} = \frac{1}{2N}\sum_{i=1}^{N}\left(\left|\frac{\Re(\Delta\hat{x}_i - \Delta x_i)}{\Re(\Delta x_i)}\right| + \left|\frac{\Im(\Delta\hat{x}_i - \Delta x_i)}{\Im(\Delta x_i)}\right|\right) \times 100\%, \tag{2.56}$$

where $\Delta\hat{x}_i$ denotes the estimated value and Δx_i denotes the true value of the i-th differential state variable, respectively. Also, we use the mean absolute error (MAE) to assess the estimation of the synchrophasors:

$$\text{MAE} = \frac{1}{2N}\sum_{i=1}^{N}\Big(\big|\Re(\hat{x}_i - x_i)\big| + \big|\Im(\hat{x}_i - x_i)\big|\Big), \tag{2.57}$$

where $\hat{x}_i$ denotes the estimated value and x_i denotes the true value of the i-th state variable, respectively.

The statistical results for performance comparison are shown in Table 2.2. We can see that the proposed method has a much lower MAPE and MAE; both in terms of its mean and its standard deviation. In 97.9% of all the scenarios, the MAPE is less for the proposed method than WLS-DSSE.

Given the sparsity in the vector of state variables, the true value for some of state variables is zero; this causes the denominator in (4.16) to be zero. As a result, if a state variable with zero value is estimated incorrectly with a non-zero value, then it would cause the MAPE to become infinity; which would make the MAPE useless. Therefore, for those rare scenarios, we consider the percentage error of the associated state variables to be 100% (as the maximum error level). Importantly, in our proposed DSSE method, the weighted Lasso optimization in (2.36) has an ℓ_1-norm penalty term whose associated weights are assigned such that the above misestimation is avoided. On the contrary, the conventional WLS-DSSE cannot prevent such misestimation; because it is not designed to recover a sparse vector. This is one of the reasons that MAPE for the WLS-DSSE is quite larger.

The average computation time, as well as the total number of divergent cases, are shown for each method in Table 2.3. We can see that the computation time of the

Table 2.2: Comparing different DSSE methods in Estimating Voltage Phasors and Estimating Differential Voltage Phasors

Method	Average MAPE	STD of MAPE	Average MAE $\times 10^{-3}$		STD of MAE $\times 10^{-3}$	
			No Calibration	Calibration	No Calibration	Calibration
Sparse Tracking with Side Information	4.51%	4.66	8.2	6.1	5.2	4.7
Sparse Tracking with Dynamic Reweighting	12.04%	12.74	9.1	-	4.6	-
The Conventional WLS-DSSE as in [57]	19.57%	9.65	28.2	-	17.4	-

proposed sparse DSSE method is much lower than that of the conventional WLS-DSSE. Also, unlike the WLS-DSSE method, the proposed method does not experience any case with divergence.

Table 2.3: Comparison of Computational Time and Convergence Ratio of DSSE Methods

Method	Average Computational Time (ms)	Number of Divergence Cases
The Proposed Sparse Tracking DSSE	764	0
The Conventional WLS-DSSE	2234	3748

2.4.2 State Estimation Profiles and Impact of Calibration

The state estimation results for an example scenario for differential voltage phasors and ordinary voltage phasors are shown in Figs. 2.4 and 2.5. Real parts of the estimated states are shown here. The results for the imaginary parts are similar and omitted due to space limitation. We can see that the proposed DSSE method significantly outperforms the other two methods. Moreover, the proposed calibration method has further enhanced the performance of voltage estimation in comparison to the case without calibration. Importantly, the calibration method is not applicable to the WLS-DSSE. That is why the curve for WLS-DSSE does not include the calibration.

Furthermore, the changes over time in the gross error in state estimation of the voltage phasors is shown in Fig. 2.6. As we can see, when the state estimation error hits the drift identification threshold β as in Eq. (2.54) at $t = t_1$, the proposed calibration mechanism is triggered and it helps the tracking DSSE method to reduce the gross error.

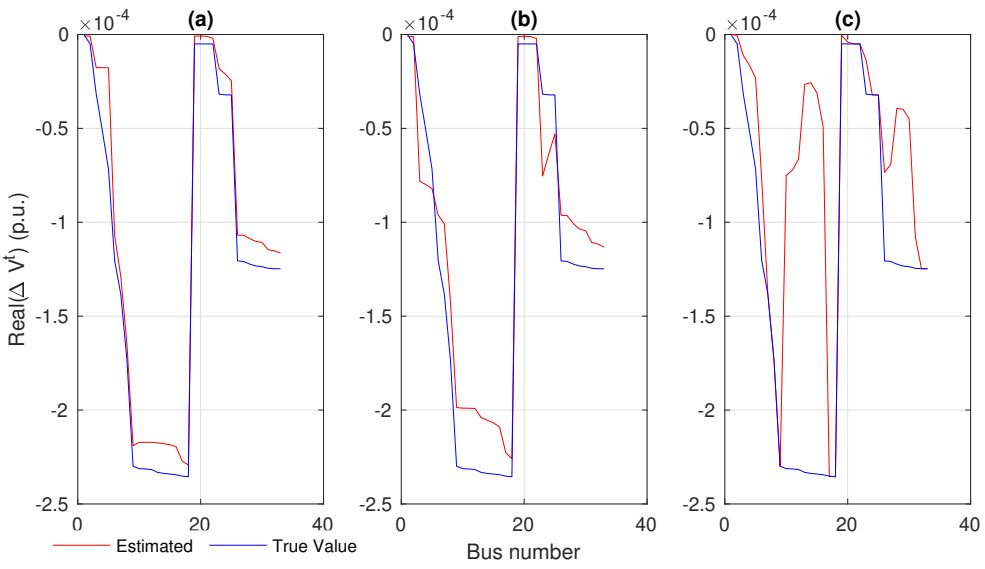

Figure 2.4: The estimation of the real part of *differential* voltage phasors at all buses in a typical scenario: a) the proposed sparse tracking DSSE with side information; b) the proposed sparse tracking DSSE with dynamic reweighting; and c) The conventional WLS-DSSE method in [57].

No need to say that this calibration can improve the performance only to a certain limit and the gross error inevitably continues to rise; until the next cycle when the measurements from the legacy meters become available at $t = t_2$. At that moment, the system becomes *momentarily fully-observable*; thus, the state estimation error is reset to zero. It is worth clarifying that, based on the setup in our case studies, the difference between using and not using the proposed calibration method is visible in Fig. 2.6 only during the period between $t = t_1$ and $t = t_2$. That is why the blue curve is visible only during this period.

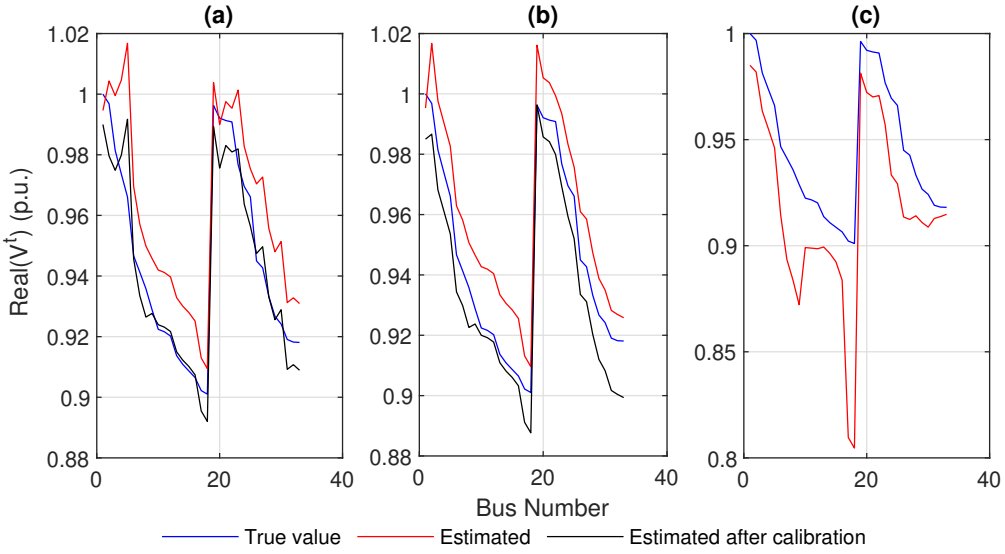

Figure 2.5: The estimation of the real part of the *ordinary* voltage phasors at all buses: a) the proposed sparse tracking DSSE with side information; b) the proposed sparse tracking DSSE with dynamic reweighting; and c) The conventional WLS-DSSE method in [57].

2.4.3 Importance of Side Information

Next, to examine how the side information improves the performance of the proposed DSSE method, we repeat the scenarios in Section 2.4.1 but this time we use non-adaptive sparse recovery, i.e., we do not use the side information. This results in using the same weights for all the unknowns in the problem formulation. The average MAPE increases from 4.51% to 15.58%; and the average MAE increases from 8.2×10^{-3} to 12.4×10^{-3}. The new error levels are still less than that of the conventional WLS-DSSE. These results show the importance of using the side information; which is one of the key contributions in this work.

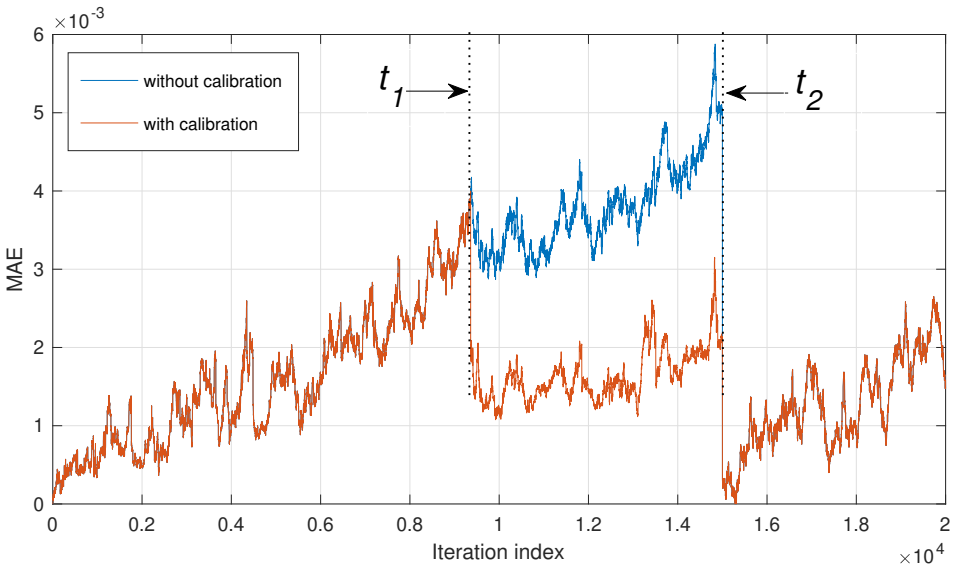

Figure 2.6: The gross estimation error of voltage phasors over time for the proposed sparse tracking DSSE method.

2.4.4 Impact of Renewable Energy Resources

In this section, we study the performance of proposed DSSE method in presence of distributed renewable energy resources. In accordance to our initial assumption that only one major event may happen at each time slot t, we assume that generation level changes only at one generation resource. Without loss of generality, suppose the distribution network has four renewable generators at nodes 14, 20, 23, and 30. Fig. 2.7 shows the DSSE results. Increasing the penetration rate of the renewable energy resources results in decreasing the accuracy of the DSSE method. The degraded accuracy in the DSSE algorithm is due to the degraded accuracy in the event zone identification. This makes the classification of state variables more challenging; although the basic sparsity features remain the same. In other words, the unknown vector is still a sparse vector even under high penetration of renewable energy resources; however, obtaining the side information about the location

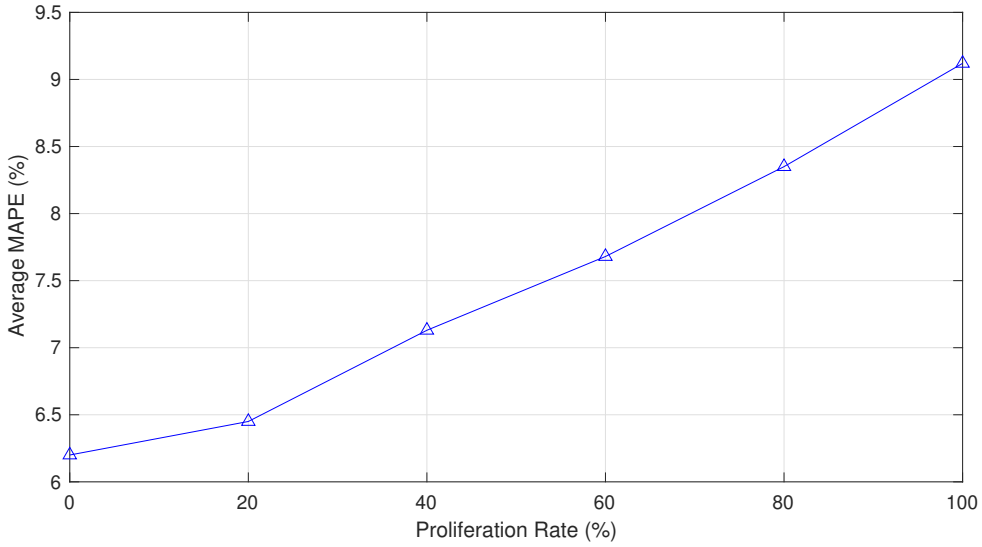

Figure 2.7: Average MAPE versus the penetration rate of renewable generation.

of the zero/non-zero entries is more difficult. This is not a surprising result; because the increasing penetration of renewable energy resources is a major challenge in practically every DSSE method; e.g., see a similar concern in [107].

2.4.5 Importance of the Number of D-PMUs

So far, we have assumed using five D-PMUs. However, increasing the number of D-PMUs can improve the performance of the proposed DSSE method; even though the network may still remain low-observable. The average MAPE with respect to the number of installed D-PMUs is shown in Fig. 2.8. Note that, the number of D-PMUs also affects the number of zones and the way that we extract the side information. However, to have a consistent comparison, we do *not* change the way that the zones are defined and identified. Based on the results in Fig. 2.8, adding more sensors slightly lowers the average estimation

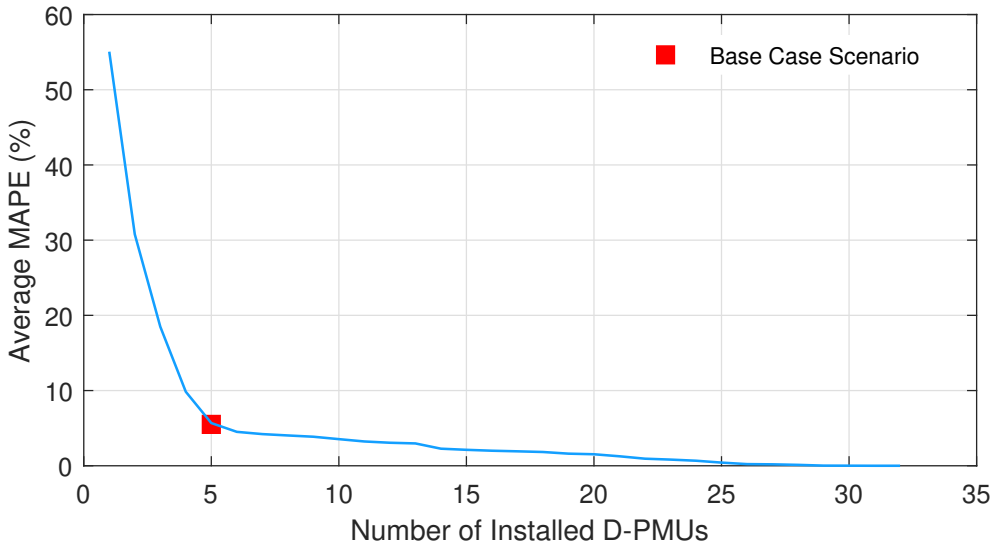

Figure 2.8: Average MAPE versus the number of newly installed D-PMUs.

error; however, we believe that for the defined sparse DSSE in a low-observable network, some points near the knee point on the curve can be assumed as the optimal number of sensors. One practical way to add new sensors is to simply equip buses which have the highest estimation error.

2.4.6 Importance of the Location of Event

To figure out the effect of the location of event on the performance of the proposed method, the distribution of average MAPE versus the node of the event is plotted in Fig. 2.9, where the red bar shows the median and the blue box shows the distribution of the first to the third quartile of MAPE. As we can see, the higher MAPE belongs to the scenarios wherein the event occurs at locations which have greater distance from the substation. This lowers the sparsity level of the unknown vector. Thus, the capability of the proposed

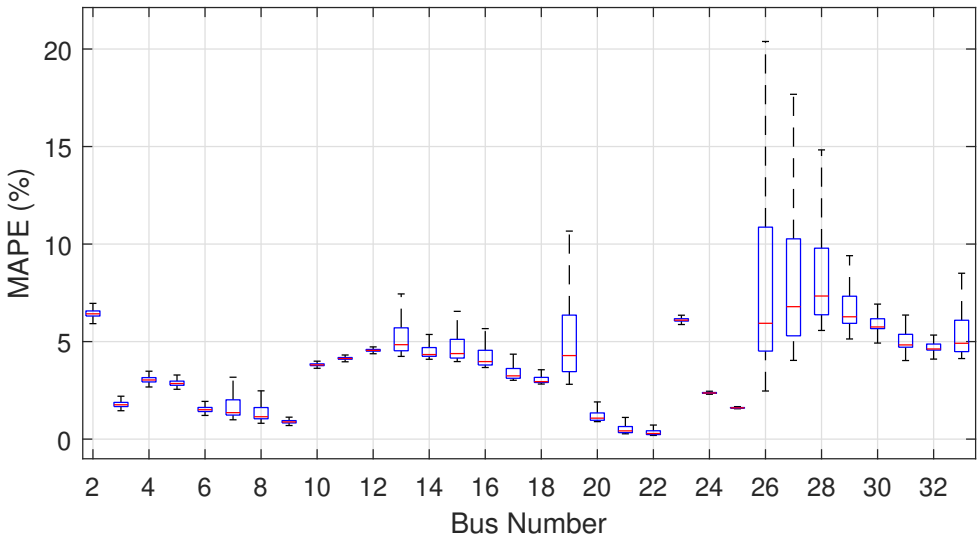

Figure 2.9: Distribution of MAPE with respect to the node of event

sparse DSSE method to recover the solution is compromised in such scenarios. Of course, the magnitude of the event also has impact on the MAPE. Events with lower magnitude cause such small changes in the operating points of the system that sparse DSSE method cannot track them. Once the magnitude is larger, the performance of proposed method is enhanced.

2.4.7 Importance of Measuring Phase Angle

A unique feature of D-PMUs is their ability to measure *phase angle* [82]. In this section, we seek to answer the following question: is it critical is to use phase angle measurements? To answer this question, we repeat all the scenarios, but we assume that phase angel at all nodes is zero. This will cause the average MAPE to increase from 4.51% to 11.76% and the average MAE to increase from 8.2×10^{-3} to to 11.39×10^{-3}. Furthermore,

removing phase angle causes a negative effect on the calibration mechanism to the extend that it increases the average MAE from 6.1×10^{-3} to 18.67×10^{-3}. Therefore, the use of D-PMUs is indeed necessary in this method.

2.4.8 Impact of Unbalanced Three-Phase Operation

In this section, we examine the IEEE 123-bus test system [2]; which is a highly unbalanced distribution network with Wye load connections. We use the percentage unbalance (PU) in current at the substation as the metric for unbalance in the network. We assume that only 19 D-PMUs are available; and they are installed at nodes 6, 11, 16, 18, 32, 39, 47, 56, 59, 66, 71, 75, 85, 86, 96, 101, 114, 250, and 450. This means that the ratio of the available measurements to the unknowns is $38/245 = 0.156$. Same as in the previous cases, we assume that each D-PMU measures one voltage phasor and one current phasor at the location where it is installed. All phasor measurements are three-phase.

We have constructed 100,000 via MATPOWER in MATLAB. We set the initial PU to 8%. Next, we modify the loads across the network to increase the PU level. As shown in Fig. 2.10, under various PU levels, the average MAPE is much lower for the proposed DSSE than for the conventional WLS-DSSE. Note that, since the DSSE formulation based on synchrophasor measurements is inherently linear; the unbalanced operation of the distribution system does not affect the core characteristics of the proposed sparse DSSE. The small increment in the MAPE is due to the increase in the neutral current as the unbalance level increases. Moreover, the estimation error in this case study is higher than that of the previous case studies based on the IEEE 33-bus test system. This is due to the higher number of small laterals for the IEEE 123-bus test system. It is not due to the unbalanced

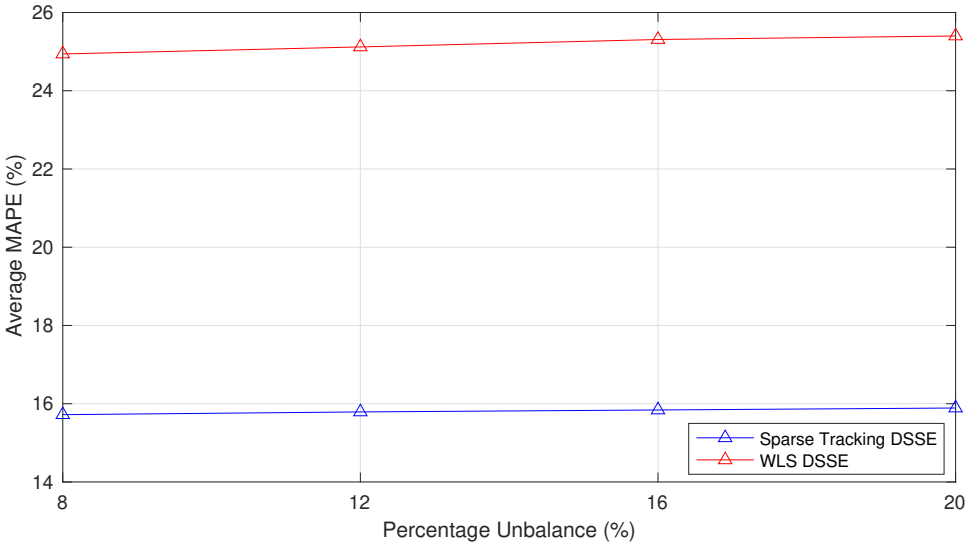

Figure 2.10: Average MAPE versus the percentage phase unbalance.

operation. The performance can be improved if we add more D-PMUs to the laterals; even while the system still remains low-observable.

2.4.9 Examining the Uniqueness of Solution

As we discussed in Section 2.3.3, the exactness of the solution in a sparse recovery problem directly depends on its uniqueness. Of interest in this context is the tightness of the theoretical sufficient conditions to assure that the obtained DSSE solution is unique. Importantly, the sufficient conditions in [113] are satisfies only in 0.8% of the scenarios. However, the sufficient conditions that we developed in Theorem 1 are satisfied in 11.3% of the scenarios. Thus, the theoretical optimality conditions have significantly improved in our analysis; compared to the conventional sufficiency conditions for a generic sparse recovery problem without side information. Of course, since these theoretical results are based on

sufficient conditions, in theory, the rest of the scenarios may or may not achieve the unique solution. Nevertheless, our numerical results confirm that the mean square error in state estimation is less than 10^{-8} in 96.4% of the scenarios.

Chapter 3

Event-Triggered Distribution System State Estimation: Sparse Kalman Filtering with Reinforced Coupling

3.1 Introduction

3.1.1 Motivations

With the transition of power distribution systems from passive non-changing networks to active time-varying types, there is a growing need to develop real-time continuous monitoring and control systems at the distribution level [105]. In practice, this has to be done by using the combination of the traditional legacy meters in SCADA systems as well

as the advanced sensors that have emerged recently, such as D-PMUs, a.k.a., micro-PMUs [82]. However, this is a challenging task, in particular when it comes to DSSE problem. In this regard, the following issues need to be addressed:

Low-Observability

Unlike the transmission level, power distribution systems are not fully-observable, i.e., there is not enough measurements redundancy to obtain the state variables [20]. Due to the expanded size of the network and limitations on the communication infrastructure, the grid operator cannot install sensors at every location in the power distribution system to maintain full-observability. Therefore, the proposed DSSE method must be capable to address the DSSE problem under the low-observability setting, and from the limited available measurements.

Dynamic and Time-Varying System

Power distribution systems are becoming increasingly dynamic and time-varying due to the high penetration of renewable energy resources with uncertain and intermittent generation. A typical sudden change in the load or in the output power of renewable generators can cause a sudden deviation in the state variables in the system. Thus, static DSSE methods cannot capture and track these frequent changes in the system; because of their inherent assumption on steady-state operation of the power distribution systems [56]. This is particularly the case in the DSSE methods that must rely on using pseudo-measurements, alongside other measurements in order to maintain full-observability [116]. To address this issue, dynamic DSSE methods are proposed as an alternative for monitoring

and control of the power distribution systems that are subject to different dynamics. A major aspect which distinguishes dynamic state estimation methods from each other is the selection of state variables [116]. In the dynamic setting, state variables are two type. The First type are the system states such as machine internal states, i.e., rotor speed, inertia, etc. The second type are the algebraic states such as voltage phasors and current phasors. At the distribution level, what matters for the gird monitoring is mainly the voltage and current as opposed to the transmission level where system states are also important, especially for dynamic stability analysis. A particular dynamic state estimation method which only includes the algebraic state variables is forecasting-aided state estimation [40]. Meanwhile, the majority of the works in the limited literature of forecasting-aided DSSE considers a pseudo-dynamic setting for the DSSE problem wherein they typically take a common but simplifying assumption that the state transition matrix is an identity matrix; e.g., see [100, 95, 27, 59].

While this simplification might work in a fully-observable network, it may further aggravate the challenges in a network with low-observability conditions. Therefore, selection of state-space model under low-observability condition is crucial; and improvement in the accuracy of state-space model is desired. To the best of our knowledge, no previous work in the literature has made an effort to obtain an accurate state-space model for the dynamic state estimation neither in the transmission level nor the distribution level.

Computation Burden

Communication issue and limited bandwidth are another hinders in the distribution feeders against achieving a reliable monitoring system. While the high reporting rate

of D-PMUs can potentially give the utility operator a unique opportunity to enhance situational awareness, processing the heavy streams of D-PMU data remains a challenge by itself. Interestingly, in practice, a major portion of the D-PMU data does not carry much useful information [7, 55]. Therefore, a well-designed DSSE mechanism that focuses on extracting the most informative pieces of measurements under the inherent low-observability in the power distribution system can help also reduce the computational and communication burden. Moreover, the state of the power distribution system may not necessarily change by small fraction of time, repetitively. An alternative solution that is proposed in the literature for such sensor networks which deal with the communication issue is the event-triggered state estimation [120]. A properly designed event-trigger mechanism only sends the measurements to the estimation center in case that a major event is detected across the network. This alone helps the network to avoid unnecessary transmission of data packets when no remarkable deviations in the operating points of the system are occurred.

3.1.2 Literature Review

Table 3.1 shows an overview summary of the related literature, in comparison with the proposed method. So far, the most common approach to address low-observability has been to increase the observability by using pseudo-measurements, whether in static DSSE methods [35, 95], or in dynamic DSSE methods [100, 95, 27]. Pseudo-measurements are typically generated from historical or proxy data and by using methods such as artificial neural networks [79] or Gaussian mixture model [98]. There have been efforts also to enhance the accuracy of pseudo-measurements, such as by considering load uncertainty [87].

Table 3.1: Summary overview and comparison with the existing DSSE methods

Method	Event-Triggered	Observability	Setting	State Transition Matrix	Physics-Based Sparsity Features
[100]	No	Full-Observability†	Dynamic	Time-Variant	Zero Injection Nodes
[95][27]	No	Full-Observability†	Dynamic	Identity	No
[59]	No	Full-Observability†	Dynamic	Identity	Temporal Correlation
[79][98]	No	Full-Observability†	Static	N/A	No
[13][43]	No	Low-Observability	Static	N/A	No
[33]	Yes	Full-Observability†	Dynamic	Identity	No
[9]	Yes	Low-Observability	Static	N/A	Sparsity in Voltage and Current
Proposed	Yes	Low-Observability	Dynamic	Time-Variant	Sparsity in Differential Power Flow Equations

†Full-observability is achieved either through sufficient measurements or by using pseudo-measurements.

A more recent approach to tackle low-observability is to use sparse recovery tools from signal processing to solve the DSSE problem without making it fully-observable [43, 13, 9, 8, 6]. However, all the existing methods in this line of work so far seek to solve the DSSE problem in a static setting, i.e., they do not use any dynamic model, such as a state transition matrix, in the formulation of the problem. Therefore, they cannot capture the dynamic nature of the modern power distribution systems.

Importantly, a method that is dedicated to address the DSSE problem, not only under low-observability conditions but also in a dynamic setting, is still missing. In [95], a dynamic DSSE method is proposed using Iterated Kalman Filter (IKF). In [27], a past-aware DSSE method is proposed which uses the estimation results from the previous time slots to correct the estimation for the current time slot through the use of Ensemble Kalman filter (EnKF). In [59], a forecasting-aided DSSE method is proposed by using a robust EnKF to increase redundancy in measurements. However, the dynamic DSSE methods in [95, 59, 27] do not address low-observability and they do not utilize the inherent sparsity in the DSSE problem.

Finally, although there are few studies that use event-triggered models in the filed of power system state estimation, they have mostly focused on wide area measurements in power transmission systems, such as in [68, 71]. Recently, a DSSE method based on event-triggered transmissions is proposed in [33]. However, the method in [33] does not address the low-observability issue, which is the primary challenge in this manuscript, and it also does not involve the learning mechanisms for the state space model; both of which are the foundation of the method that is proposed in this work.

Therefore, it is still an open problem to design a comprehensive DSSE method which can address all of the three-fold challenges mentioned in Section 3.1.1.

3.2 System Model and Assumptions

Consider a three-phase power distribution network that is represented by graph $\mathcal{G} := (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ denotes the set of three-phase nodes and $\mathcal{L}$ denotes the set of three-phase distribution lines. To avoid heavy notations, we skip using a specific index for the phases; however, all the phasors in our formulations are assumed to be in three-phase. We assume that the understudy network has a typical radial topology.

3.2.1 State Variables

Traditionally, it is very common in the power systems literature to model the power system dynamics using time-triggered state space models, in which the state variables in the system may change on fixed time intervals. However, in principle, one can model a dynamic system also by using event-triggered state space models, in which the state variables in the system may change due to the occurrences of events [66]. The advantage of considering the event-triggered paradigm is that we can run and solve the DSSE problem only when a major event is detected in the system, rather than doing so every single time that a measurement is obtained; while no significant change has happened in the state variables.

Here, an "event" is defined broadly. It can be a reliability event, such as a fault incident, or a benign event, such as a load switching, a capacitor bank switching, a sudden change in the generation level of a renewable generator. In all these cases, the event may

cause changes in the state variables in the system; which in turn triggers a new state estimation.

At each event k, let $\mathbf{V}_k$ denote the vector of voltage phasors at every node in set $\mathcal{N}$. Also, let $\mathbf{I}_k$ denote the vector of current phasors at every line in set $\mathcal{L}$. In this study, we represent all phasors in the rectangular coordinates. Thus, each phasor has two associated quantities, i.e., the real part denoted by $\Re(.)$, and the imaginary part denoted by $\Im(.)$. Similarly, let $\mathbf{S}_k$ denote the vector that contains the injected apparent power to all the nodes in $\mathcal{N}$ at event k.

Suppose $\mathbf{x}_k$ denotes the vector that contains $\mathbf{V}_k$, $\mathbf{I}_k$, and $\mathbf{S}_k$. Unlike most other formulations in the DSSE literature, we define the state variables in differential mode. We define $\Delta\mathbf{x}_k$ as the vector of differential state variables as follows:

$$\Delta\mathbf{x}_k := \mathbf{x}_k - \mathbf{x}_{k-1}. \tag{3.1}$$

Here, $\Delta\mathbf{x}_k$ denotes the vector of the changes in all the system state variables that are caused due to the occurrence of event k, in comparison with the status of the system at event $k-1$. The vector of state variables in our problem formulation contains all differential voltage phasors, all differential current phasors, and all differential apparent power injections:

$$\Delta\mathbf{x}_k := \begin{bmatrix} (\Delta\mathbf{V}_k)^\top & (\Delta\mathbf{I}_k)^\top & (\Delta\mathbf{S}_k)^\top \end{bmatrix}^\top. \tag{3.2}$$

3.2.2 Available Measurements and Network Observability

Since the state variables in our problem formulation are defined in differential mode, we need to know the initial values of state variables, such that we can add them to the estimated differential phasors to obtain the regular state variables after each event.

Hence, we assume that at $k = 0$, the initial values of the state variables are known. Subsequently, during the events $k = 1, \ldots, K$, we assume that low-observability condition holds, where the measurements are limited to the voltage and current synchrophasors that come from only a small number of D-PMUs on the power distribution feeder. We only rely on the measurements of D-PMUs, which have high reporting rates, because other types of sensor measurements, such as from legacy meters or pseudo-measurements, are inadequate to capture the sudden changes that happen in the state variables due to the events [100].

At each event k, let $\mathbf{y}_k$ denote the vector of voltage phasor measurements at the nodes in set $\mathcal{N}_m \subseteq \mathcal{N}$ and the current phasor measurements at the lines in set $\mathcal{L}_m \subseteq \mathcal{L}$. Similar to (3.1), we can define the vector of differential phasor measurements in the rectangular coordinates as:

$$\Delta \mathbf{y}_k := \mathbf{y}_k - \mathbf{y}_{k-1}. \tag{3.3}$$

Therefore, measurement vector $\Delta \mathbf{y}_k$ can be defined as:

$$\Delta \mathbf{y}_k := \begin{bmatrix} (\Delta \mathbf{V}_k^m)^\top & (\Delta \mathbf{I}_k^m)^\top \end{bmatrix}^\top, \tag{3.4}$$

where superscript m indicates the measurements.

As we can see in (3.4), we do not collect any power measurements in the proposed formulation of the DSSE problem. Yet we want to reinforce our DSSE problem formulation with physics-based couplings that exist between the differential state variables in (3.2), because of the advantages of doing so; which we will unmask throughout this work. Therefore, despite not having power measurements in (3.4), we have intentionally included the differential power phasors in the vector of state variables in (3.2), such that their values are estimated as a part of our proposed DSSE problem.

Based on the available measurements, we introduce measurement matrix $\mathbf{H}_k$ to relate the measurements to the state variables. In its basic form, matrix $\mathbf{H}_k$ has three block rows, denoted by $\mathbf{H}_k^1$, $\mathbf{H}_k^2$, and $\mathbf{H}_k^3$, as we will explain next. We will also discuss a fourth block, denoted by $\mathbf{H}_k^4$, in Section 3.2.5.

In the first set of equations, the differential voltage phasor measurements are mapped to the differential voltage phasors in the vector of state variables via an identity mapping:

$$\mathbf{H}_k^1 := \begin{bmatrix} \mathbf{U} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{3.5}$$

where $\mathbf{U}$ is the adequate identity block matrix.

In the second set of equations, the available differential current phasor measurements are mapped to the differential voltage phasors in the vector of state variables by applying the Kirchhoff's current law (KCL), as follows:

$$\Delta\mathbf{I}_k^m = \mathbf{Y}\Delta\mathbf{V}_k, \tag{3.6}$$

where $\mathbf{Y}$ is the admittance matrix. Accordingly, we have:

$$\mathbf{H}_k^2 := \begin{bmatrix} \mathbf{Y} & \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{3.7}$$

The equations in (3.6) only include a small subset of line segments, i.e., those in $\mathcal{L}_m$. For the rest of the line segments, i.e., those in $\mathcal{L}\backslash\mathcal{L}_m$, we do not have any direct measurements. Thus, we instead add the KCL equations associated with the line segments that are not equipped with D-PMUs as auxiliary equations to the DSSE problem formulation as follows:

$$\mathbf{0} = \mathbf{Y}\Delta\mathbf{V}_k - \Delta\mathbf{I}_k. \tag{3.8}$$

Accordingly, we have:

$$\mathbf{H}_k^3 := \begin{bmatrix} \mathbf{Y} & -\mathbf{U} & \mathbf{0} \end{bmatrix}. \tag{3.9}$$

3.2.3 Event-Triggered State-Space Model

In a static DSSE problem where time index is not a matter of fact, one can solve an optimization problem that leads to the solution of equation (3.19) to estimate the operating points of the system. However, the increased penetration of renewable energy resources at the the distribution level has changed the nature of power distribution systems to be dynamic and time varying. Since static DSSE methods work based on the assumption of steady-state model, they are inherently incapable of addressing the dynamic characteristics of the modern power distribution systems. To address this issue, we need to formulate the DSSE problem in a dynamic setting.

Dynamic state estimation has different variants based on the type of state variables, as well as the type of state-space model. Since in our model, our choice of state variables are the algebraic state variables, i.e., voltage and current phasors, and we look forward to modeling the state-space in a linear form, we use forecasting-aided state estimation as a sub-class of dynamic state estimation [116]. Let us start by representing the power system in differential mode based on its state space model as:

$$\mathbf{\Delta x_k} = \mathbf{A}_k \Delta \mathbf{x}_{k-1} + \mathbf{q}_k, \tag{3.10}$$

where $\mathbf{A}_k$ is the state transition matrix at event k and $\mathbf{q}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ is the zero-mean process noise with covariance $\mathbf{Q}_k$. Since our focus is on event-triggered state estimation, the DSSE problem is solved only if a major event occurs in the network. The events include

but are not limited to load switching, capacitor bank switching, a sudden change in the generation level of renewable energy resources, etc. The occurrences of major events are detected based on a pre-defined event detection criterion. In our proposed event-triggered DSSE problem formulation, measurements are sent to the control center and processed only once a major event is detected based on a pre-defined event detection criterion. In fact, there already exists a wide range of event detection methods that can be readily used to detect major events, such as from [45].

Importantly, these existing methods not only detect the occurrence of the event, but also identify the approximate location (node) of the event based on the phasor measurements from only a few D-PMUs, one at the substation, and one at the end of each lateral. In other words, they can approximate the location (node) of the event under the same low-observability assumptions that we consider in this work. Thus, for the rest of this work, we assume that the occurrence of the events and their approximate locations are known. We assume that only one event can happen at a time. The challenge is to formulate and solve the DSSE problem that is triggered by an event.

The study of events in this work is based on the findings in [118, 45] about event-detection in real-world D-PMU data. They have shown that, in practice, it is almost always the case that only one major event may happen at a time on a typical real-world distribution feeder. Accordingly, unless it is stated otherwise, we assume in this work that only one event happens at a time. However, for the special case where one major event and multiple smaller events happen at the same time, the proposed method can still perform reasonably well, as we will see in a case study in Section 3.5.6.

3.2.4 Sparsity in Voltage and Current Phasors

Consider an event that occurs on the feeder. Consider the path between the substation and the node where the event occurs. Let us denote such path by a tree $\mathcal{T} := (\mathcal{V}, \mathcal{E})$. Let us denote the rest of the network by $\mathcal{T}' := (\mathcal{V}', \mathcal{E}')$. Note that:

$$\mathcal{T} \cup \mathcal{T}' = \mathcal{G} \quad \text{and} \quad \mathcal{T} \cap \mathcal{T}' = \varnothing. \tag{3.11}$$

In previous Chapter, we presented a rigorous analysis on the sparsity features of the power distribution feeders upon the occurrence of an event. We showed that, once a major event happens at the distribution feeder, the vector of differential voltage phasors and the vector of differential current phasors become sparse. Only the differential voltage phasors at the nodes in $\mathcal{V}$, and only the differential current phasors at the line segments in $\mathcal{E}$ would be non-zero. The differential current phasors for all the line segments outside tree $\mathcal{T}$, i.e., those in $\mathcal{E}'$ would be zero. As for the nodes in $\mathcal{V}'$, group sparsity holds among the differential voltage phasors, as there would be groups of nodes whose differential voltage phasors would be either all (approximately) zero or all non-zero. That is, in each group, if the differential voltage phasor is zero for one node, then it would be zero for the rest of nodes that are on the same group. For more details please refer to [9].

3.2.5 Sparsity in Power Injections

Let us assume that all the load and generation components at each node on the power distribution feeder are collectively modeled as a constant power component. This is a common model, e.g., see [21]. An important implication of this assumption is that, once an event occurs, the differential power injection is zero at all of the nodes across the power

distribution feeder, except at the node where the event occurs. Since the event detection method gives us the location of the event, c.f. [45], we accordingly know at which nodes the differential power injection phasors are zero. By incorporating this information with the DSSE problem through the power flow equations, we can derive an additional set of equations to enhance the observability of the power distribution system.

Throughout this work, we call the above additional set of differential zero-injection equations as virtual measurements, as we do not have any direct measurement for the power in our setting. However, the outcome of our event detection method gives us the value of differential power injection at every node, but the node of event which should be estimated. While these additional equations do not necessarily make the undetermined system of equations in the DSSE problem full-rank, they reinforce the DSSE problem formulation with physics-based coupling among the state variables, which helps us with recovering the state variables at each event.

The above idea can be materialized by incorporating the power flow equations into our analysis. Since we defined the state variables in differential mode, we need to modify the standard power flow equations accordingly. After that, we need to linearize the resulting differential power flow equations. Interestingly, the linearization of the power flow equations in differential mode is justified mathematically beyond the typical linearization of the standard power flow equations. Let $S_{k,i}$ denote the injected power to node i at event k. For each event k, we can rewrite the injected power as the sum of its value at the previous event, i.e., $k-1$, and the differential power phasor at event k:

$$S_{k,i} = (P_{t-1,i} + \Delta P_{k,i}) + j(Q_{k-1,i} + \Delta Q_{k,i}). \tag{3.12}$$

Let $v_{k,i}$ denote the voltage phasor of node i at event k. We can rewrite (3.12) based on the voltage phasors as:

$$\begin{aligned} S_{k,i} = \Big[\Re\{v_{k-1,i} + \Delta v_{k,i}\} + j\Im\{v_{k-1,i} + \Delta v_{k,i}\} \Big] \times \\ \Big[\textstyle\sum_{j\in\mathcal{N}} (R_{ij} - jX_{ij}) \left(\Re\{v_{k-1,j} + \Delta v_{k,j}\}\right. \\ \left. - j\Im\{v_{k-1,j} + \Delta v_{k,j}\}\right) \Big] \end{aligned} \tag{3.13}$$

where $\Re\{.\}$ denotes the real part, $\Im\{.\}$ denotes the imaginary part, R_{ij} and X_{ij} denote the resistance and admittance of the line segment between nodes i and j, respectively. From (3.12) and (3.13), and after reordering the terms, we can obtain:

$$\begin{aligned} \Delta P_{k,i} \approx \sum_{j\in\mathcal{N}} \Big[& \Re\{\Delta v_{k,i}\} \left(R_{ij}\Re\{v_{k-1,j}\} - X_{ij}\Im\{v_{k-1,j}\}\right) \\ & + \Im\{\Delta v_{k,i}\} \left(R_{ij}\Im\{v_{k-1,j}\} + X_{ij}\Re\{v_{k-1,j}\}\right) \\ & + \Re\{\Delta v_{k,j}\} \left(R_{ij}\Re\{v_{k-1,i}\} + X_{ij}\Im\{v_{k-1,i}\}\right) \\ & + \Im\{\Delta v_{k,j}\} \left(R_{ij}\Im\{v_{k-1,i}\} - X_{ij}\Re\{v_{k-1,i}\}\right) \Big], \end{aligned} \tag{3.14}$$

$$\begin{aligned} \Delta Q_{k,i} \approx -\sum_{j\in\mathcal{N}} \Big[& \Re(\Delta v_{k,i}) \left(R_{ij}\Im\{v_{k-1,j}\} + X_{ij}\Re\{v_{k-1,j}\}\right) \\ & + \Im\{\Delta v_{k,i}\} \left(-R_{ij}\Re\{v_{k-1,j}\} + X_{ij}\Im\{v_{k-1,j}\}\right) \\ & + \Re\{\Delta v_{k,j}\} \left(-R_{ij}\Im\{v_{k-1,i}\} + X_{ij}\Re\{v_{k-1,i}\}\right) \\ & + \Im\{\Delta v_{k,j}\} \left(R_{ij}\Re\{v_{k-1,i}\} + X_{ij}\Im\{v_{k-1,i}\}\right) \Big]. \end{aligned} \tag{3.15}$$

The approximations in (3.14) and (3.15) are due to the fact that we discarded the terms that are the products of two differential voltage phasors, because such products are very small. Importantly, this approach is different from the typical Taylor expansion that is used in the literature to linearize the power flow equations. In (3.14) and (3.15), the differential

power injections and differential voltage phasors for the current event k are unknowns, while the values of voltage phasors for the previous event, i.e., $k-1$ are known.

The resulting linearized differential power flow equations are obtained in an abstract form as:

$$\mathbf{0} = \mathbf{J}_k \Delta \mathbf{V}_k - \Delta \mathbf{S}_k, \tag{3.16}$$

where $\mathbf{J}_k$ is the Jacobian matrix.

From (3.16), we can now obtain the last block row of the measurement matrix which we denote by $\mathbf{H}_k^4$, as follows:

$$\mathbf{H}_k^4 := \begin{bmatrix} \mathbf{J}_k & \mathbf{0} & -\mathbf{U} \end{bmatrix}. \tag{3.17}$$

Note, we do not use any power measurement or any power pseudo-measurement, yet we do take advantage of the power flow equations in differential mode by including the fourth block row in the measurement matrix. Thus, we catch the physics-based couplings that exist among the differential voltage phasors, differential current phasors, and the differential power injections in an event-triggered DSSE problem.

3.2.6 Event-Triggered DSSE Formulation

For notational simplicity, we introduce $\Delta \mathbf{z}_k$ as a new vector for all measurements in differential mode. It is an extension of vector $\Delta \mathbf{y}_k$ in (3.4), where we also add additional vector of zeros corresponding to the auxiliary equations in (3.8) and the virtual measurements in (3.16):

$$\Delta \mathbf{z}_k = [\, \Delta \mathbf{y}_k \quad \mathbf{0} \,]. \tag{3.18}$$

Accordingly, we can obtain the following ultimate relationship between the new vector of differential measurements and the vector of differential state variables:

$$\mathbf{\Delta z_k} = \mathbf{H}_k \, \Delta \mathbf{x}_k + \mathbf{r}_k, \tag{3.19}$$

where matrix $\mathbf{H}_k$ in (3.19) is constructed by putting together matrices $\mathbf{H}_k^1$, $\mathbf{H}_k^2$, $\mathbf{H}_k^3$, and $\mathbf{H}_k^4$. The added term $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ is the measurement noise that is assumed to be zero-mean Gaussian with covariance matrix $\mathbf{R}_k$. The measurement noises at different D-PMUs are assumed to be mutually independent.

Together, the relationships in (3.10) and (3.19) provide the core formulation for the event-triggered DSSE problem. Given the fact that the state estimation problem is dynamic, the state estimation process is done in the following two stages.

First Stage: In this stage, at the beginning of each sequence, the state variables at the current sequence are predicted based on their values in the previous sequence and by using the state-space model in (3.10). The outcome of event detection is directly taken into consideration in (3.10). An important aspect in the analysis in this first stage is to have an accurate state transition matrix as part of the event-triggered state space model in (3.10). In this work, we use Elastic Net Regression to learn the state transition matrix. We will discuss this subject in Section 3.3.

Second Stage: In this stage, the measurement model in (3.19) is used to correct the prediction of the state variables. In this work, we use a novel sparse Kalman filter/smoother, given that our focus is on sparse recovery in order to address the low-observability in the power distribution system. We will discuss this subject in Section 3.4.

3.3 Offline Learning of Transition Matrix via Elastic Net Regression

As we mentioned in Section 3.2.6, the accuracy in predicting the current state variables based on the event-triggered state space model in (3.10) depends on the accuracy of the state transition matrix $\mathbf{A}_k$. However, obtaining matrix $\mathbf{A}_k$ is a challenging task. In fact, it is common in the literature to assume that the state transition matrix in the DSSE problem is always an identity matrix [95]. While there have been some recent efforts, such as in [65], to estimate the time-varying state transition matrix, no prior study has aimed to obtain such matrix under the low-observability conditions. Also, no prior method is designed for event-triggered DSSE formulations.

3.3.1 Event-Triggered State Transition Matrix

Suppose event k is detected and its location is identified as bus s_k. Suppose the location of the previous event, i.e., event $k-1$, is bus s_{k-1}. Our goal is to obtain

$$\mathbf{A}_k = \mathbf{B}(s_{k-1}, s_k) \tag{3.20}$$

for the transition of the system in differential mode from an event at bus s_{k-1} to a subsequent event at bus s_k. In order to explain the role of matrix $\mathbf{B}$, let us first consider an example. Suppose $K = 10$, and the location of the events are identified as buses 5, 8, 15, 1, 7, 10, 4, 8, 15, 6, respectively. Notice that, at event $k = 3$, we have $s_{k-1} = 8$ and $s_k = 15$. Similarly, at event $k = 9$, we have $s_{k-1} = 8$ and $s_k = 15$. Therefore, we set

$$\mathbf{A}_3 = \mathbf{A}_9 = \mathbf{B}(8, 15). \tag{3.21}$$

Therefore, matrix $\mathbf{B}(n,m)$ is defined as the event-triggered state-transition matrix that captures the changes in the state variables in the system for any case where the previous event occurs at bus n and the current event occurs at bus m.

In total, we have $(|\mathcal{N}|-1)\times(|\mathcal{N}|-1)$ possible transitions from one event location to another event location. Therefore, there are exactly $(|\mathcal{N}|-1)\times(|\mathcal{N}|-1)$ matrices $\mathbf{B}(n,m)$.

For any two buses n and m in set $\mathcal{N}\backslash\{1\}$, we can obtain matrix $\mathbf{B}(n,m)$ by examining the set of all historical events of index κ for which we have $s_{\kappa-1}=n$ and $s_\kappa=m$. We can learn the event-triggered state transition matrix $\mathbf{B}(n,m)$ by solving the following optimization problem:

$$\mathbf{B}(n,m)=\arg\min_{\mathbf{B}}\sum_{\kappa\in\mathcal{K}}||\Delta\mathbf{x}_\kappa-\mathbf{B}\Delta\mathbf{x}_{\kappa-1}||_2^2 \tag{3.22}$$

where

$$\mathcal{K}=\{\kappa\mid s_{\kappa-1}=n,s_\kappa=m\}\,. \tag{3.23}$$

The above training is done offline by analyzing the events in the historical data. Once we obtain $\mathbf{B}(n,m)$ for all the $(|\mathcal{N}|-1)\times(|\mathcal{N}|-1)$ possible event-triggered transitions, we can obtain $\mathbf{A}_k$ for any event k by using the relation in (3.20).

3.3.2 Elastic Net Regression

While the formulation in (3.22) is conceptually valid, we may face numerical issues in solving this optimization problem. Such numerical issues are due to the inherent sparsity in matrix $\mathbf{B}$, which is due to the same reasons that we mentioned in Section 4.1.3. In fact, since each node in a radial network topology is connected to very few nodes, each row of matrix $\mathbf{B}$ is highly sparse. Therefore, we propose to formulate the optimization problem in

(3.22) as a elastic net regression [115]. The new formulation is obtained as:

$$\mathbf{B}(n,m) = \arg\min_{\mathbf{B}} \sum_{\kappa \in \mathcal{K}} ||\Delta \mathbf{x}_{\kappa} - \mathbf{B}\Delta \mathbf{x}_{\kappa-1}||_2^2 + \lambda_2 ||\mathbf{B}||_F^2 + \lambda_1 ||\mathbf{B}||_{1,1}, \tag{3.24}$$

where $||\cdot||_F$ denotes the Frobenius norm and $||\cdot||_{1,1}$ denotes the $l_{1,1}$ matrix norm. Parameters λ_1 and λ_2 are weight factors.

The reason for using the above elastic net regression in this study is the fact that, each row of matrix $\mathbf{B}$ is sparse due to the radial topology of the distribution feeder, yet the non-zero entries are correlated because of the spatial-temporal correlation of the state variables in a power distribution feeder [13]. Elastic net regression has the advantage to make a trade-off between these two properties by adjusting the penalty parameters λ_1 and λ_2. Similar to any other regression model, the values of these weight parameters are determined based on the values of data in practice. The larger λ_1 is, the more sparse and less correlated the entries of matrix $\mathbf{B}$ would be.

The solution for the elastic net regression is a matrix, whose element in row i and column j is obtained as:

$$b[ij] = \frac{\{|b_{\text{LS}}[ij]| - \lambda_1/2\}^+}{1+\lambda_2} \text{sign}\{b_{\text{LS}}[ij]\} \tag{3.25}$$

where $\{\cdot\}^+ = \max\{\cdot, 0\}$ and $b_{\text{LS}}[ij]$ is the element in row i and column j of the following matrix:

$$\sum_{\kappa \in \mathcal{K}} \Delta \mathbf{x}_{k-1} (\Delta \mathbf{x}_k)^\top. \tag{3.26}$$

In a feeder with a large number of nodes, it may not be possible to capture all the $(|\mathcal{N}|-1) \times (|\mathcal{N}|-1)$ transitions in the system. On the other hand, the distribution network has got a radial topology in which there exists a high correlation among the voltage

phasors of nodes and the current phasors of lines that are in the same proximity. Therefore, when a same event happens in different nodes that are on the same laterals, it would cause a similar general impact on the operating points of the system. Accordingly, one solution is to group multiple nodes into one zone, such that instead of finding the transition from the nodal location of one event to the nodal location of the next event, we seek to find the transition from the zonal location of one event to the zonal location of the next event.

3.4 Event-Triggered DSSE Problem Formulation as Sparse Signal Recovery and Its Solution

After learning the transition matrix, we can now solve the event-triggered DSSE problem. In this section, we provide the problem formulation and also two different solution methods.

3.4.1 Problem Formulation

At each event k, our goal is to estimate the most recent differential state variables, i.e., $\Delta\mathbf{x}_k$, as well as to refine and update the estimation of the previous differential state variables, i.e., $\Delta\mathbf{x}_1, \ldots, \Delta\mathbf{x}_{k-1}$. In this regard, suppose we stack up the vectors of differential state variables $\Delta\mathbf{x}_1, \ldots, \Delta\mathbf{x}_k$ and denote the resulting vector by $\Delta\mathbf{x}_{1:k}$. Similarly, suppose we stack up the vectors of differential measurements $\Delta\mathbf{z}_1, \ldots, \Delta\mathbf{z}_k$ and denote the resulting vector by $\Delta\mathbf{z}_{1:k}$. From (3.10) and (3.19), and given the fact that the vector of the differential state variables is a sparse vector, we can estimate $\Delta\mathbf{x}_{1:k}$ by solving the event-triggered DSSE

problem which is formulated as the following generalized group Lasso problem:

$$\begin{aligned}\underset{\Delta\mathbf{x}_{1:k}}{\text{minimize}}\ & \frac{1}{2}\sum_{\kappa=1}^{k}||\mathbf{R}_{\kappa}^{-1/2}\left(\Delta\mathbf{z}_{\kappa}-\mathbf{H}_{\kappa}\Delta\mathbf{x}_{\kappa}\right)||_2^2 \\ &+\frac{1}{2}\sum_{\kappa=2}^{k}||\mathbf{Q}_{\kappa}^{-1/2}\left(\Delta\mathbf{x}_{\kappa}-\mathbf{A}_{\kappa}\Delta\mathbf{x}_{\kappa-1}\right)||_2^2 \\ &+\frac{1}{2}||\mathbf{G}_1^{-1/2}\left(\Delta\mathbf{x}_1-\mathbf{m}_1\right)||_2^2 \\ &+\lambda\sum_{\kappa=1}^{k}\sum_{p=1}^{P}w_{\kappa,p}\|\Delta\mathbf{x}_{\kappa,p}\|_1.\end{aligned} \tag{3.27}$$

Here, $\mathbf{Q}_{\kappa}$ and $\mathbf{R}_{\kappa}$ are the covariance matrices that are used in (3.10) and (3.19), respectively; $\mathbf{m}_1$ and $\mathbf{G}_1$ are the mean and covariance of the initial differential state variable, which are assumed to be known; and λ is the sparsity regularization parameter. Subscript p is the index for each partition of the differential state variables with respect to the concept of group sparsity that we discussed in Section 4.1.3; and $w_{\kappa,p}$ is the weight associated with partition p at event κ. More details on selecting the partitions and their weights can be found in [9].

For notation simplicity, we can further rewrite the optimization problem in (3.27) in the following more compact form:

$$\begin{aligned}\underset{\Delta\mathbf{x}_{1:k}}{\text{minimize}}\ & \frac{1}{2}||\mathbf{R}^{-1/2}\left(\Delta\mathbf{z}_{1:k}-\mathbf{H}\Delta\mathbf{x}_{1:k}\right)||_2^2 \\ &+\frac{1}{2}||\mathbf{Q}^{-1/2}\left(\mathbf{\Phi}\Delta\mathbf{x}_{1:k}-\mathbf{m}\right)||_2^2 \\ &+\lambda\|\mathbf{W}\Delta\mathbf{x}_{1:k}\|_1\end{aligned} \tag{3.28}$$

where matrix $\mathbf{R}$ is the block-diagonal representation of matrices $\mathbf{R}_1,\ldots,\mathbf{R}_k$, matrix $\mathbf{H}$ is the block-diagonal representation of matrices $\mathbf{H}_1,\ldots,\mathbf{H}_k$, matrix $\mathbf{Q}$ is the block-diagonal representation of matrices $\mathbf{G}_1,\mathbf{Q}_2,\ldots,\mathbf{Q}_k$, matrix $\mathbf{\Phi}$ is the block-diagonal representation of matrices $\mathbf{A}_2,\ldots,\mathbf{A}_k$, matrix $\mathbf{W}$ is the adequate block-diagonal representation of the weights

in (3.27), and vector $\mathbf{m}$ is obtained by stacking up vector $\mathbf{m}_1$ following by $k-1$ vectors denoted by $\mathbf{0}$, which are zero vectors of the same size as $\mathbf{m}_1$. Next, we discuss two different approaches to solve the optimization problem in (3.27).

3.4.2 Approach 1: Batch ADMM Solution

A common approach to solve a generalized group Lasso optimization problem is to use the ADMM algorithm [22]. We can obtain the augmented Lagrangian function for the problem in (3.28) under constraint $\mathbf{W}\Delta\mathbf{x}_{1:k} - \boldsymbol{\mu} = 0$ as:

$$\begin{aligned}\mathcal{L}_\rho(\Delta\mathbf{x}_{1:k}, \boldsymbol{\mu}, \mathbf{u}) = &\frac{1}{2}||\mathbf{R}^{-1/2}\left(\Delta\mathbf{z}_{1:k} - \mathbf{H}\Delta\mathbf{x}_{1:k}\right)||_2^2 \\ &+ \frac{1}{2}||\mathbf{Q}^{-1/2}\left(\boldsymbol{\Phi}\Delta\mathbf{x}_{1:k} - \mathbf{m}\right)||_2^2 \\ &+ \lambda\|\boldsymbol{\mu}\|_1 + \langle \mathbf{u}, \mathbf{W}\Delta\mathbf{x}_{1:k} - \boldsymbol{\mu}\rangle \\ &+ \frac{\rho}{2}\|\mathbf{W}\Delta\mathbf{x}_{1:k} - \boldsymbol{\mu}\|_2^2\end{aligned} \tag{3.29}$$

where $\mathbf{u}$ is the vector of dual variables and ρ is a regularization parameter. For each set of the variables, ADMM alternatively solves the following minimization problems and obtains the optimal values for the $(\tau+1)$-th iteration of alternation as:

$$\begin{aligned}\Delta\mathbf{x}_{1:k}^{(\tau+1)} = \underset{\Delta\mathbf{x}_{1:k}}{\arg\min}\ &\frac{1}{2}||\mathbf{R}^{-1/2}\left(\Delta\mathbf{z}_{1:k} - \mathbf{H}\Delta\mathbf{x}_{1:k}\right)||_2^2 \\ &+ \frac{1}{2}||\mathbf{Q}^{-1/2}\left(\boldsymbol{\Phi}\Delta\mathbf{x}_{1:k} - \mathbf{m}\right)||_2^2 \\ &+ \langle \mathbf{u}^\tau, \mathbf{W}\Delta\mathbf{x}_{1:k} - \boldsymbol{\mu}^\tau\rangle \\ &+ \frac{\rho}{2}\|\mathbf{W}\Delta\mathbf{x}_{1:k} - \boldsymbol{\mu}^\tau\|_2^2,\end{aligned} \tag{3.30}$$

$$\begin{aligned}\boldsymbol{\mu}^{(\tau+1)} = \underset{\boldsymbol{\mu}}{\arg\min}\ &\lambda\|\boldsymbol{\mu}\|_1 + \langle \mathbf{u}^\tau, \mathbf{W}\Delta\mathbf{x}_{1:k}^{(\tau+1)} - \boldsymbol{\mu}\rangle \\ &+ \frac{\rho}{2}\|\mathbf{W}\Delta\mathbf{x}_{1:k}^{(\tau+1)} - \boldsymbol{\mu}\|_2^2,\end{aligned} \tag{3.31}$$

$$\mathbf{u}^{(\tau+1)} = \mathbf{u}^{\tau} + \rho \left(\mathbf{W}\Delta\mathbf{x}_{1:k}^{(\tau+1)} - \boldsymbol{\mu}^{(\tau+1)} \right). \tag{3.32}$$

By taking the derivative of the least square problem in (3.30), the optimal solution of $\Delta\mathbf{x}_{1:k}$ in iteration $(\tau+1)$ of the ADMM algorithm takes the following closed from:

$$\begin{aligned} \Delta\mathbf{x}_{1:k}^{(\tau+1)} = &[\mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{H} + \boldsymbol{\Phi}^{\top}\mathbf{Q}^{-1}\boldsymbol{\Phi} + \rho\mathbf{W}^{\top}\mathbf{W}]^{-1} \\ &[\mathbf{H}^{\top}\mathbf{R}^{-1}\Delta\mathbf{z}_{1:k} + \boldsymbol{\Phi}^{\top}\mathbf{Q}^{-1}\mathbf{m} + \rho\mathbf{W}^{\top}\boldsymbol{\mu}^{k} + \mathbf{W}^{\top}\mathbf{u}^{k}]. \end{aligned} \tag{3.33}$$

We can also solve the problem in (3.31) as:

$$\boldsymbol{\mu}^{(\tau+1)} = \mathcal{S}_{\lambda/\rho} \left(\mathbf{W}\Delta\mathbf{x}_{1:k}^{(\tau+1)} + \frac{\mathbf{u}^{\tau}}{\rho} \right), \tag{3.34}$$

where $\mathcal{S}(.)$ is the soft-thresholding operator.

3.4.3 Approach 2: Kalman Filter and Smoother

Although the ADMM algorithm is exact and effective, it can be computationally expensive; because it requires calculating products and inverses for large matrices in (3.33). Hence, in this section, we propose an alternative approach to solve (3.28).

Suppose we remove the last line in (3.28), i.e., the line with the ℓ_1-norm. What would be left in (3.28) would take the form of the optimization problem which can be solved by a conventional Kalman filter/smoother [19]. This is of interest due to the computational efficiency in Kalman filtering.

Based on the above observation, we propose to first make changes that we mentioned in the previous paragraph in (3.28), and obtain the corresponding solution for $\Delta\mathbf{x}_k$ by using Kalman filter. Next, to take into account the impact of sparsity, we project the obtained solution for $\Delta\mathbf{x}_k$ into a sparse domain. Finally, we use a conventional Kalman

smoother to refine the previous estimation for $\Delta\mathbf{x}_{1:k-1}$ that is already obtained at event $k-1$ based on the projection of solution for $\Delta\mathbf{x}_k$.

These steps will provide us with an approximate solution for $\Delta\mathbf{x}_{1:k}$, i.e., the optimal solution for (3.28). As we will discuss in next Section, the difference between the approximate solution in Approach 2 and the exact solution in Approach 1 is negligible.

Kalman Filtering

At each event $k > 1$, we can use the state space model in (3.10) to conduct a preliminary one-step ahead prediction of the state variables as follows:

$$\Delta\mathbf{x}_{k|k-1} = \mathbf{A}_k \Delta\mathbf{x}_{k-1}. \tag{3.35}$$

The covariance matrix for $\Delta\mathbf{x}_{k|k-1}$ is obtained as:

$$\mathbf{G}_{k|k-1} = \mathbf{A}_k \mathbf{G}_{k-1} \mathbf{A}_k^\top + \mathbf{Q}_k. \tag{3.36}$$

From (3.19), we can similarly make a one-step ahead prediction of the differential measurement as follows:

$$\Delta\mathbf{z}_{k|k-1} = \mathbf{H}_k \Delta\mathbf{x}_{k|k-1}. \tag{3.37}$$

We can obtain the measurement residual as $\Delta\mathbf{z}_k - \Delta\mathbf{z}_{k|k-1}$. Therefore, we can correct the preliminary one-step ahead prediction in (3.35) by applying the Kalman filter as follows:

$$\Delta\mathbf{x}_{k|k} = \Delta\mathbf{x}_{k|k-1} + \mathbf{K}_k \left(\Delta\mathbf{z}_k - \Delta\mathbf{z}_{k|k-1}\right), \tag{3.38}$$

where $\mathbf{K}_k$ is the gain of the Kalman filter:

$$\mathbf{K}_k = \mathbf{G}_{k|k-1} \mathbf{H}_k^\top \mathbf{S}_k^{-1}, \tag{3.39}$$

and $\mathbf{S}_k$ is the covariance of $\Delta\mathbf{z}_{k|k-1}$ as

$$\mathbf{S}_k = \mathbf{H}_k\mathbf{G}_{k|k-1}\mathbf{H}_k^\top + \mathbf{R}_k. \tag{3.40}$$

We can obtain the covariance matrix of $\Delta\mathbf{x}_{k|k}$ as:

$$\mathbf{G}_{k|k} = \mathbf{G}_{k|k-1} - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^\top. \tag{3.41}$$

Note that, the process to obtain $\Delta\mathbf{x}_{k|k}$ does not take into consideration the inherent sparsity in the solution.

Incorporating Sparsity

In the next step, we project the obtained estimation from Kalman filter, i.e., $\Delta\mathbf{x}_{k|k}$, to the sparse domain with respect to its corresponding weight vector $\mathbf{W}_k$ by solving the following optimization problem:

$$\underset{\Delta\mathbf{x}_k}{\text{minimize}}\, \frac{1}{2}||\Delta\mathbf{x}_k - \Delta\mathbf{x}_{k|k}||_2^2 + \lambda\|\mathbf{W}_k^\top\Delta\mathbf{x}_k\|_1. \tag{3.42}$$

Here we obtain $\Delta\mathbf{x}_k$ such that it is as close as possible to $\Delta\mathbf{x}_{k|k}$ while it meets the sparsity constraints.

Since problem (3.42) does not have a closed-form solution, we cannot obtain a closed-form expression for the covariance of its solution. Thus, we assume that the covariance of $\Delta\mathbf{x}_k$ is $\mathbf{G}_k = \mathbf{G}_{k|k}$.

Backward Smoothing

After obtaining a computationally efficient sparse solution for $\Delta\mathbf{x}_k$, we successively apply the Kalman smoother to all the previous state estimation results in order to update

$\Delta\mathbf{x}_{1:k-1}$. First, we replace $\Delta\mathbf{x}_{k-1}$ with

$$\Delta\mathbf{x}_{k-1} + \mathbf{F}_{k-1}\left(\Delta\mathbf{x}_k - \Delta\mathbf{x}_{k|k-1}\right), \tag{3.43}$$

where $\mathbf{F}_{k-1}$ is the gain of the Kalman smoother:

$$\mathbf{F}_{k-1} = \mathbf{G}_{k-1}\mathbf{A}_k^\top \mathbf{G}_{k|k-1}^{-1}. \tag{3.44}$$

Finally, we update $\mathbf{G}_{k-1}$ by replacing it with

$$\mathbf{G}_{k-1} + \mathbf{F}_{k-1}\left(\mathbf{G}_k - \mathbf{G}_{k|k-1}\right)\mathbf{F}_{k-1}^\top. \tag{3.45}$$

We then continue with applying the above backward smoother similarly to events $k-2, k-3, \ldots, 1$ to update vector $\Delta\mathbf{x}_{1:k-1}$.

3.5 Case Studies

This section includes several case studies based on both computer simulations as well as real-world measurements.

The computer simulations are done based on the IEEE 33-Bus power distribution test system [18], where we kept the default load profiles. MATPOWER in MATLAB R2018B is used to simulate the events. The magnitude of the events are set to be at most up to 50% of the associated load in the pre-event condition. The events include load switching, capacitor bank switching, and sudden major changes in the output of renewable generation units. Unless stated otherwise, we define five zones in the network. Accordingly, there are $5 \times 5 = 25$ possible choices for matrix $\mathbf{B}$ to model transitions from s_{k-1} to s_k. For each transition, we train an offline model to learn the corresponding transition matrix.

Only five D-PMUs are assumed to be available, at buses 9, 18, 22, 25, and 33. Each D-PMU reports nodal voltage and for the bus of installation line current phasors for the connected line once every 100 milliseconds. If the event detection method, as in [45], detects an event between any two consecutive readings of any D-PMU, then the proposed event-triggered DSSE is conducted. We assume that the initial value of every voltage phasor, current phasor, and complex apparent power is available to us from the legacy meters at $k = 0$. For the D-PMUs, the standard deviation in measurement error is $\sigma_V = 0.1\%$ for voltage and $\sigma_I = 1\%$ for current. The noise covariance matrices are obtained offline during the training of the event-triggered state transition matrices. We define 100 scenarios. In each scenario, we simulate $K = 1000$ events.

3.5.1 Performance Comparison

Performance comparison is done against the DSSE methods in [95, 57, 9]. The method in [95] is a dynamic DSSE based on conventional iterated Kalman filtering. It uses non-linear power flow equations. The method in [57] is a static DSSE based on conventional weighted least square analysis. Because the DSSE methods in [95] and [57] require full-observability, the initial measurements by legacy meters are used as pseudo-measurements for the rest of the events for these two methods.

We use mean absolute percentage error (MAPE) to evaluate the accuracy of estimation for differential state variables:

$$\text{MAPE} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\left|\frac{\Delta\hat{x}_{k,i} - \Delta x_{k,i}}{\Delta x_{k,i}}\right| \times 100\%, \tag{3.46}$$

where N is the length of the vector of state variables, and $\Delta x_{k,i}$ and $\Delta\hat{x}_{k,i}$ denote the true

Table 3.2: Comparing different DSSE methods in Estimating synchrophasors and differential synchrophsors

Method	Dynamic	Sparsity	MAPE $\Delta\mathbf{x}$	RMSE † $\mathbf{x}$
Proposed Method	✓	✓	3.82%	7.25
[95]	✓		21.72%	32.67
[9]		✓	4.84%	8.58
[57]			19.93%	27.78

† The RMSE values come with a 10^{-3} coefficient.

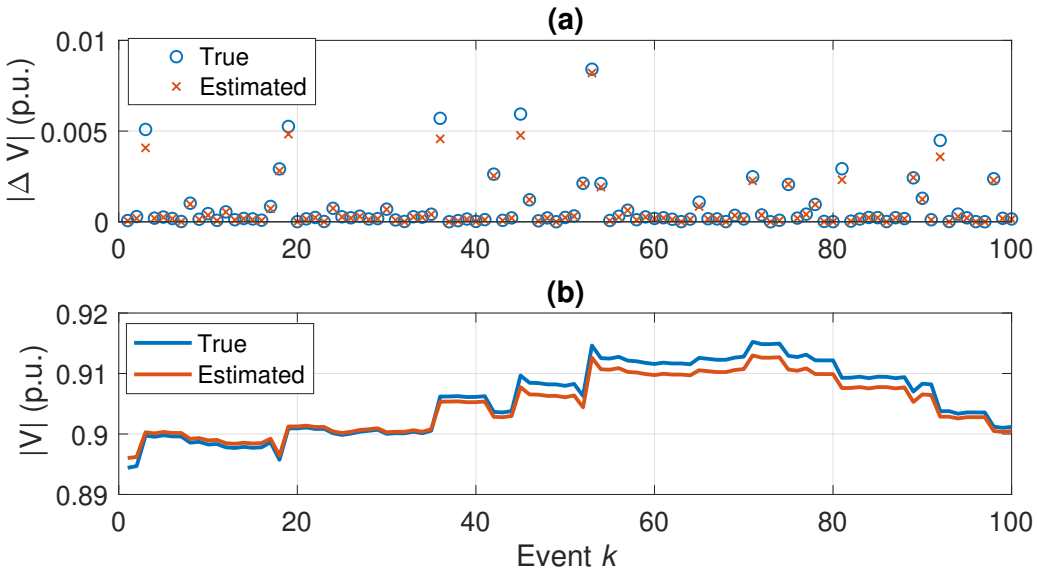

Figure 3.1: The estimated and actual values for the magnitude of voltage phasors at bus 29 for 100 events: (a) differential mode (b) regular mode.

value and the estimated value for the i-th differential state variable at event k, respectively. Also, we use root mean square error (RMSE) to evaluate the estimation of the state variables over the sequence of events:

$$\text{RMSE} = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{N}\sum_{i=1}^{N}|\hat{x}_{k,i} - x_{k,i}|^2}. \tag{3.47}$$

The statistical features of simulation results are summarized in Table 3.2. As we can see, the estimation accuracy for is better for the proposed method than the other three

methods, whether in terms of MAPE for the differential mode or in terms of RMSE for the regular mode. It is evident from Table 3.2 that our ability to integrate sparse recovery, dynamic state estimation, and virtual measurements has improved the DSSE performance across all variables.

An example for the outcome of the proposed method is shown in Fig. 3.1, where the results are at bus 29, for one random scenario, and across 100 consecutive events. The magnitude of the true and the estimated voltage phasors in differential mode are shown in Fig. 3.1(a). Notice the sparsity in the results, i.e., the fact that the differential voltage is non-zero at only a small subset of the events. Furthermore, the proposed method can identify the sparsity almost perfectly; and overall estimate the differential voltage phasors with a high accuracy. The magnitude of the true and the estimated voltage phasors in regular mode are shown in Fig. 3.1(b). We can see that the proposed method can reasonably follow the state variables.

3.5.2 Performance Comparison: IEEE 123-bus system

To show the scalability of the proposed method, we also test it over the IEEE 123-bus test system [2] which has single phase, two phase, and unbalanced three phase loads with Wye connection. We assume there are 19 D-PMUs available at nodes 6, 11, 16, 18, 32, 39, 47, 56, 59, 66, 71, 75, 85, 86, 96, 101, 114, 250, and 450 which measures the voltage and current at the location of installation as shown in Fig. 3.2. The results for this test network are shown in Table 3.3. As we can see, the performance of the proposed method under the unbalanced operation and for the larger test system is still quite well and it outperforms the conventional methods in the literature. It worth mentioning that

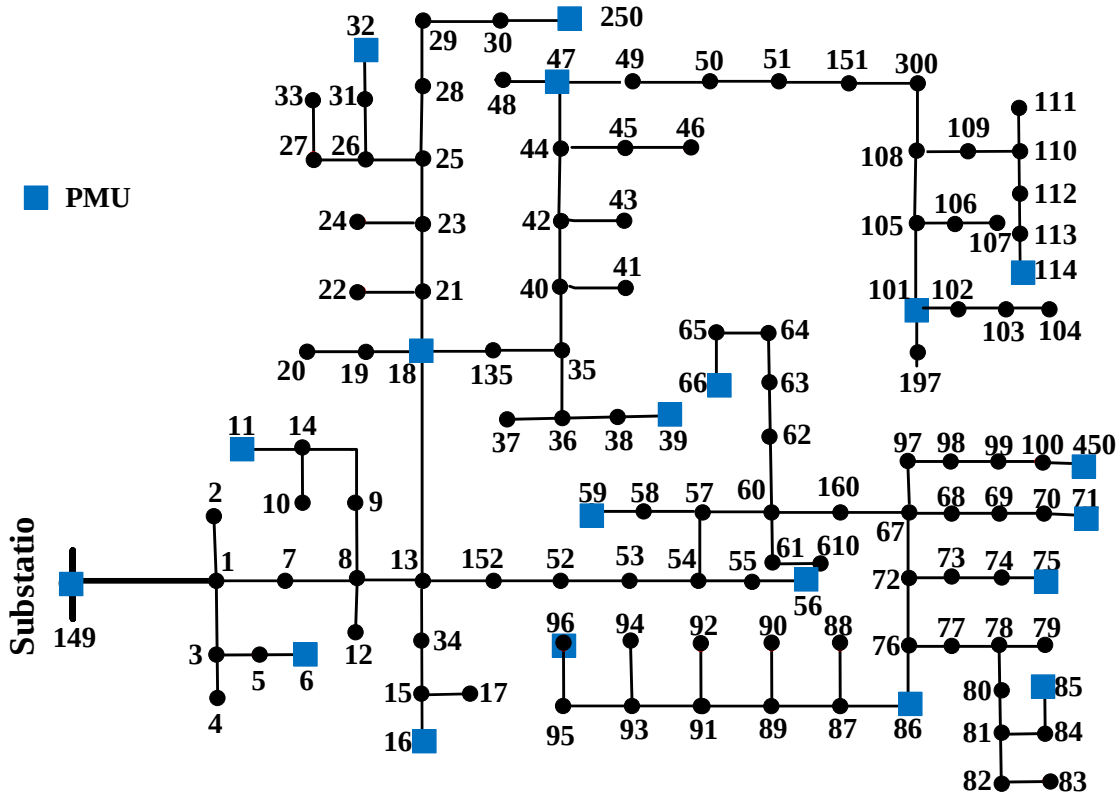

Figure 3.2: One line diagram of IEEE 123-Bus Test System.

unbalanced operation does not impact the discussed sparsity pattern in the vector of state variables and the extracted sparsity features are valid under the unbalanced operation as well.

3.5.3 Comparing Approach 1 and Approach 2

Next, we compare the performance of the two methods that we proposed in previous section, i.e., Approach 1 and Approach 2. Table 3.4 shows the results based on three different metrics. As we can see, both methods converge in 100% of the simulated random

Table 3.3: Performance Comparison for IEEE 123-Bus Test System

Method	MAPE $\Delta\mathbf{x}$	RMSE† $\mathbf{x}$
Proposed Method	7.87%	10.12
[95]	27.36%	36.05
[9]	9.14%	11.64
[57]	25.39%	32.89

† The RMSE values come with a 10^{-3} coefficient.

Table 3.4: Comparing the convergence rate, computation time, and accuracy of the two proposed DSSE approaches

Method	Convergence Rate	Computation Time	MAPE
Approach 1	100%	5.03 Seconds	3.80%
Approach 2	100%	0.75 Seconds	4.12%

scenarios. Approach 1 performs better in terms of the state estimation accuracy, but Approach 2 performs much better in terms of the computation time. The computation time in Table 3.4 corresponds to the total time for estimating the differential state variables for 1000 events.

It is worth mentioning that similar computation time for the method in [95], which uses non-linear power flow equations, is 24.56 seconds, which is much higher than both Approach 1 and Approach 2. Also, the conventional DSSE method in [95], experiences a few cases of divergence in its iterative algorithm.

3.5.4 Effect of Learning the Transition Matrix

In this section, we investigate the importance of learning the event-triggered state transition matrix. We compare two cases in Table 3.5. First, an identity state transition matrix, as in [27]. In this case, we always assume that $\mathbf{A}_k = \mathbf{I}$. Second, we learn the state transition matrix, as in this work. In this case, we set $\mathbf{A}_k$ as in (3.20). As we can see,

Table 3.5: The importance of learning the state transition matrix using discriminative elastic net regression

Method	MAPE of $\Delta\mathbf{x}_k$	MAPE of $\Delta\mathbf{x}_{k\|k-1}$
Using Identity Matrix	5.36%	18.36%
Learning the Matrix	3.98%	6.77%

learning the state transition matrix significantly improves the state estimation performance, which is understood from the improvement in MAPE of $\Delta\mathbf{x}_k$. Here, we also check the ability of the state transition matrix $\mathbf{A}_k$ in accurately predicting the differential state variables at the next event as in (3.10). From (3.35), such prediction is denoted by $\mathbf{x}_{k|k-1}$. As we can see, the MAPE of $\mathbf{x}_{k|k-1}$ can drastically improve, meaning that the accuracy of the state space model in (3.10) can drastically improve by using the proposed method to learn the state transition matrix $\mathbf{A}_k$.

We also studied how the results may change if we change the number of zones in learning the state transition matrix in Fig. 3.3. We check the ability of the state transition matrix $\mathbf{A}_k$ in accurately predicting the differential state variables at the next event in (3.10); thus we again use the MAPE of $\mathbf{x}_{k|k-1}$ as the metric. We can see that as we increase the number of zones, we can enhance the accuracy of prediction, and ultimately the accuracy of estimation. However, it would come with an increased computation time to learn the state transition matrix. Also, increasing the number of zones might lead to overfitting and higher sensitivity to the outcome of the event location identification algorithm, both of which should be avoided.

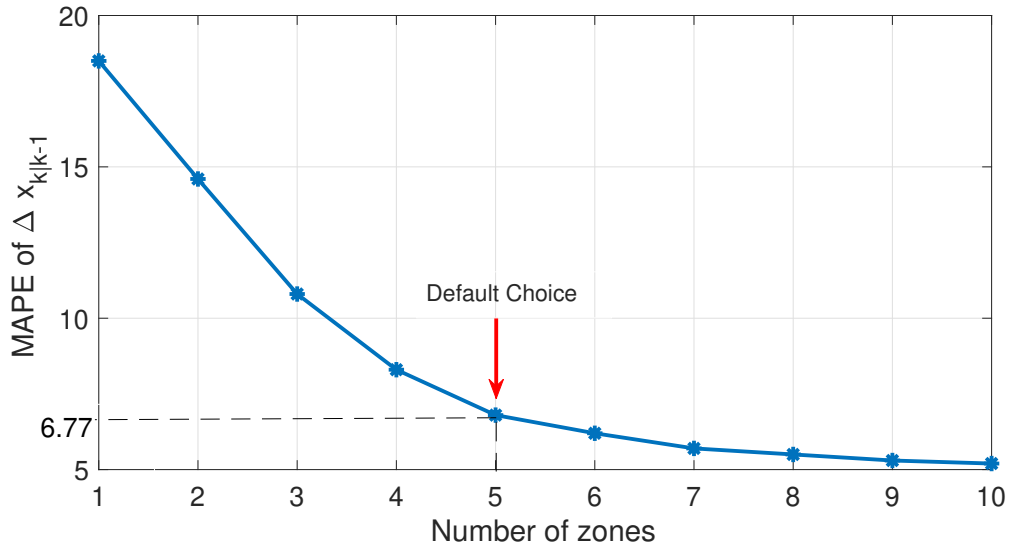

Figure 3.3: The impact of increasing the number of zones on the power distribution system on the performance of the proposed event-triggered state transition matrix in predicting the differential state variables at the next event.

3.5.5 Effect of Virtual Power Measurements

A major feature of the proposed method is the use of virtual differential power measurements in matrix $\mathbf{H}_k^4$. In order to directly evaluate the importance of this innovative feature, next, we compare the state estimation with and without use of virtual power measurements. The results are shown in Table 3.6. We have compared the results for the line power flow and nodal injection power. As we can see, the use of virtual power can significantly improve the accuracy in estimating both of these power quantities that are not being measured directly. They also help improve the overall accuracy of the DSSE results, as seen here in terms of RMSE $\mathbf{V}$ and RMSE $\mathbf{I}$.

Table 3.6: Effect of using virtual power measurements

Method	RMSE $\mathbf{S}^{\dagger}$ (Node)	RMSE $\mathbf{S}^{\dagger}$ (Line)	RMSE $\mathbf{V}^{\dagger}$	RMSE $\mathbf{I}^{\dagger}$
With Virtual Power Measurements	6.12	6.46	7.39	6.44
Without Virtual Power Measurements	8.25	7.86	7.88	7.12

† The RMSE values come with a 10^{-3} coefficient.

3.5.6 Multiple Simultaneous Events

In this section, we examine the rare case wherein multiple additional events occur simultaneously; however, those additional events are relatively minor events compared to the main major event. The main event is a sudden change in the net power of one node and between 10% to 50% of its pre-event value; while the minor events are less than 10% changes in the net power at some other nodes.

The results are shown in Fig. 3.4. As the number of additional minor events increases, the performance of the proposed method gradually degrades. However, the performance degradation is much higher for the other methods in the literature in comparison with the proposed method. Note that, there is a trade-off between adding more sensors to the network to make network fully-observable versus relying on a much lower number of sensors to obtain less exact but still accurate estimation of the state variables in a more practical setting.

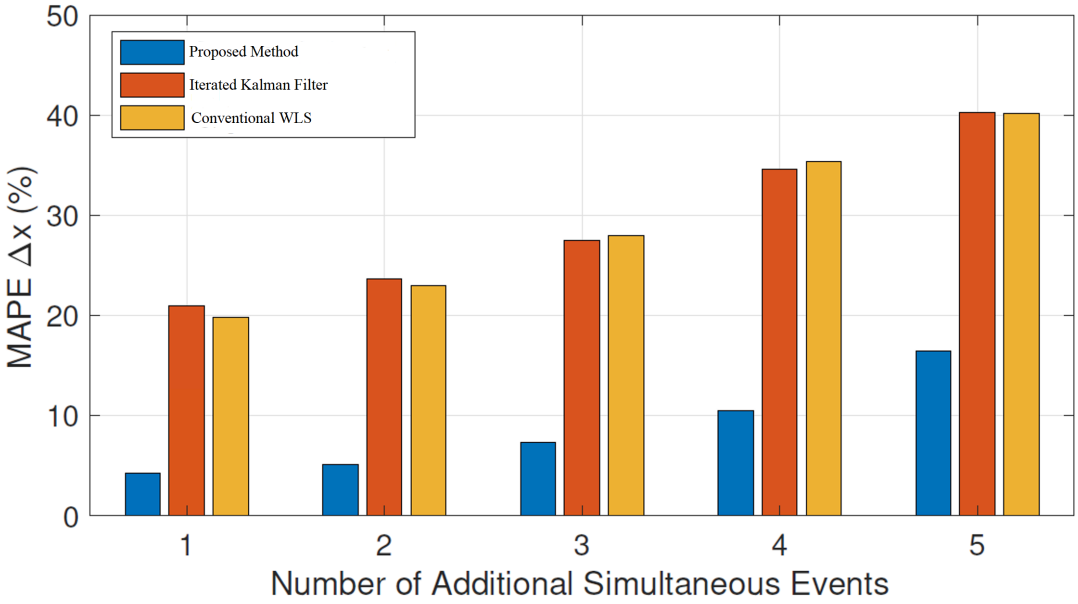

Figure 3.4: Performance evaluation of DSSE when multiple events occur simultaneously; one major event with a few additional relatively minor events.

3.5.7 Real-World Measurements

To further validate the performance of the proposed event-triggered DSSE method, in this section, we apply it to D-PMU measurements from a real-life power distribution feeder at the campus of the Swiss Federal Institute of Technology (EPFL) in Lausanne, Switzerland. The single line diagram of the understudy EPFL power distribution feeder is shown in Fig. 3.5. A total of five D-PMUs are installed on this feeder, labeled as PC-2, which is installed at the substation, and EL-L, EL-G, EL-E, and EL-A. The last one is installed at the end of the feeder. The availability of several D-PMUs in this real-world test-bed can help us validate the performance of the proposed DSSE method. Here we use the measurements from the two D-PMUs at the head and the end of the feeder to run the DSSE algorithm; and then we compare the results with the direct measurements from the

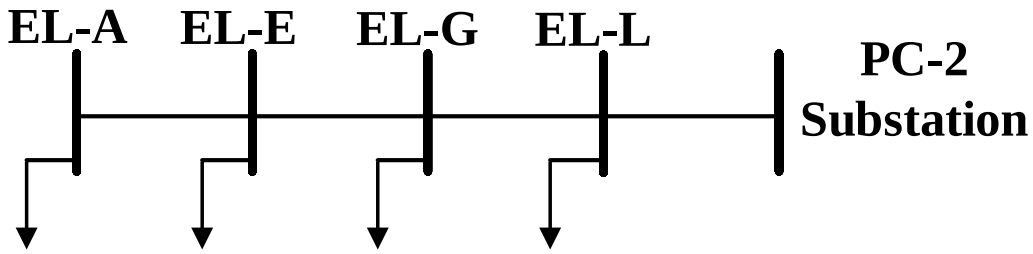

Figure 3.5: The one-line diagram of the real-world EPFL power distribution feeder with five buses and five D-PMUs, one at each bus.

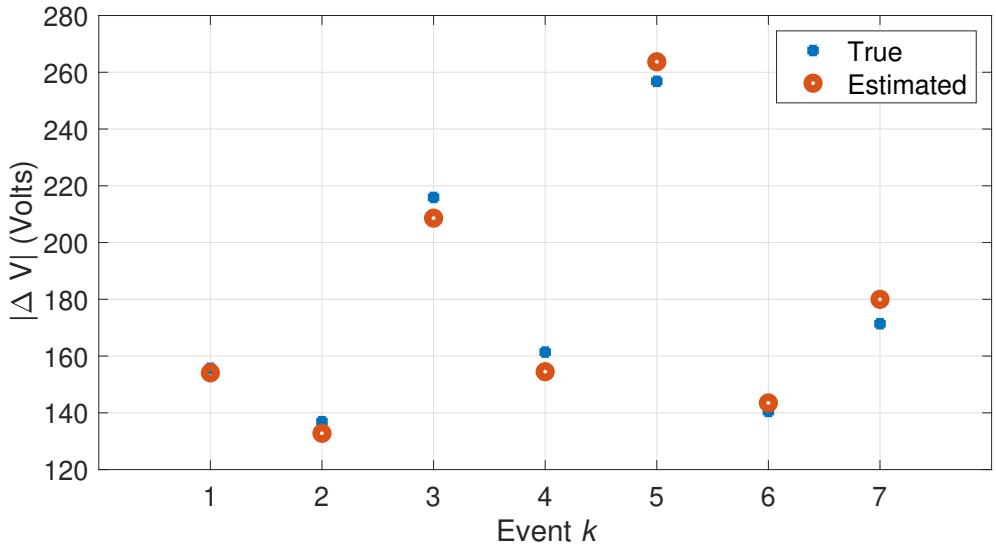

Figure 3.6: The true versus the estimated voltage phasors in differential mode at a typical bus in the real-world EPFL power distribution feeder.

rest of the D-PMUs. We only use the measurement that comes from the D-PMUs installed at node PC-2, which is the head of feeder and node EL-A at the end of feeder.

We have detected multiple major events over a period of time within the available measurements and run the proposed DSSE method. For a typical node, the true values the and estimated values of the magnitude of the voltage phasors are shown in Fig. 3.6. The high accuracy of the state estimation is evident in this real-world example.

Chapter 4

Joint Sparse Estimation of Sensitivity Distribution Factors and Power Flows in Low-Observable Power Distribution Systems

Sensitivity distribution factors (SDFs), such as injection shift factors (ISF) and power transfer distribution factors (PTDF) are powerful computational tools to help grid operators to obtain the solution of the power flow problem without the need to repeatedly solve the non-linear power flow equations. SDFs have various applications, such as in contingency analysis [111], energy dispatch [88], and voltage regulation [110].

In general, there are two different types of methods to obtain the SDF matrices: model-based and measurement-based. In model-based methods, SDF matrices are obtained from the Jacobian matrix of the power flow equations. Based on the application, the power flow equations might be in form of DC approximations [96] or non-linear AC equations [16, 30]. Because of the higher ratio of R/X in power distribution lines, DC approximation typically does not work well. As for the methods that use non-linear power flow equations, the accuracy of the analysis depends on the operating points of the system, which either may not be necessarily known or might be changing frequently. Needless to say that, non-linear power flow models also suffer from higher computation cost.

In measurement-based (i.e., data-driven) methods, the SDF matrices are obtained by applying linear regression to the available measurements [31, 84]. For instance, to estimate the ISF matrix, which is the main focus in this work, the typical required measurements include the changes in the nodal power injection at every node, and the changes in the line power flow in every line segment. One advantage of the measurement-based methods over the model-based methods is that the measurement-based methods do not need any prior knowledge about the power system topology, parameters of the lines, or the time-varying operating points of the system. However, the accuracy of the measurement-based methods highly depends on the type of the regression model, as well as the error and sufficiency of the available measurements.

The sufficiency of the available measurements can be discussed with respect to two aspects. The first one is whether or not the measurements are available at every bus and every line segment in the system, such that a regression analysis can be established

based on the available measurements. At a power transmission network, such comprehensive measurements might indeed be available due to the measurement redundancy. However, this is not the case in a typical power distribution network. Distribution feeders typically suffer from low-observability [9, 6, 20, 21], as the grid operator cannot install sensors at every bus or every line, for reasons such as larger size of the distribution grid and insufficient communication infrastructure. Therefore, unlike the existing measurement-based methods that are developed for use at the transmission level, one should take into account the issue of low-observability when it comes to estimating the SDF matrices using the measurement-based methods in power distribution systems.

To the best of our knowledge no previous work in the literature has looked into this aspect of the problem.

The second aspect to discuss about the sufficiency of the available measurements is the sufficiency of the measurement samples. Even if it is the case that measurements are available at every location, the number of measurement samples may not be enough to solve the linear regression problem through a conventional method. If there are not enough measurement samples, then the measurement matrices are not full-rank. In that case, special methods are needed to deal with rank deficiency in the problem. In [29], a method based on singular value decomposition (SVD) is proposed to obtain the sensitivity matrix from the low-rank measurement matrices. In [88], a recursive partial least square method is proposed to address the same issue. In [32], the rows of the SDF matrix are sparsified based on the electrical distance of the buses and the lines of the power transmission network, and used sparse recovery to obtain the SDF matrix from insufficient measurements.

In this work, our focus is on power distribution feeders with their typical radial topology. We show that such radial topology can create physics-based sparsity features, whose utilization in this low-observability problem makes up for the low-rank properties of the measurement matrices. Accordingly, we address both issues about the sufficiency of measurements in a low-observable power distribution network.

4.1 Problem Statement

4.1.1 System Model

Let us represent a radial power distribution feeder by graph $\mathcal{G} := (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ denotes the set of nodes and $\mathcal{E}$ denotes the set of line segments. Let Δx_i^t denote the change in the nodal power injection at node i at time t. Accordingly, let $\Delta \mathbf{x}^t \in \mathbb{R}^{1\times n}$ denote the vector that captures Δx_i^t for all buses in $\mathcal{N}$ at time t, where n is the number of nodes in the system. Also, let Δy_j^t denote the change in the power flow of line segment j at time t. Similarly, let $\Delta \mathbf{y}^t \in \mathbb{R}^{1\times l}$ denote the vector which contains Δy_j^t for all line segments in $\mathcal{L}$ at time t, where l is the number of line segments in the system. Our goal is to obtain the linear sensitivity ISF matrix such that:

$$\Delta \mathbf{y}^t = \Delta \mathbf{x}^t \mathbf{H}, \tag{4.1}$$

where $\mathbf{H} \in \mathbb{R}^{n\times l}$ is the ISF matrix. Based on equation (4.1), the change in the power flow of line segment j, for any $j \in \mathcal{L}$, can be obtained as:

$$\Delta y_j^t = \Delta x_1^t h_{1j} + \cdots + \Delta x_n^t h_{nj}. \tag{4.2}$$

Here, $h_{1j}, \ldots, h_{nj}$ are the entries of j-th column in matrix $\mathbf{H}$.

4.1.2 Problem Statement

The current measurement-based models in the literature, such as in [31, 84, 29, 32], assume that the changes in the nodal power injection, i.e., $\Delta\mathbf{x}^t$, and the changes in the line power flow, i.e., $\Delta\mathbf{y}^t$, are *known* quantities measured by a variety of sensors that are installed across the power system. Hence, they form a linear regression model to obtain the unknown entries of the ISF matrix. In this measurement-based approach, the total number of unknowns is $n \times l$. But, the number of equations in (4.1) is only l. Therefore, at least $n \times (l-1)$ more independent measurements are needed, such that the unique solution of (4.1) can be obtained.

Next, let $\Delta\mathbf{Y} \in \mathbb{R}^{T\times l}$ be the matrix which includes all the measurement samples of the changes in the line power flows, i.e., the matrix whose rows are vectors of $\Delta\mathbf{y}^t$ from time instant $t = 1$ to time instant $t = T$. In a similar way, let $\Delta\mathbf{X} \in \mathbb{R}^{T\times n}$ be the matrix which includes all the measurement samples of the change in the nodal power injection. Accordingly, we can rewrite equation (4.1) as:

$$\Delta\mathbf{Y} = \Delta\mathbf{X}\mathbf{H}. \tag{4.3}$$

If the measurement matrix $\Delta\mathbf{X}$ is a full-row-rank matrix, then the system of equations in (4.3) is overdetermined. Therefore, the solution to the linear system of equations in (4.3) can be obtained from the least square estimation as:

$$\hat{\mathbf{H}} = (\Delta\mathbf{X}^\top\Delta\mathbf{X})^{-1}\Delta\mathbf{X}^\top\Delta\mathbf{Y}. \tag{4.4}$$

In order to use the above closed-form solution, we need the two measurement matrices $\Delta\mathbf{X}$ and $\Delta\mathbf{Y}$ to be known. However, having measurements at every location

is not a common setting in power distribution systems. Therefore, while the least square estimation in (4.4) can be used for the estimation of the ISF matrix through a measurement-based model in a power transmission network with sufficient number of measurements, it cannot be used in a power distribution network which often suffers from lack of observability. Therefore, we need to develop a new model which can directly address the issue of low-observability in the estimation of the ISF matrix.

4.1.3 Sparsity Pattern of Power Flows and Distribution Factors

Consider the radial power distribution system that is shown in Fig. 4.1. As marked on the figure, suppose a change in the nodal power injection occurs at bus 31. Let us call the path that connects bus 31 to the substation as the substation connector path associated with bus 31. This path is marked with red color on the figure. Based on the Compensation Theorem in Circuit Theory [67], when a change happens in the power injection of a node, then we can replace the element that causes such change by an equivalent current source which injects a level of current to the circuit that causes the same changes on the operating points of the system. Therefore, approximately, the equal amount of current flows from bus 31 all the way up to the substation through the substation connector path, see [8]. The reason is that the impedance of the Thevenin equivalent circuit of the power grid as seen by the distribution feeder at the substation is much less than the impedance of the loads [81]. As a result, once a change in the nodal power injection happens, it only changes the power flow of those distribution line segments that are on the substation connector path, while the power flow for the rest of the line segments remains approximately unchanged.

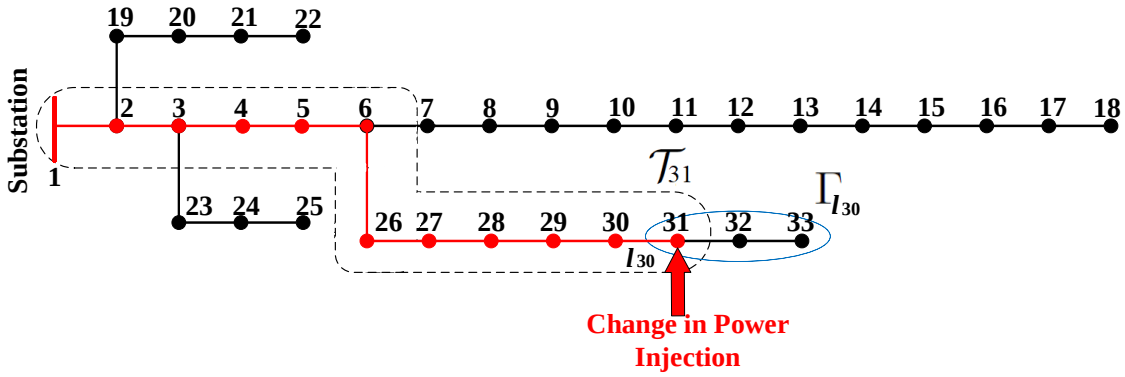

Figure 4.1: An illustrative example for substation connector path in a radial power distribution network. The red path corresponds to the injection at bus 31.

For each node $i \in \mathcal{N}$, let us define tree $\mathcal{T}_i := (\mathcal{V}_i, \mathcal{E}_i)$ as the sub-graph which includes the nodes and lines that are on the substation connector path. Based on the above discussion, once a deviation in the power injection of node i happens, it only changes the power flow in those line segments that belong to set $\mathcal{E}_i$, while the power flow in the rest of line segments remains unchanged. To take advantage of such physics-based approximation, let us denote the set which includes all the nodes that line segment j belong to their associated set $\mathcal{E}_i$ as:

$$\Gamma_j = \{\forall i \in \mathcal{N} \mid j \in \mathcal{E}_i\}. \tag{4.5}$$

A major observation here is that, if the change in the nodal power injection for every node in Γ_j is zero, then the change in the line power flow of line j would be zero. This observation plays an important role in the estimation of sensitivity distribution factors. The above statement means that the line segments which do not belong to set $\mathcal{E}_i$ are insensitive to a change in the nodal power injection of node i. Therefore, we have:

$$h_{ij} = 0, \quad \forall j \notin \mathcal{E}_i. \tag{4.6}$$

4.2 Physics-Aware Measurement-Based Model

In this section, we consider three scenarios for the availability of the measurements. We develop our proposed method to estimate the ISF matrix in each scenario accordingly.

4.2.1 Scenario 1: Nodal Power Injections and Line Power Flows are Fully Measured

Let us start with the least challenging scenario with respect to the availability of measurements. Suppose we can measure the power flow at every node and every line of the power distribution feeder. As discussed in section 4.1.1, this does not necessarily make the system of equations in (4.3) overdetermined. The reason is that the availability of measurements at every line and every node does not necessarily make the corresponding measurement matrices, i.e., $\mathbf{\Delta X}$ and $\mathbf{\Delta Y}$, full-rank, since there may not be enough measurement samples. Thus, in Scenario 1, we deal with low-rank measurement matrices, and our goal is only to obtain the entries of the ISF matrix. Based on the discussion in Section 4.1.3, we know that each column of the ISF matrix is sparse, and we have already extracted the sparsity pattern of the columns in (4.6). Therefore, we can formulate the problem in Scenario 1 as:

$$\begin{aligned} &\underset{\mathbf{H}}{\text{minimize}} \;\; ||\mathbf{\Delta Y} - \mathbf{\Delta X H}||_F^2 + \lambda \sum_{i=1}^{l} ||c(\mathbf{H}, i)||_1 \\ &\text{subject to} \;\; \mathbf{AH} = \mathbf{0}, \end{aligned} \tag{4.7}$$

where $||.||_F$ is the Frobenius norm, $\mathrm{c}(\mathbf{H}, i)$ is the operator which returns the i-th column of the matrix, λ is the sparsity regularization parameter, and matrix $\mathbf{A}$ is constructed based on the sparsity patterns in equation (4.6). The minimization problem in (4.7) is convex and can be solved by any convex optimization solver, such as the CVX toolbox [52].

4.2.2 Scenario 2: Nodal Power Injections are Fully Measured but Line Power Flows are not Fully Measured

In this scenario, we consider a more realistic and more challenging case, where the nodal power injection measurements are available for every node, but the line power flow measurements come from only a few lines in the system. This is a common scenario in practice. Due to the setting of the problem in this scenario, some entries in the measurement matrix $\Delta\mathbf{Y}$ are unknown, because they are not directly measured. Instead, they have to be simultaneously be estimated simultaneously.

To address Scenario 2, let us split the measurement matrix for line power flows into two sub-matrices as:

$$\Delta\mathbf{Y} = \Delta\mathbf{Y}_u + \Delta\mathbf{Y}_k, \tag{4.8}$$

where $\Delta\mathbf{Y}_u$ denotes the unknown part and $\Delta\mathbf{Y}_k$ denotes the known part. Based on the discussion in Section 4.1.3, we know that matrix $\Delta\mathbf{Y}$ is a sparse matrix, and its sparsity pattern is a function of the sparsity pattern of matrix $\Delta\mathbf{X}$. In particular, since we know which entries of matrix $\Delta\mathbf{X}$ are non-zero, we also know which entries of the unknown matrix

$\Delta\mathbf{Y}_u$ are zero. Therefore, we can modify the minimization problem in (4.7) as:

$$\begin{aligned} &\underset{\mathbf{H},\Delta\mathbf{Y}_u}{\text{minimize}} \ ||\Delta\mathbf{Y}_u + \Delta\mathbf{Y}_k - \Delta\mathbf{X}\mathbf{H}||_F^2 + \lambda \sum_{i=1}^{l} ||\mathrm{c}(\mathbf{H}, i)||_1 + \gamma \sum_{j=1}^{T} ||\mathrm{c}(\Delta\mathbf{Y}_u^\top, j)||_1 \\ &\text{subject to } \mathbf{A}\mathbf{H} = \mathbf{0} \\ &\qquad\quad \mathbf{B}\Delta\mathbf{Y}_u = \mathbf{0} \end{aligned} \tag{4.9}$$

where γ is the sparsity regularization parameter, and $\mathbf{B}$ is the matrix which captures the sparsity pattern of the line power flows based on the values in measurement matrix $\Delta\mathbf{X}$.

The optimization problem in (4.9) is a convex minimization problem which can be solve by a commercial solver.

4.2.3 Scenario 3: Neither the Line Power Flows nor the Nodal Power Injections are Fully Measured

In this scenario, we consider the most challenging case, where only a subset of line power flows and only a subset of nodal power injections are measured. Therefore, similar to equation (4.8), we split the matrix of nodal power injection measurements into two sub-matrices as follows:

$$\Delta\mathbf{X} = \Delta\mathbf{X}_u + \Delta\mathbf{X}_k, \tag{4.10}$$

where $\Delta\mathbf{X}_u$ denotes the unknown part and $\Delta\mathbf{X}_k$ denotes the known part. Unlike the minimization problem in (4.9), we cannot define a matrix similar to $\mathbf{B}$, because some of the values of the changes in the nodal power injection are unknown to us. Therefore, we need to develop a way such that the sparsity pattern of the power flows can still be utilized.

To do so, let us define a binary variable b_i^t. It is 1, if the node i experiences a change in the nodal power injection at time t, otherwise it is zero. At each time t, let us

stack up all the binary variables into vector $\mathbf{b}^t \in \mathbb{R}^{1\times n}$. We have:

$$-M.\mathbf{b}^t \leq \Delta\mathbf{x}^t \leq M.\mathbf{b}^t, \tag{4.11}$$

where M is a large number. Based on matrix $\Delta\mathbf{X}_k$, we already know the value for some of the binary variables. However, for the rest of the nodes that we do not measure their power injection directly, the associated binary variable is unknown.

For each line segment j, from the definition of Γ_j in (4.5) in Section 4.1.3, we know that if all of the binary variables that are associated with the nodes in set Γ_j are zero, then the change in the line power flow of line segment j is zero. We can mathematically express this observation as:

$$-M.\max\{\mathbf{b}^t_{\Gamma_j}\} \leq \Delta y^t_j \leq M.\max\{\mathbf{b}^t_{\Gamma_j}\}, \quad \forall j \in \mathcal{L}. \tag{4.12}$$

Now, we can rewrite the minimization problem in (4.9) as:

$$\begin{aligned}
\underset{\mathbf{H},\mathbf{b},\Delta\mathbf{Y}_u,\Delta\mathbf{X}_u}{\text{minimize}} \quad & ||\Delta\mathbf{Y}_u + \Delta\mathbf{Y}_k - (\Delta\mathbf{X}_u + \Delta\mathbf{X}_k)\mathbf{H}||_F^2 \\
& +\lambda \sum_{i=1}^{l} ||\mathrm{c}(\mathbf{H}, i)||_1 + \gamma \sum_{j=1}^{T} ||\mathrm{c}(\Delta\mathbf{Y}_u^\top, j)||_1 \\
& +\omega \sum_{j=1}^{T} ||\mathrm{c}(\Delta\mathbf{X}_u^\top, j)||_1 \\
\text{subject to} \quad & \mathbf{AH} = \mathbf{0} \\
& \text{Eqs. (4.11)} - \text{(4.12)}
\end{aligned} \tag{4.13}$$

where ω is the sparsity regularization parameter.

The minimization problem in (4.13), is non-convex, because of the bilinear term in the objective function. Therefore, we propose to break down the minimization problem in (4.13) into the following two optimization problems:

$$\begin{aligned}\underset{\mathbf{b},\Delta\mathbf{Y}_u,\Delta\mathbf{X}_u}{\text{minimize}} \;& ||\Delta\mathbf{Y}_u + \Delta\mathbf{Y}_k - (\Delta\mathbf{X}_u + \Delta\mathbf{X}_k)\mathbf{H}^0||_F^2 \\ & + \gamma \sum_{j=1}^{T} ||\mathrm{c}(\Delta\mathbf{Y}_u^\top, j)||_1 + \omega \sum_{j=1}^{T} ||\mathrm{c}(\Delta\mathbf{X}_u^\top, j)||_1 \qquad (4.14) \\ \text{subject to} \;& \text{ Eqs. } (4.11) - (4.12)\end{aligned}$$

and

$$\begin{aligned}\underset{\mathbf{H}}{\text{minimize}} \;& ||\Delta\mathbf{Y}^* - (\Delta\mathbf{X}^*)\mathbf{H}||_F^2 + +\lambda \sum_{i=1}^{l} ||\mathrm{c}(\mathbf{H}, i)||_1 \qquad (4.15) \\ \text{subject to} \;& \quad \mathbf{AH} = \mathbf{0}.\end{aligned}$$

In the optimization problem in (4.14), we obtain the solution of $\mathbf{b}$, $\Delta\mathbf{Y}_u$, and $\Delta\mathbf{X}_u$ for a given initial value for matrix $\mathbf{H}$, which is denoted by $\mathbf{H}^0$. Once the estimations of the power flows are obtained by solving (4.14), next, we solve the optimization problem in (4.15) in order to update matrix $\mathbf{H}$. After that, we will continue to iterate between (4.14) and (4.15) until we convergence to a solution. It is worth to mention that, the choice of the initial value for the entries of matrix $\mathbf{H}$ has impact on both convergence and the accuracy of the obtained solution. One good choice to obtain the initial matrix $\mathbf{H}^0$ is to hypothetically assume that the power distribution system is lossless, and set all the non-zero entries equal to one. Another option, which might be more accurate, is to use the estimated values for the previous time slots as the initial choice for $\mathbf{H}^0$. More discussion on this issue will be given in the next section.

4.3 Case Studies

In this section, we test the performance of our proposed method in jointly estimating both the uknown power flows and the ISF matrix. The case studies are completed by

simulating the IEEE 33-bus distribution test network which has a radial topology. We use MATPOWER in MATLAB to run the power flow simulations. We use the default values the of loads per the IEEE 33-bus test system as the operating points of the system. We assume that the measurement noise is Gaussian with zero mean and standard deviation of 0.01.

4.3.1 Performance Evaluation

For all the three scenarios in Section 4.2, we consider a range of 20 to 200 measurement samples to be available for each measurement. For Scenarios 2 and Scenario 3, we assume that between 10% to 50% of the power flow measurements are unknown. We compare our proposed method with the conventional least square estimation only for Scenario 1, as it cannot address the cases in Scenario 2 and Scenario 3 with unknown entries in measurement samples. We use Mean Absolute Percentage Error as the index for this assessment as:

$$\text{MAPE} = \frac{1}{l} \sum_{\forall j \in \mathcal{L}} \frac{1}{n} \sum_{\forall i \in \mathcal{N}} \left| \frac{\hat{h}_{ji} - h_{ji}}{h_{ji}} \right| \times 100\%, \tag{4.16}$$

where $\hat{h}_{ji}$ is the estimated value and h_{ji} is the true value of the sensitivity factor, which is the entry of the ISF matrix. The result for this comparison is shown in Fig 4.2.

As we can see, the proposed method outperforms the conventional least square method in Scenario 1. Also, from the curve of the Scenario 2 and Scenario 3 in which 30% of the measurements are unknown, we can see that missing measurements can cause a degradation on the performance of the proposed method; however, it still outperforms the conventional least square method which is not built to deal with rank deficient matrices.

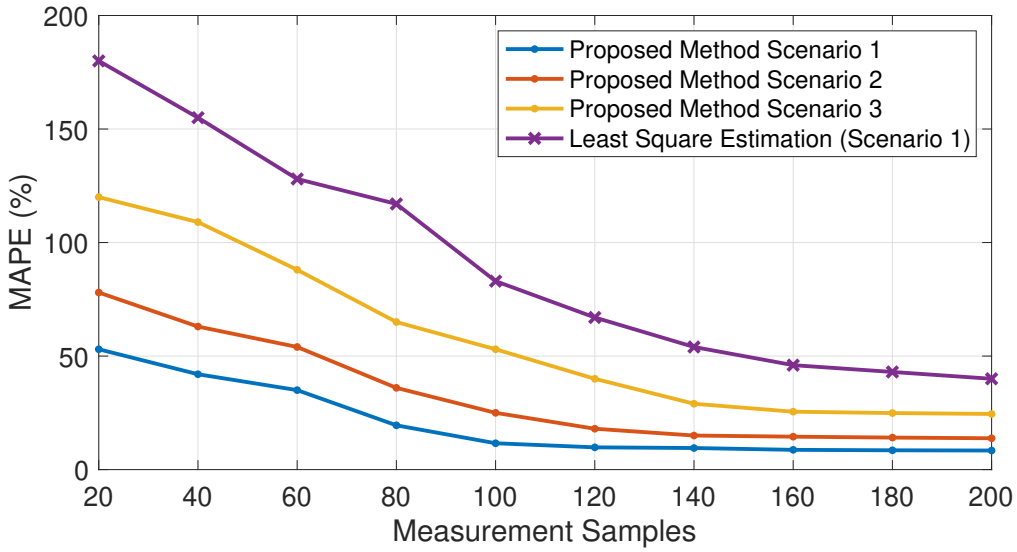

Figure 4.2: The comparison for estimation accuracy of the ISF matrix for different methods. The LS method is applicable to only Scenario 1.

Moreover, as we can see, increasing the number of measurement samples for all methods leads to improvement in the performance of estimation.

We also investigate the impact of missing date percentage on the accuracy of the proposed method in Scenario 2 and Scenario 3. To the best of our knowledge, we are the first who studies this problem under the settings in these two scenarios. Therefore, we only compare the accuracy of estimation for the missing values with respect to their true value for different level of missing data ranging from 10% to 50%. The results are shown in Fig. 4.3, where the number of measurement samples is 200. As we can see, higher percentage of missing data causes the accuracy of the proposed method to degrade.

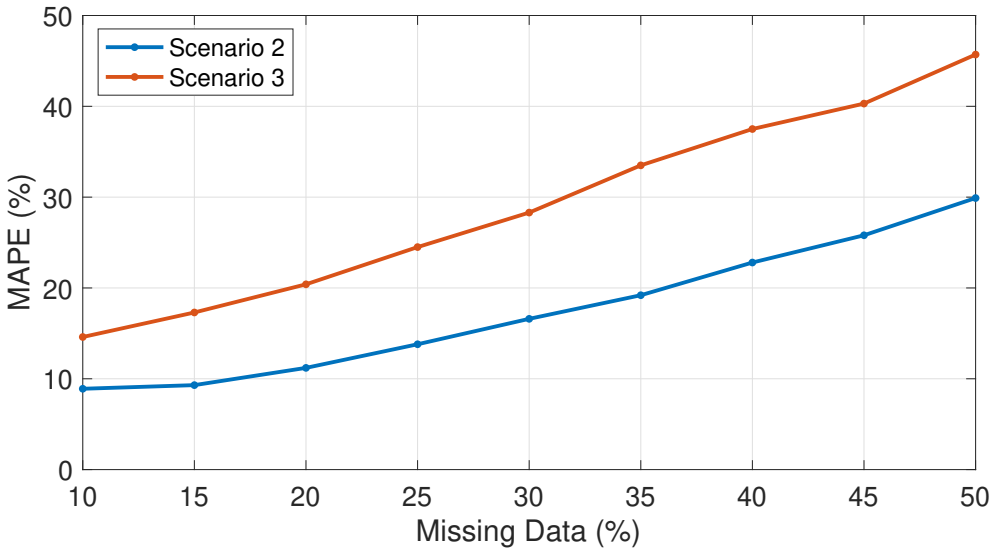

Figure 4.3: The performance of the proposed method under different levels of low-observability severance, i.e., different percentage of missing measurement data. Scenario 1 is not included; because it assumes no missing data.

4.3.2 Importance of Initial Guess for the ISF Matrix

As mentioned earlier, solving the non-linear optimization in (4.13) through the proposed iterative approach in (4.14) and (4.15) depends on the choice for the initial guess of the decision variables, in particular the choice of $\mathbf{H}^0$. An inappropriate initial guess can cause convergence issue. Also, it drastically impact the accuracy of the obtained solution. In this section, we study this impact by comparing the convergence rate and the value of the objective function for multiple initial guesses.

As explained in Section 4.2.3, a good choice for the initial guess is to start from the lossless system, and assume every non-zero entry in the ISF matrix to be one. The second choice is to use the estimated ISF matrix from the previous time instances $t = 1$ to $t = T$, as the initial estimation at time $t = T + 1$. Besides the above two options, we

Table 4.1: Impact of initial guess on the convergence and accuracy of estimation

Initial Guess	Converged	No. Iterations	MAPE
Point 1	Yes	382	31%
Point 2	Yes	566	26%
Point 3	No	-	-
Point 4	Yes	573	45%

have also randomly selected two other initial guesses as the third and the fourth selections. The performance of the above four initial guesses on the convergence and accuracy of the results are shown on Table 1, where 25% of the measurements are unknown and the total number of available measurement samples is 200. We can see that, when we get the initial guess based on estimated values of the previous time slots, proposed method converges to the solution in lower number of iterations, and additionally, its accuracy is higher. Also, random selections of initial guess can cause either to convergence and divergence.

Chapter 5

Conclusions and Future Work

5.1 Summary of Conclusions

In this thesis, we have addressed the challenges in distribution system state estimation that are caused by the low-observability of the power distribution system.

In Chapter 2, a novel sparse tracking DSSE method is proposed to address the low-observability in power distribution networks, where the measurements come from only a small number of D-PMUs. The analysis is done using differential synchrophasors and by dividing the state variables into four types based on their group sparsity properties. Prior to solving the formulated DSSE problem, an event zone identification analysis is applied to augment the proposed sparse DSSE with the side information on the support of the unknown vector. The DSSE is modeled as an adaptive group sparse recovery problem to estimate the differential synchrophasors, which are added to the initial values to estimate the voltage phasors in standard mode at each time slot. Since the proposed method has a tracking scheme, a drift identification and calibration method is developed to enhance

robustness. The proposed method is tested on the IEEE 33-bus and IEEE-123 bus power distribution feeder and the simulation results show the effectiveness of the proposed method over the conventional WLS-DSSE method which is aided by the pseudo-measurements.

In Chapter 3, a novel event-triggered state estimation method is proposed to tackle low-observability in power distribution systems, where the measurements come from a handful of D-PMUs. The proposed method simultaneously addresses three main challenges that exists in the DSSE problem. First, the event-triggered nature of our design can help to avoid imposing unnecessary burden on the operation center. Second, to properly capture the dynamic nature of the modern distribution systems under the low-observability conditions, the problem is formulated over the differential state variables as a generalized group Lasso optimization, which leverages the sparsity features that exist in the system under the event-triggered paradigm. To further improve our ability in conducting a sparse recovery, the DSSE problem formulation is reinforced by a novel set of linear differential power flow equations, and in forms of virtual measurements. Third, to improve the accuracy of the state-space model, the event-triggered state transition matrix is learned in an off-line fashion through discriminative elastic net regression. Extensive performance evaluations confirmed the effectiveness of the proposed methodologies.

In Chapter 4, a new physics-aware measurement-based approach is proposed for joint estimation of the ISF matrix and the power flows in power distribution systems with low-observability. Due to low-observability, the measurement matrices, whether for measuring nodal power injection or line power flow, have several unknown entries. The ISF matrix too is assumed to be unknown. To address low-observability, sparsity pattern of the

ISF matrix and the changes in power flows are approximated based on the radial topology of the power distribution systems. Accordingly, three scenarios are defined based on the availability of the measurements: 1) nodal power injections and line power flows are fully measured; 2) nodal power injections are fully measured but line power flows are not fully measured; 3) neither the line power flows nor the nodal power injections are fully measured. The proper optimization problem is formulated and solved in each case. The performance of the proposed method was verified through numerical results. Case studies included examining the impact of the number of measurement samples as well as the extent of the low-observability.

5.2 Future Work

While we have mainly studied the low-observability issue from the aspect of measurements, in future works, other aspects of low-observability in the power distribution systems can be studies. This can include lack of knowledge about the parameters of the system and topology of the network. Also, one can further study the problem when multiple major events happen on the distribution feeder at the same time. Also, one can investigate how adding other types of measurements, such as load measurements, or power measurements that are obtained from the voltage and current phasors at the D-PMUs, to the existing phasor measurements in our proposed tracking DSSE method may affect the problem formulation and the results. Moreover, one can study how use of machine learning techniques can help to better exploit the sparsity pattern that might exist in different problems related to the operation of power distribution systems. Finally, detection of false

data injections and developing resilient DSSE methods with respect to low-observability condition can receive some attentions, and in particular with similar use of sparse recovery tools.

Bibliography

[1] https://www.powerstandards.com/download/micropmu-data-sheet/.

[2] Radial distribution test feeders, distribution system analysis subcommittee rep. https://site.ieee.org/pes-testfeeders/resources.

[3] A. Abur and A. Gomez-Exposito. *Power system state estimation: theory and implementation*. CRC press, 2004.

[4] F. Ahmad, A. Rasool, E. Ozsoy, R. Sekar, A. Sabanovic, and M. Elitaş. Distribution system state estimation-a step towards smart grid. *Renew. Sust. Energy Rev.*, 81:2659–2671, Jan. 2018.

[5] F. Ahmadi, A. Akrami, M. Doostizadeh, and F. Aminifar. Energy pricing and demand scheduling in retail market: how microgrids' integration affects the market. *IET Smart Grid*, 3(3):309–317, Jun. 2020.

[6] F. Ahmadi Gorjayi and H. Mohsenian-Rad. Physics-aware sparse harmonic state estimation in power distribution systems. In *2022 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5, Apr. 2022.

[7] A. Ahmed, K. S. Sajan, A. Srivastava, and Y. Wu. Anomaly detection, localization and classification using drifting synchrophasor data streams. *IEEE Trans. Smart Grid*, 12(4):3570–3580, Jul. 2021.

[8] A. Akrami, M. S. Asif, and H. Mohsenian-Rad. Sparse distribution system state estimation: An approximate solution against low observability. In *Proc. IEEE PES ISGT Conf*, Washington, DC, Feb. 2020.

[9] A. Akrami, M. S. Asif, and H. Mohsenian-Rad. Sparse tracking state estimation for low-observable power distribution systems using D-PMUs. *IEEE Trans. Power Syst.*, 37(1):551–564, Jan. 2022.

[10] A. Akrami, M. Doostizadeh, and F. Aminifar. Power system flexibility: an overview of emergence to evolution. *J. Modern Power Syst. Clean Energy*, 7(5):987–1007, Sep. 2019.

[11] A. Akrami, M. Doostizadeh, and F. Aminifar. Optimal reconfiguration of distribution network using μPMU measurements: A data-driven stochastic robust optimization. *IEEE Trans. Smart Grid*, 11(1):420–428, Jan. 2020.

[12] A. K. Al-Othman and M. R. Irving. Analysis of confidence bounds in power system state estimation with uncertainty in both measurements and parameters. *Elect. power syst. research*, 76(12):1011–1018, Aug 2006.

[13] S. M. S. Alam, B. Natarajan, and A. Pahwa. Distribution grid state estimation from compressed measurements. *IEEE Trans. Smart Grid*, 5(4):1631–1642, Jul. 2014.

[14] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, and E. Vaahedi. Distribution system state estimation based on nonsynchronized smart meters. *IEEE Trans. Smart Grid*, 6(6):2919–2928, Nov. 2015.

[15] D. Atanackovic and V. Dabic. Deployment of real-time state estimator and load flow in bc hydro dms - challenges and opportunities. In *Proc. 2013 IEEE Power Energy Soc. General Meeting*, Vancouver, BC, 2013.

[16] R. Baldick. Variation of distribution factors with loading. *IEEE Trans. Power Syst.*, 18(4):1316–1323, Nov. 2003.

[17] M. E. Baran, J. Jung, and T. E. McDermott. Including voltage measurements in branch current state estimation for distribution systems. In *Proc. IEEE Power Energy Soc. General Meeting*, Calgary, AB, Jul. 2009.

[18] M. E. Baran and F. F. Wu. Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Trans. Power Del.*, 4(2):1401–1407, Apr. 1994.

[19] B. Bell. The iterated kalman smoother as a gauss–newton method. *SIAM J. Optim.*, 4(3):626–636, Aug. 1994.

[20] S. Bhela, V. Kekatos, and S. Veeramachaneni. Enhancing observability in distribution grids using smart meter data. *IEEE Trans. Smart Grid*, 9(6):5953–5961, Nov. 2018.

[21] S. Bhela, V. Kekatos, and S. Veeramachaneni. Smart inverter grid probing for learning loads: Part I — identifiability analysis. *IEEE Trans. Power Syst.*, 34(5):3527–3536, Sep. 2019.

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn*, 3(1):1–122, 2011.

[23] B. Brinkmann and M. Negnevitsky. A probabilistic approach to observability of distribution networks. *IEEE Trans. Power Syst.*, 32(2):1169–1178, Mar. 2017.

[24] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on information theory*, 52(2):489–509, 2006.

[25] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.

[26] Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Applic*, 14(5-6):877–905, 2008.

[27] C. Carquex, C. Rosenberg, and K. Bhattacharya. State estimation in power distribution systems based on ensemble kalman filtering. *IEEE Trans. Power Syst.*, 33(6):6600–6610, Nov. 2018.

[28] S. Chen, Z. Wei, G. Sun, N. Lu, Y. Sun, and Y. Zhu. Multi-area distributed three-phase state estimation for unbalanced active distribution networks. *J. Modern Power Syst. Clean Energy*, 5(5):767–776, Sep. 2017.

[29] Y. C. Chen, , J. Wang, A. D. Dominguez-Garcioa, and P. W. Sauer. Measurement-based estimation of the power flow jacobian matrix. *IEEE Trans. Smart Grid*, 7(5):2507–2515, Sep. 2016.

[30] Y. C. Chen, S. V. Dhople, A. D. Dominguez-Garcia, and P. W. Sauer. Generalized injection shift factors. *IEEE Trans. Smart Grid*, 8(5):2071–2080, Sep. 2017.

[31] Y. C. Chen, A. D. Dominguez-Garcia, and P. W. Sauer. Measurement-based estimation of linear sensitivity distribution factors and applications. *IEEE Trans. Power Syst.*, 29(3):1372–1382, May 2014.

[32] Y. C. Chen, A. D. Dominguez-Garcia, and P. W. Sauer. A sparse representation approach to online estimation of power system distribution factors. *IEEE Trans. Power Syst.*, 30(4):1727–1738, Jul. 2015.

[33] C. Cheng and X. Bai. Robust forecasting-aided state estimation in power distribution systems with event-triggered transmission and reduced mixed measurements. *IEEE Trans. Power Syst.*, 38(5):4343–4354, Sep. 2021.

[34] O. Chilard and S. Grenard. Detection of measurements errors with a distribution network state estimation function. In *Proc. 22nd Int. Conf. Exhib. Elect. Distrib.*, Stockholm, Sweden, 2013.

[35] K. A. Clements. The impact of pseudo-measurements on state estimator accuracy. In *Proc. IEEE PES General Gen. Meeting*, Detroit, MI, 2011.

[36] S. Dahale, H. S. Karimi, K. Lai, and B. Natarajan. Sparsity based approaches for distribution grid state estimation - a comparative study. *IEEE Access*, 8:198317–198327, 2020.

[37] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu. A survey on state estimation techniques and challenges in smart distribution systems. *IEEE Trans. Smart Grid*, 10(2):2312–2322, Mar. 2019.

[38] K. Dehghanpour, Y. Yuan, Z. Wang, and F. Bu. A game-theoretic data-driven approach for pseudo-measurement generation in distribution system state estimation. *IEEE Trans. Smart Grid*, 10(6):5942–5951, Nov. 2019.

[39] P. Deoliveira-Dejesus, N. Rodriguez, D. Celeita, and G. Ramos. PMU-based system state estimation for multigrounded distribution systems. *IEEE Trans. Power Syst.*, 36(2):1071–1081, Mar. 2021.

[40] M. B. Do Coutto Filho and J. C. S. de Souza. Forecasting-aided state estimation—part i: Panorama. *IEEE Trans. Power Syst.*, 24(4):1667–1677, Nov. 2009.

[41] D. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, Apr. 2006.

[42] David L Donoho, Michael Elad, and Vladimir N Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. on information theory*, 52(1):6–18, 2005.

[43] P. L. Donti, Y. Liu, A. J. Schmitt, A. Bernstein, R. Yang, and Y. Zhang. Matrix completion for low-observability voltage estimation. *IEEE Trans. Smart Grid*, 11(3):2520–2530, May 2020.

[44] G. D'Antona. Power system static-state estimation with uncertain network parameters as input data. *IEEE Trans. Instrum. Meas.*, 65(11):2485–2494, Aug. 2016.

[45] M. Farajollahi, A. Shahsavari, E. M. Stewart, and H. Mohsenian-Rad. Locating the source of events in power distribution systems using micro-PMU data. *IEEE Trans. Power Syst.*, 33(6):6343–6354, Nov. 2018.

[46] J-J Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Inf. Theory*, 50(6):1341–1344, 2004.

[47] Y. R. Gahrooei, A. Khodabakhshian, and R. A. Hooshmand. A new pseudo load profile determination approach in low voltage distribution networks. *IEEE Trans. Power Syst.*, 33(1):463–472, Jan. 2018.

[48] L. D. A. Garcia and S. Grenard. Scalable distribution state estimation approach for distribution management systems. In *Proc. IEEE PES Int. Conf. Exhib. Innovative Smart Grid Technol. Eur.*, Manchester, U.K., 2011.

[49] M. Göl and A. Abur. LAV based robust state estimation for systems measured by PMUs. *IEEE Trans. Smart Grid*, 5(4):1808–1814, Jul. 2014.

[50] A. Gomez-Exposito, A. De La Villa Jaen, C. Gomez-Quiles, P. Rousseaux, and T. Van Cutsem. A taxonomy of multi-area state estimation methods. *Electr. Power Syst. Res.*, 81(4):1060–1069, Apr. 2011.

[51] C. Gomez-Quiles, A. Gomez-Exposito, and A. V. Jaen. State estimation for smart distribution substations. *IEEE Trans. Smart Grid*, 5(4):986–995, Jun. 2012.

[52] M. Grant and S. Boyd. *CVX: Matlab software for disciplined convex programming, version 2.2,*. http://cvxr.com/cvx/, Jan, 2020.

[53] G. Gray, J. Simmins, G. Rajappan, G. Ravikumar, and S. Khaparde. Making distribution automation work: Smart data is imperative for growth. *IEEE Power Energy Mag.,*, 14(1):58–67, Jan./Feb. 2016.

[54] B. P. Hayes J. K. Gruber and M. Prodanovic. A closed-loop state estimation tool for mv network monitoring and operation. *IEEE Trans. Smart Grid*, 6(4):2116–2125, Jul. 2015.

[55] C. Hannon, D. Deka, D. Jin, M. Vuffray, and A. Y. Lokhov. Real-time anomaly detection and classification in streaming pmu data. In *2021 IEEE Madrid PowerTech*, pages 1–6, 2021.

[56] M. Hassanzadeh, C. Y. Evrenosoglu, and L. Mili. A short-term nodal voltage phasor forecasting method using temporal and spatial correlation. *IEEE Trans. Power Syst.*, 31(5):3881–3890, Sep. 2016.

[57] D. A. Haughton and G. T. Heydt. A linear state estimation formulation for smart distribution systems. *IEEE Trans. Power Syst.*, 28(2):1187–1195, May 2013.

[58] M. Huang, Z. Wei, G. Sun, and H. Zang. Hybrid state estimation for distribution systems with ami and scada measurements. *IEEE Access*, 23(7):120350–120359, Aug. 2019.

[59] M. Huang, Z. Wei, J. Zhao, R. A. Jabr, M. Pau, and G. Sun. Robust ensemble kalman filter for medium-voltage distribution system state estimation. *IEEE Trans. Instrum. Meas.*, 69:4114–4124, 2020.

[60] P. Janssen, T. Sezi, and J. C. Maun. Distribution system state estimation using unsynchronized phasor measurements. In *Proc. 3rd IEEE PES Int. Conf. Exhib. Innovative Smart Grid Technol. Eur.*, Berlin, Germany, 2012.

[61] X. Ji, Z. Yin, Y. Zhang, M. Wang, X. Zhang, C. Zhang, and D. Wang. Real-time robust forecasting-aided state estimation of power system based on data-driven models. *Int. J. Electr. Power Energy Syst.*, 125:106412, Feb. 2021.

[62] W. H. Kersting. *Distribution System Modeling and Analysis*. CRC Press, 2007.

[63] P. Khaledian, A. Aligholian, and H. Mohsenian-Rad. Event-based analysis of solar power distribution feeder using micro-PMU measurements. In *2021 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE, 2021.

[64] P. Khaledian and H. Mohsenian-Rad. Event region identification and its data-driven applications in behind-the-meter solar farms based on micro-PMU measurements. *IEEE Trans. Smart Grid*, 13(3):2094–2106, May 2019.

[65] Xiangyu Kong, Xiaopeng Zhang, Xuanyong Zhang, Chengshan Wang, Hsiao Chiang, and Peng Li. Adaptive dynamic state estimation of distribution network based on interacting multiple model. *IEEE Trans. Sustain. Energy (accepted)*, 13(2):643–652, Oct. 2021.

[66] H. Kopetz. Event-triggered versus time-triggered real-time systems. In *Operating Systems of the 90s and Beyond*, pages 86–101. Springer, 1991.

[67] K. S. Kumar. *Electric circuits and networks*. Pearson, India, 2009.

[68] S. Li et al. Event-trigger heterogeneous nonlinear filter for wide-area measurement systems in power grid. *IEEE Trans. Smart Grid*, 10(3):2752–2764, May 2019.

[69] W. M. Lin and J. H. Teng. Distribution fast decoupled state estimation by measurement pairing. *IEE Proc-Gener. Transm Distrib.*, 143(1):43–48, 1996.

[70] W. M. Lin and J. H. Teng. State estimation for distribution systems with zero-injection constraints. *IEEE Trans. Power Syst.*, 11(1):518–524, Feb. 1996.

[71] X. Liu et al. Event-trigger particle filter for smart grids with limited communication bandwidth infrastructure. *IEEE Trans. Smart Grid*, 9(6):6918–6928, Nov. 2018.

[72] A. Primadianto C. N. Lu. A review on distribution system state estimation. *IEEE Trans. Power Syst.*, 32(5):3875–3883, Dec. 2016.

[73] C. N. Lu, J. H. Teng, and W. H. Liu. Distribution system state estimation. *IEEE Trans. Power Syst.*, 10(1):229–240, Feb. 1995.

[74] W. Ma, J. Qiu, X. Liu, G. Xiao, J. Duan, and B. Chen. Unscented kalman filter with generalized correntropy loss for robust power system forecasting-aided state estimation. *IEEE Trans. Industr. Inform.*, 15(11):6091–6100, May 2019.

[75] V. Madani et al. Distribution automation strategies, challenges and opportunities in a changing landscape. *IEEE Trans. Smart Grid*, 6(4):2157–2165, Jul. 2015.

[76] R. Madbhavi, B. Natarajan, and B. Srinivasan. Enhanced tensor completion based approaches for state estimation in distribution systems. *IEEE Trans. Ind. Inform.*, 17(9):5938–5947, Sep. 2021.

[77] M. Majidi, M. Etezadi-Amoli, and H.Livani. Distribution system state estimation using compressive sensing. *In. J. Elec. Power Energy Syst.*, 88:175–186, Jun. 2017.

[78] E. Manitsas, R. Singh, B. Pal, and G. Strbac. Modelling of pseudomeasurements for distribution system state estimation. In *Proc. Int. Conf. Exhib. Elect. Distrib. Seminar: SmartGrids Distrib.*, Frankfurt, Germany, 2008.

[79] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac. Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling. *IEEE Trans. Power Syst.*, 27(4):1888–1896, Nov. 2012.

[80] L. Mili, M. G. Cheniae, and P. J Rousseeuw. Robust state estimation of electric power systems. *IEEE Trans. on Circuits Syst. I, Fundam. Theory Appl.*, 41(5):349–358, May 1994.

[81] H. Mohsenian-Rad. *Smart Grid Sensors: Principles and Applications.* Cambridge University Press, UK, 2022.

[82] H. Mohsenian-Rad, E. Stewart, and E. Cortez. Distribution synchrophasors: Pairing big data with analytics to create actionable information. *IEEE Power Energy Mag.*, 16(3):26–34, May 2018.

[83] A. Monticelli. *State estimation in electric power systems: a generalized approach.* Springer Science Businees Media, 2012.

[84] C. Mugnier, K. Christakou., J. Jaton, M. De Vivo, M. Carpita, and M/ Paolone. Model-less/measurement-based computation of voltage sensitivities in unbalanced electrical distribution networks. In *2016 Power Systems Computation Conference (PSCC)*, pages 1–7, Jun. 2016.

[85] C. Muscas, M. Pau, P. A. Pegoraro, and S. Sulis. Effects of measurements and pseudomeasurements correlation in distribution system state estimation. *IEEE Trans. Instrum. Meas.*, 63(12):2813–2823, Dec. 2014.

[86] C. Muscas, M. Pau, P. A. Pegoraro, S. Sulis, F. Ponci, and A. Monti. Multiarea distribution system state estimation. *IEEE Trans. Instrum. Meas.*, 64(5):1140–1148, Aug. 2015.

[87] D. T. Nguyen. Modeling load uncertainty in distribution network monitoring. *IEEE Trans. Power Syst.*, 30(5):2321–2328, Sep. 2015.

[88] S. Nowak, Y. C. Chen, and L. Wang. Measurement-based optimal der dispatch with a recursively estimated sensitivity model. *IEEE Trans. Power Syst.*, 35(6):4792–4802, Nov. 2020.

[89] N. Nusrat, M. Irving, and G. Taylor. Development of distributed state estimation. In *Proc. IEEE Int. Symp. Ind. Electron.*, Gdansk, Poland, 2011.

[90] T. Ochi, D. Yamashita, K. Koyanagi, and R. Yokoyama. The development and the application of fast decoupled load flow method for distribution systems with high r/x ratios lines. In *2013 IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–6. IEEE, 2013.

[91] J. Ostrometzky, K. Berestizshevsky, A. Bernstein, and G. Zussman. Physics-informed deep neural network method for limited observability state estimation. Available: arXiv:1910.06401. 2019.

[92] J. Qi, A. Taha, and J. Wang. Comparing kalman filters and observers for power system dynamic state estimation with model uncertainty and malicious cyber attacks. *IEEE Access*, 6:77155–77168, Aug. 2018.

[93] M. M. Rana and L. Li. An overview of distributed microgrid state estimation and control for smart grids. *Sensors*, 15(2):4302–4325, Feb. 2015.

[94] M. M. Rana, W. Xiang, and E. Wang. Iot-based state estimation for microgrids. *IEEE Internet Things J.*, 5(2):1345–1346, Jan. 2018.

[95] S. Sarri, M. Paolone, R. Cherkaoui, A. Borghetti, F. Napolitano, and C. A. Nucci. State estimation of active distribution networks: Comparison between WLS and iterated kalman-filter algorithm integrating PMUs. In *IEEE PES ISGT Europe*, 2012.

[96] P. W. Sauer. On the formulation of power distribution factors for linear load flow methods. *IEEE Trans. Power Appar. Syst.*, PAS-100(2):764–770, Feb. 1981.

[97] R. Singh, E. Manitsas, B. C. Pal, and G. Strbac. A recursive bayesian approach for identification of network configuration changes in distribution system state estimation. *IEEE Trans. Power Syst.*, 25(3):1329–1336, Aug. 2010.

[98] R. Singh, B. C. Pal, and R. A. Jabr. Distribution system state estimation through gaussian mixture model of the load as pseudo-measurement. *IET Gen., Transm. Distrib.*, 4(1):50–59, Jan. 2009.

[99] R. Singh, B. C. Pal, and R. A. Jabr. Statistical representation of distribution system loads using gaussian mixture model. *IEEE Trans. Power Syst.*, 25(1):29–37, Feb. 2010.

[100] J. Song, E. Dall'Anese, A. Simonetto, and H. Zhu. Dynamic distribution state estimation using synchrophasor data. *IEEE Trans. Smart Grid*, 11(1):821–831, Jan. 2020.

[101] I. Táczi, B. Sinkovics, I. Vokony, and B. Hartmann. The challenges of low voltage distribution system state estimation—an application oriented review. *Energies*, 14(17):5363, Aug. 2021.

[102] A. Tahabilder, P. K. Ghosh, S. Chatterjee, and N. Rahman. Distribution system monitoring by using micro-pmu in graph-theoretic way. In *2017 4th Int. Conf. Advances Electr. Eng. (ICAEE)*, pages 159–163. IEEE, 2017.

[103] M. Q. Tran, A. S. Zamzam, P. H. Nguyen, and G. Pemen. Multi-area distribution system state estimation using decentralized physics-aware neural networks. *Energies*, 14(11):3025, May 2021.

[104] Joel A Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans. Information Theory*, 52(3):1030–1051, 2006.

[105] A. Vargas and M. E. Samper. Real-time monitoring and economic dispatch of smart distribution grids: High performance algorithms for DMS applications. *IEEE Trans. Smart Grid*, 3(2):866–877, Jun. 2012.

[106] Y. Wang, Y. Sun, and V. Dinavahi. Robust forecasting-aided state estimation for power system against uncertainties. *IEEE Trans. Power Syst.*, 35(1):691–702, Aug. 2019.

[107] Y. Weng, R. Negi, and M. D. Ilic. Probabilistic joint state estimation for operational planning. *IEEE Trans. Smart Grid*, 10(1):601–612, Jan. 2019.

[108] A. J. Wood, B. F. Wollenberg, and G. B. Sheble. *Power generation, operation, and control*. John Wiley Sons, 2013.

[109] F. F. Wu. Power system state estimation: a survey. *Int. J. Elect. Power Energy Syst.*, 12(2):80–87, Apr. 1990.

[110] H. Xu, A. D. Dominguez-Garcia, V. V. Veeravalli, and P. W. Sauer. Data-driven voltage regulation in radial power distribution systems. *IEEE Trans. Power Syst.*, 35(3):2133–2143, May 2020.

[111] R. Yao and F. Qiu. Novel ac distribution factor for efficient outage analysis. *IEEE Trans. Power Syst.*, 35(6):4960–4963, Sep. 2020.

[112] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos. Data-driven learning-based optimization for distribution system state estimation. *IEEE Trans. Power Syst.*, 34(6):4796–4805, Nov. 2019.

[113] H. Zhang, W. Yin, and L. Cheng. Necessary and sufficient conditions of solution uniqueness in 1-norm minimization. *J. Optim. Theory Appl.*, (164):109–122, 2015.

[114] Y. Zhang, J. Wang, and Z. Li. Interval state estimation with uncertainty of distributed generation and line parameters in unbalanced distribution systems. *IEEE Trans. Power Syst.*, 35(1):762–772, Jul. 2019.

[115] Z. Zhang, Z. Lai, Y. Xu, L. Shao, J. Wu, and G. Xie. Discriminative elastic-net regularized linear regression. *IEEE Trans. Image Processing*, 26(3):1466–1481, Mar. 2021.

[116] Junbo Zhao et al. Power system dynamic state estimation: Motivations, definitions, methodologies, and future work. *IEEE Trans. Power Syst.*, 34(4):3188–3198, Jul. 2019.

[117] X. Zhou, Z. Liu, Y. Guo, C. Zhao, J. Huang, and L. Chen. Gradient-based multi-area distribution system state estimation. *IEEE Trans. Smart Grid*, 11(6):5325–5338, Jun. 2020.

[118] Y. Zhou, R. Arghandeh, and C. J. Spanos. Partial knowledge data-driven event detection for power distribution networks. *IEEE Trans. Smart Grid*, 9(5):5152–5162, Sep. 2018.

[119] P. Zhuang, R. Deng, and H. Liang. False data injection attacks against state estimation in multiphase and unbalanced smart distribution systems. *IEEE Trans. Smart Grid*, 10(6):6000–6013, Jan. 2019.

[120] L. Zou, Z. Wang, H. Gao, and X. Liu. Event-triggered state estimation for complex networks with mixed time delays via sampled data information: The continuous-time case. *IEEE Trans. Cybernetics*, 45(12):2804–2815, Dec. 2015.