Design and analysis of dynamic Compressive Sensing in distribution grids

by

Hazhar Sufi Karimi

B.S., Sahand University of Technology, Iran, 2012

M.S., Tarbiat Modares University, Iran, 2014

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Mike Wiegers Department of Electrical and Computer Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Abstract

The transition to a smart distribution grid is powered by enhanced sensing and advanced metering infrastructure that can provide situational awareness. However, aggregating data from spatially dispersed sensors/smart meters can present a significant challenge. Additionally, the lack of reliability in communication network used for aggregating this data, prevents its use for real time operations such as state estimation and control. With these challenges associated with measurement availability and accessibility, current distribution systems are typically unobservable. To cope with the unobservability issue, compressive sensing (CS) theory allows us to recover system state information from a small number of measurements provided the states of the distribution system exhibit sparsity. The spatio-temporal correlation of loads and/or rooftop photovoltaic (PV) generation results in sparsity of distribution system states. In this dissertation, we first validate this system sparsity property and exploit it to develop two (direct/indirect) voltage state estimation strategies for a three-phase unbalanced distribution network. Secondly, we focus on addressing the challenge of sparse signal recovery from limited measurements while incorporating their temporal dependence. Specifically, we implement two recursive dynamic CS approaches namely, streaming modified weighted-ℓ_1 CS and Kalman filtered CS that reconstruct a sparse signal using the current underdetermined measurements and the prior information about the sparse signal and its support set. Using practical distribution system power measurements as a case study, we quantify, for the first time, the performance improvement achievable with such recursive techniques relative to batch algorithms.

CS based signal recovery efforts typically assume that a limited number of measurements are available. However, in practice, due to communication network impairments, there is no guarantee that even this limited set of information might be available at the time of

processing at the fusion/control center. Therefore, for the first time, we investigate the impact of intermittent measurement availability and random delays on recursive dynamic CS. Specifically, we quantify the error dynamics in both sparse signal estimation and support set estimation for a modified Kalman filter-CS based strategy in the presence of measurement losses. Using input-to-state stability analysis, we provide an upper bound for the expected covariance of the estimation error for a given rate of information loss. Next, we develop a modified CS algorithm that leverages apriori knowledge of signal correlation to project delayed measurements to the current signal recovery instant. We derive a new result quantifying the impact of errors in the apriori correlation model on signal recovery error.

Lastly, we study the robustness of CS based state estimation to uncertainty in distribution network topology knowledge. Topology identification is a challenging problem in distribution systems in general and especially, when there are limited number of available measurements. We tackle this problem by jointly estimating the states and network topology via an integrated mixed integer nonlinear program formulation. By developing convex relaxations of the original formulation as well Markovian models for dynamic topology transitions, we illustrate the superior performance achieved in both state estimation and in topology identification.

In summary, this dissertation offers the first comprehensive treatment of dynamic CS in smart distribution grids and can serve as the foundation of numerous follow-on efforts related to networked state estimation and control.

Design and analysis of dynamic Compressive Sensing in distribution grids

by

Hazhar Sufi Karimi

B.S., Sahand University of Technology, Iran, 2012

M.S., Tarbiat Modares University, Iran, 2014

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Mike Wiegers Department of Electrical and Computer Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Approved by:

Major Professor
Dr. Balasubramaniam Natarajan

Copyright

Abstract

The transition to a smart distribution grid is powered by enhanced sensing and advanced metering infrastructure that can provide situational awareness. However, aggregating data from spatially dispersed sensors/smart meters can present a significant challenge. Additionally, the lack of reliability in communication network used for aggregating this data, prevents its use for real time operations such as state estimation and control. With these challenges associated with measurement availability and accessibility, current distribution systems are typically unobservable. To cope with the unobservability issue, compressive sensing (CS) theory allows us to recover system state information from a small number of measurements provided the states of the distribution system exhibit sparsity. The spatio-temporal correlation of loads and/or rooftop photovoltaic (PV) generation results in sparsity of distribution system states. In this dissertation, we first validate this system sparsity property and exploit it to develop two (direct/indirect) voltage state estimation strategies for a three-phase unbalanced distribution network. Secondly, we focus on addressing the challenge of sparse signal recovery from limited measurements while incorporating their temporal dependence. Specifically, we implement two recursive dynamic CS approaches namely, streaming modified weighted-ℓ_1 CS and Kalman filtered CS that reconstruct a sparse signal using the current underdetermined measurements and the prior information about the sparse signal and its support set. Using practical distribution system power measurements as a case study, we quantify, for the first time, the performance improvement achievable with such recursive techniques relative to batch algorithms.

CS based signal recovery efforts typically assume that a limited number of measurements are available. However, in practice, due to communication network impairments, there is no guarantee that even this limited set of information might be available at the time of

processing at the fusion/control center. Therefore, for the first time, we investigate the impact of intermittent measurement availability and random delays on recursive dynamic CS. Specifically, we quantify the error dynamics in both sparse signal estimation and support set estimation for a modified Kalman filter-CS based strategy in the presence of measurement losses. Using input-to-state stability analysis, we provide an upper bound for the expected covariance of the estimation error for a given rate of information loss. Next, we develop a modified CS algorithm that leverages apriori knowledge of signal correlation to project delayed measurements to the current signal recovery instant. We derive a new result quantifying the impact of errors in the apriori correlation model on signal recovery error.

Lastly, we study the robustness of CS based state estimation to uncertainty in distribution network topology knowledge. Topology identification is a challenging problem in distribution systems in general and especially, when there are limited number of available measurements. We tackle this problem by jointly estimating the states and network topology via an integrated mixed integer nonlinear program formulation. By developing convex relaxations of the original formulation as well Markovian models for dynamic topology transitions, we illustrate the superior performance achieved in both state estimation and in topology identification.

In summary, this dissertation offers the first comprehensive treatment of dynamic CS in smart distribution grids and can serve as the foundation of numerous follow-on efforts related to networked state estimation and control.

Table of Contents

List of Figures

List of Tables

Acknowledgments

There are many people who have helped me along the way on this journey. I want to take this opportunity to thank them.

Firstly, I would like to express my sincere gratitude to my advisor Dr. Bala Natarajan for his dedicated support, knowledge and guidance. Dr. Bala continuously provided encouragement and was always willing and enthusiastic to assist in any way he could throughout my Ph.D study.

My special appreciation to my doctoral committee members: Dr. Sanjoy Das, Dr. Hongyu Wu, Dr. Bacim Alali, and Dr. Vinod Kumarappan, for their constructive advices and insightful comments that help to improve the quality of this dissertation. Also, I would like to thank Dr. Anil Pahwa for his guidance during this research.

Special thanks are extended to my friends in WiCom group, Alaleh, Solmaz, Reza B., Wenji, Kumar, Rahul, Kexing, Sai, Shweta, Biswajeet, Mohammad and Dylan for their help, encouragement and friendship.

I would like to thank my dear and loyal friends at K-STATE, Heman, Behnaz, Liam, Reza M., Hojjat, Sahar (Rest In Peace), Farzaneh, Jalal, Aram, Erfan and Javad for their support and generous care. I will always miss the unforgettable moments and joys that you shared with me.

I am extremely grateful to my parents and my parents-in-law, for their unconditional trust, timely encouragement, and endless patience. In spite of all the hardship due to long distance and being away from home, they have always supported me to achieve my goals.

Special thanks to my supportive family members: Avin, Hemn, Soodabeh, Smko, Shadi, Hiwa, Sahar, Eiliya, Zhowan and Hezha for their kindness, motivation and unwavering belief in me.

I would like to express my endless love to my caring, loving, and supportive fiancé, Katayoun (Katy). You put up with me being distracted and missing many events. Katy,

I'm glad to find my soulmate and my bestfriend within you. "I will always lean my heart as close to your soul as I can". Thank you for your patience, support and encouragement.

This material is based upon work supported by the Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), Solar Energy Technologies Office, under Award Number DE-EE0008767.

Dedication

To my parents, my parents-in-law, my siblings

&

Katy

for their support, encouragement and unconditional love.

Abbreviations

AMI	Advanced metering infrastructure
BPDN	Basis pursuit denoising
CMR	Compressed measurement ratio
CRN	Cognitive radio networks
CS	Compressive sensing
DCT	Discrete cosine transform
DSSE	Distribution system state estimation
IMM	Interacting multiple model
INAE	Integrated normalized absolute error
ISS	Input-to-state stability
KF	Kalman filter
KF-Mod-CS	Kalman filtered modified compressive sensing
MILP	Mixed integer linear programming
MINLP	Mixed integer nonlinear programming
MIQP	Mixed integer quadratic programming
MMSE	Minimum mean square error
NP-Hard	Non-deterministic polynomial-time hardness
PV	Photo-voltaic
RIP	Restricted isometry property
SMW-ℓ_1	Streaming modified weighted-ℓ_1
SR	Sparsity ratio
WLS	Weighted least square

Chapter 1

Introduction

1.1 Overview and Motivation

Smart distribution grids consist of several components including distributed generation (e.g., solar, wind), advanced metering infrastructure (AMI) and SCADA (e.g., voltage, current, switch status, etc.) sensors. The sensing and smart metering infrastructure provides previously unavailable insights into the distribution grid all the way to the home level. Data from all the sensors and smart meters can possibly be used for situational awareness as well as real time control. While the physical scale of the system results in a large amount of raw measurements [1], the sampling rate of commercially available smart meters and SCADA sensors can vary in the range of a few seconds to the order of minutes contributing to a data deluge [2]. Furthermore, it is usually challenging to aggregate all the data reliably as even under optimal conditions, the capacity of the underlying communication network is inadequate to keep up with the volume and velocity of information flowing from the smart meters and sensors [3]. Therefore, currently, the smart meter data is primarily used for billing purposes and not for real time operations. Since optimal control and operation of a distribution grid is contingent upon the ability to estimate the system states via the supporting cyber infrastructure, it is critical to investigate methods to estimate the states with a small number of measurements. This is the fundamental problem we seek to study in the

context of three phase unbalanced distribution networks.

1.2 Related Work and Challenges

State estimation in a power grid is a classic problem that has garnered interest over many decades. Various estimation strategies including their advantages and the key challenges are summarized in [4]-[6]. State estimation in a traditional distribution grid is difficult due to the limited number of measurements at that level making the system unobservable [7]. However, the underdetermined system can be transformed into an overdetermined system by generating pseudo-measurements (based on historical data) as discussed in [8]. Then, a weighted least square (WLS) approach can be used to estimate the states for the underlying nonlinear system [9]. The estimator variance can be further reduced by exploiting the load correlation as discussed in [10]. In addition to correlation among loads, with the increasing penetration of rooftop PV (photo-voltaic) generation in smart grids, one can anticipate a significant correlation in generation. This underlying correlation typically results in sparsity in a transformation basis (e.g., wavelet basis) as demonstrated in [11]. Therefore, compressive sensing strategies that enable us to recover a larger dimensional signal from fewer measurements in the presence of sparsity can play a critical role in efficient data aggregation and processing [12].

Although, application of compressive sensing for smart distribution grids is not as mature as for speech, video, or image compression applications, there have been some recent efforts in this direction. [13] presents a summary of efforts involved in the space of compressive sensing in distribution grids. In [14], the authors propose a fuzzy based transformation with a least-squares method to compress signals resulting in a higher accuracy relative to principal component analysis and discrete wavelet transformation approaches (under certain conditions). A general framework for data compression of AMI data is provided in [15]. Here, sparse binary measurement matrices are utilized for joint reconstruction of the spatio-temporal load profile. A singular value decomposition based data compression approach is investigated for a smart distribution grid in steady state [16]. A novel distribution system

state estimation for three phase systems has been provided in [17], which uses micro-phasor measurement units and (1D) spatial compressive sensing. However, [17] does not consider smart meter data and given the fact that the number of smart meters is typically far more than micro PMUs, the proposed approach is limited in its practical application. To address the problem of power line outage detection, [18] exploits compressive system identification which is well-known for being a time-efficient approach in complex network analysis. Recently, a comprehensive survey on the compression methods for smart meter big data is presented in [19] that evaluates the methods and discusses the existing challenges in data compression of smart grids. In the initial efforts, the physical grid model (and the associated power flow equations) was not incorporated as part of the analysis. Spatio-temporal correlation between loads and distributed generation is exploited in [20] that recovers the power measurements across the grid and estimates the voltages using noiseless CS. The mentioned techniques deal with single phase systems. However, three phase networks may not be balanced, (i.e., each phase could have different load and generation profiles) and exhibit coupling across phases. Therefore, states of a three phase system cannot be directly and trivially estimated by generalizing the proposed method for single phase systems. Following this limitation, the first question we seek to address in this dissertation is:

Question 1: *How can a compressive sensing method be designed to indirectly/directly estimate the states of a three phase unbalanced distribution system with limited measurement?*

Most of the above mentioned research efforts employ static sparse recovery methods where they only apply current measurements for signal recovery purposes. Even, with a dynamic scheme, the existing approaches implement batch algorithms that consider the entire streaming signal as a single sparse spatiotemporal signal where time sparsity is assumed in the Fourier domain or wavelet domain [21]. However, these assumptions may not hold for many cases. Additionally, batch algorithms suffer from other disadvantages including (1) offline implementation that is slow, and (2) a large memory requirement that increases linearly with the sequence length [22]. To cope with this problem, the following question needs to be addressed:

Question 2: *Can we implement dynamic CS methods for signal recovery that effectively*

address the drawbacks of batch algorithms?

The existing body of work on CS based estimation assume that the required measurements arrive on-time. In practice, the control fusion center may receive intermittent or delayed measurements due to impairments in the underlying communication link [23, 24]. To address these issues in distribution systems, [25] provides a distributed dynamic state estimation strategy when the measurement packets are lost. In [26], a modified re-iterated Kalman filter is proposed that estimates the states of the smart grid in the presence of both lost and delayed measurements. [27] presents a flexible hybrid state estimation algorithm when the underlying communication network suffers from irregularities. While the impacts of such network induced measurement loss/delay on KF based state estimation have been extensively studied and well understood [28]-[32] , the corresponding effects on dynamic CS is still unknown. In this regard, there are a two key research questions that need to be properly addressed:

Question 3: *How a dynamic compressive sensing technique can be modified in presence of lossy and/or delayed measurements? What is the effect of the losses/delays on performance of the estimator?*

Many of the distribution system state estimation (DSSE) approaches rely on the assumption that the underlying topology is accurately identified. Therefore, a reliable topology identification method must be executed for control/estimation purposes. Furthermore, the time-varying nature of distribution systems implies that the topology identification process must be frequently performed if reliable results are expected from distribution systems operations. Conventional methods of topology identification suffer from high computational complexity [33]. To overcome this drawback, some other strategies propose to simultaneously estimate the switch statuses and system states [34]. However, these techniques require a considerable amount of measurements. Based on this challenge, we aim to answer the following question:

Question 4: *How can we formulate a compressive sensing method with low complexity that simultaneously estimates both the states and network topology?*

1.3 Contributions

To address the open questions highlighted in the previous section, this dissertation contains fundamental advances that are summarized below:

1. The goal of research Question 1 is to design a compressive sensing method for state estimation in a three phase unbalanced distribution system. To address Question 1, we extend the compressive sensing approach in [20] to a three phase unbalanced system with the following contributions.

- A forward backward algorithm based voltage state estimation for the three phase system is implemented using the reconstructed power measurements. Furthermore, we update our minimization problem with a linearized model of the power flow that relates the voltage phasors to the power values. In addition to directly estimating the voltages, this technique enables us to simultaneously employ both power and voltage measurements in our minimization problem.

- We evaluate the proposed approach using the IEEE 34-node and 37-node test systems that include different impedances for capturing the relationship between phases throughout the network. Additionally, we demonstrate that the choice of both sparsifying basis and projection matrix play critical roles in reliable data recovery. Our work demonstrates the feasibility of efficient smart meter/sensor data integration and its use for state estimation.

More details related to these contributions are in chapter 2 and has been published in the following articles:

[35] Hazhar Sufi Karimi and Balasubramaniam Natarajan. "Compressive sensing based state estimation for three phase unbalanced distribution grid." *In GLOBECOM 2017-2017 IEEE Global Communications Conference,* pages 1-6. IEEE, 2017

[36] Shweta Dahale, Hazhar Sufi Karimi, Kexing Lai, and Balasubramaniam Natarajan. "Sparsity based approaches for distribution grid state estimation a comparative study." *IEEE Access,* under review (2020).

2. Research Question 2 asks about implementing dynamic CS methods for distribution grids. To address this question, we provide the following contributions:

- we first exploit slow support set change and slow sparse signal change properties of the underlying data. To validate our assumptions, we incorporate practical home-level data from an actual distribution system.

- Then, we employ two recursive algorithms: 1- Streaming modified weighted-ℓ_1, 2- Kalman filtered CS. The simulation results (based on two different data framework) validate the efficiency of the recursive CS techniques where the performance of CS is significantly improved relative to the classic CS approach presented in [20].

These contributions are discussed in detail in chapter 3 and in the following article:
[37] Hazhar Sufi Karimi and Balasubramaniam Natarajan. "Recursive dynamic compressive sensing in smart distribution systems." *In 2020 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference (ISGT),* pages 1-5. IEEE, 2020

3. Research Question 3 deals with dynamic CS in the presence of delayed/lossy measurements. Our main contributions associated with addressing Question 3 are summarized below:

- To address dynamic CS in presence of lossy measurement, we consider a KF-Mod-CS based algorithm (**Algorithm 6** in [38]) where the goal is to recursively recover a sparse signal and estimate its corresponding support set.The KF-Mod-CS based algorithm consists of two main steps at each time: a compressive sensing algorithm to determine the support set using Kalman filter state estimates, and state estimation by Kalman filter using the estimated support set obtained from the modified CS. Therefore, the error in each process directly affects the other step. In this dissertation, we first quantify the dynamics of error in each step while considering loss of measurements.

- Using input-to-state stability analysis for discrete nonlinear systems, we derive an upper bound for the expected variances of the estimated signal. This upper bound enables us to find a critical value for the rate of receiving measurements ($\bar{\lambda}^c$) that

ensures the convergence of error in the KF-Mod-CS based algorithm. That is, if the probability of measurement loss is smaller than $1 - \bar{\lambda}^c$, we can ensure that the error in signal estimation is bounded. It is important to note that error and convergence analysis has not been investigated in the previous works, even for the case of Kalman filter CS without lossy measurements [38].

- We derive a more precise upper bound for the expected error variances when the underlying support set remains unaltered for a period of time. In this case, a reduced order Kalman filter is employed for estimation of the sparse signal.

- We uncover a steady state upper bound for the error covariance matrix if the underlying support set is constant and the reduced order model satisfies full observability condition. Simulation results provided in section 4.5 demonstrate the validity of the new fundamental theoretical results. We particularly evaluate our theoretical result using the IEEE 37-node test feeder with practical load/generation data.

- We propose a modified compressive sensing technique that incorporates the delayed measurements in the estimation process. To this end, we employ a well-known model for temporal correlation among states that enable us to project the delayed measurements to the current time.

- The dynamic model of the sparse states may not accurately capture the exact temporal correlation due to stochastic behavior of the signal as well as error in identification of the model parameters. Therefore, the uncertainty in the dynamic model affects the performance of the proposed signal recovery approach. To complete the analysis, we quantify the error in the recovered signal based on the errors in the models for temporal correlation.

More details related to these contributions are provided in chapter 4 and chapter 5 and published in the following articles:

[39] Hazhar Sufi Karimi and Balasubramaniam Natarajan. "Kalman filtered compressive sensing with intermittent observations." *Signal Processing,* 163:49,58, 2019.

[40] Hazhar Sufi Karimi and Balasubramaniam Natarajan. "Dynamic signal recovery in distribution grids using compressive lossy measurements." *IET Smart Grid*, under review (2020).

[41] Hazhar Sufi Karimi and Balasubramaniam Natarajan. "Dynamic signal recovery with compressive delayed measurements." *IEEE Signal Processing Letters,* under review (2020).

4. Research Question 4 seeks to find a method for joint state estimation and topology identification in distribution grids. To address this question, we provide the following contributions.

- For concurrent estimation of states and topology, we first formulate an optimization problem with states of systems and switch statuses as decision variables. Our framework projects the switch statuses on the underlying admittance matrix. The proposed formulation contains nonlinear relationships and integer variables leading to a mixed integer nonlinear programming (MINLP) problem.

- In order to guarantee the solvability of the optimization formulations and reduce MINLP complexity, we modify the nonlinear constraints to yield a mixed integer linear programming (MILP) problem. To this end, we transform the existing nonlinearity by introducing auxiliary variables.

- Although we could fix the nonlinearity issue by introducing MILP, the proposed technique is still a nonconvex problem. In other words, there is no guarantee to reach the optimal solution. To cope with this issue, we relax the nonlinearity of MINLP by replacing integer variables with continuous decision variables. Then, we apply an alternative minimization approach that improves both state estimation and topology identification.

- Finally, we introduce a hybrid dynamic framework that incorporates prior information about system topology for current state/switch status estimation. More precisely, a Markov jump model is proposed for switch status that helps topology identification especially for high levels of compression. Besides, we employ the previous information

about the support set that improves the fidelity of our method.

- Simulation results provided in section 6.5 demonstrate the performance of the proposed approaches on the IEEE 37-node test feeder with practical load/generation data.

More details related to these contributions are in chapter 6 and in the following article: [42] Hazhar Sufi Karimi and Balasubramaniam Natarajan. "Joint topology identification and state estimation in unobservable distribution grids." *IEEE Transactions on Smart Grid,* to be submitted (2020).

1.4 Organization of This Dissertation

The remainder of this dissertation is organized as follows. Chapter 2 describes the fundamental idea of sparsity-based data aggregation and voltage state estimation in three phase distribution grids. In chapter 3, we exploit the basic properties for implementing dynamic CS in distribution systems. Then, two recursive dynamic CS methods are employed for signal recovery. Chapter 4 is dedicated to the analysis of Kalman filter modified CS in the presence of lossy measurements. In chapter 5, we design a dynamic CS method for the system with delayed measurements. Then, chapter 6 provides two approaches for joint state estimation and topology identification in distribution systems. The summary of this dissertation and future ideas constitute chapter 7.

Chapter 2

Background

The power distribution grid is typically unobservable due to a lack of measurements. While deploying more sensors can alleviate this issue, it also presents new challenges related to data aggregation and the underlying communication infrastructure. Therefore, developing state estimation methods that enhance situational awareness at the grid edge with compressed measurements is critical. In this chapter, we focus our attention on fundamental idea of the sparse-aware data aggregation and voltage state estimation in three phase distribution grids.

2.1 Related Work and Contributions

Recently, the challenges posed by unobservability and limited measurement availability at the grid edge has been addressed by sparse recovery methods. All these methods exploit the underlying smoothness or sparsity of the raw or linearly transformed measurements/system states. These methods exploit the network structure to perform state estimation at current levels of data availability and observability. Therefore, the requirement of creating pseudo-measurements for ensuring observability is eliminated. Compressive Sensing (CS) based distribution system state estimation (DSSE) was the first class of solutions proposed where the sparsity of measured data in a linear transformation basis was exploited to compress measurements. The spatio-temporal correlation between loads and distributed generation in

a single-phase distribution system is exploited in the estimation strategies proposed in [20]. A singular value decomposition based data compression approach is investigated for a smart distribution grid in steady state [16]. A DSEE method has been provided in [17], which uses micro-phasor measurement units and (1-D) spatial compressive sensing. However, [17] does not consider smart meter data and given the fact that the number of smart meters is typically far more than micro PMUs, the proposed approach is limited in its practical application. A temporal compression scheme is employed in [43] on three publicly available datasets for spatial and spatio-temporal compression. A robust tensor completion based state estimation method is proposed in [44]. Recursive dynamic state estimation is implemented in [37] by exploiting the sparsity in distribution grid data.

In contrast to the single phase systems [20], three phase networks may not be balanced, (i.e., each phase could have different load and generation profiles) and exhibit coupling across phases. Thus, the analysis of a single phase system cannot be trivially generalized to a three phase system. In this chapter, we extend the compressive sensing approach to a three phase unbalanced system. Specifically, we exploit the underlying correlation among the electrical parameters (power/voltage) and the resulting sparsity in a discrete wavelet transform to reconstruct these parameter values for the entire grid using limited number of measurements. A forward backward algorithm based voltage state estimation for the three phase system is implemented using the reconstructed power measurements. Furthermore, we update our minimization problem with a linearized model of the power flow that relates the voltage phasors to the power values. In addition to directly estimating the voltages, this technique enables us to simultaneously employ both power and voltage measurements in our minimization problem. We evaluate the proposed approach using the IEEE 34-node and 37-node test systems that include different impedances for capturing the relationship between phases throughout the network. Additionally, we demonstrate that the choice of both sparsifying basis and projection matrix play critical roles in reliable data recovery. Specifically, we show that a Bernoulli projection matrix allows for more accurate reconstruction relative to a Gaussian projection matrix that was used in [20]. Furthermore, among the discrete wavelet sparsifying bases, "Haar" mother wavelet shows considerably

better performance comparing to other bases. Our work demonstrates the feasibility of efficient smart meter/sensor data integration and its use for state estimation even in the presence of communication network impairments (e.g., missing data) that currently limit the use of this information in real time operations.

The rest of this chapter is organized as follows: Section 2.2 introduces the three phase distribution grid model, linearized power flow, and the underlying correlation structure. In section 2.3 a brief introduction of compressive sensing and relevant mathematical fundamentals are provided. Section 2.4 focuses on state estimation in three phase unbalanced distribution grids by presenting two methods. Then, section 2.5 provides simulation results that demonstrate the potency of the proposed approach in the context of IEEE 34-node and IEEE 37-node benchmark systems. The summary and direction of the next chapter constitute section 2.7.

2.2 Distribution Grid Model

Consider a general three phase distribution grid with $\mathcal{J}$ nodes. For example, Figure 2.1 shows a diagram of the IEEE 34-node test feeder. Another example of distribution network is IEEE-37 node test feeder shown in Figure 2.2. The substation/source node is considered the 0^{th} node (slack bus) where the voltage level is assumed to be fixed or regulated (typically 1 per unit). In this three phase system, the voltage vector at i^{th} node denoted as a vector $\mathbf{V}_i^{abc}$, corresponds to:

$$\mathbf{V}_i^{abc} = \begin{bmatrix} \mathbf{V}_i^a \\ \mathbf{V}_i^b \\ \mathbf{V}_i^c \end{bmatrix} = \begin{bmatrix} \mid \mathbf{V}_i^a \mid e^{j\theta_i^a} \\ \mid \mathbf{V}_i^b \mid e^{j\theta_i^b} \\ \mid \mathbf{V}_i^c \mid e^{j\theta_i^c} \end{bmatrix}, \tag{2.1}$$

where, $\mathbf{V}_i^*$ represents voltage phasor of phase *. $\mid \mathbf{V}_i^* \mid$ and θ_i^* are magnitude of the voltage and angle of the voltage, respectively. It should be noted that $\mathbf{V}_0^{abc}$ is fixed or regulated, thus it is assumed to be known. The ultimate goal is to estimate both magnitude and angle of voltages for each node based on power measurements collected from smart meters/sensors.

Therefore, the voltage state vector for the entire distribution grid corresponds to a $6\mathcal{J}$-dimensional vector,

$$\mathcal{V} = \begin{bmatrix} |\mathbf{V}_1^a| & |\mathbf{V}_1^b| & |\mathbf{V}_1^c| & \dots & |\mathbf{V}_{\mathcal{J}}^a| & |\mathbf{V}_{\mathcal{J}}^b| & |\mathbf{V}_{\mathcal{J}}^c| & , & \theta_1^a & \theta_1^b & \theta_1^c & \dots & \theta_{\mathcal{J}}^a & \theta_{\mathcal{J}}^b & \theta_{\mathcal{J}}^c \end{bmatrix}' \quad (2.2)$$

The relationship between voltages and currents of a $\mathcal{J}$-bus distribution system can be easily expressed as:

$$\begin{bmatrix} \mathbf{V}_1^{abc} \\ \mathbf{V}_2^{abc} \\ \vdots \\ \mathbf{V}_{\mathcal{J}}^{abc} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{11}^{abc} & \mathbf{Z}_{12}^{abc} & \cdots & \mathbf{Z}_{1\mathcal{J}}^{abc} \\ \mathbf{Z}_{21}^{abc} & \mathbf{Z}_{22}^{abc} & \cdots & \mathbf{Z}_{2\mathcal{J}}^{abc} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{Z}_{\mathcal{J}1}^{abc} & \mathbf{Z}_{\mathcal{J}2}^{abc} & \cdots & \mathbf{Z}_{\mathcal{J}\mathcal{J}}^{abc} \end{bmatrix} . \begin{bmatrix} \mathbf{I}_1^{abc} \\ \mathbf{I}_2^{abc} \\ \vdots \\ \mathbf{I}_{\mathcal{J}}^{abc} \end{bmatrix} \quad (2.3)$$

where $\mathbf{Z}_{ij}^{abc}$ is the impedance matrix between i^{th} node and j^{th} node, and $\mathbf{I}_i^{abc}$ is the vector of three phase current corresponding to i^{th} node. For every phase at each node, we assume that smart meters can measure the real and reactive power. Of course, the measured real and reactive power in each phase, at particular node-i depends on the load and generation at that node as given by: $\mathbf{P}_i^a = Real(\mathbf{S}_i^a) = Real(\mathbf{S}_{Gi}^a - \mathbf{S}_{Li}^a)$ and $\mathbf{Q}_i^a = Imag(\mathbf{S}_i^a) = Imag(\mathbf{S}_{Gi}^a - \mathbf{S}_{Li}^a)$. Here, $\mathbf{S}_{Gi}^a$ represents apparent power generation and $\mathbf{S}_{Li}^a$ is the load at i^{th} node for phase-a. The apparent power is related to voltage and current as $\mathbf{S}_i^a = \mathbf{V}_i^a(\mathbf{I}_i^a)^*$. Similar expressions can be written for phases b and c, for each node. Accordingly, the power state vector contains both real and reactive power at all nodes represented as a $6N$-dimensional vector,

$$\mathcal{P} = \begin{bmatrix} \mathbf{P}_1^a & \mathbf{P}_1^b & \mathbf{P}_1^c & \dots & \mathbf{P}_{\mathcal{J}}^a & \mathbf{P}_{\mathcal{J}}^b & \mathbf{P}_{\mathcal{J}}^c & , & \mathbf{Q}_1^a & \mathbf{Q}_1^b & \mathbf{Q}_1^c & \dots & \mathbf{Q}_{\mathcal{J}}^a & \mathbf{Q}_{\mathcal{J}}^b & \mathbf{Q}_{\mathcal{J}}^c \end{bmatrix}' . \quad (2.4)$$

Here, let $\mathcal{X}$ denote the entire state vector and it is defined as $\mathcal{X} = \begin{bmatrix} \mathcal{P}', \mathcal{V}' \end{bmatrix}'$. Obviously, DSSE techniques require an exact knowledge about the relationship between the power values and voltage phasors of the different nodes. However, the function f_{PF} that captures

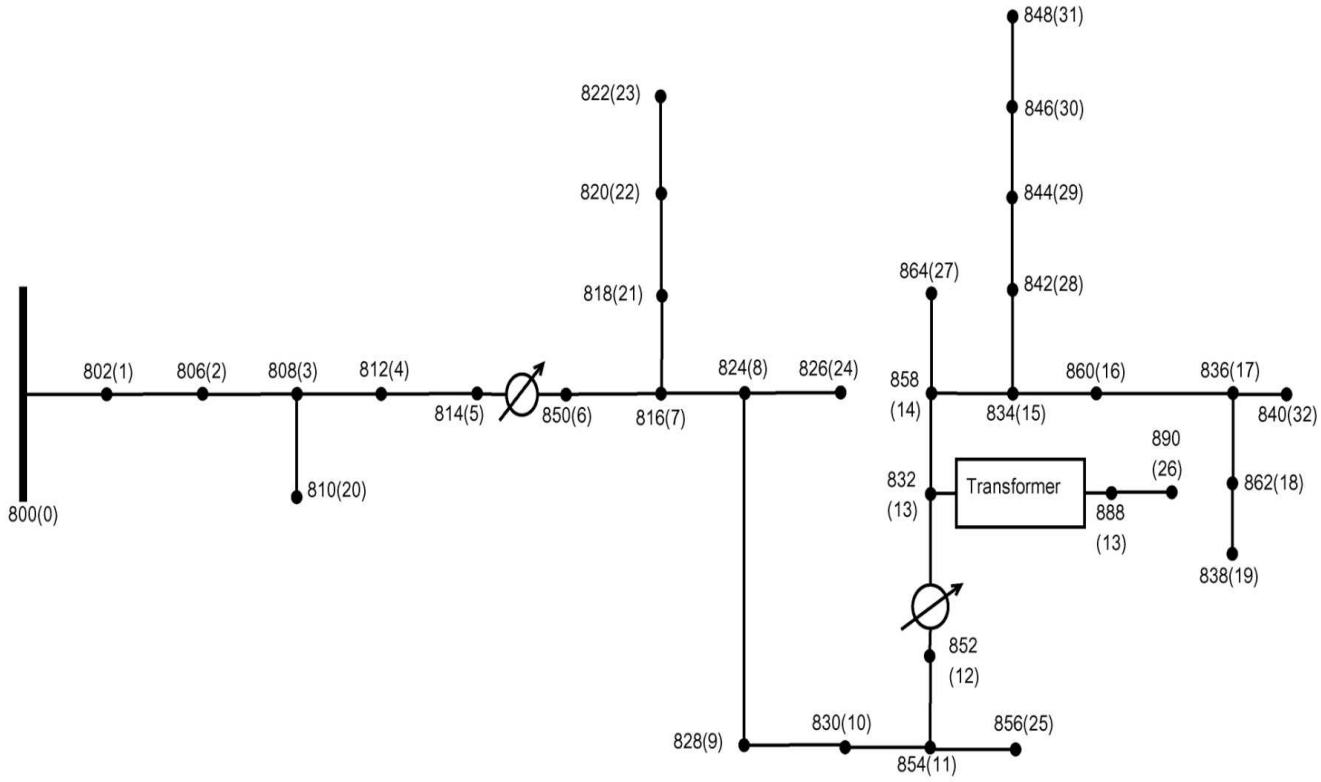

Figure 2.1: *IEEE 34-node Test Feeder*

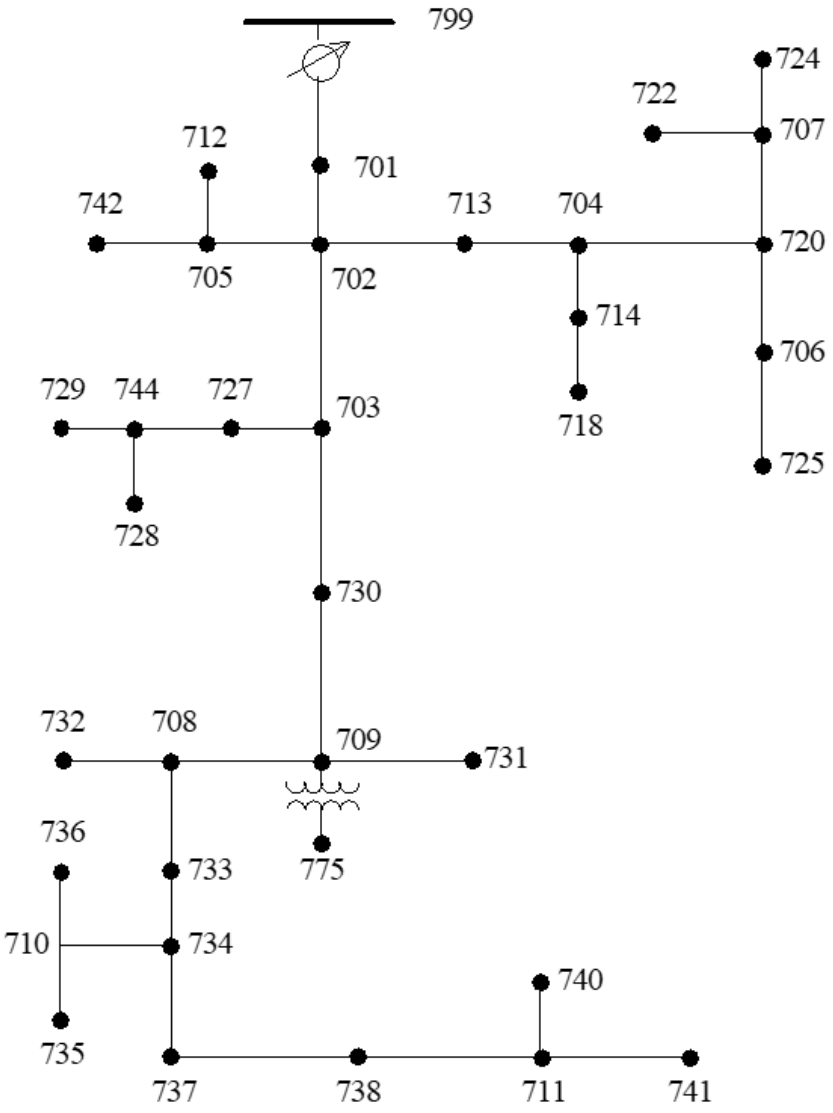

Figure 2.2: *IEEE 37-node Test Feeder*

the relationship between voltage states and power injected at the nodes is not linear.

$$\mathcal{V} = f_{PF}(\mathcal{P})$$

The nonlinearity of the AC power-flow equations poses significant challenges for the development of computationally affordable state estimation tasks. To avoid such complexities, we employ a linearized approximation that is applicable to multiphase distribution networks [45]. First, we write the voltages in a complex form rather than a phasor/polar format; Thus, if we have $\mathbf{v}_i^* = Real(\mathbf{V}_i^*) + jImage(\mathbf{V}_i^*)$, the voltage vector is:

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_1^a & \mathbf{v}_1^b & \mathbf{v}_1^c & \dots & \mathbf{v}_{\mathcal{J}}^a & \mathbf{v}_{\mathcal{J}}^b & \mathbf{v}_{\mathcal{J}}^c \end{bmatrix} \tag{2.5}$$

For a three-phase distribution system with $\mathcal{J}$ buses, $\mathbf{Y}$ denotes the admittance matrix: $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{00} & \mathbf{Y}_{0L} \\ \mathbf{Y}_{L0} & \mathbf{Y}_{LL} \end{bmatrix}$ where, $\mathbf{Y}_{00} \in \mathbb{C}^{3\times3}$, $\mathbf{Y}_{L0} \in \mathbb{C}^{3(\mathcal{J}-1)\times3}$, $\mathbf{Y}_{0L} \in \mathbb{C}^{3\times3(\mathcal{J}-1)}$ and $\mathbf{Y}_{LL} \in \mathbb{C}^{3(\mathcal{J}-1)\times3(\mathcal{J}-1)}$ are submatrices of the admittance matrix. Denoted by $\tilde{\mathbf{v}}$, the linearized voltage is:

$$\mathbf{v} \approx \mathbf{M}\mathcal{P} + \mathbf{w} \tag{2.6}$$

where,

$$\mathbf{M} = \begin{bmatrix} \mathbf{Y}_{LL}^{-1} diag(\tilde{\hat{\mathbf{v}}})^{-1} & -j\mathbf{Y}_{LL}^{-1} diag(\tilde{\hat{\mathbf{v}}})^{-1} \end{bmatrix} \tag{2.7}$$

$\mathbf{w} = -\mathbf{Y}_{LL}^{-1}\mathbf{Y}_{L0}\mathbf{v}_0$. Here, $\mathbf{v}_0$ is the slack bus voltage where the voltage level is assumed to be fixed or regulated (typically 1 per unit). $\tilde{\hat{\mathbf{v}}}$ denotes voltage in an operating point. One may consider $\mathbf{w} = \tilde{\hat{\mathbf{v}}}$.

The goal of state estimation is to estimate $\mathcal{X}$ using the available measurements. If all the elements in $\mathcal{P}$ are available without errors, the distribution grid is fully observable and estimation of $\mathcal{X}$ is trivial [7]. However, the communication network imposes two main con-

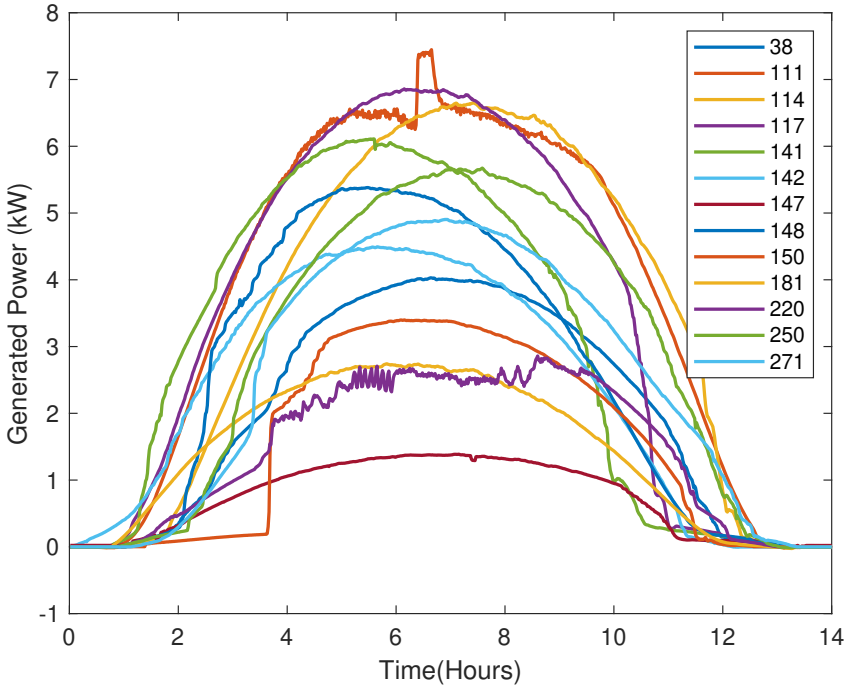

Figure 2.3: *PV generations of 13 representative homes*

straints in this process of estimation of $\mathcal{X}$. First, as the size of the network grows (i.e., $\mathcal{J}$ increases), the volume of data that needs to be aggregated grows with it. This puts a significant stress on the underlying communication network. Secondly, if the communication network is unreliable, there could be missing measurements making the system underdetermined. Both these issues can be effectively dealt by exploiting the underlying correlation in measurements. In a distribution grid, nodes may possess various electrical characteristics, e.g., some nodes may have only loads, whereas some other nodes, in addition to loads, may have photovoltaic (PV) panels or wind generators (both representing renewable energy sources). As a result of short distances between nodes in a distribution grid, one can expect a certain level of correlation between generation characteristics of the nodes. To verify the presence of spatial correlation in a distribution system, we collect daily PV generation profiles of 13 representative homes with 1-min time resolution from eGauge website [2], representing a typical day in Summer 2019. Figure 2.3 presents 14-hour PV generation profiles of 13 representative homes, where we can observe a similar pattern for different homes.

The correlation across space and time of contributed generation leads to sparsity in a wavelet domain which can be exploited to effectively estimate $\mathcal{X}$ from regular/compressed measurements. It is important to note that the proposed compressive sensing approach described in the following section is applicable to any underlying correlation and load models. However, for our analysis, we consider a previously suggested correlation [46] and load model [47] as an example for demonstrating the potency of our proposed approach. As in [46], the correlation among the generated powers at i^{th} node and j^{th} node can be expressed via an exponential function of distance, parameterized with α_1 and α_2:

$$\mathcal{R}_{ij} = exp\left\{-\left(\frac{\mathcal{D}_{ij}}{\alpha_1}\right)^{\alpha_2}\right\};\ \alpha_1 > 0,\ \alpha_2 \in (0, 2] \tag{2.8}$$

where, $\mathcal{R}_{ij}$ is the coefficient of correlation and $\mathcal{D}_{ij}$ is the Euclidean distance between the i^{th} and j^{th} node.

Similarly, a temporal correlation is common for both generated power and loads. Figure 2.4 shows PV generation profiles of Home 141 for three consecutive days during summer 2019. To provide a comprehensive and quantitative correlation analysis, table 2.1 shows the correlation matrix for PV generation at five consecutive hours (13:01 to 13:05) of home 141. This correlation is calculated based on the data collected from 112 days.

Table 2.1: *Temporal correlation of PV generation at home 141*

Time	13:01	13:02	13:03	13:04	13:05
13:01	1	0.90993	0.861113	0.811539	0.73272
13:02	0.90993	1	0.940018	0.861275	0.777457
13:03	0.861113	0.940018	1	0.936477	0.843463
13:04	0.811539	0.861275	0.936477	1	0.928608
13:05	0.73272	0.777457	0.843463	0.928608	1

Therefore, we model the distributed generation and load as first order auto-regressive (AR(1)) process [47] given by:

$$\mathbf{S}_{t+1} - \mathbf{D}_{t+1} = \alpha_3(\mathbf{S}_t - \mathbf{D}_t) + \mathcal{W}_t \tag{2.9}$$

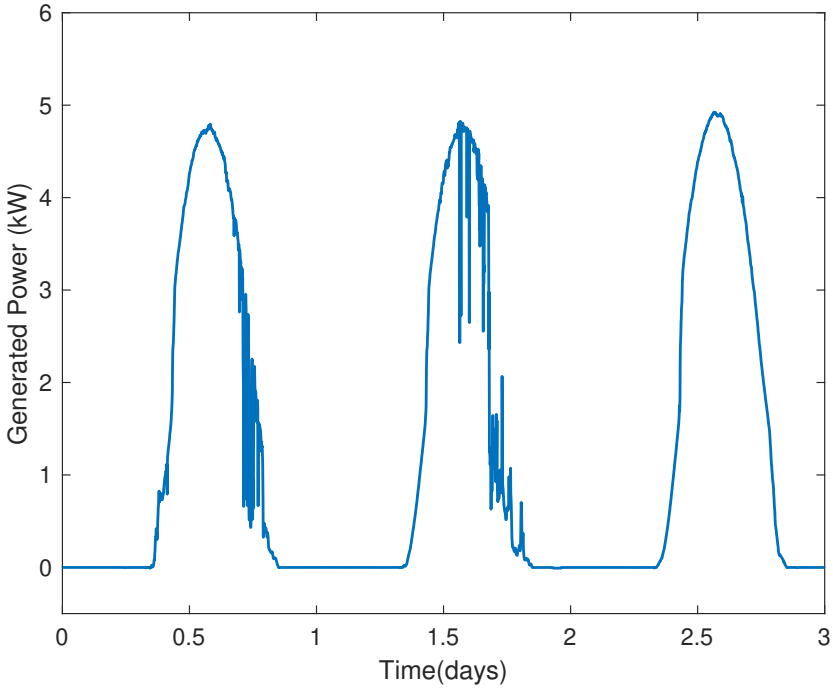

Figure 2.4: *PV generations Home 141 for three consecutive days during Summer 2019*

where, $\mathbf{S}_t$ denotes apparent power for both generation and loads at time t; and α_3 denotes the AR(1) coefficient; $\mathcal{W}_t$ is the corresponding noise or model uncertainty component. $\mathbf{D}_t$ is the deterministic trend component of the time series that also captures any periodic variation that may exist.

2.3 Information Aggregation in Distribution Grid

The necessary condition for using sparsity-based data recovery methods is that the signal of interest exhibits sparsity or must be approximately sparse in a linear transformation basis.

2.3.1 1-D Compressive Sensing

The following theorem describes the concept of 1-D compressive sensing and its proof can be found in [48]:

Theorem 1. *Let* $\mathbf{z} \in \mathbb{R}^n$ *be the signal of interest, which is sparse in a linear transformation*

basis $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$ *such that,*

$$\mathbf{z} = \mathbf{\Psi x}$$

where, $\mathbf{x}$ *has at most* $S \ll n$ *significant coefficients i.e.,* $\mathbf{z}$ *is* S*-sparse in the sparsifying basis* $\mathbf{\Psi}$*. If the sensing mechanism is such that:*

$$\mathbf{h} = \mathbf{\Phi z} \tag{2.10}$$

where, $\mathbf{h} \in \mathbb{R}^m$ *is the available measurement vector and* $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ *is a random measurement/projection matrix (e.g., matrix elements distributed as i.i.d. Gaussian random variables with mean* 0 *and variance* $1/m$ *or Bernoulli random variables), then the original signal* $\mathbf{z}$ *can be reconstructed by solving the following* ℓ_1 *minimization problem,*

$$\hat{\mathbf{x}} = arg \min_{\mathbf{b}} \| \mathbf{b} \|_1 \quad subject\ to\ \mathbf{h} = \mathbf{\Phi\Psi b}$$

$$\hat{\mathbf{z}} = \mathbf{\Psi}\hat{\mathbf{x}} \tag{2.11}$$

The result of the optimization problem in (2.11) provides an exact reconstruction with overwhelming probability if there exists a $\delta \in (0, 1)$ such that,

$$(1 - \delta) \| \mathbf{x} \|_2^2 \leq \| \mathbf{\Phi\Psi x} \|_2^2 \leq (1 + \delta) \| \mathbf{x} \|_2^2,$$

holds for all K-sparse signal $\mathbf{x}$. This is called the Restricted Isometry Property (RIP) of order K. Correspondingly, the measurement dimension m is bounded in the following order

[12],

$$\begin{aligned} m &= \mathcal{O}\left(\frac{1}{\delta^2} K \; log\left(\frac{n}{K}\right)\right) \\ \Rightarrow \left(\frac{m}{n}\right) &= \mathcal{O}\left(\frac{1}{\delta^2}\frac{K}{n} \; log\left(\frac{n}{K}\right)\right) \\ \Rightarrow CMR &= \mathcal{O}\left(\frac{1}{\delta^2} SR \; log\left(\frac{1}{SR}\right)\right) \end{aligned} \tag{2.12}$$

where CMR and SR are compressed measurement and sparsity ratio, respectively.

Remark. *In practice, the available measurements have different level of errors due to measurement noises. Therefore, instead of* $\mathbf{h}$*, the available measurement vector corresponds to,*

$$\mathbf{y} = \mathbf{h} + \boldsymbol{\epsilon} \tag{2.13}$$

where, $\boldsymbol{\epsilon}$ *represents the measurement noise vector. That implies that the equality constraint is not valid anymore. Instead, the difference between* $\mathbf{y}$ *and* $\mathbf{\Phi\Psi b}$ *must be as low as possible. To impose such a condition, we modify (2.11) in to the following form:*

$$\hat{\mathbf{x}} = arg \min_{\mathbf{b}} \| \mathbf{b} \|_1 + \lambda \| \mathbf{y} - \mathbf{\Phi\Psi b} \|_2 \tag{2.14}$$

where, λ *is a tuning parameter that controls the trade-off between sparsity of the underlying signal and the error residual.*

2.3.2 2-D Compressive Sensing

If compressive sensing is applied both spatial and temporal signals simultaneously, the following corollary is applicable [49]:

Corollary 2. *Let* $\mathbf{W} \in \mathbb{R}^{n_{space} \times n_{time}}$ *be the spatio-temporal data over* n_{space} *nodes and for* n_{time} *number of observations.* $\mathbf{W}$ *is sparse in sparsifying basis* $\mathbf{\Psi}_{n_{space}}$ *and* $\mathbf{\Psi}_{n_{time}}$ *such that,*

$$\mathbf{W} = \mathbf{\Psi}_{n_{space}} \mathbf{A} \mathbf{\Psi}^T_{n_{time}}$$

The spatio-temporal compressed sensing of $\mathbf{W}$ *is,*

$$\mathbf{H} = \mathbf{\Phi}_{space}\mathbf{W}\mathbf{\Phi}_{time}^{T},$$

$$\mathbf{\Phi}_{space} \in \mathbb{R}^{m_{space} \times n_{space}},\ \mathbf{\Phi}_{time} \in \mathbb{R}^{m_{time} \times n_{time}}$$

$$m_{space} \ll n_{space}\ and\ m_{time} \ll n_{time} \tag{2.15}$$

where, $\mathbf{\Phi}_{space}$ *and* $\mathbf{\Phi}_{time}$ *are random measurement/projection matrices (e.g., matrices with elements distributed as i.i.d. Gaussian random variables with zero mean and respective variance of* $1/{m_{space}}^2$ *and* $1/{m_{time}}^2$ *or Bernoulli random variables). The spatio-temporal data is recovered if the following* ℓ_1 *minimization problem is solved,*

$$\hat{\mathbf{X}} = arg\min_{\mathbf{z}} \parallel \mathbf{Z} \parallel_1$$

$$subject\ to\ vec(\mathbf{H}) = \left(\mathbf{\Phi}_{space} \otimes \mathbf{\Phi}_{time}\right)\left(\mathbf{\Psi}_{n_{space}} \otimes \mathbf{\Psi}_{n_{time}}\right)\mathbf{Z} \tag{2.16}$$

Here, $vec(.)$ *represents the vectorized version of a matrix and* $\otimes$ *represents the kronecker product. The reconstructed state is therefore,*

$$vec(\hat{\mathbf{W}}) = \mathbf{\Psi}_{n_{space}} \otimes \mathbf{\Psi}_{n_{time}}\hat{\mathbf{X}} \tag{2.17}$$

According to Theorem 1, a random projection matrix $\mathbf{\Phi}_{projection}\colon \mathbb{R}^n \rightarrow \mathbb{R}^m$ is utilized to make individual compressive measurements of real and reactive powers for an $\mathcal{J}$-node power distribution system (Here, $n = 3\mathcal{J}$); i.e., $\mathbf{h}_P = \mathbf{\Phi}_{projection}\begin{bmatrix}\mathbf{P}_1^{abc} & \cdots & \mathbf{P}_{\mathcal{J}}^{abc}\end{bmatrix}'$; $\mathbf{h}_Q = \mathbf{\Phi}_{projection}\begin{bmatrix}\mathbf{Q}_1^{abc} & \cdots & \mathbf{Q}_{\mathcal{J}}^{abc}\end{bmatrix}'$; $\mathbf{h}_{\mathbf{v}} = \mathbf{\Phi}_{projection}\begin{bmatrix}\mathbf{v}_1^{abc} & \cdots & \mathbf{v}_{\mathcal{J}}^{abc}\end{bmatrix}'$. Due to the inherent (spatial, temporal, spatio-temporal) correlation between real and reactive powers, these measurements are approximately sparse on respective linear transformation basis $\mathbf{\Psi}_P$ and $\mathbf{\Psi}_Q$ (e.g. a wavelet basis [11]). Therefore, the relationship between $\mathbf{h}$, $\mathbf{h}_P$, $\mathbf{h}_Q$ and $\mathbf{h}_{\mathbf{v}}$ can be

written down as: $\mathbf{h} = \begin{bmatrix} \mathbf{h}_P{}' & \mathbf{h}_Q{}' & \mathbf{h}_\mathbf{v}{}' \end{bmatrix}'$, and $\mathbf{h} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{a}$. Where,

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_P & \mathbf{O}_{m\times n} & \mathbf{O}_{m\times n} \\ \mathbf{O}_{m\times n} & \mathbf{\Phi}_Q & \mathbf{O}_{m\times n} \\ \mathbf{O}_{m\times n} & \mathbf{O}_{m\times n} & \mathbf{\Phi}_\mathbf{v} \end{bmatrix}, \ \mathbf{\Psi} = \begin{bmatrix} \mathbf{\Psi}_P & \mathbf{O}_{m\times n} & \mathbf{O}_{m\times n} \\ \mathbf{O}_{m\times n} & \mathbf{\Psi}_Q & \mathbf{O}_{m\times n} \\ \mathbf{O}_{m\times n} & \mathbf{O}_{m\times n} & \mathbf{\Psi}_\mathbf{v} \end{bmatrix}.$$

Directly applying theorem 1, i.e., ℓ_1 minimization approach, we can reconstruct an n-dimensional power measurements based on a m-dimensional ($m \ll n$) projection information. We elaborate the impact of actual implementation of this idea using two examples related to spatial compressive sensing.

First, let's assume our $\mathbf{\Phi}$ matrix consists of i.i.d. Gaussian random variables with zero mean and variance $1/m$. in this case, the data aggregation process is very well suited to what is referred to as serial information fusion. Here, each sensor or smart meter can take their individual measurement and weight it using random Gaussian weights (initiated by the control center) and send it to the next sensor/aggregation unit. The next sensor can linearly combine their randomly weighted measurements to the information received and send the fused lower dimensional data onto the next aggregation point. This stagewise aggregation/fusion process prevents the data explosion problem that is common in large scale networks, and reduces the stress on communication bandwidth/rate. The compressed/fused information at the control center can be used for the complete data recovery based on the ℓ_1 minimization approach.

Alternately, if we let our $\mathbf{\Phi}$ matrix consists of Bernoulli distributed random variables (with a parameter equal to 0.5) elements, then we have a data aggregation process well suited to a parallel information fusion set up. Here, each sensor has the ability to directly send their measurements to the control center. However, the control center randomly picks a subset of measurements to create a compressed measurement space which can be used to recover entire measurement set based on the ℓ_1 minimization approach. This parallel information fusion is especially effective when the underlying communication network has faults, dropped packets, etc. that result in missing data. The randomness of such communication failures

can be naturally handled in this form of data aggregation.

2.4 Voltage State Estimation

In distribution systems, voltage estimation is a more difficult task compared to power estimation since the voltage measurements are extremely rare at distribution level. Therefore, we employ two compressed sensing based techniques to estimate voltages from a smaller number of measurements. In the first method, all the power values are first reconstructed. Then, the voltages are indirectly estimated by a forward-backward algorithm. In the second approach, the linearized power flow presented by (2.6) is incorporated in the ℓ_1 minimization problem, where we directly obtain the voltage estimates.

2.4.1 Indirect Method

Algorithm 1 describes a forward-backward approach to estimate the complex voltage phasors using reconstructed power information. First, all the recovered real and reactive powers are embedded in a vector $\mathbf{S}^{abc}$. Since we assume that there is no prior information about node voltages, a flat initial profile is considered. In the first iteration, all phase currents flowing through each node are calculated using the corresponding recovered apparent power and the initialized voltage values. Using Kirchhoff current law, currents flowing between nodes are then calculated. Finally, voltage states are estimated starting from the substation node to the end node using the computed currents and impedances between nodes. The new voltage serves as the basis for current calculations for the next iteration. These iterations continue until the voltage states do not change significantly (indicating convergence) when new current values are computed.

2.4.2 Direct Method

In the direct method, we consider both the power values and voltages in our signal of interest. Here, we also consider the linearized power flow. Therefore, both the power and voltage can

Algorithm 1 Iterative Estimation of Voltage

Step 1. Collect the CS based reconstructed data into a vector $\mathbf{S}^{abc}$

Step 2. Assume a flat profile for all node voltage vectors in the first iteration, that is:
$\mathbf{V}_1^{abc(0)} = ... = \mathbf{V}_{32}^{abc(0)} = \mathbf{V}_0^{abc(0)}$

Step 3. Set iteration counter $k = 0$.

Step 4. Find $\mathbf{I}_1^{abc(k)}, ..., \mathbf{I}_i^{abc(k)}, ..., \mathbf{I}_{32}^{abc(k)}$ from the following equation:

$$\mathbf{I}_i^{abc(k)} = \begin{bmatrix} \mathbf{I}_i^{a(k)} \\ \mathbf{I}_i^{b(k)} \\ \mathbf{I}_i^{c(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{S}_i^{a} / \mathbf{V}_i^{a(k)} \\ \mathbf{S}_i^{b} / \mathbf{V}_i^{b(k)} \\ \mathbf{S}_i^{c} / \mathbf{V}_i^{c(k)} \end{bmatrix} \tag{2.18}$$

Step 5. Find $\mathbf{I}_{ij}^{abc(k)}$ by using $\mathbf{I}_i^{abc(k)}$ and $\mathbf{I}_j^{abc(k)}$.

Step 6. Determine the voltage drop in each feeder section by proceeding from source to load to determine $\mathbf{V}_j^{abc(k+1)}$ using:
$\mathbf{V}_j^{abc(k+1)} = \mathbf{V}_i^{abc(k+1)} - [Z_{P_{ij}}].\mathbf{I}_{ij}^{abc(k)}$

Step 7. Check if the error between current value and the previous value is smaller than a threshold. i.e:
$\epsilon_i^{abc(k)} = \mid \mathbf{V}_i^{abc(k+1)} - \mathbf{V}_i^{abc(k)} \mid$

if $\epsilon_i^{abc(k)}$ is smaller than the tolerance value, the final value is the estimated value. If not, set $k = k + 1$, then repeat steps from 4 to 7.

be estimated by solving the following optimization method.

$$\hat{\mathbf{x}} = arg\min_{\mathbf{b}} \parallel \mathbf{b} \parallel_1 + \lambda \parallel \mathbf{y} - \mathbf{\Phi\Psi b} \parallel_2 + \mu \parallel \mathbf{v} - \mathbf{M}\mathcal{P} - \mathbf{w} \parallel_2 \tag{2.19}$$

where, μ is a tuning parameter that controls enforcement of the linearized power flow. It should be noted that the vector $\mathbf{y}$ can contain both power and voltage measurements or only power measurements

2.5 Simulation Results

In this section, we evaluate the proposed estimation methods by considering the IEEE 34-node for indirect methods and 37-node distribution test feeder for direct method. For each of the cases, we take different number of random measurements and use (2.11) and (2.17) to reconstruct the original signals. The ℓ_1 minimization problems in (2.11) and (2.17) are solved using `cvx` [50]. Initially, we use the "Haar" mother wavelet as the sparsifying basis $\mathbf{\Psi}$. We also examine other wavelet bases and their performance is compared later in this section. In addition to the sparsifying basis, selection of an appropriate projection matrix also plays a critical role in achieving an acceptable fidelity in signal recovery.

We investigate the performance of the three proposed compressed sensing based reconstruction approaches by performing 150 Monte Carlo simulation at different CMRs. The performance of spatial, temporal and spatio-temporal compressed sensing is compared using a percentage ratio, called *Integrated Normalized Absolute Error (INAE)*. For 1-D (spatial or temporal) case, *INAE* is defined as,

$$INAE = \frac{\sum_{j=1}^{n} |\mathbf{x}_j - \hat{\mathbf{x}}_j|}{\sum_{j=1}^{n} |\mathbf{x}_j|} \times 100 \tag{2.20}$$

where, $\mathbf{x}_j$ is the actual real power at j^{th} node (spatial) or at the j^{th} time instant (temporal) and $\hat{\mathbf{x}}_j$ is the corresponding estimate of $\mathbf{x}_j$ recovered by compressive sensing. similarly, For 2-D (spatio-temporal) case, *INAE* corresponds to

$$INAE = \frac{\sum_{j=1}^{n_{space}} \sum_{k=1}^{n_{time}} |\mathbf{x}_{jk} - \hat{\mathbf{x}}_{jk}|}{\sum_{j=1}^{n_{space}} \sum_{k=1}^{n_{time}} |\mathbf{x}_{jk}|} \times 100 \tag{2.21}$$

where, $\mathbf{x}_{jk}$ is the actual real power generated at j^{th} node, which is measured at the j^{th} time instant. In this thesis, we sometimes use the following metric, too.

$$MAPE = \frac{1}{n} \sum_{j=1}^{n} \left| \frac{\mathbf{x}_j - \hat{\mathbf{x}}_j}{\mathbf{x}_j} \right| \times 100 \tag{2.22}$$

$$MIAE = \sum_{j=1}^{n} \frac{|\mathbf{x}_j - \hat{\mathbf{x}}_j|}{n} \times 100 \tag{2.23}$$

More precisely, MAPE is the average magnitude of the normalized errors produced by our estimation technique. Now that we have to scale everything by the actual value, MAPE is undefined for data points where the value is 0 or it can grow unexpectedly large if the actual values are exceptionally small themselves. This problem is especially highlighted for the sparse states which most of them are small values. Unlike MAPE, INAE does not require the signal values to be non-zero. Similar to MAPE, The lower INAE is, the better is the reconstruction performance. Both statistic measure show a kind of percentage error.

2.5.1 Indirect Method

We assume that the base kVA and base kV for IEEE 34-node are 100 and 24.9, respectively. For IEEE 37-node, the base kVA is 100 and the base kV is 48. We assume that every node has distributed generation with a deterministic trend $\mathbf{D}_t = 10{+}5\sin(6.4\pi t){+}4\cos(6.4\pi t - \pi)$ and the deterministic trend for load is $\mathbf{D}_t = 14{+}5\sin(6.4\pi t - 0.1\pi){+}0.025\cos(6.4\pi t - 0.8\pi)$. We consider spatial, temporal and spatio-temporal compressed sensing approaches to effectively aggregate and recover the power measurements. As the name suggests, in spatial (temporal) compressive sensing, we only exploit the correlation of data across the spatial (temporal) domain. In spatio-temporal compressive sensing, the correlation across both spatial and temporal domain are exploited.

Fig.2.5 demonstrates INAE performance associated with spatial compressive sensing as

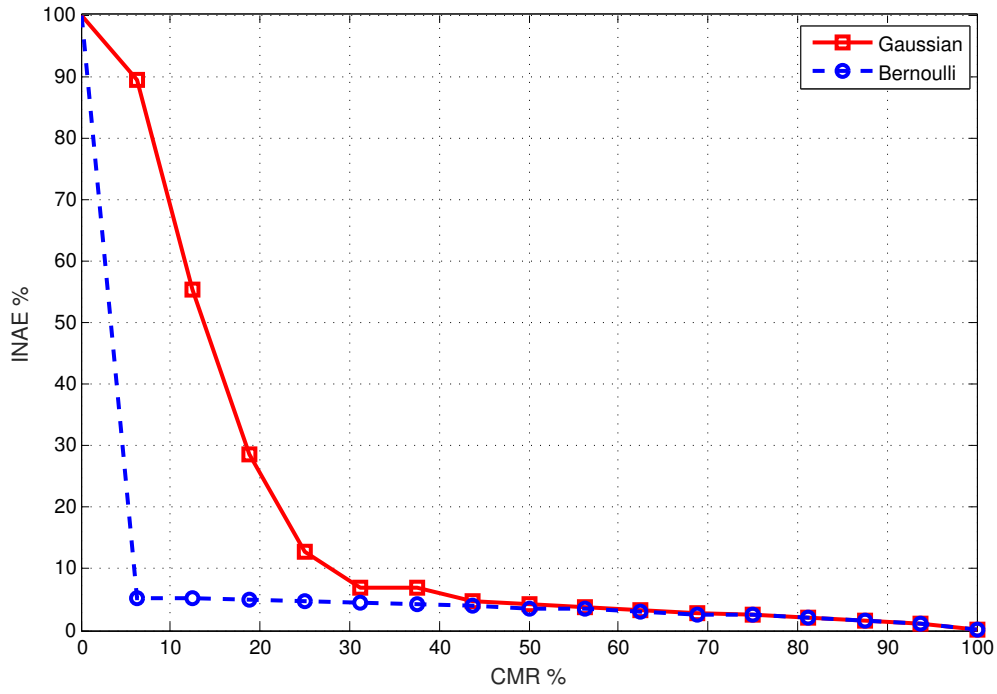

Figure 2.5: *Spatial compressive sensing for different projection matrix*

a function of CMR showing that a Bernoulli projection matrix is more effective than a Gaussian projection matrix, as even with very low CMR values, i.e., significantly fewer data values, the CS based reconstruction is very effective. Fig.2.6 shows the results based on temporal compressive sensing. In the temporal case, we have a better error performance than the spatial case as expected. This is because our model assumptions impose a higher degree of correlation over time than space. Once again, the Bernoulli projection matrix outperforms Gaussian projection matrix especially at low CMR levels. In Fig.2.7 performance of compressive sensing for different CMR has been depicted when data recovery is performed across spatial and temporal dimensions.

Table 2.2: *Performance of three different wavelet mother bases when 60% of data is available*

mother wavelet	**"Haar"**	**"Beylkin"**	**"Vaidyanathan"**
INAE %	**4.1**	**42.6**	**76.9**

Our goal is to estimate the voltage states from compressed power measurements. The idea is to measure power randomly based on a Bernoulli projection matrix, and then estimate the voltages after performing a (1D or 2D) compressive sensing based data recovery. For

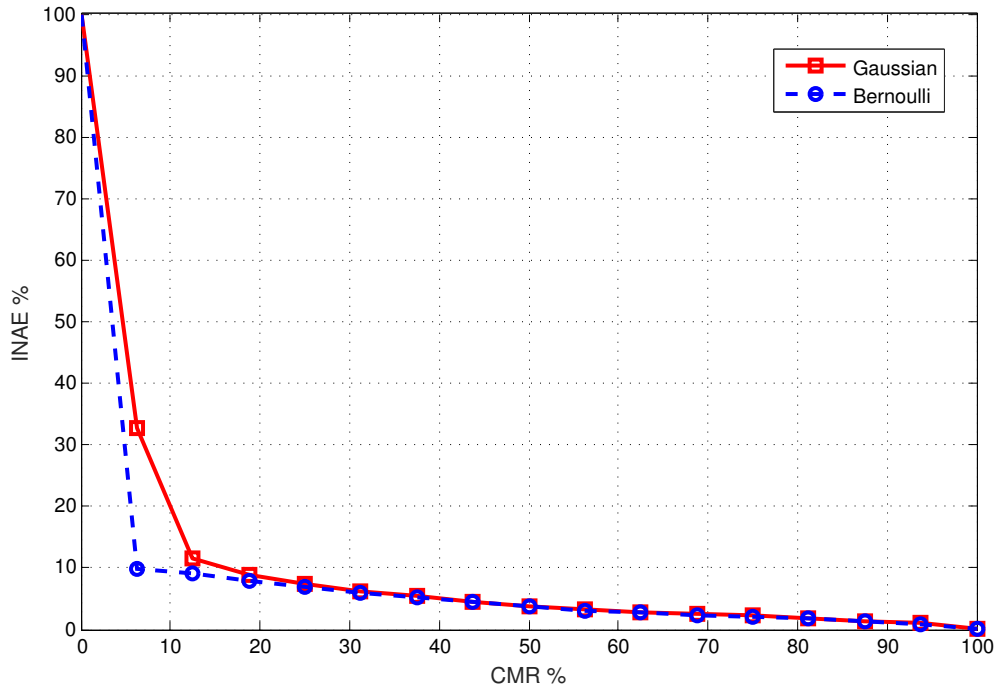

Figure 2.6: *Temporal compressive sensing for different projection matrix*

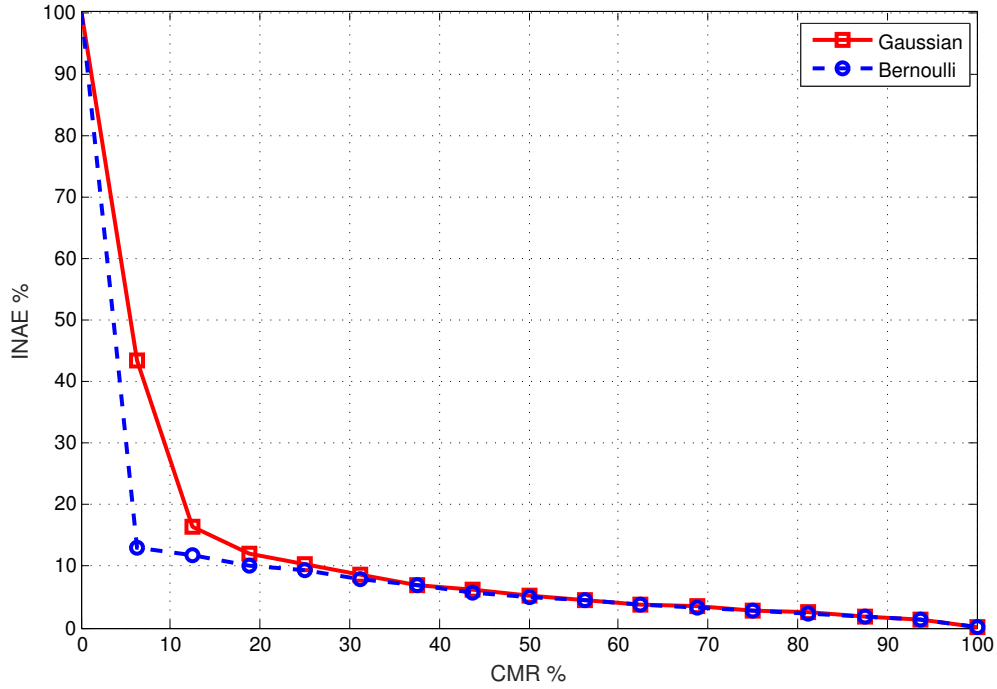

Figure 2.7: *Spatio-temporal compressive sensing for different projection matrix*

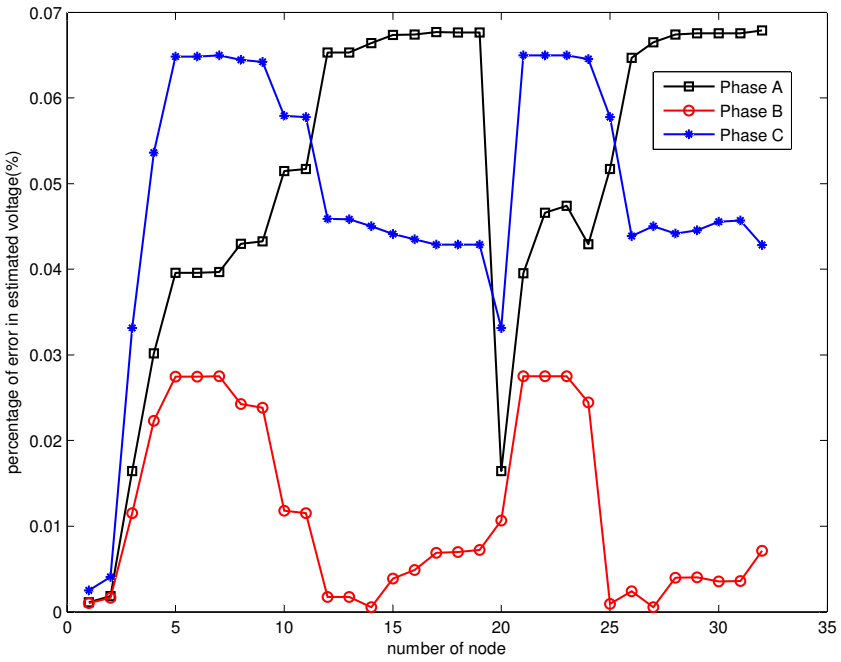

Figure 2.8: *Error in voltage estimation of three-phase unbalanced system recovered from 60% of information*

voltage state estimation, we use the forward backward algorithm described in section IV. Fig.2.8 depicts the estimation error of voltages for the IEEE 34-node three phase unbalanced system. The estimate is based on power measurement that were recovered from only sixty percent of original data. From Fig.2.8, it is evident the resulting estimates are of high fidelity with error varying only up to 0.07 percent. As mentioned earlier, the quality of reconstruction is dependent both on the choice of sparsifying basis and projection matrix. Table.2.2 compares the impact of two other wavelet bases on the error performance. It is clear that "Haar" as a mother wavelet is better choice for sparsifying basis Ψ in comparison with "Vaidyanathan" and "Beylkin".

2.5.2 Direct Method

We assume that the base kVA and base kV for IEEE 37-node are 100 and 48, respectively. Different case studies are presented for evaluating the performance of the compressive sens-

ing techniques. In the first case we consider both the power and voltage measurements while case II performs only based on power measurements.

Case I- Power and Voltage measurements

Fig. 2.9 and Fig. 2.10 show the comparative performance of spatial approaches with a subset of both power and voltage measurements in the IEEE 37 bus test system. Also, Fig. 2.11 and Fig. 2.12 demonstrate the 2-D CS in estimation magnitude and angle of voltages. Although, at higher CMR, the system becomes more observable and CS methods perform better but the techniques still shows a promising results even with low CMR.

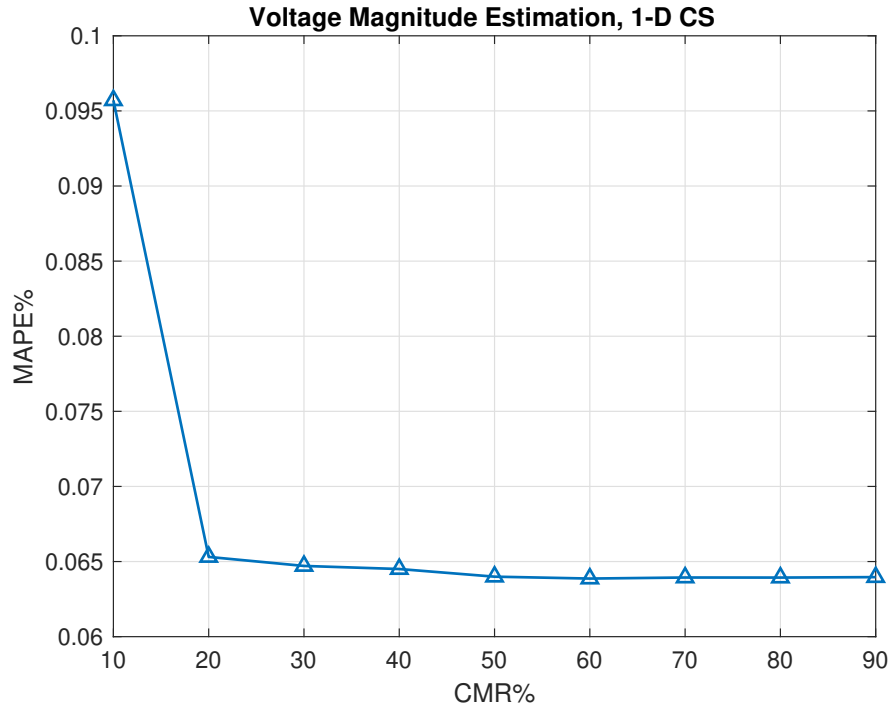

Figure 2.9: *Case I: Voltage magnitude performance of 1-D CS*

Case II- Only power measurements

In reality, voltage measurements at buses may be unavailable. To simulate this scenario, we demonstrate the performance of DSSE methods with only power measurements. Fig. 2.13 and Fig. 2.14 corresponds to 1-D CS, while Fig. 2.15 and Fig. 2.16 show performance of 2-D CS. It should be noted that both 1-D CS and 2-D CS based methods reconstruct the states

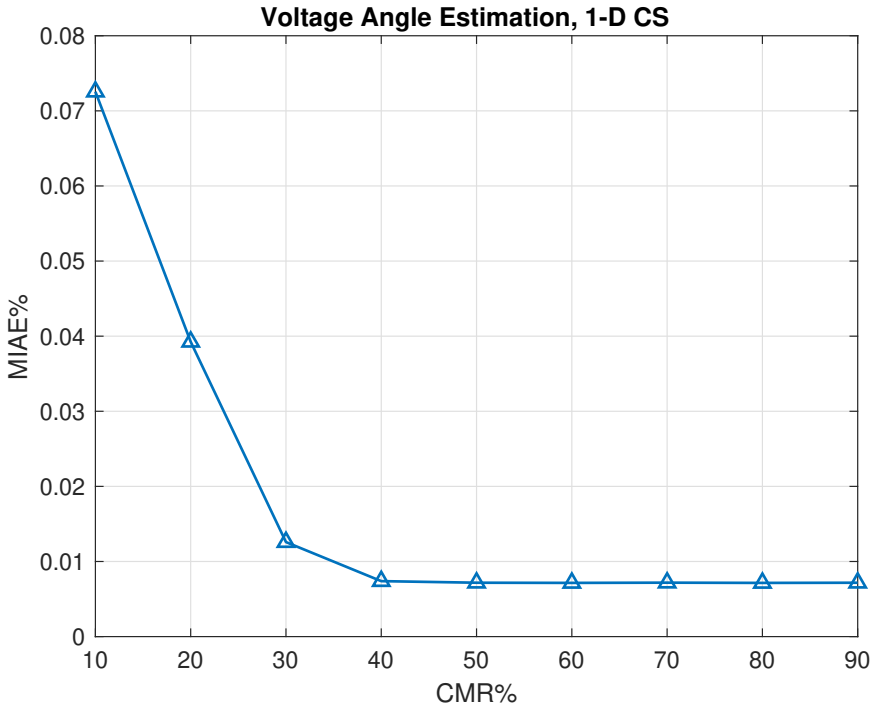

Figure 2.10: *Case I: Voltage angle performance of 1-D CS*

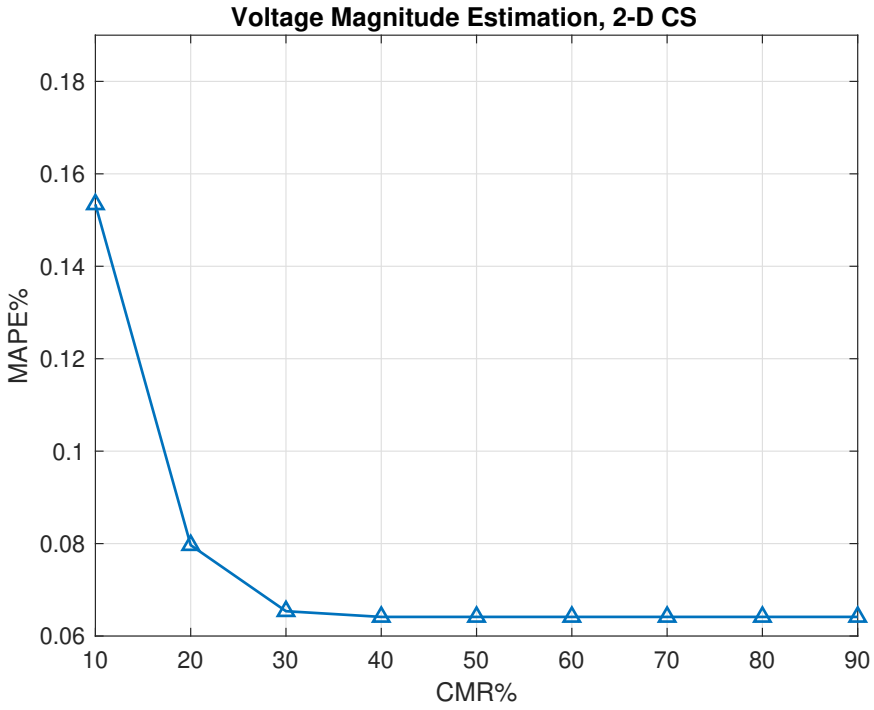

Figure 2.11: *Case I: Voltage magnitude performance of 2-D CS*

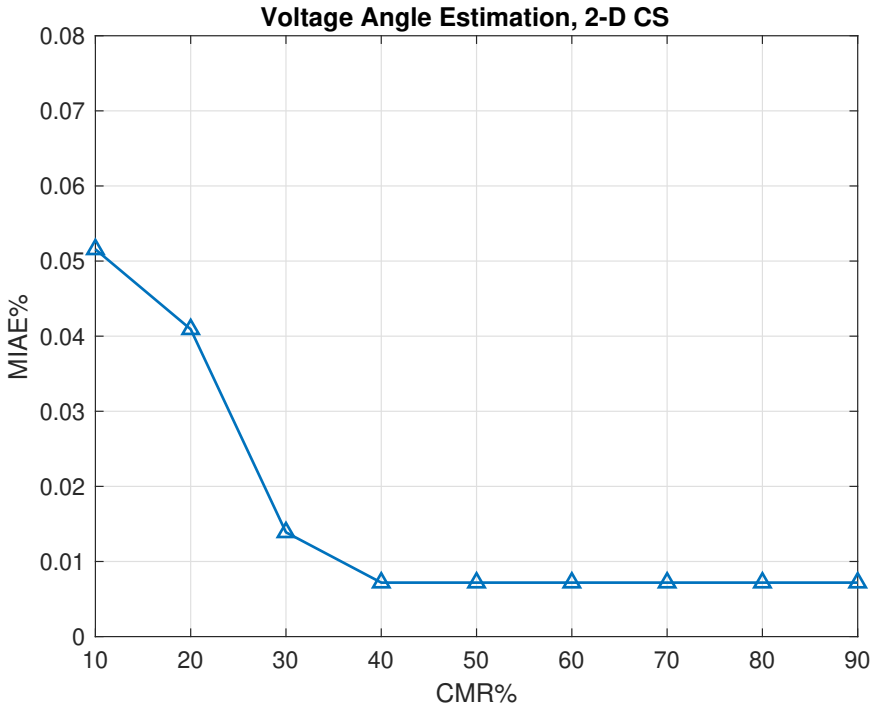

Figure 2.12: *Case I: Voltage angle performance of 2-D CS*

with high fidelity even at 40% CMRs.

2.6 Computational complexity

1-D CS based method involves an l_1 minimization which is solved by linear programming. Let m be the number of variables and n the number of constraints. The computational complexity of solving one newton step is $\mathcal{O}(mn.min(m,n))$ [51]. 2-D CS involves Kronecker product [52] of its sparsifying basis and measurement matrics for each of its d-sections. It solves a single higher dimensional optimization problem of complexity $\mathcal{O}(\prod_{d=1}^{D}(mn.min(m,n))^d)$.

The time required for each of these sparsity based methods for one Monte-Carlo simulation run and one CMR is tabulated in Table 2.3. These time calculations are performed on i5 core, 8 GB RAM processor with CVX Mosek package. In accordance with the complexity analysis , the time complexity of 1-D CS is lower than the time complexity of 2-D CS.

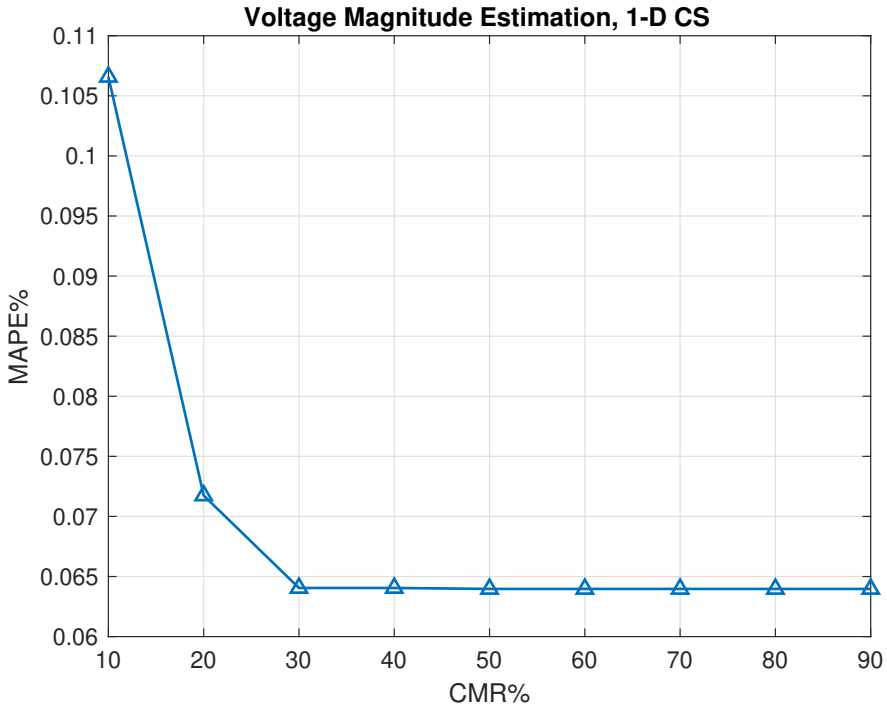

Figure 2.13: *Case II : Voltage magnitude performance of 1-D CS*

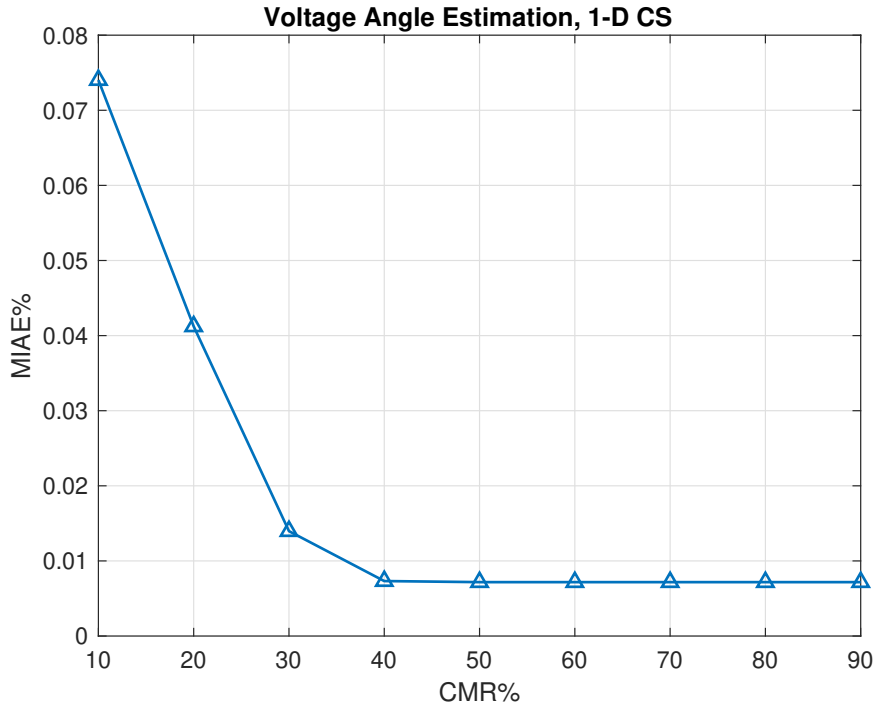

Figure 2.14: *Case II : Voltage angle performance of 1-D CS*

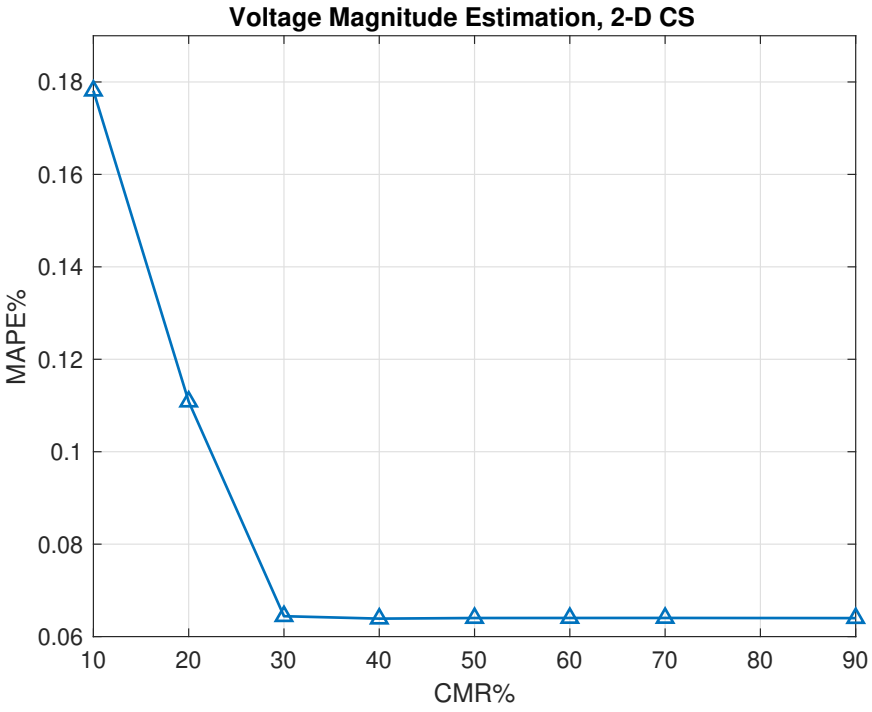

Figure 2.15: *Case II : Voltage magnitude performance of 2-D CS*

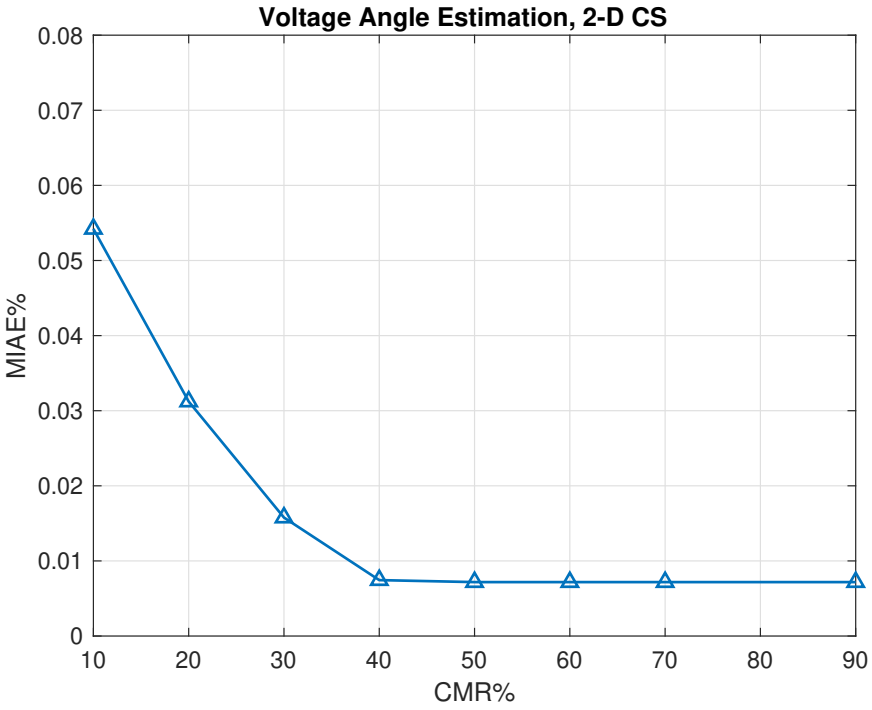

Figure 2.16: *Case II : Voltage angle performance of 2-D CS*

Table 2.3: *Time requirement for 1 run, 1 CMR*

Sparsity based methods	Time
1-D CS	3.6s
2-D CS	83.9s

2.7 Summary

In this chapter, we studied the issue of efficient data aggregation in a power distribution grid. With the increase in number of smart meters and sensors, the volume and velocity of the data impose a significant stress on the communication network. To alleviate this stress, we proposed compressive sensing approaches to recover the full dimensional power measurement from lower dimensional compressed data in three-phase unbalanced distribution networks. The proposed approaches leveraged the underlying spatio-temporal correlation in generation and load that results in sparsity in the wavelet domain. The recovered power measurements were then used to estimate the voltage states with high fidelity as demonstrated via simulation of the IEEE 34 node test system. With each projection matrix used in compressed sensing representing a specific modality of data aggregation, we demonstrated that a Bernoulli projection matrix outperforms Gaussian projection matrix, especially at low compression ratios. Furthermore, we proposed a direct methods that incorporated a linearized power flow constraint in to the ℓ_1 minimization which in turn enabled us to estimate both power and voltage, simultaneously. We validated the efficiency of this technique using IEEE 37-node test system. Consequently, impressive performance of compressive sensing has motivated us to address some critical questions in this field, particularly with their application in distribution systems. Next chapter is dedicated to recursive dynamic CS in distribution systems that improves power signal recovery in distribution systems.

Chapter 3

Dynamic CS in distribution systems

All the compressive sensing approaches discussed in chapter 2 employ static sparse recovery methods where only the current measurements are used for signal recovery purposes. Recently, it has been shown that if the sparse signal satisfies certain conditions, one can better reconstruct the current signal using the current and previous information. [22] presents a survey of recursive dynamic CS strategies which outperform batch algorithms and offer advantages such as on-line implementation, fast operation and a smaller memory requirement. In the current literature, recursive CS methods have not been employed in smart grids. Therefore, we aim to pursue this direction for the first time. To this end, we first exploit slow support set change and slow sparse signal change properties of the underlying data. To validate our assumptions, we incorporate practical home-level data from an actual distribution system. Then, we employ two recursive algorithms: 1- Streaming modified weighted-ℓ_1, 2- Kalman filtered CS. The simulation results (based on two different data framework) validate the efficiency of the recursive CS techniques where the performance of CS is significantly improved relative to the classic CS approach presented in [20].

3.1 Distribution System Signal Characteristic

In order to guarantee a certain signal recovery fidelity, it is obvious that the number of measurements can not be arbitrarily reduced (i.e., m must be large enough to achieve an acceptable level of reconstruction). However, the number of required measurements can be further reduced if the underlying sparse signal satisfies certain conditions. Let $\mathbf{x}_t$ be equivalent to the sparse signal $\mathbf{x}$ at discrete time instant t. Let N_t denote the support set of a sparse signal $\mathbf{x}_t$. That is, $N_t = \{i_1, i_2, ..., i_{S_t}\}$ where i_k are the non-zero coordinates of x_t and $S_t = |N_t|$, i.e.,

$$N_t := supp(\mathbf{x}_t) = \{i : (\mathbf{x}_t)_i \neq 0\}.$$

Definitely, signal sparsity implies that $S_t \ll n$. Let, $T_t = \hat{N}_t$ denotes an estimate of the actual support set.

Most of the dynamic sparse recovery techniques are based on two main assumptions.

Assumption 1: Sparsity patterns are assumed to be changing slowly over time. Slow support changes are practically observed in many applications such as [22]. Here, we assume that the maximum change in the support set size at time t is small relative to the size of the actual support set. In other words if S_t^m denotes the size of maximum changes in the support set, $S_t^m = max\big[|N_t \backslash N_{t-1}| + |N_{t-1} \backslash N_t|\big] < min\ S_t$ must hold for all times t. If this assumption is valid for a dynamical system, intuitively, one can employ the knowledge about the previous support set to estimate current support set and the sparse signal, simultaneously. For example a new version of basis pursuit denoising (BPDN) is introduced in [53], prior support set information is used to recover the current signal.

Assumption 2: Similar to slow support changes, sparse signals can be slowly changing over time. Here, we assume that the difference between current signal $\mathbf{x}_t$ and the previous signal $\mathbf{x}_{t-1}$ is significantly smaller than the signal $\mathbf{x}_t$, i.e., $\| \mathbf{x}_t - \mathbf{x}_{t-1} \|_2 \ll \| \mathbf{x}_t \|_2$. Using this information, it is possible to fit a dynamical model for the sparse signal $\mathbf{x}_t$. For example, one may consider a Gaussian random walk model with constant variance σ_{sys}^2 in all directions [54].

To evaluate the validity of the above assumptions, we use the load data of 272 homes

in an actual distribution system [55]. This data is publicly available which is collected and cleaned by Pecan Street research group for off-line analysis. Specifically, we consider 24 hours load data averaged in 15 minutes intervals on March 1^{st} 2018. Here, we cluster the loads of every 17 homes together, therefore, we have 16 lumped load data at each time $t \in \{1, 2, ..., 96\}$. We randomly select 75% of data without any measurement noises. Hence, the dimensions are $\mathbf{x}_t \in \mathbb{R}^{16}$, $\mathbf{h}_t \in \mathbb{R}^{12}$ and $\mathbf{\Phi} \in \mathbb{R}^{12\times16}$. Then, we reconstruct the spatial signal using the simple compressive sensing approach described by Theorem 1. At current time t_1, both the sparse signal and its support set are estimated without considering any temporal correlations with the previous times $t < t_1$. Fig 3.1 demonstrates the changes in the support set during the entire day. This figure shows both addition and deletion in the support set ($|N_t \backslash N_{t-1}| + |N_{t-1} \backslash N_t|$). It is evident that these support set changes are slow for this distribution system. Furthermore, Fig 3.2 shows the second norm of the sparse signal and its variations over a day. Fig 3.2 suggests that slow signal changes are typical in a distribution system. Therefore, Assumption 1 and 2 are reasonable and can help motivate the design of a new state estimator that exploit these properties.

3.2 Dynamic Recursive Compressive Sensing

Dynamic CS refers to the techniques that aim to recover a time sequence of sparse signals. Although the batch algorithms can deal with some dynamic CS problems, they suffer from serious drawbacks such as being offline, slow operation and a huge memory requirement. Dynamic recursive CS techniques overcome these drawbacks and provide superior performance relative to classic CS approaches with equal number of measurements. Recursive methods rely on one or both the Assumptions 1 and 2 stated in section 3.1. However, the problem of dynamic CS is significantly more difficult than the static set up because the support set N_t changes with time and is unknown. Furthermore, large errors in signal estimation may propagate and lead to poor performance of future signal recovery. Therefore, all recursive techniques typically assume that the initial estimates of support set and sparse signal are accurate.

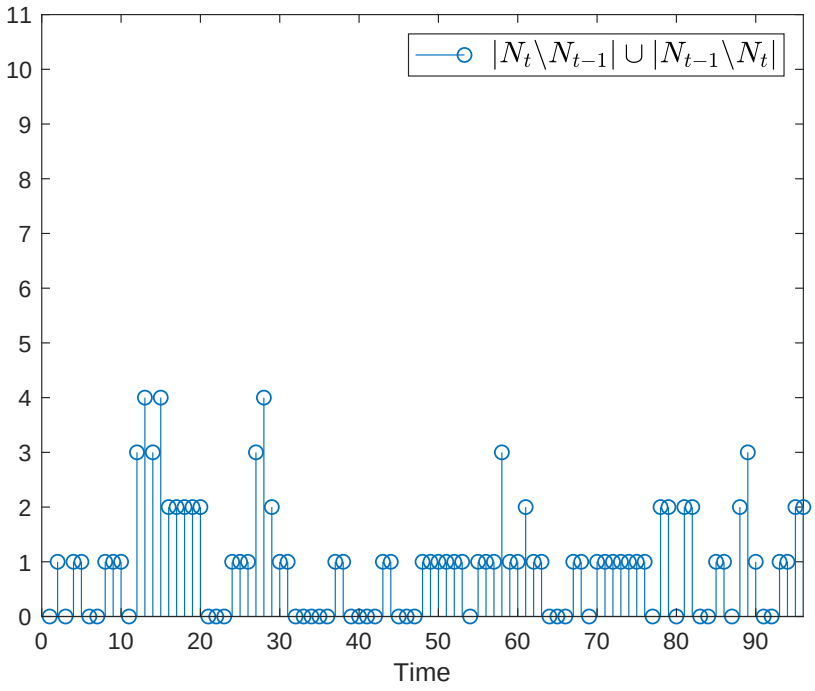

Figure 3.1: *Changes in a support set over time in a smart distribution system*

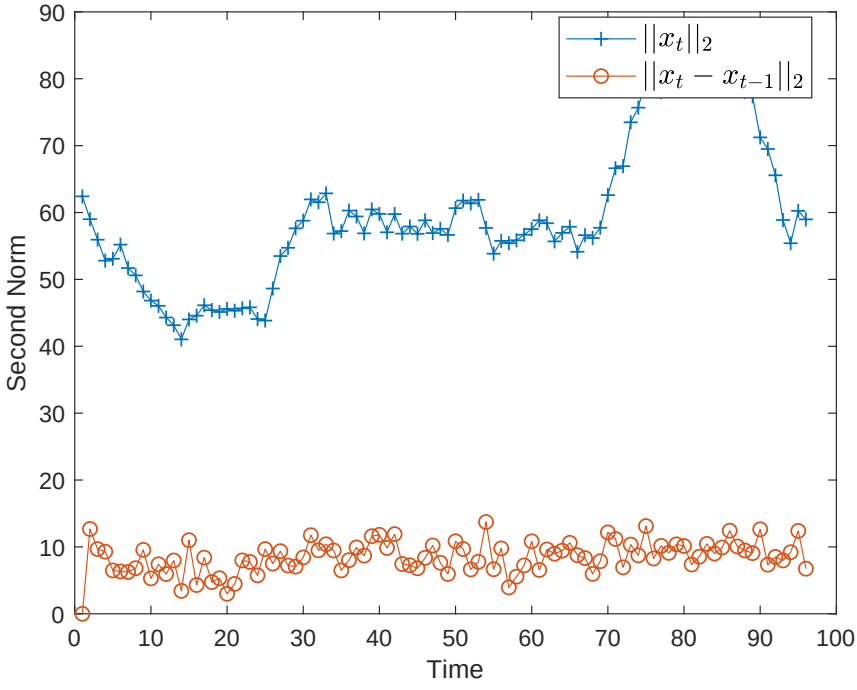

Figure 3.2: *Changes in a sparse signal over time in a smart distribution system*

In this chapter, we assume that the sparse signal changes slowly and its dynamic corresponds to:

$$(\mathbf{x}_t)_i = \begin{cases} (\mathbf{x}_{t-1})_i + (\boldsymbol{\omega}_t)_i, & \text{if } i \in N_t \\ (\mathbf{x}_{t-1})_i, & \text{otherwise} \end{cases} \tag{3.1}$$

where, $\boldsymbol{\omega}_t \in \mathbb{R}^n$ is a white Gaussian noise with covariance matrix $\mathbf{Q}_t$, i.e., $\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{Q}_t)$. Covariance matrix $\mathbf{Q}_t$ plays an important role in accuracy of reconstruction when we employ Kalman filter (KF) based methods. However, historical data can be a used source to estimate covariance matrix $\mathbf{Q}$.

A majority of existing CS approaches for smart grids assume that the measured signal is noiseless. However, this assumption is unrealistic and degrades the accuracy of signal recovery. Therefore, in this chapter, we assume that an additive noise vector is associated with measurements,

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \boldsymbol{\nu}_t \tag{3.2}$$

where, $\mathbf{y}_t \in \mathbb{R}^m$ $(m < n)$ is the available measurement at time t, $\mathbf{C} = \boldsymbol{\Phi}\boldsymbol{\Psi}$, and $\boldsymbol{\nu}_t \in \mathbb{R}^m$ is a white Gaussian noise with covariance matrix $\mathbf{R}_t$, i.e., $\boldsymbol{\nu}_t \sim \mathcal{N}(0, \mathbf{R}_t)$ and $\mathbb{E}\big[\boldsymbol{\omega}_t \boldsymbol{\nu}_j^{'}\big] = 0; \forall j, t$. The objective of dynamic recursive CS is to observe $\mathbf{y}_t,\ t \in \{1, 2, ..., t\}$ and estimate $\mathbf{x}_t$. In the following subsections, we describe dynamic recursive CS approaches to estimate $\mathbf{x}_t$.

3.2.1 Weighted BPDN

Streaming modified weighted-ℓ_1 method is summarized in Algorithm 2. As a generalization of the modified BPDN idea, a weighting scheme is introduced in [56]. Instead of 0 or 1 weighting policy, the weights can be selected to be any value between 0 and 1. If i does not belong to support set (i.e., $(\hat{\mathbf{x}}_{t-1})_i \approx 0$), the corresponding weight $(\mathbf{W}_t)_{ii}$ tends to be γ; otherwise, it takes smaller values. The weights are updated such that they become inversely proportional to the magnitude of the entries of $\hat{\mathbf{x}}_{t-1}$. Therefore, this approach incorporates

the slow support change assumption while using the signal values to update the weights. It should be noted that Step 1 in Algorithm 2 can use any CS approach as long as it returns an accurate estimate of the initial signal.

Algorithm 2 Streaming modified weighted-ℓ_1 (SMW-ℓ_1) [22]

Step 1: At $t = 1$: Solve BPDN with sufficient measurements, i.e., compute $\hat{\mathbf{x}}_0$ as the solution of $min_{\mathbf{b}}\ \gamma \parallel \mathbf{b} \parallel_1 + \parallel \mathbf{y}_1 - \mathbf{C}\mathbf{b} \parallel^2$

Step 2: For $t > 1$, set,

$$(\mathbf{W}_t)_{ii} = \frac{\gamma}{\beta|(\hat{\mathbf{x}}_{t-1})_i| + 1}$$

where $\beta = n\dfrac{\parallel \hat{\mathbf{x}}_{t-1} \parallel_2^2}{\parallel \hat{\mathbf{x}}_{t-1} \parallel_1^2}$,

Step 3: Compute $\hat{\mathbf{x}}_t$ as the solution of

$$\underset{\mathbf{b}}{\text{argmin}}\ \parallel \mathbf{W}_t\mathbf{b} \parallel_1 + \parallel \mathbf{y}_t - \mathbf{C}\mathbf{b} \parallel_2^2$$

Increment t and go to step 2

3.2.2 Kalman Filter CS

As a first solution to recursive dynamic CS problem, Kalman Filtered CS (KFCS) was introduced in [54]. Recently, [22] implements the Mod-BPDN-residual to capture possible changes in support set with the sparse signal being estimated by a regular KF. This method exploits both Assumption 1 and Assumption 2. Optimality and convergence analyses of the KFCS approaches remain as open questions. Recently, [39] provide statistical error analysis of KF-Mod-CS in presence of lossy measurements. When accurate prior knowledge of the signal values is available, KF-Mod-CS outperforms the modified-BPDN. Algorithm 3 summarizes Kf-ModCS. It should be noted that the parameters including $\mathbf{P}_0$, σ_{sys} and $\mathbf{R}$ matrices are extremely critical to find an accurate solution; otherwise, Algorithm 3 fails to converge.

Algorithm 3 Kalman Filtered modified Compressive Sensing [22]

Initialization: Set $\mathbf{x}_0 = 0$, $\mathbf{P}_0$, N_0 =empty (if unknown) or equal to the known/partially known support.
For $t > 0$, do,
Step 1: Set $T = \hat{N}_{t-1}$
Step 2: Mod-BPDN residual:
$\hat{\mathbf{x}}_{t,mod} = \hat{\mathbf{x}}_{t-1} + \left[\underset{\mathbf{b}}{\text{argmin}}\ \gamma \parallel \mathbf{b}_{T^c} \parallel_1 + \parallel \mathbf{y}_t - \mathbf{C}\hat{\mathbf{x}}_{t-1} - \mathbf{C}\mathbf{b} \parallel^2 \right]$
Step 3: Support Estimation - Simple thresholding:

$$\hat{N}_t = \{i : |(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a\}$$

Step 4: Modified Kalman Filter:
$\hat{\mathbf{Q}}_t = \sigma^2_{sys}\mathbf{I}_{\hat{N}_t}\mathbf{I}'_{\hat{N}_t}$
$\mathbf{K}_t = (\mathbf{P}_{t-1} + \hat{\mathbf{Q}}_t)\mathbf{C}'(\mathbf{C}(\mathbf{P}_{t-1} + \hat{\mathbf{Q}}_t)\mathbf{C}' + \mathbf{R})^{-1}$
$\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_t\mathbf{C})(\mathbf{P}_{t-1} + \hat{\mathbf{Q}}_t)$
$\hat{\mathbf{x}}_t = (\mathbf{I} - \mathbf{K}_t\mathbf{C})\hat{\mathbf{x}}_{t-1} + \mathbf{K}_t\mathbf{y}_t$
Increment t and go to step 1

3.3 Simulation Results

In this section, we evaluate the performance of the algorithms introduced in section 3.2. As described in section 3.1, the signal of interest $\mathbf{z}$ is not a sparse vector itself. However, we transform it to a sparse signal $\mathbf{x}$ by linear transformation $\mathbf{\Psi}$. Here, we apply the "Haar" mother wavelet as the sparsifying basis $\mathbf{\Psi}$ as it has been shown to outperform other basis choices [35]. A random measurement matrix $\mathbf{\Phi}$ consists of random Bernoulli entries which selects $\frac{m}{n} \times 100\%$ of the data is used. We establish our experiments based on two data frameworks: 1-IEEE 34-test feeder system; 2-Pecan Street data.

3.3.1 IEEE 34-bus system

[20] and [35] provides details of the IEEE 34 node distribution test feeder system considered. Here, we assume that every node has a stochastic trend $D_t = 0.05\sin(\frac{\pi}{12}t) + 0.3\mathbf{w}_t$ for load, where $\mathbf{w}_t \sim \mathcal{N}(0, I)$. The measurements are also associated with white noises where $\boldsymbol{\nu}_t \sim \mathcal{N}(0, 0.4I_{m\times m})$. In Algorithm 2, we consider $\gamma = 1.5$. In Algorithm 3, we set $\gamma = 0.075$ and $\alpha_a = 0.25$. Then, we recover the data using half ($m = 16$, $n = 32$) of the measurements.

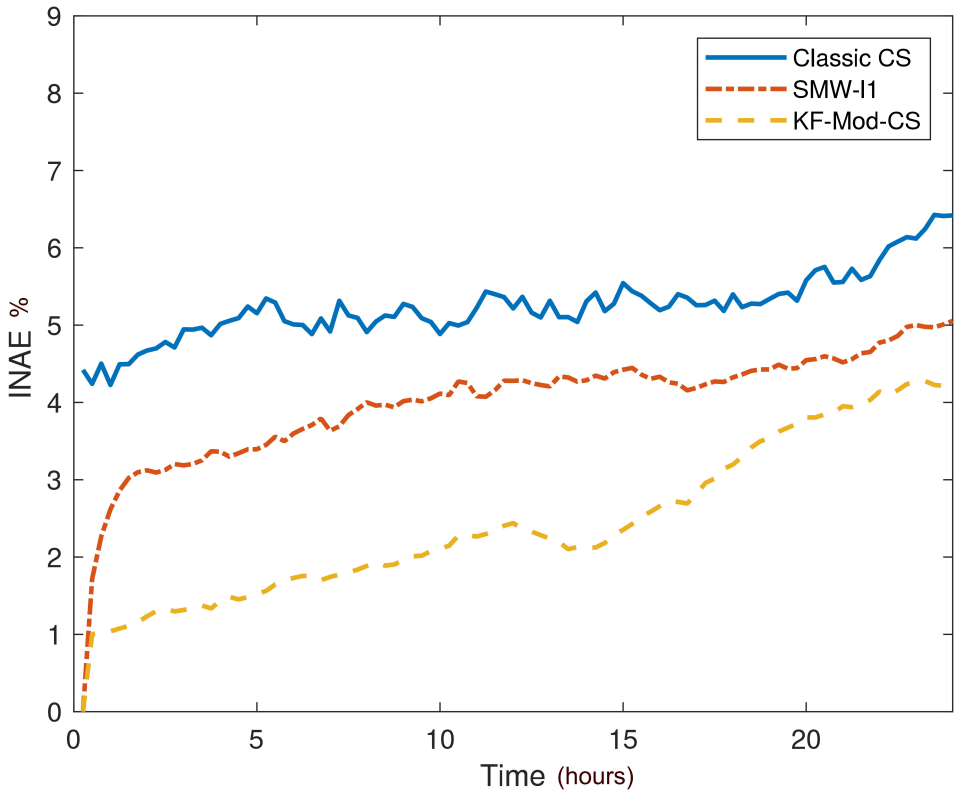

Figure 3.3: *INAE % of Algorithms 2 and 3 versus classic CS (Theorem 1)*

As it is shown by Fig 3.3, the Algorithms 2 and 3 (the recursive approaches) which rely on Assumption 1 and 2 outperform the conventional CS method suggested by Theorem 1. This improvement in performance can be justified by the facts that the underlying sparse power signals and their supports slowly vary over time.

3.3.2 Pecan Street Data

Pecan Street Inc provides electricity and water consumptions of more than 1000 volunteer customers. Furthermore, different types of the load data is available (e.g., HVAC use, refrigerator use, solar generation, etc.). The data is recorded at two time scales 1 minute and 1 hour. Here, we employ the one-minute load data relevant to 272 homes collected by Pecan Street Inc [55]. Specifically, we consider 24 hours load data averaged in 15 minutes intervals on March 2^{nd} 2018. As the loads associated 17 homes are lumped together, we aim to recover

Table 3.1: *Average INAE % of different CS approaches for Pecan Street Data*

Method	*INAE % (CR=75%)*	*INAE % (CR=50%)*
Classic CS	17.48	24.61
SMW-ℓ_1(Algorithm 2)	14.74	20.23
KF-Mod-CS(Algorithm 3)	12.38	17.54

16 data points at each time $t \in \{1, 2, ..., 96\}$. Hence, the dimensions are $\mathbf{x}_t \in \mathbb{R}^{16}$, $\mathbf{h}_t \in \mathbb{R}^8$ and $\mathbf{\Phi} \in \mathbb{R}^{8\times16}$. We set $\gamma = 2$ in Algorithm 2 and $\gamma = 0.3$, $\alpha_a = 0.2$ in Algorithm 3. Since, we don't have the exact information about the underlying sparse signal dynamics, we approximate the covariance matrix $\mathbf{Q}$ using prior data collected on March 1^{st} 2018. Table 3.1 summarizes the averages of *INAE* over the entire day resulting from the use of classic CS and the algorithms introduced in section 3.2. In this table, we consider different levels of compression ratio. Similar to the previous results for IEEE-34 test feeder system, the results are significantly improved by the recursive approaches which exploit slow signal and support changes.

3.4 Summary

In this chapter, our goal was to dynamically reconstruct a power signal from a fewer number of measurements. To do so, we employed two recursive dynamic CS techniques known as streaming modified weighted-ℓ_1 CS and Kalman filtered CS. Using a practical power data from an actual distribution system, we validated the underlying assumptions (slow signal and support change) required by dynamic recursive CS approaches. Since the recursive methods use the prior information about the sparse signal and its support set, they could improve the previous results obtained by classic CS. The superior performance of the two recursive CS methods were validated via two examples (the IEEE 34 node test feeder system and PecanStreet data). Although, real-time signal recovery in distribution systems was remarkably improved using the recursive dynamic techniques, the recovery process may be subject to real-time data aggregation issues such as measurement loss and delay. In the next chapter, we aim to study the KF-ModCS approach in the presence of lossy measurements.

Chapter 4

Updated KF-ModCS with lossy measurements

In existing studies related to dynamic compressive sensing, although number of measurements is much less than the dimension of the signal, it is assumed that measurements are always available. In this chapter, we are interested in understanding the impact of information loss on a specific type of dynamic CS based signal recovery. Specifically, many practical cyber-physical systems rely on wired or wireless communication infrastructure to aggregate measurements. Due to constraints such as limited bandwidth, multiple access interference and channel impairments, measurements could be lost. For example, let consider cognitive radio networks (CRN). Here, dynamic compressive sensing [57], [58] has been used ro exploit the two-fold sparsity in spectrum and location in order to address the spectrum sensing and localization problem in CRNs. Then, a KF-CS [59] is implemented to reconstruct the sparse signal of length N using M random measurements, where $M < N$. However, due to network constraints, CRNs are susceptible to packet losses [60], [61]. In other words, the availability of the required M measurements can not be guaranteed and varies randomly with network state. Therefore, the dynamic compressive sensing process is impaired due to intermittent measurements.

Similar to CRNs, smart grids may also suffer from lossy measurements due to unreliable

communication network between smart meters/SCADA sensors and the control center[62], [63]. While, all of the mentioned techniques assume that the required measurements arrive on-time, the control center may suffer from intermittent measurements due to impairments in the underlying communication link [23]. In a distribution system, [25] introduces a distributed dynamic state estimation strategy when the measurement packets are lost. In [26], a modified re-iterated Kalman filter is proposed that estimates the states of the smart grid in the presence of both lost and delayed measurements. [27] presents a flexible hybrid state estimation algorithm when the underlying communication network suffers from irregularities. As a result, even the compressed measurements for sparse recovery could be completely lost. These examples highlight the problem of lossy measurements in dynamic compressive sensing. While the impact of such network induced measurement loss on KF based state estimation has been extensively studied and well understood [28], [29], [30], the corresponding effect on dynamic CS is still unknown. This chapter aims to bridge this gap. To this end, we first establish our analysis for a general model that also fits for distribution systems. Then, as an example, we evaluate dynamic sparse recovery of distribution systems states impacted by packet losses. This chapter is organized as follows: Section 4.1 describes the system model of interest. Section 4.2 provides the updated modified kalman filter compressive sensing (Updated KF-Mod-CS) that can be used for state estimation of a sparse signal over time, when there are missing observations. The corresponding error analysis is presented in section 4.3. In section 4.4, convergence analysis of the algorithm is discussed. Additionally, the proposed upper bound is further improved for special cases. In section 4.5, two numerical simulations demonstrate the proposed method and validate the upper bound derived for the case of lossy measurements.

4.0.1 Contributions

We consider a KF-Mod-CS based algorithm (**Algorithm 6** in [38]) where the goal is to recursively recover a sparse signal and estimate its corresponding support set. Within this framework, our main contributions are summarized below:

- The KF-Mod-CS based algorithm consists of two main steps at each time: a compressive sensing algorithm to determine the support set using Kalman filter state estimates, and state estimation by Kalman filter using the estimated support set obtained from the modified CS. Therefore, the error in each process directly affects the other step. In this chapter, we first quantify the dynamics of error in each step while considering loss of measurements.

- Using input-to-state stability analysis for discrete nonlinear systems, we derive an upper bound for the expected variances of the estimated signal. This upper bound enables us to find a critical value for the rate of receiving measurements ($\bar{\lambda}^c$) that ensures the convergence of error in the KF-Mod-CS based algorithm. That is, if the probability of measurement loss is smaller than $1 - \bar{\lambda}^c$, we can ensure that the error in signal estimation is bounded. It is important to note that error and convergence analysis has not been investigated in the previous works, even for the case of Kalman filter CS without lossy measurements [38].

- We derive a more precise upper bound for the expected error variances when the underlying support set remains unaltered for a period of time. In this case, a reduced order Kalman filter is employed for estimation of the sparse signal.

- We uncover a steady state upper bound for the error covariance matrix if the underlying support set is constant and the reduced order model satisfies full observability condition. Simulation results provided in section 4.5 demonstrate the validity of the new fundamental theoretical results. We particularly evaluate our theoretical result using the IEEE 37-node test feeder with practical load/generation data.

4.0.2 Notation

Let t denote the discrete time index; for a set T, we use T^c to indicate the complement of T; the set operation $\setminus$ denotes set difference; $\mathbf{M}'$, $\mathbf{M}^*$, $\mathbf{M}^\dagger$ and $\mathbf{tr}(\mathbf{M})$ denote transpose of $\mathbf{M}$, adjoint of $\mathbf{M}$, Moore-Penrose pseudo-inverse of $\mathbf{M}$ and trace of $\mathbf{M}$, respectively. We

use $\emptyset$ to denote the empty set. Also, $||\boldsymbol{\nu}||_k$ denotes the ℓ_k norm of a vector $\boldsymbol{\nu}$. We use $supp(\boldsymbol{\nu}) = \{i \ : \ \boldsymbol{\nu}_i \neq 0\}$ to denote the support set of a sparse vector $\boldsymbol{\nu}$, i.e., the set of indices of its nonzero entries. For a matrix $\mathbf{A}$, $\mathbf{A}_T$ denotes the sub-matrix obtained by extracting the columns of $\mathbf{A}$ corresponding to the indices in T. We use $\mathbf{I}$ to denote the identity matrix. $diag(\mathbf{A})$ denotes a vector consisting of the diagonal elements of $\mathbf{A}$. $(\mathbf{x}_t)_i$ denote i^{th} element of vector $\mathbf{x}_t$, $N_{i,t}$ is the i^{th} member of set N_t, and $\mathbf{e}^{\mathbf{Q}}_{i,t}$ is the i^{th} elements on the diagonal of matrix $\mathbf{e}^{\mathbf{Q}}_t$. Also the signum function is defined as $sgn(x) = -1$ if $x < 0$, $sgn(x) = 0$ if $x = 0$, $sgn(x) = 1$ if $x > 0$. A function $\xi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is called $\mathcal{K}$-function if it is continuous, strictly increasing and $\xi(0) = 0$. The composition of two functions ξ_1 and ξ_2 is denoted by $\xi_1 \circ \xi_2$. $\mathcal{Q}$-function for a random variable X with standard normal distribution is defined as $\mathcal{Q}(x) = \mathbf{Pr}(X > x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{s^2}{2}} ds$.

4.1 System Model

In this chapter, we consider a general system model for recovery of the signal of interest. Similar to the notation of previous chapter, let $(\mathbf{z}_t)_{n\times 1}$ denote an n-dimensional spatial signal at time t. Our goal is to recover $(\mathbf{z}_t)_{n\times 1}$ from an m-dimensional observation vector $\mathbf{y}_t = \boldsymbol{\Phi}\mathbf{z}_t + \boldsymbol{\omega}_t$ at time t, where $\boldsymbol{\Phi}$ is an $m \times n$ measurement matrix with $m < n$; $\boldsymbol{\omega}_t$ is i.i.d. Gaussian measurement noise with mean 0 and covariance $\sigma^2_{obs}\mathbf{I}$. We assume that $\mathbf{z}_t$ is sparse in a known basis $\boldsymbol{\Psi}_{n\times n}$ (e.g, wavelet or Fourier). That is $\mathbf{x}_t = \boldsymbol{\Psi}'\mathbf{z}_t$, is a sparse vector, (i.e, just S_t elements of $\mathbf{x}_t$ are non-zero with $S_t << n$). Furthermore, for a given sparsity basis, the measurements must be incoherent. Incoherency implies that the correlation between the columns of $\mathbf{C} = \boldsymbol{\Phi}\boldsymbol{\Psi}$ has to be small and we can assert that, any S-column sub-matrix of $\mathbf{C}$ is approximately orthonormal for all $S \geq S_t$ (with its nonzero singular values between $(1-\delta)^{1/2}$ to $(1+\delta)^{1/2}$ for $\delta < 1$) [59], [64]. Figure 4.1 shows the relationship between signal of interest, the underlying sparse signal as well as the KF-CS estimation process. The updated measurement model is:

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \boldsymbol{\omega}_t, \tag{4.1}$$

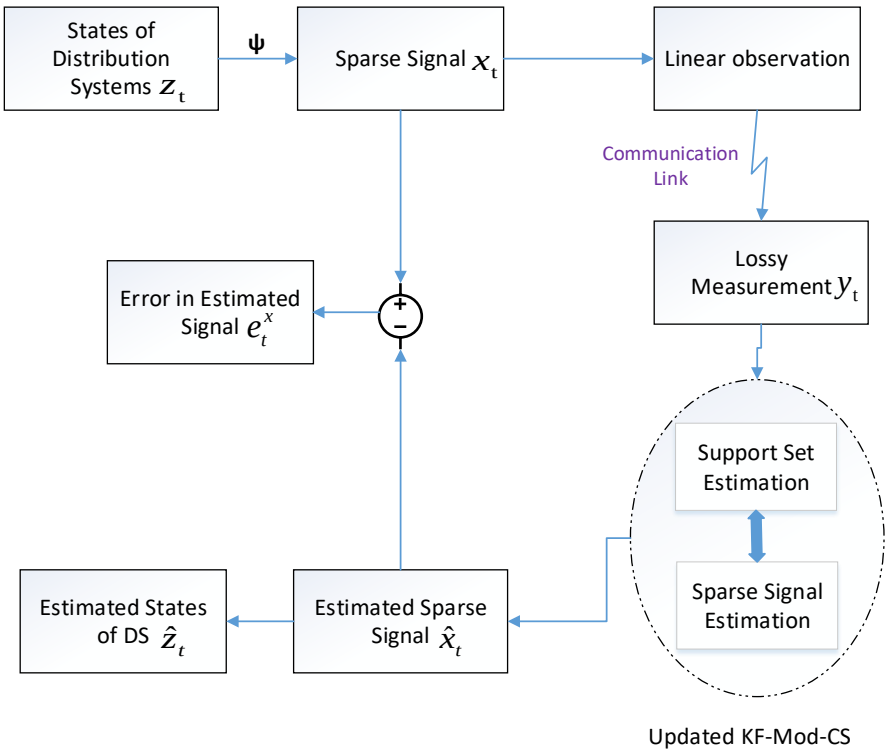

Figure 4.1: *Estimation Process.*

From a state space point of view, $\mathbf{x}_t$ is the state of the system at time t. Our goal is to reconstruct the signal of interest $\mathbf{z}_t$ from a smaller number of available measurements. Equivalently, the goal can also be interpreted as a causal state estimation of $\mathbf{x}_t$. As stated in the previous chapter, we assume that two assumptions hold for the system of interest.

For the dynamics of the sparse signal $\mathbf{x}_t$, we consider a Gaussian random walk model with constant variance σ^2_{sys} in all directions. The state equations for our system therefore corresponds to:

$$(\mathbf{x}_t)_i = \begin{cases} (\mathbf{x}_{t-1})_i + (\boldsymbol{\nu}_t)_i, & \text{if } i \in N_t \\ (\mathbf{x}_{t-1})_i, & \text{otherwise} \end{cases} \quad \forall t = 1, 2, ... \tag{4.2}$$

where, $(\boldsymbol{\nu}_t)_i \sim \mathcal{N}(0, \sigma^2_{sys})$. The updated state and measurement model for $t = 1, 2, ...$

with $x_0 \sim \mathcal{N}(0, P_0)$ can be written as:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{x}_{t-1} + \boldsymbol{\nu}_t, \\ \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t + \boldsymbol{\omega}_t, \end{aligned} \tag{4.3}$$

In the above equations, t is the discrete-time index which is an integer, $\mathbf{x}_t \in \mathbb{R}^n$ is the state vector, $\mathbf{y}_t \in \mathbb{R}^n$ is the observation vector with a Gaussian noise $(\boldsymbol{\omega}_t)_i \sim \mathcal{N}(0, \sigma^2_{obs})$. $\boldsymbol{\omega}_t \in \mathbb{R}^m$, $\boldsymbol{\omega}_t \in \mathbb{R}^m$ and $\boldsymbol{\nu}_t \in \mathbb{R}^n$ are zero-mean Gaussian random vectors with $\mathbb{E}\big[\boldsymbol{\nu}_t \boldsymbol{\nu}_j^{'}\big] = \delta_{tj}\mathbf{Q}_t,\ \mathbf{Q}_t > 0$, $\mathbb{E}\big[\boldsymbol{\omega}_t \boldsymbol{\omega}_j^{'}\big] = \delta_{tj}\mathbf{R}_t,\ \mathbf{R}_t \geq 0$, $\mathbb{E}\big[\boldsymbol{\omega}_t \boldsymbol{\nu}_j^{'}\big] = 0$; $\forall j, t$, where $\delta_{tj} = 0$ if $t \neq j$ and $\delta_{tj} = 1$ otherwise. Without loss of generality, we assume $\mathbf{R}_t$ is an constant matrix $\mathbf{R}_t = \mathbf{R}$. The initial state x_0 is assumed to be zero-mean Gaussian with covariance $\mathbf{P}_0 \geq 0$ and is uncorrelated with $\boldsymbol{\omega}_t$ and $\boldsymbol{\nu}_t$ for all $t > 0$. Additionally, $\mathbf{Q}_t$ is defined as follow:

$(\mathbf{Q}_t)_{N_t, N_t} = \sigma^2_{sys}\mathbf{I}$

$(\mathbf{Q}_t)_{N_t^c, N_t^c} = 0$

If N_t is known at each t, the system model is completely defined, and therefore the MMSE estimate of $\mathbf{x}_t$ from $\{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t\}$ can be obtained by a reduced order Kalman filter. It is important to remember that the set N_t is unknown and time-varying making the estimation problem challenging. If σ^2_{sys}, σ^2_{obs} are known, we can obtain the best estimates of N_t and $\mathbf{x}_t$ at each t using all available observations $\mathbf{Y}_t \triangleq \{\mathbf{y}_0, ..., \mathbf{y}_t\}^{'}$. Kalman Filtered CS-residual (KF-CS) was the first solution introduced for solving this recursive dynamic CS problem in [59]. Recently, in [38], authors use a modified-CS approach known as Kalman Filtered Modified-CS-residual (KF-ModCS) to address this problem. Analysis and simulations demonstrated in [38], show that this new technique outperforms conventional KF-CS. Both the KF-CS and KF-ModCS approaches first estimate the support set N_t and then share the traditional Kalman filter process as described bellow:

$$\hat{\mathbf{x}}_{t+1|t+1} = \hat{\mathbf{x}}_{t+1|t} + \mathbf{K}_{t+1}(\mathbf{y}_{t+1} - \mathbf{C}\hat{\mathbf{x}}_{t+1|t}) \tag{4.4}$$

$$\mathbf{P}_{t+1|t+1} = \mathbf{P}_{t+1|t} - \mathbf{K}_{t+1}\mathbf{C}\mathbf{P}_{t+1|t}. \tag{4.5}$$

where,

$$\hat{\mathbf{x}}_{t|t} \triangleq \mathbb{E}\{\mathbf{x}_t|\mathbf{y_t}\} \tag{4.6}$$

$$\mathbf{P}_{t|t} \triangleq \mathbb{E}\{(\mathbf{x}_t - \hat{\mathbf{x}}_t)(\mathbf{x}_t - \hat{\mathbf{x}}_t)^{'}|\mathbf{Y}_t\} \tag{4.7}$$

$$\hat{\mathbf{x}}_{t+1|t} \triangleq \mathbb{E}\{\mathbf{x}_{t+1}|\mathbf{Y}_t\} = \hat{\mathbf{x}}_{t|t} \tag{4.8}$$

$$\mathbf{P}_{t+1|t} \triangleq \mathbb{E}\{(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1})^{'}|\mathbf{Y_t}\} = \mathbf{P}_{t|t} + \mathbf{Q}_t \tag{4.9}$$

$$\hat{\mathbf{y}}_{t+1|t} \triangleq \mathbb{E}\{\mathbf{y}_{t+1}|\mathbf{Y}_t\}. \tag{4.10}$$

Here, $\mathbf{K}_{t+1} = \mathbf{P}_{t+1|t}\mathbf{C}^{'}(\mathbf{C}\mathbf{P}_{t=1|t}\mathbf{C}^{'} + \mathbf{R})^{-1}$ is the Kalman gain matrix.

As discussed in the introduction, dropped packets in the underlying communication network may result in lost measurements at some time instances [25], [24]. At time instance t, it is assumed that either all the elements in a measurement vector are received or lost. Therefore, we can model the availability of an observation as an i.i.d Bernoulli random process with $\mathbf{Pr}(\lambda_t = 1) = \bar{\lambda}$. More precisely, if the measurement vector $\mathbf{y}_t$ is fully received at time t, we have $\lambda_t = 1$; otherwise $\lambda_t = 0$. In this chapter, to recover the sparse states we propose the use of the updated KF-Mod-CS presented in [39]. Figure 4.1 shows the relationship between the signal of interest, the underlying sparse signal as well as the KF-CS estimation process.

As discussed in the introduction, it is possible that measurements may be lost due to dropped packets in a networked system. We can model the availability of a measurement as an i.i.d Bernoulli random process with $\mathbf{Pr}(\lambda_t = 1) = \bar{\lambda}$. In the presence of lossy measurements, a modified Kalman filter can be used to estimate the states as follows: Prediction is performed through (4.8) and (4.9). For the update steps, the standard equations are rewritten as follows:

$$\hat{\mathbf{x}}_{t+1|t+1} = \hat{\mathbf{x}}_{t+1|t} + \lambda_{t+1}\mathbf{K}_{t+1}(\mathbf{y}_{t+1} - \mathbf{C}\hat{\mathbf{x}}_{t+1|t}) \tag{4.11}$$

$$\mathbf{P}_{t+1|t+1} = \mathbf{P}_{t+1|t} - \lambda_{t+1}\mathbf{K}_{t+1}\mathbf{C}\mathbf{P}_{t+1|t}. \tag{4.12}$$

where, $\hat{\mathbf{x}}_{t|t} \triangleq \mathbb{E}\{\mathbf{x}_t|\mathbf{Y}_t,\ \lambda_\mathbf{t}\}$, $\mathbf{P}_{t|t} \triangleq \mathbb{E}\{(\mathbf{x}_t - \hat{\mathbf{x}_t})(\mathbf{x}_t - \hat{\mathbf{x}_t})'|\mathbf{Y}_t,\ \lambda_\mathbf{t}\}$, $\hat{\mathbf{x}}_{t+1|t} \triangleq \mathbb{E}\{\mathbf{x}_{t+1}|\mathbf{Y}_t,\ \lambda_\mathbf{t}\}$, $\mathbf{P}_{t+1|t} \triangleq \mathbb{E}\{(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1})(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1})'|\mathbf{Y}_t,\ \lambda_\mathbf{t}\}$, $\hat{\mathbf{y}}_{t+1|t} \triangleq \mathbb{E}\{\mathbf{y}_{t+1}|\mathbf{Y}_t,\ \lambda_\mathbf{t}\}$, here we define a Bernoulli random vector $\lambda_\mathbf{t} \triangleq \{\lambda_0, ..., \lambda_t\}'$. It should be noted that the main difference between the modified Kalman filter and the standard Kalman filter formulation is that both $\mathbf{P}_{t|t}$ and $\mathbf{P}_{t+1|t}$ are now random matrices due to λ_t.

Remark. *It should be clarified that the measurements may be reduced in two ways. First, we linearly compress the observations, i.e., we only transmit a certain percentage of the measurements. This process is depicted by "linear observation" block in Figure 4.1. Second, even that compressed measurement vector may be totally lost at some time instances due to communication link impairments ("lossy measurement" block in Figure 4.1).*

The next section provides a brief description of the modified Kalman filter based algorithm that estimates the sparse states in the presence of lossy measurements.

4.2 Updated KF-Mod-CS

The problem of dynamic CS is significantly more difficult than the static set up because the support set N_t changes with time and is unknown. [65] proposes a recursive algorithm known as modified Basis Pursuit Denoising (Mod-BPDN)-residual for estimating the sparse signal. In order to improve signal estimation and tracking changes in support set, [38] presents a Kalman filter based algorithm (KF-ModCS) similar to the approach provided in [59]. [38] implements the Mod-BPDN-residual to capture possible changes in support set with the sparse signal being estimated by a regular KF. Unlike the regular KF for a fixed dimensional linear Gaussian state space model, KF-CS or KF-ModCS do not enjoy any optimality properties. When accurate prior knowledge of the signal values is available, KF-ModCS outperforms the modified-BPDN.

We consider a KF-Mod-CS based algorithm (**Algorithm 6** in [38]) which is compatible with lossy measurements. In order to track the possible changes of support set estimation and also to improve the previous support set estimations, we need to implement compressive

sensing on the residual. If measurements are available, we first find a sparse vector $\hat{\mathbf{x}}_{t,mod}$ by solving an ℓ_1 optimization problem. Then, we can update the support set estimation when measurements are available. However, if the measurements are lost ($\lambda_t = 0$), the previous estimated support set is chosen as the current estimated support set. Equation (4.13) shows the revised support set estimation:

$$\hat{N}_t = \lambda_t\{i : |(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a\} + (1 - \lambda_t)\hat{N}_{t-1} \tag{4.13}$$

where, $\hat{N}_t$ is an estimation of support set N_t, and α_a is a threshold. Finally, instead of a standard Kalman filter, sparse signals $\mathbf{x}_t$ are estimated by a modified kalman filter [28] using the estimated support set, available measurements and previous estimated signals. We refer to this revision of KF-Mod-CS as updated Kalman filtered modified compressive sensing (Updated KF-Mod-CS) and we summarize it in Algorithm 4. As a recursive method, computational complexity of Algorithm 4 is much lower than the batch algorithms [38]. Specifically, the computational complexity of Step 2 in Algorithm 4 is $\mathcal{O}(m^3 + m^2 n)$ [66]. Since $m > n$ holds, the computational complexity becomes $\mathcal{O}(m^3)$. Basic implementation of Step 4 results in complexity of $\mathcal{O}(m^3)$. However, the computational complexity of the Kalman filter step can be reduced using various optimized practical algorithms [67]. For example, [68] recently proposed an approach to implement Kalman filter based on rank-one update.

In order to recover the sparse signal in the presence of lossy measurements, we consider the updated KF-Mod-CS based approach (**Algorithm** 4). In many of the recursive algorithms, we need to precisely know the support set (at least an accurate approximation of the support set). Since the support set does not change quickly, we use the previous support set in step 1 of Algorithm 4. If measurements are available and we estimate the possible changes in support set, we can update the previous support set. To this end, we first need to implement compressive sensing on the residual. More precisely, we obtain a sparse vector $\hat{\mathbf{x}}_{t,mod}$ by solving the ℓ_1 optimization problem in Step 2. This step is based on the algorithm presented in [32] known as modified Basis Pursuit Denoising (Mod-BPDN)-residual for estimating the

Algorithm 4 Updated Kalman Filtered Modified Compressive Sensing

Initialization: Set $\hat{\mathbf{x}}_0 = 0$, $\hat{\mathbf{P}}_0$, $\hat{N}_0$ =empty (if unknown) or equal to the known/partially known support.
For $t > 0$, do,
Step 1: Set $T = \hat{N}_{t-1}$
Step 2: If measurement is available ($\lambda_t = 1$), run Mod-BPDN residual:
$\hat{\mathbf{x}}_{t,mod} = \hat{\mathbf{x}}_{t-1} + \left[\underset{b}{\text{argmin}}\ \gamma||\mathbf{b}_{T^c}||_1 + ||\mathbf{y}_t - \mathbf{C}\hat{\mathbf{x}}_{t-1} - \mathbf{C}\mathbf{b}||^2\right]$
Step 3: Support Estimation - Simple thresholding:

$$\hat{N}_t = \lambda_t\{i : |(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a\} + (1 - \lambda_t)\hat{N}_{t-1}$$

Step 4: Modified Kalman Filter:
$\hat{\mathbf{Q}}_t = \sigma^2_{sys}\mathbf{I}_{\hat{N}_t}\mathbf{I}'_{\hat{N}_t}$
$\mathbf{K}_{t+1} = (\mathbf{P}_{t-1} + \hat{\mathbf{Q}}_t)\mathbf{C}'(\mathbf{C}(\mathbf{P}_{t-1} + \hat{\mathbf{Q}}_t)\mathbf{C}' + \mathbf{R})^{-1}$
$\mathbf{P}_t = (\mathbf{I} - \lambda_t\mathbf{K}_t\mathbf{C})(\mathbf{P}_{t-1} + \hat{\mathbf{Q}}_t)$
$\hat{\mathbf{x}}_t = (\mathbf{I} - \lambda_t\mathbf{K}_t\mathbf{C})\hat{\mathbf{x}}_{t-1} + \lambda_t\mathbf{K}_t\mathbf{y}_t$
Increment t and go to step 1.

sparse signal. Then, the support set is estimated by a simple thresholding in Step 3. However, if the measurements are lost ($\lambda_t = 0$), we skip Step 2 and the previous estimated support set is selected as the current estimated support set in Step 3. Finally, we estimate the states by a modified Kalman filter in Step 4. If measurements are available, the previous states are updated within the Kalman filter; otherwise, we use the previous estimated states for the current time. It should be noted that the main difference between the modified Kalman filter and the standard Kalman filter formulation is that both error covariance matrices $\mathbf{P}_{t|t}$ and $\mathbf{P}_{t+1|t}$ are now random due to λ_t.

As stated earlier and based on practical distribution system data, the voltage and power signals seldom experience significant changes in support set. In other words, the support set N_t remains constant over a time interval. Therefore, if the initial support set N_0 is accurately estimated, the support set N_t is reasonably known at each t. This implies that the system model is completely defined, and therefore the MMSE estimate of $\mathbf{x}_t$ from $\{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t\}$ can be obtained by a reduced order Kalman filter. In this case, The Step 1 to 3 of Algorithm 4 are skipped and we only run Step 4 of Algorithm 4.

4.3 Error analysis

Missing observations at some time instances leads to increase in error associated with computations where measurements are used. In Algorithm 4, both Mod-BPDN compressive sensing and Kalman filter updates rely on current measurements. Therefore, we have two dynamical interactive error processes, where errors in support set estimation and signal estimation impact each other at both current and future time instances. For example, let's look at equation (4.13) which corresponds to support set estimation. Error in the first term of (4.13) originates from the previous error in the Kalman filter step because $\hat{\mathbf{x}}_{t_s,mod}$ directly depends on $\hat{\mathbf{x}}_{t_s-1}$. Furthermore consider that measurements are lost at time t_s. Thus, $\hat{N}_{t_s}$ is equal to $\hat{N}_{t_s-1}$. This implies that $\hat{N}_{t_s}$ is impacted by the error associated with $\hat{N}_{t_s-1}$ and changes in support set can not be tracked in this time instant. This error in support estimation leads to errors in $\hat{\mathbf{Q}}_{t_s}$, which is a critical term in the Kalman filter equations. Consequently, $\hat{\mathbf{x}}_{t_s}$ has errors due to both missing measurements and errors in support estimation. In other words, we have a dynamic process where both signal and support estimation errors evolve with time and interact with each other. We attempt to quantify these error dynamics via Theorem 3 and Theorem 4 along with Lemma 1 and Lemma 2.

Let $N_t = \hat{N}_t + e_t^N$, where N_t is the actual support set, $\hat{N}_t$ is the estimated support set, and e_t^N is the error in support set estimation at time t. Similarly, we define $\mathbf{x}_t = \hat{\mathbf{x}}_t + \mathbf{e^x}_t$, where $\mathbf{x}_t$, $\hat{\mathbf{x}}_t$ and $\mathbf{e^x}_t$ are sparse signal of the system, estimated sparse signal and error in sparse signal estimation, respectively. We also denote $\mathbf{x}_{t,mod} = \hat{\mathbf{x}}_{t,mod} + \mathbf{e}_t^{mod}$, where $\mathbf{x}_{t,mod} = \mathbf{x}_{t-1} + \left[\underset{\mathbf{b}}{\text{argmin}}\ \gamma||\mathbf{b}_{N^c}||_1 + ||\mathbf{y}_t - \mathbf{C}\mathbf{x}_{t-1} - \mathbf{C}\mathbf{b}||^2\right]$, $\hat{\mathbf{x}}_{t,mod}$ is the estimate of $\mathbf{x}_{t,mod}$, and $\mathbf{e}_t^{mod}$ is the error in this estimation. Furthermore, we define $\hat{\mathbf{b}}$ as the solution of ℓ_1 optimization, i.e., $\hat{\mathbf{b}}_t = \underset{\mathbf{b}}{\text{argmin}}\ \left[\gamma||\mathbf{b}_{T^c}||_1 + ||\mathbf{y}_t - \mathbf{C}\mathbf{x}_{t-1} - \mathbf{C}\mathbf{b}||_2^2\right]$. Now, consider an estimation of $\hat{\mathbf{b}}_t$ such that $\tilde{\mathbf{b}}_t = \underset{\mathbf{b}}{\text{argmin}}\ \left[\ \gamma||\mathbf{b}_{T^c}||_1 + ||\mathbf{y}_t - \mathbf{C}\hat{\mathbf{x}}_{t-1} - \mathbf{C}\mathbf{b}||_2^2\right]$. Please note that unlike $\hat{\mathbf{b}}_t$, $\tilde{\mathbf{b}}_t$ minimizes the residual based on $\hat{\mathbf{x}}_{t-1}$. Error in this estimation corresponds to $\mathbf{e^b}_t = \hat{\mathbf{b}}_t - \tilde{\mathbf{b}}_t$. Obviously, we can conclude that $\mathbf{e}_t^{mod} = \mathbf{e^x}_{t-1} + \mathbf{e^b}_t$.

4.3.1 Support Estimation Error Dynamics

To estimate the support set, the Mod-BPDN-residual provides the sparse signal $\hat{\mathbf{x}}_{t,mod}$ in Step 2 of Algorithm 4. Theorem 3 reveals the relationship between the error in Mod-BPDN-Residual and the sparse signal estimation error. As a precursor to Theorem 3, we first provide Lemma 1 that quantifies the relationship between $\mathbf{e^b}_t$ and $\mathbf{e^x}_t$. Note that $\hat{\mathbf{x}}_{t,mod}$ is a sum of the previous estimated signal and the outcome of an ℓ_1 optimization.

Lemma 1. *Given measurement* $\mathbf{y}_t$ *and estimated sparse signal* $\hat{\mathbf{x}}_t$*, if* $\mathbf{C'C}$ *is an invertible matrix, the* $\mathbf{e^b}_t$ *corresponds to:*

$$\mathbf{e^b}_t = \mathbf{e^x}_{t-1} - \gamma.\mathbf{\Gamma}^{\dagger}\mathbf{\Upsilon}_t \tag{4.14}$$

where,

$$\mathbf{\Gamma} = \begin{bmatrix} \sum_{j=1}^{m} 2c_{j1}c_{j1} & \sum_{j=1}^{m} 2c_{j1}c_{j2} & \cdots & \sum_{j=1}^{m} 2c_{j1}c_{jn} \\ \sum_{j=1}^{m} 2c_{j2}c_{j1} & \sum_{j=1}^{m} 2c_{j2}c_{j2} & \cdots & \sum_{j=1}^{m} 2c_{j2}c_{jn} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{j=1}^{m} 2c_{jn}c_{j1} & \sum_{j=1}^{m} 2c_{jn}a_{j2} & \cdots & \sum_{j=1}^{m} 2c_{jn}c_{jn} \end{bmatrix}, \tag{4.15}$$

and c_{ij} *is the* $(i,j)^{th}$ *element of matrix* $\mathbf{C}$*, and* $\mathbf{\Upsilon}_t$ *is an* $n \times 1$ *vector, which if* $i \in T_t^c$*, then* $(\mathbf{\Upsilon}_t)_i = sgn(\hat{\mathbf{b}}_i) - sgn(\tilde{\mathbf{b}}_i)$*; otherwise* $(\mathbf{\Upsilon}_t)_i = 0$*.*

Proof. Consider,

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}}\ \gamma||\mathbf{b}_{T^c}||_1 + ||\mathbf{y}_t - \mathbf{C}\mathbf{x}_{t-1} - \mathbf{C}\mathbf{b}||_2^2 = \underset{\mathbf{b}}{\operatorname{argmin}}((\sum_{i=1}^{n} \gamma|\mathbf{b}_i|,\ i \in T^c) + \sum_{i=1}^{m}(\mathbf{y}_i - \left[c_{i1}x_{1,t-1} + ... + c_{im}x_{n,t-1}\right] - \left[c_{i1}b_1 + ... + c_{im}b_n\right])^2) \tag{4.16}$$

Since this is an unconstrained convex optimization problem, the optimal value can be ex-

tracted by setting $\bigtriangledown f(\mathbf{b}) = 0$, where $f(\mathbf{b})$ is:

$$f(\mathbf{b}) = \gamma||\mathbf{b}_{T^c}||_1 + ||\mathbf{y}_t - \mathbf{C}\mathbf{x}_{t-1} - \mathbf{C}\mathbf{b}||_2^2 \tag{4.17}$$

Here,

$$\frac{\partial f}{\partial \mathbf{b}_i} = \gamma\frac{\mathbf{b}_i}{|\mathbf{b}_i|} + \sum_{j=1}^{m}(-2c_{ji})(\mathbf{y}_j - \big[c_{j1}x_{1,t-1} + ... + c_{jm}x_{n,t-1}\big] - \big[c_{j1}\mathbf{b}_1 + ... + c_{jm}\mathbf{b}_n\big]) \tag{4.18}$$

Setting $\frac{\partial f}{\partial \mathbf{b}_i} = 0$, $i \in \{1, 2, ..., n\}$, we get:

$$\sum_{j=1}^{m}(-2c_{ji})(\mathbf{y}_j - \big[c_{j1}\mathbf{x}_{1,t-1} + ... + c_{jm}\mathbf{x}_{n,t-1}\big]) = \sum_{j=1}^{m}(-2c_{ji})\big[c_{j1}\hat{\mathbf{b}}_1 + ... + c_{jm}\hat{\mathbf{b}}_n\big] - \gamma\frac{\hat{\mathbf{b}}_i}{|\hat{\mathbf{b}}_i|},\ i \in \{1, 2, ..., n\} \tag{4.19}$$

Now, consider a case that the estimated sparse signal of the system has some errors due to KF, i.e, $\mathbf{x}_t = \hat{\mathbf{x}}_t + \mathbf{e}^{\mathbf{x}}{}_t$. We can rewrite the corresponding equation for $\tilde{\mathbf{b}}_t$ in terms of actual sparse signal and its error as: $\tilde{\mathbf{b}}_t = \underset{\mathbf{b}}{argmin}\ \gamma||\mathbf{b}_{T^c}||_1 + ||\mathbf{y}_t - \mathbf{C}\mathbf{x}_{t-1} - \mathbf{C}\mathbf{b} - \mathbf{C}\mathbf{e}^{\mathbf{x}}{}_{t-1}||^2$. Let $\tilde{f} = \gamma||\mathbf{b}_{T^c}||_1 + ||\mathbf{y}_t - \mathbf{C}\mathbf{x}_{t-1} - \mathbf{C}\mathbf{b} - \mathbf{C}\mathbf{e}^{\mathbf{x}}{}_{t-1}||^2$. The optimal values $\tilde{\mathbf{b}}_t$ is computed by: $\{\tilde{\mathbf{b}}_t \mid \frac{\partial \tilde{f}}{\partial \mathbf{b}_i} = 0\ ,\ i \in \{1, ..., n\}\}$. Here, $\frac{\partial \tilde{f}}{\partial \mathbf{b}_i} = \gamma\frac{\mathbf{b}_i}{|\mathbf{b}_i|} + \sum_{j=1}^{m}(-2c_{ji})(\mathbf{y}_j - \big[c_{j1}\mathbf{x}_{1,t-1} + ... + c_{jn}x_{n,t-1}\big] - \big[c_{j1}\tilde{\mathbf{b}}_1 + ... + c_{jn}\tilde{\mathbf{b}}_n\big] - \big[c_{j1}\mathbf{e}^{\mathbf{x}}{}_{1,t-1} + ... + c_{jn}\mathbf{e}^{\mathbf{x}}{}_{n,t-1}\big]) = 0$.

Since $\frac{\partial f}{\partial \mathbf{b}_i} = \frac{\partial \tilde{f}}{\partial \mathbf{b}_i} = 0$, we have,

$$\sum_{j=1}^{m}(-2c_{ji})(\big[c_{j1}\hat{\mathbf{b}}_1 + ... + c_{jn}\hat{\mathbf{b}}_n\big] - \big[c_{j1}\tilde{\mathbf{b}}_1 + ... + c_{jn}\tilde{\mathbf{b}}_n\big] - \big[c_{j1}\mathbf{e}^{\mathbf{x}}{}_{1,t-1} + ... + c_{jn}\mathbf{e}^{\mathbf{x}}{}_{n,t-1}\big]) = \gamma(\frac{\hat{\mathbf{b}}_i}{|\hat{\mathbf{b}}_i|} - \frac{\tilde{\mathbf{b}}_i}{|\tilde{\mathbf{b}}_i|}),\ i \in \{1, 2, ..., n\} \tag{4.20}$$

Using $\mathbf{e}^{\mathbf{b}}_{i,t} = \hat{\mathbf{b}}_{i,t} - \tilde{\mathbf{b}}_{i,t}$, we can rewrite (4.20) as,

$$\sum_{j=1}^{m}(-2c_{ji})(\left[c_{j1}\mathbf{e}^{\mathbf{b}}{}_{1,t} + ... + c_{jn}\mathbf{e}^{\mathbf{b}}{}_{n,t}\right] = \sum_{j=1}^{m}(-2c_{ji})\left[c_{j1}\mathbf{e}^{\mathbf{x}}{}_{1,t-1} + ... + c_{jn}\mathbf{e}^{\mathbf{x}}{}_{n,t-1}\right]) + \gamma(sgn(\hat{\mathbf{b}}_i) - sgn(\tilde{\mathbf{b}}_i))\ i \in \{1, 2, ..., n\} \quad (4.21)$$

Using our definition of $\mathbf{\Gamma}$ we can rewrite (4.21) as,

$$-\mathbf{\Gamma}.\begin{bmatrix}\mathbf{e}^{\mathbf{b}}{}_{1,t}\\ \mathbf{e}^{\mathbf{b}}{}_{2,t}\\ \vdots\\ \mathbf{e}^{\mathbf{b}}{}_{n,t}\end{bmatrix} = -\mathbf{\Gamma}.\begin{bmatrix}\mathbf{e}^{\mathbf{x}}{}_{1,t-1}\\ \mathbf{e}^{\mathbf{x}}{}_{2,t-1}\\ \vdots\\ \mathbf{e}^{\mathbf{x}}{}_{n,t-1}\end{bmatrix} + \gamma\mathbf{\Upsilon}_t. \quad (4.22)$$

Thus, if $\mathbf{C}'\mathbf{C}$ is invertible, $\mathbf{e}^{\mathbf{b}}{}_t$ corresponds to: $\mathbf{e}^{\mathbf{b}}{}_t = \mathbf{e}^{\mathbf{x}}{}_{t-1} - \gamma.\mathbf{\Gamma}^{\dagger}\mathbf{\Upsilon}_t$. □

The possible error in Step 2 of algorithm 4 ($\mathbf{e}^{mod}_t$) is quantified by theorem 3 and an interpretation of this result is provided following proof of the theorem.

Theorem 3. *If measurements exist at time t, i.e.* $\lambda_t = 1$, *and* $\mathbf{C}'\mathbf{C}$ *is an invertible matrix, we have:*

$$\mathbf{e}^{mod}_t = 2\mathbf{e}^{\mathbf{x}}{}_{t-1} - \gamma.\mathbf{\Gamma}^{\dagger}\mathbf{\Upsilon}_t. \quad (4.23)$$

Proof. $\mathbf{x}_{t,mod} = \mathbf{x}_{t-1} + \hat{\mathbf{b}}_t$. Since the errors associated with the $\mathbf{x}_{t-1}$ and $\hat{\mathbf{b}}_t$ are additive, we have: $\mathbf{e}^{mod}_t = \mathbf{e}^{\mathbf{x}}{}_{t-1} + \mathbf{e}^{\mathbf{b}}{}_t$. Using this equation and Lemma 1, we can infer equation (4.23). □

We can characterize the distribution of $\mathbf{e}^{mod}_t$ by considering the distribution of $\mathbf{e}^{x}_t$. The error in sparse signal estimation ($\mathbf{e}^{\mathbf{x}}{}_t$) is a random vector with a Gaussian distribution, $\mathbf{e}^{\mathbf{x}}{}_t \sim \mathcal{N}(0, \mathbf{P}_{t|t-1})$, since $\mathbf{e}^{\mathbf{x}}{}_t = \mathbf{x}_t - \hat{\mathbf{x}}_t$ is a summation of two independent normal random vectors. The mean of $\mathbf{e}^{\mathbf{x}}{}_t$ corresponds to $\mathbb{E}(\mathbf{e}^{\mathbf{x}}{}_t) = \mathbb{E}(\mathbf{x}_t) - \mathbb{E}(\hat{\mathbf{x}}_t)$. Since $\hat{\mathbf{x}}_t = \mathbb{E}(\mathbf{x}_t)$, we can infer $\mathbb{E}(\mathbf{e}^{\mathbf{x}}{}_t) = 0$. The covariance of $\mathbf{e}^{\mathbf{x}}{}_t$ can be written as $Cov(\mathbf{e}^{\mathbf{x}}{}_t) = \mathbb{E}\left[(\mathbf{e}^{\mathbf{x}}{}_t)(\mathbf{e}^{\mathbf{x}}{}_t)'\right] = \mathbf{P}_{t|t-1}$.

For the second part of $\mathbf{e}_t^{mod}$, while we don't know the actual signs of $\hat{\mathbf{b}}_i$ and $\tilde{\mathbf{b}}_i$, we know that $sgn(\hat{\mathbf{b}}_i) - sgn(\tilde{\mathbf{b}}_i)$ can only take on three values $\{-2, 0, 2\}$. Hence, one possible model is to treat $\boldsymbol{\Upsilon}_t$ as a vector with each element corresponding to a uniform discrete random variable on $\Omega = \{-2, 0, 2\}$. Therefore, $\boldsymbol{\Gamma}^\dagger \boldsymbol{\Upsilon}_t$ represents a weighted summation of discrete random variables. In [69], an exact distribution for sum of discrete uniform variables is derived. However, we can invoke Lindeberg-Feller central limit theorem to approximate the elements of $\boldsymbol{\Gamma}^\dagger \boldsymbol{\Upsilon}_t$ as normal random variables. Thus, if we find mean and covariance of the elements of $\boldsymbol{\Gamma}^\dagger \boldsymbol{\Upsilon}_t$, we can approximate the distribution of the second term in (4.23). Consequently, since $\mathbf{e}^{\mathbf{x}}{}_t$ and $\boldsymbol{\Gamma}^\dagger \boldsymbol{\Upsilon}_t$ are independent Gaussian random vector, we infer that the distribution of $\mathbf{e}_t^{mod}$ is a combination of two Gaussian random vectors.

In step 3 of Algorithm 4, we use $\mathbf{x}_{t,mod}$ to estimate the support set N_t. Since $\mathbf{x}_{t,mod}$ can not be exactly estimated, support set estimation involves some errors. Lemma 2 determines the probability of error in support estimation.

Lemma 2. *Let T_t represent the estimated support set; $\mathbf{e}_t^N$ denote the error in support set estimation such that $\mathbf{e}_t^N = (\{N_t\} \setminus \{T_t\}) \cup (\{T_t\} \setminus \{N_t\})$. Then, probability of error in support set estimation, when the measurements are available ($\lambda_t = 1$) is:*

$$\begin{aligned}\mathbf{Pr}(\mathbf{e}_{i,t}^N \neq \emptyset) = \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a)\big].\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}| > \alpha_a) \mid |(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a\big] + \\ \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a)\big].\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}| < \alpha_a) \mid |(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a\big]\end{aligned} \tag{4.24}$$

Proof. Error in support set estimation can change due to two phenomena. First, when the actual support set has i as its member, but support estimation process has not detected i as a member of T. The second source of error is when the actual support set does not have i as its member, but support estimation process has wrongly identified i as a member of T. These situations are captured in the error vector characterization, $\mathbf{e}_t^N = (\{N_t\} \setminus \{T_t\}) \cup (\{T_t\} \setminus \{N_t\})$. For support set, we have $N_t = \{i : |(\mathbf{x}_{t,mod})_i| > \alpha_a\}$, and for the estimated support set, we use $\hat{\mathbf{x}}_{t,mod}$ to determine T_t as $T_t = \{i : |(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a\}$. Now, we can define the two sources

of error in support estimation as $T_t = \{i : |(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a\}$. Now, we can define the two sources of error in support estimation as: $\{N_t\} \setminus \{T_t\} = \{i : |(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a,\ |(\tilde{\mathbf{x}}_{t,mod})_i| < \alpha_a\} = \{i : |(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a,\ |(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}| < \alpha_a\}$ and $\{T_t\} \setminus \{N_t\} = \{i : |(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a,\ |(\tilde{\mathbf{x}}_{t,mod})_i| > \alpha_a\} = \{i : |(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a,\ |(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}| > \alpha_a\}$. Therefore, the probability of $\mathbf{e}_t^N$ corresponds to : $\mathbf{Pr}(\mathbf{e}_{i,t}^N \neq \emptyset) = \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a) \cap (|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}| > \alpha_a)\big] + \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a) \cap (|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}| < \alpha_a)\big]$. Writing the joint probability in terms of conditional probability yields the result in (4.24).

□

Similar to Theorem 3 that relates the error in step 2 of Algorith 4 ($\mathbf{e}_t^{mod}$) in terms of error in the previous step ($\mathbf{e}^{\mathbf{x}}{}_{t-1}$), Lemma 2 provides the probability of error in Step 3 ($\mathbf{e}_t^N$) of Algorithm 4 in terms of the error resulting from the previous step ($\mathbf{e}_t^{mod}$).

4.3.2 Sparse Signal Estimation Error Dynamics

At the fourth step of Algorithm 4, the sparse signal is estimated by a Kalman filter that uses the estimated support set, previous estimated signal and currently available measurements. However, support set estimation only appears in determining matrix $\hat{\mathbf{Q}}_t$. Thus, $\hat{\mathbf{Q}}_t$ has an additive error due to error in support set estimation. So, let define $\hat{\mathbf{Q}}_t = \tilde{\mathbf{Q}}_t + \mathbf{e}^{\mathbf{Q}}{}_t$ which $\hat{\mathbf{Q}}_t$, $\tilde{\mathbf{Q}}_t$ and $\mathbf{e}^{\mathbf{Q}}{}_t$ are actual covariance matrix, estimated covariance matrix from support estimation and error of covariance matrix estimation, respectively. Theorem 4 provides the covariance of sparse signal prediction in step 4 of Algorithm 4.

Theorem 4. *Assume* $\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}'$ *is an invertible matrix. Also, if the following conditions hold for a Hilbert space* H*:*

- $(\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^\dagger$ *and* $(\mathbf{P}_{t|t-1}\mathbf{C}')^\dagger$ *are bounded,*
- $\mathbf{C}^*H$ *and* $\mathbf{P}_{t|t-1}^*H$ *are invariant under* $(\mathbf{P}_{t|t-1}\mathbf{C}')(\mathbf{P}_{t|t-1}\mathbf{C}')^*$ *and* $\mathbf{C}'\mathbf{C}'^*$*, respectively.*
- $\mathbf{C}^*H \cap Ker\ (\mathbf{P}_{t|t-1}\mathbf{C}')^*$ *and* $\mathbf{P}_{t|t-1}^*H \cap Ker\ (\mathbf{C}')^*$ *are invariant under* $\mathbf{C}^*\mathbf{C}$ *and* $\mathbf{P}_{t|t-1}^*\mathbf{P}_{t|t-1}$*, respectively.*

For the Kalman filter provided in step 4 of Algorithm 4 with arrival observation index λ_t, the covariance of prediction corresponds to:

$$\mathbf{P}_{t+1|t} = (1-\lambda_t)\mathbf{P}_{t|t-1} + \tilde{\mathbf{Q}}_t - \lambda_t \frac{1}{1+g}\mathbf{C}^{\dagger}\mathbf{R}\mathbf{C}'^{\dagger}, \tag{4.25}$$

where, $g = \mathbf{tr}(\mathbf{R}(\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^{-1})$.

Proof. According to the Kalman filter equations, covariance of sparse signal prediction is :

$$\mathbf{P}_{t+1|t} = \mathbf{P}_{t|t-1} + \tilde{\mathbf{Q}}_t - \lambda_t \mathbf{P}_{t|t-1}\mathbf{C}'\left[\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}' + \mathbf{R}\right]^{-1}\mathbf{C}\mathbf{P}_{t|t-1}. \tag{4.26}$$

Using the results of [70] we can assert that if $\mathbf{A}$ and $\mathbf{A}+\mathbf{B}$ are invertible matrices, inverse of summation of the two matrices can be computed by:

$$(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1} - \frac{1}{1+G}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}, \tag{4.27}$$

where $G = \mathbf{tr}(\mathbf{B}\mathbf{A}^{-1})$. According to (4.27), if $\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}'$ and $\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}' + \mathbf{R}$ are invertible matrices, we can rewrite $\left[\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}' + \mathbf{R}\right]^{-1}$ as: $\left[\mathbf{C}\mathbf{P}^{-}_{t|t-1}\mathbf{C}' + \mathbf{R}\right]^{-1} = (\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^{-1} - \frac{1}{1+g}(\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^{-1}\mathbf{R}(\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^{-1})$, where $g = \mathbf{tr}(\mathbf{R}(\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^{-1})$. We use the fact that if inverse of $\mathbf{A}$ exists, then $\mathbf{A}^{\dagger} = \mathbf{A}^{-1}$ [71]. Thus we can assert: $(\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^{-1} = (\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^{\dagger}$. According to *Theorem 3.1.* provided in [72], we have $(\mathbf{A}\mathbf{B})^{\dagger} = \mathbf{B}^{\dagger}\mathbf{A}^{\dagger}$, if and only if, the following conditions hold for a Hilbert space H:

- $(\mathbf{A}\mathbf{B})^{\dagger}$ is bounded,
- $\mathbf{A}^{*}H$ is invariant under $\mathbf{B}\mathbf{B}^{*}$
- $\mathbf{A}^{*}H \cap Ker\ \mathbf{B}^{*}$ is invariant under $\mathbf{A}^{*}\mathbf{A}$.

Therefore, under mentioned circumstances[72], the following equation holds:

$$(\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^{\dagger} = \mathbf{C}'^{\dagger}\mathbf{P}^{\dagger}_{t|t-1}\mathbf{C}^{\dagger}. \tag{4.28}$$

Applying this relationship to (4.26), we have:

$$\mathbf{P}_{t+1|t} = \mathbf{P}_{t|t-1} + \tilde{\mathbf{Q}}_t - \lambda_t \mathbf{P}_{t|t-1}\mathbf{C}^{'}\left[\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}^{'} + \mathbf{R}\right]^{-1}\mathbf{C}\mathbf{P}_{t|t-1} =$$

$$\mathbf{P}_{t|t-1} + \tilde{\mathbf{Q}}_t - \lambda_t \mathbf{P}_{t|t-1}\mathbf{C}^{'}((\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}^{'})^{-1} - \frac{1}{1+g}(\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}^{'})^{-1}(\mathbf{R})(\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}^{'})^{-1})\mathbf{C}\mathbf{P}_{t|t-1} =$$

$$\mathbf{P}_{t|t-1} + \tilde{\mathbf{Q}}_t - \lambda_t \mathbf{P}_{t|t-1}\mathbf{C}^{'}(\mathbf{C}^{'\dagger}\mathbf{P}^{\dagger}_{t|t-1}\mathbf{C}^{\dagger} - \frac{1}{1+g}C^{'\dagger}\mathbf{P}^{\dagger}_{t|t-1}\mathbf{C}^{\dagger}(\mathbf{R})\mathbf{C}^{'\dagger}\mathbf{P}^{\dagger}_{t|t-1}\mathbf{C}^{\dagger})\mathbf{C}\mathbf{P}_{t|t-1} \quad (4.29)$$

Simplification of equation (4.29) leads to find the covariance of prediction as $\mathbf{P}_{t+1|t} = (1 - \lambda_t)\mathbf{P}_{t|t-1} + \tilde{\mathbf{Q}}_t - \lambda_t \frac{1}{1+g}\mathbf{C}^{\dagger}\mathbf{R}\mathbf{C}^{'\dagger}$. □

In summary, Theorem 4 quantifies the dynamics of error covariance in sparse signal prediction. In the next section, we study the stochastic stability of this dynamical error process.

4.4 Main Results

Accuracy of some of the sparse reconstruction methods has been studied in the previous papers. For example in [73], it is shown that the Basis Pursuit (BP) approach exactly recovers a sparse noiseless signal under certain assumptions. **Theorem 1.3** in [73] provides an upper bound for second norm of error when the sparse signal is noisy and reconstructed by Noisy-BP. Time invariant bounds for error of recovery in Modified-CS is obtained in [74]. Quantifying the efficiency of the proposed method in terms of the least possible error of signal recovery is one of the goals of this dissertation. To this end, we provide an upper bound for variances of error in sparse signal estimation using $\mathbf{P}_{t|t-1}$ and $\mathbf{e}_t^{mod}$ derived in the previous section. The focus on the diagonal elements (i.e., variances) of $\mathbf{P}_{t|t-1}$ is motivated by our system model assumption wherein: (1) the different elements of the sparse signal do not affect each other, i.e., $(\mathbf{x}_{t+1})_i$ is just a function of $(\mathbf{x}_t)_i$ in (4.3); (2) $\mathbf{Q}_t$ is assumed to be a diagonal matrix. In step 4 of Algorithm 4, the estimated support set determines the diagonal elements of $\tilde{\mathbf{Q}}$ and therefore, $\mathbf{e}^{\mathbf{Q}}{}_t$ is also a diagonal matrix. This implies that the support set estimation just impacts the diagonal elements of $\mathbf{P}_{t|t-1}$. Hence, we embed the

diagonal elements of $\mathbf{P}_{t|t-1}$ in a vector and its dynamics correspond to,

$$diag(\mathbf{P}_{t+1|t}) = (1-\lambda_t)diag(\mathbf{P}_{t|t-1}) + diag(\tilde{\mathbf{Q}}_t) - \lambda_t \frac{1}{1+g} diag(\mathbf{C}^{\dagger}\mathbf{R}\mathbf{C}'^{\dagger}). \tag{4.30}$$

Furthermore, it should be noted that $\mathbf{e^Q}_t$ is a diagonal matrix with diagonal values depending on the element membership status within the support set. More precisely, if $\{i \in N_t, T_{i,t} \neq \emptyset\}$, then $\mathbf{e^Q}_{i,t} = 0$; otherwise $\mathbf{e^Q}_{i,t} = -\sigma^2$. If $\{i \notin N_t, T_{i,t} = \emptyset\}$, then $\mathbf{e^Q}_{i,t} = 0$; otherwise $\mathbf{e^Q}_{i,t} = \sigma^2$. Thus, it is reasonable to model $\mathbf{e^Q}_{i,t}$ as a Bernoulli random variable.

The analysis of $\mathbf{P}_{t+1|t}$ is challenging since $diag(\mathbf{P}_{t+1|t})$ consists of three random parts; $\mathbf{P}_{t|t-1}$, λ_t and $\mathbf{e^Q}_t$. Although, λ_t is independent of $\mathbf{P}_{t|t-1}$ and $\mathbf{e^Q}_t$, the Bernoulli random vector $\mathbf{e^Q}_t$ is dependent on $\mathbf{P}_{t|t-1}$. This complicates the evolution of the distribution of $\mathbf{P}_{t+1|t}$. Therefore, we limit our analysis to study the mean of $\mathbf{P}_{t+1|t}$ with respect to λ_t and $\mathbf{e^Q}_t$. A similar idea was developed for standard Kalman filter in [28]. If we take the mean of $P_{t|t-1}$ with respect to λ_t and $\mathbf{e^Q}_t$, we have:

$$\begin{aligned} diag(\mathbb{E}_{\lambda_t,\mathbf{e^Q}}\mathbf{P}_{t+1|t}) = (1-\bar{\lambda})diag(\mathbb{E}_{\lambda_t,\mathbf{e^Q}}\mathbf{P}_{t|t-1}) + \mathbb{E}_{\lambda_t,\mathbf{e^Q}}diag(\hat{\mathbf{Q}}_t) \\ -\bar{\lambda}diag(\mathbb{E}_{\lambda_t,\mathbf{e^Q}}\frac{1}{1+g}\mathbf{C}^{\dagger}\mathbf{R}\mathbf{C}'^{\dagger})) + diag(\mathbb{E}_{\lambda_t),\mathbf{e^Q}}(\mathbf{e^Q}_t), \end{aligned} \tag{4.31}$$

where $\bar{\lambda} = \mathbb{E}_{\lambda_t}(\lambda_t)$ is the rate of missing measurements. Using lemma 2, we conclude that: if $\{i \in N_t\}$, then $\mathbb{E}_{\lambda_t,\mathbf{e^Q}}\mathbf{e^Q}_{i,t} = -\sigma^2\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a)\big].\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}^{mod}_{i,t}| < \alpha_a) \mid |(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a\big]$; and similarly, if $\{i \notin N_t\}$, then, $\mathbb{E}_{\lambda_t,\mathbf{e^Q}}\mathbf{e^Q}_{i,t} = \sigma^2\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a)\big].\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}^{mod}_{i,t}| > \alpha_a) \mid |(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a\big]$. In order to evaluate the above probabilities, the following two lemmas are provided for the two different cases.

Lemma 3. *For $\{i \in N_t\}$, if there exist α_a and $U = \frac{\alpha_a^2}{4\mathbf{P}_{i,t}+\mathbf{D}_{i,t}}$, such that $\alpha_a U^4 + 2\alpha_a U^3 + (2-4\alpha_a)U^2 + (4-6\alpha_a)U + 2 + 3\alpha_a \geq 0$ and measurements are available at time t, i.e., $\lambda_t = 1$, then,*

$$|\mathbb{E}_{\lambda_t,\mathbf{e^Q}}(\mathbf{e^Q}_{i,t})| \leq \sigma^2\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a)\big].(1 - \sqrt{\frac{2}{\pi}}\frac{h}{1+h^2}e^{-\frac{h^2}{2}}), \tag{4.32}$$

where $h = \dfrac{\alpha_a}{\sqrt{4\mathbb{E}_{\lambda_t,\mathbf{e^Q}}\mathbf{P}_{i,t|t-1} + \mathbf{D}_{i,t}}}$, *and* $\mathbf{D}_{i,t}$ *is variance of* $(\hat{\mathbf{x}}_{t,mod})_i - \gamma.(\mathbf{\Gamma}^\dagger \mathbf{\Upsilon}_t)_i$.

Proof. See A.1 □

Lemma 4. *For* $\{i \notin N_t\}$, *if there are* α_a *and* $q = \frac{\alpha_a}{(4\mathbf{P}_{i,t}+\mathbf{D}_{i,t})^{\frac{1}{2}}}$, *such that* $q^2 < 3$, *and measurements are available at time* t, *i.e.,* $\lambda_t = 1$, *then,*

$$\mathbb{E}_{\lambda_t,\mathbf{e^Q}}(\mathbf{e^Q}_{i,t}) \leq \sqrt{\frac{2}{\pi}}\sigma^2 \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a)\big].\int_h^\infty e^{-\frac{s^2}{2}} ds. \tag{4.33}$$

Proof. See A.2 □

It is important ti note that these two lemmas hold only when the measurements are available. If at time t, the measurements are lost ($\lambda_t = 0$), we use the previous estimated support set to determine $\hat{\mathbf{Q}}_t$. Therefore, we can generalize the above two lemmas: if $\{i \in N_t\}$, then $|\mathbb{E}_{\lambda_t,\mathbf{e^Q}}(\mathbf{e^Q}_{i,t})| \leq \bar{\lambda}\sigma^2 \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a)\big](1 - \sqrt{\frac{2}{\pi}}\frac{h}{1+h^2}e^{-\frac{h^2}{2}}) + (1-\bar{\lambda})|\mathbb{E}_{\lambda_t,\mathbf{e^Q}}(\mathbf{e^Q}_{i,t-1})|$; and similarly if $\{i \notin N_t\}$, then $\mathbb{E}_{\lambda_t,\mathbf{e^Q}}(\mathbf{e^Q}_{i,t}) \leq \bar{\lambda}\sqrt{\frac{2}{\pi}}\sigma^2 \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a)\big].\int_h^\infty e^{-\frac{s^2}{2}} ds + (1-\bar{\lambda})|\mathbb{E}_{\lambda_t,\mathbf{e^Q}}(\mathbf{e^Q}_{i,t-1})|$. Furthermore, if we consider the maximum value of $|\mathbb{E}_{\lambda_t,\mathbf{e^Q}}(\mathbf{e^Q}_{i,t})|$, the inequalities presented in lemma 3 and lemma 4 become equalities. To simplify the notations, for the rest of this chapter, we consider $\mathbb{E}_{\lambda_t,\mathbf{e^Q}}\mathbf{P}_{i,t|t-1} \leq \bar{\bar{P}}_t$, and $\mathbb{E}_{\lambda_t,\mathbf{e^Q}}\mathbf{e^Q}_{i,t} \leq \bar{\bar{e}}^Q_t$, and $\mathbf{l}_t = diag(\hat{\mathbf{Q}}_t)_{i,i} - \bar{\lambda}\mathbb{E}_{\lambda_t,\mathbf{e^Q}} diag(\frac{1}{1+g}\mathbf{C}^\dagger \mathbf{R}\mathbf{C}'^\dagger)_{i,i}$, where,

$$\bar{\bar{P}}_{t+1} = (1-\bar{\lambda})\bar{\bar{P}}_t + \bar{\bar{e}}^Q_t + \mathbf{l}_{i,t}$$

$$\bar{\bar{e}}^Q_{t+1} = \begin{cases} (1-\bar{\lambda})\bar{\bar{e}}^Q_t + \bar{\lambda}K_1(1 - \frac{h}{1+h^2}e^{-\frac{h^2}{2}}), & \text{if } i \in N_t \\ (1-\bar{\lambda})\bar{\bar{e}}^Q_t + \bar{\lambda}K_2 \int_h^\infty e^{-\frac{s^2}{2}} ds, & \text{if } i \notin N_t. \end{cases} \tag{4.34}$$

Here, $K_1 = \sigma^2 \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a)\big]$, and $K_2 = \sqrt{\frac{2}{\pi}}\sigma^2 \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a)\big]$. From (4.34) it is evident that $\bar{\bar{P}}_t$ and $\bar{\bar{e}}^Q_t$ form a coupled nonlinear dynamical system. It should be recalled that both $\bar{\bar{P}}_t$ and $\bar{\bar{e}}^Q_t$ are scalars. Obviously, if $\mathbf{l}_{i,t} = 0$ (i.e., unforced system), the system is stable for all the $\bar{\lambda} > 0$. However, since $\mathbf{l}_{i,t} \neq 0$, we can study the input-to-state stability of

the coupled nonlinear system (4.34) using the results presented in [75]. For $(\mathbf{x}_t)_i$, where i is a member of the support set (i.e., $i \in N_t$), Theorem 5 provides input-to-stability for the first and second equation of (4.34).

Theorem 5. *The interconnected system of* (4.34) *is input-to-state stable, for all* $\bar{\lambda} > 0$ *and* $i \in N_t$, *if* $\xi_1 \circ \xi_2(s) < s$ *for all* $s > 0$, *where* $\xi_1(s) = \frac{s}{\bar{\lambda}}$, $\xi_2(s) = K_1(1 - \frac{h_1(s)}{1+h_1^2(s)} e^{-\frac{h_1^2(s)}{2}} - \tau_0)$, $h_1(s) = \frac{\alpha_a}{\sqrt{4s + \mathbf{D}_i}}$ *and* $\tau_0 = 1 - \frac{h_1(0)}{1+h_1^2(0)} e^{-\frac{h_1^2(0)}{2}}$.

Proof. According to Example 3.4. of [75], $\bar{\bar{P}}_{t+1} = (1 - \bar{\lambda})\bar{\bar{P}}_t + v_1(t) + u_1(t)$ is input-to-state stable, with two inputs $v_1(t) = \bar{\bar{e}}_t^Q$ and $u_1(t) = l_t$. Additionally,

$$|\bar{\bar{P}}_t| \leq (1 - \bar{\lambda})^t \bar{\bar{P}}_0 + \xi_1(||v_1||) + \xi_1(||u_1||) \tag{4.35}$$

where $\xi_1(r) = \frac{r}{\bar{\lambda}}$ is a $\mathcal{K}$-function. In order to prove input-to-stability of the second equation in (4.34), we select the input-to-state stability gain function $\xi_2(s)$ to be a $\mathcal{K}$-function. Similar to $\bar{\bar{P}}_t$, $\bar{\bar{e}}_{t+1}^Q = (1 - \bar{\lambda})\bar{\bar{e}}_t^Q + \bar{\lambda} K_1(1 - \frac{h_1(v_2)}{1+h_1^2(v_2)} e^{-\frac{h_1^2(v_2)}{2}})$ is input-to-state stable with $v_2(t) = \bar{\bar{P}}_t$ and $u_2(t) = \bar{\lambda} K_1 \tau_0$ as its inputs. Thus we have,

$$|\bar{\bar{e}}_t^Q| \leq (1 - \bar{\lambda})^t \bar{\bar{e}}_0^Q + \xi_2(||v_2||) + \xi_1(||u_2||) \tag{4.36}$$

where $\xi_2(v_2) = K_1(1 - \frac{h_1(v_2)}{1+h_1^2(v_2)} e^{-\frac{h_1^2(v_2)}{2}} - \tau_0)$ is its ISS-gain function. According to the nonlinear small gain theorem (Theorem 2.) of [75], if the first and second equation of (4.34) are individually input-to-state stable and $\xi_1 \circ \xi_2(s) < s$ for all $s > 0$, then the interconnected system of (4.34) is input-to-state stable. □

If i is not a member of the support set (i.e., $i \notin N_t$), the following theorem shows input-to-state stability of (4.34).

Theorem 6. *The interconnected system of* (4.34) *is input-to-state stable, for all* $\bar{\lambda} > 0$ *and* $i \notin N_t$, *if* $\xi_1 \circ \xi_3(s) < s$ *for all* $s > 0$, *where* $\xi_3(s) = K_2(\int_{h_1(s)}^{\infty} e^{-\frac{s^2}{2}} ds - \varkappa_0)$, $\varkappa_0 = \int_{h_1(0)}^{\infty} e^{-\frac{s^2}{2}} ds$

and $u_3(t) = \bar{\lambda} K_2 \varkappa_0$. *Additionally,*

$$
\begin{aligned}
|\bar{\bar{P}}_t| &\leq (1-\bar{\lambda})^t \bar{\bar{P}}_0 + \xi_1(||v_1||) + \xi_1(||u_1||) \\
|\bar{\bar{e}}_t^Q| &\leq (1-\bar{\lambda})^t \bar{\bar{e}}_0^Q + \xi_3(||v_2||) + \xi_1(||u_3||)
\end{aligned} \tag{4.37}
$$

Proof. The proof is similar to the proof of Theorem 5. □

Remark 1. While we consider $\mathbf{l}_{i,t}$ as an input independent from the states, this is not actually true. Their dependency is captured in g that appears in the denominator of the second term: $\bar{\lambda}\mathbb{E}_{\lambda_t,\mathbf{e_Q}}(\frac{1}{1+g} C^\dagger \mathbf{R} \mathbf{C}'^\dagger))$. Since $\mathbf{R} = \sigma_{sys}^2 \mathbf{I}$, we can rewrite g as $g = \mathbf{tr}(\sigma_{sys}^2 (\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^{-1})$. Now assume we design a $\mathbf{C}$ such that $\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}' > 0$. If $\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}' > 0$ we can say $(\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^{-1}$ is a positive definite (p.d) matrix. This implies that $g = \mathbf{tr}((\mathbf{C}\mathbf{P}_{t|t-1}\mathbf{C}')^{-1})$ is always nonnegative. Thus, if we assume $\hat{\mathbf{l}}_t = diag(\hat{\mathbf{Q}}_t) - \bar{\lambda} diag(\mathbf{C}^\dagger \mathbf{R} \mathbf{C}'^\dagger))$, Theorem 5 and Theorem 6 still hold, since $\hat{\mathbf{l}}_t \geq \mathbf{l}_t$.

Theorem 7. *If all the mentioned conditions hold,* (4.38) *provides an upper bound for the two possible cases,*

$$
\begin{cases}
|\bar{\bar{P}}_t| \leq \Xi_1^{-1}((1-\bar{\lambda})^t \bar{\bar{P}}_0 + \dfrac{(1-\bar{\lambda})^t}{\bar{\lambda}} \bar{\bar{e}}_0^Q + \dfrac{K_1 \tau_0}{\bar{\lambda}} + \xi_1(||\hat{\mathbf{l}}_t||)), & \text{if } i \in N_t \\
|\bar{\bar{P}}_t| \leq \Xi_2^{-1}((1-\bar{\lambda})^t \bar{\bar{P}}_0 + \dfrac{(1-\bar{\lambda})^t}{\bar{\lambda}} \bar{\bar{e}}_0^Q + \dfrac{K_2 \varkappa_0}{\bar{\lambda}} + \xi_1(||\hat{\mathbf{l}}_t||)) & \text{if } i \notin N_t
\end{cases} \tag{4.38}
$$

where, $\Xi_1(s) = s - \xi_1 \circ \xi_2(s)$ *for* $s > 0$ *and* $\Xi_2(s) = s - \xi_1 \circ \xi_3(s)$ *for* $s > 0$.

Proof. Inequality (4.38) is directly obtained from the results of Theorem 5 and Theorem 6. □

Remark 2. Analytical inverse of the functions $\Xi_1(s)$ and $\Xi_2(s)$ may not be straightforward. Since $\xi_1 \circ \xi_3(s) < s$, one possible approximation is to treat $\xi_1 \circ \xi_2(s)$ as linear function, which considerably simplifies the calculation based on $\Xi_1^{-1}(s)$.

The upper bounds given by Theorem 7 enable us to find a critical value for the rate of receiving measurements ($\bar{\lambda}^c$) that ensures the stability of error in Algorithm 4. That is, if

the probability of measurement loss is smaller than $1-\bar{\lambda}^c$, we can ensure that error in signal estimation is bounded.

4.4.1 Unaltered Support Set

In some special situations, the underlying support set may remain constant during a time interval. For example, in smart distribution systems, power and voltage signals may not change very quickly during a certain time interval. This fact suggests that the support set remains constant for a period of time. In this case, we only need to accurately estimate the support set once. That is, Step 2 and Step 3 in Algorithm 4 can be skipped. Consequently, error analysis is limited to the Kalman filter step (Step 4). Therefore, we use our previous analysis to find a tighter bound for $\hat{P}_t$ when we assume the exact support set is known. In this case, we have $\tilde{\mathbf{Q}} = \hat{\mathbf{Q}}$ and $\mathbf{e}_t^{\mathbf{Q}} = 0$ always holds. This modifies (4.30) as,

$$\begin{aligned} diag(\mathbf{P}_{t+1|t}) = (1-\lambda_t) diag(\mathbf{P}_{t|t-1}) + diag(\hat{\mathbf{Q}}_t) \\ -\lambda_t \frac{1}{1+g} diag(\mathbf{C}^\dagger \mathbf{R} \mathbf{C}'^\dagger). \end{aligned} \tag{4.39}$$

Theorem 8 updates the upper bound for the expected variance of the estimation error for a given rate of information loss.

Theorem 8. *If the underlying support set is fixed and precisely known, the system in* (4.40) *is input to state stable with the following bound,*

$$|\hat{P}_t| \leq (1-\bar{\lambda})^t \hat{P}_0 + \xi_1(||\hat{l}_t||), \tag{4.40}$$

Proof. According to Example 3.4. of [76], $\bar{\bar{P}}_{t+1} = (1-\bar{\lambda})\bar{\bar{P}}_t + u_1(t)$ is input-to-state stable, with input $u_1(t) = \hat{l}_t$. Additionally,

$$|\hat{P}_t| \leq (1-\bar{\lambda})^t \bar{\bar{P}}_0 + \xi_1(||u_1||) \tag{4.41}$$

where $\xi_1(r) = \frac{r}{\bar{\lambda}}$ is a $\mathcal{K}$-function. □

The upper bounds given by Theorem 7 and Theorem 8 enable us to find a critical value for the rate of receiving measurements ($\bar{\lambda}^c$) that ensures the stability of error in Algorithm 4. That is, if the probability of measurement loss is smaller than $1 - \bar{\lambda}^c$, we can ensure that error in signal estimation is bounded. The following subsection analyzes steady state behavior of the covariance matrix for a special type of measurement matrix $\mathbf{C}$.

4.4.2 Reduced Order Case

Let $\tilde{N}$ denote the accurate support estimation of the sparse vector $\mathbf{x}_0$. Then, assume that the support set does not change over a period of time (i.e., $N_t = \tilde{N}$ where $size(\tilde{N}) = r$). Therefore, the system in (4.2) is not time-varying anymore since N_t is constant over that period. This implies that $(\mathbf{x}_t)_i = 0$, $\forall i \notin \tilde{N}$. Let $\mathbf{x}_t^r \in \mathbb{R}^r$ denote a vector that contains nonzero elements of $\mathbf{x}_t$ (i.e., $\mathbf{x}_t^r = \{(\mathbf{x}_t)_i | i \in \tilde{N}\}$). Consequently, the dimension of the covariance matrix of process $\mathbf{Q}_t$ is reduced to a r by r diagonal matrix with nonzero elements on its diagonal ($\mathbf{Q}^r = \sigma_{sys}\mathbf{I}_{r\times r}$). Furthermore, we only keep the i^{th} columns of the measurement matrix, where $i \in \tilde{N}$. To clarify this modification, let rewrite the observation matrix as

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & ... & \mathbf{C}_i & ... & \mathbf{C}_m \end{bmatrix}, \tag{4.42}$$

where, $\mathbf{C}_i \in \mathbb{R}^n$ is i^{th} vector of $\mathbf{C}$. Therefore, the reduced order measurement matrix is $\mathbf{C}^r = \{[\mathbf{C}_i] | i \in \tilde{N}\}$ and $\mathbf{R}^r$ denote covariance of measurement noise corresponding to $\mathbf{C}^r$.

Lemma 5. *If the support set remains unaltered and $rank(\mathbf{C}^r) \geq r$, then we have the following steady-state upper bound for state covariance matrix.*

$$lim_{t\to\infty} \mathbb{E}_{\lambda_t} \mathbf{P}_{t+1|t} \leq \mathbf{F}, \tag{4.43}$$

where $\mathbf{F}$ *can be calculated by solving the following equation:*

$$\mathbf{Q} = \bar{\lambda}\mathbf{F}\mathbf{C}^{r^{'}}(\mathbf{C}^{r}\mathbf{F}\mathbf{C}^{r^{'}} + \mathbf{R})^{-1}\mathbf{C}^{r}\mathbf{F}. \tag{4.44}$$

Proof. We first consider observability and controllability of the reduced order system. Pair $(\mathbf{I}_{r\times r}, \mathbf{Q}^r)$ is always controllable, since all the diagonal elements of $\mathbf{Q}^r$ are nonzero and consequently, $rank(\mathbf{Q}^r) = r$. Similarly, for pair $(\mathbf{I}_{r\times r}, \mathbf{C}^r)$ the corresponding observability matrix is:

$$\mathcal{O}_r = \begin{bmatrix} \mathbf{C}^r \\ \mathbf{I}_{r\times r}\mathbf{C}^r \\ \vdots \\ (\mathbf{I}_{r\times r})^{r-1}\mathbf{C}^r \end{bmatrix}$$

Obviously, $rank(\mathcal{O}_r) = rank(\mathbf{C}^r)$. To satisfy observability $rank(\mathcal{O}_r) \geq r$ must hold that is equivalent to $rank(\mathbf{C}^r) \geq r$. According to *Theorem 4* in [77], if the pairs $(\mathbf{I}_{r\times r}, \mathbf{Q}^r)$ is controllable and $(\mathbf{I}_{r\times r}, \mathbf{C}^r)$ is observable, the following bound holds:

$$lim_{t\to\infty}\mathbb{E}_{\lambda_t}\mathbf{P}_{t+1|t} \leq \mathbf{F},$$

where $\mathbf{F}$ is the fixed point of equation (4.45) (i.e., $\mathbf{F} = g_{\bar{\lambda}}(\mathbf{F})$).

$$g_{\bar{\lambda}}(\mathbf{X}) = \mathbf{A}\mathbf{X}\mathbf{A}^{'} + \mathbf{Q}^r - \bar{\lambda}\mathbf{A}\mathbf{X}\mathbf{C}^{r^{'}}(\mathbf{C}^{r}\mathbf{X}\mathbf{C}^{r^{'}} + \mathbf{R})^{-1}\mathbf{C}^{r}\mathbf{X}\mathbf{A}^{'} \tag{4.45}$$

By substituting $\mathbf{A} = \mathbf{I}$, the simplified equation is equal to (4.44). □

4.5 Simulation Results

Here, we evaluate our theoretical results using two examples. We first consider a time sequence of a sparse signal $\mathbf{x}_t$ with length $n = 256$, where the maximum sparsity $S_{max} = 20$. We use the $C_{m\times n}$ matrix that is simulated in [64] by generating $m \times n$ i.i.d. Gaussian entries (with $m = 128$) and normalizing each column of the resulting matrix. The variance

of observation noise is $\sigma^2_{obs} = \frac{1}{\sqrt{3}m} S_{max}$. The dynamic model of $\mathbf{x}_t$ is represented by (4.3) with $\sigma^2_{sys} = 1$. N_1 (the support set of $\mathbf{x}_1$) is obtained by generating $S_{max} - 12$ unique indices randomly from $\big[1 : n\big]$. Then, the support set is increased at $t = 10$. That is, $N_t = N_1$, $\forall t < 10$, while at $t = 10$, we add four more elements to the support set. Thus, $N_t = N_{10}$, $\forall 10 < t < 20$. Similarly, there are four additions in support set at $t = 20$ and $t = 30$. Therefore, at time $t = 30$, we reach the maximum size of sparsity $S_{max} = 20$.

We apply the online updated KF-CS algorithm (Algorithm 4), to compute the estimate of $\mathbf{x}_t$ at each t (i,e,. $\hat{\mathbf{x}}_t$). The normalized error of state estimation corresponding $\frac{||\mathbf{x}_t - \hat{\mathbf{x}}_t||_2}{||\mathbf{x}_t||_2}$ is used as the metric to illustrate estimation performance. As stated earlier, measurements losses are modeled as a Bernoulli random variable with three different rates $\bar{\lambda} = 0.5$, $\bar{\lambda} = 0.75$ and $\bar{\lambda} = 1$ (all measurements are available). Also, we choose $\gamma = 4$ and the thresholding parameter $\alpha_a = 0.1$. Discussion about parameter setting is provided in [38]. The results averaged over 200 Monte-Carlo simulation runs that are presented in Figures 4.2 through 4.4.

Figure 4.2 illustrates the normalized error in state estimation for our system with three different rates of lossy measurements. It is important to note that the normalized error of state estimation is below 16%, even when only 50% of the measurements are available over time. Obviously, when $t = 10$, 20 and 30, there are sharp increases in the error due to changes in actual support set at the mentioned times. However, this error is refined as time proceeds, since the support set is estimated more precisely. Figure 4.3 depicts the error in support set estimation ($\mathbf{e}_t^N$), i.e., the corresponding number for the members of support set that are not detected by the compressive sensing step (Step 3 of Algorithm 4) or are detected incorrectly. For initialization, we assume that we know the exact support set. It should be noted that since we plot the average values over multiple runs, the number of errors is not an integer value. Figure 4.4 illustrates the behavior of the variance of the 6^{th} state element (i.e., the 6^{th} element on diagonal of covariance matrix $\mathbb{E}_{\lambda_t, \mathbf{e_Q}} \mathbf{P}_{t|t-1}$). We assume that $\mathbf{P}_{0|0} = 0$. Also consider that $(\hat{\mathbf{Q}}_t)_{6,6} = sig^2_{sys}$ which means the sixth element belongs to the support set ($(\mathbf{x}_t)_6 \in N$). We also approximate Ξ_1 by a linear function. Obviously, simulated $(\bar{\bar{P}}_t)_{6,t}$

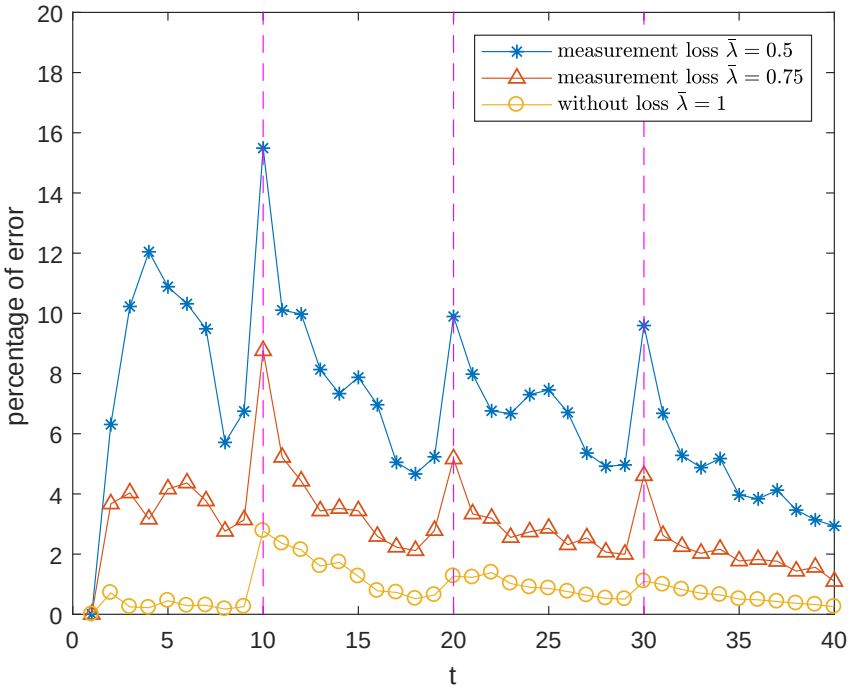

Figure 4.2: *Normalized error in sparse signal estimation*

always lies under the theoretical bound provided by Theorem 7. The simulation results validate our theoretical results(bounds) and confirm that the proposed approach is robust in the presence of lossy measurements.

4.5.1 Distribution System Example

As we mentioned earlier, the signal of interest $\mathbf{z}$ is not a sparse vector itself. However, we transform it to a sparse signal $\mathbf{x}$ by a linear transformation $\mathbf{\Psi}$. Here, we apply the DCT as the sparsifying basis $\mathbf{\Psi}$. It should be noted that the transformation bases may vary for different signals in distribution systems. Here, we apply a random measurement matrix $\mathbf{\Phi}$ consisting of random Bernoulli entries which selects $\frac{m}{n} \times 100\%$ of the data. We implement Algorithm 4 on IEEE 37-test feeder system that is a three phase unbalanced distribution system. For a realistic simulation setup, we collect data from actual consumers that is publicly available at eGauge website [2]. Here, we use the strategy described in [78] where we employ the load data of particular homes to a specific node in IEEE 37 test feeder system. The measurements

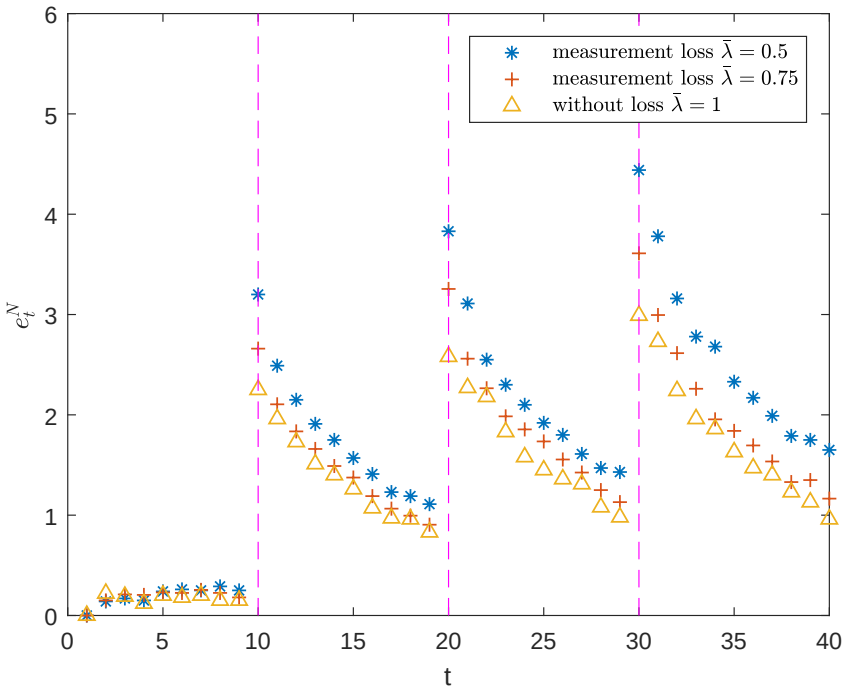

Figure 4.3: *Error in support set estimation*

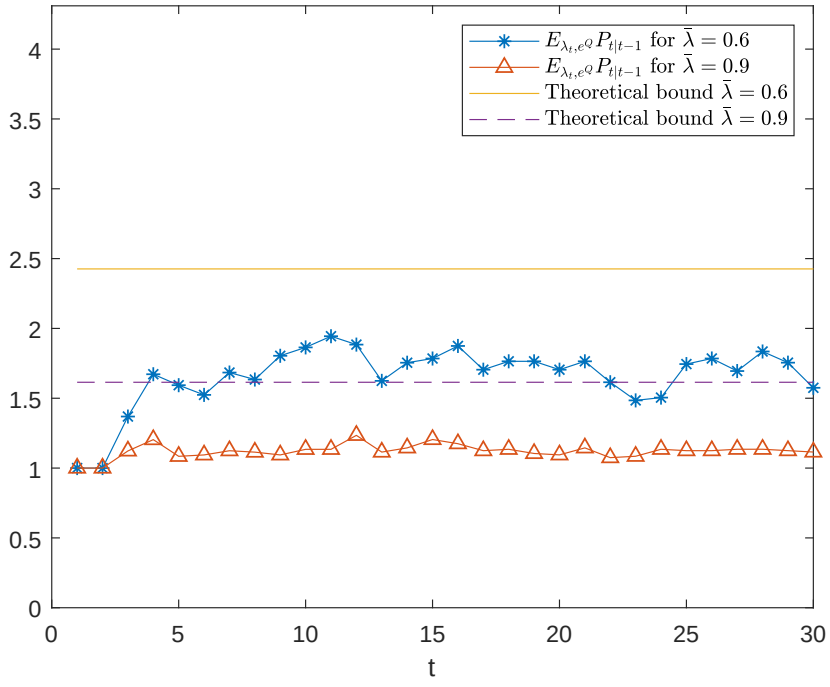

Figure 4.4: $E_{\lambda_t,\mathbf{e}^{\mathbf{Q}}}\mathbf{P}_{6,t|t-1}$ *and the theoretical bound of* $\bar{\bar{P}}_{6,t}$

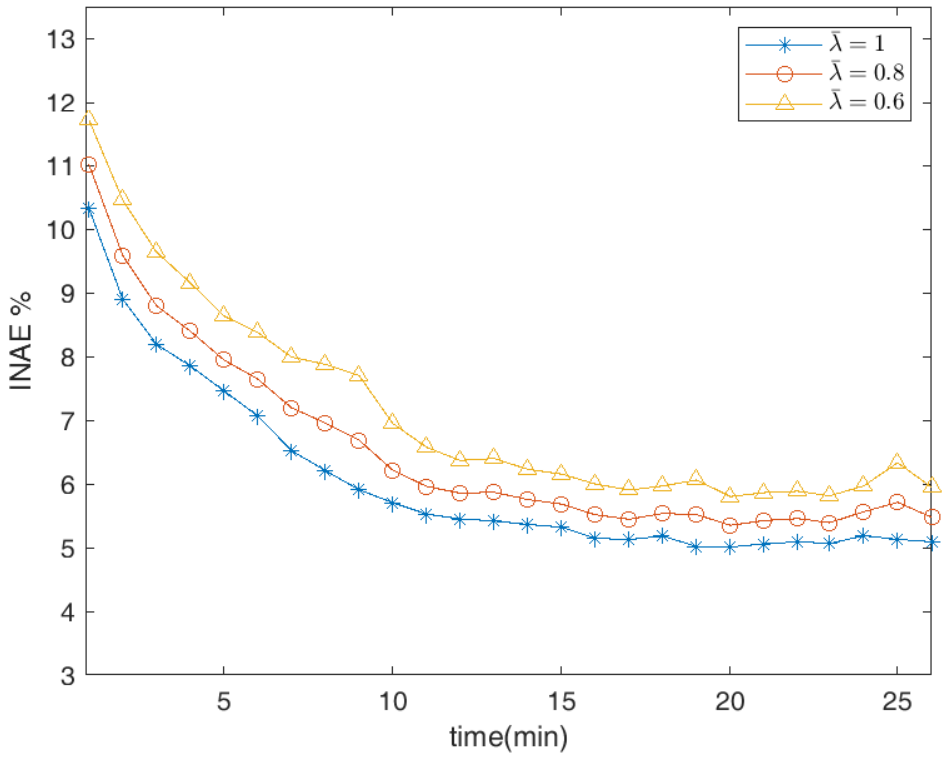

Figure 4.5: *INAE corresponding to state estimation for different* $\bar{\lambda}$

are noisy with $\boldsymbol{\nu}_t \sim \mathcal{N}(0, 0.4\mathbf{I}_{m\times m})$. In Algorithm 4, we set $\gamma = 0.075$ and $\alpha_a = 0.25$.

Due to random nature of information loss, we perform Monte-Carlo simulations and show the averaged results. Here, we provide the averages across 3000 simulations. Fig 4.5 shows *INAE* performance when $CMR = 35\%$ with different measurement loss rates $\bar{\lambda}$. Here, we can make two important observations. First, the *INAE* settles down after approximately 15 minutes. Second, not surprisingly, the error increases with decreasing availability of measurements.

Figure 4.6 illustrates the behavior of the variance of the 4^{th} state element (i.e., the 4^{th} element on diagonal of covariance matrix $\mathbb{E}_{\lambda_t,\mathbf{e}^{\mathbf{Q}}}\mathbf{P}_{t|t-1}$). We assume that $\mathbf{P}_{0|0} = 10$. Also consider that $(\hat{\mathbf{Q}}_t)_{4,4}$ has a value that has been correctly identified. This means the fourth state belongs to the support set ($(\mathbf{x}_t)_4 \in N$). As discussed in Remark 1, we approximate Ξ_1 by a linear function. As seen in Figure 4.6, the simulated $(\bar{\bar{P}}_t)_{4,t}$ always lies under the theoretical bound provided by Theorem 7. The simulation results validate our theoretical

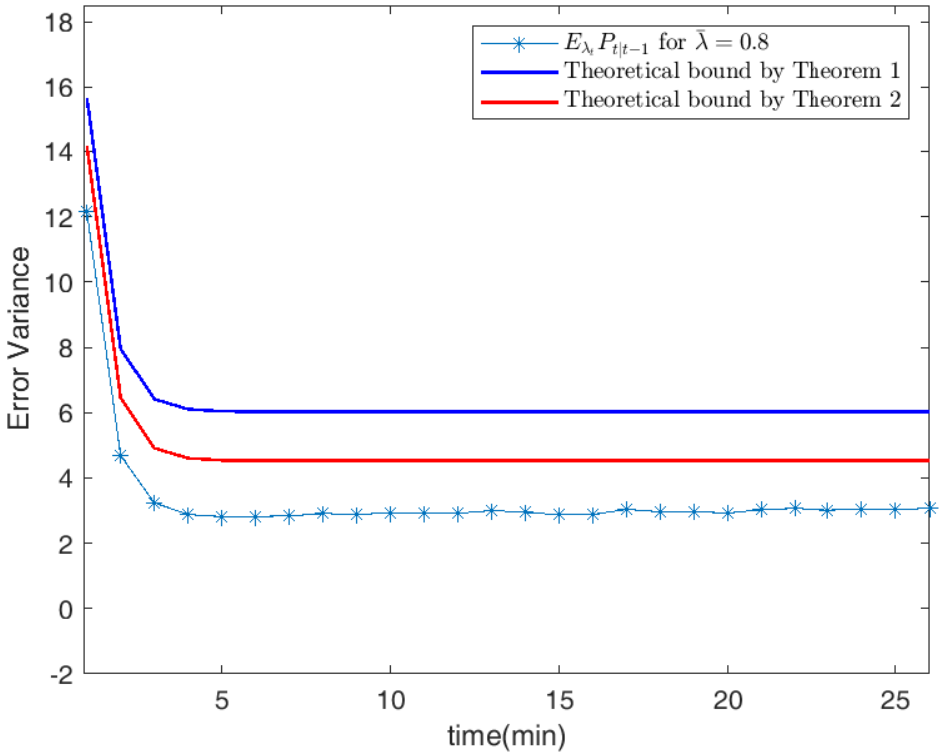

Figure 4.6: $E_{\lambda_t, \mathbf{eQ}} \mathbf{P}_{4,t|t-1}$ *for* $\bar{\lambda} = 0.8$*'and two theoretical bounds of* $\hat{P}_4$ *suggested by Theorem 1 and 2.*

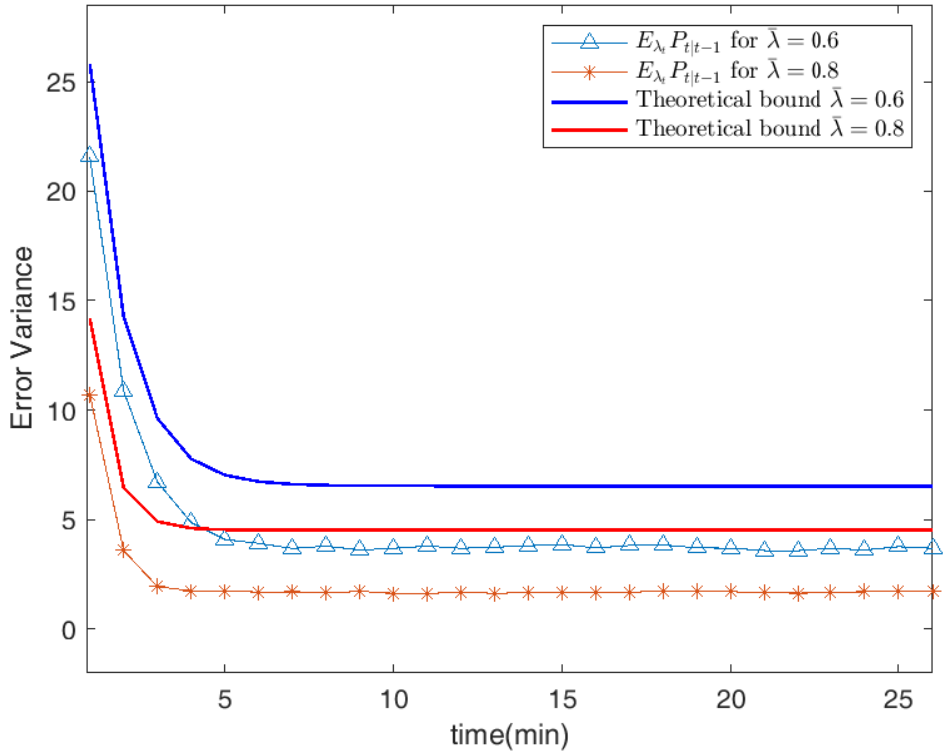

Figure 4.7: *$E_{\lambda_t}\mathbf{P}_{4,t|t-1}$ for two different measurement arrival rates and their theoretical bounds of $\hat{\mathbf{P}}_4$.*

results (bounds) and confirm that the proposed approach is robust in the presence of lossy measurements. However, since the support set remains unaltered, the bound could be tightened as suggested by Theorem 2. Here, we design the measurement matrix to ensure that the necessary conditions for the special case 4.4.2 holds. In this case, the steady state bound for the 4^{th} element is 4.13 (i.e. $(\mathbf{F})_{4,4} = 4.13$). Figure 4.7 demonstrates the effect of measurement arrival rate on $E_{\lambda_t}\mathbf{P}_{4,t|t-1}$ and its theoretical bound for a CMR of 50%. Once again, the variance of error increases with more packet losses (smaller measurement arrival rate).

Figure 4.8 evaluates behavior of the expected variance $E_{\lambda_t}\mathbf{P}_{4,t|t-1}$ with two different CMRs. It is evident that the two expected variances and their bounds are very close to each other. Therefore, we can infer that performance of the estimator with 50% of measurements is very similar to the case of $CMR = 75\%$.

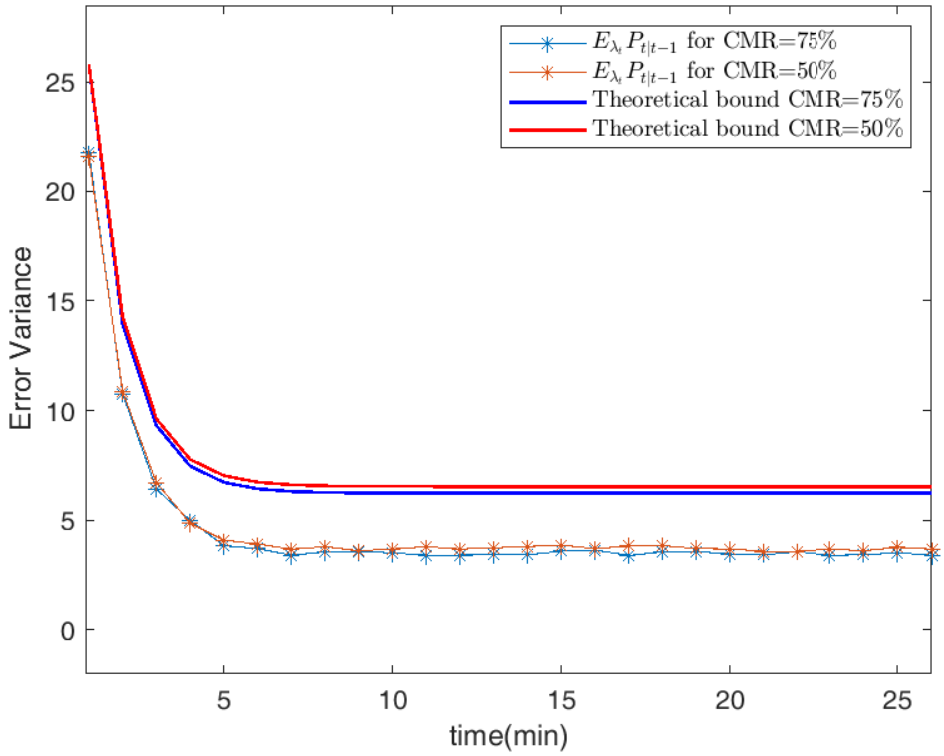

Figure 4.8: $E_{\lambda_t}\mathbf{P}_{4,t|t-1}$ *for two different CMR and their theoretical bound of* $\hat{\mathbf{P}}_4$.

4.6 Summary

In this chapter, we studied dynamic sparse signal recovery in the presence of lossy measurements. We first modified an existing approach for Kalman filtered compressive sensing (KF-CS) to obtain both sparse signal estimate and support set estimate while considering measurement losses. Then, we quantified the error dynamics in both sparse signal estimation and support set estimation. We provided an upper bound for the expected covariance of the estimation error for a given rate of information loss using nonlinear input-to-state stability analysis. The upper bound was updated for two special cases: (1) if the support set remained constant over time, and (2) in addition to the constant support set, a reduced model was observable. Consequently, the critical value for loss in measurements that ensures convergence of error in the KF-CS based algorithm was calculated by the proposed bounds. These upper bound in turn allowed us to evaluate the critical value for loss in measurements that ensured a desirable bound for error in Algorithm 4. Simulations validated both theo-

retical results and highlighted the efficiency of the proposed recursive estimation of a sparse system with lossy measurements. Similar to packet losses, the underlying communication infrastructures are susceptible to delayed measurement due to existing obstacles in the communication networks. Next chapter is dedicated to the study of dynamic compressive sensing in the presence of delayed measurements.

Chapter 5

Dynamic CS with delayed measurements

As discussed in chapter 4, the underlying communication network may not support the transmission of sensor measurements in a regular timely manner. To understand the effect of network limitations on CS based signal recovery, we analyzed the behavior of a sparse recovery algorithm in the presence of lossy measurements in chapter 4. Another issue resulting from network congestion relates to measurements arriving with different random delays. In this chapter, an approach is proposed for sparse recovery of the signals with random delayed measurements.

5.1 Related Work and Contributions

For many decades, state estimation in the presence of delayed measurements has attracted a lot of attention. In fact, all estimation and control processes that aggregate measurements via a communication network may encounter delayed measurements. If the delays are precisely known, the estimation process becomes trivial. However, in most of the cases, the delays are unknown and random leading to poor estimation performance. Most of the existing methods that address this issue attempt to modify the basic state estimation techniques in

a way that the delayed measurements are included in the estimation process. One of the initial efforts in this area, presents a recursive linear filtering algorithm that modifies the conventional model to be compatible with randomly delayed sensor data [79]. For discrete-time stochastic parameter systems, [80] presents a reduced-order linear unbiased estimator. Specifically, the authors propose a criterion that tunes the estimator gains to obtain a certain estimation error covariance. Exponential stability of a linear discrete-time partially observed system perturbed by white noises is analyzed in [81]. [82] designs a robust estimator to deal with linear uncertain discrete-time stochastic systems with randomly varying sensor delay. [31] provides an upper bound for the mean of error in a Kalman filter based estimate based on the statistical properties of the measurement delays. Furthermore, [83] designs a robust Kalman filter for the systems with random delays in both measurements and state dynamics. The mentioned techniques assume that all the sensor data have the same delay properties, i.e., the delays vary only between different time instances. However, in most of the practical situations, the measurements corresponding to a certain sensor can arrive with different delays. To understand state estimation with this type of measurement delays, [84] formulates a Kalman filter for discrete-time systems with instantaneous and l-time delayed measurements by using re-organized innovation analysis. Using covariance information, [85] designs an innovation-based approach to estimate the signals in a system with unknown state space model and randomly delayed measurements. While the impact of such network induced measurement delay on Kalman filter based state estimation has been extensively investigated and well addressed , the corresponding effect on dynamic CS is still unknown. This chapter aims to bridge this gap. Within this framework, the main contributions of this chapter are summarized below:

- We propose a modified compressive sensing technique that incorporates the delayed measurements in the estimation process. To this end, we employ a well-known model for temporal correlation among states that enables us to project the delayed measurements to the current time.

- The dynamic model of the sparse states may not accurately capture the exact temporal

correlation due to stochastic behavior of the signal as well as error in identification of the model parameters. Therefore, the uncertainty in the dynamic model affects the performance of the proposed signal recovery approach. To complete the analysis, we quantify the error in the recovered signal based on the errors in the model for temporal correlation.

The rest of this chapter is organized as follows. Section 5.2 describes the system model of interest. In section 5.3, we introduce the compressive sensing formulation for sparse recovery of a signal sequence using delayed measurements. Section 5.4 quantifies the effect of mismatch in the dynamic model on the estimated signal. Section 5.5 provides numerical simulations that demonstrate the performance of the proposed method.

5.2 System Model

Due to communication network impairments and geographical characteristics of sensors, the sensor data arriving at a fusion center may be delayed. Obviously, if the delays are correctly identified, the estimation process at the fusion center is trivial and can be easily carried out by an augmented system. Another estimation method assumes that all the sensors have the same delay value, at time t. delay. However, in practice, different sensors could be impacted by different delays. Here, we assume that all the sensors are time-stamped, i.e., our fusion center processor knows that a measurement corresponds to a specific time. For example, assume that the sensors synchronously generate their measurements, one every minute. In this chapter, we assume that the sensor data at time t arrives with different random delays. Figure 5.1 shows the delayed measurement sequence. For example, assume the sensors sample and transmit their measurement at time instant t. According to this model, the measurements at time t arrive in a period of time between t and $t + t_d$. Here, t_d denotes the maximum delay. Obviously, the maximum delay t_d is unknown and can take different values at different time instances. To better understand this phenomenon, assume that the sensor data (measurements) at time t constitute the vector $\mathcal{Y}(t) \in \mathbb{R}^n$, where $\mathcal{Y}(t) =$

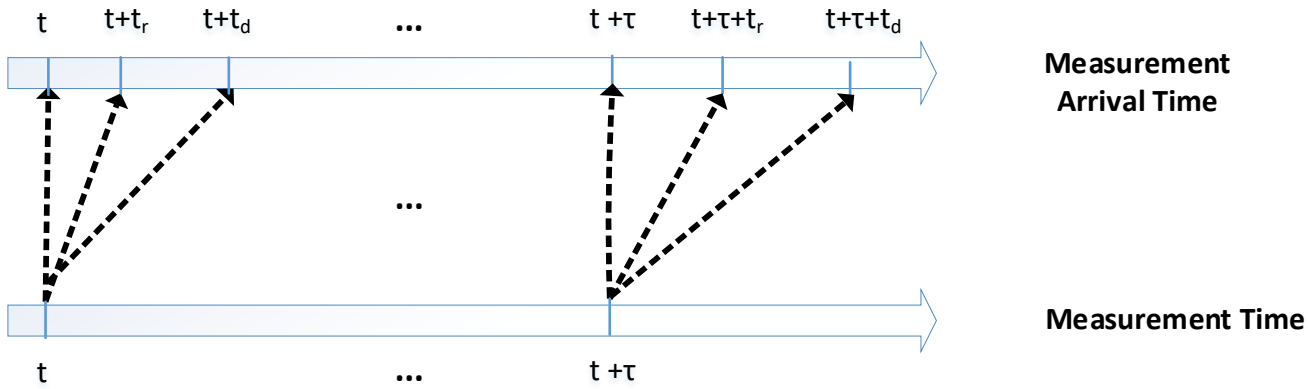

Figure 5.1: *Delayed measurements sequence*

$\begin{bmatrix} y_1(t) & y_2(t) & \dots & y_i(t) & \dots & y_j(t) & \dots & y_k(t) & \dots & y_l(t) & \dots & y_n(t) \end{bmatrix}'$. According to our model, $\begin{bmatrix} y_1(t) & y_2(t) & \dots & y_i(t) \end{bmatrix}'$ arrives at time t; $\begin{bmatrix} y_{j+1}(t) & \dots & y_k(t) \end{bmatrix}'$ arrives at time $t+t_r$; and $\begin{bmatrix} y_{l+1}(t) & \dots & y_n(t) \end{bmatrix}'$ arrives at time $t+t_d$. The rest of data in the measurement vector $\mathcal{Y}(t)$ arrive at a time between t and $t+t_d$.

Let $\mathbf{x}(t) \in \mathbb{R}^n$ denotes the sparse signal/state at time t. Our ultimate goal is to estimate the states $\mathbf{x}(t)$ using the available measurements. To this end, we arrange the available measurement vectors at time t as,

$$\mathbf{y}_a(t) = \mathbf{C}_a(t)\mathbf{x}(t_a) + \boldsymbol{\nu}_a(t) \quad a = 0, 1, ..., l \tag{5.1}$$

where $t_a = t - \kappa_a$, and κ_a is the delay of a^{th} set of the sensors. If a set of sensors does not have any delay (i.e.,$\kappa_0 = 0$, the measurement vector $\mathbf{y}_0(t)$ contains the information about the state at time t; otherwise for $a > 0$ the measurement vector $\mathbf{y}_a(t)$ is a delayed measurement that carries the information of time $t - \kappa_a$. The measurement vectors have different dimensions, $\mathbf{y}_a(t) \in \mathbb{R}^{m_a(t)}$. Here, $m_a(t)$ denotes the dimension of $\mathbf{y}_a(t)$ that can be different at each time due to random nature of the delays ($\sum_{a=0}^{l} m_a(t) = m \ , \ \forall t$). Here, $\boldsymbol{\nu}_a(t) \in \mathbb{R}^{m_a(t)}$ represents the measurement noise vectors. $\mathbf{C}_a(t) \in \mathbb{R}^{m_a(t) \times n}$ is the measurement matrix at time $t - \kappa_a$,

that maps $\mathbf{x}(t_a)$ to $\mathbf{y}_a(t)$.

For the dynamics of the sparse signal $\mathbf{x}_t$, we consider the model presented in [86] that captures the temporal correlation of the sparse states. Here, it is assumed that all the elements of the state vector are mutually decoupled and each element in the underlying sparse signal is temporally correlated. The temporal correlation is modeled via an independent stationary steady-state Gauss-Markov processes, i.e.,

$$\mathbf{x}(t_a) = \boldsymbol{\zeta}(t) + \Theta(\mathbf{x}(t_a - 1) - \boldsymbol{\zeta}(t)) + \boldsymbol{\omega}_t \tag{5.2}$$

where $\boldsymbol{\zeta}(t) \in \mathbb{R}^n$ is the mean of the process. Unlike the original model in [86], we assume that the mean is a function of time and it shows the temporal trend of the state. $\boldsymbol{\omega}_t \in \mathbb{R}^n$ is an i.i.d. Gaussian noise. The temporal correlation is controlled by parameter Θ. At one extreme, if $\Theta = 1$, the states are fully correlated, (i.e., $\mathbf{x}(t_a) = \mathbf{x}(t_a - 1)$), while at the other extreme, $\Theta = 0$, the states evolve according to an uncorrelated Gaussian random process with mean $\boldsymbol{\zeta}(t)$ [86].

As shown in equation (5.1), some information related to $\mathbf{x}(t - \kappa_a)$ arrives at time t. Therefore, we are interested in the relationship between the states at time t and $t - \kappa_a$. Based on (5.2), we can write this relationship as,

$$\mathbf{x}(t - \kappa_a) = \mathbf{G}_a \mathbf{x}(t) + \mathbf{f}_a(t) + \boldsymbol{\varrho}_a(t) \tag{5.3}$$

where, $\mathbf{G}_a \in \mathbb{R}^{n \times n}$ determines the correlation of the signal vector at time $t - \kappa_a$ with current time t. Since the states are decoupled, $\mathbf{G}_a$ is a diagonal matrix. $\mathbf{f}_a(t) \in \mathbb{R}^{n \times 1}$ captures the deterministic trend of the signal. In addition, $\boldsymbol{\varrho}_r(t)$ is the process noise that determines the uncertainty in the model. The advantage of this model is that we can search for the impact of delayed measurements, since every measurement is a linear combination of the elements in $\mathbf{x}$. However, if parameters of the model are not accurately identified, this model can deteriorates the performance of our proposed signal recovery scheme. To understand the possible errors, section 5.4 quantifies the effect of erroneous model on the estimated states.

5.3 Proposed Approach

All the existing compressive sensing techniques including static or dynamic methods rely on availability of the current measurements. When a delayed measurement arrives at the data processing/fusion center, it is unlikely to be incorporated in the estimation process. This policy could result in ignoring a notable amount of information. Particularly in state estimation based on compressive sensing, losing measurements could drastically degrade performance of the estimation since the number of measurements is already limited to begin with. Therefore, even ignoring a small portion of the concise compressed measurements can impact the functioning of the CS methods. Therefore, it is very critical to utilize the delayed measurements in the sparse recovery process to boost the performance of the CS approaches.

The temporal correlation model suggested by (5.3) implies that the delayed measurements are related to the current states, and therefore, can be included in the ℓ_1 minimization problem. Based on the original BPDN problem (2.14), we aim to minimize $\| \mathbf{y}_a(t) - \mathbf{C}_a(t)\mathbf{x}(t-\kappa_a) \|_2^2$. However, since our objective is real-time state estimation, this term must be projected to current time using (5.3). Accordingly, the term $\| \mathbf{y}_a(t) - \mathbf{G}_a\mathbf{C}_a(t)\mathbf{x}(t) - \mathbf{C}_a(t)\mathbf{f}_a(t) \|_2^2$ would be a part of our objective function for minimization. Definitely, a weighting scheme helps to prioritize the effects of each parts on the objective function. To implement this concept, we assign a coefficient λ_a that controls the enforcement of the ℓ_1 minimization problem on the $\| \mathbf{y}_a(t) - \mathbf{G}_a\mathbf{C}_a(t)\mathbf{x}(t) - \mathbf{C}_a(t)\mathbf{f}_a(t) \|_2^2$.

In addition, the support set is assumed to be slowly changing over time. As it was extensively discussed in chapter 3, the partial support set information could enhance performance of the CS techniques. The original problem setup (2.14) tries to minimize the first norm of the whole sparse vector $\mathbf{b}$. However, when the estimated support set at the previous time is T, the modified CS tries to find the vector $\mathbf{b}$ that is sparsest outside the set T among all vectors that fulfill the measurement constraint., i.e., instead of considering all the members in $||\mathbf{b}||_1$, we aim to minimize $||\mathbf{b}_{T^c}||_1$ in our problem. Therefore, we propose to estimate the states at time t using the following minimization problem which includes the delayed

measurements.

$$\hat{\mathbf{b}}_t = \underset{\mathbf{b}}{\operatorname{argmin}} \; ||\mathbf{b}_{T^c}||_1 + \lambda_0 ||\mathbf{y}_0(t) - \mathbf{C}_0(t)\mathbf{b}||_2^2 + \sum_{a=1}^{l} \lambda_a ||\mathbf{y}_a(t) - \mathbf{G}_a\mathbf{C}_a(t)\mathbf{b} - \mathbf{C}_a(t)\mathbf{f}_a(t)||_2^2. \tag{5.4}$$

where λ_a is a weighting/forgetting factor that controls the effect of delayed measurements. Obviously, the most recent measurements should have a bigger impact on the estimation, while the measurements from the distant past will carry less useful information about the current signal. Accordingly, the λ_a relevant to the current measurements should be selected to be relatively larger than the λ_a corresponding to the delayed measurements.

Algorithm 5 Dynamic compressive sensing with delayed measurements

Step 1: At $t = 1$: Employ a suitable CS method with sufficient measurements to find a precise estimation of the support set $\hat{N}_1$.

For $t > 1$, set:
Step 2: Set $T = \hat{N}_{t-1}$

Step 3: Compute $\hat{\mathbf{x}}_t$ as the solution of

$$\hat{\mathbf{x}}_t = \underset{\mathbf{b}}{\operatorname{argmin}} ||\mathbf{b}_{T^c}||_1 + \lambda_0 ||\mathbf{y}_0(t) - \mathbf{C}_0(t)\mathbf{b}||_2^2 + \sum_{a=1}^{l} \lambda_a ||\mathbf{y}_a(t) - \mathbf{G}_a\mathbf{C}_a(t)\mathbf{b} - \mathbf{C}_a(t)\mathbf{f}_a(t)||_2^2$$

Step 4: Support Estimation - Simple thresholding:

$$\hat{N}_t = \{i : |(\hat{\mathbf{x}}_t)_i| > \alpha_a\}$$

Increment t and go to step 2.

Algorithm 5 summarizes our proposed approach for Dynamic CS with delayed measurements. Since the proposed algorithm is a recursive method, we must ensure that the method knows the exact initial support set. To this end, a sufficient amount of measurement must be provided. Since the support set does not change quickly, we use the previous support set in Step 2 of Algorithm 5. Then, Step 3 provides the estimates of the states by solving equation (5.4). Finally, we employ a thresholding criterion to estimate the current support set $\hat{N}_t$ which will be used in the next time instant.

5.4 Error Analysis

Uncertainty in the temporal correlation model (5.3) leads to increase in error associated with the estimation. In addition to the random term $\boldsymbol{\varrho}_a(t)$, there is an uncertainty with identification of $\mathbf{f}_a(t)$. Therefore, $\mathbf{x}(t-\kappa_a) = \mathbf{G}_a\mathbf{x}(t) + \mathbf{f}_a(t) + \mathbf{e^f}_a(t) + \boldsymbol{\varrho}_a(t)$ must be replaced in the regularization of (5.4). Therefore, if the correlation model is accurate, the optimal estimation is obtained from:

$$\tilde{\mathbf{b}}_t = \underset{\mathbf{b}}{\operatorname{argmin}} \; ||\mathbf{b}_{T^c}||_1 + \lambda_0||\mathbf{y}_0(t) - \mathbf{C}_0(t)\mathbf{b}||_2^2 + \sum_{a=1}^{l} \lambda_a||\mathbf{y}_a(t) - \mathbf{G}_a\mathbf{C}_a(t)\mathbf{b} - \mathbf{C}_a(t)(\mathbf{f}_a(t) + \mathbf{e^f}_a(t) + \boldsymbol{\varrho}_a(t))||_2^2. \tag{5.5}$$

Let $\mathbf{e^b}_t = \hat{\mathbf{b}}_t - \tilde{\mathbf{b}}_t$ denote the error in the estimation. Also, $\mathbf{e}_t^c$ indicates the difference between the correlation models used in (5.4) and (5.5), i.e., $\mathbf{e}_t^c = \mathbf{e^f}_a(t) + \boldsymbol{\varrho}_a(t)$. With respect to this notation, Theorem 9 quantifies the relationship between $\mathbf{e^b}_t$ and $\mathbf{e}_t^c$.

Theorem 9. *If* $(\mathscr{J} + \sum_{a=1}^{l} \mathscr{K}_a)$ *is invertible, the error in the estimated signal due to correlation model mismatch corresponds to:*

$$\mathbf{e^b}_t = \big[(\mathscr{J} + \sum_{a=1}^{l} \mathscr{K}_a)^{-1} \sum_{a=1}^{l} \mathscr{L}_a\big]\mathbf{e}_t^c + \big[\mathscr{J} + \sum_{a=1}^{l} \mathscr{K}_a\big]^{-1}\boldsymbol{\Upsilon}_t. \tag{5.6}$$

where, $\boldsymbol{\Upsilon}_t$ *is an* $n \times 1$ *vector such that if* $i \in T_t^c$*, then* $(\boldsymbol{\Upsilon}_t)_i = sgn(\hat{\mathbf{b}}_i) - sgn(\tilde{\mathbf{b}}_i)$*; otherwise* $(\boldsymbol{\Upsilon}_t)_i = 0$*. The other parameters are defined as follows:*

$$\mathscr{J} = 2\lambda_0 \begin{bmatrix} \sum_{j=1}^{m_0} \mathbf{C}_0(j,1)\mathbf{C}_0(j,:) \\ \vdots \\ \sum_{j=1}^{m_0} \mathbf{C}_0(j,i)\mathbf{C}_0(j,:) \\ \vdots \\ \sum_{j=1}^{m_0} \mathbf{C}_0(j,n)\mathbf{C}_0(j,:) \end{bmatrix}, \; \mathscr{K}_a = 2\lambda_a \begin{bmatrix} \sum_{j=1}^{m_0} \mathbf{H}_a(j,1)\mathbf{H}_a(j,:) \\ \vdots \\ \sum_{j=1}^{m_0} \mathbf{H}_a(j,i)\mathbf{H}_a(j,:) \\ \vdots \\ \sum_{j=1}^{m_0} \mathbf{H}_a(j,n)\mathbf{H}_a(j,:) \end{bmatrix},$$

$$\mathscr{L} = 2\lambda_a \begin{bmatrix} \sum_{j=1}^{m_0} \mathbf{H}_0(j,1)\mathbf{C}_0(j,:) \\ \vdots \\ \sum_{j=1}^{m_0} \mathbf{H}_0(j,i)\mathbf{C}_0(j,:) \\ \vdots \\ \sum_{j=1}^{m_0} \mathbf{H}_0(j,n)\mathbf{C}_0(j,:) \end{bmatrix}.$$

Here, $\mathbf{C}_a(i,:)$ refers to i^{th} row of $\mathbf{C}_a$; and $\mathbf{C}_a(i,j)$ indicates the elements on the i^{th} row and j^{th} column of $\mathbf{C}_a$.

Proof. Since the objective function (5.4) is an unconstrained convex optimization problem, the optimal value can be extracted by setting $\nabla\mathscr{F}(\mathbf{b}) = 0$, where $\mathscr{F}(\mathbf{b}) : \mathbb{R}^n \to \mathbb{R}$ is:

$$\mathscr{F}(\mathbf{b}) = ||\mathbf{b}_{T^c}||_1 + \lambda_0||\mathbf{y}_0(t) - \mathbf{C}_0(t)\mathbf{b}||_2^2 + \sum_{a=1}^{l} \lambda_a||\mathbf{y}_a(t) - \mathbf{G}_a\mathbf{C}_a(t)\mathbf{b} - \mathbf{C}_a(t)\mathbf{f}_a(t)||_2^2 \quad (5.7)$$

For simplification, we ignore the time index t for the rest of this proof. Based on this notation, we rewrite (5.7),

$$\mathscr{F}(\mathbf{b}) = ||\mathbf{b}_{T^c}||_1 + \lambda_0||\mathbf{y}_0 - \mathbf{C}_0\mathbf{b}||_2^2 + \sum_{a=1}^{l} \lambda_a||\mathbf{y}_a - \mathbf{H}_a\mathbf{b} - \mathbf{C}_a\mathbf{f}_a||_2^2 =$$
$$||\mathbf{b}_{T^c}||_1 + \lambda_0 \sum_{j=1}^{m_0} \big(\mathbf{y}_{0,j} - \mathbf{C}_0(j,:)\mathbf{b}\big)^2 + \sum_{a=1}^{l} \lambda_a \sum_{j=1}^{m_a} \big(\mathbf{y}_{a,j} - \mathbf{H}_a(j,:)\mathbf{b} - \mathbf{C}_a(j,:)\mathbf{f}\big)^2.$$

Here, gradient of the $\mathscr{F}(\mathbf{b})$ is equal to : $\nabla\mathscr{F}(\mathbf{b}) = \begin{bmatrix} \frac{\partial\mathscr{F}}{\partial\mathbf{b}_1} & \dots & \frac{\partial\mathscr{F}}{\partial\mathbf{b}_i} & \dots & \frac{\partial\mathscr{F}}{\partial\mathbf{b}_n} \end{bmatrix}'$. To find the optimal solution, each element in the gradient should be set to zero. Therefore,

$$\frac{\partial\mathscr{F}}{\partial\mathbf{b}_i}|_{\hat{\mathbf{b}}} = \frac{\mathbf{b}_i}{|\mathbf{b}_i|} + \lambda_0 \sum_{j=1}^{m_0} -2\mathbf{C}_0(j,i)\big(\mathbf{y}_{0,j} - \mathbf{C}_0(j,:)\mathbf{b}\big) +$$
$$\sum_{a=1}^{l} \lambda_a \sum_{j=1}^{m_a} -2\mathbf{H}_0(j,i)\big(\mathbf{y}_{a,j} - \mathbf{H}_a(j,:)\mathbf{b} - \mathbf{C}_a(j,:)\mathbf{f}\big) = 0 \quad (5.8)$$

Now, consider the actual case where we know the precise correlation model. The objective function relevant to this situation is denoted by $\hat{\mathscr{F}}(\mathbf{b}) : \mathbb{R}^n \to \mathbb{R}$ and corresponds to,

$$\hat{\mathscr{F}}(\mathbf{b}) = ||\mathbf{b}_{T^c}||_1 + \lambda_0 \sum_{j=1}^{m_0} \big(\mathbf{y}_{0,j} - \mathbf{C}_0(j,:)\mathbf{b}\big)^2 + \sum_{a=1}^{l} \lambda_a \sum_{j=1}^{m_a} \big(\mathbf{y}_{a,j} - \mathbf{H}_a(j,:)\mathbf{b} - \mathbf{C}_a(j,:)\mathbf{f} - \mathbf{C}_a(j,:)\mathbf{e}^c\big)^2.$$

The optimal solution of this objective function is denoted by $\tilde{\mathbf{b}}$ and is computed by,

$$\frac{\partial \mathscr{F}}{\partial \mathbf{b}_i}|_{\tilde{\mathbf{b}}} = \frac{\mathbf{b}_i}{|\mathbf{b}_i|} + \lambda_0 \sum_{j=1}^{m_0} -2\mathbf{C}_0(j,i)\big(\mathbf{y}_{0,j} - \mathbf{C}_0(j,:)\mathbf{b}\big) +$$
$$\sum_{a=1}^{l} \lambda_a \sum_{j=1}^{m_a} -2\mathbf{H}_0(j,i)\big(\mathbf{y}_{a,j} - \mathbf{H}_a(j,:)\mathbf{b} - \mathbf{C}_a(j,:)\mathbf{f} - \mathbf{C}_a(j,:)\mathbf{e}^c\big) = 0 \tag{5.9}$$

Using equations (5.8) and (5.9), we have:

$$\frac{\tilde{\mathbf{b}}_i}{|\tilde{\mathbf{b}}_i|} - \frac{\hat{\mathbf{b}}_i}{|\hat{\mathbf{b}}_i|} + \lambda_0 \sum_{j=1}^{m_0} -2\mathbf{C}_0(j,i)\mathbf{C}_0(j,:)\big(\tilde{\mathbf{b}}_i - \hat{\mathbf{b}}_i\big) +$$
$$\sum_{a=1}^{l} \big(\lambda_a \sum_{j=1}^{m_a} -2\mathbf{H}_0(j,i)\mathbf{H}_a(j,:)\big(\tilde{\mathbf{b}}_i - \hat{\mathbf{b}}_i\big)\big) - \sum_{a=1}^{l} \big(\lambda_a \sum_{j=1}^{m_a} 2\mathbf{H}_0(j,i)\mathbf{C}_a(j,:)\mathbf{e}^c\big) = 0 \tag{5.10}$$

Using $\mathbf{e}_i^{\mathbf{b}} = \tilde{\mathbf{b}}_i - \hat{\mathbf{b}}_i$ and the defined matrices, we can rewrite (5.10) as,

$$\boldsymbol{\Upsilon} - \mathscr{J}\,\mathbf{e}^{\mathbf{b}} + \sum_{a=1}^{l} \mathscr{K}_a \mathbf{e}^{\mathbf{b}} + \sum_{a=1}^{l} \mathscr{L}_a \mathbf{e}^c = 0 \tag{5.11}$$

Thus, if $(\mathscr{J} + \sum_{a=1}^{l} \mathscr{K}_a)$ is invertible, ${\mathbf{e}^{\mathbf{b}}}_t$ corresponds to:
${\mathbf{e}^{\mathbf{b}}}_t = \big[(\mathscr{J} + \sum_{a=1}^{l} \mathscr{K}_a)^{-1} \sum_{a=1}^{l} \mathscr{L}_a\big]\mathbf{e}_t^c + \big[\mathscr{J} + \sum_{a=1}^{l} \mathscr{K}_a\big]^{-1} \boldsymbol{\Upsilon}_t.$ □

5.5 Simulation Results

To evaluate our approach, we simulated a time sequence of sparse $n = 70$ length signal $\mathbf{x}_t$, with maximum sparsity $S_{max} = 10$. Overall, the length of time sequence is 30 instances. We design the $\mathbf{C}$ matrix generating $n \times m$ i.i.d. Gaussian entries (with $n = 28$) and normalizing

each column of the resulting matrix. This means that the CMR is 40% in a normal situation with no delays. We assume that $\boldsymbol{\varrho}_a(t)$ is a diagonal matrix with σ^2_{sys} on its diagonal. The observation noise variance is $\sigma^2_{obs} = 0.24$. For modeling $\mathbf{x}_t$, we consider the diagonal elements of $\mathbf{G}$ to be between 0.6 and 1, which implies a strong correlation between the signal samples. $\mathbf{f}_i = diag(10 + \tilde{\alpha} sin(\frac{\pi}{16}t))$, where $\tilde{\alpha} \in \big[1, 2\big]$.

We design the simulation in a way that 50% of the compressed measurements are randomly delayed, after time instant $t = 4$. In fact, if these delayed measurements are not utilized in the estimation process, it is equal to the situation where the $CMR = 20\%$. We then, investigate the performance of Algorithm 5 by performing 1000 Monte Carlo simulations. Figure 5.2 shows the affect of delayed measurements on the performance of CS methods. The blue curve shows the INAE value when there are no delayed measurements. This case corresponds to the best performance due to complete measurement availability. In this situation, we simply use the classic CS. Now, if we still employ classic CS but only 50% of the measurements are available on-time, we can not use the other 50% of data (delayed measurements). As a consequence, the red curve shows a degradation in performance of the signal recovery. Finally, the yellow curve demonstrates our proposed method (Algorithm 5) in presence of delay where a significant improvement is achieved by incorporating the delayed measurements and the correlation model.

5.6 Summary

In this chapter, we studied impact of measurement arriving at the fusion center with different random delays. First, We proposed a modified compressive sensing technique that incorporates the delayed measurements in the sparse recovery process. To this end, we employed an existing model that captures the temporal correlation of the sparse signals. Consequently, we could utilize the information carried by the delayed measurements for estimation at the current time. In addition to the stochastic nature of sparse signal evolution over time, the parameters of the underlying dynamic model may not be precisely identified. Therefore, we quantified the error in the recovered signal based on the errors in the modeling of the under-

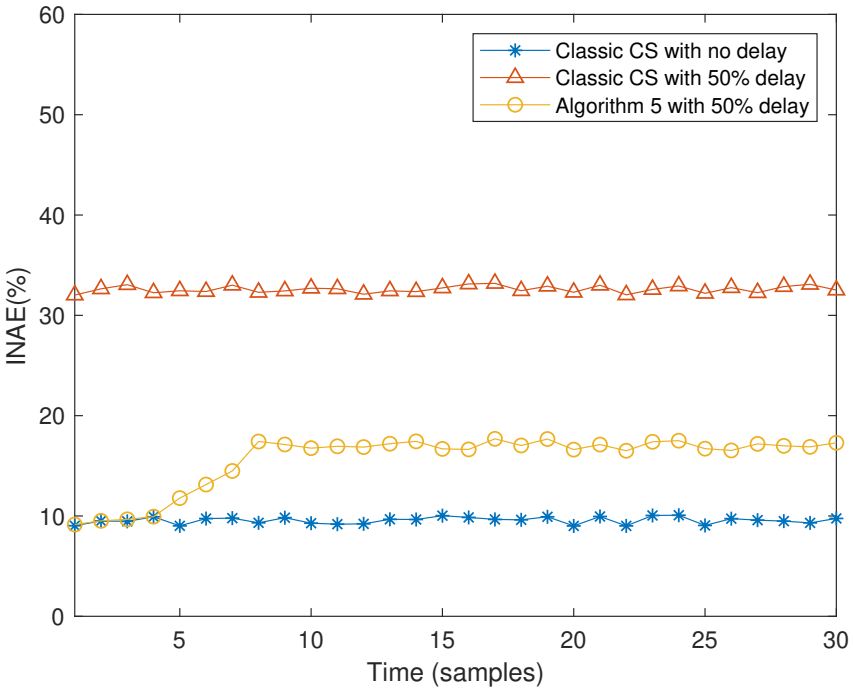

Figure 5.2: *Signal recovery with delayed measurements*

lying dynamic correlation. Applicability of our proposed method was verified by simulation results. In the next chapter, we address the problem of joint state estimation and topology identification in distribution systems.

Chapter 6

Joint Topology Identification and State Estimation

Distribution system topology identification has become a popular research topic due to its critical role in state estimation, Volt-VAR control, fault detection, demand response, etc. [87]. Many of the proposed solutions to challenges associated with situational awareness, control, and optimization in distribution systems are not feasible if the underlying distribution grid is not accurately identified [88]. Furthermore, the time-varying nature of distribution systems implies that the topology identification process must be frequently performed if reliable results are expected from distribution systems operations. However, a timely topology monitoring may not be completely possible since the required measurements for the topology identification tasks are either unavailable or unreliable [89]. Therefore, this chapter tries to address the joint topology identification problem along with state estimation for highly unobservable distribution systems.

6.1 Related Work and Contributions

The existing methods for topology identification can be classified into different categories. Traditional approaches employ all the available measurements for state estimation based

on every possible topology configuration. Then, the topology that results in the minimum residual state estimation error is selected [33]. A notable drawback of these techniques is the high computational complexity that grows with the possible topologies. Instead of considering all possible topologies, other techniques focus on the simultaneous estimation of switch statuses and system states [34], [90] and [91]. Although, these methods could alleviate the burden of computational complexity, they require a large number of measurements. For a low-voltage distribution network, [92] proposes a data-driven approach to identify the underlying network topology using time series of energy measurements. In [93], a mixed integer quadratic programming (MIQP) based topology identification model is proposed, that is suitable for radially operated distribution grids. To detect network topology, [94] presents a time-series signature verification approach which is based on the analysis of synchronized voltage phasor data. While most of the topology identification techniques rely on voltage and power information, [88] provides a new topology identification algorithm that is based on measurements from a few line current sensors, together with available pseudo-measurements. To continuously monitor the underlying topology of a smart distribution network, real-time topology identification has been targeted in some of recent works [89], [95].

As comprehensively explained in chapter 2, the state estimation processes based on compressive sensing employ power flow equations that rely on the identified topology. Since the compressive sensing techniques are typically implemented for undersampled systems, it is likely that the required topology knowledge is impaired due to lack of observability. Consequently, the compressive sensing based state estimation fails to provide the correct states since the underlying topology is not accurately identified. On the other hand, the topology identification methods are not functional due to the lack of accurate state information. Therefore, our goal is to simultaneously estimate the states and the network topology with limited available measurements. Distribution system operators usually have access to information about physical distances among the buses, lines connecting buses and their parameters. Therefore, the topology relevant to the fixed lines is always known. However, in addition to the fixed lines that are always energized, there are lines with switches that establish a connection between the buses based on their statuses (open or close). Therefore,

topology identification can be mapped to the problem of switch status estimation. Within this framework, the main contributions of this chapter are summarized below:

- For concurrent estimation of states and topology, we first formulate an optimization problem with states of systems and switch statuses as decision variables. Our framework projects the switch statuses on the underlying admittance matrix. The proposed formulation contains nonlinear relationships and integer variables leading to a mixed integer nonlinear programming (MINLP) problem.

- In order to guarantee the solvability of the optimization formulations and reduce MINLP complexity, we modify the nonlinear constraints to yield a mixed integer linear programming (MILP) problem. To this end, we transform the existing nonlinearity by introducing auxiliary variables.

- Although we could fix the nonlinearity issue by introducing MILP, the proposed technique is still a nonconvex problem. In other words, there is no guarantee to reach the optimal solution. To cope with this issue, we relax the nonlinearity of MINLP by replacing integer variables with continuous decision variables. Then, we apply an alternative minimization approach that improves both state estimation and topology identification.

- Finally, we introduce a hybrid dynamic framework that incorporates prior information about system topology for current state/switch status estimation. More precisely, a Markov jump model is proposed for switch status that helps topology identification especially for high levels of compression. Besides, we employ the previous information about the support set that improves the fidelity of our method.

- Simulation results provided in section 6.5 demonstrate the performance of the proposed approaches on the IEEE 37-node test feeder with practical load/generation data.

The rest of this chapter is organized as follows. Section 6.2 describes the system model of interest and linearized power flow formulation. In section 6.3, we introduce the compressive

sensing formulation for joint topology identification and state estimation. Then, we propose two methods to alleviate the complexities of the original MINLP problem. Section 6.4 presents dynamical joint state estimation and topology identification in distributions systems. Section 6.5 provides numerical simulations that demonstrate the performance of the introduced methods.

6.2 System Model

In distribution grids, the conventional state estimation methods or topology identification techniques need to aggregate all the available sensor/smart meter measurements. This requirement implies that a large volume of measurements must be gathered and transmitted over a communication network in an actual distribution system. To this end, a communication network with large bandwidth and high reliability should be established. However, providing such a communication infrastructure is not feasible or affordable in most of the situations. Furthermore, the grid infrastructure may not be equipped with all of these reliable sensors. Therefore, our goal is to alleviate the burden on the communication network by reducing the required number of measurements for topology identification and state estimation while maintaining a reasonable fidelity in estimating both. As suggested by the main theme of this dissertation, we employ compressive sensing theory that enables us to recover a signal from a smaller number of random measurements [96].

Distribution system state estimation techniques with high accuracy rely on accurate power flow methods which in turn require precise knowledge about the system, including the system topology, parameters, power injection data, etc.. As a reminder, our direct compressive sensing based state estimation method is based on the following optimization problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \parallel \mathbf{x} \parallel_1 + \lambda \parallel \mathbf{y} - \mathbf{\Phi}\mathbf{\Psi}\mathbf{x} \parallel_2$$

$$\text{subject to :}$$

$$\mathbf{v} = f_{PF}(\mathbf{P}, \mathbf{Q})$$

where, f_{PF} is the nonlinear power flow function that captures the relationship between voltage states and power injected at the nodes. Nonlinearity of the AC power-flow equations poses significant challenges for the development of computationally affordable state estimation tasks. To avoid such a complexity, we employ a linearized approximation model that is applicable to multiphase distribution networks [45]. For a three-phase distribution system with $\mathcal{J}$ buses, $\mathbf{Y}$ denotes the admittance matrix: $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{00} & \mathbf{Y}_{0L} \\ \mathbf{Y}_{L0} & \mathbf{Y}_{LL} \end{bmatrix}$ where, $\mathbf{Y}_{00} \in \mathbb{C}^{3\times3}$, $\mathbf{Y}_{L0} \in \mathbb{C}^{3(\mathcal{J}-1)\times3}$, $\mathbf{Y}_{0L} \in \mathbb{C}^{3\times3(\mathcal{J}-1)}$ and $\mathbf{Y}_{LL} \in \mathbb{C}^{3(\mathcal{J}-1)\times3(\mathcal{J}-1)}$ are submatrices of the admittance matrix. Denoted by $\tilde{\mathbf{v}}$, the linearized voltage is:

$$\mathbf{v} \approx \mathbf{M} \begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix} + \mathbf{w} \tag{6.1}$$

where,

$$\mathbf{M} = \begin{bmatrix} \mathbf{Y}_{LL}^{-1} diag(\tilde{\hat{\mathbf{v}}})^{-1} & -j\mathbf{Y}_{LL}^{-1} diag(\tilde{\hat{\mathbf{v}}})^{-1} \end{bmatrix} \tag{6.2}$$

$\mathbf{w} = -\mathbf{Y}_{LL}^{-1}\mathbf{Y}_{L0}\mathbf{v}_0$. Here, $\mathbf{v}_0$ is the slack bus voltage where the voltage level is assumed to be fixed or regulated (typically 1 per unit). $\tilde{\hat{\mathbf{v}}}$ denotes voltage in an operating point. One may consider $\mathbf{w} = \tilde{\hat{\mathbf{v}}}$.

If the topology is correctly identified along with accurate data and parameters, one can implement different power flow methods. Most of the existing distribution systems are equipped with switches and breakers that result in various topology configurations based on the status of the switches/breakers. Figure. 6.1 demonstrates IEEE 37-node test feeder as an example. Here, the system is equipped with six switches. Node 1 serves as slack bus and the source of power. Let us consider node 27. In the normal situation, s_1 is closed

and s_2 is open, therefore, node 27 is energized from node 3. Another possibility is when switch s_1 is open and s_2 is closed. This implies that node 27 is fed by node 8. Also, both s_1 and s_2 can be closed or open, simultaneously. Consequently, different voltage profiles will result based on the flow of power in the network. Impact of such a switching configuration is captured in the underlying admittance matrix. In other words, if we are able to identify the correct admittance matrix, we can determine the exact topology of the system. However, the number of possible topologies dramatically increases with the number of switches, i.e., if there are d switches in the network, 2^d different topology configurations are possible. Such large numbers of different possibilities impose a big load on the required computations; or they can lead to inaccurate and unreliable results. Therefore, estimation of switch statuses is the only reasonable approach for topology identification [90], [88]. To this end, we write the admittance matrix as a function of the status of the switches as

$$\mathbf{Y}_{LL} = \bar{\mathbf{Y}} + s_1\mathbf{Y}_1 + s_2\mathbf{Y}_2 + ... + s_d\mathbf{Y}_d, \tag{6.3}$$

where $bar\mathbf{Y}$ is the admittance matrix between the nodes with fixed permanent connection (i.e., the lines with no switch/breaker). $s_i \in \{0, 1\}$ denote the status of the i^{th} switch. If the switch s_1 is closed, then the corresponding line is activated. Mathematically, this implies that $s_1 = 1$. Therefore, the submatrix $\mathbf{Y}_1$ (the admittance matrix between the two nodes that are connected by s_1) is added to the general admittance matrix $\mathbf{Y}_{LL}$; otherwise, $\mathbf{Y}_{LL}$ does not contain submatrix $\mathbf{Y}_1$.

In this chapter, we are interested in estimating the state vector $\mathbf{z} = \begin{bmatrix}\mathbf{P}^T & \mathbf{Q}^T & \mathbf{v}^T\end{bmatrix}^T$, where $\mathbf{P} \in \mathbb{R}^{3n}$, $\mathbf{Q} \in \mathbb{R}^{3n}$ and $\mathbf{v} \in \mathbb{C}^{3n}$ are three phase real power, reactive power and voltage phasor, respectively. Here, we create vector $\mathbf{S} = \begin{bmatrix}\mathbf{P}^T & \mathbf{Q}^T\end{bmatrix}^T$ containing the real and reactive power values. The measurement mechanism is such that:

$$\begin{aligned}\mathbf{y}^{\mathbf{S}} &= \mathbf{\Phi}^{\mathbf{S}}\mathbf{S} + \boldsymbol{\nu}_{\mathbf{S}} \\ \mathbf{y}^{\mathbf{V}} &= \mathbf{\Phi}^{\mathbf{V}}\mathbf{V} + \boldsymbol{\nu}_{\mathbf{V}}\end{aligned} \tag{6.4}$$

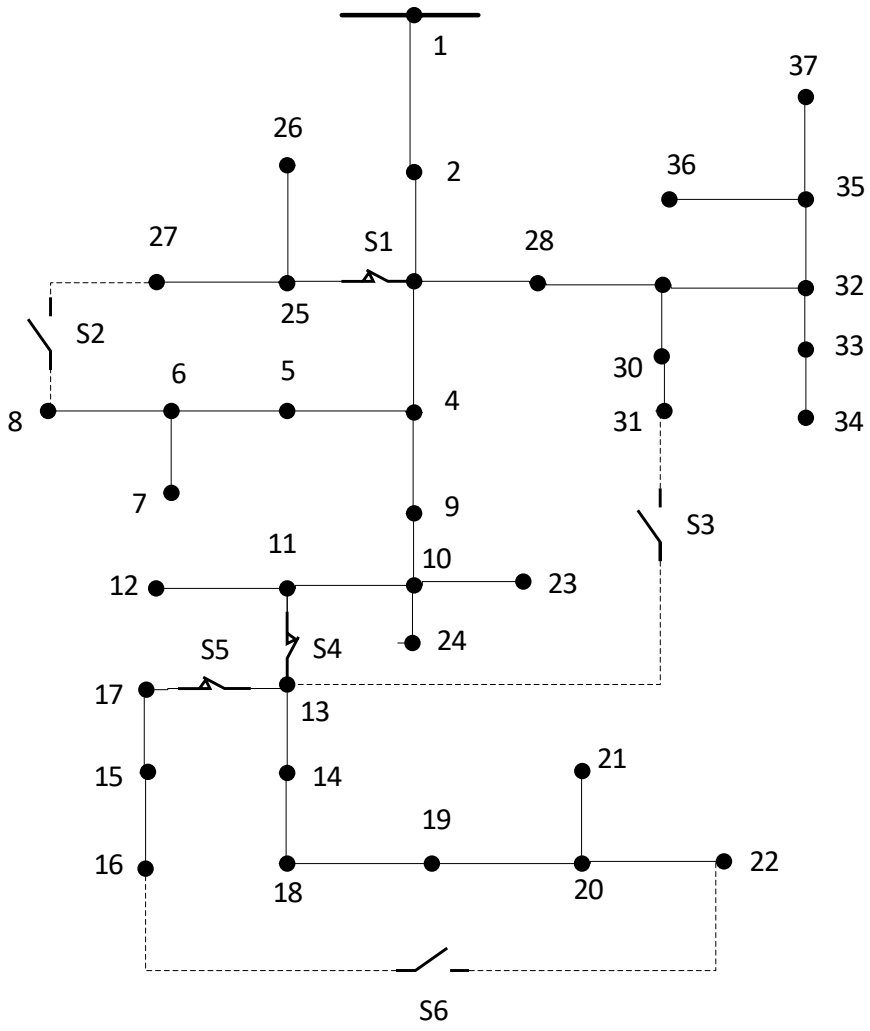

Figure 6.1: *IEEE 37-node test feeder with six switches*

where, $\mathbf{y^S} \in \mathbb{R}^{m_\mathbf{S}}$ and $\mathbf{y_V} \in \mathbb{C}^{m_\mathbf{V}}$ are the power measurements and voltage measurement, respectively. Here, $\boldsymbol{\nu}_\mathbf{S}$ and $\boldsymbol{\nu}_\mathbf{V}$ are the measurement noise vectors which are zero mean white Gaussian noises. $\mathbf{\Phi^S} \in \mathbb{R}^{m_\mathbf{S} \times 6n}$ and $\mathbf{\Phi^V} \in \mathbb{R}^{m_\mathbf{V} \times 3n}$ are measurement/projection matrices. These matrices are determined by the underlying measurement mechanisms (e.g., matrix elements distributed as i.i.d. Gaussian random variables with mean 0 and variance $1/m_\mathbf{S}$ or Bernoulli random variables).

6.3 Proposed Approach

The necessary condition for implementing compressive sensing approaches is that the signal of interest exhibits sparsity or must be approximately sparse in a linear transformation basis. Let the signal of interest $\mathbf{z} \in \mathbb{R}^{9n}$ be a sparse vector in a linear transformation basis $\mathbf{\Psi} \in \mathbb{R}^{9n \times 9n}$ such that,

$$\mathbf{z} = \mathbf{\Psi x} = \begin{bmatrix} \mathbf{\Psi^S} & \mathbf{0}_{6n \times 3n} \\ \mathbf{0}_{3n \times 6n} & \mathbf{\Psi^V} \end{bmatrix} \mathbf{x}$$

where, $\mathcal{A}$ has at most $S \ll 9n$ significant coefficients i.e., $\mathbf{z}$ is S-sparse in the sparsifying basis $\mathbf{\Psi}$. Previous literature has validated the sparsity assumption for distribution systems [20], [37]. Along with the sparsity condition, the linearized relationship (6.1) between power and voltage serves as the link between the states and the network topology. However, as indicated by equation (6.2), both $\mathbf{M}$ and $\mathbf{w}$ contain inverse of the admittance matrix which complicates the optimization problem since admittance matrix $\mathbf{Y}$ is influenced by the switch status and additional terms. To avoid such a complexity, we multiply equation (6.2) by $\mathbf{Y}_{LL}$. Consequently, equation (6.5) represents the linearized power flow constraint in our optimization problem:

$$\mathbf{Y}_{LL}\mathbf{v} + \mathbf{Y}_{L0}\mathbf{v}_0 - \mathbf{M}_{yl}\mathbf{S} = 0 \tag{6.5}$$

where, $\mathbf{M}_{yl} = \begin{bmatrix} diag(\tilde{\hat{\mathbf{v}}})^{-1} & -jdiag(\tilde{\hat{\mathbf{v}}})^{-1} \end{bmatrix}$.

It should be noted that in many practical situations, opening/closing of a switch may

depend on status of other switches. For example, let us consider the IEEE 37-node test feeder in Figure 6.1. If switch s_1 (normally closed) is opened, nodes 25, 26 and 27 are not energized anymore. Obviously, to cope with this outage, switch s_2 must be closed, where the islanded nodes (25-27) can be connected to the grid and obtain their energy through node 8. To enhance resiliency in distribution systems, most of the utilities equip the distribution grids with line switches [97], [98]. Definitely, in such grids, the switch statuses are not completely independent. Therefore, we introduce a set of constraints that captures the relationship between the switches as follow:

$$\mathcal{G}(s_1, s_2, ..., s_d) = \mathcal{C} \tag{6.6}$$

where, $\mathcal{C}$ is a vector with constant integer variables. For example, if the nodes (25-27) must be always energized and no mesh structure is allowed in the system, the constraint (6.6) is written as, $s_1 + s_2 = 1$.

Using the information about the system, we setup the minimization (6.7) to jointly reconstruct the sparse signal $\mathcal{A}$ and status of the switches s_i,

$$\underset{\mathcal{A}, s_i}{\operatorname{argmin}} \parallel \mathcal{A} \parallel_1 + \lambda^{\mathbf{S}} \parallel \mathbf{y}^{\mathbf{S}} - \mathbf{\Phi}^{\mathbf{S}} \mathbf{\Psi}^{\mathbf{S}} \mathcal{A} \parallel_2 + \lambda^{\mathbf{V}} \parallel \mathbf{y}^{\mathbf{V}} - \mathbf{\Phi}^{\mathbf{V}} \mathbf{\Psi}^{\mathbf{V}} \mathcal{A} \parallel_2 \tag{6.7}$$

$$\text{subject to } (6.5), (6.6)$$

Obviously, to guarantee a certain state estimation fidelity and accurate topology identification, the number of measurements can not be arbitrarily reduced (i.e., if we excessively reduce $m_{\mathbf{S}}$ and $m_{\mathbf{V}}$, both topology identification and state estimation may fail to perform accurately). It should be noted that choosing an appropriate transformation matrix $\mathbf{\Psi}$ plays an important role in the performance of compressive sensing problem. [35] and [99] demonstrate effect of different transformation/sparsifying matrices on accuracy of power signal reconstruction.

The introduced minimization problem (6.7) is nonlinear and nonconvex due to the mul-

tiplication of s_i and $\mathbf{v}$. Moreover, since the decision variables s_i are binary, the minimization problem (6.7) is a mixed integer programming formulation. Therefore, our proposed compressive sensing based mechanism is a mixed integer nonlinear program (MINLP). In the remainder of this section, we aim to alleviate the complexity of the proposed optimization scheme by converting the problem into similar but less complex optimization problems.

6.3.1 MILP

Although, both MILP and MINLP are considered as NP-hard problems [100] and they share similar computational complexity behaviors, the availability of more advanced solvers motivates us to transform our MINLP formulation to MILP. To this end, the nonlinearity of (6.5) needs to be removed. Therefore, we consider an auxiliary variable $\mathbf{U}_i = s_i\mathbf{v}$. Now, if the corresponding switch is closed (i.e., $s_i = 1$) we have $\mathbf{U}_i = \mathbf{v}$, otherwise $\mathbf{U}_i = 0$. This relationship can be formulated using the linear constraints:

$$-(1 - s_i)F \leq \mathbf{U}_i - \mathbf{v} \leq (1 - s_i)F \tag{6.8}$$

$$-Fs_i \leq \mathbf{U}_i \leq Fs_i \tag{6.9}$$

Here, F should be large enough to impose the correct constraints. If $s_i = 0$, (6.8) is not binding and (6.9) enforces $\mathbf{U}_i$ to be 0. Similarly, if $s_i = 1$, (6.8) enforces $\mathbf{U}_i = \mathbf{v}$ while the bound (6.9) is loose to allow $\mathbf{U}_i$ to have flexibility. Algorithm 6 summarizes the procedure of MILP based topology identification and state estimation.

6.3.2 Convex Relaxation

Both the MINLP and MILP formulation are categorized as nonconvex problems due to the integer decision variables. We aim to relax the nonconvexity by replacing the integer variables with continuous variables. Let assume that a continues variable represents the switch status

Algorithm 6 MILP Compressive Sensing based Joint Topology Identification and State Estimation

Step 1. Collect the compressed measurements into two vectors $\mathbf{y^S}$ and $\mathbf{y^V}$.
Step 2. Solve the following minimization problem

$$\hat{\mathcal{A}}, \hat{s}_i = \underset{\mathcal{A}, s_i}{\operatorname{argmin}} \parallel \mathcal{A} \parallel_1 + \lambda^{\mathbf{S}} \parallel \mathbf{y^S} - \mathbf{\Phi^S \Psi^S} \mathcal{A} \parallel_2 + \lambda^{\mathbf{V}} \parallel \mathbf{y^V} - \mathbf{\Phi^V \Psi^V} \mathcal{A} \parallel_2$$

subject to

$$\begin{aligned} \mathbf{Y}_{LL}\mathbf{v} + \mathbf{Y}_{L0}\mathbf{v}_0 - \mathbf{M}_{yl}\mathbf{S} &= 0 \\ -(1-s_i)F \leq \mathbf{U}_i - \mathbf{v} &\leq (1-s_i)F \\ -Fs_i \leq \mathbf{U}_i &\leq Fs_i \\ \mathcal{G}(s_1, s_2, ..., s_d) &= \mathcal{C} \end{aligned} \tag{6.10}$$

Step 3. The original state vector is: $\hat{\mathbf{z}} = \mathbf{\Psi}\hat{\mathcal{A}}$, and the estimation of switch status is: $\hat{s}_i,\ i \in \{1, 2, ..., k\}$

(i.e., $(\bar{s}_i = s_i)$), where:

$$0 \leq \bar{s}_i \leq 1 \tag{6.11}$$

We impose the power flow constraint in a second norm:

$$||\mathbf{Y}_{LL}\mathbf{v} + \mathbf{Y}_{L0}\mathbf{v}_0 - \mathbf{M}_{yl}(\mathbf{P}^T, \mathbf{Q}^T)^T||_2 \leq \epsilon \tag{6.12}$$

Obviously, the constraint (6.12) is convex, since the norm function is a convex function defined on a convex set. Although, this formulation resolves the complexity issue due to the integer variables, the constraint (6.12) is still a nonlinear problem. To tackle the nonlinearity, we employ an alternating minimization technique. According to the alternating minimization method, once we consider the switch statuses $\bar{s}_i$ as constant values. Then we solve the minimization problem to find the voltage values. At the next time, the power and voltage values are constant while we seek for an optimal solution with switch statuses as decision variables.

Algorithm 7 summarizes the convex technique for joint estimation of states and switch

statuses. In the first step, we arrange the power and voltage measurements in two separate vectors. Then, we initialize the continuous switch statuses to be 0.5 in Step 2. This initialization policy implies that we have no idea or prior information about status of the switches. Then, we start the iteration process. In Step 3, we find the states by solving the minimization problem while the $\hat{\bar{s}}_i^{j-1}$ are constant. Using the estimated states, we find the switch statuses in Step 4. In Step 5, we decide whether we obtain a convergence or we should still continue the iteration process. It should be noted that the estimated switch status in Step 4 is a continuous value (i.e. a value between 0 and 1). However, a switch status should be an integer value. Therefore, Step 6 finalizes the estimation of switch statuses.

Algorithm 7 Convex Compressive Sensing based Joint Topology Identification and State Estimation

Step 1. Collect the compressed measurements into two vectors $\mathbf{y^S}$ and $\mathbf{y^V}$.

Step 2. Initialize the parameters $\hat{\bar{s}}_i^0 = 0.5 \quad \forall i \in \{1, 2, ..., k\}$; $\hat{\mathbf{v}}_i^0 = 1\angle 0$.

Set $j = 1$ and do:

Step 3. Fix the switch status by using $\hat{\bar{s}}_i^{j-1}$. Then estimate the voltage $\hat{\mathbf{V}}^j$ using:

$$\hat{\mathcal{A}} = \underset{\mathcal{A}}{\operatorname{argmin}} \parallel \mathcal{A} \parallel_1 + \lambda^{\mathbf{S}} \parallel \mathbf{y^S} - \mathbf{\Phi^S \Psi^S} \mathcal{A} \parallel_2 + \lambda^{\mathbf{V}} \parallel \mathbf{y^V} - \mathbf{\Phi^V \Psi^V} \mathcal{A} \parallel_2$$

$$\text{subject to (6.6), (6.12)}$$

Step 4. In the following minimization, implement the voltage vector $\hat{\mathbf{V}}^j$ as fixed values and estimate switch status $\hat{\bar{s}}_i^j$ by solving:

$$\hat{\mathcal{B}} = \underset{\mathcal{B}}{\operatorname{argmin}} \parallel \mathcal{B} \parallel_1 + \lambda^{\mathbf{S}} \parallel \mathbf{y^S} - \mathbf{\Phi^S \Psi^S} \mathcal{B} \parallel_2$$

$$\text{subject to (6.6), (6.11), (6.12)}$$

Step 5. If $||\hat{\mathbf{v}}^j - \hat{\mathbf{v}}^{j-1}|| < \delta$ holds, $\hat{\mathbf{z}} = \mathbf{\Psi} \hat{\mathcal{A}}$ and proceed to Step 6.
Otherwise, $j \leftarrow j + 1$ and return to Step 3.

Step 6. Round the continuous switch status to 0 and 1:
$\hat{s}_i = \begin{cases} 1 & \hat{\bar{s}}_i \geq 0.5 \\ 0 & \hat{\bar{s}}_i < 0.5 \end{cases}$

The estimated power/voltage is equal to the values calculated at the last iteration.

6.4 Dynamic Hybrid Estimation

In this section, we focus on dynamic state estimation as well as real-time topology identification. Here, the possible topologies can be interpreted as discrete modes. That is, if the system operates in mode q, there is a unique topology relevant to that mode. Therefore, the terms "mode estimation" and "topology identification" can be used, interchangeably. In dealing with dynamical estimation of continuous states with discrete modes (hybrid systems), Interacting Multiple Model (IMM) is one of the famous technique where it combines all the acquired estimations based on the underlying models [101]. However, implementing such a criterion for our problem may not be possible due to the large number of discrete modes. For example, let us consider the d switches system that lead to 2^d discrete modes. Obviously, real-time estimation of 2^d systems requires significant computation resources, particularly for larger values of d (number of switches). Furthermore, IMM algorithms estimate the probability of discrete modes. In this case, we have 2^d probability values that could be similar. Therefore, distinguishing the true discrete mode (topology) is not a straightforward task.

Similar to the small changes in the sparse signal and its support (explained in chapter 3), the switch status may hold a correlation with the previous times. Here, we assume that we only know a probability of the switch status based on the previous statuses. To quantify this assumption, we consider the transition probability matrix denoted by (6.13). $s_{i,t}$ denote the i^{th} switch status at time t, respectively. Here, $s_{i,t}$ can be considered as a discrete-time two-state Markov chain as demonstrated by Figure 6.2. Prior works also model the changes in a distribution system as a hierarchical Markov model [102]. The transition probability rates satisfy $0 \leq \alpha_i,\ \beta_i \leq 1$ and the transition matrix for this discrete time chain corresponds to:

$$\boldsymbol{\pi}_i = \begin{bmatrix} 1-\alpha_i & \alpha_i \\ \beta_i & 1-\beta_i \end{bmatrix}, \tag{6.13}$$

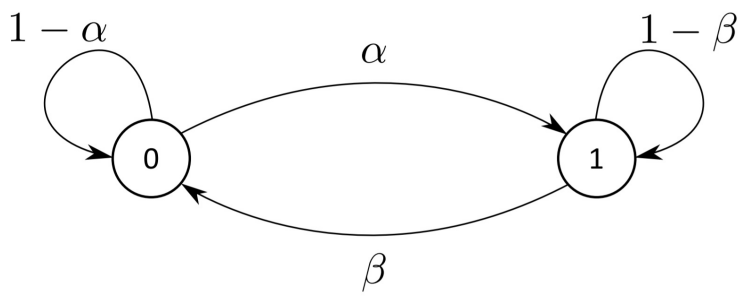

Figure 6.2: *Discrete time Markov chain model for switch status.*

where, $(\boldsymbol{\pi}_i)_{jk} = \mathbf{Pr}\big(s_{i,t} = k-1 | s_{i,t-1} = j-1\big)$. In practice, the transition probability rates (α and β) are identified using the historical data and events in the distribution grid.

Remark. *The transition probability matrix* (6.13) *suggests that probabilities of switch status changes are completely independent. However, as previously stated by* (6.6)*, the switch statuses are not completely independent. To capture this dependency, first let* $\mathcal{F}$ *to indicate the status of other all the switches except the* i^{th} *switch (i.e.,* $\mathcal{F} = \big\{s_j | j \neq i\big\}$*). Then, the transition probability matrix of the* i^{th} *switch based on status of other switches corresponds to,*

$$\boldsymbol{\pi}_{i|\mathcal{F}} = \begin{bmatrix} 1-\alpha_{i|\mathcal{F}} & \alpha_{i|\mathcal{F}} \\ \beta_{i|\mathcal{F}} & 1-\beta_{i|\mathcal{F}} \end{bmatrix}, \tag{6.14}$$

where, $\alpha_{i|\mathcal{F}} \backslash \beta_{i|\mathcal{F}}$ *are the probability rates of closing\opening of the* i^{th} *switch based on status of other switches.*

Therefore, we can obtain a probability of every switch status at each time t. Then, we directly use this probability to start the alternative minimization. More precisely, the $\hat{s}_i^1$ is not selected to be 0.5 anymore. Instead, the initial estimation of i^{th} switch status corresponds to $\hat{s}_{i,t}^1 = \mathbf{Pr}(s_{i,t} = 1 | \hat{s}_{i,t-1})$. In other words, we use the level of our confidence about a switch status to find more accurate guesstimate for switch statuses that lead to a better voltage estimation at the first iteration. Furthermore, we devise the partial knowledge about support of the sparse signal to enhance the accuracy of the estimation process. With

the assumption of slow support change, we use $\hat{N}_{t-1}$ in the ℓ_1 minimization problem:

$$\hat{\mathcal{A}} = \underset{\mathcal{A}}{\operatorname{argmin}} \parallel \mathcal{A}_{(\hat{\mathbf{N}}_{\mathbf{t-1}})^{\mathbf{c}}} \parallel_1 + \lambda^{\mathbf{S}} \parallel \mathbf{y}^{\mathbf{S}} - \mathbf{\Phi}^{\mathbf{S}} \mathbf{\Psi}^{\mathbf{S}} \mathcal{A} \parallel_2 + \lambda^{\mathbf{V}} \parallel \mathbf{y}^{\mathbf{V}} - \mathbf{\Phi}^{\mathbf{V}} \mathbf{\Psi}^{\mathbf{V}} \mathcal{A} \parallel_2$$

Algorithm 8 summarizes our proposed approach for dynamic joint topology identification and state estimation.

Algorithm 8 Dynamic Joint Topology Identification and State Estimation

Step 1. Collect the compressed measurements into two vectors $\mathbf{y}_{\mathbf{t}}^{\mathbf{S}}$ and $\mathbf{y}_{\mathbf{t}}^{\mathbf{V}}$.
Step 2. Initialize the parameters $\hat{\bar{s}}_i^0 = \mathbf{Pr}\big(s_{i,t} = 1 | \hat{s}_{i,t-1}\big) \quad \forall i \in \{1, 2, ..., k\}$; $\hat{\mathbf{V}}_i^0 = 1 \angle 0$. Set $j = 1$.
Step 3. Fix the switch status by using $\hat{\bar{s}}_{i,t}^{j-1}$. Then estimate the voltage $\hat{\mathbf{V}}_t^j$ using:

$$\hat{\mathcal{A}} = \underset{\mathcal{A}}{\operatorname{argmin}} \parallel \mathcal{A} \parallel_1 + \lambda^{\mathbf{S}} \parallel \mathbf{y}^{\mathbf{S}} - \mathbf{\Phi}^{\mathbf{S}} \mathbf{\Psi}^{\mathbf{S}} \mathcal{A} \parallel_2 + \lambda^{\mathbf{V}} \parallel \mathbf{y}^{\mathbf{V}} - \mathbf{\Phi}^{\mathbf{V}} \mathbf{\Psi}^{\mathbf{V}} \mathcal{A} \parallel_2$$
$$\text{subject to (6.6), (6.12)}$$

Step 4. In the following minimization, implement the voltage vector $\hat{\mathbf{V}}_t^j$ as fixed values and estimate switch status $\hat{\bar{s}}_{i,t}^j$ by solving:

$$\hat{\mathcal{B}} = \underset{\mathcal{B}}{\operatorname{argmin}} \parallel \mathcal{B} \parallel_1 + \lambda^{\mathbf{S}} \parallel \mathbf{y}_t^{\mathbf{S}} - \mathbf{\Phi}^{\mathbf{S}} \mathbf{\Psi}^{\mathbf{S}} \mathcal{B} \parallel_2$$
$$\text{subject to (6.6), (6.11), (6.12)}$$

Step 5. If $||\hat{\mathbf{v}}_t^j - \hat{\mathbf{v}}_t^{j-1}|| < \delta$ holds, $\hat{\mathbf{z}}_t = \mathbf{\Psi}\hat{\mathcal{A}}$ and proceed to Step 6.
Otherwise, $j \leftarrow j + 1$ and return to Step 3.
Step 6. Round the continuous switch status to 0 and 1:
$\hat{s}_{i,t} = \begin{cases} 1 & \hat{\bar{s}}_{i,t} \geq 0.5 \\ 0 & \hat{\bar{s}}_{i,t} < 0.5 \end{cases}$

6.5 Simulation Results

To evaluate our proposed methods, we establish our experiments based on IEEE 37-test feeder system with six switches that is described in 6.2. The test system has an unbalanced three-phase framework. We apply Monte-Carlo simulation strategy where the proposed algorithms are executed multiple times for different topologies. Then, we present the averaged

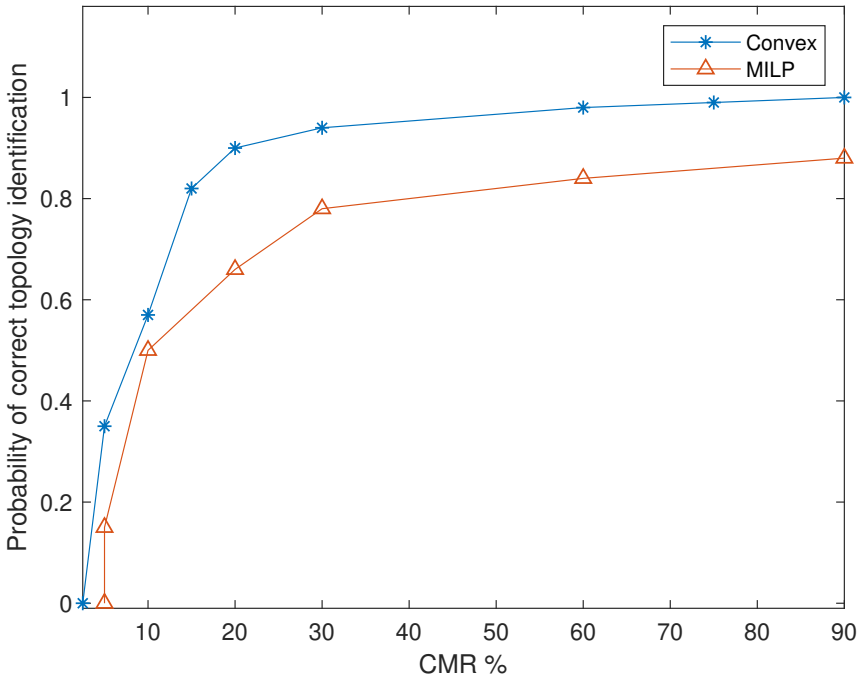

Figure 6.3: *Convex vs MILP topology identification for 6 switches system.*

performances. Here, we execute 2000 simulations. Sparsifying matrix for both power and voltage vectors ($\mathbf{\Psi^S}$ and $\mathbf{\Psi^V}$) are DCT matrices. Here, we apply random measurement matrices $\mathbf{\Phi}^S$ and $\mathbf{\Phi}^V$ consisting of random Bernoulli entries which selects $\frac{m}{n} \times 100\%$ of the data. Similar to chapter 4, we collect data from actual consumers that are publicly available at eGauge website [2].

Based on different CMR (percentage of measurements), Figure 6.3 demonstrates and compares the probabilities of accurate topology identification performed by the convex method (Algorithm 7) and MILP approach (Algorithm 6). Obviously, the convex method outperforms the MILP technique, i.e., Algorithm 7 has been more successful relative to the Algorithm 6 in terms of accurate topology identification. Similarly, the convex approach shows a better performance compared to the MILP method, in terms of state estimation accuracy. It should be noted that the simulation time is shorter for the convex method. Table 6.1 demonstrates the required times for implementing each method.

The dynamic topology identification technique (Algorithm 8) is evaluated based on the

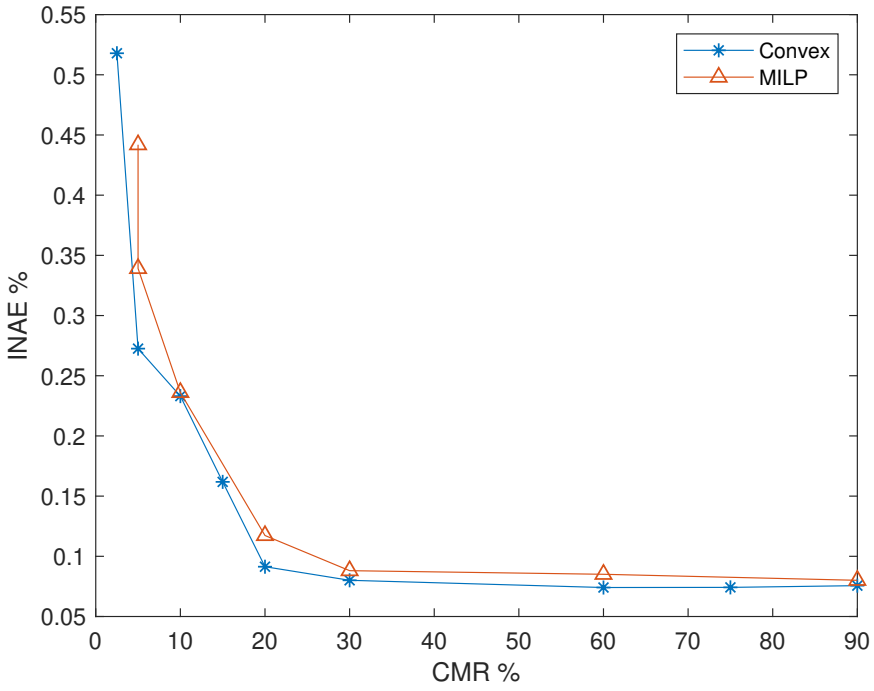

Figure 6.4: *Topology identification performance with different CMR*

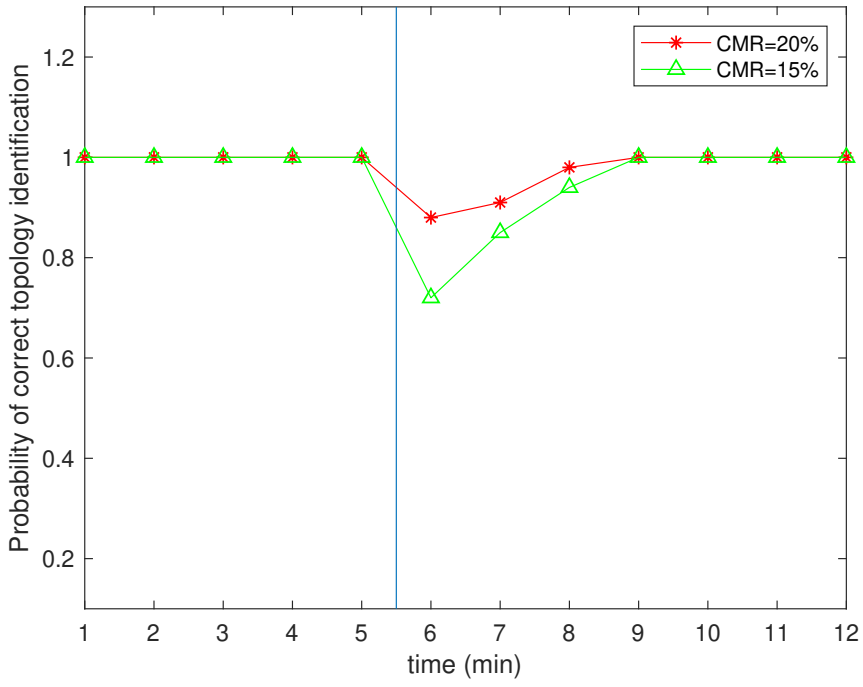

Figure 6.5: *Dynamic TI for two different CMR when the topology changes.*

Table 6.1: *Time requirement for 1 run, 1 CMR*

Algorithm 6 and 7	Time
MILP	29.2s
Convex	10.4s

practical load data that is measured every minute. We assume that we have an accurate knowledge about the states and switches. Furthermore, we assume that the topology changes between the fifth and sixth minutes (i.e., $5 < t < 6$). Figure 6.5 demonstrates the result of this setup. Interestingly, when we have partial information about the topology, the performance is improved. However, we witness a degradation in performance after the change in topology configuration of the IEEE 37 nodes test feeder. However, after three minutes the topology could be correctly identified. As a result, the dynamic technique outperforms the static techniques when the partial knowledge of switch statuses is available.

6.6 Summary

In this chapter, we proposed an MINLP framework that aims to jointly estimate the states and topology of the distribution system. We removed the nonlinearity in the MINLP formulation by introducing auxiliary variables that led to an MILP. Furthermore, to provide a faster solution, we relaxed the nonlinearity of MINLP by replacing the integer variables with continuous decision variables. Then, we apply an alternative minimization approach that improves both state estimation and topology identification. Finally, we introduced a hybrid dynamic framework that incorporates the previous information about the switch statuses for current time estimation. More precisely, a Markov jump model was proposed for switch status that helps to improve topology identification especially when the compression ratio is low. Besides, we employed the previous information about the support set that further improved the performance of our method. Simulation results validate the applicability of our techniques.

Chapter 7

Conclusion and future works

In this chapter, we summarize the contributions of this dissertation and discuss future research directions.

7.1 Summary

In this dissertation, we addressed the issue of efficient data aggregation in smart distribution systems. Future smart distribution grids will incorporate several different components including sensors, renewable generation (e.g., wind, solar) and AMI. Along with situational awareness, real time operations and control require the aggregation of data generated by all sensors and smart metering equipments. This results in a data deluge due to the physical scale of the system and relatively high sampling rate. This large volume and the rate of streaming data from the smart meters and sensors will impair the data aggregation process as the underlying network will not be able to keep up with the large amount of information. To alleviate this stress, compressive sensing has been an effective approach to recover the full dimensional state vector from lower dimensional compressed data in distribution networks. With this background, we summarize the contributions of this dissertation where four research questions are investigated in the field of sparse recovery in distribution grids.

- In chapter 2, the proposed approaches leveraged the underlying spatio-temporal correlation in generation and load that results in sparsity in the wavelet domain. The recovered power measurements were then used to estimate the voltage states with high fidelity as demonstrated via simulation of the IEEE 34 node test system. With each projection matrix used in compressed sensing representing a specific modality of data aggregation, we demonstrated that a Bernoulli projection matrix outperforms Gaussian projection matrix, especially at low compression ratios. Furthermore, we proposed a direct methods that incorporated a linearized power flow constraint in to the ℓ_1 minimization which in turn enabled us to estimate both power and voltage, simultaneously. We validated the efficiency of this technique using IEEE 37-node test system. Consequently, the impressive performance of compressive sensing has motivated us to address some critical questions in this field, particularly with their application in distribution systems.

- In chapter 3, we proposed an approach to dynamically reconstruct a power signal from a smaller number of measurements. To this end, we implemented two recursive dynamic CS methods known as streaming modified weighted-ℓ_1 CS and Kalman filtered CS. We, employed a practical power data from an actual distribution system to validate the underlying assumptions (slow signal and support change) required by dynamic recursive CS techniques. Since the recursive methods use the prior information about the sparse signal and its support set, they could improve the previous results obtained by classic CS. The superior performance of the two recursive CS methods were validated via two examples (the IEEE 34 node test feeder system and PecanStreet data).

- Chapter 4 and chapter 5 investigated the effect of communication network impairment (losses/delays) on compressive sensing. In chapter 5, we studied dynamic sparse signal recovery in the presence of lossy measurements. We first modified an existing approach for Kalman filtered compressive sensing (KF-CS) to obtain both sparse signal estimate and support set estimate while considering measurement losses. Then, we quantified the error dynamics in both sparse signal estimation and support set estimation. Using

nonlinear input-to-state stability analysis, we provided an upper bound for the expected covariance of the estimation error for a given rate of information loss. Then, we updated the upper bound for two special cases: (1) if the support set remained constant over time, and (2) in addition to the constant support set, a reduced model was observable. As a result, we could calculate the critical value for loss in measurements that ensures convergence of error in the KF-CS based algorithm. The provided upper bounds in turn allowed us to evaluate the critical value for loss in measurements that ensured a desirable bound for error in Algorithm 4.

- In chapter 5, we addressed the problem of compressive sensing in presence of measurements with random delays. To this end, we first proposed a modified compressive sensing technique that incorporates the delayed measurements in the sparse recovery process. To accomplish this goal, we employed an existing model that captures the temporal correlation of the sparse signals. Consequently, we could utilize the information carried by the delayed measurements for estimation at the current time. In addition to the stochastic nature of sparse signal evolution over time, the parameters of the underlying dynamic model may not be precisely identified. Therefore, we quantified the error in the recovered signal based on the errors in the modeling of the underlying dynamic correlation. Simulations validated both theoretical results and highlighted the efficiency of the proposed recursive estimation approaches in the presence of lossy/delayed measurements.

- In chapter 6, we addressed the problem of joint state estimation and topology identification in distribution systems. To this end, we first formulated an MINLP framework. In addition, we extracted an MILP formulation from the MINLP by removing its nonlinearity using auxiliary variables. Furthermore, to provide a faster solution, we relaxed the nonlinearity of MINLP by replacing the integer variables with continuous decision variables. Subsequently, an alternative minimization technique is employed that enhances performance of both state estimation and topology identification. Finally, a Markov jump model was proposed for switch status that helps to improve topology

identification especially when the compression ratio is low. Besides, we employed the previous information about the support set that further improved the performance of our method. Simulation results demonstrated the applicability of our techniques.

7.1.1 Future Directions

Based on the work accomplished in this dissertation, some follow-on efforts and possible future research directions are highlighted in the following:

- In chapter 3 and 4, we only consider power values in the dynamic signal of interest. Similar to chapter 2 and 5 that estimate the voltage values by incorporating the power flow relationship, it would be complementary if we aim to estimate both the power and voltage signals. To accomplish this, we need to apply the power flow constraints (2.6) and the error analysis in chapter 4 must be revised with this new setup.

- For the dynamics of the sparse signal, we consider a Gaussian random walk model with a constant covariance matrix. Although this model has been extensively employed in many existing literature, this model may not capture some dynamical aspect of the sparse signal. For example, [103] develops a KF-CS technique based on a more realistic model of dynamic sparse signal. Therefore, it would be interesting to investigate the dynamic methods in chapter 3 and error analysis in chapter 4 for other realistic models that capture true dynamics of the sparse signal.

- All the CS techniques applied in this dissertation are based on data aggregation at a centralized location which may result in latency and other data aggregation challenges. While compressive sensing idea presented earlier provides a way to address this issue, it is advisable to implement a decentralized state estimator to minimize latency and overcome the challenge of data aggregation.

- In smart distribution grids, measurements are acquired from various sources such as micro-PMUs, SCADA sensors and AMI. The corresponding sensors and smart meters may be asynchronous with different sampling rates that turn the dynamic CS to a

more challenging task. To cope with the possible challenges, it would be interesting to develop a systematic approach that integrates asynchronous multi-time scale measurements from different sources.

- In this dissertation, we separately attacked the problem of dynamic CS in the presence of delayed and lossy measurements. However, in a practical situation, a sequence of measurements may encounter with both problems of delayed and lossy measurements. Although, this issue is well understood in discrete-time linear systems [104], the mentioned problem has not been addressed in dynamic CS framework. Therefore, it would be helpful to design and analyze a technique that incorporates both lossy and delayed measurements in a dynamic CS framework.

Bibliography

[1] Kazunori Nagasawa, Charles R Upshaw, Joshua D Rhodes, Chris L Holcomb, David A Walling, and Michael E Webber. Data management for a large-scale smart grid demonstration project in austin, texas. In *ASME 2012 6th International Conference on Energy Sustainability collocated with the ASME 2012 10th International Conference on Fuel Cell Science, Engineering and Technology*, pages 1027–1031. Citeseer, 2012.

[2] egauge. URL http://egauge141.egaug.es/.

[3] Piyush Gupta and Panganmala R Kumar. The capacity of wireless networks. *IEEE Transactions on information theory*, 46(2):388–404, 2000.

[4] Yih-Fang Huang, Stefan Werner, Jing Huang, Neelabh Kashyap, and Vijay Gupta. State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid. *IEEE Signal Processing Magazine*, 29(5):33–43, 2012.

[5] Anggoro Primadianto and Chan-Nan Lu. A review on distribution system state estimation. *IEEE Transactions on Power Systems*, 32(5):3875–3883, 2016.

[6] Kaveh Dehghanpour, Zhaoyu Wang, Jianhui Wang, Yuxuan Yuan, and Fankun Bu. A survey on state estimation techniques and challenges in smart distribution systems. *IEEE Transactions on Smart Grid*, 10(2):2312–2322, 2018.

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

[8] Efthymios Manitsas, Ravindra Singh, Bikash C Pal, and Goran Strbac. Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling. *IEEE Transactions on Power Systems*, 27(4):1888–1896, 2012.

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

[10] Gustavo Valverde, Andrija T Saric, and Vladimir Terzija. Stochastic monitoring of distribution networks including correlated input variables. *IEEE Transactions on Power Systems*, 28(1):246–255, 2012.

[11] Alexandre Ciancio, Sundeep Pattem, Antonio Ortega, and Bhaskar Krishnamachari. Energy-efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm. In *Proceedings of the 5th international conference on Information processing in sensor networks*, pages 309–316. ACM, 2006.

[12] Marco F Duarte and Yonina C Eldar. Structured compressed sensing: From theory to applications. *IEEE Transactions on signal processing*, 59(9):4053–4085, 2011.

[13] Michel P Tcheou, Lisandro Lovisolo, Moises V Ribeiro, Eduardo AB Da Silva, Marco AM Rodrigues, João MT Romano, and Paulo SR Diniz. The compression of electric signal waveforms for smart grids: State of the art and future trends. *IEEE Transactions on Smart Grid*, 5(1):291–302, 2013.

[14] Vincenzo Loia, Stefania Tomasiello, and Alfredo Vaccaro. Fuzzy transform based compression of electric signal waveforms for smart grids. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(1):121–132, 2016.

[15] Zhen Hu, Salman Mohagheghi, and Mina Sartipi. Flexible data acquisition, compression, and reconstruction in advanced metering infrastructure. In *2016 Clemson University Power Systems Conference (PSC)*, pages 1–6. IEEE, 2016.

[16] Julio Cesar Stacchini de Souza, Tatiana Mariano Lessa Assis, and Bikash Chandra Pal. Data compression in smart distribution systems via singular value decomposition. *IEEE Transactions on Smart Grid*, 8(1):275–284, 2015.

[17] M Majidi, M Etezadi-Amoli, and H Livani. Distribution system state estimation using

compressive sensing. *International Journal of Electrical Power & Energy Systems*, 88: 175–186, 2017.

[18] Mohammad Babakmehr, Farnaz Harirchi, Ahmed Al-Durra, SM Muyeen, and Marcelo Godoy Simões. Compressive system identification for multiple line outage detection in smart grids. *IEEE Transactions on Industry Applications*, 55(5):4462–4473, 2019.

[19] Lulu Wen, Kaile Zhou, Shanlin Yang, and Lanlan Li. Compression of smart meter big data: A survey. *Renewable and Sustainable Energy Reviews*, 91:59–69, 2018.

[20] SM Shafiul Alam, Balasubramaniam Natarajan, and Anil Pahwa. Distribution grid state estimation from compressed measurements. *IEEE Transactions on Smart Grid*, 5(4):1631–1642, 2014.

[21] Michael Wakin, Jason N Laska, Marco F Duarte, Dror Baron, Shriram Sarvotham, Dharmpal Takhar, Kevin F Kelly, and Richard G Baraniuk. Compressive imaging for video representation and coding. In *Picture Coding Symposium*, volume 1, page 13, 2006.

[22] Namrata Vaswani and Jinchun Zhan. Recursive recovery of sparse signal sequences from compressive measurements: A review. *IEEE Transactions on Signal Processing*, 64(13):3523–3549, 2016.

[23] Ioan Serban, Sandra Céspedes, Corneliu Marinescu, Cesar A Azurdia-Meza, Juan S Gómez, and Doris Sáez Hueichapan. Communication requirements in microgrids: A practical survey. *IEEE Access*, 8:47694–47712, 2020.

[24] Md Masud Rana, Li Li, and Steven W Su. Distributed state estimation of smart grids with packet losses. *Asian Journal of Control*, 19(4):1306–1315, 2017.

[25] Masud Rana, Li Li, and Steven W Su. Distributed state estimation over unreliable communication networks with an application to smart grids. *IEEE Transactions on Green Communications and Networking*, 1(1):89–96, 2017.

[26] Imran Mohammed, Sreenath Jayakumar Geetha, Swapnil Sadashiv Shinde, Ketan Rajawat, and Saikat Chakrabarti. Modified re-iterated kalman filter for handling delayed and lost measurements in power system state estimation. *IEEE Sensors Journal*, 20 (7):3946–3955, 2019.

[27] Vanja G Švenda, Aleksandar M Stanković, Andrija T Sarić, and Mark K Transtrum. Flexible hybrid state estimation for power systems with communication irregularities. *IET Generation, Transmission & Distribution*, 14(11):2111–2119, 2020.

[28] Bruno Sinopoli, Luca Schenato, Massimo Franceschetti, Kameshwar Poolla, Michael I Jordan, and Shankar S Sastry. Kalman filtering with intermittent observations. *IEEE transactions on Automatic Control*, 49(9):1453–1464, 2004.

[29] Dominik Sierociuk, Inés Tejado, and Blas M Vinagre. Improved fractional kalman filter and its application to estimation over lossy networks. *Signal Processing*, 91(3): 542–552, 2011.

[30] Hong Lin, Hongye Su, Peng Shi, Zhan Shu, Renquan Lu, and Zheng-Guang Wu. Optimal estimation and control for lossy network: stability, convergence, and performance. *IEEE Transactions on Automatic Control*, 62(9):4564–4579, 2017.

[31] Ling Shi, Lihua Xie, and Richard M Murray. Kalman filtering over a packet-delaying network: A probabilistic approach. *Automatica*, 45(9):2134–2140, 2009.

[32] Wei Lu and Namrata Vaswani. Modified compressive sensing for real-time dynamic mr imaging. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 3045–3048. IEEE, 2009.

[33] Felix F Wu and W-HE Liu. Detection of topology errors by state estimation (power systems). *IEEE Transactions on Power Systems*, 4(1):176–183, 1989.

[34] George N Korres and Petros J Katsikas. Identification of circuit breaker statuses in wls state estimator. *IEEE Transactions on Power Systems*, 17(3):818–825, 2002.

[35] Hazhar Sufi Karimi and Balasubramaniam Natarajan. Compressive sensing based state estimation for three phase unbalanced distribution grid. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–6. IEEE, 2017.

[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. doi: 10.1109/ACCESS.2020.3035378.

[37] Hazhar Sufi Karimi and Balasubramaniam Natarajan. Recursive dynamic compressive sensing in smart distribution systems. In *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE, 2020.

[38] Namrata Vaswani and Jinchun Zhan. Recursive recovery of sparse signal sequences from compressive measurements: A review. *IEEE Trans. Signal Processing*, 64(13): 3523–3549, 2016.

[39] Hazhar Sufi Karimi and Balasubramaniam Natarajan. Kalman filtered compressive sensing with intermittent observations. *Signal Processing*, 163:49–58, 2019.

[40] Hazhar Sufi Karimi and Balasubramaniam Natarajan. Dynamic signal recovery in distribution grids using compressive lossy measurements. *IET Smart Grid*, under review (2020).

[41] Hazhar Sufi Karimi and Balasubramaniam Natarajan. Dynamic signal recovery with compressive delayed measurements. *IEEE Signal Processing Letters*, under review (2020).

[42] Hazhar Sufi Karimi and Balasubramaniam Natarajan. Joint topology identification and state estimation in unobservable distribution grids. *IEEE Transactions on Smart Grid*, to be submitted (2020).

[43] Amit Joshi, Laya Das, Balasubramaniam Natarajan, and Babji Srinivasan. A framework for efficient information aggregation in smart grid. *IEEE Transactions on Industrial Informatics*, 15(4):2233–2243, 2018.

[44] R. Madbhavi, H. S. Karimi, B. Natarajan, and B. Srinivasan. Tensor completion based state estimation in distribution systems. In *2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5, 2020.

[45] Andrey Bernstein and Emiliano Dall'Anese. Linear power-flow models in multiphase distribution networks. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6. IEEE, 2017.

[46] James O Berger, Victor De Oliveira, and Bruno Sansó. Objective bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96(456): 1361–1374, 2001.

[47] Peder Bacher, Henrik Madsen, and Henrik Aalborg Nielsen. Online short-term solar power forecasting. *Solar energy*, 83(10):1772–1783, 2009.

[48] Emmanuel Candes and Terence Tao. Decoding by linear programming. *arXiv preprint math/0502327*, 2005.

[49] Yair Rivenson and Adrian Stern. Compressed imaging with a separable sensing operator. *IEEE Signal Processing Letters*, 16(6):449–452, 2009.

[50] Michael Grant and Stephen Boyd. Cvx: Matlab software for disciplined convex programming, version 2.1, 2014.

[51] Richard G Baraniuk. Compressive sensing [lecture notes]. *IEEE signal processing magazine*, 24(4):118–121, 2007.

[52] Marco F Duarte and Richard G Baraniuk. Kronecker compressive sensing. *IEEE Transactions on Image Processing*, 21(2):494–504, 2011.

[53] Wei Lu and Namrata Vaswani. Regularized modified bpdn for noisy sparse reconstruction with partial erroneous support and signal value knowledge. *IEEE Transactions on Signal Processing*, 60(1):182–196, 2011.

[54] Namrata Vaswani. Kalman filtered compressed sensing. In *2008 15th IEEE International Conference on Image Processing*, pages 893–896. IEEE, 2008.

[55] Pecan street. URL https://www.pecanstreet.org/.

[56] M Salman Asif and Justin Romberg. Sparse recovery of streaming signals using ℓ_1-homotopy. *IEEE Transactions on Signal Processing*, 62(16):4209–4223, 2014.

[57] Fatima Salahdine, Naima Kaabouch, and Hassan El Ghazi. A survey on compressive sensing techniques for cognitive radio networks. *Physical Communication*, 20:61–73, 2016.

[58] Lanchao Liu, Zhu Han, Zhiqiang Wu, and Lijun Qian. Collaborative compressive sensing based dynamic spectrum sensing and mobile primary user localization in cognitive radio networks. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pages 1–5. Citeseer, 2011.

[59] Namrata Vaswani. Kalman filtered compressed sensing. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 893–896. IEEE, 2008.

[60] Xiao Ma, Seddik M Djouadi, and Husheng Li. State estimation over a semi-markov model based cognitive radio system. *IEEE Transactions on Wireless Communications*, 11(7):2391–2401, 2012.

[61] Xianghui Cao, Xiangwei Zhou, Lu Liu, and Yu Cheng. Energy-efficient spectrum sensing for cognitive radio enabled remote state estimation over wireless channels. *IEEE Transactions on Wireless Communications*, 14(4):2058–2071, 2015.

[62] Masud Rana, Li Li, and Steven W Su. Distributed state estimation over unreliable communication networks with an application to smart grids. *IEEE Transactions on Green Communications and Networking*, 2017.

[63] Md Masud Rana, Li Li, and Steven W Su. Distributed state estimation of smart grids with packet losses. *Asian Journal of Control*, 19(4):1306–1315, 2017.

[64] Emmanuel Candes, Terence Tao, et al. The dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351, 2007.

[65] Wei Lu and Namrata Vaswani. Modified compressive sensing for real-time dynamic mr imaging. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 3045–3048. IEEE, 2009.

[66] Bradley Efron, Trevor Hastie, Iain Johnstone, Robert Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

[67] Matti Raitoharju and Robert Piché. A survey of code optimization methods for kalman filter extensions. *arXiv preprint arXiv:1512.03077*, 2015.

[68] Christopher D'Souza and Renato Zanetti. Information formulation of the udu kalman filter. *IEEE Transactions on Aerospace and Electronic Systems*, 2018.

[69] CAMILA CS Caiado and PUSHPA N Rathie. Polynomial coefficients and distribution of the sum of discrete uniform variables. In *Eighth Annual Conference of the Society of Special Functions and their Applications, Pala, India, Society for Special Functions and their Applications*, 2007.

[70] Harold V Henderson and Shayle R Searle. On deriving the inverse of a sum of matrices. *Siam Review*, 23(1):53–60, 1981.

[71] Josef Stoer and Roland Bulirsch. *Introduction to numerical analysis*, volume 12. Springer Science & Business Media, 2013.

[72] Richard Bouldin. The pseudo-inverse of a product. *SIAM Journal on Applied Mathematics*, 24(4):489–495, 1973.

[73] Emmanuel J Candes. The restricted isometry property and its implications for compressed sensing. *Comptes rendus mathematique*, 346(9-10):589–592, 2008.

[74] Jinchun Zhan and Namrata Vaswani. Time invariant error bounds for modified-cs-based sparse signal sequence recovery. *IEEE Transactions on Information Theory*, 61 (3):1389–1409, 2015.

[75] Zhong-Ping Jiang and Yuan Wang. Input-to-state stability for discrete-time nonlinear systems. *Automatica*, 37(6):857–869, 2001.

[76] Zhong-Ping Jiang and Yuan Wang. Input-to-state stability for discrete-time nonlinear systems. *Automatica*, 37(6):857–869, 2001.

[77] Bruno Sinopoli, Luca Schenato, Massimo Franceschetti, Kameshwar Poolla, Michael I Jordan, and Shankar S Sastry. Kalman filtering with intermittent observations. *IEEE transactions on Automatic Control*, 49(9):1453–1464, 2004.

[78] Ahmad Reza Malekpour and Anil Pahwa. Radial test feeder including primary and secondary distribution network. In *2015 North American Power Symposium (NAPS)*, pages 1–9. IEEE, 2015.

[79] Asok Ray, LW Liou, and JH Shen. State estimation using randomly delayed measurements. 1993.

[80] Edwin Yaz and Asok Ray. Linear unbiased state estimation under randomly varying bounded sensor delay. *Applied Mathematics Letters*, 11(4):27–32, 1998.

[81] Alexey S Matveev and Andrey V Savkin. The problem of state estimation via asynchronous communication channels with irregular transmission times. *IEEE transactions on automatic control*, 48(4):670–676, 2003.

[82] Zidong Wang, Daniel WC Ho, and Xiaohui Liu. Robust filtering under randomly varying sensor delay with variance constraints. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 51(6):320–326, 2004.

[83] B Chen, L Yu, and W-A Zhang. Robust kalman filtering for uncertain state delay

systems with random observation delays and missing measurements. *IET Control Theory & Applications*, 5(17):1945–1954, 2011.

[84] Xiao Lu, Huanshui Zhang, Wei Wang, and Kok-Lay Teo. Kalman filtering for multiple time-delay systems. *Automatica*, 41(8):1455–1461, 2005.

[85] Raquel Caballero-Águila, Aurora Hermoso-Carazo, José D Jiménez-López, Josefa Linares-Pérez, and Seiichi Nakamori. Signal estimation with multiple delayed sensors using covariance information. *Digital Signal Processing*, 20(2):528–540, 2010.

[86] Justin Ziniel and Philip Schniter. Dynamic compressive sensing of time-varying signals via approximate message passing. *IEEE transactions on signal processing*, 61(21): 5270–5284, 2013.

[87] Hazhar Sufi Karimi, Kumarsinh Jhala, and Balasubramaniam Natarajan. Impact of real-time pricing attack on demand dynamics in smart distribution systems. In *2018 North American Power Symposium (NAPS)*, pages 1–6. IEEE, 2018.

[88] Mohammad Farajollahi, Alireza Shahsavari, and Hamed Mohsenian-Rad. Topology identification in distribution systems using line current sensors: An milp approach. *IEEE Transactions on Smart Grid*, 11(2):1159–1170, 2019.

[89] Guido Cavraro, Andrey Bernstein, Vassilis Kekatos, and Yingchen Zhang. Real-time identifiability of power distribution network topologies with limited monitoring. *IEEE Control Systems Letters*, 4(2):325–330, 2019.

[90] NS Da Silva, A Simões Costa, KA Clements, and E Andreoli. Simultaneous estimation of state variables and network topology for power system real-time modeling. *Electric Power Systems Research*, 133:338–346, 2016.

[91] Mohammad Mohammadian, Hamidreza Momeni, Javad Zahiri, and Hazhar Sufi Karimi. Switched adaptive observer for structure identification in gene regulatory networks. In *28th Iranian Conference on Electrical Engineering (ICEE 2020)*, pages 1–6. IEEE, 2020.

[92] Satya Jayadev Pappu, Nirav Bhatt, Ramkrishna Pasumarthy, and Aravind Rajeswaran. Identifying topology of low voltage distribution networks based on smart meter data. *IEEE Transactions on Smart Grid*, 9(5):5113–5122, 2017.

[93] Zhuang Tian, Wenchuan Wu, and Boming Zhang. A mixed integer quadratic programming model for topology identification in distribution network. *IEEE Transactions on Power Systems*, 31(1):823–824, 2015.

[94] Guido Cavraro and Reza Arghandeh. Power distribution network topology detection with time-series signature verification method. *IEEE Transactions on Power Systems*, 33(4):3500–3509, 2017.

[95] Nan Zhou, Lingen Luo, Gehao Sheng, and Xiuchen Jiang. Power distribution network dynamic topology awareness and localization based on subspace perturbation model. *IEEE Transactions on Power Systems*, 35(2):1479–1488, 2019.

[96] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[97] Shanshan Ma, Liu Su, Zhaoyu Wang, Feng Qiu, and Ge Guo. Resilience enhancement of distribution grids against extreme weather events. *IEEE Transactions on Power Systems*, 33(5):4842–4853, 2018.

[98] Roy Billinton and Satish Jonnavithula. Optimal switching device placement in radial distribution systems. *IEEE transactions on power delivery*, 11(3):1646–1651, 1996.

[99] Amit Joshi, Laya Das, Balasubramaniam Natarajan, and Babji Srinivasan. Effect of transformation in compressed sensing of smart grid data. In *2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia)*, pages 177–182. IEEE, 2019.

[100] Ravindran Kannan and Clyde L Monma. On the computational complexity of integer programming problems. In *Optimization and Operations Research*, pages 161–172. Springer, 1978.

[101] Henk AP Blom and Yaakov Bar-Shalom. The interacting multiple model algorithm for systems with markovian switching coefficients. *IEEE transactions on Automatic Control*, 33(8):780–783, 1988.

[102] Richard Eric Brown, Shalini Gupta, Richard D Christie, Subrahmanyam S Venkata, and R Fletcher. Distribution system reliability assessment using hierarchical markov modeling. *IEEE Transactions on power Delivery*, 11(4):1929–1934, 1996.

[103] Avishy Carmi, Pini Gurfil, and Dimitri Kanevsky. Methods for sparse signal recovery using kalman filtering with embedded pseudo-measurement norms and quasi-norms. *IEEE Transactions on Signal Processing*, 58(4):2405–2409, 2009.

[104] Shuli Sun. Optimal linear filters for discrete-time systems with randomly delayed and lost measurements with/without time stamps. *IEEE Transactions on Automatic Control*, 58(6):1551–1556, 2012.

Appendix A

Proofs of Chapter 4

A.1 Proof of lemma 3

According to lemma 2, if $\{i \in N_t\}$, and measurements are available: $\mathbf{Pr}(\{N_{i,t}\} \setminus \{T_{i,t}\}) = \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a)\big].\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}| < \alpha_a) \mid |(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a\big]$. As we discussed in section 4.4, $\mathbf{e}_{i,t}^{\mathbf{Q}}$ is a Bernoulli random variable, thus, $\mathbb{E}_{\lambda_t,\mathbf{e}^{\mathbf{Q}}}(\{N_{i,t}\}\setminus\{T_{i,t}\}) \leq -\sigma^2\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a)\big].\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}| < \alpha_a)\big]$. Since $(\hat{\mathbf{x}}_{t,mod})_i$ is a Gaussian random variable, it is easy to find to find $\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a)\big]$. From Theorem 3 and the discussion after that, we conclude that $(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}$ is a Gaussian random variable with variance $4\mathbf{P}_{i,t} + \mathbf{D}_{i,t}$. We compute $\mathcal{Q}(q(\mathbf{P}_{i,t}, \mathbf{D}_{i,t}))$, where $q(\mathbf{P}_{i,t}, \mathbf{D}_{i,t}) = \frac{\alpha_a}{\sqrt{4\mathbf{P}_{i,t}+\mathbf{D}_{i,t}}}$ is a transformation function to convert $(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}$ to a standard normal distribution. However, since we take $\mathbb{E}_{\lambda_t,\mathbf{e}^{\mathbf{Q}}}\big[\mathcal{Q}(q(\mathbf{P}_{i,t}, \mathbf{D}_{i,t}))\big]$ and according to Theorem 4, $\mathbf{P}_{i,t}$ is represented in terms of λ_t and $\mathbf{e}^{\mathbf{Q}}{}_t$, we can not apply the $\mathcal{Q}$-function directly. Therefore, we seek for a bound in terms of $\mathbb{E}_{\lambda_t,\mathbf{e}^{\mathbf{Q}}}(\mathbf{P}_{i,t})$.

Consider $\mathbf{Z}_{i,t} = \frac{(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}}{\sqrt{4\mathbf{P}_{i,t}+\mathbf{D}_{i,t}}}$. Obviously, $\mathbf{Z}_t \sim N(0,1)$. Thus, $\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}| > \alpha_a)\big] = \mathbf{Pr}\big[|\mathbf{Z}_{i,t}| > q\big]$, where $q = \frac{\alpha_a}{\sqrt{4\mathbf{P}_{i,t}+\mathbf{D}_{i,t}}}$. This probability is bounded by $\sqrt{\frac{2}{\pi}}\frac{q}{1+q^2}e^{-\frac{q^2}{2}} \leq \mathbf{Pr}\big[|\mathbf{Z}_{i,t}| > q\big] \leq \sqrt{\frac{2}{\pi}}\frac{1}{q}e^{-\frac{q^2}{2}}$. Hence, we conclude that $\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}| < \alpha_a)\big] =$

$1 - \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}_{i,t}^{mod}| > \alpha_a)\big] = 1 - \mathbf{Pr}\big[|\mathbf{Z}_{i,t}| > q\big] \leq 1 - \sqrt{\frac{2}{\pi}}\frac{q}{1+q^2}e^{-\frac{q^2}{2}}$. Therefore,

$$|\mathbb{E}_{\lambda_t,\mathbf{e}^{\mathbf{Q}}}(\{N_{i,t}\} \setminus \{T_{i,t}\})| \leq \sigma^2 \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a)\big].(1 - \sqrt{\frac{2}{\pi}}\mathbb{E}_{\lambda_t,\mathbf{e}^{\mathbf{Q}}}\frac{q}{1+q^2}e^{-\frac{q^2}{2}}). \quad (A.1)$$

Lemma 6. *If there exists an α_a and $U = \frac{\alpha_a^2}{4\mathbf{P}_{i,t}+\mathbf{D}_{i,t}}$, such that $\alpha_a U^4 + 2\alpha_a U^3 + (2-4\alpha_a)U^2 + (4-6\alpha_a)U + 2 + 3\alpha_a \geq 0$, we have:*

$$\frac{h}{1+h^2}e^{-\frac{h^2}{2}} \leq \mathbb{E}_{\lambda_t,\mathbf{e}^{\mathbf{Q}}}\frac{q}{1+q^2}e^{-\frac{q^2}{2}} \quad (A.2)$$

Proof. For simplification, let denote $r = \mathbf{P}_{i,t}$ and $d = \mathbf{D}_{i,t}$. First, we prove that $f(r) = \frac{q}{1+q^2}e^{-\frac{q^2}{2}}$ is a convex function. To this end, we calculate the second derivative of $f(r)$. From the chain rule for the derivation of the composition of two function, we have: $\frac{\partial^2 f(r)}{\partial r^2} = \frac{\partial^2 f}{\partial q^2}(\frac{\partial q}{\partial r})^2 + \frac{\partial f}{\partial q}\frac{\partial^2 q}{\partial r^2}$. Here,

$$\frac{\partial q}{\partial r} = -2\alpha_a(r+d)^{-\frac{3}{2}}, \quad (A.3)$$

$$\frac{\partial^2 q}{\partial r^2} = 12\alpha_a(r+d)^{-\frac{5}{2}}, \quad (A.4)$$

$$\frac{\partial f}{\partial q} = (1+q^2)^{-1}e^{-\frac{q^2}{2}} + q(-2q)(1+q^2)^{-2}e^{-\frac{q^2}{2}} + q(1+q^2)^{-1}(-q)e^{-\frac{q^2}{2}} = (1+q^2)^{-1}e^{-\frac{q^2}{2}}(1-q^2-2q^2(1+q^2)^{-1}), \quad (A.5)$$

$$\begin{aligned}\frac{\partial^2 f}{\partial q^2} &= -2q(1+q^2)^{-2}e^{-\frac{q^2}{2}}(1-q^2-2q^2(1+q^2)^{-1}) + \\ &(1+q^2)^{-1}(-qe^{-\frac{q^2}{2}})(1-q^2-2q^2(1+q^2)^{-1}) + \\ &(1+q^2)^{-1}e^{-\frac{q^2}{2}}(-4q(1+q^2)^{-1}+4q^3(1+q^2)^{-1}-2q) = \\ &-q(1+q^2)^{-1}e^{-\frac{q^2}{2}}((1-2q^2(1+q^2)^{-1}-q^2)(2(1+q^2)^{-1}+1)+2). \quad (A.6)\end{aligned}$$

Using these results, we have $\frac{\partial^2 f(r)}{\partial r^2} = 4\alpha_a(4r+d)^3(1+q^2)^{-1}e^{-\frac{q^2}{2}}(-\alpha_a q(1-2q^2(1+q^2)^{-1}-q^2)(2(1+q^2)^{-1}+1)+2+3\alpha_a q^{-1}(1-2q^2(1+q^2)^{-1}-q^2))$. For convexity, $\frac{\partial^2 f(r)}{\partial r^2} \geq 0$ must hold. Thus it is sufficient to check if $-\alpha_a q^2(1-2q^2(1+q^2)^{-1}-q^2)(2(1+q^2)^{-1}+1)+2+3\alpha_a(1-2q^2(1+q^2)^{-1}-q^2) \geq 0$. This is equivalent to $(1-q^4-2q^2)(-\alpha_a q^2(3+q^2)+3\alpha_a(1+q^2))+2(1+q^2)^2 \geq 0$. Finally, we can see that the sufficient condition for convexity is $\alpha_a q^8+2\alpha_a q^6+(2-4\alpha_a)q^4+(4-6\alpha_a)q^2+2+3\alpha_a \geq 0$. If we take $U=q^2$, convexity of $f(r)$ is equal to: $\alpha_a U^4+2\alpha_a U^3+(2-4\alpha_a)U^2+(4-6\alpha_a)U+2+3\alpha_a \geq 0$. □

If the sufficient condition for convexity holds, we apply Jensen's inequality that leads to inequality (A.2). Using inequality (A.1) and inequality (A.2), we have: $|\mathbb{E}_{\lambda_t,\mathbf{e}^\mathbf{Q}}(\{N_{i,t}\} \setminus \{T_{i,t}\})| \leq \sigma^2 \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| > \alpha_a)\big].(1-\sqrt{\frac{2}{\pi}}\frac{h}{1+h^2}e^{-\frac{h^2}{2}})$

A.2 Proof of lemma 4

According to lemma 2, if $\{i \notin N_t\}$, and measurements are available: $\mathbf{Pr}(\{T_{i,t}\} \setminus \{N_{i,t}\}) = \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a)\big].\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}^{mod}_{i,t}| > \alpha_a) \mid |(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a\big]$. As we discussed in section 4.4, $\mathbf{e}^\mathbf{Q}_{i,t}$ is a Bernoulli random variable, thus, $\mathbb{E}_{\lambda_t,\mathbf{e}^\mathbf{Q}}(\{T_{i,t}\} \setminus \{N_{i,t}\}) \leq \sigma^2 \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a)\big].\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}^{mod}_{i,t}| > \alpha_a)\big]$, where $\mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a)\big]$ is easily computed, as $(\hat{\mathbf{x}}_{t,mod})_i$ is a Gaussian random variable. Since $(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}^{mod}_{i,t}$ is a Gaussian random variable with zero mean and variance $4\mathbf{P}_{i,t}+\mathbf{D}_{i,t}$, we have $\mathbf{Pr}\big[|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}^{mod}_{i,t}| > \alpha_a\big] = 2\mathbf{Pr}\big[(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}^{mod}_{i,t} > \alpha_a\big] = 2\mathcal{Q}(q(\mathbf{P}_{i,t},\mathbf{D}_{i,t})) = \frac{2}{\sqrt{2\pi}}\int_q^\infty e^{-\frac{s^2}{2}}ds$. The following lemma provides an upper bound for $\mathbb{E}_{\lambda_t,\mathbf{e}^\mathbf{Q}}\mathbf{Pr}\big[|(\hat{\mathbf{x}}_{t,mod})_i - \mathbf{e}^{mod}_{i,t}| > \alpha_a\big]$.

Lemma 7. *If there exists an α_a and $q = \frac{\alpha_a}{(4\mathbf{P}_{i,t}+\mathbf{D}_{i,t})^{\frac{1}{2}}}$, such that $q^2 < 3$, we have:*

$$\mathbb{E}_{\lambda_t,\mathbf{e}^\mathbf{Q}}\int_q^\infty e^{-\frac{s^2}{2}}ds \leq \int_h^\infty e^{-\frac{s^2}{2}}ds = \int_{\frac{\alpha_a}{\sqrt{4\mathbb{E}_{\lambda_t,\mathbf{e}^\mathbf{Q}}\mathbf{P}_{i,t|t-1}+\mathbf{D}_i}}}^\infty e^{-\frac{s^2}{2}}ds. \tag{A.7}$$

Proof. we first prove that $f(\mathbf{P}_{i,t}) = \int_q^\infty e^{-\frac{s^2}{2}}ds$ is a concave function. Similar to Appendix A,

let us consider $r = \mathbf{P}_{i,t}$ and $d = \mathbf{D}_{i,t}$. Thus, $f(r) = \int_{\frac{\alpha_a}{(4r+d)^{\frac{1}{2}}}}^{\infty} e^{-\frac{s^2}{2}} ds$. From the Leibniz's rule, we have: $\frac{\partial f}{\partial r} = 2\alpha_a(4r+d)^{-\frac{3}{2}} e^{-\frac{\alpha_a^2}{2(4r+d)}}$, and $\frac{\partial^2 f}{\partial r^2} = (-12\alpha_a(4r+d)^{-\frac{5}{2}} + 4\alpha_a^3(4r+d)^{-\frac{7}{2}}) e^{-\frac{\alpha_a^2}{2(4r+d)}}$. If the $\frac{\partial^2 f}{\partial r^2} < 0$, $f(r)$ is a concave function. Therefore, the sufficient condition for concavity of $f(r)$ is $\frac{\alpha_a^2}{4r+d} < 3$. Jensen's inequality for concave functions implies $\mathbb{E}\big[f(r)\big] \leq f(\mathbb{E}\big[r\big])$. Therefore, $\mathbb{E}_{\lambda_t, \mathbf{e}^{\mathbf{Q}}} \int_q^{\infty} e^{-\frac{s^2}{2}} ds \leq \int_{\sqrt{4\mathbb{E}_{\lambda_t, \mathbf{e}^{\mathbf{Q}}} \mathbf{P}_{i,t|t-1} + \mathbf{D}_i}}^{\infty} e^{-\frac{s^2}{2}} ds$. □

Using the result of lemma 7, we have:
$\mathbb{E}_{\lambda_t, \mathbf{e}^{\mathbf{Q}}}(\mathbf{e}^{\mathbf{Q}}{}_{i,t}) \leq \sqrt{\frac{2}{\pi}} \sigma^2 \mathbf{Pr}\big[(|(\hat{\mathbf{x}}_{t,mod})_i| < \alpha_a)\big] . \int_h^{\infty} e^{-\frac{s^2}{2}} ds.$