COMPRESSIVE POWER SYSTEMS: APPLICATIONS OF COMPRESSIVE SENSING AND SPARSE RECOVERY IN THE ANALYSIS OF SMART POWER GRIDS

by:
Mohammad Babakmehr
A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Electrical Engineering).

Golden, Colorado
Date ________________

Signed: _________________________
Mohammad Babakmehr

Signed: _________________________
Marcelo G. Simões
Thesis Advisor

Golden, Colorado
Date ________________

Signed: _________________________
Prof. Atef Z. Elsherbeni
Professor and Head
Department of Electrical Engineering and Computer Science, Division Electrical Engineering
ABSTRACT

During the last two decades, an intelligence revolution intensively changed the technology of electrical power networks, forming a new generation of power systems called smart grids (SGs). In general, “the term SG refers to electricity networks that can intelligently integrate the behavior and actions of all parameters and users connected to them”. This revolution transformed the traditional power grid from a single-layer physical system into a huge cyber-physical network, using a layer of information that flows through the system. This information includes the status of several parameters in the network such as bus voltages, branch currents, and load consumption. In addition to the actuating or controlling commands fed back to the network from controllers and decision making units. Nowadays, the Supervisory Control and Data Acquisition (SCADA) system in addition to the Wide Area Mentoring System (WAMS) provide electrical data for each local system in near real time. The most popular sensing technology used widely in SG data collection systems is the high sampling rate synchronous Phasor Measurement Unit (PMU). Nevertheless, collecting, storing, transferring, and analyzing the huge amount of data flowing through the information layer of the SG, together with the uncertainty caused by renewable-based distributed generators and unpredictable load characteristics, challenge the standard methods for security, monitoring, and control.

In this thesis, we aim to exploit the inherent sparse nature of both structure and data in SGs to introduce new, fast and reliable techniques to address the challenges related to real time data analysis, monitoring and security in smart power systems. Our work is primarily inspired by a new paradigm in the field of signal processing widely known as the theory of Compressive Sensing and Sparse Recovery (CS-SR). Generally, CS-SR implies that a sufficiently sparse phenomenon can be recovered from a small set of randomly collected measurements. In our early chapters, combining the sparse sampling theory from the field of CS with concepts borrowed from graph theory, we introduce a set of sparse recovery-based mathematical formulations to address famous global monitoring challenges such as power line outage localization, network topology identification, network dynamic behavior modeling and tracking. Lastly, we develop a modified sparse representation-based classification approach to deal with a challenging local monitoring problem widely known as power quality events recognition.
TABLE OF CONTENTS

ABSTRACT ................................................................................................................................... iv
LIST OF FIGURES AND TABLES .............................................................................................. ix
ACKNOWLEDGMENTS ............................................................................................................ xii
DEDICATION ............................................................................................................................. xiv

CHAPTER 1  INTRODUCTION ....................................................................................................1
  1.1 From Power Networks to Smart Grids: Challenges vs. Opportunities .........................1
  1.2 Full Observability from Partial Information ..................................................................3
  1.3 Complex Structure but Sparse Nature of Power Systems .............................................5
  1.4 Overview and Contributions .........................................................................................6

CHAPTER 2  BACKGROUND ....................................................................................................12
  2.1 General Mathematical Preliminaries ..........................................................................12
    2.1.1 Notations ................................................................................................................12
    2.1.2 Linear Algebra ........................................................................................................12
  2.2 Power Engineering and Graph Theory Preliminaries ....................................................14
    2.2.1 Notations ................................................................................................................14
    2.2.2 Power Network Model: Corresponding Graph and Matrices ...............................14
    2.2.3 Corresponding Matrices of a Graph ....................................................................14
    2.2.4 DC Power Flow Model and Corresponding Graph of Power Network ............17
  2.3 Fundamentals of Compressive Sensing and Sparse Recovery ......................................18
    2.3.1 Sparse Recovery Problem ....................................................................................18
    2.3.2 Theorems and Definitions ....................................................................................18
CHAPTER 3  SPARSE RECOVERY-BASED POWER LINE OUTAGE IDENTIFICATION.........................................................24

3.1 Introduction and Literature Review .................................................24
3.2 Mathematical Formulation of POI-SRP..............................................26

3.3 Exploiting Sparsity in Line Outage Identification: Challenges and Solutions ..........29

3.3.1 POI-SRP in Case of Limited Number of PMUs ..................................29
3.3.2 Signal Dynamic Range Issue and Binary POI-SRP ................................31
3.3.3 Sensing Matrix Coherence: Challenges and Solutions ............................32
3.3.4 Support Selection Modification: BLO-based Techniques for Coherent POI-SRP .................................................................34
3.3.5 Structured POI-SRP and COMP ..................................................36
3.3.6 An Axillary PMU Placement Criterion ............................................37

3.4 Simulations and Discussion ..................................................................38

CHAPTER 4  CS-BASED SMART GRID TOPOLOGY IDENTIFICATION .........45

4.1 Introduction and Literature Review ......................................................45
4.2 Compressive Topology Identification ...................................................47

4.2.1 Gaussian Distribution of Parameters ............................................47
4.2.2 Sparse Setup of Topology Identification Problem ...............................48

4.3 Exploiting Sparsity in Compressive Topology Identification .....................51

4.3.1 Exploiting Special Sparse Structures in Nodal-Admittance Matrices .........51
4.3.2 Reweighted $l_1$-Minimization Based CPN-TI ...................................52
4.3.3 Noisy Reweighted $l_1$-Minimization Based PNTI ..............................53
4.3.4 Data Correlation Issue ....................................................................54

4.4 Simulations and Discussion ..................................................................54
CHAPTER 5 MODELING AND MONITORING THE DYNAMIC BEHAVIOR OF SMART GRIDS USING STRUCTURED SPARSITY ............................................... 71

5.1 Introduction and Literature Review ................................................................. 71

5.2 Power Network Modeling and Current wave Dynamic Analysis ...................... 73
  5.2.1 Traveling Wave Analysis in Transmission Lines ....................................... 73
  5.2.2 Power Network Modeling under Dynamic Behavior ............................. 78
  5.2.3 Mathematical Formulation of SGDBM .................................................... 79

5.3 Structured Sparse Dynamic Modeling .............................................................. 81
  5.3.1 Sparse Dynamic Modeling ....................................................................... 81
  5.3.2 Block OMP Recovery Algorithm for Block Sparse Signals ...................... 82
  5.3.3 BOMP Guarantee Conditions .................................................................. 82
  5.3.4 Line Dynamic Order Variation and Clustered Sparsity ......................... 84
  5.3.5 Data Correlation Issue ............................................................................. 84
  5.3.6 Band-excluded Locally Optimized BOMP (BLOBOMP) ....................... 85

5.4 Simulations and Discussion ............................................................................ 86

CHAPTER 6 SPARSE REPRESENTATION-BASED PQ EVENTS CLASSIFICATION ...... 91

6.1 Introduction and Literature Review .................................................................. 91
  6.1.1 Literature Review and State of the Art Approaches ............................... 92
  6.1.2 Technical Challenges .............................................................................. 95
  6.1.3 Contributions ......................................................................................... 97
  6.1.3 Related Works ....................................................................................... 98

6.2 Compressive-Informative Power Quality Events Classification .................... 99
6.2.1 Mathematical symbols and notations for PQC ......................................................99

6.2.2 PQC: Problem Definition .....................................................................................100

6.3 Methodology.............................................................................................................100

6.3.1 Interpretation of the Sparse Representation-based PQ Events Classification .....104

6.3.2 Feature Extraction and Dimensionality Reduction for SRC of PQ Events .........105

6.3.3 Informative Data Points Selection .......................................................................105

6.4 Simulations and Discussion .....................................................................................108

CHAPTER 7 CONCLUSIONS AND FUTURE RESEARCH DIRECTION ....................119

REFERENCES CITED ...........................................................................................................121
LIST OF FIGURES

Figure 1.1 From Traditional Electrical Grid to Smart Grid .........................................................2

Figure 2.1 Corresponding graph of IEEE Standard 30 BUS ..........................................................15

Figure 2.2 (a) Degree, (b) Incidence, (c) Laplacian and (d) Weighted Laplacian matrices for the corresponding graph of IEEE Standard 118 Bus ..................................................16

Figure 2.3 (a) 6 sparse signal, (b) (25,5)-block-sparse signal, (c) (21,5)-clustered-sparse signal ..........................................................................................................................20

Figure 3.1 (a)-(b) Sensing matrix $A$ and $A_{\text{svd}}$ of the IEEE BUS-118. (c)-(d) the corresponding Grammian matrices ..........................................................................................................................34

Figure 3.2 Comparison of recovery performance of BPOI-BLOOMP, BPOI-OMP, POI-OMP and POI-BLOOMP approaches, for (a) IEEE 300-BUS PN in case of 1000 randomly selected 20 simultaneous line outages, and (b) IEEE 2736-BUS system in case of 1000 randomly selected 50 simultaneous line outages vs noise level, respectively .............................................................................................................39

Figure 3.3 Comparison of recovery performance of BPOD-BLOOMP, BPOD-OMP, POD-OMP and POD-BLOOMP approaches, for (a) fully measurable 118-BUS IEEE PN, (b) partially measurable PN, in case of 1000 randomly selected 10 simultaneous line outages, vs noise level, respectively ........................................41

Figure 3.4 Comparison of recovery performance of (a) OMP vs (b) MCOMP method for a sample (15,3)-C-SOV in IEEE 118-BUS a-b..............................................................................42

Figure 3.5 Comparison of the recovery performance of (a) OMP vs (b) MCOMP vs (c) BLOCOMP methods for a sample (33,10)-C-SOV in IEEE 118-BUS.1.1 ........43

Figure 4.1 Corresponding normalized Laplacian matrices of (a) IEEE Standard-14 Bus and (b) IEEE Standard-30 Bus, (c) IEEE Standard-57 Bus and, and (d) IEEE standard IEEE 118 Bus ..........................................................................................................................48

Figure 4.2 (a) The average value of Coherence $\mu_A$ vs # of measurements $M$ over all 118 corresponding sensing matrices $'A'$ for all the nodes in the IEEE 118-Bus over 100 realizations of the network. (b) Average coherence versus number of measurements over all corresponding sensing matrices $A_i$ of all the nodes in IEEE Standard-30 Bus system, where the curve is averaged over 100 realizations of the network ...............................................................................................................55
Figure 4.3  Recovery rate comparison of nodes (a) 3 (in-degree 2), (b) 10 (in-degree 6) in the IEEE standard 30 Bus, respectively. Success rate is calculated over 100 realizations of the network for a given number of measurements ..........................56

Figure 4.4  Recovery rate comparison of nodes (downward): 1 (in-degree 3), 2 (in-degree 4), 15 (in-degree 6), 3 (in-degree 7), 130 (in-degree 9) respectively. Success rate is calculated over 100 realizations of the IEEE Standard-300 Bus for a given number of measurements ........................................................................58

Figure 4.5  Recovery performance for 2 SRP solvers over all of the buses of the standard IEEE Standard-30 Bus: (a) BP, (b) $R_w l_1$, (c) $R_w l_1 - BP$ ........................................59

Figure 4.6  Whole network topology recovery performance for 2 SRP solvers for (a) the IEEE Standard-30 Bus, (b) the IEEE Standard-300 Bus over 100 realizations of the network, and (c) the standard IEEE Standard-2383 Bus over 100 realizations of the network .............................................................................61

Figure 4.7  Recovery rate comparison of (a) node 3 and (b) node 10 in the IEEE 30 bus system. These nodes have in-degree 2 and in-degree 6, respectively .........................63

Figure 4.8  Recovery rate comparison of nodes 74, 59, and 49 in the IEEE 118-Bus, with in-degree 2, 5, and 9 respectively ...........................................................................64

Figure 4.9  Recovery rate comparison of (a) nodes 3 and (b) 1920 in the IEEE standard 2383 network. Nodes 3 and 1920 have in-degree 3 and 9, respectively ......................66

Figure 4.10  Recovery rate comparison of nodes (a) 37 and (b) 80 in the IEEE 118-Bus. An initial value of $m = 4$ is chosen in MCOMP and BLOMCOMP ......................67

Figure 4.11  Recovery performance for 3 SRP solvers over all of the buses of the standard IEEE 118 Bus system: (a) OMP, (b) LASSO, (c) MCOMP .....................................68

Figure 4.12  Whole network topology recovery performance for 3 SRP solvers for the standard IEEE 118 Bus system over 1000 realizations of the network ....................70

Figure 5.1  (a) Single-phase two-wire lossless transmission line with source and load terminals. (b) A Section of length $\Delta x$ of the line .................................................74

Figure 5.2  A typical super-node/bus model under PN KCL law .................................................................78

Figure 5.3  Recovery performance of the BOMP for buses: 74, degree 2, 49, degree 9, and 59, degree 6, within the IEEE-118 Bus network (for LDO = 4) over 100 realization of the network ........................................................................87

Figure 5.4  Coherence measures of IEEE 30 Bus vs number of measurements (curves are averaged over 100 realizations of the network) ..............................................87
Figure 5.5  Recovery performance for (a), (c) OMP and (b), (d) BOMP over all of the 
buses of the standard IEEE 30 Bus for LDO = 2 and 4, respectively .........................89

Figure 5.6  Recovery performance for (a) OMP and (b) BOMP (c) COMP over all of the 
buses of the IEEE 30 Bus network for Unknown LDOs .............................................90

Figure 6.1  Five basic steps in a PR problem: 1) pre-processing, 2) detection, 3) features 
extraction, 4) feature Selection, and 5) classification ................................................93

Figure 6.2  The role of the Feature Extraction and Feature Selection ...................................97

Figure 6.3  A geometrical comparison between $l_1$-norm vs. $l_2$-norm minimizations. 
Since the $l_2$-ball does not have sharp edges along the axes, the corresponding 
solution to $P_2$ is almost never sparse, while regarding sharp edges along the 
axes of $l_1$-ball we may arrive at the true sparse solution if all other sparse 
recovery conditions are satisfied ...............................................................................103

Figure 6.4  A visualization of a typical SRC procedure for an interruption PQ event ..........104

Figure 6.5  Informative samples (Golden circles) are the corresponding vertices of the 
associated convex hull of each data cluster within the feature space .....................107

Figure 6.6  Distribution of the first 100 dominant SVD coefficients over each of the 14 
classes of PQ events within the corresponding training sub dataset 
$Y_j \in \mathbb{R}^{2560\times 400}$ for $j = 1: 14$. ..................................................................................111

Figure 6.7  (a)-(p), corresponding convex hull approximation error vs number of selected 
informative samples or equivalently convex hull vertices or extreme points ........113
LIST OF TABLES

Table 2.1  Power engineering quantities notation .................................................................14
Table 3.1  Time cost comparison: BLOOMP vs OMP in POI-SRP ....................................41
Table 4.1  Recovery performance of 3 SRP solvers using 30 measurements for the most
          significant clustered sparse buses in the IEEE 118-Bus system ..........................69
Table 6.1  Visualization of the PQ Events Classes .............................................................109
Table 6.2  Classification results of the CIC-PQ (CONFUSIONMATRIX): Descriptions:
          id-c: identified class, a-c: actual class, im: impulse, n: notch, in: interruption,
          sw: swell, s:sag, t:transient oscillatory, f:flicker, h: harmonic................................110
Table 6.3  Classification results of the CIC-PQ (Confusion Matrix) ...............................110
Table 6.4  Selected Number of Informative Data Points for Each PQ Class ......................112
ACKNOWLEDGMENTS

During my Ph.D program, I have been privileged to work with Prof. Marcelo Godoy Simões a wonderful scientist with a great personality. His help and support gave my wife Farnaz and I a unique opportunity to work together during our PhD program at the Colorado School of Mines. The course of ethics I learned from him, however, is by far more valuable for the rest of my life. Not only is he a great advisor but he truly acts as a compassionate father for his students. Dear Marcelo, thanks for your patience, motivation, immense knowledge and all of your kindness.

I was also extremely fortunate to meet, Prof. Michael B. Wakin, whose continued support and presence have significantly inspired me during the past few years. Your outstanding professionalism and exemplary behavior would be a unique gift I can never repay you for; I will be forever thankful for all those Compressive Discussions we had! Furthermore, I would like to thank Prof. Tyrone L. Vincent not only for his insightful comments and encouragement to pursue my research, but also for the hard questions which gave me an incentive to widen my research from various perspectives.

I would also like to express my sincere gratitude to our co-advisors in petroleum institute Dr. Ahmed Al Durra and Dr. S.M Muyeen for their continuous support of my Ph.D study and related research. I should also sincerely appreciate the guidance from Dr. Salam Mohagheghi, Dr. Qi Han and Dr. Raveal Ammerman that assisted me during all my time of research and working on this PhD thesis. Besides them I would like to thank the chair of my thesis committee member at large, Prof. Steffen Rebannak, for his time and support during this project.

I am also very thankful to Dr. Armin Eftekharri, Dr. Luis Tenorio, Dr. Gongguo Tang and Dr. Bernard Bialecki for all the useful mathematical and technical discussions we had. My sincere thanks also goes to many of my colleagues (the new generation of Doctors in Electrical Engineering) Dr. Borhan Sanandaji, Dr. Pirooz Javanbakht, Dr. Hossein Sartipizadeh and, my brother-in-law Dr. Farshad Harirchi. I would like to extend my appreciation to all of my other colleagues and officemates: Bananeh, Abdullah, Moein, Khaled, Hisham, Rammah, Andrew, Dehui, Zihui, Abdulhakeem, Ali, Varaprasad, Ricardo, Sepideh and Alejandro.

Thanks to all my friends: Mahmood, Elham, Naser, Shirin, Neema, Azita, Abbas, Zoheir, Zahra, Kamran, Mona, Sayeh, Mohammad, Ensieh, Soheil, Taraneh, Morteza, Arash all Alis! and many
more for all the thanksgiving and Persian new year gatherings, mountain hiking together, vacations and weekly volleyball events.

I was very honored to be a part of and work as a member of the IEEE Denver Section during past two years. Dear Murali, Jim, Tim, Pankaj Ian, and Marc thanks for all shared memories including the night time Wednesdays and early Saturday morning meetings.

I would like to thank all of my teachers and mentors, from early ages of my elementary school all the way to my graduate school, Amirkabir university of technology, especially, Dr. AhmadAli Ashrafian. Thanks also to my dear bAjenAgh Yasaman, and my relatives, uncle Karim, Mahtab, Ali, Rosa, Yevan, Shahin, Dr. Kenrangi, and lovely Yashar for their great hospitality and kind support which have significantly facilitated our life during our time in the USA.

My deepest gratitude belongs to my parents, Mahroo and Khosro, my parents-in-law, Shadi and Farhad, my lovely sister Sadaf, for their unconditional love and support during my life. I owe them a lot and this thesis is dedicated to all these beloved dears. Finally, to my beloved wife and the best colleague in the world Farnaz: “Thanks for your patience, endless love and support, during the past five years. I could not have imagined having a better companion for my Ph.D study! I don’t know how long I am going to live but I know that every single second will be worth a lifetime because it will be spent with you. I love you”.
This is dedicated to my uncle Karim, may his soul rest in peace
and My family,
Mahroo, Khosro, Shadi, Farhad, and Sadaf
for their boundless support and love

To my beloved companion,
Farnaz
CHAPTER 1
INTRODUCTION

In this chapter, we present an overview on the technology of smart power grids, opportunities and challenges, in addition to the motivation and contributions of our work.

1.1 From Power Networks to Smart Grids: Challenges vs. Opportunities

An intelligence revolution is happening in the technology of power networks (PN), supporting the development of so-called smart grids (SGs). Roughly speaking, “the term SG refers to electricity networks that can intelligently integrate the behavior and actions of all parameters and users connected to them” [1]. This revolution transforms a traditional power grid from a single-layer physical system into a huge cyber-physical network, using a layer of information that flows through the system (Figure 1.1). This information includes the status of several parameters in the network such as bus voltages, active and reactive powers, phase angles, branch currents, and load consumption, in addition to the actuating or controlling commands fed back to the network from controllers and decision making units [2].

Another important feature of next generation SGs is the aggregation of renewable energy micro-sources in distributed generation technology [3]. Despite the environmental advantages bring in by utilizing the renewable sources of energy, the nature of the injected power from renewables as wind turbines and solar panels is inherently random. This randomness, along with the uncertainty in the load, lead to probabilistic behavior in the load-power flow at buses [4]-[5]. Collecting, transferring, and storing the huge amount of data flowing through the information layer of large scale SGs, together with the uncertainty caused by distributed power generators and load characteristics, challenge the standard methods for security, monitoring, and control. Other important issues are the time cost, efficiency and accuracy of real-time data analysis within these huge interconnected and complex networks [6].

Nowadays, the Supervisory Control and Data Acquisition (SCADA) system reports the electrical parameters of buses almost within the range of couple of seconds. Moreover, Wide Area Monitoring System (WAMS) uses the data measured by phasor measurement units (PMUs) to

---

1 In the rest of this thesis we may alternatively use terms power network and smart grid in exchange (or equivalently PN vs SG).
report the electrical parameters at the monitored buses by exploiting a global positioning system (GPS) for each local system in near real time [7]. Currently, the system data exchange (SDX) module of the North American Electric Reliability Corporation (NERC) provides network-wide status on an hourly basis [8], however, to transform the PN into a smart network a near real-time monitoring of system status, in all the operational levels, is mandatory [9].

Figure. 1.1 From Traditional Electrical Grid to Smart Grid [100]
In terms of power engineering technology, we can split the monitoring challenges into two major groups: 1. Global or network-wide level problems, such as network topology identification, power system state estimation, and power line outage localization. 2. Local level or, the signal processing problems, this second category is mostly related to the power signals quality monitoring and analysis, such as nodal load or price estimation, electrical equipment monitoring, voltage and current disturbances detection and classification.

1.2 Full Observability from Partial Information

In general, most of the monitoring approaches, within local operational levels, are highly dependent on the quality of the available data from system parameters and poorly perform in case of corrupted measurements where only partial information may be available [10]. On the other hand, most of the classical global level monitoring methodologies have been adapted in case of simple structures, these methods are usually suffering from high computational cost and nonlinearity in formulation which results in intractable solutions in case of real time applications [11]. These computational issues are specifically highlighted in case of large-scale PNs, where usually a huge amount of data is arrived at the control and operational rooms in every time slot. In many practical cases (considering the time cost, storage capacity limitations and power efficiency), operators may have to intelligently reduce the amount of data needed to be processed and still accurately extract the whole of the useful information for system analysis applications.

These inherent technical challenges motivated us to investigate the applications of Compressive Sensing and Sparse Recovery (CS-SR) in the analysis of smart power networks to deal with the big data analysis issues in SGs.

Compressive sensing is a new technique in signal processing that has revolutionized the data analysis science. Due to its time efficiency, the CS-SR framework has found a critical role in the large scale data/system modeling/analyzing and significantly affected the era of Big Data. From the signal processing viewpoint, CS is a novel sensing/sampling technique that abjures the prevalent wisdom in signal and data acquisition. Based on the Shannon sampling theorem and regarding the Nyquist sampling rule, the minimum sampling frequency (number of samples per cycle) needed for capturing the complete information from a phenomenon should exceed two times of the phenomenon bandwidth (please also refer to [12]). Roughly speaking, the CS theorem tries to answer the following question:
Can a sensor be designed that records fewer samples of a very high dimensional signal $(f \in \mathbb{R}^N)$, while missing information (or samples) can be reconstructed from the collected ones?

The answer is yes if the phenomenon under study has an inherent information level that could be well modelled utilizing only $K$ parameters or degrees of freedom$^2$. A comprehensive study in CS literature has revealed the fact that practically most of the natural and industrial $N$-sample signals have far smaller than $N$ degrees of freedom. This fact suggests that, under appropriate conditions, one may exploit the signal's information level $K$ instead of its ambient dimension $N$, to design an efficient measuring, and modeling framework for high-dimensional signal analysis. In other words, the fundamental idea that underlies CS-SR theorem is that ‘if a signal has an intrinsic information level $K$ it can be possible to record a small number of linear measurements of the signal, $M$, proportional to $K^3$; “compressive sensing”, and from those measurements reconstruct the complete set of all, $N$, samples that a conventional sensor would have recorded, “sparse recovery”’ as stated in [13].

Not only low dimensional behavior can be captured in many real-world signals, but also many system models rely on a low dimensional (sparse) structure. This idea has been further developed and resulted in arising of a new branch in the CS-SR field, widely known as Compressive System Identification (CSI), as termed in [14], where an inherent sparsity can be observed in the system structure/model, or when a typical problem represents a sparse behavior in its formulation. In compressive systems, the sparse recovery techniques can be utilized to address the large-scale system modeling problems in a fast and time efficient manner. A famous class of CSI problems is related to graph structures analysis.

Intuitively, a sparse recovery problem (SRP) is an optimization problem in which the goal is to recover a $K$-sparse signal $\mathbf{x} \in \mathbb{R}^N$ from a set of observations $\mathbf{y} = \mathbf{A} \mathbf{x} \in \mathbb{R}^M$ where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the sensing protocol (matrix) with $M < N$ (in many cases $M << N$) [15]. Due to the underdetermined nature of this recovery problem (since $M < N$), the null space of the matrix $\mathbf{A}$ is non-trivial; therefore, there exist infinitely many candidate solutions for this problem. However, under certain conditions on the sensing matrix $\mathbf{A}$, various sparsity based recovery methods can be guaranteed to

---

$^2$ In general, the number of $K$ can be much fewer than $N$.

$^3$ A $K$-sparse signal $\mathbf{x} \in \mathbb{R}^N$ is a signal of length $N$ with $K$ nonzero entries where $K < N$ (in many cases $(K << N)$). The sparsity level of a signal $\mathbf{x}$ is denoted by the $l_0$ norm $||\mathbf{x}||_0$. 
efficiently find the candidate solution that is sufficiently sparse. Examples of such conditions are the restricted isometry property (RIP), the exact recovery condition, and low coherence \[16\]. In essence, what these conditions require is that any two small subsets of columns of the sensing matrix \( \mathbf{A} \) must be almost orthogonal to each other. In the second chapter of this thesis we present an introduction to the CS-SR definitions, theorems and concepts.

1.3 Complex Structure but Sparse Nature of Power Systems

At first glance, most of the power networks, either in the transmission or in the distribution level, may look like a huge interconnected system with a high level of complexity and extensive number of electrical parameters. However, most of these systems are less complicated than what is suggested. For example, in terms of transmission networks, one may easily figure out most of the standard PN models are forming a sparse interconnected graph. This sparsity in the structure helps us to reformulate the power network topology identification as a sparse recovery problem that can be solved using a small set of measurements in a fast and accurate way using sparse recovery solvers (Chapter.4). This sparsity assumption can be proved based on our observations on a survey of standard power grid models found on databases and toolboxes such as MATPOWER \[17\]. Considering a variety of standard PN models, it can be observed that the regular maximum connectivity level of an electrical bus in a grid is typically less than 5-10\%, especially in case of large-scale PNs \[18\]. As suggested in \[18\], the same fact can be, more or less, observed in case of distribution networks as well.

Now we discuss another possible sparse monitoring problem where the fact of sparsity is not easily observable from the beginning. In the network-wide operational level of SGs, fast and reliable localization of power line outages is categorized on top of the critical tasks. Power line outages can result in heavy damages such as blackouts \[19\], while massive blackouts can be prevented if we have a comprehensive situational awareness of the structure of the power grid. Initially, one may not be able to figure out a sparsity in the power line outage detection problem, however, there is an inherent sparsity pattern in this problem, that is in the most of the practical scenarios the number of simultaneous outages may happened within a power network is a very small portion of the total number of active transmission lines \[39\]. Consider a PN with a set of \( L \) transmission lines, assume a set \( \zeta_{out} \) of \( L_0 \) lines have been damaged within a failure phenomenon in the network. Assign an outage index to each line and collect them in an outage vector \( \mathbf{x} \in \mathbb{R}^L \)
where \( x_i \) for \( i = 1:L \) is equal to 1 if \( i \in \zeta_{out} \) and 0 otherwise. Obviously, the corresponding outage vector \( \mathbf{x} \) would have a sparse structure with most of its elements equal to zero except for those on the corresponding outage locations. Thus, one may think of developing a possible formulation where the outage vector \( \mathbf{x} \) can be recognized using a set of measurements from network parameters (Chapter 3). A variety of other possible sparse recovery-based applications may also be revealed in the global monitoring level problems by manipulating the mathematical formulation of power systems models.

On the other hand, in the local monitoring level, we majorly face with the power signal analysis problems for sake of stability and protection. Naturally, most of the electrical signals (such as currents and voltages) are traveling through the power systems in the shape of a single-frequency sinusoidal wave. Due to system variations, such as load uncertainties, this fundamental sinusoidal shape may temporarily (such as impulsive events during arcs) or steadily (such as long term harmonic integrations) change during the time. However, it can easily be shown that under almost all operational conditions a typical time segment of a power signal can still be sparsely represented in terms of another signal domain (such as frequency domain). As a result, one may exploit this natural sparsity to address the power signals analysis in terms of a CS-SR problem to reduce the time cost and improve the system reliability for online applications.

1.4 Overview and Contributions

In this thesis, we aim to develop new analytical methods and problem formulations to exploit the underlying sparsity in the monitoring and analysis of large-scale smart grids. We begin in Chapter 2 by introducing our mathematical preliminaries in addition to specific power engineering area abbreviations and notations, followed by an introduction to CS-SR and the graph theory concepts.

In the following three chapters (Chapter 3-5), we particularly focus on the following three global level monitoring and security challenges:

- Power line outage detection and localization
- Power network topology identification
- Network-wide modeling and tracking the dynamical behavior of PNs

In addition to developing sparse recovery-based mathematical formulations, we will discuss the corresponding technical issues in solving these security problems (from both power engineering
and CS-SR viewpoints) and adapt couple of alternative sparse recovery techniques to address the inherent challenges regarding our sparse-based formulations.

In Chapter 3 we consider the problem of power line outage localization from a set of measurements of voltage phasor angles. We represent the PN as a graph $G(S_N, S_E)$, consisting of a set of $N$ nodes $S_N = \{1, \ldots, N\}$, where each node $i$ represents an arbitrary bus of the SG, and a set of $L$ edges or transmission lines $S_E = \{l_{ij}; i,j \in S_N\}$, which connect certain nodes and form the general structure or topology of the grid. We assume that the data from the nodes’ parameters (such as real power and phase angles) are obtained from a network of PMU sensors.

Inspired by the idea originated in [39], incorporating the linear DC power flow model\(^4\) (Section 2.2.4), $\mathbf{p} = B\mathbf{\theta} \in \mathbb{R}^N$, into some linear algebraic matrices properties we model a failure event (with couple of damaged lines forming an outage set $\zeta_{out}$) in terms of a set of linear systems of equations termed as Power Outage Identification-Sparse Recovery Problem (POI-SRP), such that in the proposed POI-SRP formulation we have: $\mathbf{y} = A\mathbf{x} + \mathbf{n}$, where $\mathbf{y} \in \mathbb{R}^N$ is termed the measurement vector and, the matrix $A \in \mathbb{R}^{N \times L}$ is termed the sensing matrix. both parameters can be obtained from a set of measurements from voltage phasor angles $\mathbf{\theta}$ for electrical buses in addition to an initial information of pre-failure structural matrices such as nodal-admittance matrix, $B$, and incidence matrix, $H \in \mathbb{R}^{N \times L}$, of the corresponding PN. The resulting vector $\mathbf{x} \in \mathbb{R}^L$ can be considered as a sparse vector\(^5\) with most of its elements equal to zero except on the corresponding $L_{out}$ outage positions. We term this vector as the Sparse Outage Vector (SOV) with the following mathematical definition: $\mathbf{x} = \begin{cases} x_l & \text{if } l \in \zeta_{out} \\ 0 & \text{otherwise} \end{cases}$. Clearly, recovering such an SOV, $\mathbf{x}$, one can easily obtain the outage lines index set $\zeta_{out}$ by looking at the nonzero positions. Due to sparse nature of this recovery problem, sparse recovery solvers may be exploited to correctly recover the outage vector $\mathbf{x}$ from $\mathbf{y} = A\mathbf{x} + \mathbf{n}$ . However, as we further investigate in Chapter 3, there is a variety of challenges (regarding sparse recovery guarantee conditions) which may prevent us from ending up with the correct solution. These problems include: high coherence of the sensing matrix $A$, number of available measurements from the nodes, and high signal dynamic range\(^6\). To this end,

---

\(^4\) This DC power flow model is a popular mathematical system of equations which has been widely used in the fundamental mathematical formulation of the most of the POI methods developed in the state of the art. In this linear system of power flow equations, $\mathbf{\theta} \in \mathbb{R}^N$ stand for vectors formed by the active power and the voltage phase angle values of all the electrical buses, respectively, and $B \in \mathbb{R}^{N \times N}$ is called the nodal-admittance matrix of the PN.

\(^5\) Also widely termed as an “$l_{out}$-sparse signal” in compressive sensing and sparse recovery literature, where $l_{out}$ indicates the number of nonzero elements in the corresponding sparse signal $\mathbf{x}$.

\(^6\) By definition, for a signal $\mathbf{x} = [x_1, \ldots, x_L] \in \mathbb{R}^L$, the dynamic range is defined as follows:
we introduce alternative formulation named Binary POI-SRP where the SOV, $x$, is changed into a binary sparse vector, $x_b$, where $x_b = \begin{cases} 1 & \text{if } \ell \notin \text{out} \\ 0 & \text{otherwise} \end{cases}$. This binary representation helps us to address the signal dynamic range issue. Another important factor in multiple line outage scenarios is the presence of a possible structure in the pattern of outages. We discuss that in such a situation, one can remodel the POI-SRP as a Structured-POI-SRP and solve a new problem which uses the structure assumption on the outages distribution. In this work, we use the idea of clustered sparsity [20] and modify the Clustered OMP\textsuperscript{7} algorithm [21] to solve the Structured-POI-SRP. Moreover, couple of alternative approaches are investigated to diminish the high coherence issue, such as matrix decomposition substitution or applying Band-excluded Locally Optimized SRP solvers such as BLOOMP [22]. Finally, we discuss how the technical challenges regarding limited number of available PMUs can be addressed within the sparse-based POI framework and how this sparse-based formulation can be used as an auxiliary PMU placement approach in network observability analysis.

In Chapter 4, we would go one step ahead and try to generalize the POI problem where we address a dynamic framework for the whole PN topology identification (PNTI) instead of focusing on outage lines localization. In another word, PNTI problem can be interpreted as the generalized version of POI problem, where the goal is to reveal all changes in the structure of the grid caused by outages, unreported switching phenomena and, so on. Most of the state of the art in POI have been developed based on the DC power flow model and similar to our approach in Chapter 3 require hourly basecase grid topology information (such as nodal-admittance matrix $B$) in addition to the measurements from system parameters to identify the location of outage lines. However, a notable issue in such formulations is the rank deficiency of the matrix $B$ [23]. To avoid the problems arising from this rank deficiency, one bus is normally considered as the reference bus and its corresponding row and column are removed from the matrix $B$, so the matrix becomes full rank. However, due to cyber-attacks and other precision deficiency issues, whenever the records from the reference bus are affected by the bad data, the overall procedure can be corrupted. On the other hand, the mathematical formulation of most of these approaches is dependent on the

\[ \text{Signal}_{\text{dynamic-range}} = \frac{\max(x_i)}{\min(x_i)} \quad \text{and} \quad x_i \neq 0 \quad \text{for} \quad i = 1:L. \]

\text{It has been discussed in the literature [22] that sparse signals with smaller dynamic ranges are easier to be recognized using any sparse recovery solver.}

\text{\textsuperscript{7} Orthogonal Matching Pursuit (OMP) is a famous and popular sparse recovery solver, which is widely used in many sparse recovery problems.}
connectivity of the power grid graph, where it is assumed that the power outages may not be allowed to result in an islanding situation in any part of the network.

In this chapter, we introduce a general and robust topology identification approach, in that the presented reformulation relies only on the measurements from system parameters and does not need any a priori information about the topology of the network. Although we use the DC power flow model as the basis of our mathematical formulation, we alternatively reformulate the PNTI problem in such a way that the output of the optimization problem is the structure of the topology matrix $B$. As a result, this framework is capable of dealing with the aforementioned imperfections while no connectivity assumption is needed over the network structure, this is an essential factor in case of natural disasters where we may face with partial islanding in different areas in the PN.

In our formulation, utilizing the DC power flow model we develop a set of linear equations such that $y_i = A_i x_i + n$, where the columns of matrix $A_i$ is formed by concatenating the difference of phase angles between bus $i$ vs all other buses in the network across $M$ samples of time. The measurement vector $y_i$ includes the active power measurements of $i^{th}$ bus within time series of $M$ sample times and the sparse vector $x_i$ represent the $i^{th}$ column of the matrix $B$. Regarding sparse connectivity pattern in PNs, each column of the matrix $B$, indicates by a sparse vector $x_i$, with all its elements equal to zero except a small portion which are in those positions corresponding the (few) neighbors of node $i$. Solving this SRP with respect to entire nodes of the PN we can form the new structure of the matrix $B$, we call this formulation as PNTI-SRP. We also discuss the role of the data correlation and inherent challenges in PNTI-SRP and develop a generalized BLO-based algorithm termed as Bound-excluded Locally Optimized COMP (BLOCOMP) to deal with this issue. Due to special structure of the corresponding sparse signals to be recovered in PNTI-SRP, we apply the COMP and re-weighted $l_1$ minimization approaches to improve the final recovery performance using even a smaller set of measurements.

Chapter 5 contains another major contribution of this work. Within this chapter, we introduce a new general framework for modeling the dynamic behavior of the entire PN, using the analysis of current waves propagation. Inspired by the same voltage wave propagation analysis in [24], the current waves are mathematically formulated using partial differential equations and the lumped model of the transmission lines. Once a current wave propagates in a transmission line, some small amplitude delayed copies of the original wave arise and move along the line due to reflections and line capacitive-inductive properties. These small waves are damped quickly for each sample of the
time; however, their effect is always presented in several future time samples, and can be captured using high sampling rate PMUs. Incorporating a new port-Hamiltonian model of the power network [25] with this wave propagation modeling, a set of new dynamic parameters are introduced and termed as the dynamic coefficients. We interpret the PN dynamic behavior modeling as a real-time estimation of these coefficients. We describe how this problem can be formulated as a block-structured sparse recovery problem and is solved using modified block-sparse recovery algorithms such as Block-OMP or COMP. We also suggest another alternative structured-based sparse recovery method termed as Bound-excluded Locally Optimized Block-OMP (BLOBOMP) to address the corresponding data correlation issue.

In Chapter 6, we consider a famous local level monitoring problem widely known as power quality events recognition. Roughly speaking, any electrical signal shape deviation from a pure sinusoidal is termed as a power quality (PQ) event [74]. Due to extensive increase in sensitive and critical loads usage (such as digital economy, continuous process manufacturing industries and fabrication and, essential services), the PQ events classification is defined as a critical task beyond the security assessments of future power systems. Like most of the usual pattern recognition (PR) problems, the state of the art in this area can be summarized in terms of the following 5-step procedure [27]:

- Data Preprocessing
- Potential Pattern Detection
- Feature Extraction (FE)
- Feature Selection (FS)
- Clustering/Classification

A variety of technical challenges are appeared within optimal stepwise PR framework selection, for example an optimal FE-FS may result in a widely separated feature space for different PQ event classes so a simple classifier may perfectly distinguish the classes. However, if an optimal FE-FS procedure is not achievable, complex and usually nonlinear classification approaches such as artificial neural networks or model-based classifiers are needed to learn the nonlinear and fuzzy classes boundaries within the corresponding feature space. Beside these general difficulties couple of specified challenges are assigned to this PR problem such as data measuring, storing and transferring protocol, data processing and analyzing time cost, measurement noise, uncertainty
injected by renewable generation sources (such as wind turbines or photovoltaic cells), and finally, multiple events or overlapped/simultaneous events.

In this chapter, we investigate the possibility of PQ events classification using a modified version of the sparse representation-based classification (SRC) approach [28], where despite usual classifiers no training procedure is needed. Moreover, we study how a random feature selection can be applied to decrease the feature space dimension while preserve and even improve the classification accuracy. On the other hand, we describe how selecting an optimal set of training samples can be interpreted as a high dimensional convex hull approximation. This latter method is used to find the most important training samples; we term as informative samples. We show how a small set of the informative samples from each class can be exploited to decrease the data space dimension while preserve the classification accuracy. We discuss how the overall proposed framework could address a variety of the imperfections within the previous approaches in terms of the PQ classification problem. We end up with our final discussion and directions for the future research in Chapter 7. The major contribution of this thesis has been published or is under preparation within the following set of articles: [26], [30], [31], [94]-[98].
CHAPTER 2
BACKGROUND

In this chapter, first we briefly describe the mathematical tools that we utilize in the rest of this thesis. Next, we present a short overview on the useful and basic concepts that we borrowed from the field of power engineering and graph theory followed by an introduction to the related CS-SR concepts, fundamental theories and techniques to be used throughout the rest of this work.

2.1 General Mathematical Preliminaries

2.1.1 Notations

The letters $C_1, C_2, ...$ are reserved in this work to represent universal positive constants. All the vectors will be distinguished by small letters but boldface $\mathbf{x}$. Matrices will be defined by capital letters such as $B$ for nodal-admittance or power network structure matrix. Parameter “$\varepsilon$” stands for a very small constant value, and “$\eta$” indicates the noise related parameters except otherwise mentioned. Here $\mathbb{C}^N$ and $\mathbb{R}^N$ stands for the set of complex and real numbers in $N$ dimensional space, respectively. Moreover, “$\cdot^T$” operator, stands for either matrix or vector transpose. We also reserve $\beta(\cdot)$ to represent the coherence band related concepts.

The $i^{th}$ entry of the vector $\mathbf{x}$ is denoted by $x_i$ or $\mathbf{x}[i]$. Similarly, the $(i,j)^{th}$ entry of the matrix $B$ is denoted either by $B_{i,j}$ or $B_{ij}$. Unless otherwise noted, we would deal with the real-valued, discrete and finite domain signals throughout the rest of this thesis. Thus, by default, we model our signals as a vector in $\mathbb{R}^N$ and occasionally use the terms signal and vector interchangeably.

2.1.2 Linear Algebra

- $l_p$ Norms

For a vector $\mathbf{x} \in \mathbb{C}^N$, the $l_p$ norm of $\mathbf{x}$ is defined as follows:

$$
\|\mathbf{x}\|_{l_p(\mathbb{C}^N)} = \begin{cases} 
(\sum_i |x_i|^p)^{1/p}, & 0 < p < \infty \\
\max_i |x_i|, & p = \infty \\
\sum_i 1_{x_i \neq 0}, & p = 0
\end{cases}
$$

(2.1)
Here $x_i$ is the $i^{th}$ entry of the vector $\mathbf{x}$ and, $\mathbf{1}$ denotes the indicator function. Without loss of generality we simplify the norm notation from $\|x\|_{p(\mathbb{C}^N)}$ to $\|x\|_p$. Actually, for $0 < p < 1$, the above measures are not defined as norm operators. Moreover, $l_0$ norm counts the number of non-zero elements for a vector $\mathbf{x} \in \mathbb{C}^N$. This $l_0$ norm will play an important role in compressive sensing and sparse recovery theorems and concepts.

- Matrix Decomposition: Singular Value Decomposition vs QR Decomposition.

Consider $A$ as a $m \times n$ matrix whose elements come from either the field of real numbers or complex numbers. Then there exists a factorization of the form

$$A = U\Sigma V^H$$

(2.2)

where $U$ is an $m \times m$ unitary matrix over field $K$ (orthogonal matrix if $K = \mathbb{R}$), $\Sigma$ is a $m \times n$ diagonal matrix with non-negative real numbers on its main diagonal, and the $n \times n$ unitary matrix $V^H$ denotes the conjugate transpose (Hermitian) of the $n \times n$ unitary matrix $V$. Such a factorization is called a singular value decomposition of $A$, also widely known as SVD decomposition. Most notable roles of the SVD decomposition are in solving least square problems, dealing with ill-conditioned systems, pattern recognition and system identification.

In linear algebra, a QR decomposition represents an $m \times n$ matrix $A$ in terms of a product of two other metrics $Q$ and $R$ such that $A = QR$, where $Q$ is an $m \times m$ unitary matrix, and $R$ is an upper triangular $m \times n$ matrix. QR decomposition is often used to solve the linear least squares problem. As it has been discussed in the literature, the fact that, each column $k$ of the original matrix $A$ only depends on the first $k$ columns of the unitary matrix $Q$, forces the triangular format of the matrix $R$. It has been shown in the literature that within a similar approach, one may alternatively define the QL, RQ, and LQ decompositions, with $L$ to be formatted as a lower triangular matrix.

- Grammian Matrix

In linear algebra, for a set of vectors $\alpha_1, ..., \alpha_n$ lying over an inner product space the corresponding Grammian matrix is defined as the Hermitian matrix of the inner products, with
entries equal to $G_{ij} = < a_i, a_j >$. In case of finite-dimensional real vectors, defining a matrix $A$ whose columns are the vectors $a_k$ the Grammian matrix is simply defined as $G = A^T A$.

2.2 Power Engineering and Graph Theory Preliminaries

2.2.1 Notations

The following capital letters, indicated in the Table 2.1, are used in this work to present the power signals and quantities.

<table>
<thead>
<tr>
<th>Table 2.1. Power engineering quantities notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Power</td>
</tr>
<tr>
<td>Reactive Power</td>
</tr>
<tr>
<td>Nodal-Admittance Matrix</td>
</tr>
<tr>
<td>Electrical Voltage Phasor Angle</td>
</tr>
<tr>
<td>Electrical Voltage</td>
</tr>
<tr>
<td>Electrical Current</td>
</tr>
</tbody>
</table>

2.2.2 Power Network Model: Corresponding Graph and Matrices

In the rest of this thesis a PN is modelled as a graph $G(S_N, S_E)$, consisting of a set of $N$ nodes $S_N = \{1, \ldots, N\}$, where each node $i$ represents an arbitrary bus of the SG, and a set of $L$ edges or transmission lines $S_E = \{l_{i,j}: i, j \in S_N\}$, which connect certain nodes and form the general structure or topology of the grid. As an example, Figure 2.1 illustrates the structure of the corresponding graph of the IEEE Standard-30 Bus network. This graph representation and its corresponding matrices are playing an important role in our sparse-based mathematical formulations.

2.2.3 Corresponding Matrices of a Graph

For a graph with a set of $N$ nodes (buses) $S_N$ and a set of $L$ edges (Lines), $S_E$, $G(S_N, S_E)$, a set of four important structural matrices are defined as follows [29]: 1) Adjacency matrix, $\text{Adj}(G(S_N, S_E))$, is an $N \times N$ matrix that represents\textsuperscript{8}, which nodes of a graph are adjacent to which

---

\textsuperscript{8} In the standard definition, each row $i$ of the adjacency matrix is a vector of length $N$ with entries equal to 1 if node $j \in S_N$ is directly connected to node $i$ and zero otherwise.

14
other nodes. 2) Incidence matrix, \( Inc(G(S_N, S_E)) \), is an \( N \times L \) matrix (where \( N \) and \( L \) are the number of Nodes and Lines) that shows which nodes of a graph are connected to which edges. 3) Degree matrix, \( Deg(G(S_N, S_E)) \), is an \( N \times N \) diagonal matrix which contains information about the degree of each node -that is, the number of edges attached to each node. 4) Laplacian matrix \( Lap(G(S_N, S_E)) \), the Laplacian matrix, nodal admittance matrix, Kirchhoff matrix or discrete Laplacian matrix is an \( N \times N \) matrix representation of a graph that is defined as follows:

\[
Lap(G(S_N, S_E)) = Deg(G(S_N, S_E)) - Adj(G(S_N, S_E))
\] (2.3)

Figure 2.1 Corresponding graph of IEEE Standard 30 BUS

\(^9\) In the standard definition, each row \( i \) of this matrix is a vector of length \( L \) with entries equal to 1 if line \( leS_E \) is directly connected to node \( i \) and zero otherwise.
One of the most important relationships between the Laplacian and the Incidence matrices is as follows:

\[
\text{Lap}(G(S_N, S_E)) = \text{Inc}(G(S_N, S_E)) \times \text{Inc}(G(S_N, S_E))^T
\]

(2.4)

From this point to the end of this work we will show the corresponding Laplacian matrix of the graph by “\(B\)” and Incidence matrix as “\(H\)”, so regarding (2.4) we have: \(B = HH^T\).

Figure 2.2: (a) Degree, (b) Incidence, (c) Laplacian and (d) Weighted Laplacian matrices for the corresponding graph of IEEE Standard 118 Bus

We will discuss that the nodal-admittance (or structure) matrix of a power network can be represented in terms of a weighted version of the Laplacian matrix of its corresponding graph. Please also refer to e.g., [29] for more information on graph theory and corresponding matrices of
a graph. A visualization of the incidence, Laplacian, degree and weighted Laplacian (nodal-admittance) matrices for the corresponding graph of the IEEE Standard 118 Bus are presented in Figure 2.2

2.2.4 DC Power Flow Model and Corresponding Graph of PN

The AC power-flow model, also known as the AC load-flow model, is a numerical technique in power engineering, applied to a power system in order to analyze the behavior of different forms of AC power (i.e. voltage amplitudes, voltage angles, active (real) power and reactive (imaginary) power) under normal steady-state operation. In such a model two distinct sets of non-linear equations (called power balance equations) indicate the relationship between active and reactive power (injected at bus $i$ from bus $j$) and the voltage magnitude and phase angle as follows [23]-[24]:

$$P_{ij} = g_{i,j}V_i^2 - g_{i,j}V_i V_j \cos(\theta_i - \theta_j) + b_{i,j}V_i V_j \sin(\theta_i - \theta_j)$$  \hspace{1cm} (2.5)

$$Q_{ij} = b_{i,j}V_i^2 - b_{i,j}V_i V_j \cos(\theta_i - \theta_j) + g_{i,j}V_i V_j \sin(\theta_i - \theta_j).$$  \hspace{1cm} (2.6)

Here $P_{ij}$ and $Q_{ij}$ are the active and reactive powers injected from node $j$ to node $i$, respectively; $V_i$ and $\theta_i$ represent the amplitude and phase of the voltage on node $i$, respectively; $g_{i,j}$ is the real part of the admittance of the line $l_{i,j}$ (or the conductance); and $b_{i,j}$ is the imaginary part of the admittance of line $l_{i,j}$ (or the susceptance).

In [23]-[24], it has been shown that when the system is stable, under a normal steady-state condition, the phase angle differences are small, meaning the term $\sin(\theta_i - \theta_j) \approx \theta_i - \theta_j$. i.e. for small phase-shift the power flow is proportional to the phase shift, instead of the sine of this phase difference. Using this simplification and the DC power flow method discussed in [24],[23]-[6], the power injected to a distinct bus $i$ follows the superposition law; this means that for each node $i$ in the network we have the two following approximated summations over the active and reactive powers (however, in practical cases equation (2.8)/(2.10) cannot be exactly validated):

$$P_i = \sum_j P_{ij} = \sum_{j \in N_i} b_{i,j} \left(\theta_i - \theta_j\right)$$  \hspace{1cm} (2.7)

$$Q_i = \sum_j Q_{ij} = \sum_{j \in N_i} b_{i,j} \left(V_i - V_j\right).$$  \hspace{1cm} (2.8)
Here $N_i$ is the set of neighbor buses connected directly to bus $i$. It is useful to rewrite these summations in a matrix-vector format, where we have:

\begin{align}
p &= B\theta \\
q &= Bv
\end{align}

(2.9) \quad (2.10)

In this notation, the voltage amplitude and voltage phasor angle values from all the nodes are collected in two vectors $\mathbf{v}, \mathbf{\theta} \in \mathbb{R}^N$, respectively. Also, the active and reactive power values of all the nodes are stored in the vectors $\mathbf{p}$ and $\mathbf{q}$, respectively. The matrix $B \in \mathbb{R}^{N \times N}$ is called the nodal-admittance matrix, describing a power network of $N$ buses, and its elements can be represented in the following format:

\[
B_{ij} = \begin{cases} 
-b_{i,j} & \text{if } l_{i,j} \in S_E \\
\sum_{j \in N_i} b_{i,j} & \text{if } i = j \\
0 & \text{otherwise}
\end{cases}
\]

(2.11)

The nodal-admittance matrix of the IEEE Standard-118 Bus is shown in Figure 2.2(d). We see that this matrix has a sparse structure. Moreover, it has been shown in the literature that this matrix can be viewed as a weighted version of the Laplacian matrix of the graph $G(S_N, S_E)$ [39] and [23].

2.3 Fundamentals of Compressive Sensing and Sparse Recovery

2.3.1 Sparse Recovery Problem: SRP

The sparse recovery problem (SRP) can be interpreted as a paradigm for recovering an unknown signal from a set of underdetermined linear measurements. The ability to solve an SRP requires the assumption of sparsity for the signal to be recovered, and it requires the measurement or sensing matrix to satisfy certain conditions.

2.3.2 Theorems and Definitions

**Definition 2.12:** A $K$-sparse signal $\mathbf{x} \in \mathbb{R}^N$ is a signal of length $N$ with $K$ nonzero entries where $K < N$ (in many cases ($K \ll N$)). The sparsity level of a signal $\mathbf{x}$ is denoted by the $l_0$ norm $\|x\|_0$. 

18
Intuitively, an SRP is an optimization problem in which the goal is to recover a $K$-sparse signal $x \in \mathbb{R}^N$ from a set of observations $y = Ax \in \mathbb{R}^M$ where $A \in \mathbb{R}^{M \times N}$ is the sensing matrix or measurement protocol with $M < N$ (in many cases $M \ll N$). Due to the underdetermined nature of this recovery problem (since $M < N$), the null space of the matrix $A$ is non-trivial; therefore, there exist infinitely many candidate solutions for this problem. However, under certain conditions on the sensing matrix $A$, various sparsity based recovery methods (solvers) can be guaranteed to efficiently find the candidate solution that is sufficiently sparse. Examples of such conditions are the restricted isometry property (RIP), the exact recovery condition, and low coherence. In essence, what these conditions require is that any two small subsets of columns of the sensing matrix $A$ must be almost orthogonal to each other.

**Definition 2.13** [36]: An $M$ by $N$ sensing matrix $A$ is said to satisfy the Restricted Isometry Property (RIP) of order $K$ if there exists a constant $\delta_K \in (0,1)$ such that:

$$\left(1 - \delta_K\right)\|x\|_2^2 \leq \|Ax\|_2^2 \leq \left(1 + \delta_K\right)\|x\|_2^2$$

(2.14)

holds for all $K$-sparse vectors $x$. The parameter $\delta_K$ is known as the isometry constant of order $K$.

**Definition 2.15** [38]: The coherence of an $M \times N$ matrix $A$ is the maximum normalized inner product (correlation) between any two distinct columns of $A$:

$$\mu_A = \max_{1 \leq m, n \leq N, m \neq n} \frac{|\langle a_m, a_n \rangle|}{\|a_m\|_2 \|a_n\|_2}.$$  \hspace{1cm} (2.16)

It has been shown that the coherence of an $M$ by $N$ matrix $A$ is a value in the interval $[\frac{1}{\sqrt{N}}, 1]$.

**Definition 2.17**: In addition to the coherence, another useful correlation metric presented in the literature is the average coherence criterion [33]. For an $M \times N$ sensing matrix $A$, the average coherence value is defined as follows:

$$\nu_A = \frac{1}{N-1} \max_{i \in \{1, \ldots, N\}} \left| \sum_{j=1 \atop j \neq i}^N \langle a_i, a_j \rangle \right|.$$  \hspace{1cm} (2.18)
Roughly speaking, this metric shows the average level of correlation between sensing matrix columns. This metric can further be used to represent the correlation issue among sparse formulations in power system problems; such as correlation in measurements from electrical bus voltage phase angles.

**Definition 2.19 (Block sparsity):** We refer to a signal $\mathbf{x} \in \mathbb{R}^N$ as $(K; b)$-block sparse if the total number of nonzero coefficients is $K$ and all nonzero coefficients are distributed within $b$ disjoint block each with the same length $d = K/b$ (the locations of blocks are arbitrary) [34].

**Definition 2.20 (Clustered sparsity):** We refer to a signal $\mathbf{x} \in \mathbb{R}^N$ as $(K; c)$-clustered sparse if the total number of nonzero coefficients is $K$ and all nonzero coefficients are distributed within $c$ disjoint clusters (the locations and sizes of the clusters are arbitrary) [20].

Figure 2.3 shows the difference between the general structure of a $K$-sparse signal (Figure 2.3(a)), $(K; b)$-block sparse signal (Figure 2.3(b)), and a $(K; c)$-clustered sparse signal (Figure 2.3(c)).

![Graph](a) 6 sparse signal, (b) $(25,5)$-block-sparse signal, (c) $(21,5)$-clustered-sparse signal.

**Theorem 2.21 $l_1$-Minimization Based Sparse Recovery:** If the $M$ by $N$ sensing matrix $A$ satisfies the RIP of order $2K$ with any isometry constant less than $1$, then the following
optimization problem can recover the original $K$-sparse signal $\mathbf{x}$ from the set of measurements $\mathbf{y} = A\mathbf{x}$ [32] and [35]-[36]:

$$P_0: \quad \hat{x} = \arg\min_{\hat{x}} \|\hat{x}\|_0 \quad \text{subject to} \quad y = Ax'. \quad (2.22)$$

In general, this $l_0$-minimization problem is known to be NP-hard [35] and [16]. Fortunately, there is a relaxed version of this $l_0$-minimization problem, that can still guarantee the recovery of sparse signal, named as “$l_1$-minimization”.

$$P_1: \quad \hat{x} = \arg\min_{\hat{x}} \|\hat{x}\|_1 \quad \text{subject to} \quad y = Ax'. \quad (2.23)$$

Here, $\|\hat{x}\|_1 = \sum_{n=1}^{N} |\hat{x}(n)|$. Since the $l_1$-norm is convex, this results in a tractable convex optimization problem, widely known as Basis Pursuit (BP).

**Theorem 2.24 Noisy $l_1$-Minimization:** In the case of noisy measurements, the following alternative problem can be substituted with $P_1$, to arrive at the correct solution for $\mathbf{x}$ [35]:

$$NP_1: \quad \hat{x} = \arg\min_{\hat{x}} \|\hat{x}\|_1 \quad \text{subject to} \quad \|y - A\hat{x}\|_2 < \eta. \quad (2.25)$$

This problem is widely known as Basis Pursuit De-Noising (BPDN). Suppose $A$ satisfies the RIP condition of order $2K$ with isometry constant $\delta_{2K} < 0.4651$. Let $y = Ax + n$ be the set of noisy measurements of any vector $\mathbf{x}$. If $\eta \geq \|n\|_2$, then the solution $\hat{x}$ to (2.25) obeys:

$$\|x - \hat{x}\|_2 \leq C_1 \frac{\|x - x_K\|_1}{\sqrt{K}} + C_2 \eta, \quad (2.26)$$

where $C_1$ and $C_2$ depend only on $\delta_{2K}$.

**Theorem 2.27 (Exact recovery condition) [36]:** An alternative, but weaker property known as coherence (2.16) is easier to check in practice. For a $K$-sparse vector $\mathbf{x}$ and a vector of

---

10 The vector $x_K$ is the closest $K$-sparse approximation to $\mathbf{x}$. 

---
measurements $y = Ax$, if $\mu_A < \frac{1}{2K-1}$, then, $P_1$ can recover the original $K$-sparse vector $x$ from the set of measurements $y$.

The smaller the coherence, the larger the permitted value of $K$, and the broader the class of sparse vectors $x$ that can be recovered. The minimum number of measurements needed for perfect recovery is related to both the original dimension $N$ of the signal $x$ and the sparsity level $K$; specifically, $M$ must be at least proportional to $K \log(N/K)$. Successful recovery with $K \log(N/K)$ is generally possible (with high probability) when the sensing matrix $A$ is generated randomly, such as with independent and identically distributed (i.i.d.) Gaussian entries.

Within the literature, a variety of other alternative optimization-based formulations have been developed to solve SRPs such as $NP_1$. A famous example is a convex optimization algorithm known as LASSO. The LASSO estimator is defined as follows [37]:

$$\hat{x} = \arg \min_y \| y - Ax \|_2^2 + \lambda \| x \|_1.$$  \hspace{1cm} (2.28)

This can be viewed as the Lagrangian form of the BPDN problem (2.24). Due to solving a recovery problem over a sparse vector $x$ (in all the aforementioned formulations), throughout the rest of this thesis, we call such an optimization problem as a sparse recovery problem or SRP.

2.3.3 Greedy Methods:

Beside the general optimization-based approaches, a family of alternative SRP solvers have been developed in the literature widely known as Greedy methods. Instead of searching among a set of candidate solutions to $P_1$, they try to explicitly find a sparse solution to the equation.

**Theorem 2.29 ([38]):** Suppose that the columns of the sensing matrix $A$ have unit norm, and $\mu_A < \frac{1}{2K-1}$. Then for any $K$-sparse signal $x$, given the set of noise-free measurements $y = Ax$, the stepwise procedure detailed in Algorithm 2.1 can correctly return the correct sparse solution $\hat{x} = x$. This greedy algorithm is widely known as Orthogonal Matching Pursuit (OMP).

In summary, the OMP recovery algorithm attempts to correct the errors of the least square estimation of the sparse vector $x$ caused by minor possible correlations among the columns of sensing matrix $A$. OMP uses an iterative procedure to identify the correct “support (location of the nonzero elements)” of the $K$-sparse vector $x$. The OMP is detailed in Algorithm 1. (Please also...
Various expanded forms of OMP have been developed in the literature which are considering an extra level of information regarding the structure of the signal $x$ such as block sparsity or clustered sparsity. We will widely use these set of greedy methods through the rest of this thesis and will introduce their specifications whenever they are needed within a certain application.

**Definition 2.30:** Within the rest of this thesis we define the true support of a $K$-sparse signal $x$ to be a set of $K$ indexes corresponding to the location of nonzero coefficients along vector $x$.

---

Algorithm 1. Orthogonal Matching Pursuit (OMP)

```plaintext
require: matrix A, measurements $y = Ax + n$, stopping criterion

initialize: $r^0 = y, x^0 = 0, l = 0, SUP = \emptyset$

repeat

1. match: $h^l = A^* r^l$

2. identify support indicator:

   $sup^l = \{argmax_j |h^l(j)|\}$

3. update the support:

   $SUP^{l+1} = SUP^l \cup sup^l$

4. update signal estimate:

   $x^{l+1} = argmin_{z: SUP(z) \subseteq SUP^{l+1}} \|y - Az\|_2$

   $r^{l+1} = y - A x^{l+1}$

   $l = l + 1$

Until stopping criterion met

Output: $\hat{x} = x^l$
```
CHAPTER 3
SPARSE RECOVERY BASED POWER LINE OUTAGE DETECTION AND LOCALIZATION (IDENTIFICATION)

Nowadays, fast and accurate detection and localization of power line outages are beyond the most important monitoring tasks in smart grids. Reliable power line monitoring can prevent serious damages in PNs, such as huge blackouts. Fortunately, the real-time data collected using sensing technologies such as SCADA or WAMS from system parameters can be utilized to develop an online mathematical solution for POI problem. In this chapter\textsuperscript{11} exploiting such a real-time data (available from sensor networks), we investigate the idea of sparse-based power line outage identification and introduce variety of alternative reformulations and sparse recovery approaches to address the technical challenges within such a framework.

3.1 Introduction and Literature Review

In order to implement an accurate POI framework, a variety of methods and algorithms have been developed so far [39]-[51]. However, most of the conventional methods are mathematically complex and inappropriate for real-time monitoring applications. This time complexity is especially critical in the case of large-scale PNs and is destructive in emergency situations such as pre-blackouts. Beside the time and complexity issues, the susceptibility in multiple line outage detection is another important limitation on the performance of the most of these algorithms [39].

Recently, some new approaches have been adapted in order to address the multiple line outages and the challenges therein. The method which has been introduced in [5] adapts a Gauss-Markov graphical model of the PN to deal with the multiple line outages. An ambiguity group-based location recognition algorithm has been proposed in [51]. In [41] a new method for solving the POI problem based on the theory of quickest change detection is developed. Also, in [42] a global stochastic optimization technique based on cross-entropy optimization has been presented. Zhao and Song [43] presented a distributed framework to perform the identification locally at each phasor data concentrator. In [40], Wu et al considered the same problem under scenarios with a limited number of PMUs. The effect of bad data on outage detection is addressed in [45].

\textsuperscript{11} This section is based on the following set of papers: [94] and [97].
Moreover, a non-iterative method for wide-area fault location has been presented in [46] which applies the substitution theorem. Each of these methods has its own advantages and disadvantages. However, due to time cost or algorithmic complexity, most of these methods have been adapted for simple structures and their performance only evaluated within small number of outages, e.g. 1, 2, 3 and 4 outages in [47], [48], [39] and [42], respectively. As another example, a non-linear approach has been recently reported in [50], where the method is only verified over small scale systems such as IEEE 14, 39 and 118 Bus for at most 8 simultaneous line outages. Besides, various assumptions on the proficiency of real time data collection from all buses in the network is needed to ensure the accuracy of the proposed algorithm. Moreover, no time cost analysis has been reported to evaluate the algorithm from the time complexity viewpoint, which is a critical factor in case of larger power networks. Among recent techniques, exploiting the sparse nature of the POI problem in PNs, a new approach has been initialized in [39], where the POI problem has been formulated as a sparse recovery problem and the performance of the algorithm has been reported over small number of simultaneous outages (up to 3).

In this chapter, inspired by this new formulation, we are going to investigate the challenges and capabilities of the sparse recovery technique over POI problem. We term this formulation as POI-SRP. It is shown that due to the sparse nature of the POI-SRP formulation the POI problem can efficiently be solved in a short period. We start the chapter by reformulation the POI problem as a SRP using the graph model of the PN, where a failure phenomenon in the PN is modelled using a set of 3 subnetworks namely, 1) the initial, pre-event or old structure grid, contains $N$ nodes and $L$ or $L_{old}$ active transmission lines; 2) the secondary, post-event or new structure grid, contains $N$ nodes and $L_{new}$ active transmission lines; 3) the difference or the outage grid contains $N$ nodes but only $L_{out} = L_{old} - L_{new}$ lines ($L_{out}$ indicates the number of outages or the lines which have been disconnected). Next, the linear DC power flow model, $p = B\theta \in \mathbb{R}^N$ is integrated into some alternative matrices representations to end up in a set of linear systems of equations termed as Power Outage Identification-Sparse Recovery Problem (POI-SRP), $\mathbf{y} = A\mathbf{x} + n$, where $\mathbf{y} \in \mathbb{R}^N$ is termed the measurement vector and, the matrix $A \in \mathbb{R}^{N \times L}$ is termed the sensing matrix. As it is discussed further in the chapter the resulted vector $\mathbf{x} \in \mathbb{R}^L$ can be considered as a sparse vector$^{12}$ with most of its elements equal to zero except on the corresponding $L_{out}$ outage positions. We

\[12\] Also widely termed as an “$L_{out}$-sparse signal” in compressive sensing and sparse recovery literature, where $L_{out}$ indicates the number of nonzero elements in the corresponding sparse signal $\mathbf{x}$. 25
term this vector as the Sparse Outage Vector (SOV). Under ideal sparse recovery conditions, the OMP (Algorithm 1), can perfectly recover the structure of the SOVs. However, as we will discuss later, a variety of technical challenges may arise from sparse recovery perspective, such as high coherence of the sensing matrix $A$, number of available measurements from the nodes, and high signal dynamic range. To this end, we introduce alternative mathematical formulations, and exploit the modified sparse recovery techniques capabilities to handle these issues (such as matrix decomposition substitution or applying Band-excluded Locally Optimized SRP solvers such as BLOOMP). Another important factor to be considered in case of multiple line outages is the presence of a possible structure in the pattern of outages. We discuss that in such a situation, one can remodel the POI-SRP as a Structured-POI-SRP and solve a new problem which uses the structure assumption about the outages distribution. In this work, we use the idea of clustered sparsity and modify the Clustered OMP algorithm to solve the Structured-POI-SRP. Finally, we discuss how the technical challenges regarding limited number of available PMUs can be addressed within the sparse-based POI framework and how this sparse-based formulation can be interpreted as an auxiliary PMU placement approach in network observability analysis.

3.2 Mathematical Formulation of POI-SRP

In this section, we represent the mathematical formulation of POI-SRP, Structured-POI-SRP and Binary POI-SRP, discussed in a work by Zhu and Giannakos [39] and our papers [94] and [97], respectively. Moreover, the mathematical representation of the weighted Laplacian matrix of the PN, in addition to SOV definition and its corresponding formats are presented.

Consider a PN with a set of $N$ buses and a set of $L$ transmission lines. As it has been mentioned before, we can represent the DC power flow model in the following matrix-vector format:

$$
p = B\theta \in \mathbb{R}^N. \quad (3.1)$$

**Remark 3.2:** In a PN which has been modelled as a graph $G(S_n, S_e)$, the nodal-admittance matrix, $B$, can be viewed as a weighted version of the Laplacian matrix such that [39-related ref]:

$$B = \sum_l^L b_l h_l h_l^T = HW_bH^T, \quad (3.3)$$
where \( h_l \in \mathbb{R}^K \) for \( l = 1, \ldots, L \) represents the \( l^{th} \) column of incidence matrix \( H \), and \( W_b \) is a diagonal \( L \times L \) matrix with the following format:

\[
W_b = \begin{bmatrix} b_1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & b_L \end{bmatrix},
\]

where \( b_l \) indicates the susceptance along the \( l^{th} \) line. Now, suppose that during a disturbance, several line outages occur in the grid. This results in a partial change in the structure of the network and, a new nodal-admittance matrix \( B_{\text{new}} \) for the PN. For sake of clarity we model a failure event using a set of three sub-grids: 1) The initial grid, pre-event or old structure, contains \( N \) nodes and \( L \) or \( L_{old} \) lines; 2) The secondary grid, post-event or new structure, contains \( N \) nodes and \( L_{\text{new}} \) lines; 3) The difference or (namely) the outage grid contains \( N \) nodes but only \( L_{\text{out}} = L_{old} - L_{\text{new}} \) lines (\( L_{\text{out}} \) indicates the number of outages or the lines which have been damaged). The following notation is used in order to identify the corresponding specifications of each of these three networks: \( \mathbf{p}_{\text{old}}, \mathbf{p}_{\text{new}}, \theta_{\text{old}}, \theta_{\text{new}}, \theta_{\text{out}} \in \mathbb{R}^N \) and \( B_{\text{old}}, B_{\text{new}}, B_{\text{out}} \in \mathbb{R}^{N \times N} \). It has been shown in the literature, following an outage incident, the post-event power vector of the system can be modelled as follows [47]:

\[
\mathbf{p}_{\text{new}} = \mathbf{p}_{\text{old}} + \eta, \quad \text{where } \eta \text{ is an } N \times 1 \text{ vector that accounts for the perturbation (residual).}
\]

Assuming \( \theta_{\text{diff}} = \theta_{\text{new}} - \theta_{\text{old}} \), following a simple mathematical substitution we can reach to (3.5). Moreover, fill in the format (3.3) for \( B_{\text{diff}} \) in (3.5), we can conclude (3.6) as well.

\[
p_{\text{new}} = p_{\text{old}} + \eta \Rightarrow B_{\text{old}} \theta_{\text{out}} = B_{\text{out}} \theta_{\text{new}} + \eta \tag{3.5}
\]

\[
B_{\text{old}} \theta_{\text{out}} = \sum_{l \in \zeta_{\text{out}}} x_l h_l + \eta, \tag{3.6}
\]

where \( \zeta_{\text{out}} \) represents the set of all \( L_{\text{out}} \) outage lines and \( x_l := h_l^T \theta_{\text{new}} b_l \forall l \in \zeta_{\text{out}} \). Solving (3.6) with respect to \( \theta_{\text{out}} \) we have:

\[
\theta_{\text{out}} = B_{\text{old}}^{-1} \sum_{l \in \zeta_{\text{out}}} x_l h_l + B_{\text{old}}^{-1} \eta. \tag{3.7}
\]
To find the location of outage lines, an exhaustive search (ES) approach has been developed in [47], where the set of outage indices, \( \zeta_{out} \), is found throughout a \( l_2 \)-norm minimization approach over the least-squares (LS) error of (3.7) for each individual line index numbers. However, due to high computational complexity, such an ES method would be inappropriate in case of large PNs. Moreover, such an ES approach is only applicable in case of single or double line outages [47]-[48].

**Definition 3.8 Sparse Outage Vector (SOV):** Summing over all \( le\zeta_{out} \) in (3.7), we can arrive at the following matrix-vector format:

\[
\theta_{out} = B^{-1}_{old}Hx + B^{-1}_{old} \eta \rightarrow y = Ax + n. \tag{3.9}
\]

where \( y \in \mathbb{R}^M \) is termed the measurement vector\(^{13}\) and, the matrix \( A \in \mathbb{R}^{M \times L} \) is termed the sensing matrix. Practically, we can consider the total number of outage lines, (cardinality of the set \( \zeta_{out} \)) \( |\zeta_{out}|_0 = L_{out} \), that may simultaneously occur in a power grid is a small subset of total number of lines \( (L_{out} \ll L \text{ or } L_{old}) \). As a result, the vector \( x \in \mathbb{R}^L \) can be considered as a sparse vector\(^ {14}\) with most of its elements equal to zero except on the corresponding \( L_{out} \) outage positions (Figure 2.3(a)). We term this vector as the Sparse Outage Vector (SOV) with the following mathematical definition:

\[
x = \begin{cases} x_l & \text{if } le\zeta_{out} \\ 0 & \text{otherwise} \end{cases}. \tag{3.10}
\]

Clearly, recovering such an SOV, \( x \), form (3.9), one can easily obtain the outage line index set \( \zeta_{out} \) looking at the nonzero positions.

**Definition 3.11 Clustered Sparse Outage Vector (C-SOV):** In some of the practical situations (especially in case of cascade failures during pre-blackout stage), due to local load-flow congestion and localized natural disasters behavior, the outages are likely to be expanded to the neighbor lines rather than those which are far from. This means that line outages can develop with a kind of structure. From sparse recovery viewpoint, it means that SOV, \( x \), will have a clustered-shape format

---

\(^{13}\) Initially, we consider that the PMU measurements are available from all of buses in the entire of PN so the number of measurements \( M \) is equal to the number of buses \( N \).

\(^ {14}\) Also widely termed as an “\( l_{out} \)-sparse signal” in compressive sensing and sparse recovery literature, where \( l_{out} \) indicates the number of nonzero elements in the corresponding sparse signal \( x \).
(Figure 2.3(c)). In this situation one can generalize the SOV to a Clustered SOV (C-SOV). We term such a situation as structured multiple outages. This possible structure in outage expansion pattern (if happens) can be considered as a new challenge from sparse recovery viewpoint.

Given that in (3.9) the number of measurements $M$ is less than the dimension of the unknown vector$^{15} \mathbf{x} \in \mathbb{R}^L$ and $\mathbf{x}$ is an $L_{out}$-sparse vector, solving (3.9) with respect to SOV $\mathbf{x}$ can be interpreted as an SRP termed POI-SRP, moreover, we alternatively rename the POI-SRP as Structured POI-SRP when we are dealing with C-SOVs. As it has been mentioned in Section 2.4, while the sensing matrix $A$ is fat, the solution of the POI-SRP is not unique, however, under certain conditions (guarantee conditions) on the sensing matrix $A$, $L_{out}$-sparse signal $\mathbf{x}$ can be theoretically recovered from a number of measurements $M$ which should be at least proportional merely to $L_{out} \left( \log \left( \frac{L}{L_{out}} \right) \right)$. This number of measurements can be significantly smaller than the number of unknown parameters (namely $L$ in POI-SRP), and is only greater than the fundamental limit $L_{out}$ by a logarithmic factor$^{16}$.

3.3 Exploiting Sparsity in Line Outage Identification: Challenges and Solutions

Regarding sparse recovery theorems (Chapter.2), an SRP can efficiently be solved and the sparse vector $\mathbf{x}$ can perfectly be recovered if:

1. Enough number of measurements $M$ is available.
2. The sparse signal dynamic range is bounded and relays over a small range of variations.
3. The coherence of the sensing matrix $A$ is small (or in general, sparse recovery guarantee conditions are satisfied).

In this section, we are going to introduce and investigate the technical challenges regarding satisfaction of these 3 perquisites in the case of line outage identification problem (3.9), specifically in terms of large-scale multiple outages and try to address them within alternative adequate frameworks.

3.3.1 POI-SRP in Case of Limited Number of PMUs

Within the POI-SRP formulation the number of measurements $M$ is forced by the number of PMUs installed in the PN. Mathematically speaking, in case of a PN with partially installed PMUs,

---

$^{15}$ Usually in a PN the number of transmission lines, $L$, exceeds the number of measurable buses, $M$, where a PMU has been installed.

$^{16}$ Intuitively, this logarithmic factor is the price one pays for not knowing the locations of the nonzero coefficients in advance.
problem (3.9) is affected by decreasing the number of measurements and a reduction in the corresponding dimension of the sensing matrix A. Consider a situation where only \( M = N_f \) buses out of \( N \) total buses of the network have been guarded by a PMU. By removing the corresponding unmeasurable bus phase angles from measurement vector \( \mathbf{y} \) (or equivalently \( \theta_{\text{out}} \)) in addition to corresponding rows in the inverse of the nodal-admittance matrix \( B_{\text{old}}^{-1} \), (3.9) can be reformed as follows:

\[
\theta_{\text{out}}' = [B_{\text{old}}^{-1}]' \sum_{t \in \xi_{\text{out}}} x_t h_t + [B_{\text{old}}^{-1}]' \eta \tag{3.9I}
\]

where \( [B_{\text{old}}^{-1}]' \) is constructed from the rows corresponding to \( N_f \) measurable buses of the inverse matrix. In such a situation, the sensing matrix \( A \) will change its dimension to \( R^{N_f \times L} \). From sparse recovery viewpoint, this means that we should solve an SRP with even smaller number of measurements. However, theoretically, the perfect recovery for POI-SRP is still possible if the minimum number of measurable buses remains at least proportional to \( (L_{\text{out}} \log \left( \frac{L}{L_{\text{out}}} \right)) \). Optimal PMU placement was one of the hot topics in power network monitoring so far\(^{17} \) [52]. In some literature, it is found that if about 30\% of buses are equipped with PMU sensors, the full system is recoverable without loss in observability. On the other hand, in many practical SRPs, \( M = 4 * S \), has been found to be the minimum required number of measurements (where \( S \) is the sparsity level of the vector to be recovered, equal to \( L_{\text{out}} \) in our notation). Considering the discussion on system observability, almost \( M \approx 39, 100 \) and 794 number of measurements should be at least accessible in the IEEE standard 118, 300 and 2383 Bus systems, respectively. These networks are including \( L = 186, 411 \) and 2896 transmission lines. Thus, sparse recovery theorem indicates that from measurements number perspective, the maximum number of theoretical detectable simultaneous outages (in the presence of at least 30\% installed PMUs in the network) using POI-SRP is \( L_{\text{out}} \approx 10, 25 \) and 198 respectively. In case of full measurable power network these values is increased to \( L_{\text{out}} \approx 29, 75 \) and 595, respectively. However, these numbers are calculated for ideal noise free situations. Our case studies indicate that, in practice, besides the minimum number of measurements, the sparse recovery guarantee conditions such as sensing matrix coherence and signal dynamic range, in addition to corresponding noise level are limiting the overall recovery

---

\(^{17}\) Nowadays vast amount of work is undergoing in PMU technology. In near future this technology will change from a hardware to a virtualized software-based technology with much lower cost of units [70]-[71]. As a results, in near future the next generation of cost efficient PMUs can be implemented on all of the buses in the network.
performance. In what comes next, we will discuss these two challenges and their behavior and role in POI-SRP in addition to our solutions to address these issues. Moreover, we discuss how an axillary sparse-based optimal PMU placement criterion can be defined based on the coherence minimization.

3.3.2 Signal Dynamic Range Issue and Binary POI-SRP

**Definition 3.12 Signal Dynamic Range:** by definition, for a signal $\mathbf{x} = [x_1, \ldots, x_L] \in \mathbb{R}^L$, the dynamic range is defined as follows:

$$
\text{Signal}_{\text{dynamic-range}} = \frac{\max(x_i)}{\min(x_i)} \quad \text{and} \quad x_i \neq 0 \quad \text{for} \quad i = 1: L
$$

(3.13)

Considering $x_i = h_i^T \theta_{new} b_i$, the dynamic range of the SOV $\mathbf{x}$, may vary in a big range in the PN depends on the voltage phase angle behavior in addition to the susceptance of the transmission lines. In order to address this issue, we propose the following alternative POI-SRP formulation where the SOV $\mathbf{x}$ has a binary format so the $\text{Signal}_{\text{dynamic-range}} = 1$. Inspired by the former formulations [39],[45] and considering $\theta_{new} = \theta_{out} + \theta_{old}$, $B_{new} = B_{old} - B_{out}$ and using notation (3.3) for $B_{out}$ we have:

$$
B_{out} = \sum_{l \in \xi_{out}} b_l h_l h_i^T = HW_b \text{diag}(\mathbf{x}_b) H^T,
$$

(3.13)

where $\mathbf{x}_b \in \{0,1\}^L$ such that:

$$
\mathbf{x}_b = \begin{cases} 
1 & \text{if} \quad l \in \xi_{out} \\
0 & \text{otherwise}
\end{cases}
$$

(3.14)

fill in (3.5) for $B_{out}$ (3.13) results in:

$$
B_{old} \theta_{out} = B_{out} \theta_{new} + \eta = HW_b \text{diag}(\mathbf{x}_b) H^T \theta_{new} + \eta = HW_b \text{diag}(H^T \theta_{new}) \mathbf{x}_b + \eta.
$$

(3.15)

Solving this equation with respect to $\theta_{out}$ we have:
\[
\theta_{out} = B^{-1}_{old}HW_b \text{diag}(H^T \theta_{new})x_b + B^{-1}_{old}\eta.
\] (3.16)

Now considering \(A_b = B^{-1}_{old}HW_b \text{diag}(H^T \theta_{new})\), we can reach to the following new sparse formulation for POD, (3.17), where the SOV \(x_b\) to be recovered is a binary sparse signal.

\[
y_b = A_b x_b + n.
\] (3.17)

Index “\(b\)” stands for the binary representation of the POI-SRP so we call this problem as the Binary POI-SRP (BPOI-SRP). This new formulation can be assumed as an alternative SRP to the POI-SRP. We discuss how this new formulation can help sparse recovery solvers to deal with the dynamic range issue.

### 3.3.3 Sensing Matrix Coherence: Challenges and Solutions

As it has been mentioned in the Chapter.2, one of the most famous sparse recovery guarantee conditions on the sensing matrix \(A\) is the small coherence. If this condition is satisfied within an SRP, it can be guaranteed that with high probability the greedy based sparse recovery algorithms such as Orthogonal Matching Pursuit (OMP) can find the true \(S\)-sparse vector \(x\) (Theorem.2.1). Based on Theorem.2.2, the smaller the coherence metric, the larger the permitted value of \(S\) (\(L_{out}\) in POI-SRP), and the broader the number of simultaneous outages (SOVs) that can be recovered.

To study the specific format and properties of the corresponding sensing matrix \(A\) in terms of POI-SRP we use the IEEE Standard 118 Bus system as a case study. Figure 3.1(a) and 3.1(c) visualize the structure of the normalized sensing matrix \(A_{118}\) in addition to its corresponding Grammian matrix, respectively. From this figure, one may easily observe that the average pairwise correlation among the columns of the sensing matrix \(A_{118}\) is high that, regarding (2.15)-(2.17), results in a high coherence and high average coherence accordingly. Surprisingly, not only the coherence metric of this sensing matrix is high but is equal to “1” that, per the Definition 2.7, is the highest possible coherence value (in another word, there is at least a pair of identical columns in \(A_{118}\)).

Our observations over the structure of this PN and its corresponding sensing matrix, \(A_{118}\), reveals that, the following identical pairs of columns can be found in \(A_{118}\): \{(66,67), (75,76), (85,86), (98,99), (123,124), (138,139), (141,142)\}. Take a brief look at the physical structure of
the IEEE BUS-118 we see that each pair of these lines are connecting a same pair of neighbor buses to each other, which means that the following set of buses: \( NR = \{(49,42), (49,54), (56,59), (49,66), (77,80), (89,90), (89,92)\} \) are connected through two distinct transmission lines instead of a single line. Thus, the associated positions of these node-line combinations determine that the absolute value of the corresponding element within the incidence matrix \( H_{118} \) is equal to 2 instead of 1. Finally, while \( A_{118} = B_{old_{118}}^{-1} H_{118} \); this fact generates identical columns in the structure of the corresponding sensing matrix. To avoid from further issues within true signal support selection in sparse recovery solvers one may implement a line-model modification where, based on the superposition admittance law for two parallel lines, the equivalent single line representation would be replaced instead.

Although this can be considered as a mandatory preprocessing step in case of POI-SRPs but still a high average pairwise correlation remains between the matrix columns. A matrix substitution idea has been reported in [39] for whitening the noise distribution. However, this idea can be used to make a meaningful change in the structure of the sensing matrix \( A \) in aim to decrease the correlation between columns. Replacing the \( B_{old}^{-1} \) in (3.9) and (3.16) with its SVD decomposition (or alternatively QR decomposition [97]), \( B^{-1} = U_{(B^{-1})} \Sigma_{(B^{-1})} V_{(B^{-1})}^{H} \), and multiply the both side with \( \Sigma_{(B^{-1})}^{-1} U_{(B^{-1})}^{-1} \) we can reach to the following new POD and BPOD SRPs:

\[
y = V_{(B^{-1})}^{H} H x + V_{(B^{-1})}^{H} \eta = A_{QR} x + n, \tag{3.18}
\]
\[
y_{b} = V_{(B^{-1})}^{H} H W_{b} diag(H^{T} \theta_{new}) x_{b} + V_{(B^{-1})}^{H} \eta = A_{bQR} x_{b} + n; \tag{3.19}
\]

where the new sensing matrix \( A_{svd} \) has the following definition; \( A_{svd} = V_{(B^{-1})}^{H} H \), and \( y = \Sigma_{(B^{-1})}^{-1} U_{(B^{-1})}^{-1} \theta_{diff} \), similarly, the binary sensing matrix \( A_{bsvd} = V_{(B^{-1})}^{H} H W_{b} diag(H^{T} \theta_{new}) \). For instance, the structure of the normalized version of the sensing matrix \( A_{svd} \) and its corresponding Grammian matrix for IEEE BUS-118 have been illustrated in Figure 3.1(b) and 3.1(d), respectively. As it can be seen, applying the matrix decomposition substitution the structure of the sensing matrix, \( A \), is modified in such a way that we can reduce the correlation among sensing matrix columns, reducing the average coherence, and improving the performance of the SRP solvers, respectively. These approaches may be helpful in case of small scale PNs and in the
presence of very small number of outages (single or double line outages). Our observations revealed that, although in case of multiple outages these modifications are still necessary but in general not sufficient specifically in case of the large-scale and more complex PNs. One possible alternative approach to deal with the high coherence issue has been reported in [22]. Roughly speaking, this technique tries to modify the performance of a sparse recovery solver algorithm by modifying the support selection step of the SRP solver instead of making changes in the structure of the corresponding sensing matrix. This modified sparse recovery algorithm is named Band-excluded locally optimized (BLO) technique.

Figure 3.1(a)-(b). Sensing matrix $A$ and $A_{QR}$ of the IEEE BUS-118. (c)-(d) the corresponding Grammian matrices.

3.3.4 Support Selection Modification: BLO-based Techniques for Coherent POI-SRP

The high coherence issue is highlighted and affects the sparse recovery solvers performance over large-scale multiple outages. In another word, the more the number of outage lines we have, the
higher the probability of highly correlated columns appearance in the signal support, and the more confusion within true support selection step in SRP solvers (e.g. step.2 in OMP).

In case of POI-SRP, this challenge can be addressed by exploiting the BLO-based OMP algorithm named BLOOMP (Algorithm 2) which has been developed in [22]. In a nutshell, this technique improves the support identification step in greedy algorithm OMP to deal with high coherence issue. Within the BLO technique a new concept termed the coherence-bound is defined: for some tolerance $\alpha \geq 0$,

$$B_\alpha (\lambda) = \{i | \mu(i, \lambda) > \alpha\}$$  \hspace{1cm} (3.20)

$$B_\alpha (\Lambda) = \cup_{\lambda \in \Lambda} B_\alpha (\lambda)$$  \hspace{1cm} (3.21)

$$B_\alpha^{(2)} (\lambda) = B_\alpha (B_\alpha (\lambda)) = \cup_{j \in B_\alpha (\lambda)} B_\alpha (j)$$  \hspace{1cm} (3.22)

$$B_\alpha^{(2)} (\Lambda) = B_\alpha (B_\alpha (\Lambda)) = \cup_{\lambda \in \Lambda} B_\alpha^{(2)} (\lambda).$$  \hspace{1cm} (3.23)

Intuitively speaking, if within a sparse signal $\mathbf{x}$ the non-zero objects are not in each other’s coherence bound, then it should be possible to localize the objects “approximately” within their respective coherence bound, no matter how large the mutual coherence is [22]. Within bound exclusion step, we make the following changes into the support selection step of OMP (Algorithm 1 step.2):

$$sup^l = \{argmax_j |h^l(j)|\}, \text{where } h^l_j = A^T_r r^l, j = 1, \ldots, L \text{ s.t. } j \notin B_\alpha^{(2)} (SUP^{l-1}), l = 1,2, \ldots$$  \hspace{1cm} (3.24)

This means that the double $\alpha$-coherence-bound of the estimated support in the previous iteration is avoided in the current search and thus the highly-correlated columns would not be selected within the true support of the signal. Similar to the most of the SRP solvers such as OMP, the major drawback with bound-exclusion method is the signal dynamic range issue. It has been observed that the recovery performance of bound-exclusion technique (also conventional OMP) affected by big changes in the signal dynamic range. In order to deal with the dynamic range issue, a local optimization (LO) framework has been added to the bound exclusion technique. We will use the BLO-based OMP (BLOOMP) method in order to address the high coherence issue in POI-SRP. One notable remark is that, since in BPOI-SRP the sparse signal to be recovered is a binary
signal the dynamic range will always be equal to 1, this will help any SRP solver including BLOOMP to have a better performance. The detailed version of the BLOOMP algorithm can be found in Algorithm 2. as follows (for more information please refer to [22]):

Algorithm.2: Band-excluded Locally Optimized OMP (BLOOMP)

require: matrix \( A \), measurements \( y \), Coherence band \( \alpha > 0 \), maximum of cluster size \( m \), stopping criterion
initialize: \( r^0 = p, y^0 = 0, l = 0, SUP = \emptyset \)
repeat
1. match: \( h^l = A^T r^l \)
2. identify support indicator: \( supp = \{ \argmax_j |h^l(j)| \} \ , j \notin B^{(2)}_\alpha (SUP^{l-1}) \)
3. update the support: \( SUP^{l+1} = LO(SUP^l \cup supp) \)
4. update signal estimate: \( y^{l+1} = argmin_{z \in supp(z) \subseteq SUP^{l+1}} ||p - Az||_2 \)
5. update residual vector: \( r^{l+1} = p - Ay^{l+1}, \)
6. increase index \( l \) by 1: \( l = l + 1 \)
until stopping criterion met
output: \( \hat{y} = y^l \)

Local Optimization Procedure

require: \( A, y \), Coherence band \( \alpha > 0 \), \( SUP^0 = \cup \{ supp^i \} \) for \( i = 1: k \)
repeat: for \( i = 1: k \)
1. \( y_i = argmin_{z \in supp(z) = (SUP^{l-1} \setminus supp^i) \cup \{ \} \} ||p - Az||_2 \)
2. \( SUP^i = supp(y_i) \)
output: \( SUP^k \)

3.3.5 Structured-POI-SRP and COMP

In case of structured multiple outages, one may use the idea of Clustered-OMP to deal with the C-SOVs. COMP modifies the support selection step of the OMP considering the pre-knowledge that the nonzero coefficients are locating within a limited number (sparse) of disjoint clusters, with each of maximum length \( m \), among the true support of the C-SOV \( x \). (Algorithm 3). In each iteration once a candidate position is found, a cluster of neighbor positions (with length \( m \)) is formed around the candidate position and this set is added to the true support of the C-SOV \( x \). Within the simulation results we have shown how exploiting the pre-knowledge of structured sparsity, a modified version of COMP named MCOMP (Algorithm. 1) improves the identification results vs OMP in case of Structured-POI-SRP for multiple large number of outages. In case of structured outages, due to the clustered pattern of the sparsity, the true support of clusters within the corresponding C-SOV may not be reached using BLOOMP. To deal with this problem we suggest a combined algorithm composed of the BLOOMP and the MCOMP that we termed as
BLOCOMP. This algorithm adds the pre-knowledge of clustered sparsity to the BLOOMP. This modification prevents us from losing nonzero neighbor elements within a cluster. This BLOCOMP technique is detailed in Algorithm 4.

Algorithm 4: Bound-excluded Locally Optimized OMP (BLOCOMP)

<table>
<thead>
<tr>
<th>Require: matrix $A$, measurements $y$, Coherence-bound, $\alpha &gt; 0$, maximum of cluster size $m$, stopping criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialize: $r^0 = p, y^0 = 0, l = 0, SUP^0 = \emptyset$</td>
</tr>
<tr>
<td>Repeat</td>
</tr>
<tr>
<td>1. Match: $h^l = A^T r^l$</td>
</tr>
<tr>
<td>2. Identify support indicator: $sup^l = { \text{argmax}_j</td>
</tr>
<tr>
<td>3.1. Finding the boundaries of the support</td>
</tr>
<tr>
<td>Upper Bound = $\min((sup^l + m - 1), 1))$</td>
</tr>
<tr>
<td>Lower bound = $\max(sup^l + m + 1), L)$</td>
</tr>
<tr>
<td>3.2. Modified extend support: $sup^l = { \text{Lower bound, ..., sup}^l, ..., \text{Upper Bound} }$</td>
</tr>
<tr>
<td>4. Update the support: $SUP^{l+1} = SUP^l \cup sup^l$</td>
</tr>
<tr>
<td>5. Update signal estimate: $y^{l+1} = \text{argmin}_z { sup^{l+1} \subseteq SUP, | y - Az |_2 }$</td>
</tr>
<tr>
<td>6. Update residual vector: $r^{l+1} = p - \text{Ay}^{l+1}$</td>
</tr>
<tr>
<td>7. Increase index $l$ by 1: $l = l + 1$</td>
</tr>
<tr>
<td>Until stopping criterion met</td>
</tr>
<tr>
<td>Output: $x = x^l$</td>
</tr>
</tbody>
</table>

3.3.6 An Axillary PMU Placement Criterion

Within the literature, a vast amount of work has been developed to address the optimal PMU placement in a PN [52]. Beyond these approaches some end up with a couple of candidate acceptable solutions where a cost function or a criterion is needed to select the best final
configuration. As it has been discussed in Section 3.3.1 in case of limited number of PMUs the structure of the corresponding sensing matrix $A$ will be affected by extracting the corresponding rows for unmeasurable buses in the inverse of the nodal-admittance matrix $B_{old}^{-1}$. Accordingly, this forced matrix dimensionality reduction can affect the distribution of the elements along each individual column and finally change the final coherence value of the matrix. As a result, we suggest the lowest coherence value for the sensing matrix $A$ to be considered as an axillary criterion in the final optimal PMU network configuration selection.

3.4 Simulations and Discussion

In this section, we test the proposed sparse-based POI frameworks for multiple power line outage identification. We use the IEEE standard 118 Bus, 300 Bus and 2736 Bus test-beds, with 186, 411, and 3504 transmission lines, respectively, to model the PN. The MATPOWER toolbox have been utilized to solve the power flow models and the resulting voltage phase angles have been entered to the SRP solvers to form the corresponding measurement vector $y$ and the sensing matrix $A$. In the following set of simulations, we consider the modified POI-SRPs (3.18)-(3.19). Moreover, to comply with the approaches in [47] and [39], following a failure event, we assume that the system settles down to a quasi-stable state. Another assumption, especially in the case of multiple outages, is that the outage lines do not result in an islanding situation for any individual bus in the PN [47] and [39]. This means that any random combination of lines cannot be considered as a valid failure occurrence.

**Multiple Line Outage with Fully Measurable Network**

Within the first part of the simulations, we initially consider the PNs to be fully measurable where at least one PMU has been installed on each electrical bus. We compare the following set of combined approaches:

<table>
<thead>
<tr>
<th>SRP Formulation</th>
<th>POD-SRP</th>
<th>BPOD-SRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRP Solver</td>
<td>POD-OMP</td>
<td>BPOD-OMP</td>
</tr>
<tr>
<td>OMP</td>
<td>POD-BLOOMP</td>
<td>BPOD-BLOOMP</td>
</tr>
</tbody>
</table>
Figure 3.2(a), presents the recovery performance percentage of each of these 4 approaches vs. noise level for 1000 randomly chosen 20-sparse SOV (20 simultaneous line outages) in the IEEE 300 Bus system. Figure 3.2(b), illustrates the recovery performance percentage of each of these 4 approaches vs. noise level for 1000 randomly chosen 50-sparse outage vectors (50 simultaneous line outages) in the IEEE 2736 Bus system. In both two case studies, the standard deviation of the noise (residual term $\eta$) has been set to zero (account for the noise-free case) and $1\%$, $2\%$, $3\%$ or $5\%$, of the average pre-event power injection.

![Bar chart](image)

Figure 3.2: Comparison of recovery performance of BPOI-BLOOMP, BPOI-OMP, POI-OMP and POI-BLOOMP approaches, for (a) IEEE 300-BUS PN in case of 1000 randomly selected 20 simultaneous line outages, and (b) IEEE 2736-BUS system in case of 1000 randomly selected 50 simultaneous line outages vs noise level, respectively.
As it can be observed from these two figures, due to high number of outages once the noise level increases, the high coherence issue in addition to high possible signal dynamic range are highlighted, and thus the identification performance of the OMP is may significantly decreased in the case of POI formulation. The overall performance of the OMP is relatively better in case of BPOD formulation where the signal dynamic range issue has been addressed within the BPOD formulation. On the other hand, modifying the support selection step, BLOOMP outperforms the OMP algorithm in both POD and BPOD scenarios. Finally, considering these two case studies, one can conclude that a combined approach including the BPOD formulation in addition to BLOOMP is the best sparse recovery-based framework to address the large-scale multiple power outages. However, if the nonzero elements are locating within each other’s coherence bounds, we will be limited in the final recovery performance.

**Multiple Line Outage with Partially Measurable Network**

To investigate the partially measurable situations where a smaller set of buses have been guarded with a PMU, the IEEE standard 118 Bus system has been used. The full measurable network has been compared vs a partially measurable network including the following set of measurable buses\(^{18}\): \{1-45, 113,114,115,117\}. Figure 3.3(a) and 3.3(b), illustrate the recovery performance percentage of each of the four aforementioned sparse recovery-based approaches (vs. noise level) for 1000 randomly chosen 10-sparse outage vectors (10 simultaneous line outages) in case of fully measurable and partially measurable IEEE 118 Bus system, respectively. As it can be seen, the smaller the number of PMUs installed within a PN, the lower the identification proficiency. However, comparing these results for 10 simultaneous outages, vs the identification results reported over 3-4 outages in [42], and considering the inherent challenges within a partially measurable system, a good improvement can be clearly observed within the final power line outage identification performance. Moreover, most of the state of the art, such as the very recent work in [50], do not account for the partially measurable networks at all and consider the data to be available from all the buses at each time instance. Finally, the operational time complexity of each of the SRP algorithms have been reported in Table 3.1 for a sample case of a large-scale multiple outages in IEEE 118, 300 and 2383 Bus systems. Compare to the time cost reported over small

\(^{18}\) This is a famous standard example which has been widely used in the previous works. For more info, please refer to [47],[39] and the references therein.
scale outages ($L_{out} < 4$), this time cost is significantly lower than the approaches which have been developed in [47]-[42] or [50].

<table>
<thead>
<tr>
<th>SRP Solver</th>
<th>BLOOMP</th>
<th>OMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-outage 118Bus</td>
<td>0.1564 sec</td>
<td>0.0127 sec</td>
</tr>
<tr>
<td>20-outage 300Bus</td>
<td>0.1722 sec</td>
<td>0.0174 sec</td>
</tr>
<tr>
<td>50-outage 2736Bus</td>
<td>3.8 sec</td>
<td>1.24 sec</td>
</tr>
</tbody>
</table>

Considering the number of possible topologies over 50 simultaneous outages in the IEEE 2736 Bus system (which is extremely large), most of the state of the art approaches such as the exhaustive search [47] or cross-entropy optimization [42] (which has a time complexity almost half of the ES approach) if not impossible to be implemented are really challenging techniques from time complexity and reliability viewpoint.

![Figure 3.3: Comparison of recovery performance of BPOD-BLOOMP, BPOD-OMP, POD-OMP and POD-BLOOMP approaches, for (a) fully measurable 118-BUS IEEE PN, (b) partially measurable PN, in case of 1000 randomly selected 10 simultaneous line outages, vs noise level, respectively.](image-url)

Figure 3.3: Comparison of recovery performance of BPOD-BLOOMP, BPOD-OMP, POD-OMP and POD-BLOOMP approaches, for (a) fully measurable 118-BUS IEEE PN, (b) partially measurable PN, in case of 1000 randomly selected 10 simultaneous line outages, vs noise level, respectively.
Structured Multiple Line Outages

In this part, we examine the performance of the BLOCOMP vs. MCOMP and OMP in case of large-scale multiple Structured-POI-SRP. Figure 3.4(a) and 3.4(b) illustrate the recovery performance of a sample structured power outage for OMP and MCOMP algorithms, respectively. Due to exploitation of pre-knowledge of structured sparsity, MCOMP outperforms the OMP in case of this (15, 3)-clustered sparse outage vector, where the noise level is set to 1%.

Figure 3.4. Comparison of recovery performance of (a) OMP vs (b) MCOMP method for a sample (15,3)-C-SOV in IEEE 118-BUS a-b.
Figure 3.5. Comparison of the recovery performance of (a) OMP vs (b) MCOMP vs (c) BLOCOMP methods for a sample (33,10)-C-SOV in IEEE 118-BUS.

Increment in the number of outage-clusters will results in augmentation in the probability of highly correlated clustered-columns inclusion. Correspondingly, this will affect the performance
of the MCOMP algorithm. Moreover, the dynamic range of the signal may also change in a big range that makes the problem tougher. To deal with these issues, the BLOCOMP algorithm is dropped in to the challenge. Figure 3.5 illustrates the performance of BLOCOMP, vs. OMP, and MCOMP, in a typical highly complex situation where the outage vector $\mathbf{x}$ to is a $(33, 10)$-C-SOV (which considering the number of possible measurements, is a relatively high sparsity level for an outage vector within the IEEE 118Bus). In addition to the challenges the MCOMP recovery performance may be affected by high within-cluster dynamic range of the nonzero elements in each cluster in the signal as well. BLOCOMP can overcome this challenge utilizing the LO procedure to peak the best center for each cluster. Although the BLOCOMP illustrates adequate performance, we can still improve the results using BPOD formulation if needed. Regarding Algorithm 3 and Algorithm 4, the MCOMP and BLOCOMP need the maximum size of the clusters to be known, however, this can be easily determined from the pre-knowledge of the local structure of the incidence matrix of the PN graph. Moreover, in most of the existing works the number of line outages is a priori information for the algorithm.
CHAPTER 4

COMPRESSIVE SENSING-BASED
SMART GRID TOPOLOGY IDENTIFICATION

Similar to our POI-SRP formulations in Chapter 3, most of the state of the art in the line change monitoring area, are adapting their formulations in such a way that the final output of the algorithm indicates the possible position of the outage lines following a failure event. As it has been mentioned before these approaches are usually rely on a pre-knowledge on the corresponding structure of the PN before any intentional or unintentional structural change may happen in the network. Within most of the recent approaches this structural awareness is tied with our accessibility to the structure of the nodal-admittance matrix $B$. However, such a pre-structural information may not be available in advance due to different reasons. As a result, in this chapter\footnote{This section is based on the following set of papers: [26],[30]-[31] and [98].} we will introduce an alternative compressive sensing-based dynamic mathematical reformulation in order to reveal the structure of the nodal-admittance matrix itself using a small set of measurements from system parameters.

4.1 Introduction and Literature Review

The system data exchange (SDX) module of the North American Electric Reliability Corporation (NERC) provides such a network-wide topology status on an hourly basis [8]; however, to transform the PN into a smart network a near real-time monitoring of system topology and transmission lines status is mandatory. Besides, in case of emergency, such as pre-blackouts, where the structure of the PN may change again and again within less than couple of minutes such an hourly-base information may totally go useless and even misleading.

Moreover, another important problem to be considered is the rank deficiency of the matrix $B$, which is widely used in terms of DC power flow model within most of the recent POI formulations [39]-[42]. To shun the problems arising from this rank deficiency, one of the network buses is normally considered as the reference bus and its corresponding row and column are removed from the matrix $B$; as a result, the nodal-admittance matrix becomes full rank. Nevertheless, whenever the records from the reference bus are affected by bad data (due to cyber-attacks or any other...
precision deficiency issues,), the overall procedure can be depraved. Another important issue is with the mistakenly unreported switching events. These types of errors can affect the configuration of the nodal-admittance matrix $B$. This means that, if the last reported status of the topology of the grid has not been stated within a suitable time interval the results can be disturbed. Moreover, to the best of our knowledge, in most of the aforementioned methods the random behavior applied by renewable energy sources and the uncertainty in loads are not accounted within the problem formulation.

Finally, as we also developed our POI-SRP formulation based on the similar assumptions, the mathematical formulation of some methods is dependent on the connectivity of the power grid graph, and quasi steady state behavior of the network following a failure event [39]. However, some of these assumptions may not be totally reliable in terms of emergencies such natural disasters.

Although POI problem has been addressed in various articles but the entire power network (PN) topology identification (TI) problem has been rarely considered in the literature. Recently the TI problem has been specifically addressed in [53] based on a consuming or economic approach. In this work, we introduce an efficient and low cost solution for TI (that can also be used to solve the POI problem) using a few measurements of the system parameters.

Our work combines sparse sampling theory from the field of Compressive Sensing with concepts borrowed from graph theory. Although formulated based on DC power flow model, our approach is completely different from previous formulations, in that the presented reformulation relies only on the measurements from system parameters and does not need any a priori information about the topology of the network. In fact, we reformulate the topology identification problem in such a way that the nodal-admittance or structure matrix, $B$, itself, is the direct output of the optimization problem. As a result, this framework is capable of dealing with the aforementioned imperfections.

Case studies using standard IEEE test-beds indicate that the proposed method represents a promising new strategy for topology identification, line change, fault detection, and monitoring issues in SGs. This problem and challenges therein are efficiently solved using modified sparse recovery algorithms.
4.2 Compressive Topology Identification

Due to our discussion in Chapter 2, since the Laplacian matrix has a direct relationship with other important structural matrices of a graph, the nodal-admittance matrix $B$ can give a full description of the power network structure.

As a result, our goal is to design an appropriate and fast method to determine the structure of the matrix $B$ from a small (compressive) set of measurements of the PN parameters. A key assumption in our work is that the SG can be considered as a sparse interconnected system. This assumption is based on a survey of articles and standard power network models found in software and toolboxes such as MATPOWER.

The nodal-admittance matrix of the IEEE Standard-14, 30, 57 and 118 Bus are represented in Figure 4.1 (a)-(d). A major observation from this set of figures is that these matrices have a sparse structure, which indicates our assumption on the sparsity of the IEEE Standard PN models. Here we develop our compressive topology identification formulation based on a dynamic interpretation of the DC power flow model, $P_i = \sum_j p_{ij} = \sum_{j \in N_i} b_{i,j} (\theta_i - \theta_j)$ and its corresponding matrix vector format $\mathbf{p} = B \mathbf{\theta}$.

4.2.1 Gaussian Distribution of Parameters

In [4]-[5] it has been thoroughly discussed that due to the load uncertainty in large-scale transmission networks and the increasing contribution of distributed renewable sources in smaller scale PNs, injected active and reactive powers can be modeled as random variables. Moreover, “the injected power can be modeled as a Gaussian random variable since it models the superposition of many independent factors (e.g., loads) [54]”, [4].

The aforementioned linear relationship in (3) suggests that the difference of phasor angles $(\theta_i - \theta_j)$ for each sample of time across a bus can be approximated by a Gaussian random variable as well [4]-[5]. Moreover, allowing for the fixed phasor at the slack bus, under steady-state conditions individual phasor angle measurements $\theta_i$ can be modeled as Gaussian random variables.

Consequently, we will assume that during the observation window of a TI problem, the measurements of the phase angle of node $i$ can be well approximated by Gaussian random variables. As we will discuss later in this chapter, this random behavior has an impact on the
structure of the sensing matrix in the TI problem, and it is beneficial to the performance of sparse recovery techniques.

Figure 4.1 Corresponding normalized Laplacian matrices of (a) IEEE Standard-14 Bus and (b) IEEE Standard-30 Bus, (c) IEEE Standard-57 Bus and (d) IEEE standard IEEE 118 Bus.

4.2.2 Sparse Setup of Topology Identification Problem

Given an interconnected PN of \( N \) buses, let the phase angle and active power measurements of \( i^{th} \) bus be associated with the following time series of \( M \) sample times:

\[
\begin{align*}
    p_i(t) & \quad \text{for } t = 1, 2, \ldots, M \\
    \theta_i(t) & \quad \text{for } t = 1, 2, \ldots, M .
\end{align*}
\]

As mentioned before, in the DC model, the (active or reactive) power injected into a distinct bus \( i \) follows the superposition law (2.9) and (2.11). Since (2.11) cannot be validated properly in some
practical situations, we prefer to use (2.9) to formulate the sparse TI problem. This means that, for
each node \( i \) in the network and at each sample time \( t \), we have:

\[
p_i(t) = \sum_j p_{ij}(t) = \sum_{j \in N_i} b_{i,j} (\theta_i(t) - \theta_j(t)). \tag{4.3}
\]

As mentioned before, \( b_{i,j} \) is the susceptance along the line \( l_{i,j} \) under the DC model, and \( N_i \) is
the set of neighbor buses connected directly to bus \( i \). If we write \( b_{i,j} = 0 \) for \( j \not\in N_i \), we can extend
(4.3) as follows:

\[
p_i(t) = \sum_j p_{ij}(t) = \sum_{j \in S_N^i} b_{i,j} (\theta_i(t) - \theta_j(t)) + u_i(t) + e_i(t). \tag{4.4}
\]

where \( S_N^i \) is the set of all nodes in the network except node \( i \) (so the set \( S_N^i \) contains \( N - 1 \) nodes),
\( u_i \) is a possible leakage of active power in node \( i \) itself, and \( e_i \) is the measurement noise. Since we
assume data is collected at \( M \) sample times, we can use matrix-vector notation to write

\[
p_i = A_i y_i + u_i + e_i, \tag{4.5}
\]

where, \( p_i, u_i, e_i \in \mathbb{R}^M \). In equation (4.5) we have

\[
A_i = \begin{bmatrix} a_{1,i}^T, \ldots, a_{i-1,i}^T, a_{i+1,i}^T, \ldots, a_{N,i}^T \end{bmatrix} \in \mathbb{R}^{M \times N-1} \tag{4.6}
\]

\[
a_{j,i}^T = (\theta_j(t) - \theta_i(t)) \in \mathbb{R}^M \quad \text{for} \quad t = 1, 2, \ldots, M \tag{4.7}
\]

\[
y_i = \begin{bmatrix} b_{i,1}, \ldots, b_{i,i-1}, b_{i,i+1}, \ldots, b_{i,N} \end{bmatrix}^T \in \mathbb{R}^{N-1}. \tag{4.8}
\]

Each column of the matrix \( A_i \) represents the difference of phase angles between node \( i \) and one
node \( j \in S_N^i \) in the network across \( M \) samples of time. As discussed in Section 4.2.1, this difference
can be modeled as a Gaussian random variable.

The vector \( y_i \) is a sparse vector, with all its elements equal to zero except a small portion, which
are located in those positions corresponding the (few) neighbors of node \( i \). Summing over \( u_i \) and
\( e_i \) (modeled as a vector of white Gaussian noise), we reach the following equation for each
individual bus:
\[ p_i = A_i y_i + \eta_i . \quad (4.9) \]

The TI problem can be viewed as the estimation of all vectors \( \{y_i\}_{i=1}^{N} \) that best match the observed measurements \( \{p_i\}_{i=1}^{N} \). In order to solve such a problem, we can define the following optimization problem:

\[
\min_{\{ \hat{y}_i \}_{i=1}^{N}} \sum_{i=1}^{N} \|A_i \hat{y}_i - p_i\|_2^2 . \quad (4.10)
\]

Problem (4.10) can be solved individually for each node \( i \). If \( A_i \) is full rank, the number of measurements \( M \) exceeds the number of unknowns \( N - 1 \), and the variance of the noise vector, \( \eta_i \), is small enough, this optimization problem can be solved using Least Square (LS) techniques. Since the value of \( N \) depends on the number of buses in the grid, a large number of measurements \( M \) will be needed to solve this problem in the case of large power grids. In order to avoid this problem, we suggest using sparse recovery techniques to solve for the vectors \( \{y_i\}_{i=1}^{N} \).

Since each vector \( y_i \) is sparse, we can view the goal in (4.9) as recovering an \( S \)-sparse vector \( y_i \in \mathbb{R}^{N-1} \) from a set of observations \( p_i \in \mathbb{R}^{M} \), where \( A_i \in \mathbb{R}^{M \times N-1} \) is the sensing matrix, \( \eta_i \in \mathbb{R}^{M} \) is a vector of white Gaussian noise, and \( S \) is the number of nodes which are directly connected to the one individual node \( i \) for which we are solving the problem.

In general, we want \( M \ll N \) so the problem can be solved using a reasonable set of observations in huge power grids. After solving this problem for each sparse vector \( y_i \) corresponding to each node \( i \), we can concatenate all the sparse vectors together, form the weighted Laplacian matrix \( B \), and the process is completed. Because \( \theta_i(t) - \theta_j(t) = 0 \) for \( i = j \), \( S_N^i \) should include \( N-1 \) nodes; in other words, we should keep the node \( i \) out of \( S_N^i \) since it produces a vector of zeros in the corresponding column of the matrix \( A_i \). Thus, we cannot directly calculate the value of the parameter \( B_{ii} \) from the recovered vector \( y_i \); however, regarding the aforementioned definition of the matrix \( B \),

\[
B_{ii} = \sum_{j \in N_i} b_{i,j} . \quad (4.11)
\]
Thus, after recovering the vector $y_t, B_{ii}$ can be easily calculated. We term this SRP as compressive power network topology identification (CPN-TI). In the next section, we investigate the unique specifications of this SRP and address couple of challenges and capabilities within solving the CPN-TI.

4.3 Exploiting Sparsity in Compressive Topology Identification

In general, the standard techniques for solving SRPs can be categorized into two major groups: greedy algorithms and convex optimization algorithms. We have already utilized couple of basic and modified greedy algorithms in Chapter 3 each of which may be implemented based on a specific configuration in the SRP under study. In this chapter beside implementing greedy based algorithms we would examine the capabilities of convex optimization recovery approaches as well. We would specifically, incorporate $P_1$ and $NP_1$ formulations, LASSO estimator and a generalized version of $P_1$ widely known as Reweighted $l_1$-minimization.

4.3.1 Exploiting Special Sparse Structures in Nodal-Admittance Matrices

Take a brief look at Figure 4.1, one may easily extract the following two facts regarding the structure of the columns in the nodal admittance matrix $B$ (or equivalently the set of sparse vectors to be recovered through CPN-TI):

1. By inspection of the columns of the Laplacian matrices of standard IEEE test-beds we see that the columns tend to exhibit a clustered sparse structure: the column vectors not only contain few nonzero elements (i.e., they are sparse), but the nonzero elements tend to occur in nearly adjacent positions. In the case of the IEEE Standard-14 Bus or IEEE Standard-30 Bus, columns 2, 5 and 6 or columns 6 and 10 can be taken as clear examples, respectively. Since in a real PN, nodes and lines are ordered based on their position in the network, and commonly the neighbor nodes are connected through neighbor lines (except in special cases), it is realistic in practice to assume that nodes have an ordering that lends itself to clustered sparsity. This fact motivated us to use this structural pre-knowledge and try to solve the CPN-TI using COMP solver.

2. On the other hand, we can see that due to the roughly sequential numbering of neighbor buses within a PN, the positions of the nonzero elements in this set of matrices exhibit a certain structure. In particular, Figure 4.1 shows that most of the nonzero entries concentrate close to the main
diagonal. This fact motivated us to implement a weighted version of $l_1$-minimization formulation that takes the diagonal concentration fact into the account.

4.3.2 Reweighted $l_1$-Minimization Based CPN-TI

In the literature, some specialized convex optimization methods have been introduced which can exploit extra prior information to solve the SRP more efficiently and more accurately from even fewer measurements than $P_1$ or $NP_1$.

In terms of CPN-TI formulation, Figure 4.1 shows that most of the nonzero entries (within the standard nodal-admittance matrices) concentrate close to the main diagonal. This means that in each sparse vector $\mathbf{y}$, certain entries are more likely than others to be nonzero. To exploit this anticipated behavior, we can replace $P_1$ with the following weighted optimization problem instead [55]:

$$WP_1: \hat{\mathbf{y}} = \min_{\mathbf{x}} \sum_{n=1}^{N} \omega_{(n)}|\hat{y}_{(n)}| \quad \text{subject to} \quad \mathbf{p} = A\mathbf{y},$$

(4.12)

where $\omega_{(1)}, \omega_{(2)}, \ldots, \omega_{(N)}$ are non-negative weights. Similar to $P_1$, $WP_1$ is a convex optimization problem that can be solved as a linear program (LP). By assigning smaller weights near the diagonal elements, we can encourage those elements to be selected in the recovered sparse signal $\mathbf{y}$. $WP_1$ can be solved within a single step using our prior knowledge for setting the weights. However, in order to increase the accuracy of the final results we have extended this formulation through an iterative procedure called Reweighted $l_1$ minimization. This iterative algorithm ($Rwl_1$) updates the weights in each step based on the estimated sparse vector magnitudes from the previous step (see Algorithm 5-[55]).

In the Algorithm 5, $\varepsilon$ is defined as a stabilizer parameter that is used in order to obviate the effect of zero-valued components in $\mathbf{y}$. It has been shown that in general the $Rwl_1$ recovery process tends to be reasonably robust to the choice of this parameter ($\varepsilon > 10^{-3}$ is suggested for practical situations). The diagonal matrix $W^l$ is defined as follows:

$$W^l = \begin{bmatrix} \omega_{(1)}^l & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \omega_{(N)}^l \end{bmatrix}.$$ 

(4.13)
Algorithm 5: Reweighted $l_1$-minimization ($Rwl_1$) CPN-TI

require: stopping criterion, phase angle measurements, active power measurements, each of $M$ sample times

$p_i(t), \theta_i(t)$ \hspace{1cm} for \hspace{0.1cm} $t = 1, 2, \ldots, M$.

1. form: measurement matrix $A \in R^{N-1 \times M}$, and measurement vector $y$

2. Set counter $l$ to zero and initialize $\omega_{(n)}^0$ for $n = 1: N - 1$

3. solve:

$$y^l = \text{argmin} \|W^{(l)}y\|_1 \hspace{0.5cm} \text{subject to} \hspace{0.5cm} p = Ay$$

4. update weights:

$$\omega_{(n)}^{l+1} = \frac{1}{|y_{(n)}^l| + \epsilon} \hspace{0.5cm} \text{for} \hspace{0.5cm} n = 1: N - 1$$

5. Go back to step 3. Until stopping criterion met.

6. output: $\hat{y} = y^l$

4.3.3 Noisy Reweighted $l_1$-Minimization Based PNTI

In short, to adapt the $Rwl_1$ algorithm in the case of noisy measurements the following change should be made in the 3rd step:

$$y^l = \text{argmin} \|W^{(l)}y\|_1 \hspace{0.5cm} \text{subject to} \hspace{0.5cm} \|p - Ay\|_2 < \eta. \hspace{1cm} (4.14)$$

Additional discussion of $WP_1$ and $Rwl_1$ is contained in [55]. There are couple of optimization packages that can be used to solve $P_1$, $NP_1$, $WP_1$ and $Rwl_1$; examples include CVX [56] and NESTA [57]. Our justification for this modeling assumption of quasi-sequential node numbering was based on the structure observed in IEEE standard test beds as well as some real-world power networks. We believe that this modeling assumption will be valid in a considerable number (but not all) networks. Moreover, whenever a new area is constructed and is added to a power network the new buses will usually be numbered similarly to their neighbors, so some sequential numbering may be preserved. In case such a sequential numbering is not present within a network then one may see only limited improvement from weighted $l_1$ minimization. Even in such a situation, one may still use ordinary sparsity-based optimization approaches (including the iterative re-weighted $l_1$-minimization algorithm, but with weights initialized to 1) to solve the problem, although with a slightly greater number of measurements.
4.3.4 Data Correlation Issue

Even though graph data analytics are essential for gaining insight into big networks, large-scale interconnected network processing is complex because of its graph-specific challenges, including complicated correlations among data entities, highly skewed distributions, and so on. Considering large-scale interconnected power networks, the possible correlation among the recorded data from the nodes (especially from neighboring nodes) is an issue in sparse TI since it may result in highly coherent sensing matrices $A$. As noted in Theorem 2.29, high coherence can have an undesirable effect on the performance of SRP solvers. On the other side, as it has been mentioned before, in TI the positions of the nonzero elements of a signal vector are typically clustered. Since correlation in a sensing matrix tends to be highest for columns corresponding to neighboring nodes, this means that applying the BLOOMP method may cause us to miss some nonzero elements of the signal vector. In order to deal with this problem while addressing the possible correlation among the sensing matrix columns, we suggest using the BLOMCOMP as an alternative solution (Algorithm 4).

4.4 Simulation Results and Discussion

4.4.1 Setup

In this section, we test the proposed method for recovering the topology of an SG using compressive observations ($M < K$) collected from the system parameters. We see that the probability of successful recovery varies for different nodes in the network based on the local pattern of sparsity. In these simulations, we use the IEEE Standard-30 Bus, 118 Bus and, 2383 as case studies. These three PNs include 30 Buses and 47 power transmission lines, 118 Buses and 186 lines, and 2383 Buses and 2896 lines, respectively, and their detailed specifications have been fully described in the MATPOWER toolbox [17]. To generate the data, we first feed the system with Gaussian demand and simulate the PN. The MATPOWER toolbox is used for solving the power flow equations in various demands and the resulting phase angle and active power measurements are applied as the input to the SRP. Finally, 1-2% white noise has been added to measurement vectors randomly. Each column of the matrix $B$ represents one of the sparse vectors $\mathbf{x}_i$ which we attempt to recover by solving an SRP. For the IEEE Standard-118 Bus and 30 Bus, Figure 2.1(a), and 2.1(b), show how the coherence (2.16) of the corresponding sensing matrix $A$
changes as the number of measurements $M$ increases; curves are averaged over 100 realizations of the network and over all of the 118/30 nodes. This coherence measure appears to approach a non-zero asymptote as the number of measurements increases. As has been mentioned before, the smaller the coherence metric, the larger the sparsity $S$ of signals that can be recovered. One class of matrices well known to have low coherence are random Gaussian matrices. As mentioned in Section 4.1.3 and in the literature such as [4]-[5], due to the load uncertainty and the aggregation of the renewables in SGs, the difference of the phase angles across a bus can be approximated by a Gaussian random variable. Moreover, in our SRP formulation, the sensing matrix is formed by the variation in bus phase angles. Thus, the closer the behavior of the phase angles is to being Gaussian, the better one might expect the sparse recovery algorithms to perform.

Figure 4.2: (a) The average value of Coherence $\mu_A$ vs # of measurements $M$ over all 118 corresponding sensing matrices 'A' for all the nodes in the IEEE 118-Bus over 100 realizations of the network. (b) Average coherence versus number of measurements over all corresponding sensing matrices $A_i$ of all the nodes in IEEE Standard-30 Bus system, where the curve is averaged over 100 realizations of the network.

4.4.2 Results

*Optimization-based Algorithms in IEEE Standard-30, 300, and 2383 Bus Networks*

Within the first set of simulation we tested the proposed method for recovering the topology of a SG using compressive observations, collected from the system parameters. In these simulations, we used the IEEE Standard-30 Bus and IEEE Standard-300, IEEE Standard-2383 Bus as case
studies. These power networks include 30, 300, and 2383 buses and 47, 411, and 2896 power transmission lines, and their detailed specifications have been fully described in MATPOWER toolbox. The MATPOWER toolbox is used for solving the power flow equations in various demands and the resulting phase angle and active power measurements are applied as the input to the sparse solver. In order to generate the data, first, we fed the system with Gaussian demands and simulated the PN.

Figure 4.3 Recovery rate comparison of nodes (a) 3 (in-degree 2), (b) 10 (in-degree 6) in the IEEE standard 30 Bus, respectively. Success rate is calculated over 100 realizations of the network for a given number of measurements.

Within the network graphs, SRP solvers have different recovery performance for different nodes, mainly because of the sparsity level of the signal (or in our PN, in-degree or the number of incoming transmission lines to an individual bus). Figure 4.3(a), and 4.3(b) show the recovery performance of the $BP$ and $Rwl_1$ solutions for nodes 3 and 10 of the IEEE Standard-30 Bus,
respectively. These buses are distinguished from each other by their in-degrees. The success rate is calculated over 100 realizations of the network for a given number of measurements. From the sparsity level viewpoint, the 6-sparse signal $x_{10}$ corresponds to one of the most complicated signals $x_i$ to be recovered in the IEEE Standard-30 Bus. The 2-sparse signal $x_3$ (corresponding to the double in-connection node 3) is more likely to be recovered using a smaller number of measurements than the 6-sparse signal $x_{10}$ (corresponding to the 6 in-connection node 10). In Figure 4.4, the same trends can be observed by looking at the recovery rate over the 1st, 2nd, 15th, 3rd and 130th buses, with sparsity levels (in-degrees) 3, 4, 6, 7, and 9, respectively, in the IEEE Standard-300 Bus.

We occasionally observe that nodes with similar in-degree actually have a different recovery performance. For example, nodes 3 and 15 (corresponding to the 7 and 6-sparse signals $x_3$ and $x_{15}$, respectively) have close in-degrees, but Figure 4.4 shows that the recovery rate curve of $x_3$ is closer to that of the 9-sparse signal $x_{130}$, i.e., it requires a larger number of measurements for correct recovery. This difference can be caused by the network-wide location of the bus and also the structure of the incoming transmission lines to each bus. Moreover, due to the close geographical position and similarity in load pattern within an interconnected network, the parameters of different nodes may share a level of correlation. This will affect the coherence of the resulting sensing matrix and may result in different recovery performance over the network nodes. Figures 4.3 and 4.4 also show how the presence of structured sparsity helps the $R_{wl_1}$ algorithm to outperform $BP$, especially in case of the IEEE Standard-300 Bus where the in-degree of the nodes is larger.

Figure 4.5(a)-(c) provide a node-by-node comparison of the recovery performance for the 2 aforementioned SRP approaches over 100 realizations of the IEEE Standard-30 Bus. The vertical axis indicates the number of measurements $M$, while the horizontal axis represents the bus number from 1 to 30. A color spectrum ranging from dark blue (corresponding to 0% recovery) up to dark red (corresponding to 100% recovery) has been used to illustrate the recovery performance percentage. Results indicate that, for almost all nodes, both of the algorithms can arrive at full

---

20 For our success criterion we first check to see whether the support (the positions of the non-zeros) of the true signal is correctly identified, and if so, we then check to see whether the recovery error $\|x - x_0\|_2/\|x_0\|_2$ is within a certain bound $\varepsilon$. Both conditions must be satisfied for recovery to be considered successful. We picked $\varepsilon = 0.05$ to illustrate the acceptable recovered signals.

21 Or equivalently, the column number from the nodal-admittance matrix.
recovery performance using 20-25 measurements. In general, however, the $Rwl_1$ algorithm demonstrates better performance with lower numbers of measurements, especially for columns with high congestion around the diagonal elements.

Figure 4.4. Recovery rate comparison of nodes (downward): 1 (in-degree 3), 2 (in-degree 4), 15 (in-degree 6), 3 (in-degree 7), 130 (in-degree 9) respectively. Success rate is calculated over 100 realizations of the IEEE Standard-300 Bus for a given number of measurements.

Finally, Figure 4.6(a)-(c) demonstrate the network-wide topology recovery performance for BP and $Rwl_1$ on the IEEE Standard-30, 300, and 2383 Bus networks, over 100 realizations of each
network, respectively. For each curve, the vertical axis represents the percentage of trials in which all 30, 300, or 2383 columns of the corresponding nodal-admittance matrix $B$ (and, as a result, the network-wide topology) are successfully recovered.

As has been discussed, theoretically, the number of measurements required for full recovery $M$ should be at least proportional to $S \log(N/S)$. In the case of the IEEE Standard-30 Bus, node 6 has the highest in-degree of 7; as a result the whole network-wide recovery is expected at roughly...
\[ M = 7 \log(30/7) \approx 10 \text{ or more.} \]  Moreover, based on the discussion in Section 2.4.1, as the number of buses of a particular network, \( N \), grows, more measurements are needed. For example, in case of the IEEE Standard-300 Bus, node 276 has in-degree 12, and full recovery is theoretically possible when roughly \( M = 12 \log \left( \frac{300}{12} \right) \approx 39 \).

In the case of the IEEE Standard-2383 network, node 1920 has in-degree 9; as a result, full recovery should be expected at \( M = 9 \log \left( \frac{2383}{9} \right) \approx 50 \) measurements or more. However, Figure 4.5 and 4.6 indicate that in practice the whole network topology for these two systems can typically be recovered from \( M \approx 20, 60 \) or \( 90 \) measurements per bus, respectively. This is likely due to the coherence of the sensing matrices (because the elements are not perfectly independent Gaussian random variables) and the additional noise.

An important factor is that, although the number of measurements needed for full recovery increases with the network size \( N \), this is not a linear relationship. As can be seen, although the IEEE Standard-2383 and 300 Bus are almost 80 and 10 times larger than IEEE Standard-30 Bus in scale, respectively, they require less than 3 and 5 times more measurements for full recovery, respectively.

This fact highlights the suitability of the sparse CPN-TI setup especially for large scale power networks. In general, a lack of measurements reduces the recovery performance for any SRP solver; however, the reweighted \( l_1 \) minimization algorithm (Rw\( l_1 \)), which uses additional structural knowledge on the sparse vectors, is less affected and exhibits a better recovery performance when compared with \( BP \). Finally, we do note that it is not possible to recover the nodes with higher in-degrees until \( M \) is large enough that the PN coherence metric reaches a suitably small level. If such a level were never to be reached, we might be limited in the degree of nodes that we could recover with this technique.
Figure 4.6: Whole network topology recovery performance for 2 SRP solvers for (a) the IEEE Standard-30 Bus, (b) the IEEE Standard-300 Bus over 100 realizations of the network, and (c) the standard IEEE Standard-2383 Bus over 100 realizations of the network.
Greedy-based Algorithms vs LASSO in IEEE Standard-118, and 2383 Bus Networks

Similar to our conclusions over convex optimization algorithms, within the network graph, SRP solvers exhibit different recovery performance on different nodes for a given number of measurements, mainly because of the following two factors.

The first factor is the sparsity level of the signal $y_l$ (or in our PN, the in-degree, or the number of connected links to a bus), and the second factor is the presence of a possible structure in the position of the non-zero components. Figure 4.7, 4.8, 4.9, and 4.10 show the recovery performance of the OMP (green), LASSO (blue) and MCOMP (red) algorithms over a selected set of nodes for 1000 realizations of three IEEE networks used as case studies.

In the first case, the IEEE standard-30 Bus system has been used. In Figure 4.7(a) and 4.7(b), recovery results are illustrated for nodes 3 and 10 (corresponding to the easiest and the hardest cases in the IEEE 30 bus network), respectively. At each specific number of measurements $M$ ($M \in \{4, \ldots, 60\}$), the vertical axis shows the percentage of trials in which perfect recovery is achieved, over 1000 realizations of the network. Nodes 3 and 10 are distinguished from each other by their in-degrees. We see that node 3, with in-degree 2, is more frequently recovered correctly than is node 10, which has in-degree 6. This is to be expected: in the corresponding SRPs, the signal $y_3$ has a sparsity level of 2, while the signal $y_{10}$ has sparsity 6; signals with higher sparsity levels are, in general, more difficult to recover.

Another important factor, which affects the SRP solver’s performance, is the existence of a possible clustered structure in the sparse signals to be recovered. As we have mentioned, MCOMP is an extended version of OMP, which has been adapted to exploit the knowledge that the non-zero entries of the signal appear in clusters.

The structured sparsity in the columns of the matrix $B$ can be used to improve the recovery performance. Figure 4.7 show how the presence of clustered sparsity helps the MCOMP algorithm to outperform the LASSO and OMP, especially in the case of node 10, where $y_{10}$ is a $(6, 4)$-clustered sparse signal and the in-degree of the node (or the number of nonzero elements of the signal to be recovered) is larger.
Figure 4.7 Recovery rate comparison of (a) node 3 and (b) node in the IEEE 30 bus system. These nodes have in-degree 2 and in-degree 6, respectively.

In order to compare the performance over more complicated cases, the IEEE Standard-118 Bus and 2383 Bus systems have also been used for modeling the PN. In the IEEE 118 bus, we focus on recovering the connections to nodes 74 (which has in-degree 2), 49 (in-degree 9), and 59 (in-degree 6). In the IEEE 2383 bus, we focus on nodes 3 (in-degree 3) and 1920 (in-degree 9). From the sparsity level viewpoint, $y_{49}$ and $y_{1920}$ correspond to the most complicated signals $y_i$ to be recovered in the two networks, respectively. Recovery results are again shown over 1000 realizations of the corresponding network.
Figure 4.8. Recovery rate comparison of nodes 74, 59, and 49 in the IEEE 118-Bus, with in-degree 2, 5, and 9 respectively.
Figure 4.8(a)-(c) show the recovery results for the IEEE 118 Bus network. As might be expected, we see that node 74 (with in-degree 2) is more likely to be recovered from a given number of measurements than is node 59 (in-degree 6). Node 49 (with in-degree 9) is the most difficult to recover. These plots also show how MCOMP outperforms LASSO and OMP in the case of node 49, where $y_{49}$ is a (9, 5)-clustered sparse signal and the in-degree of the node (or the number of nonzero elements of the signal to be recovered) is larger. On node 59 (where the corresponding signal $y_{59}$ is a (6, 3)-clustered sparse signal) MCOMP again outperforms OMP and LASSO. On node 74 (where $y_{74}$ has no clustered sparsity), LASSO is the best performing algorithm.

Moreover, Figure 4.9(a) and 4.9(b) are representing the recovery performance of each of 3 SRP solvers for 3-sparse node 3 and 9-sparse node 1920 in IEEE 2383 Bus network. Due to the larger scale of this network, we generally need more measurements to achieve the perfect recovery for any individual node compared to smaller networks like the IEEE 118 Bus. As it has been mentioned before, the number of measurements needed for full recovery depends on both the sparsity level and the original dimension of the signal $y$ (here, the dimension $N$ equals the number of buses in our sparse TI problem); specifically, $M$ must be at least proportional to $S \log(N/S)$. As a rough rule of thumb, in many applications, when the sensing matrix has low coherence, one observes perfect recovery when $M \approx 4S$. However, as the dimension $N$ grows, it becomes necessary to take more measurements. For example, in the case of the IEEE 118 Bus, node 49 has in-degree 9, and perfect recovery is possible when $M \approx 50$. In the case of the IEEE 2383 network, node 1920 again has in-degree 9, but perfect recovery is not possible until $M \approx 80$.

Comparing the results from different size networks number of measurements required for perfect recovery is affected less by the network size than by the sparsity level of the corresponding signal $y_i$ to be recovered. Taking fewer measurements reduces the recovery performance of any SRP solver; however, the LASSO and MCOMP generally show better recovery performance than OMP. Moreover, we cannot recover the nodes with higher in-degrees (such as node 49, where $y_{49}$ is a (9, 5)-clustered sparse signal) until $M$ is large enough that the PN coherence metric reaches a suitably small level, and if such a level were never to be reached, we might be limited in the degree of nodes that we could recover with these techniques.
Figure 4.9 Recovery rate comparison of (a) nodes 3 and (b) 1920 in the IEEE standard 2383 network. Nodes 3 and 1920 have in-degree 3 and 9, respectively.

Finally, in order to highlight the effect of the data correlation issue in addition to the impact of clustered sparsity, we have examined the recovery results for two other clustered sparse signals in the IEEE 118 bus: signal $\mathbf{y}_{37}$ which is a (6, 2)-clustered sparse signal, and signal $\mathbf{y}_{80}$ which is an (8, 4)-clustered sparse signal. In general, as one should expect $\mathbf{y}_{37}$ to be easier to recover compared to $\mathbf{y}_{80}$. In order to evaluate the results, we define the interconnection order ($I_D$) for node $i$ as follows: $I_D(i) = \sum_{j \in N_i} \text{degree}(j)$. 

$$I_D(i) = \sum_{j \in N_i} \text{degree}(j)$$
Node 37 has 6 direct neighbors and $I_O(37) = 17$; node 80 has 8 direct neighbors and $I_O(80) = 22$. The interconnection order can be interpreted as a factor that can reflect the level of data correlation within a specific bus versus the rest of the network; however, it will not be the only factor. Figure 4.10 illustrates how BLOMCOMP outperforms MCOMP, OMP and LASSO in the case of node 80, where we face a more complicated and interconnected local topology and a higher coherence in the corresponding sensing matrix.

Figure 4.10: Recovery rate comparison of nodes (a) 37 and (b) 80 in the IEEE 118-Bus. An initial value of $m = 4$ is chosen in MCOMP and BLOMCOMP.
Figure 4.11 Recovery performance for 3 SRP solvers over all of the buses of the standard IEEE 118 Bus system: (a) OMP, (b) LASSO, (c) MCOMP.
Figure 4.11(a)-(c) illustrate the recovery performance for 3 SRP solvers bus-by-bus (i.e., for each of the columns of the nodal-admittance matrix $B$) over 100 realizations of the IEEE 118-Bus network. The vertical axis represents the number of measurements $M$ taken for sparse recovery of each sparse signal from the following set: \{10, 20, 30, 40, 50\}. The horizontal axis shows the bus number from 1 to 118, or equivalently, the column number from the nodal-admittance matrix. The recovery performance is illustrated by a color spectrum ranging from dark blue (corresponding to 0% recovery) up to dark red (corresponding to 100% recovery). As can be seen, almost all the algorithms can reach 100% recovery performance with 50 measurements. However, LASSO and MCOMP generally exhibit better performance when the number of measurements is lower, and MCOMP generally has better performance than LASSO, especially in the case of clustered sparse signals. Table 4.1 compares the recovery performance of 3 SRP solvers using 30 measurements for some of the most significant structured columns in the IEEE 118-Bus system.

Finally, Figure 4.12 illustrates the whole network topology recovery performance for 3 SRP solvers for the standard IEEE 118 Bus system, over 1000 realizations of the network. The vertical axis shows the percentage of trials in which all 118 columns of the nodal-admittance matrix $B$ (and, as a result, the whole network topology) are perfectly recovered. This figure also shows that the whole network topology can typically be recovered from 50 or fewer measurements per bus.

<table>
<thead>
<tr>
<th>Bus (Column) #</th>
<th>OMP (%)</th>
<th>LASSO (%)</th>
<th>MCOMP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>90</td>
<td>91</td>
<td>100</td>
</tr>
<tr>
<td>11</td>
<td>92</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>12</td>
<td>52</td>
<td>84</td>
<td>51</td>
</tr>
<tr>
<td>15</td>
<td>85</td>
<td>93</td>
<td>100</td>
</tr>
<tr>
<td>17</td>
<td>74</td>
<td>71</td>
<td>88</td>
</tr>
<tr>
<td>56</td>
<td>89</td>
<td>92</td>
<td>100</td>
</tr>
<tr>
<td>62</td>
<td>92</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>85</td>
<td>74</td>
<td>92</td>
<td>99</td>
</tr>
<tr>
<td>100</td>
<td>31</td>
<td>32</td>
<td>78</td>
</tr>
<tr>
<td>105</td>
<td>90</td>
<td>90</td>
<td>100</td>
</tr>
</tbody>
</table>
Figure 4.12. Whole network topology recovery performance for 3 SRP solvers for the standard IEEE 118 Bus system over 1000 realizations of the network.
CHAPTER 5
MODELING AND ANALYSING THE DYNAMIC
BEHAVIOUR OF SMART GRIDS USING STRUCTURED SPARSITY

In this chapter\textsuperscript{22} we introduce an alternative structured sparse recovery-based mathematical reformulation for power network dynamic behavior modeling and tracking using the wave propagation analysis in transmission lines.

5.1 Introduction and Literature Review

Regarding the high mathematical modeling complexity, network dynamic behavior modeling, and transmission line dynamic parameter estimation is classified as an intricate monitoring problem in SGs [58]-[59]. Climate situations and ambient temperature are continuously affecting and changing transmission line parameters, or in general their behavior. As a result, developing a real-time (dynamic) line monitoring methodology, is a crucial challenge in the future SGs [58]. Due to the high analyzing time cost, lack of online measurement accessibility, and involute mathematical modeling, most of the conventional methods are working based on the geometry of line conductors or exploiting offline measurements through offline-based methodologies [58]. Moreover, most of the calculation techniques provided to measure the parameters are adapted for de-energized lines. However, the advancement of power system sensing technologies has created the opportunity for online estimation and tracking of line behavior when a line is energized as well [59].

Nevertheless, the typical sampling rate of the well-known Supervisory Control and Data Acquisition (SCADA) system is still not enough for capturing the fast dynamic events in online monitoring applications. In order to analyze and track the dynamic operation of PNs, phasor measurement units (PMU) were introduced in the mid-1980s\textsuperscript{23}. Due to high sampling rate recordings, PMU technology is well suited for many online monitoring and analysis of dynamic events in PNs [60]. Specifically, the synchronized phasor data, i.e. the PMU data, has been used for transmission line behavior tracking and parameter estimation in [58]-[59], [61]-[68] and the references therein.

\textsuperscript{22} This section is based on the following set of papers: [96].

\textsuperscript{23} Nowadays vast amount of work is undergoing in power sensor technology especially on high frequency PMU technology. In near future this technology will change from a fully hardware structure to a software-based technology with much lower cost of units. Low cost PMU architecture in addition to virtualizing the PMUs have been addressed in [28]-[29] and the references therein.
Recently, a SCADA measurement-based approach has been adapted in [69], however, most of the modern methods are taking advantage of the PMU technology in their frameworks [58]-[68]. In [61] an idea has been developed based on the implementation of distributed line model for estimating transmission-line positive-sequence parameters incorporating nonlinear estimation theory. Authors in [62], have addressed an algorithm using moving-window total least squares, however, the treating parameters have been considered to remain constant during the estimation time window. In [63] the effect of bias and none-bias noises of PMU records have been discussed on the estimation accuracy for short transmission-lines. Moreover, Bi et.al investigated the effect of conductor sag on the line behavior in [64]. In another try, Yang et.al addressed a recursive least-square algorithm for estimating the impedance matrix for a power network [65]. Also in [66], a Newton–Raphson solution-based approach has been proposed. In [67] the actual line resistance has been calculated by incorporating the actual line temperature and reference resistance through a recursive extended Kalman filter framework. Recently, a frequency-dependent approach has been addressed in [68] for transmission line behavior tracking and parameter estimation. More recently, a combined data approach has been developed in [58] that utilized both the SCADA and PMU data in order to model the line parameter behavior problem.

The majority of the aforementioned approaches are suffering from the drawbacks which are enforced by the nonlinearity and mathematical complexity of the presented formulation. In case of large scale PNs with thousands of lines, this nonlinearity may cause serious timing issues in the online network-wide monitoring applications. As a result, the performance of most of these methods have been only reported for individual lines. On the other hand, most of these methods require additional and perquisite information, calculations and assumptions to be satisfied, in order to have an accurate modeling and estimation of the line behavior. Within most of these approaches, the exact length of the line, \( l \), is needed to be a known factor, where the accuracy of the line length measurement should be in a level that the line propagation constant can be calculated with a high resolution [59]. However, the line length can itself be affected by climate situations and the ambient temperature that results in a change in the line parameter values [58]. Thus, considering this value as a constant known factor could be an unrealistic assumption in practical situations. Besides, most of the available literatures are focusing on \( \pi \) model for long length transmission lines, although, some other work have been adapted for medium/short length lines and for distribution lines as well [69]. However, most of these approaches have been developed based on
typical line models which are basically adapted for steady state conditions and may not be validated in case of dynamic behavior modeling [24].

In this chapter, our main goal is to introduce a new dynamic framework for real-time monitoring and tracking the transmission lines behavior in the entire PN. We introduce a mathematical formulation for transmission line dynamic behavior modeling using detailed dynamic differential equation model of the current wave propagation in the transmission lines. Recently, it has been suggested in [25], that a power system can be modelled as a collection of port-Hamiltonian sub-systems. Combining the current wave propagation analysis with an understanding of this port-Hamiltonian model we define a new mathematical formulation for the transmission line dynamic behavior modeling and tracking. Like our previous approaches in Chapters 3 and 4, in this part the PN is modeled as an interconnected graph, where each bus is modeled as a node and each transmission line as an edge. Exploiting the sparse nature of the PN graphs, in addition to CS-SR techniques, we introduce a new structured sparse recovery-based (SSRP) mathematical formulation for the smart grid dynamic behavior monitoring (SGDBM) that, due to its linear programming format, can be solved within entire network lines over a considerably short amount of time. For sake of notational simplicity, we abbreviate this problem as SGDBM-SSRP.

This SGDBM-SSRP introduces a new dynamic monitoring framework to track the transmission line behavior in SGs through defining a new intermediate parameter, termed as the line dynamic index coefficient (LDIC). Finally, this SGDBM-SSRP is solved for LDIC vectors, using the extended greedy sparse recovery methods, named Block Orthogonal Matching Pursuit (BOMP) [34] and COMP. The simulation results over IEEE standard testbeds indicate that the presented framework can be developed under a fast and accurate fashion among the PN.

5.2 Power Network Modeling and Current Wave Dynamic Analysis

5.2.1 Traveling Wave Analysis in Transmission Lines

Before presenting the SGDBM-SSRP formulation, we need to mathematically model the transmission line parameters dynamic behavior by using electrical current wave propagation through a transmission line. We start with the single-phase two-wire lossless transmission line model with source and load terminals as shown in Figure 5.1(a), where \( Z_R(s) \) is the impedance at
the receiving end, \( Z_C(s) \) is the characteristic impedance of the line, and the source bus is modeled with its Norton equivalent current source (\( I_N(s) \)) and Norton equivalent impedance (\( Z_N(s) \)).

A Section of length \( \Delta x \) of the line, including its parametrical specifications, has been illustrated in Figure 5.1(b), where \( C \) represents the parallel capacitor and \( L \) stands for the line inductance. It has been shown in [24-please refer to Chapter 13]\textsuperscript{24}, that under dynamic modeling conditions (considering time as an effective argument in the transmission line behavior modeling), by implementing the KCL and KVL equations for the circuit in Figure 5.1(b), in addition to Laplace transform properties, we arrive at the following equation for each point \( x \) along the transmission line:

\[
\frac{d^2 i(x,s)}{dx^2} - s^2 LC I(x,s) = 0
\]  

\textsuperscript{24} Inspired by a similar voltage analysis in [20], and for sake of simplicity we illustrate the mathematical analysis on lossless line model, however, a similar approach can be taken for lossy models where one may conclude with a similar final result as also mentioned in the same reference [20 Ch.13].
solving this 2nd order ODE equation we have

\[ I(x, s) = I^+(s)e^{-sx/\theta} + I^-(s)e^{sx/\theta}. \]  

(5.2-1)

Similarly for the voltage we have:

\[ V(x, s) = V^+(s)e^{-sx/\theta} + V^-(s)e^{sx/\theta}, \]  

(5.2-2)

where \( \theta = \frac{1}{\sqrt{(LC)}} \) m/s is about the light speed in the line and \( I^+, I^-, V^+, V^- \), are calculated from boundary conditions. Using (5.2) and considering (from KVL and KCL equations):

\[ \frac{dV(x,s)}{dx} = -sLI(x, s) \]  

(5.3)

we can accordingly find out:

\[ \frac{s}{\theta} \left[ -V^+(s)e^{-sx/\theta} + V^-(s)e^{sx/\theta} \right] = -sL[I^+(s)e^{-sx/\theta} + I^-(s)e^{sx/\theta}] \]  

(5.4)

equating the coefficients \( e^{sx/\theta} \) and \( e^{-sx/\theta} \) on the both side of (5.4) we can reach to the following equivalent boundary conditions for voltage in terms of current:

\[
\begin{cases} 
V^+(s) = \theta LI^+(s) = I^+(s)\sqrt{\frac{L}{C}} = I^+(s)Z_c, \\
V^-(s) = -I^-(s)Z_c
\end{cases}
\]  

(5.5)

where \( Z_c = \sqrt{\frac{L}{C}} \Omega \), is the characteristic impedance of the line. Finally, we will have the following equivalent equation for voltage (5.2-2):

\[ V(x, s) = Z_c[I^+(s)e^{sx/\theta} - I^-(s)e^{-sx/\theta}]. \]  

(5.6)
Using the general model in Figure 5.1(a) and applying boundary conditions at the receiving end (5.7), we will have:

\[ V(l, s) = Z_R(s)I(l, s) \]  

(5.7)

where \( Z_R(s) \) is the impedance at the receiving end and \( l \) is the line length. Applying (5.2-1) and (5.6) in (5.7):

\[
Z_c \left[ I^+(s)e^{-\frac{sl}{\sigma}} - I^-(s)e^{\frac{sl}{\sigma}} \right] = Z_R(s) \left[ I^+(s)e^{-\frac{sl}{\sigma}} + I^-(s)e^{\frac{sl}{\sigma}} \right],
\]

(5.8)

solving for \( I^-(l, s) \) we have the following boundary condition at the receiving end: \( V(l, s) = Z_R(s)I(l, s) \); where \( l \) is the line length. Applying (5.1) in this condition and solving for \( I^-(l, s) \) we will have:

\[
\begin{aligned}
I^-(l, s) &= \Omega_R(s)I^+(s)e^{-2s\tau} \\
\Omega_R(s) &= \frac{Z_c(\frac{Z_R}{Z_R^{(b)}} - 1)}{Z_c + 1} \quad \text{p. u.}
\end{aligned}
\]

(5.9)

where the line’s time constant is \( \tau = \frac{l}{\sigma} \). Using (5.2), (5.6) and (5.9) we have:

\[
\begin{aligned}
I(x, s) &= I^+(s)\left[ e^{-\frac{sx}{\sigma}} + \Omega_R(s)e^{\frac{x}{\sigma} - 2\tau} \right] \\
V(x, s) &= I^+(s)Z_c \left[ e^{-\frac{sx}{\sigma}} - \Omega_R(s)e^{\frac{x}{\sigma} - 2\tau} \right].
\end{aligned}
\]

(5.10)

Using the KCL rule in the sending end, we can reach the following boundary condition: \( I_N(s) - \frac{V(0, s)}{Z_N(s)} = I(0, s) \). Using (5.10) in this boundary condition, we arrive at:
\[
I_N(s) = I^+(s) \left[ 1 + \frac{Z_c}{Z_N(s)} \right] [1 - \Omega_R(s) \Omega_U(s) e^{-2st}] \\
\Omega_U(s) = \frac{\left( \frac{Z_c}{Z_N(s)^{-1}} \right)}{\left( \frac{Z_c}{Z_N(s)^{+1}} \right)}.
\]

Finally, applying (5.9) and (5.11) in (5.2-1) results in:

\[
I(x, s) = I_N(s) \left[ \frac{Z_N(s)}{Z_N(s) + Z_c} \right] \left[ e^{\frac{sx}{\varpi + \Omega_R(s)} - \Omega_U(s) e^{\frac{s(x - 2t)}{\varpi}} - \Omega_U(s)} \right].
\]

Thus, the time domain representation of the current wave equation at point \( x \) along the transmission line is determined by the inverse Laplace transform of (5.12). Based on the involved impedance values, the corresponding time domain representation would take different forms at the receiving bus \( (x = l) \). However, applying series expansion to factorize (5.12) in addition to considering the actual conditions and estimations from power system perspective, the general time domain representation can be approximated and expressed in terms of the following superposition format (please also refer to [24- Chapter 13], for more info and several practical examples over a similar formulation for voltage wave propagation analysis):

\[
i(l, t) = \sum_{k=1}^{K} x_k i_N(t - \alpha_k \tau);
\]

We term the sequence of coefficients, \( x = \{ x_k \text{ for } k = 1, \ldots, K \} \), as line dynamic index coefficients vector. Evidently, these index coefficients are associated to the dynamic behavior of the line parameters, and network impedances. The total number of contained delayed terms, \( K \), that emerges within the measurements, pertains to the recording sampling rate and time constant of the line, which itself mainly depends on the line length. To capture the effect of these dynamic delayed terms in the measurements, the sampling frequency of the employed sensor should be comparable to the line time constant, \( \tau \). Nowadays, the common sampling rate of industrial PMUs is around ‘1-10’ kilo samples/second, which means the dynamic phenomena with time constant of order around \( \tau \geq \frac{3.47}{2} \times 10^{-4} \text{ s} \), can be recorded sans aliasing. Assuming \( \vartheta \) to be approximately equal to the light speed, and inserting the above bound on the \( \tau \)’s value in \( \tau = \frac{l}{\vartheta} \), we discover that
under this sampling rate, the minimum line length, \( l \), that still can illustrate such a superposition behavior (5.13) in its current measurements, is approximately 35 km; which is almost around the lower bound of the standard small size transmission lines [24]. Therefore, currently in most of the PMU records among the power grid these dynamic effects can already be recorded. Moreover, using new technologies of Micro-PMUs with sampling rates of about 100 kHz - 1 MHz [70]-[71], one is able to capture these effects in the measurements from much shorter lines even within distribution level scale. In the following section, we describe how this superposition format (5.13) helps us to present a new framework for SGDBM.

5.2.2 Power Network Modeling under Dynamic Behavior

In the rest of this chapter we assume a PN as a typical graph, consisting of a set of \( N \) nodes, where each node \( i \) represents a bus and a set of edges representing the transmission lines, \( \{ l_{(i,j)} | i, j \in \{1,...,N\} \} \). We call a bus a generator/load bus if a generator is connected to it/or if it is connected to a load. Following the same approach in [25], we do not account for the interior buses in this study.

![Figure 5.2 A typical super-node/bus model under PN KCL law.](image)

Inspired by the KCL interconnection law of the power grid, which has been discussed in [25], in this work each bus is modeled as a node with the configuration that has been illustrated in Figure 5.2. In other words, the input-output current relation of node \( i \) is considered as follows:

\[
I_i(t) = \sum_{p=1}^{P} I_{i,p}(t) = \sum_{j \in L_i} z_{ji}(t-1) + b_i(t) + c_{ji}(t),
\]

(5.14)
where $z_{ji}$ is the input current wave from each individual bus $j$ to bus $i$ and, $P$ indicates the total number of output terminals from bus $i$. Based on the bus class (load-generator), $b_i$ represents the generator or load currents, arriving/injecting at/from the bus $i$; the $c_{ji}$ is the conforming current to the capacitor of the line $l_{ji}$ connected to the bus $i$; finally, $L_i$, stands for the set of indexes of all of the buses in the PN which are physically connected to the bus $i$. The cardinality of $L_i$ is termed as the degree of the bus $i$.

5.2.3 Mathematical Formulation of SGDBM

Using equation (5.13), the receiving current waves $z_{ij}$’s at bus $i$ can be represented as the following superposition of $K$ delayed terms of the sending current wave, $I_{ji}$, at the sending bus $j$:

$$z_{ji}(t) = \sum_{k=1}^{K} x_{ji,k} l_{ji}(t - k + 1),$$

(5.15)

We initially consider that all of the line lengths are within a certain range. This assumption results in an ideal situation where all of the lines will follow the same total number of time delay terms, we term this number as the line dynamic order (LDO). We investigate how one can address the challenge of variable LDO, using clustered sparsity technique in our case studies. From (5.12) and (5.13), the LDIC values, $x_{ji,k}$, mainly depend on the dynamic behavior of source, sink and characteristic impedances of the line $l_{ji}$ as well as the source bus current signal behavior, thus, real time estimation/tracking of the values of these LDICs, and the study of their behavior under different PN conditions can be interpreted as an alternative approach for tracking the dynamic behavior of the PN. In what comes next, we describe how the LDICs estimation problem can be expressed in terms of an optimization problem by exploiting the measurements of the $p.u$ PMU current waves. Replacing (5.15) in (5.14), we have

$$I_i(t) = \sum_{j\in L_i} \sum_{k=1}^{K} x_{ji,k} l_{ji}(t - k) + b_i(t) + c_{ij}(t).$$

(5.16)

\[\text{For ease of expansion we will use the discrete sample-index notation of the delay terms, so } \alpha_k \tau = k.\]
For time samples, \( t = 1, ..., M \), setting \( x_{ji} = 0 \) for \( j \notin L_i \), assuming \( l_i(t) = 0 \) for \( t \leq 0 \) and considering no feed-through term, \( z_{ji}(0) = 0 \), then per equation (5.16), the output of each node \( I_i \in \mathbb{R}^M \) can be written as

\[
I_i = \sum_{j=1}^{N} A_{ji} x_{ji} + b_i + c_i \quad i = 1, 2, ..., N, \tag{5.17}
\]

where \( N \) is the total number of buses in the PN, \( A_{ji} \) is an \( M \times K \) Toeplitz matrix, \( x_{ji} \in \mathbb{R}^K \) and \( b_i, c_i \in \mathbb{R}^M \). For each single node \( i \), equation (5.17) can be rewritten in the following matrix-vector format:

\[
I_i = A^i x^i + b_i + c_i \tag{5.18}
\]

where \( A^i = [A_{1i} \ldots A_{ji} \ldots A_{Ni}] \), \( x^i = [x_{1i} \ldots x_{ji} \ldots x_{Ni}]^T \), \( I_i \in \mathbb{R}^M \), \( x^i \in \mathbb{R}^{NK} \), and \( A^i \in \mathbb{R}^{M \times NK} \) (called the sensing or measurement matrix) is formed by the concatenation of \( N \) Toeplitz matrices.

As we discussed before, the SGDBM problem can be interpreted as calculating the set of dynamic vectors \( \{ I_i \}_{i=1}^N \), from a given set of PMU current wave measurements at each electrical bus in the PN with a proper sampling rate. We consider the available set of measurements to include all the output currents \( \{ I_i \}_{i=1}^N \) and the known current terms \( \{ b_i \}_{i=1}^N \). Roughly speaking, since the capacitor currents are usually much smaller than other currents and are unknown, we model the \( c_i \) as a vector of random variables with a Gaussian distribution. Moreover, replacing \( I_i - b_i = y_i \), and \( c_i = \eta_i \) for each individual bus we end up with \( y_i = A^i x^i + \eta_i \). As a result, the SGDBM can be defined as the estimation of the index coefficient sub-vectors \( \hat{b}_{ji} \) that best matches the observed measurements. The following optimization problem can be used to solve such an estimation problem:

\[
\min_{\{x^i\}_{i=1}^N} \sum_{i=1}^{N} \| A^i x^i - y_i \|_2^2. \tag{5.19}
\]
This problem can be solved individually over each bus \( i \). However, due to structural similarity in the matrix \( A^i \) for all buses, without loss of generality, we simplify the index-notation \( A^i = A \), \( x^i = x \), \( y^i = y \) and focus on the following general optimization problem:

\[
\min_{\hat{x}} \|A\hat{x} - y\|_2^2.
\]  
(5.20)

Defining \( L = N \times K \), we have \( y \in \mathbb{R}^M \), \( x \in \mathbb{R}^L \), and \( A \in \mathbb{R}^{M \times L} \). From algebraic viewpoint, if the sensing matrix \( A \) is full rank, the number of measurements, \( M \), exceeds the number of unknown parameters, \( L \), and the noise variance, \( \eta \), is fairly small, this optimization problem can be solved by exploiting the Least Square (LS) methods. Based on our problem formulation, the value of \( L \) depends on the number of buses in the network in addition to the dynamic order of lines. Thus, in the case of large scale PNs one may need a large number of measurements to solve such a problem. This will result in a serious time and complexity issue within dynamic behavior monitoring applications. To avoid such an unwanted subject, we suggest using the sparse recovery techniques.

5.3 Sparse Structured Dynamic Modeling

In this section, we will show how the SGDBM problem (5.20) can be represented as a mathematical structured sparse recovery problem. In the following, also a brief overview on the related compressive sensing concepts and structured sparse recovery techniques is given (Although some of the following descriptions may have an overlap with the general CS-SR notation we introduced earlier in Chapter 2, we found this repetition mandatory to describe the similar concepts in terms of Block-sparse recovery).

5.3.1 Sparse Dynamic Modeling

Regarding (5.19), the corresponding dynamic coefficient vector \( x \), over each bus, can be considered as a concatenation of \( N \) blocks as follows:

\[
x = [x_1^T, x_2^T, ..., x_N^T]^T.
\]  
(5.21)
The sub-block $\mathbf{x}_j^T = \mathbf{0} \in \mathbb{R}^K$ if, within the network structure, bus $i$ is not physically connected to the bus $j$. Now, consider $S = ||L_i||_0$, to be defined as the cardinality of the index set of neighbor buses of bus $i$ which are directly connected to this bus. The corresponding vector $\mathbf{x}^i$, is called $S$-block-sparse if $S \times K << L$. Such a vector would contain $S$ non-zero block with each of length $K$ and, the total number of nonzero elements is signed as $||\mathbf{x}||_0 = S \times K = F$, and named as the sparsity level. Regarding our formulation, the number of nonzero blocks in the LDIC vector $\mathbf{x}$ is limited to the number of physically connected neighbor buses, that is always a small fraction of total buses in the PN (refer to our discussions in Chapter 1 and 2). As a result, $\mathbf{x}$ can be represented as an $S$-block-sparse vector over all of the buses. Obviously, whenever $M < L$ in (5.20), due to non-trivial null space of the sensing matrix $A$, we face with infinitely many candidate solutions. However, it has been shown in the literature that within certain recovery guarantee conditions on matrix $A$ (such as RIP and mutual coherence - See [15]-[16] for more info), there exists modified CS-SR based recovery techniques (such as Block OMP) which are able to efficiently find the true block-sparse candidate solution [34]. In general, we can interpret the CS-SR recovery problem as recovery of a $F$-sparse signal $x \in \mathbb{R}^F$ ($||x||_0 = F$) from a set of measurements $y = A\mathbf{x} \in \mathbb{R}^M$ where $A \in \mathbb{R}^{M \times L}$ with $M << L$. Whenever the sparse signal to be recovered follows a special structure such as block-sparse signals, the problem is termed as the structured sparse recovery problem (SSRP).

5.3.2 Block OMP Recovery Algorithm for Block Sparse Signals

A variety of extended greedy sparse recovery algorithms have been proposed to consider the additional information regarding the structure in the sparse vector to be recovered. In between, the Block OMP (BOMP) is especially introduced to exploit the block-structured sparsity [34]. Due to block sparse nature of the optimization variable $\mathbf{x}$ in SGDBM (20), we use BOMP (Algorithm 6) as the desirable recovery method in this work and compare its performance vs OMP.

5.3.3 BOMP Guarantee Condition

**Definition 5.22:** The block-coherence of matrix $A$ (with normalized columns) is defined as:

$$\mu_{\text{block}}(A) = \max_{i,j \neq i} \frac{1}{K} ||A_i^T A_j||_2.$$  (5.23)
Definition 5.24: The sub-block-coherence of matrix $A$ (with normalized columns) is defined as:

$$
\mu_{\text{sub-block}}(A) = \max_n \max_{i,j \neq l} |a_{ni}^T a_{nj}|,
$$

(5.25)

where $a_{ni}$ and $a_{nj}$ are the columns of $n^{th}$ block of matrix $A$ or $A_n$ (5.17). A sufficient condition has been provided in [34] to guarantee the recovery of any $S$-block sparse signal $x$ from the set of measurements via BOMP method.

Algorithm 6 The Block Orthogonal Matching Pursuit (BOMP)

require: matrix $A$, measurements $y$, stopping criterion
initialize: $r^0 = y, x^0 = 0, l = 0, SUP = \emptyset$
repeat
1. match: $h_j^l = A_j^T r^l \; \text{for} \; j = 1, \ldots, N$
2. identify support indicator: $\mathcal{I}^l = \{ \arg \max_j \| h_j^l \|_2 \}$
3. update the support: $SUP^{l+1} = SUP^l \cup \mathcal{I}^l$
4. update signal estimate: $x^{l+1} = \arg \min_{x: SUP(x) \subseteq SUP^{l+1}} \| y - Ax^{l+1} \|_2$
5. update the residual vector: $r^{l+1} = y - Ax^{l+1}$
6. increase index $l$ by 1: $l = l + 1$
Until stopping criterion met
Output: $\hat{x} = x^l$

Theorem 5.26: [34] For an $S$-block-sparse vector $x$ with $S$ blocks of length $K$, and a vector of measurements $y$ such that $y = Ax$, if $S \times K < \mu_T$, then the BOMP can recover original vector $x$ from the set of measurements $y$, where

$$
\mu_T = \frac{1}{2} \left( \frac{1}{\mu_{\text{block}}} + K - (K - 1) \frac{\mu_{\text{sub-block}}}{\mu_{\text{block}}} \right).
$$

(5.27)

In case of the noisy measurements, $y = Ax + n$, the arguments can be extended to the robust recovery in noisy settings. While $S \times K < \mu_T$ should be satisfied as the guarantee condition, the maximum block-sparsity level $S$ of the signal, that can be recovered using BOMP, is limited by
the block-coherence of the sensing matrix $A$. The higher the block-coherence metrics, the smaller the permitted value of $S$, and the broader the sparsity level of the signal that can be recovered using BOMP [34].

5.3.4 Line Dynamic Order Variation and Clustered Sparsity

In this work, we initially assumed that all the line lengths are perchng within a certain range that results in the same dynamic order (LDO). If dynamic orders of all lines ($l_{(i,j)}$ for ($j \in L_i$)) connected to the bus $i$, or equivalently the number of affective delayed terms of $I_{ji}(t - k)$ in $I_i(t)$ is fixed and equal to $K$, sparse vector $x$ is a $S$-block sparse signal with $S$ non-zero blocks of length $K$. However, since line physical properties can vary in a big range in power networks this assumption can be dilapidated in the realistic situations. This means that each non-zero block $x_{ji}$ (of the sparse LDIC vector $x^i$) can have an arbitrary number of nonzero elements. From sparse recovery viewpoint, the corresponding sparse signal $x^i$ changes its structure from a block-sparse to a clustered-sparse vector.

As it has been mentioned before, in another generalized version of OMP (named COMP), the support identification step has been modified to be adapted to situations that we expect a clustered sparsity pattern to be happened in the support of the signal to be recovered (please refer to Chapter 3). The recovery performance of this algorithm (MCOMP Algorithm 3) will be compared vs BOMP in case of arbitrary dynamic orders in the SGDBM problem.

5.3.5 Data Correlation Issue

As it can be seen from (5.19), the structure of the corresponding sensing matrix $A$, is formed by the measurements from transmission lines electrical flows. Since the overall behavior of loads in neighboring areas is expected to be similar to each other, it is highly likely to have a meaningful correlation within the current behavior in neighbor lines. This correlation itself can result in a higher coherence in the corresponding sensing matrices within the SGDBM-SSRP. To deal with this correlation issue, a bound excluded locally optimized approach is suggested to address the high coherence issue. Since in SGDBM-SSRP we face with block-sparse signals we develop a BLO-based block orthogonal greedy algorithm termed as Band-excluded Locally Optimized Block Orthogonal Matching Pursuit (BLOBOMP). The BLOBOMP is summarized in Algorithm 7.
5.3.6 Band-excluded Locally Optimized BOMP (BLOBOMP)\(^\text{26}\)

In this section, we try to extend the program laid out in Faninjiang and Liao [22] to the block-sparse setting. It has been shown in the literature that, there are practical scenarios (such as SGDBM) that involve vectors with nonzero entries rising in block-structured positions in the sparse signal \(x\). Combining the idea of BLO technique to the Block-OMP settings, we introduce a new method (Algorithm 7) to deal with the highly coherent block-sparse recovery problems.

### Algorithm 7. BLOBOMP

```
require: matrix \(A\), measurements \(y\), stopping criterion, \(\eta_b\), block-size \(K\)
initialize: \(r^0 = y, x^0 = 0, l = 0, \text{SUP}^0 = \emptyset\)
repeat
  1. match: \(h_j^l = A_j^T y^l\) for \(j = 1, ..., N\)
  2. identify support indicator: \(\text{sup}^l = \{\arg\max_j \|h_j^l\|_2\} \& \Delta \notin \beta_{\eta_b}^{\text{SUP}^{l+1}(2)}\)
  3. update the support: \(\text{SUP}^{l+1} = \text{LO}(\text{SUP}^l \cup \text{sup}^l)\)
  4. update signal estimate: \(x^{l+1} = \arg\min_{x, \text{SUP}^l(\text{SUP}^l \cup \text{SUP}^{l+1})} \|y - Ax^{l+1}\|_2\)
  5. update the residual vector: \(r^{l+1} = y - Ax^{l+1}\), adding step no by 1: \(l = l + 1\)
Until stopping criterion met
Output: \(\hat{x} = x^l\)
```

Local Optimization Procedure for block-sparse signals

```
require: \(A, x\), Coherence band \(\eta_b > 0\), \(\text{SUP}^0 = \cup \{\text{sup}^l\}\) for \(l = 1: k\)
repeat: for \(l = 1: k\)
  1. \(x^l = \arg\min_{x, \text{sup}(x) = (\text{sup}^{l-1} \setminus \text{sup}^l) \cup \{x\}, \lambda \in \beta_{\eta_b}^{\text{SUP}^l}} \|y - Az\|_2\)
  2. \(\text{SUP}^l = \text{sup}(x^l)\)
output: \(\text{SUP}^k\)
```

Here we present the related definitions we use in order to implement this algorithm. Assume the \(\eta_b > 0\), a constant value, we define \(\eta_b\)-block-coherence band of the block index \(\lambda\) to be the set:

\[
\beta_{\eta_b}^\lambda = \{i|\mu_b(i, \lambda) > \eta_b\} \quad \text{and} \quad \mu_b(i, j) = \frac{1}{K} \|A_i^T A_j\|_2, \quad (5.28)
\]

define \(\eta_b\)-block-coherence band of the block index set \(\Lambda\) to be the set:

\[
\beta_{\eta_b}^\Lambda = \bigcup_{\lambda \in \Lambda} \beta_{\eta_b}^\lambda. \quad (5.29)
\]

Moreover, denote “double \(\eta_b\)-block-coherence band” of the block index \(\lambda\) and the block index set \(\Lambda\) as follows:

---

\(^{26}\) Since this algorithm is not a direct contribution of this work, (inspired by the BLO technique, developed in [22]) in this thesis we will only suggest a general approach that should be developed including required mathematical definitions, theorems and the general greedy format for this sparse recovery solver. Full mathematical details and proofs will be provided in our future work.

85
\begin{align}
\beta_{\eta_b}^{(2)} &= \beta_{\eta_b}^\lambda (\beta_{\eta_b}^\lambda) = \mathcal{U}_{j \in \eta_b} \beta_{\eta_b}^j \\
\beta_{\eta_b}^{(2)} &= \beta_{\eta_b}^\lambda (\beta_{\eta_b}^\lambda) = \mathcal{U}_{\lambda \in \eta_b} \beta_{\eta_b}^{(2)}
\end{align}

5.4 Simulations and Discussion

In this section, we test the proposed method for recovering the LDIC vectors $x^i$ (modeling the overall dynamic behavior of the PN) utilizing compressive observations. In the following set of simulations, we use the structure of the IEEE 30 and 118 Bus networks as case studies and generate the data through a Monte Carlo approach. Based on recent PMU standards, the measurements SNR is set to 20-100 $\text{db}$, randomly. The aforementioned PNs include 30, 118 Buses, 47, 186 transmission lines respectively [17].

**Constant Line Dynamic Order**

Regarding casualties in physical systems, the amplitude of the dynamic coefficients is subject to fast decay. Thus, in simulations the dynamic orders $K = 2$ & $4$ are considered as reasonable choices. The relevant coherence measures of IEEE 30 Bus system (vs number of measurements) are illustrated in Figure 5.4 for $K = 4$ (curves are averaged over 100 realizations of the network). Consequently, the coherence measures are asymptotically decreased when measurement number increases. In another word, LDIC vectors with higher number of non-zero blocks can be recovered from more number of measurements. As a rule of thumb, within the proposed SGDBM-SSRP the BOMP solver has different recovery performance over different buses. The differences are mainly related to the sparsity level of the corresponding LDIC vector. As mentioned before, within the SGDBM-SSRP the sparsity level of the vector $x^i$ is forced by the degree of the bus, $\|L_i\|_0$ and the LDO of the connected transmission lines, $K$.

Figure 5.3, illustrates the recovery performance for the following set of selected buses: 74, with degree 2, 49, with degree 9, and 59, with degree 6 within the IEEE-118 Bus network (for LDO = 4). This set of specific buses were selected since they represent the lowest, highest and a moderate degree level for an individual bus in the IEEE-118 Bus system. Since the LDO is assumed to be fixed in this case the probability of successful recovery changes based on the degree of the bus. Consequently, as one might expect, the double in-connection bus 74 is more likely to be successfully recovered utilizing a smaller number of measurements than the 6 in-connection bus
87

59. Same argument is valid comparing the 6-degree bus 59 vs 9-degree bus 49, however, the
difference is not considerable in this case. Hence, there should exist other affective factors which
are effecting the overall recovery performance.

Figure 5.3 Recovery performance of the BOMP for buses: 74, degree 2, 49, degree 9, and 59,
degree 6, within the IEEE-118 Bus network (for LDO = 4) over 100 realization of the network.

Figure 5.4 Coherence measures of IEEE 30 Bus vs number of measurements (curves are
averaged over 100 realizations of the network).
In the SGDBM-SSRP formulation, the sensing matrix $A$ is shaped by the current wave measurements (5.18). If the temporal correlation among the incoming currents $I_{ji}$ (for $j \in L_i$) to bus $i$ is high, the correlation among corresponding block-columns $A_{ji}$ of the matrix $A$ increases, consequently, and we will face with higher coherence metrics. On the other hand, guarantee conditions of BOMP indicates that this high coherence decreases the recovery performance. This temporal correlation may especially happen in neighbor buses due to similar load and weather conditions within neighbor urban or rural areas.

Another important factor that can affect the sensing matrix specifications is the local structure of the corresponding graph. As a conclusion, in addition to sparsity level, graph topology and temporal correlation in neighbor areas are playing important roles in the final recovery performance of the proposed sparse modeling technique. We will address these challenges within our future work.

Figure 5.5 illustrates a network-wide evaluation over recovery performance of the OMP and BOMP algorithms vs bus number for IEEE-30 Bus network with dynamic orders $K = 2$ and $4$.\(^{27}\) The recovery performance of each sparse signal $x^i$ for $i = 1: 30$, has been illustrated by a color spectrum from dark blue corresponding to 0% recovery up to dark red corresponding to 100% recovery vs selected number of measurements over 100 realization of the network.

Although BOMP illustrates a better performance in general, but due to the lower sparsity level, both OMP and BOMP can achieve to an acceptable recovery with small number of measurements for LDO, $K = 2$. In contrast, exploiting the structured sparsity pattern, BOMP considerably outperforms the OMP in the case of LDO, $K = 4$.

**Variable Line Dynamic Order**

In the next set of simulations, and in order to investigate the corresponding challenges regarding the variable LDOs, we modeled the network taking different LDOs for different lines. In this case, a random LDO value between 1 to 4 has been randomly selected and assigned to each individual line. Consequently, the sparse LDIC vector $x^i$ changes its structure from a block-sparse signal into a clustered-sparse signal with maximum cluster size of 4.

\(^{27}\) The vertical axis represents the number of measurements taken for sparse recovery of each sparse signal $x^i$ from the following set: {5:5:100}. The horizontal axis shows the bus number from 1 to 30.
Figure 5.5: Recovery performance for (a), (c) OMP and (b), (d) BOMP over all of the buses of the standard IEEE 30 Bus for LDO = 2 and 4, respectively.

The OMP, MCOMP (with cluster size 4) and BOMP (with block size 4) algorithms were applied and final network-wide results over 100 realization of the IEEE 30 Bus have been compared in Figure 5.6. As it can be seen MCOMP algorithm represents a better performance than BOMP and OMP when the exact LDO cannot be estimated.
Figure 5.6: Recovery performance for (a) OMP and (b) BOMP (c) COMP over all of the buses of the IEEE 30 Bus network for Unknown LDOs.
CHAPTER 6
SPARSE REPRESENTATION-BASED POWER QUALITY EVENTS CLASSIFICATION USING INFORMATIVE SAMPLES

6.1 Introduction and Literature Review

Nowadays, ensuring the quality of the electrical energy delivered to consumers is an important task in power system operations. Regarding increasing usage of sensitive and critical loads in various commercial/residential applications, real time tracking and optimizing the behavior of the quality of the electrical power (traveling among the power network) is one of the major challenges in the future of smart power grids [72].

Various power quality (PQ) events (disturbances)\(^{28}\) may happen in a power system due to different faults and distortions. These events are happening with different probabilities and some of them may cause serious damages if not recognized within a suitable range of time [73]. On the other hand, power quality events may have different levels of importance for utility companies vs. residential or commercial customers. Thus, fast and reliable identification and classification of these events would be an important issue not only for power utilities but also for industrial companies, and critical loads such as banking systems or databases [72].

One widely standardized method for evaluating the power quality is to study the behavior of the electrical voltage signal within critical observation points in the PN. The new technology of fast recording sensor networks provides (almost) online measurements from the voltage behavior. However, extracting useful information from the corresponding huge amount of data (collected from thousands of sensors within large-scale SGs) would result in a complex and time consuming processing and analyzing procedure which raises new technical challenges. The principal standards in this field are IEC 61000, and IEEE 1159 [74]-[75]. Regarding IEEE standard 1159-1995, the PQ disturbances include a wide range of PQ phenomena, namely transient (impulsive and oscillatory), short duration variations (interruption, sag and swell), frequency variations, long duration variations (sustained under-voltages and sustained over-voltages) and steady state variations (harmonics, notch and flicker) with a time scale which ranges from tens of nanoseconds

\(^{28}\) Within the rest of this manuscript, we may use these two terms: disturbance/event, interchangeably.
to steady state\textsuperscript{29}. Broad classes of the disturbances that may occur in a power system are described as follows:

- Voltage Dips: The major cause of voltage dips on a system is local and remote faults, inductive loading, and switch on of large loads.
- Voltage surges: The major cause of Voltage surges on a system is Capacitor switching, Switch off of large loads and Phase faults.
- Overvoltage: The major cause of overvoltage on a system is Load switching, Capacitor switching, and System voltage regulation.
- Harmonics: The major cause of Harmonics on a system is Industrial Furnaces Nonlinear Loads Transformers/generators, and Rectifier equipment.
- Power frequency variation: The major cause of Power frequency variation on a system is Loss of generation and Extreme loading conditions.
- Voltage fluctuation: The major cause of Voltage fluctuation on a system is AC motor drives, Inter-harmonic current components, and Welding and arc furnaces.
- Rapid voltage change: The major cause of Rapid voltage change on a system is Motor starting, Transformer tap changing.
- Voltage imbalance: The major cause of Voltage imbalance on a system is unbalanced loads, and unbalanced impedances.
- Short and long voltage interruptions: The major cause of Short and long voltage interruptions on a system are Power system faults, Equipment failures, Control malfunctions, and CB tripping.
- Undervoltage: The major cause of Undervoltage on a system is Heavy network loading, Loss of generation, Poor power factor, and Lack of VAR support.
- Transients: The major cause of Transients on a system is Lightning, Capacitive switching, Nonlinear switching loads, System voltage regulation” \[74].

6.1.1 Literature Review and State of the Art Approaches

In the literature, usually the PQ events classification is considered as a general pattern recognition (PR) problem. Commonly, a PR problem can be divided into the following standard 5-steps procedure [27], including: 1) phenomena recoding and raw data preprocessing, 2) potential patterns detection/data segmentation, 3) feature extraction (FE), 4) feature selection (FS) or dimensionality reduction, and 5) classification or clustering. A variety of mathematical methods has been developed to implement each of these steps in the PR literature. From simple linear to highly complex and nonlinear FE, FS and classification approaches are available and one may select the

\textsuperscript{29} We refer to each of these event types (such as impulse, sag, flicker and etc.) as a PQ event class.
optimal approach based on the data specifications and characteristics. Figure 6.1 shows the overall structure of a typical PR problem.

![Figure 6.1](image)

**Figure 6.1:** Five basic steps in a PR problem: 1) pre-processing, 2) detection, 3) features extraction, 4) feature Selection, and 5) classification.

Like any other PR problem, many consolidated approaches have been implemented in the literature and state of the art for PQ events classification. Within the 1\textsuperscript{st} or preprocessing step, usually, a combination of denoising, DC offset removal and alignment techniques are incorporated to purify the data from the background noise and other unwanted components.

In the 2\textsuperscript{nd} step, parts of the raw signal which can potentially be considered as PQ events are isolated from the raw data. This step, can be interpreted as another preprocessing technique that divides the data sequence into many transition segments, where events are defined as the segments
in between transition segments. Most of the recent approaches can be categorized into the following two categorize: 1) parametric (or model based) methods, such as Kalman filter (KF) and auto-regressive (AR) models, and 2) nonparametric (or transform based) methods, such as short-term Fourier transform (STFT) and wavelet transform (WT) [72]. Very recently a filter bank-based approach with rapidly varying features is applied for the decomposition of voltage variation events to perform the detection of occurrence time [76].

Although the time domain representation may be useful for classifying some of the detected PQ event patterns, most of the algorithms are usually implementing a 3\textsuperscript{rd} analysis step, named feature extraction. Within the FE procedure, usually a signal processing technique is applied (on each of the individual event segments) to extract a set of features expected to reveal a hidden layer of information. Hopefully, the set of extracted features (information) is unique enough to better distinguish between different types of disturbances (compare to the initial time domain representation). From a mathematical perspective, the union space composed of all feature sets (corresponding to all detected PQ events), forms a geometrical representation that is called the \textit{feature space} (Figure 6.2), where the corresponding number of extracted features is called the \textit{feature space dimension}. In the literature, various features have been extracted to reveal the specific characteristic of each PQ event class [73]. Some prior researchers have used primary features such as event height, event width or peak to peak amplitude of the event. From algorithmic complexity viewpoint, forming such a low dimensional feature space simplifies the classification procedure at the expense of decreasing the distinguishing ability. In the modern schemes, mathematical transforms such as Fourier coefficients, wavelet transform-based features [77] and, S-transform [78] have found much interest to improve the final performance of the algorithm. Recently a new approach has been developed in [79], where statistical indices (such as minimum, maximum, variance, etc.) have been used in combination with sine and cosine discrete transforms.

From geometrical viewpoint, the corresponding shape of an event class within the feature space (or shape of its corresponding cluster), depends on the distribution of the extracted features (Figure 6.2(a)-(b)).

Although lots of attentions and efforts have been made on feature extraction step, the role of the feature selection step (step 4\textsuperscript{th}) has been widely ignored in the literature. Especially in case of online and real time applications, applying an appropriate FS method can significantly decrease the time complexity (by reducing the feature space dimension). Moreover, an intelligent FS
approach can modify the classes cluster shapes and form a more clustered-compact feature space (Figure 6.2(c)). To reduce the dimensionality of the feature space, PCA-based methods have been implemented in some of the literature [72]-[73].

Finally, in the 5th step, various classifiers have been applied to find the correct between-class boundaries in the feature space. Naturally, the overall performance of a typical classifier would be highly dependent on the prior FE and FS procedures. Geometrically, if all classes have a compact distribution (and are widely separated from each other), a simple linear classifier should be able to perfectly recognize a set of linear boundaries between these classes and accurately assign the correct class label to each of the events (inside the feature space). As a result, the optimal classifier selection should be made based on the specifications of the implemented FE and FS algorithms. A large set of supervised, semi-supervised and unsupervised approaches have been already examined in the previous work, including rule-based classification, Warping classifier, the common feed-forward neural network, time delay neural network, Hidden Markov models [82] and vector quantization, recurrence quantification analysis, Multi-way principal component analysis, nearest neighbor rule, inductive inference approach [72,73 and references therein], probabilistic NN, balanced neural tree, the dual neural network based methodology, the radial basis function (RBF) neural network [80]-[81], and support vector machines [83]. In case of simultaneous (overlapped) events, fuzzy rules have been applied to make a final decision regarding the occurrence and classification of the type of the disturbance [84].

In general, most of the aforementioned techniques are implemented based on a train-test framework, where a set of labeled events is used to train a classifier or to find an optimal map between a set of feature vectors and data classes (boundaries). Within the test procedure, a new unknown/unlabeled pattern or data segment (in the form of a feature vector) is presented to the classifier and the classifier (if suitably trained) will predict the true class of the new sample.

6.1.2 Technical Challenges

In this part, we highlight the following technical challenges (we mentioned before), in addition to some practical issues regarding PQ events classification in an ordered manner.

✓ Most current approaches are implementing complicated FE techniques to find a more compact and distinguishing representation of the data to improve the final classification performance. However, there is no known guideline for the optimal FE technique selection.
Optimal feature selection is a challenging issue highly depends on the nature of the data and less investigated within the PQ classification literature.

Conventionally, the overall performance of the traditional classifiers is critically dependent on the choice of features. However, no magic instruction manual is available. If an ordinary classifier cannot perfectly handle the classification step, one may need to utilize sophisticated (usually nonlinear) classifiers such as advanced artificial neural networks (ANN), fuzzy classifiers, generalized support vector machines or generative models (such as Markov models, Gaussian mixture models, etc.) to identify the nonlinear boundaries between data classes inside the feature space. Each of these approaches may itself contain a complicated model selection and parameter estimation procedure.

Another major challenge associated with the classifier training step is the choice of optimal number of training data samples for perfect classifier training (while decreasing the training time cost). This number is generally unknown and usually is found within a try and error procedure subject to minimizing a criterion such as those training errors used in ANNs training.

Alongside these usual PR problematic challenges, one may consider the following notable practical issues in the PQ event classification problem:

- Sensing and data transferring protocol.
- Data processing and analyzing time cost.
- Level of the measurement noise.
- Uncertainty injected by renewable generation sources, such as wind turbines or Photovoltaic cells, that may result in unexpected changes in the data characteristics.
- Multiple events or overlapped/simultaneous events.

Figure 6.2 illustrates a typical visualization from (a): a bad FE-FS procedure that resulted in a totally overlapped and hardly distinguishable class clusters within the feature space; (b): a good FS that resulted in a low dimensional with separable class clusters feature space, however, an unsuitable FE method resulted in highly nonlinear boundaries between class clusters. Finally, Figure 6.2 (c) illustrates a perfect FE-FS framework that resulted in linearly separable class clusters within the feature space. One notable point is that, even if a perfect approach is designed and implemented for a PR problem, there is always a small set of outliers which are hard to be correctly assigned to an individual class (basically resulting from the noise or other sources of uncertainty).
6.1.3 Contributions

In this work, we investigate the possibility of the PQ events classification using a modified sparse representation-based classification (SRC) framework, where despite usual classifiers no

Bad FE will form overlapped data clusters within the feature space with no identifiable boundaries between classes

Good FE-FS will form data clusters with linear boundaries between classes

Bad FS will form data clusters with nonlinear boundaries between classes

Figure 6.2 The role of the Feature Extraction and Feature Selection
training procedure is needed. While no particular classifier is trained, our approach can easily be
geneneralized and adapted to the future changes in the nature of the data. Within our proposed
approach, we will study how a random FS operation can be applied to decrease the feature space
dimension while preserve or even improve the classification accuracy. On the other hand, we
describe how selecting an optimal set of training samples can be interpreted as a high dimensional
convex hull approximation. This latter method is used to find the most important training samples
(we term as the informative samples). We show how a small set of informative training samples
(chosen from each event class) can be exploited to decrease the data space dimension while
preserves the classification accuracy. We also suggest several alternative approaches to deal with
the aforementioned imperfections related to the PQ events classification.

6.1.4 Related Works

Recently in [79], a PQ events classification approach has been developed using sparse signal
decomposition on hybrid dictionaries. This work basically tries to exploit the sparse nature of the
power signals (to define a sparse-based signal decomposition approach) in order to extract suitable
features. In this method, a hybrid dictionary is formed by concatenating the corresponding impulse,
Discrete cosine transform and Discrete sine transform functions (basis). Next, this dictionary is
used to provide better decomposition of PQ disturbance signals. This method can be interpreted as
a sparse-based feature extraction method that can be compared vs other FE techniques such as
wavelet transforms. A very similar approach has been developed in [86], that considers a variety
of other transforms such as S transform and wavelet transform as well (in designing the
corresponding hybrid dictionary).

In another work Wang et.al proposed a compressive sensing based approach, where authors
claim to sample power signals via a robust and semi-supervised compressive measure [85]. Next,
the set of obtained measurements are directly implemented as features for the subsequent online
classification (via an Online Sequential Learning Algorithm). This approach is also categorized
within the usual PR framework, where a set of features are collected and a classifier is trained.
However, exploiting the sparse nature of the power signals a set of compressive-based features are
utilized within the overall framework.
6.2 Compressive-Informative Power Quality Events Classification

In this section, first we introduce the specific mathematical notations that we use in the rest of this chapter. Next, we define the problem of PQ events Classification (PQC) with respect to our specific mathematical framework.

6.2.1 Mathematical symbols and notations for PQC

**Definition 6.1 Data Point:** We define the vector \( \mathbf{y}_l \in \mathbb{R}^M \) to be a measurement vector from an electrical voltage signal including a PQ event, we call this vector as a *data point*.

**Definition 6.2 Data class:** We call each of the following PQ events as a *data class* and assign them with a label \( c_j, j = 1, 2, \ldots, J \). PQ classes in this study include: 1) impulsive, 2) oscillatory, 3) sag, 4) swell, 5) flicker, 6) interruption, 7) harmonics and, 8) Notch. By composing a subset of these 8 classes we also generate 6 combined classes of events including 4 double and 2 leash events (Section 6.4).

**Definition 6.3 Feature vector:** We can consider each time sample of each data point \( \mathbf{y}_l^m, l = 1: N, m = 1: M \) to be a representative feature for each PQ event \( \mathbf{y}_l \). However, as we discussed before (in the area of pattern recognition) we usually extract and select more useful and distinguishing features from detected events and utilize them for the classification purpose (for example, the set of Fourier coefficients, or a set of statistical indices, or wavelet coefficients). We use the vector \( \mathbf{f}_l \in \mathbb{R}^L \) to represent the corresponding set of extracted features from initial time domain representation, \( \mathbf{y}_l \) (of each PQ event). Usually, the number of features selected to represent a data point, \( L \), is much smaller than the original dimension \( M \).

**Definition 6.4 Data space dimension vs feature space dimension:** Consider a data matrix \( \mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathbb{R}^{M \times N} \), composed of \( N = \sum_{j=1}^{J} n_j \) data points, each of length \( M \) samples (from \( J \) different classes of PQ events) have been recorded, where \( n_j \) represents the contribution of each data class in the whole dataset. Now, a set of \( L \) features are extracted from each data point \( \mathbf{y}_l \) resulting to a subsequent feature vector \( \mathbf{f}_l \). One may consider the representation of the initial dataset \( \mathbf{Y} \) in terms of feature space as a new matrix, called feature matrix, \( \mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N] \in \mathbb{R}^{L \times N} \). We call \( N \) as the data space dimension while \( L \) is termed as the feature space dimension.
6.2.2 PQC: Problem Definition

Consider for a 110V-50Hz power system a set of \( N = \sum_{j=1}^{J} n_j \) labeled\(^{30}\) sinusoidal electrical voltage segments (each of length 10-cycles) has been recorded from \( J \) classes of PQ events with a 12.8 kHz sampling frequency (256 sample/cycle), and is available as a training set\(^{31}\). We consider each of these known-class vectors, \( y_i \in \mathbb{R}^M, i = 1, 2, \ldots, N \), and \( M = 2560 \), to be a PQ event pattern. The final goal is to develop a framework that takes a feature vector, \( f_i^{test} \), extracted from a new data point, \( y_{i}^{test} \), and assign or map it into an individual class of PQ events \( c_j, j = 1, 2, \ldots, J \). We illustrate this combined mapping as a mathematical function \( c_i = A(D(y_i^{test})) \), \( c_i \in C = \{c_j | j = 1, 2, \ldots, J\} \).

Roughly speaking operator \( D \) stands for the data-feature mapping resulting from FE and FS procedure, while the feature-class (classifier) mapping is represented by the operator \( A \) (such a mapping is usually done by training a classifier). As mentioned before, in general, the optimal framework for designing these mappings is unknown, and may change based on the data nature and the specific application requirements and priorities.

In this work, we exploit the idea of sparse representation-based classification \(^{[87]}\) to decrease the size of the feature vectors in the mapping \( D \) while keeping the FE and FS steps as simple as possible. On the other hand, we develop a strategy to select the most informative data points (or their corresponding feature vectors) to be used in the realization of the mapping \( A \). Since within our framework we reduce both the data and feature space dimensions we call this approach as Compressive-Informative PQ events Classification (CI-PQ C).

6.3 Methodology

In this section, we introduce our sparse-based PQ events classification approach based on the theory of SRC \(^{[87]}\). Moreover, we briefly give an overview on the related sparse recovery background and concepts utilized in interpreting and solving CI-PQ C problem.

Consider a feature matrix \( F = [f_1, f_2, \ldots, f_N] \in \mathbb{R}^{L \times N} \), (including \( N \) feature vectors from \( N \) PQ data points of \( J \) PQ classes) is formed as follows:

\(^{30}\) Usually, a training dataset is labelled or manually classified by an expert person.

\(^{31}\) Regarding \(^{[74]}\)-\(^{[75]}\), a window of length 10-cycles captured by a sampling frequency about 12.8kHz is an appropriate segment of signal to represent a PQ event for most of the PQ event classes.
where \( n_j \) represents the contribution of each data class in the whole dataset. Regarding the theory of sparse representation-based classification [87], if the training samples from each single class, \( F_j \) for \( j = 1:J \) do lie on a low dimensional subspace, the corresponding data class, \( c_j \), of a new data point, \( y_{\text{test}}^i \), can correctly be recognized using the optimization procedure summarized in Algorithm.8.

Algorithm 8. Sparse Representation-Based Classifier

\[
\text{require: training data matrix } Y \in \mathbb{R}^{M \times N}, \text{ test sample } y_{\text{test}}^i \in \mathbb{R}^M
\]

1. Extract the desirable set of features from each data point, \( y_i \) using feature transform \( D \in \mathbb{R}^{L \times M}, f_i = D \times y_i \in \mathbb{R}^L \).
2. Form: measurement matrix \( A = [A_1, A_2, \ldots, A_J] \in \mathbb{R}^{L \times N} \), and feature test vector \( f_{\text{test}}^i \in \mathbb{R}^L \).
3. Solve \( P_1 \) or \( NP_1 \) for sparse vector \( x_i: \hat{x}_1 = \text{argmin}_{\hat{x}} \| \hat{x} \|_1 \text{ subject to } f = A\hat{x} \)
4. Compute \( J \) purified vectors \( \hat{x}_1^j \) for \( j = 1:J \), using indicator function \( g(\hat{x}_1): \in \mathbb{R}^L \rightarrow \mathbb{R}^L \), such that \( \hat{x}_1^j = g(\hat{x}_1) \), is a new vector whose only nonzero entries are the entries in \( \hat{x}_1 \) that are associated with class \( c_j \).
5. Compute residual \( r_j = \| f_{\text{test}}^i - A\hat{x}_1^j \|_2 \) for \( j = 1:J \)

output: Class (\( y_{\text{test}}^i \)) \( \triangleq c_j \) corresponds to the minimum \( r_j \)

This result comes from the fact that the corresponding feature vector, \( f_{\text{test}}^i \), of any new data point, \( y_{\text{test}}^i \), can be approximated as a linear combination within the space spanned by the training samples associated with its actual data class:

\[
f_{\text{test}}^i = \alpha_1^j f_1^j + \alpha_2^j f_2^j + \ldots + \alpha_n^j f_n^j, \quad (6.6)
\]
equivalently:

\[
f_{\text{test}}^i = F_j, x_{ij}, \quad (6.7)
\]
where $x_{ij} = [\alpha_1^j, \alpha_2^j, \ldots, \alpha_n^j]$. Alternatively, one may rewrite the linear representation of $f_{test}^i$ in terms of all training samples as:

$$f_{test}^i = F.x_i,$$  \hspace{1cm} (6.8)

where $x_i = [0,0, \ldots, 0, \alpha_1^i, \alpha_2^i, \ldots, \alpha_n^i, 0,0, \ldots, 0]^T \in \mathbb{R}^N$. Since most of the feature transformations include linear operations (or approximately so), the projection from the time domain into the feature space can be represented as a matrix $D \in \mathbb{R}^{L \times M}$. Thus, we may rewrite the whole procedure in terms of original time domain data points as follows:

$$D.y_i = D.Y.x_i \leftrightarrow f_i = F.x_i.$$  \hspace{1cm} (6.9)

Consider a valid test sample $y_{test}^i$ (from class $c_i$) can be adequately represented by its corresponding training data points from the same class, thus, the representation coefficient vector $x_i$ is naturally sparse if the number of object classes $J$ is reasonably large. If $L \geq N$, then the system of linear equations (6.9) can be solved in terms of the following $l_2$-norm optimization problem $(P_2)$ using pseudoinverse of $A$ (replacing $F$ with a general matrix $A$ we have)

$$P_2: \quad \tilde{x}_2 = \arg \min_{\tilde{x}} \| \tilde{x} \|_2 \quad \text{subject to} \quad f = Ax.$$

However, due to its general dense format (in another word, a solution with large nonzero entries corresponding to training samples from many different classes) the solution $\tilde{x}_2$ is not certainly informative to recognize the corresponding class of the test data point $y_{test}^i$. Thus, one may need to develop an alternative formulation to find the corresponding sparse representation as follows:

$$P_0: \quad \tilde{x}_0 = \arg \min_{\tilde{x}} \| \tilde{x} \|_0 \quad \text{subject to} \quad f = Ax.$$  \hspace{1cm} (6.11)

\[32\] For ease of notation we ignore the index notation, $i$, and consider the following set of linear equations: $f = Ax$. 

102
Where here again \( l_0 \)-norm stands for the number of non-zero elements of vector \( \mathbf{x} \). As we already discussed in Chapter 2, fortunately, the theory of sparse recovery and compressed sensing introduced the following alternative \( l_1 \)-norm relaxed version of this problem that can still correctly approximate the true sparse solution.

\[
P_1: \quad \hat{x}_1 = \arg\min_{\hat{x}} \|x\|_1 \quad \text{subject to} \quad f = A\hat{x}.
\]

Figure 6.3 represents a geometrical reasoning on why \( l_1 \)-norm results in the desired sparse solution while \( l_2 \)-norm solution is almost never sparse.

Figure 6.3 a geometrical comparison between \( l_1 \)-norm vs. \( l_2 \)-norm minimizations. Since the \( l_2 \)-ball does not have sharp edges along the axes, the corresponding solution to \( P_2 \) is almost never sparse, while regarding sharp edges along the axes of \( l_1 \)-ball we may arrive at the true sparse solution if all other sparse recovery conditions are satisfied.

On one hand a major goal is to reduce the feature space dimension, \( L \), (as much as possible), on the other hand, the sum of required number of training data points, \( N \), usually exceeds the feature space dimensionality. Thus, \( L \) is usually much smaller than \( N \), \( L < N \). Subsequently, solving \( P_1 \) with respect to \( \mathbf{x} \) is interpreted as an underdetermined system of linear equations that may result in a set of infinitely many candidate solutions \( \mathbf{x} \)'s for a sample test data point, \( \mathbf{y}^{\text{test}} \). Fortunately,
if $x_i$ is sparse enough (Chapter 2) with most of its elements close to zero except a small set of them, $K$ (usually a subset of $\alpha_k^i$, $k = 1:K, K < n_j$ coefficients are enough to represent each $f_{test}^i$ in terms of corresponding subspace of PQ event class $j$), this sparse recovery problem is efficiently solved by means of Theorem 2.22. In terms of noisy measurements, one can solve the following BPDN problem instead:

$$NP_1: \hat{x} = \underset{x}{\text{argmin}} \|x\|_1 \quad \text{subject to} \quad \|f - Ax\|_2 < \eta. \quad (6.13)$$

6.3.1 Interpretation of the Sparse Representation-based PQ Events Classification

Consider for a set of $J$ classes of PQ events a training matrix $A \in \mathbb{R}^{L \times N (= \sum_{j=1}^{J} n_j)}$ has been built as discussed above, and for a test PQ data point, $y_{test}^i$, a set of selected representative features have been extracted and stored in a feature vector $f_{test}^i$. Next, $P_1$ or $NP_1$ has been solved for corresponding sparse representation vector $x_i$, resulted in a solution $\hat{x}_i$. Now, one may use different criteria to decide on the true class of $y_{test}^i$. We may count the number of nonzero elements $e_j$ within each interval $n_{j-1} + 1: n_j$ and peak the $c_j$ corresponds to the maximum $e_j$. Another possibility is to calculate the corresponding $l_1$-nrom of each subvector $x_j = [\alpha_1^j, \alpha_2^j, ..., \alpha_n^j]$, $n_j = \|x_j\|_1$, for $j = 1:J$, and choose the $c_j$ corresponds to the maximum $n_j$. Another approach which has been used in [87], is to purify the vector $\hat{x}_i$ for each class $j$ using an indictor function and find the minimum reconstruction residual as described in Algorithm 8.

Figure 6.4 is a visualisation of a typical SRC procedure for a PQ event.

![Figure 6.4](image.png)
6.3.2 Feature Extraction and Dimensionality Reduction for SRC of PQ Events

As mentioned before, the final accuracy of the conventional classifiers is highly dependent on the shape of class clusters inside the feature space. Thus, optimal and expert FE would be critical for regular classification approaches to generate an easily separable feature space. Moreover, an intelligent FS reduces the feature space dimensionality and decreases the overall time cost. However, a surprising fact is suddenly get into the sparse representation-based classification that totally affects the meaning of FE and FS.

“If the solution \( x_{l} \) is sparse enough, then with overwhelming probability, it can be correctly recovered via \( P_{1} \) or \( NP_{1} \) from any sufficiently large number \( L \) of linear measurements \( f_{i} = D.Y.{x}_{i} \), precisely, for \( L \geq 2K\log(N/L) \)” [87]. Surprisingly, there is no need for an expert procedure to choose these \( L \) features and one may only select them randomly! As a result, the choice of an optimal feature transformation is no longer critical! [87]. This surprising phenomenon has been dubbed the “blessing of dimensionality” [88]. Thus, one may peak the transformation \( D \) whose entries are independently sampled from a zero-mean normal distribution, and each row is normalized to unit length. As a result, in this work we consider the following two sets of features for PQ classification:

1. The time samples of each PQ event, such that \( f_{i} = y_{l} \in \mathbb{R}^{M} \).
2. A random map from each vector \( y_{l} \), such that: \( f_{i} = \tilde{y}_{l} = D y_{l} \in \mathbb{R}^{L} \); where \( D \in \mathbb{R}^{L \times M} \) is a random Gaussian matrix (\( L \ll M \)).

6.3.3 Informative Data Points Selection

So far, we described the general sparse representation-based PQ events classification framework and later investigated a surprisingly random FE-FS approach to optimize the feature space geometry while reducing its dimensionality. In this section, we would like to go one step ahead and investigate the possibility of reducing the data space dimension \( N \). In another word, we would like to see if we can somehow optimize (minimize) the number of training data points needed from each class, \( n_{j} \), to reduce the overall size of the matrix \( A \), and reduce the classification producer time cost while preserve the classification accuracy. Considering the training matrix or dictionary \( A \) as a dictionary of atoms (columns), such an idea can be categorized as a sparse dictionary learning problem [89]-[90], where one would like to optimize the structure of the dictionary \( A \) to
recover the sparsest representation vector $\mathbf{x}$. The notable point is that, in case of CI-PQC, dictionary atoms are selected from our training data points, $y_i$’s.

If there is a subset of data points from each class $c_j$ where all other data points from the same class can be represented in terms of a linear combination of those few informative points, then such a subset should be enough to cover the whole corresponding cluster of class $c_j$ within the feature space. Consequently, one should be able to replace the corresponding $n_j$ training samples with a smaller subset of $n_j^{\text{informative}}$ ($\ll n_j$) data points, within the measurement dictionary (or matrix $A$) and still get the same classification accuracy. Applying such an informative data points selection procedure (over all classes of training subsets, in all data classes) results in a new measurement matrix $A^{\text{informative}} \in \mathbb{R}^{L \times N^{\text{informative}}}$ where $N^{\text{informative}} = \sum_{j=1}^{J} n_j^{\text{informative}}$, and $N^{\text{informative}} \ll N$. This can significantly reduce the data storage size and computational complexity for any classification algorithm and not only for SRC.

From geometrical perspective, finding a subset of informative points from a set of data points in an $N$-dimensional space, $\mathbb{R}^N$, is interpreted as finding the extreme points or vertices of the corresponding convex hull that covers the whole of the data points. From PR point of view, the corresponding convex hull of these data points represents the location of their corresponding cluster in the feature space (Figure 6.5).

However, a major technical issue in high dimensional feature spaces would be the time cost of the convex hull approximation. Within the literature [91]-[92 and references therein] a variety of high dimensional convex hull approximation algorithms have been developed so far, mostly with a time complexity that changes in an exponential order of dimension, $L$ [92]. Beside their specific limitations, most of them can’t converge once the dimensionality gets larger and larger. Moreover, most of these methods are not able to optimize the approximation quality once new vertices are identified, thus, usually, use many vertices for initialization. Recently, in [92], a new greedy convex hull approximation algorithm was presented with time complexity of $O\left(d^3N^2 \log(d) \right)$ where $d$ stands for the number of iterations of the proposed algorithm. As argued by the authors this number should be close to the number of extreme point or vertices of the approximated convex hull, where in most of the practical data processing problems is much smaller than $L$.

---

33 In [89], a similar approach has been adapted in case of representative face recognition.

34 such as famous quick hall technique; with worst case time complexity in high dimensions of $O\left(N^{\frac{12}{5}}\right)$;
Regarding our feature selection argument is section 6.3.2, the initial dimension of the feature space is equal to the number of time samples captured by the voltage sensor per 10 cycles, that is 2560 sample/event. Based on our observations from different case studies (in average) a sparsity level $50 < K < 100$ is appeared within the true support of the vector $x$ for most of the correctly classified $y^{test}$’s. Thus, considering the bound $L \geq 2K\log(N/L)$ within random FE-FS approach, we can reduce the feature space dimension up to a number between 305-508. While this is still a high dimensional feature space, we decided to exploit the aforementioned greedy approximate convex hull computation algorithm (Please refer to [92]- Algorithm 1 and 2) in order to find the vertices of the corresponding convex hull of our data class clusters within our feature space. We call these vertices as the most informative data points.
6.4 Simulations and Discussion

To examine the recognition performance of the CI-PQC approach a synthetized database has been generated using the corresponding IEEE standard mathematical equation (Table 6.1 [99]) of 8 well-known PQ events. As a rule of thumb within the state of the art and power quality standards (refer to section 6.2.2) a 110V-50Hz power system has been considered and a set of 400 training PQ events per class, has been generated for 8 single and 6 combined event classes. Thus, a dataset $Y \in \mathbb{R}^{2560 \times (400 \times 14)}$ has been considered to form the initial sensing matrix $A$. Finally, a set of 100 test data points (with the same specifications) have been generated for each class of PQ events.

Table 6.2 illustrates the final classification results using all data points and full length feature vectors $f_i = y_i \in \mathbb{R}^{2560}$. In the next set of simulation results we implemented the same framework on a set of 507 randomly selected time samples from 2560 total number of samples per event pattern. Results have been reported in Table 6.3.

As it can be observed from Table 6.2, the CI-PQC approach illustrates an acceptable performance both in single and combined events classes. Table 6.3 indicates that the same accuracy performance is achieved with only 20% randomly selected time samples!

Our achievements are partially coming from the fact that the corresponding set of data points to each single PQ event class, approximately lie on a low dimensional subspace. One may check the fact of low dimensionality in subspace by evaluating the corresponding distribution of the singular value decomposition coefficients over each sub dataset matrix $Y_j \in \mathbb{R}^{2560 \times 400}$ for $j = 1: 14$.

Figure 6.6, illustrates the first 100 largest SVD coefficients for each data class (except the first one, for sake of better scaling). Regarding the linear algebra concepts, roughly speaking, the number of a set of SVD coefficients with much higher value than the rest (dominant coefficients), indicates the corresponding real dimension of the data points or dimensionality of the corresponding low dimensional subspace. As it can be observed, most of these curves are decreasing quickly, which indicates the fact that almost all of the PQ data classes are relying over a low dimensional subspace (with a dimensionality much smaller than the original dimension). In another word, the smaller the number of dominant SVD coefficients the lower the dimension of the corresponding subspace.
Table 6.1 PQ Events Classes

<table>
<thead>
<tr>
<th>Typical behavior</th>
<th>Mathematical equation</th>
<th>Representative parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Notch</strong></td>
<td>$y(t) = 1.11v_0(e^{-7.5 \times 10^3(t-T_0)} - e^{-3.44 \times 10^4(t-T_0)}) \times (u(t-T_d) - u(t-T_d))$</td>
<td>$0.2 \leq v_0 \leq 1$</td>
</tr>
<tr>
<td></td>
<td>$\omega = 2\pi \times 50, \rho = 0.5$</td>
<td>$T \leq T_d \leq 7T$</td>
</tr>
<tr>
<td></td>
<td>$k = 1.7$</td>
<td>$T_s = T_d + 1,\text{ms}$</td>
</tr>
<tr>
<td><strong>Interruption</strong></td>
<td>$y(t) = \sqrt{2}v_0 \sin(2\pi \times 50t + \varphi) \times (u(t-T_d) - u(t-T_d))$</td>
<td>$-\pi/4 \leq \varphi \leq \pi/4$</td>
</tr>
<tr>
<td></td>
<td>$0.05 \leq v_0 \leq 0.1$</td>
<td>$T_s = T, 2T, ..., 8T$</td>
</tr>
<tr>
<td></td>
<td>$T_e - T_s = T, 2T, ..., 9T$</td>
<td>$T_e = T, 2T, ..., 9T$</td>
</tr>
<tr>
<td><strong>Short</strong></td>
<td>$y(t) = \sqrt{2}v_0 \sin(2\pi \times 50t + \varphi) \times (u(t-T_d) - u(t-T_d))$</td>
<td>$-\pi/4 \leq \varphi \leq \pi/4$</td>
</tr>
<tr>
<td></td>
<td>$0.1 \leq v_0 \leq 0.9$</td>
<td>$T_s = T, 2T, ..., 8T$</td>
</tr>
<tr>
<td></td>
<td>$T_e - T_s = T, 2T, ..., 9T$</td>
<td>$T_e = T, 2T, ..., 9T$</td>
</tr>
<tr>
<td><strong>Flapper</strong></td>
<td>$y(t) = \sqrt{2}v_0 \sin(2\pi \times 50t + \varphi) \times [1 + \sqrt{2}v_0 e^{-\alpha \varphi(t-T_d)} \sin(2\pi f_0(t-T_d) + \varphi)]$</td>
<td>$0 \leq \varphi \leq 2\pi$</td>
</tr>
<tr>
<td></td>
<td>$0.05 \leq v_0 \leq 0.1$</td>
<td>$T_s = T, 2T, ..., 7T$</td>
</tr>
<tr>
<td></td>
<td>$300 \leq f_0 \leq 900$</td>
<td>$150 \leq \alpha \leq 1000$</td>
</tr>
<tr>
<td><strong>Radominger</strong></td>
<td>$y(t) = \sqrt{2}[\sin(2\pi \times 50t + \varphi) + \cdots + \alpha_{2k} \sin(2\pi(2k) \times 50t + \varphi_{2k}) + \alpha_{2k+1} \sin(2\pi(2k + 1) \times 50t + \varphi_{2k+1}) + \cdots]$</td>
<td>$0.015 \leq \alpha_{2k} \leq 0.03$</td>
</tr>
<tr>
<td></td>
<td>$0.03 \leq \alpha_{2k+1} \leq 0.06$</td>
<td>$-\pi \leq \varphi_{2k} \leq \pi$</td>
</tr>
<tr>
<td></td>
<td>$-\pi \leq \varphi_{2k+1} \leq \pi$</td>
<td>$k = 1, 2, ..., 10$</td>
</tr>
<tr>
<td></td>
<td>$-\pi \leq \varphi \leq \pi, \text{THD} \geq 5%$</td>
<td>$-\pi \leq \varphi \leq \pi$</td>
</tr>
</tbody>
</table>
Table 6.2 Classification results CI-PQC (CONFUSION MATRIX): Descriptions: id-c: identified class, a-c: actual class, im: impulse, n: notch, in: interruption, sw:swell, s:sag, t:transient oscillatory, f:flicker, h: harmonic

<table>
<thead>
<tr>
<th></th>
<th>im</th>
<th>n</th>
<th>in</th>
<th>sw</th>
<th>s</th>
<th>t</th>
<th>f</th>
<th>h</th>
<th>f/sw</th>
<th>h/sw</th>
<th>h/s</th>
<th>s/h</th>
<th>t/sw</th>
<th>t/sw/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>im</td>
<td></td>
<td>83</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>in</td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sw</td>
<td></td>
<td></td>
<td></td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>3</td>
<td>89</td>
<td>4</td>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>1</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>f</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>94</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>f/sw</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>h/s</td>
<td>1</td>
<td>95</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sw/h</td>
<td>1</td>
<td>91</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s/h</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>92</td>
</tr>
<tr>
<td>t/sw</td>
<td>5</td>
<td>9</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>81</td>
</tr>
</tbody>
</table>

Table 6.3 Classification results of the CIC-PQ (Confusion Matrix)

<table>
<thead>
<tr>
<th></th>
<th>im</th>
<th>n</th>
<th>in</th>
<th>sw</th>
<th>s</th>
<th>t</th>
<th>f</th>
<th>h</th>
<th>f/sw</th>
<th>h/sw</th>
<th>h/s</th>
<th>s/h</th>
<th>t/sw</th>
<th>t/sw/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>im</td>
<td></td>
<td>95</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>n</td>
<td>2</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>in</td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sw</td>
<td></td>
<td></td>
<td></td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>1</td>
<td>91</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f/sw</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h/s</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>97</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sw/h</td>
<td>1</td>
<td>94</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s/h</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t/sw</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>86</td>
</tr>
</tbody>
</table>

110
Figure 6.6: Distribution of the first 100 dominant SVD coefficients over each of the 14 classes of PQ events within the corresponding training sub dataset $Y_j$ for $j = 1:14$. 
In what comes next, we examine the idea of informative data point selection over the same set of training dataset. For each of the 14 classes of events the corresponding sub training dataset $Y_j \in \mathbb{R}^{2560 \times 400}$ have been provided to the convex hull approximator (please refer to Algorithm-2 [92]) and the algorithm is asked to find the most informative 150 data points from each data class. The approximation error, $e_i$, has been set to $10^{-4}$, which means that if such an error level has been reached using smaller number of data points the algorithm will stop and the output will be the first $k < 150$ data points. Figure 6.7 illustrate the approximation error vs number of the selected vertices for the corresponding convex hull (or in our interpretation number of informative data points) of a couple of PQ event classes. As it can be seen some data classes are more compressible in terms of their general behavior, thus the number of informative samples needed to cover their corresponding convex hull (or equivalently corresponding cluster within the feature space) is much smaller. For example, due to the repetitious nature of harmonics within a limited range of sinusoidal frequencies, a small number of 20 informative data points are adequate to span the whole corresponding data cluster of the harmonic events. In contrast, due to semi-random behavior of impulsive events, even 150 data points can only cover up to 50% of the whole cluster intra-space. These are the two extreme cases in the 14 data classes under study. The rest of the PQ event classes would illustrate a behavior in between. For instance, interruption events, can be covered using almost 44 informative data points while swell events need about 90 vertices (informative data points) to span their cluster inside the feature space. Using the following number (Table 6.4) of informative data points (out of 400 available data points for each class) we reached to the same classification performance (reported in Table 6.2 and 6.3) with at most 2% variation in the classification accuracy.

Table 6.4 Selected Number of Informative Data Points for Each PQ Class

<table>
<thead>
<tr>
<th>Class</th>
<th>Impulse</th>
<th>Notch</th>
<th>Interruption</th>
<th>Swell</th>
<th>Harmonics</th>
<th>Sag</th>
<th>Sag-Harm</th>
<th>Harm-Swell</th>
<th>Harm-Sag</th>
<th>Harm-Swell-Harm</th>
<th>Sag-Swell-Harm</th>
<th>Sag-Swell-Transient</th>
<th>Flick-Sag-Swell</th>
<th>Flicker</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>400</td>
<td>200</td>
<td>50</td>
<td>90</td>
<td>20</td>
<td></td>
<td>150</td>
<td>120</td>
<td>120</td>
<td>150</td>
<td>250</td>
<td>120</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

112
Figure 6.7 (a)-(k): corresponding convex hull approximation error vs number of selected informative samples or equivalently convex hull vertices for couple of PQ event classes.
Figure 6.7: continued.
Figure 6.7: continued.
Figure 6.7: continued.
HARMONIC SWELL

(i)

HARMONIC SAG

(j)

Figure 6.7: continued.
Figure 6.7: continued.
In this thesis, we investigated the opportunities and challenges related to exploiting the theory of compressive sensing and sparse recovery to address some of the major monitoring challenges in the technology of smart grids. We demonstrated in Chapter 3 that since, usually, a small ratio of simultaneous power line outages may happen during a failure, one may define (based on the DC power flow model) a sparse recovery-based formulation to localize the position of damaged lines. We further investigated the technical challenges regarding this POI-SRP and suggested a variety of alternative solutions. In Chapter 4, we introduced a new compressive sensing-based approach for power network topology identification that can also be used to address the POI problem. Due to a special structure of the sparse vectors in our formulation (CS-TI), we exploited the capabilities of a variety of generalized sparse recovery solvers to further decrease the time cost and the number of measurements needed for full topology identification in a very short range of time. In Chapter 5, we introduced a new network-wide framework for modeling and tracking the dynamic behavior of the transmission lines, or alternatively the entire power network. Based on the mathematical differential equation of the current wave propagation in the transmission lines we termed new dynamic indexes, line dynamic order, and line dynamic coefficients. We further discussed how a structured sparse recovery problem (SGDBM-SSRP) can be developed to quickly estimate and track the behavior of these parameters from online measurements of electrical currents. In these three chapters, exploiting the idea of Bound-exclusion and local optimization (initiated in [22]), we developed new alternative greedy algorithms to address the challenges resulting from data correlation issue.

Finally, in Chapter 6, we introduced a modified sparse representation-based classification approach to address the power quality events recognition. The IEEE standard formulation of 14 individual classes of PQ events has been utilized to generate a dataset of 500 events per each PQ class. Unlike most proposed state of the art algorithms, no detection procedure is mandatory. Moreover, since the data points from each single PQ event class lie approximately on a low dimensional subspace, we achieved an acceptable classification performance using only the time samples from each PQ event without any further feature extraction analysis.
In addition, we showed that thanks to the blessing of dimensionality fact, a simple random feature selection can be practically implemented to decrease the dimensionality of the corresponding feature space. Finally, we interpreted the problem of finding the most informative training data points in terms of approximating the corresponding convex hull or equivalently approximating the shape of the corresponding data cluster within the feature space. Due to the high dimensionality of the corresponding features space (in our CI-PQC formulation) a new greedy algorithm for approximating the convex hull [92] has been successfully applied. The results indicate that exploiting only a subset of informative samples would be enough to reach a similar accuracy performance with the sparse representation-based approach.

At the completion time of this PhD dissertation, a variety of other researches had also started investigating the capabilities and opportunities of CS-SR techniques for various power engineering areas [101]. Thus, there is still huge potential for developing and exploring the other fields of the power system from the CS-SR perspective. Security problems such as attack recognition, Microgrids construction, sparse matrix converters, and most power signal processing problems are still waiting for fresh and compressible minds! Particularly, regarding our own contributions, we suggest further investigation of the sparse classification approach through a signal processing framework where one may decide to implement the same idea on a set of wavelet-based extracted features. Another idea would be related to robustness vs. low SNR. Recently, a new idea has been suggested in [93] to improve the sparse representation-based classification in terms of low SNR in electric machinery.
REFERENCES CITED


