Model-Based Machine Learning for the Power Grid

by

Raksha Ramakrishna

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved September 2020 by the
Graduate Supervisory Committee:

Anna Scaglione, Chair
Douglas Cochran
Andreas Spanias
Vijay Vittal
Junshan Zhang

ARIZONA STATE UNIVERSITY

December 2020

ABSTRACT

The availability of data for monitoring and controlling the electrical grid has increased exponentially over the years in both resolution and quantity leaving a large data footprint. This dissertation is motivated by the need for equivalent representations of grid data in lower-dimensional feature spaces so that machine learning algorithms can be employed for a variety of purposes. To achieve that, without sacrificing the interpretation of the results, the dissertation leverages the physics behind power systems, well-known laws that underlie this man-made infrastructure, and the nature of the underlying stochastic phenomena that define the system operating conditions as the backbone for modeling data from the grid.

The first part of the dissertation introduces a new framework of graph signal processing (GSP) for the power grid, Grid-GSP, and applies it to voltage phasor measurements that characterize the overall system state of the power grid. Concepts from GSP are used in conjunction with known power system models in order to highlight the low-dimensional structure in data and present generative models for voltage phasors measurements. Applications such as identification of graphical communities, network inference, interpolation of missing data, detection of false data injection attacks and data compression are explored wherein Grid-GSP based generative models are used.

The second part of the dissertation develops a model for a joint statistical description of solar photo-voltaic (PV) power and the outdoor temperature which can lead to better management of power generation resources so that electricity demand such as air conditioning and supply from solar power are always matched in the face of stochasticity. The low-rank structure inherent in solar PV power data is used for forecasting and to detect partial-shading type of faults in solar panels.

DEDICATION

To my family

ACKNOWLEDGEMENTS

I am deeply indebted to my advisor, Prof. Anna Scaglione for her wonderful guidance, patience and support. This work has greatly benefited from her scrutiny and insight. Her creativity in formulating problems, courage in venturing out seeking to connect seemingly different areas of study while doing multidisciplinary work and her excitement for research will continue to motivate me. I also thank my dissertation committee- Prof. Douglas Cochran, Prof. Andreas Spanias, Prof. Vijay Vittal and Prof. Junshan Zhang for their input, feedback and collaboration. I am also grateful to all the Professors I have taken courses with at ASU who helped me establish a strong technical foundation.

My first brush with research was at the Signal Processing for Communications (SPC) lab at the Indian Institute of Science under the tutelage of Prof. Chandra Murthy. I want to thank him for providing me the opportunity to explore various facets of research ranging from writing papers, data analysis to developing software. Huge thanks to all the members of SPC lab I overlapped with and particularly to Karthik and Geethu for being my very first collaborators and co-authors.

I would like to thank all my collaborators and internship hosts Prof. Emiliano Dall'Anese, Dr. Andrey Bernstein, Dr. Sean Peisert and Dr. Aram Shumavon for their ideas, encouragement and support.

A big thank you to all the past and present SINE lab members-To, Reinhard, Lorenzo, Mahdi, Kari, Eran, Bita, Sissi, Tech, Nikhil, Nurullah, Ignacio and Shammya. They made work and leisure so much fun and I made friends for life. You guys are simply the best!

Thanks to my longtime friends-Ayesha, Saraswati, Manoj, Anisha, Ashwin, Sushma, Vivek, Shivarama, Suhas, Suleman and Sunil for believing in me and offering me kind words of encouragement from time to time.

I am also extremely grateful to all my friends in Tempe and roommates-Karthigayani, Anushree, Vaishnavi, Subramanian, Niveditha, Suhas, Saranya, Pratyusha, Rakshith, Parth and Dheeraj. They made Tempe feel like home, and were my local family!

I thank my parents, Ramakrishna and Mangala, sister Varsha and my in-laws Surendra Rao and Vijayalakshmi for their relentless cheering and great support. Last but not the least, I would not have made it this far if not for my husband Harish's love, patience, understanding and optimism. Thank you my dear, for enduring our long periods of separation and for moving countries for me!

Lastly, I want to thank the ASU ECEE Graduate advisors, Business Operations Specialists, ASU IT, Tax advisors and the International Scholars and Student Center who helped a great deal with all the bureaucratic tasks. I am also grateful for the expansive ASU library and the Sun Devil Fitness Center that hosted great group fitness classes which kept me in good health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

Algorithm Page

Chapter 1

INTRODUCTION

1.1 Overview

Electric power systems are moving from using traditional system of fossil-fuel based generation to serve passive loads, to renewable generation and smart active loads. In order to have this energy system compliant with the traditional system and address energy imbalances, there is need for advancing wide area sensing and control in the power grid. This advanced management of the grid has led to the collection of data from large new networks of sensors such as phasor measurement units (PMU) [1, 2] in the transmission grid and advanced meters in the distribution grid. Since the invention of PMUs, their deployment has steadily increased in size worldwide and have become an important tool in wide area measurement for monitoring the electrical grid [3] while the advanced metering infrastructure (AMI) has become an important tool for electric utilities to monitor the possibly negative injections on distribution feeders that increasingly incorporate solar generation and storage [4].

PMUs produce data with high temporal resolution since they sample at the rate of 60 or 120 samples per second [3]. It is then natural that this wide area sensing system will produce data with a large footprint, putting the study of electrical power grid in the same league as fields that rely heavily on data-driven methods for decision making. While this abundance of data from the power grid seems daunting, the main tenet of the first part of this dissertation is that such data can be represented in a low-dimensional subspace. Intuitively, correlation in the data is induced by the physical laws of the grid and in part by collection of redundant measurements. The goal of this dissertation is to go beyond intuition and empirical observation to justify the black box models. Instead it is shown how it is possible to utilize the underlying physics as well as the knowledge of electrical-grid that is a man-made infrastructure to guide the process of abstraction and support the design of model-based machine learning algorithms for power systems.

This dissertation explores the topic of modeling data from power systems in a way that their low-dimensional representations become apparent. Then, these models are utilized to address several problems in both transmission and distribution of power systems such as forecasting renewable energy generation and anomaly detection. Well-established theory of power systems, tools from statistical signal processing and validation through a large amount of data from the power grid are the ingredients for the model-based machine learning algorithms proposed in this dissertation.

The first part of the dissertation focuses on low-dimensional modeling techniques to voltage phasor measurements that characterize the overall system 'state' of the power grid. Given that data from the power grid lies on a graph that is the electric grid, it introduces the framework of graph signal processing (GSP) in general and specializes the tools and ideas from GSP to voltage measurements. It establishes that voltage measurements have a generative model in the realm of GSP which is suggestive of the underlying low-dimensionality in the data.

The second part of this dissertation focuses on AMI measurements of power injection from solar photo-voltaic (PV) power generation and power consumption of thermostatically controlled loads (TCL) while modeling the statistical dependency that exists between these data.

Since both solar PV power and temperature are dependent on weather and the sun, there is natural correlation not just between the two quantities but also in space due to geographical proximity between residential consumers. This induces

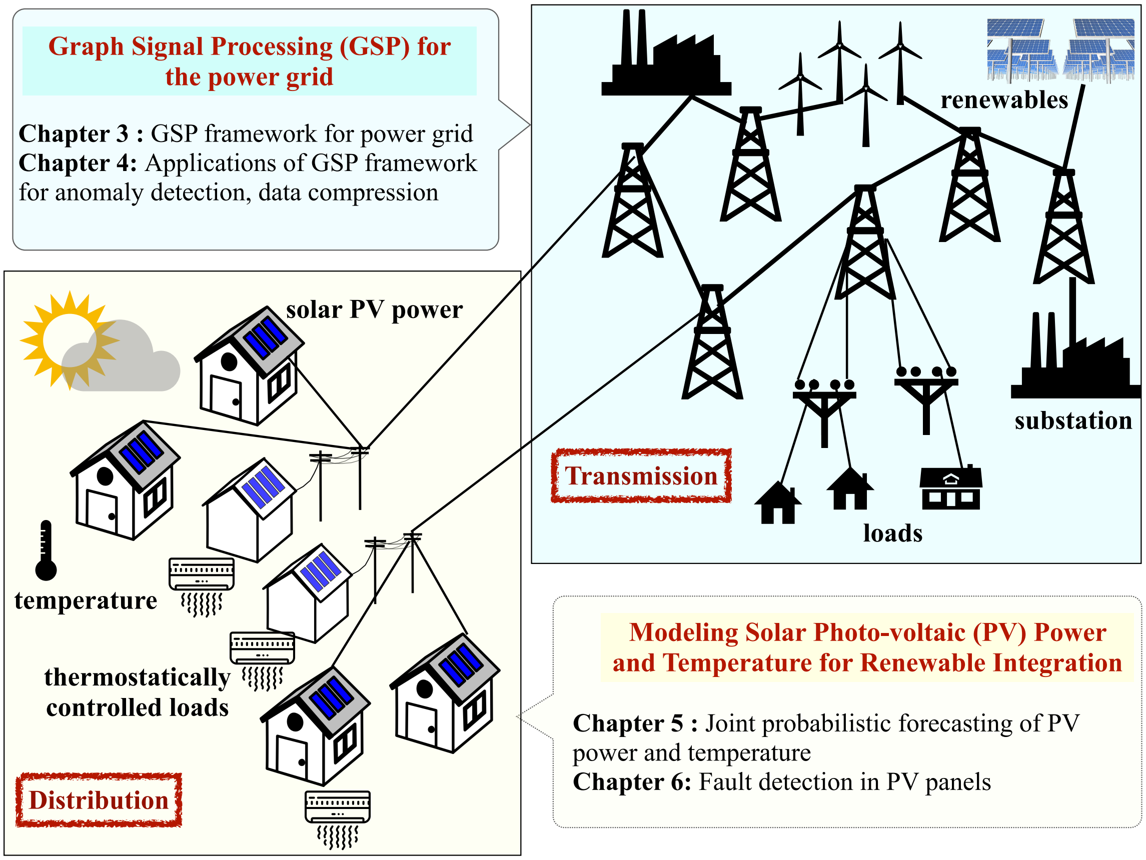


Figure 1.1.1: An Overview Of This Dissertation And The Corresponding Chapters.

correlation in the variability that lies in the heating or cooling power demand and renewable power production. Therefore an equivalent low-dimensional representation of these quantities is possible and the joint modeling of solar PV power and outdoor temperature is studied. In fact, prior works have already applied concepts from GSP to model the spatial correlation in solar radiation [5] and outdoor temperature [6]. The graph considered is based on the geographical distance between multiple locations. This dissertation fill the gap that exists in modeling the temporal structure and coupling between these data sets modeling the correlation between solar PV power and temperature at one location while accounting for temporal stochasticity and then considering the spatial model for locations in close geographical proximity.

The models introduced in this dissertation make the low-dimensional representation explicit and utilize these observations to address problems such as forecasting, data compression and fault detection. Fig. 1.1.1 presents an overview of the dissertation.

Section 1.2 introduces the first part of the dissertation including GSP, class of GSP models to which power systems data belongs to and details the prior work and contributions. In Section 1.3, second part of the dissertation is introduced wherein joint models for solar PV power and temperature are motivated and prior work in this area are discussed.

1.2 Graph Signal Processing for the Power Grid

The power grid is one of the foremost examples of a large-scale man-made network. The nodes of the associated graph are the grid *buses* and its edges are its *transmission lines*. It is therefore natural to see data from the power grid such as voltage phasor as signals that reside on a graph or 'graph signals'. Although this perspective is not new, the field of graph signal processing (GSP) [7,8] has the ability to provide further insight into data from the electrical grid since it extends ideas from discrete-time signal processing to data that reside on a graph. In a similar vein as linear time-invariant (LTI) filters in discrete-time signal processing, a graph filter can be classified as either low-pass, band-pass, or high-pass, depending on its graph frequency response.

It is well known that voltage phasor data tend to be confined to a much smaller dimension compared to the size of the data record in both space and time. Many papers have adopted this observation of dimensionality reduction for missing data interpolation [9, 10], correcting bad data [11] and to detect faulty events [12–15]. Unlike the aforementioned papers that only discuss the reduction in dimensionality, it is possible to explicitly put forth the structure of this low-dimensional subspace using a GSP-based generative model.

This characteristic of dimensionality reduction of voltage data can be well modeled by a *low-pass* graph filter which captures a smoothing operation applied to an input graph signal. As it turns out, this is a common property of processes observed in many physical/social systems. Chapter 2 briefly reviews the concepts entailing low-pass GSP and provides a set of particular examples highlighting the fact that low-pass graph signals often appear in different application domains.

Once the general concepts from low-pass GSP are introduced, Grid-GSP: a GSP framework for the power grid is discussed in Chapter 3. It builds upon the existing system-level knowledge of power systems to create a solid foundation to analyze power-grid measurements using tools from GSP. Well-known results for sampling, data interpolation and network-inference in GSP are extended to power system data without losing the associated physical interpretation. The core idea in this chapter is derived by rewriting the differential algebraic equations (DAE) [16] in a way often done in transient stability analysis of power systems, to reveal that the inherent structure in voltage phasor measurements can be explained using a low-pass graph filter, whose inputs are the generator voltages, as a generative model.

In Chapter 4, few applications of Grid-GSP are studied namely-identification of community structure in the grid which is a consequence of the low-pass nature of voltage graph signals, fault localization using under-sampled grid, data data compression and the detection of false data injection attacks using Grid-GSP.

1.2.1 Prior Works and Contributions

Several papers have used insights from spectral and algebraic graph theory for problems such as optimal placement [17–20], for generating statistically accurate topologies [21], for grid topology identification [22, 23] and to defend against attacks [24, 25]. GSP concepts have been leveraged in [26, 27] to detect false data injection (FDI) attacks. Prior work in [28] dealt with performance limits on fault localization with inadequate number of PMUs and connected it with graph signal sampling theory and optimal placement of PMUs for best possible resolution of fault localization in this under-sampled regime.

The Kron-reduced network among the generator buses and the associated properties are used in [29, 30] to detect low-frequency oscillations as well as the resulting islanding patterns. In [31], the authors have shed light on the relationship that exists between Graph Laplacian and modes in power systems. Recently, a comprehensive review of graph-theoretical concepts in power systems is presented in [16].

Additionally, there have been several papers adopting graphical models for state estimation [32], topology estimation [33] and optimal power flow [34]. While the modeling approach is valid, the graphical models capture correlation whereas GSP-based method models the underlying cause for that correlation structure thereby opening the door for statistical and non-statistical approaches.

Note that with the exception of [26, 27], no other papers make the connection with GSP and even in the aforementioned work, GSP is used in an empirical manner without establishing a foundation, as done in preliminary work [35, 36].

Contributions The overarching goal of Chapters 3 and 4 are to lay down the foundations of Grid-GSP and elucidate properties of power grid signals using tools from GSP. In particular:

1. A generative model for voltage measurements is established which states that voltage phasors are the result of an excitation to a low-pass graph filter whose graph shift operator (GSO) is defined using a function of the system admittance matrix. This was partly explored in previous work in [35,36] and used for blind community detection.

2. Study the spatio-temporal structure of the sparse excitation that, at a fast time scale, dominated by the generators dynamics. It is shown that the excitation can be modeled as an auto-regressive graph filter [37] (AR-2) for the input signals from the generator internal bus.
3. Known GSP based algorithms for sampling and reconstruction, interpolation and denoising and network inference in the context of signal-processing voltage data are revisited.
4. Specific applications of Grid-GSP are discussed namely:
    - Nature of the voltage graph signal is used to inspect the underlying community structure in the electrical grid by utilizing tools from blind community detection [38].
    - Fault localization in the under-sampled grid regime [28] and optimal placement of sensors under such conditions is studied.
    - A sequential lossy voltage data compression algorithm that exploits the spatio-temporal structure in the generative model is studied to arrive at an empirical rate-distortion curve.
    - Detection of False data injection attacks (FDI) by explicitly using the structure of the subspace specified by the low-pass graph filter for voltage data.

1.2.2 Related Publications

Work discussed in the first part of the dissertation can be found in

- Raksha Ramakrishna and Anna Scaglione ,"Grid-Graph Signal Processing (Grid-GSP): A Graph Signal Processing Framework for the Power Grid", (under review) in *IEEE Transactions on Signal Processing*
- Raksha Ramakrishna, Hoi-To Wai, Anna Scaglione, "A User Guide to Low-

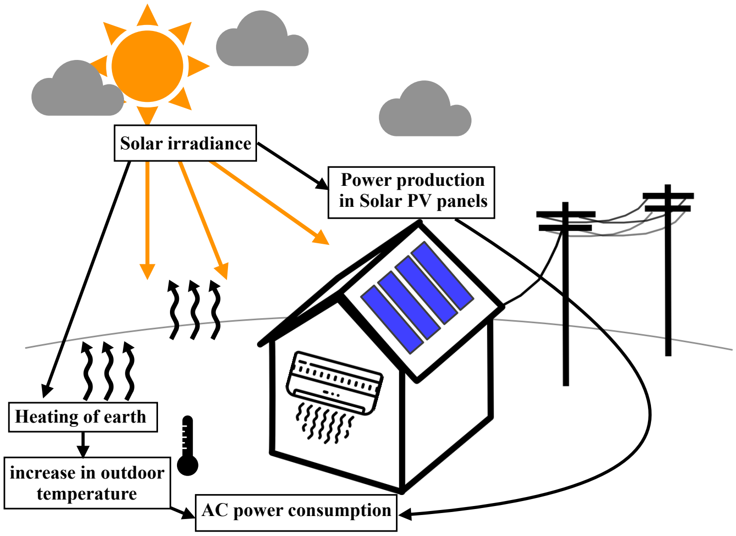


Figure 1.3.1: Figure highlighting the relationship between solar PV power and thermostatically controlled loads (TCL).

Pass Graph Signal Processing and its Applications", *IEEE Signal Processing Magazine* (accepted)

- Mahdi Jamei, Raksha Ramakrishna, Teklemariam Tesfay, Reinhard Gentz, Ciaran Roberts, Anna Scaglione, Sean Peisert , "Phasor Measurement Units Optimal Placement and Performance Limits for Fault Localization", *IEEE Journal on Selected Areas in Communications*, vol. 38, no.1, pp. 180-192, Jan 2020
- Raksha Ramakrishna and Anna Scaglione, "Detection of False Data Injection Attack using Graph Signal Processing for the Power Grid", *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2019*
- Raksha Ramakrishna and Anna Scaglione, "On Modeling Voltage Phasor Measurements as Graph Signals", *IEEE Data Science Workshop (DSW), 2019*

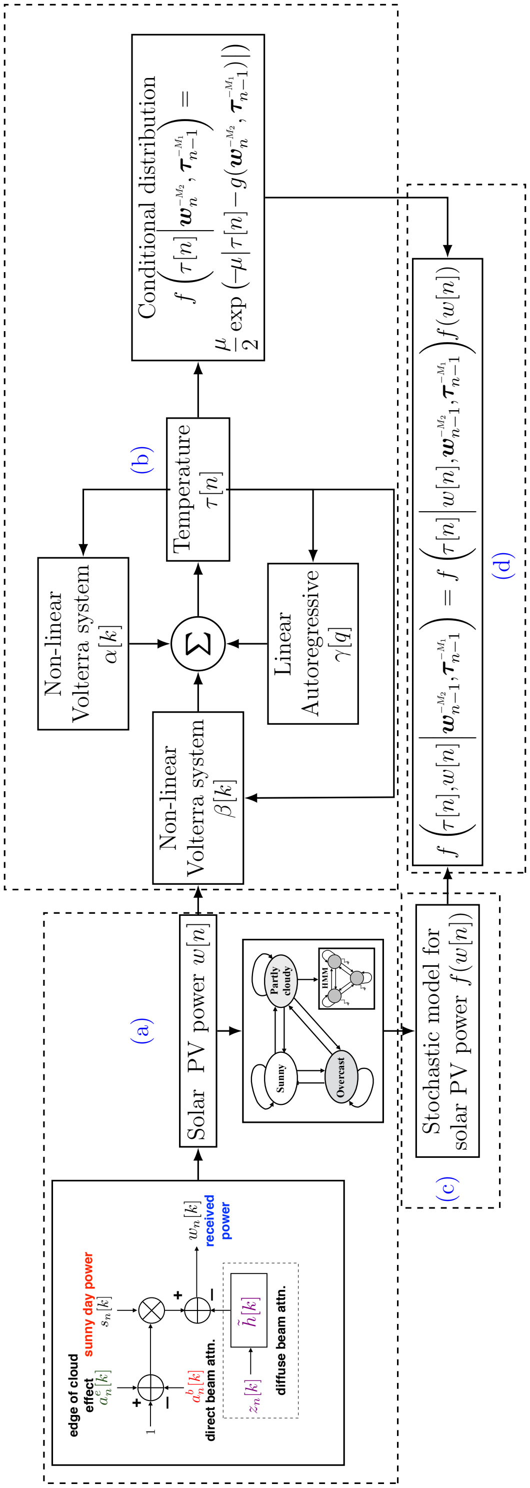


Figure 1.3.2: The block diagram describing the overall joint system modeling approach in Chapter 5. (a) A deterministic system model for solar PV power was introduced in [39]. (b) A joint system model for both solar PV power and temperature was introduced in [40]. (c) Stochastic model for solar PV power is introduced in [41] so that it can be used for (d) joint probabilistic forecasting methods.

1.3 Model for Solar Photo-voltaic Power and Temperature for Renewable Integration

Solar power generation, both from solar photo-voltaic (PV) farms and roof-top solar panel installations is on the rise and its impact is increasingly felt in traditional energy markets. Being a weather-dependent natural resource, attached to this renewable energy production is certain level of variability and uncertainty that cannot be controlled. This calls for greater amounts of reserves to be online to compensate for renewable generation shortage. On the other hand, electricity consumption is also affected by the weather. In fact, the so-called thermostatically controlled loads (TCL), i.e. the appliances used for heating and/or cooling, have a consumption pattern that has direct dependence on outdoor temperature. Naturally, solar irradiation affects both solar power production as well as temperature, making supply of solar power and TCL demand statistically dependent. This is illustrated in Fig.1.3.1. These underlying dependencies, however, are not leveraged in traditional power systems operations. Today, there is significant interest in introducing stochastic optimization tools in power systems operations [42, 43] and the challenge is the explosion of complexity associated with representing such uncertainty. Typically, accounting for uncertainty involves constructing scenario trees [44] i.e., quantization of multidimensional random processes with random variables being load, renewable generation and electricity prices. Treating solar production and TCL load, or demand response, as independent random variables is inaccurate and inefficient besides increasing the domain of the associated random variables. By capturing the relationship between temperature and solar power, using the model presented in this work, one can relate renewable infeed (solar PV power) with TCLs (see Fig.1.3.1). Such modeling leads to reduction in the size of the sample space by excluding the unlikely joint events of occurrence of a certain PV output power and TCL consumption. A smaller sample space neglects low-probability outcomes and limits the computational complexity while computing expected cost which is the objective function in a stochastic optimization problem. [42]

The main goal of Chapter 5 is to provide a mathematical model for the joint uncertainty in solar PV power and temperature which can also be used to produce probabilistic forecasts.The modeling approach is motivated by the physics of the problem which helps in understanding the underlying phenomenon and provides an easier interpretation of the results obtained. The proposed forecasting method provides *parametric* predictive densities which are useful for creating scenario trees for stochastic optimization to take decisions under uncertainty. The premise is that solar irradiation is the cause of both solar PV power and temperature and therefore acts as a cofounding or lurking variable [45]. Solar PV power exclusively depends on solar irradiance which means that the random variable of solar power is a direct transformation of irradiance. The uncertainty in solar power or irradiance is dictated almost entirely by the cloud cover. The residual uncertainty in temperature given solar power is then captured by the means of a conditional density function. A vast array of literature exists in the area of forecasts for solar power. On the other hand, there are very few papers that model the relationship that solar PV power has with TCLs. These are reviewed separately in the subsequent paragraphs.

1.3.1 Prior Works and Contributions

Solar PV power forecasting: A majority of approaches taken to model solar power generation can be broadly classified as being physical, statistical or a hybrid of the two methods (see e.g. [46] for a review). Physical methods employ astronomical relationships [47], meteorological conditions and numerical weather predictions

(NWPs) for an improved forecast [48–51]. Other papers use static images of clouds recorded by a total sky imager (TSI) [52] recording cloud motion [53] to predict solar power. These models rely on a deterministic mapping given additional information to produce an estimate of the power generated by the panel.

One of the most prevalent statistical methods for solar power forecasting include time-series modeling such as using autoregressive (AR) models [54–56] that are designed to model stationary processes. They assume that some transformation such as dividing by the clear sky power time series makes the series stationary. Such assumptions may not be sufficient to fully capture the non-stationarity of solar power production. Alternatively, there are other approaches such as autoregressive integrated moving average (ARIMA) and autoregressive with exogenous input (ARX) [46,57,58] based non-stationary methods for solar power prediction. In the same class of statistical methods there exist other works that capture variability in solar PV power in a multitude of ways [59–61] including the use of black box methods like artificial neural networks (ANN) [62, 63]. Additionally, there are also methods assuming that solar PV power is a Markov process [64, 65] for forecasting. The body of work on probabilistic forecasting of solar PV power is relatively small. In [66], a non-parametric kernel density estimation method is used to fit a probability distribution to PV power. In [67], a higher order Markov chain is used to characterize solar PV power and operating points based on temperature are also used to classify different PV systems and then Gaussian mixture models (GMM) are used for probabilistic forecasts. In [68], a combination of beta distribution, quantile regression and Markov model is used for probabilistic forecasting. More recently, producing parametric predictive densities was undertaken in [69] where beta and two-sided power distributions were used to provide an aggregated forecast distribution. A review of approaches for probabilistic forecasting of solar PV power can be found in [70].

Joint modeling of solar PV power and temperature: Empirical studies such as in [71,72] and data driven approaches like in [73] demonstrated a relationship between temperature and PV power. In [74], ARMA models are utilized to model PV power generation, load and wind generation separately, and their profiles are used to investigate reliability of the resultant hybrid power system which consists of renewable infeed and flexible loads. In the area of stochastic optimization, clustering methods such as k-nearest neighbors (k-NN) along with auxiliary variables such as temperature are used in [75] to create k scenarios for PV power and TCL load. Correlation was accounted for in this case; however, equal probabilities were assigned to all k scenarios. Self-consumption of photo-voltaic power in buildings, i.e. usage of rooftop solar PV power production for satisfying the load requirements within a building has also been studied to some extent. See [76], for a literature review.

The goal of Chapter 5 is to present a unified forecasting approach using a joint stochastic model for solar PV power and temperature that is physically inspired. To highlight this, Fig. 1.3.2 illustrates the block diagram of the joint stochastic model. The deterministic models for solar PV power and temperature were introduced in [39] and [40] respectively. In this chapter, a stochastic model for solar PV power is introduced and a forecasting methodology is developed. In combination with the model for temperature [40], a joint stochastic model is presented.

The proposed model for solar PV power at a macro-level defines a regime switching process [77] that classifies the PV production under one of the following three settings: *sunny, partly cloudy, overcast.* The stochastic models for *sunny* and *overcast* are Gaussian distributions whereas for the *partly cloudy* regime, a hidden Markov model (HMM) is proposed. Such an approach simplifies the understanding of temporal variations in solar PV power by examining each regime separately. It is also important to note that no assumption of stationarity is made while describing the regime switching process and no attempt is made to estimate this time-varying transition probability. In this manner, the proposed method uniquely captures the non-stationarity in solar power, which is not just due to its diurnal structure, and maps that onto the non-stationarity of the temperature process. The key contributions of the chapter are:

1. The proposition of a regime-switching process for solar PV power that consists of periods that can be classified as *sunny*, *overcast* and *partly cloudy* and development of stochastic models for the three regimes.
2. A HMM for the *partly cloudy* regime whose latent states are the support of sparse parameters pertaining to attenuation of power.
3. A classification algorithm to identify the present regime using solar power data and use it for solar power prediction. No other auxiliary data such as wind speed is used.
4. A Volterra model to capture the correlation between temperature and solar PV power for a joint forecast of temperature and solar PV power in the probabilistic sense using the reduced size uncertainty model that emerges from the proposed modeling approach.

In Chapter 5, a spatial model for solar PV power is proposed wherein the fact that all the panels in a small enough geographical area see the same levels of irradiance and cloudiness. This observation is consistent with the modeling of solar PV power and temperature as graph signals [5] and helps to tackle the problem of partial shading in solar panels as explained in the following subsection.

1.3.2 Partial Shading in Solar Panels

Partial shading type of faults in solar panels reduce the power output from solar panels to a value below their operating point since panels are shaded due to dirt, soiling, buildings or tree shadows [78, 79]. As a consequence, regular maintenance is of utmost importance in solar PV panel arrays. However, there could be thousands of panels and human inspection to determine which panels are soiled or shaded is an impossible task. Besides, such soiling errors can go unnoticed for a very long time and slowly degrade the performance of the system.

In Chapter 6, a distributed algorithm that can be applied both in the context of a utility-scale array of solar panels or multiple residential installations of solar panels within a certain geographical region is proposed. The underlying principle is simple: all the panels in a small enough geographical area see the same levels of irradiance and cloudiness as corroborated by real-data and are consistent with the graph signal model. This lends to the problem a low-rank structure which we utilize to detect the fraction of *partial-shading* or soiling of solar panels in multiple installations. Since it is assumed that very few panels are shaded, the infrequent nature of shading can be modeled by assuming a prior distribution that promotes sparsity, such as that of independent truncated exponentially distributed shading coefficients. Near-neighbor communication based methods are a natural extension to the existing state of the art which use just local information for fault detection. Additionally, such communications could be multipurpose and be used not just for fault detection but for general monitoring and participation in demand-response programs as an aggregate load. This serves as motivation for a decentralized algorithm that is used to find the common feature of cloud-induced shading among the nodes with fewer number of data exchanges whereas the specific partial-shading parameter is computed locally.

Under these assumptions, the problem of partial shading estimation can be cast as a decentralized matrix decomposition into a rank-1 matrix, where the left singular vector represents the common attenuation pattern due to the clouds, and it is estimated in a decentralized way via average consensus among agents, while the right singular vector (fractions of partial shading) is *sparse* and it is estimated locally, alternating between the two estimates. It is noted that existing methods, such as the decentralized power iteration methods [80] could be applied, but they would not leverage fully the structure of the problem.

Related works: Many papers address the fault detection problem. They range from utilizing panel specific electrical characteristics such as their operating points that depend on solar irradiance levels [81] to applying machine learning algorithms [82–84] including k-means clustering [79] and minimum covariance determinant [78,85] methods and general statistical methods like multi-resolution signal analysis [86, 87]. A review of machine learning techniques for fault detection in solar panels can be found in [88]. Also, a more recent paper [89] detects faults using concepts from graph signal processing (GSP). Most of the aforementioned studies focus mainly on a single installation with an array of solar PV panels. However, in this work, structure in solar data from different installations is used to estimate the fraction of partial shading instead of using additional data like irradiance or temperature.

1.3.3 Related Publications

Work discussed in the second part of the dissertation can be found in

- Raksha Ramakrishna and Anna Scaglione ,"A Bayesian lower bound for parameters with bounded support priors", *2020 54th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, USA, 2020, pp. 1-6
- Raksha Ramakrishna, Anna Scaglione, Vijay Vittal, Emiliano Dall'Anese,Andrey

Bernstein, “A Model for Joint Probabilistic Forecast of Solar Photo-voltaic Power and Outdoor Temperature”, *IEEE Transactions on Signal Processing*, vol. 67 no. 24, pp. 6368-6383, Dec 2019

- Raksha Ramakrishna, Anna Scaglione, Andreas Spanias, Cihan Tepedelenlioglu, “Distributed Bayesian Estimation with Low-rank Data: Application to Solar Array Processing”, *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*
- Raksha Ramakrishna, Andrey Bernstein, Emiliano Dall’Anese, Anna Scaglione, “Joint Probabilistic Forecasts of Temperature and Solar Irradiance”, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018*
- Raksha Ramakrishna and Anna Scaglione,“A Compressive Sensing Framework for Solar Photo-Voltaic Power”, in *Conference Record of the Fiftieth Asilomar Conference on Signals, Systems and Computers, 2016*

1.4 Notation

In this dissertation, lowercase letters x are used to denote scalars, boldfaced lowercase letters are used for vectors, $\boldsymbol{x}$ and uppercase for matrices, $\mathbf{A}$. Transpose of vector or matrix is $\boldsymbol{x}^\top, \mathbf{A}^\top$ and conjugate transpose as $\mathbf{A}^H$. $[\boldsymbol{x}]_{\mathcal{M}}$ denotes the new vector that contains elements of the vector indexed by the set $\mathcal{M}$. The operation $\Re\{.\}, \Im\{.\}$ denote the real and imaginary parts of the argument. Pseudo-inverse of a matrix is indicated as $\mathbf{A}^\dagger$. The operation $\texttt{Diag}(.)$ implies extraction of the diagonal values of a matrix $\mathbf{A}$ when the input is a matrix and creating a diagonal matrix with elements from a vector if the input is a vector $\boldsymbol{x}$.

Chapter 2

GRAPH SIGNAL PROCESSING

The need for *processing* graph data has led to the emerging field of graph signal processing (GSP), which takes a deterministic and system theoretic approach to justify the properties of graph data and to inspire the associated signal processing algorithms. A cornerstone of GSP is the formal definition of graph filter, which extends the notions of linear time invariant (LTI) filtering of time series signals to processing data defined on a graph, a.k.a. graph signals. Classical signal processing tools such as frequency analysis have been successfully applied with analogous interpretation to graph data, generating new insights for data science. The notion of graph filters can be used to define generative models for graph data. In fact, the data obtained from many examples of network dynamics may be viewed as the output of a graph filter. With this interpretation, the specific class of graph data, where the generating graph filters are low-pass, i.e., the filter attenuates contents in the higher graph frequencies while retaining contents in the lower frequencies is studied in this chapter. This choice is motivated by the prevalence of low-pass models in application domains such as social networks, financial markets, and power systems. As a motivating example, in Fig. 2.0.1, a few real datasets with such models from social networks, power systems and financial market, and show the eigenvalues of their sample covariance matrices. A salient feature observed is that these sample covariance matrices are low-rank, thus displaying an important symptom of low-pass filtered graph signals. Also, in Fig. 2.0.2, eigenvalues of sample covariance matrices of solar PV power and temperature are shown for an area around San Francisco, California to highlight again the low-rank nature of sample covariance matrices which in turn are indicative of the fact that

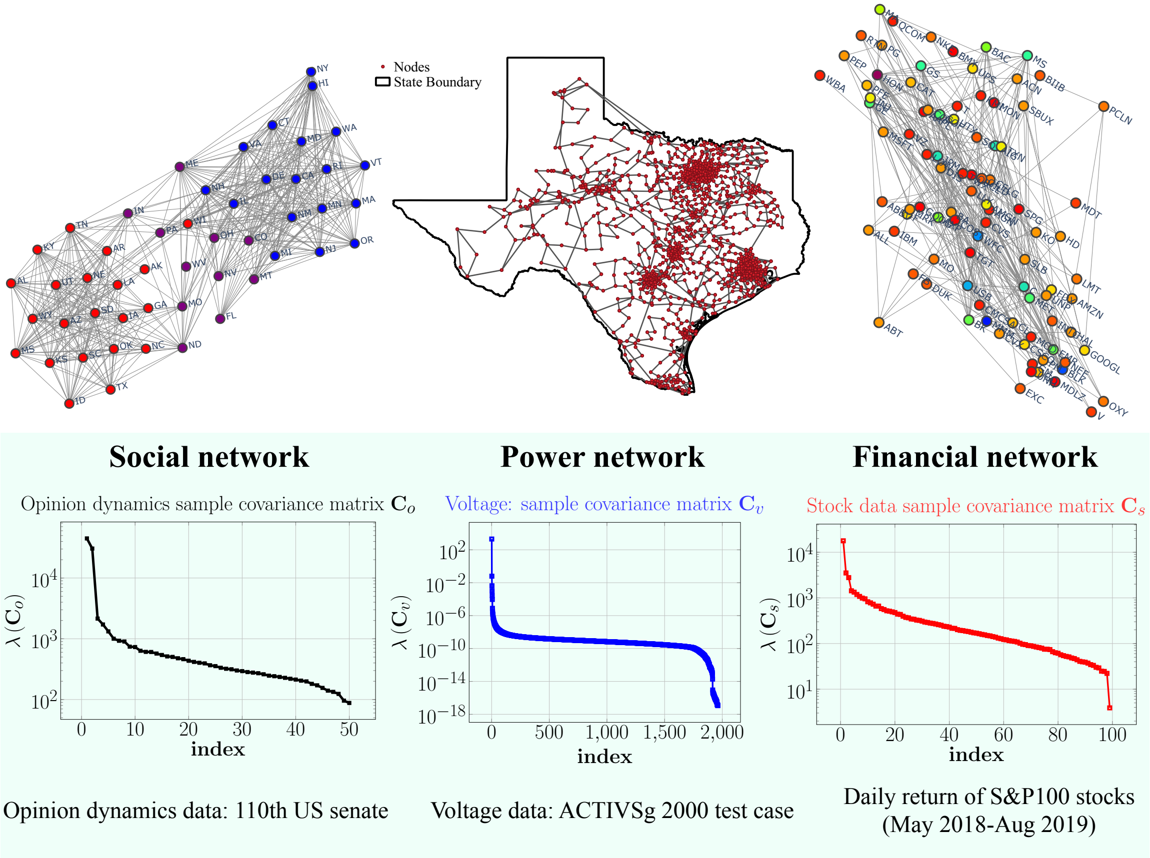


Figure 2.0.1: Illustrating the eigenvalues or spectra of sample data covariance matrix of voltage, Senate rollcall, and financial stock data. These data admit physical/social models that can be regarded as low-pass filtered graph signals. A salient feature of their low-pass nature is observed as the low-rank property of the sample covariance matrices.

solar PV power and temperature can also be modeled as low-pass graph signals. The low-pass property of solar radiation has been studied in [5] and outdoor temperature in [6] and used for data compression.

In this chapter, the basics of Graph Signal Processing (GSP) are briefly reviewed specifically from the standpoint of low-pass GSP in Section 2.1. Then, a few examples of graph data general processes that can be modeled as outputs of low-pass graph filters are discussed in Section 2.2. Lastly, in Section 2.3, complex-valued GSP is discussed to lay down the foundation on which the GSP framework for the power grid is built in Chapter 3.

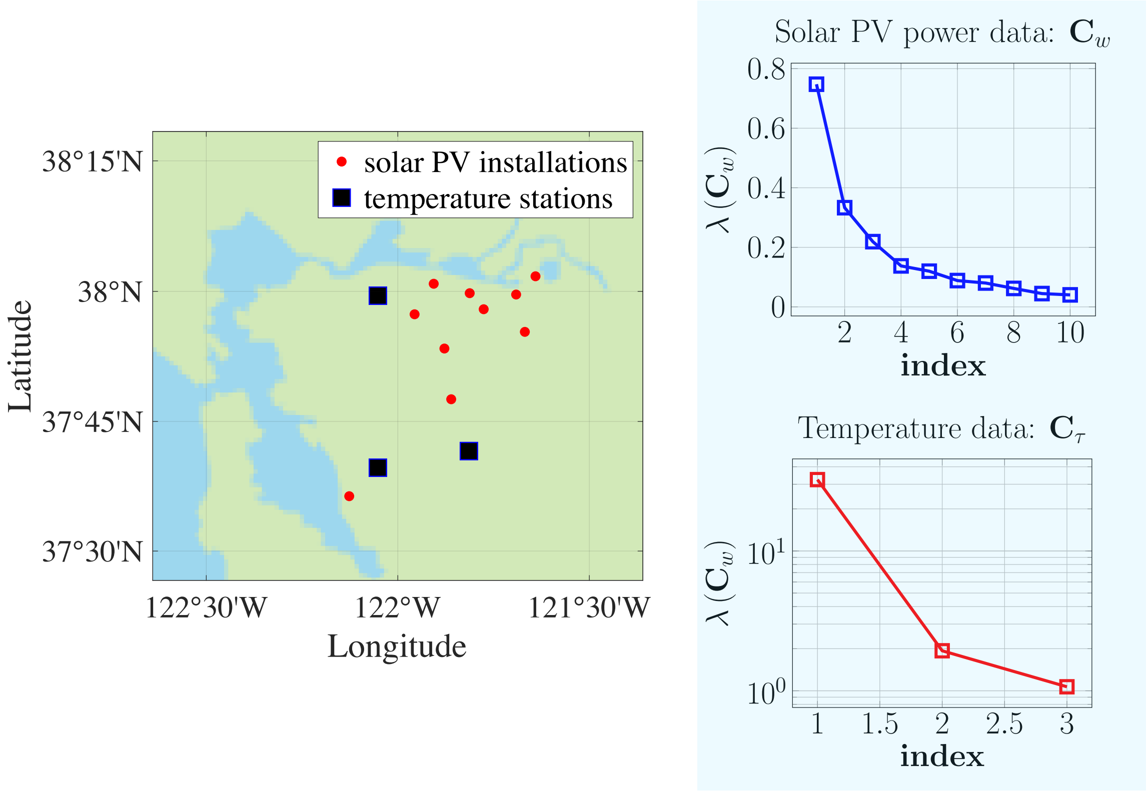


Figure 2.0.2: Eigenvalues or spectra of sample data covariance matrix of solar PV power and temperature. Solar PV power data is obtained from SolarCity dataset that is also used in [39, 41]. Power data is normalized by the sunny day pattern at the location to remove panel related variability. Temperature data obtained from National Oceanic and Atmospheric Administration (NOAA) weather stations.

2.1 Basics of Graph Signal Processing

In this section, fundamental concepts of GSP and a formal definition of low-pass graph filters/signals are reviewed for real-valued signals. For more details, please refer to the excellent prior overview articles such as [7, 8].

Consider a weighted undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with n nodes such that $\mathcal{N} = \{1, \ldots, n\}$ and $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ is the edge set. A graph signal is a function $x : \mathcal{N} \rightarrow \mathbb{R}$ which can be represented by a n-dimensional vector $\boldsymbol{x} = (x(i))_{i \in \mathcal{N}}$. The set of nodes connected to node i is called the *neighborhood* of i and denoted as $\mathcal{N}_i$. GSP generalizes the notion of *discrete time shift* for a time series by introducing the notion of graph shift operator (GSO):

Definition 1. *A graph shift operator (GSO) is a linear neighborhood operator, such that each entry of the shifted graph signal is a linear combination of the graph signal neighbors' values [90].*

Generally, the GSO is a matrix $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ satisfying $[\boldsymbol{S}]_{ij} \neq 0$ if and only if $i = j$ or $(i, j) \in \mathcal{E}$. When multiplied by a graph signal $\boldsymbol{x}$, each entry of the shifted graph signal is a linear combination of the one-hop neighbors' values, therefore 'shifting' the graph signal with respect to the graph topology. For the simplicity of exposition, Laplacian matrix is regarded as the GSO although it is not the only choice. The Laplacian matrix is defined as $\boldsymbol{L} := \boldsymbol{D} - \boldsymbol{A}$, where $\boldsymbol{A}$ is the weighted symmetric adjacency matrix of $\mathcal{G}$, and $\boldsymbol{D} = \texttt{Diag}(\boldsymbol{A}\mathbf{1})$ is a diagonal matrix of the weighted degrees. It is also common to take the GSO as the normalized Laplacian matrix, or the adjacency matrix [6].

Using the definition of GSO, one can measure the smoothness of graph signals and analyze their content in the graph frequency domain. Recall that if a signal is smooth in time, the norm of its time derivative is small. For a graph signal $\boldsymbol{x}$, its *graph derivative* is defined as

$$[\nabla \boldsymbol{x}]_{ij} = \sqrt{A_{ij}}(x_i - x_j).$$

The squared Frobenius norm of graph derivative, a.k.a. the *graph quadratic form* [7], provides an idea of the smoothness of the graph signal $\boldsymbol{x}$:

$$\mathrm{S}_2(\boldsymbol{x}) := \frac{1}{2}\|\nabla \boldsymbol{x}\|_{\mathrm{F}}^2 = \boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \sum_{i,j} A_{ij}(x_i - x_j)^2. \tag{2.1.1}$$

Observe that if $x_i \approx x_j$ for any neighboring nodes i, j, then $\mathrm{S}_2(\boldsymbol{x}) \approx 0$. A graph signal is *smooth* if $\mathrm{S}_2(\boldsymbol{x})/\|\boldsymbol{x}\|_2$ is small.

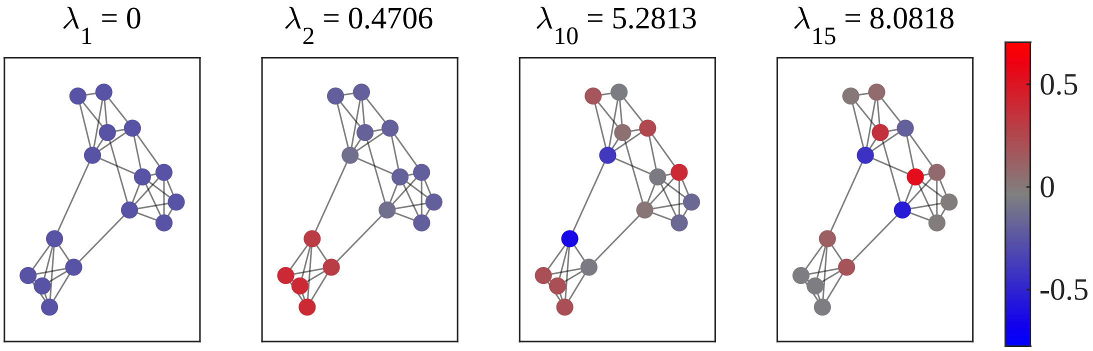


Figure 2.1.1: The GFT basis, $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_{10}, \mathbf{u}_{15}$, associated to the graph Laplacian of an undirected, unweighted graph with 15 nodes. As the eigenvalue increases, the eigenvectors tend to be more oscillatory.

Let the eigendecomposition of the Laplacian matrix be $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ and assume that it has eigenvalues of multiplicity one ordered as $\boldsymbol{\Lambda} = \texttt{Diag}(\lambda_1, \dots, \lambda_n)$ with $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_n$, and $\boldsymbol{U} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_n)$ with $\mathbf{u}_i \in \mathbb{R}^n$ being the eigenvector for λ_i. Observe that for any $\boldsymbol{x} \in \mathbb{R}^n$, it holds $\frac{\mathrm{S}_2(\boldsymbol{x})}{\|\boldsymbol{x}\|_2} \geq \frac{\mathrm{S}_2(\mathbf{u}_1)}{\|\mathbf{u}_1\|_2} = \lambda_1$, and for any $\boldsymbol{x}$ orthogonal to $\mathbf{u}_1$, it holds $\frac{\mathrm{S}_2(\boldsymbol{x})}{\|\boldsymbol{x}\|_2} \geq \frac{\mathrm{S}_2(\mathbf{u}_2)}{\|\mathbf{u}_2\|_2} = \lambda_2$ and so on for the other eigenvectors. The observation indicates that the larger the eigenvalue, the more oscillatory the eigenvector is over the vertex set. In particular, the smallest eigenvalue $\lambda_1 = 0$ is associated with the flat, all-ones eigenvector $\mathbf{u}_1 = (1/\sqrt{n})\mathbf{1}$, as seen in Fig. 2.1.1. The above motivates the definition of graph frequencies as the eigenvalues $\lambda_1, \dots, \lambda_n$ and of the Graph Fourier Transform (GFT) basis as the set of eigenvectors $\boldsymbol{U}$ [6]. Therefore, the i^{th} frequency component of $\boldsymbol{x}$ is defined as the inner product between $\mathbf{u}_i$ and $\boldsymbol{x}$:

$$\widetilde{x}_i = \mathbf{u}_i^\top \boldsymbol{x}, \ i = 1, \dots, n, \tag{2.1.2}$$

and $\widetilde{\boldsymbol{x}} = \boldsymbol{U}^\top \boldsymbol{x}$ is called the GFT of $\boldsymbol{x}$. The magnitude of the GFT vector $|\widetilde{\boldsymbol{x}}|$ is the 'spectrum' of the graph signal $\boldsymbol{x}$, where $|\widetilde{x}_i|^2$ represents the signal power at the λ_ith frequency.

It is worth noting that the GSO and Fourier transforms do not have in general important properties that are found in their conventional counterparts for time series. One notable fact is that the spectrum of $\boldsymbol{Lx}$ does not have the same amplitude as the spectrum of $\boldsymbol{x}$. In fact, the effect of the GSO is closer to that of a derivative, since each of the GFT coefficients is rescaled by the corresponding frequency.

An important concept to modeling data with GSP is the graph filtering operation. To this end, consider the definition of a shift-invariant operator and consequently the graph filter. Given a GSO $\boldsymbol{L}$, a *shift-invariant operator* $\mathcal{H}$ acting on a graph signal is such that:

$$\mathcal{H} : \boldsymbol{x} \mapsto \boldsymbol{y}, \qquad \Longleftrightarrow \qquad \mathcal{H} : \boldsymbol{Lx} \mapsto \boldsymbol{Ly}. \tag{2.1.3}$$

As it turns out, these operators must be linear and, more specifically, matrix polynomials of the GSO $\boldsymbol{L}$. Hence:

Definition 2. *A graph filter is a linear operator:*

$$\boldsymbol{y} = \mathcal{H}(\boldsymbol{L})\boldsymbol{x}, \quad \mathcal{H}\left(\boldsymbol{L}\right) = \sum_{p=0}^{P-1} h_p \boldsymbol{L}^p = \boldsymbol{U}\left(\textstyle\sum_{p=0}^{P-1} h_p \boldsymbol{\Lambda}^p \right)\boldsymbol{U}^\top, \tag{2.1.4}$$

where P is the *filter order* and $\{h_p\}_{p=0}^{P-1}$ are the *filter's coefficients.* As a convention, use $\boldsymbol{L}^0 = \boldsymbol{I}$ and $\lambda_0^0 = 0^0 = 1$. Note that by Cayley-Hamilton theorem, any matrix polynomial can be represented as in (2.1.4) with $P \leq n$. From (2.1.4), one can see that the graph filter has similar interpretation as an LTI filter in discrete-time signal processing where the former replaces the time shifts by powers of the GSO. Meanwhile, the second expression in (2.1.4) defines the *frequency response* as the diagonal matrix $h(\boldsymbol{\Lambda}) := \sum_{p=0}^{P-1} h_p \boldsymbol{\Lambda}^p$. A graph signal $\boldsymbol{y}$ is said to be filtered by $\mathcal{H}(\boldsymbol{L})$ with the input excitation $\boldsymbol{x}$ when

$$\boldsymbol{y} = \mathcal{H}(\boldsymbol{L})\boldsymbol{x}. \tag{2.1.5}$$

To better appreciate the effects of the graph filter, note that the ith frequency component of $\boldsymbol{y}$ is:

$$\widetilde{y}_i = h(\lambda_i) \cdot \widetilde{x}_i, \; i = 1, \ldots, n, \tag{2.1.6}$$

where $h(\lambda) := \sum_{p=0}^{P-1} h_p \lambda^p$ is the transfer function of the graph filter, or equivalently, $\widetilde{\boldsymbol{y}} = h(\boldsymbol{\lambda}) \odot \widetilde{\boldsymbol{x}}$. It is similar to the convolution theorem in discrete-time signal processing.

Inspired by (2.1.6), the ideal low-pass graph filter with a cut-off frequency λ_k is defined through setting the transfer function as $h(\lambda) = 1, \lambda \leq \lambda_k$ and 0 otherwise [91]. Alternatively, one can say that a graph filter is low-pass if its frequency response is concentrated on the low graph frequencies. In this chapter, the following definition from [38] is adopted:

Definition 3. *For any* $1 \leq k \leq n-1$*, define the ratio*

$$\eta_k := \frac{\max\{|h(\lambda_{k+1})|, \ldots, |h(\lambda_n)|\}}{\min\{|h(\lambda_1)|, \ldots, |h(\lambda_k)|\}}. \tag{2.1.7}$$

The graph filter $\mathcal{H}(\boldsymbol{L})$ *is* k*-low-pass if and only if the low-pass ratio* η_k *satisfies* $\eta_k \in [0, 1)$*.*

The integer parameter k characterizes the *bandwidth*, or the cut-off frequency of the low-pass filter is at λ_k. The ratio η_k quantifies the ‘strength’ of the low-pass graph filter. Upon passing a graph signal through $\mathcal{H}(\boldsymbol{L})$, the high frequency components (above λ_k) are attenuated by a factor of less than or equal to η_k. Using this definition, the ideal k-low-pass graph filter has the ratio $\eta_k = 0$, whose filter order has to be at least $P \geq n - k + 1$ and transfer function $h(\lambda)$ has $\{\lambda_{k+1}, \ldots, \lambda_n\}$ as its roots.

Finally, a *k-low-pass graph signal* refers to a graph signal that is the output of a k-low-pass filter, subject to a ‘well-behaved’ excitation, i.e. does not possess strong high-pass components, which includes, but is not limited to, the white noise.

2.1.1 The Impact of Graph Topologies on Filter Order

From Definition 3, observe that the low-pass ratio η_k of a graph filter depends on the filter's coefficients $\{h_p\}_{p=0}^{P-1}$ and the graph Laplacian matrix's spectrum $\lambda_1, \ldots, \lambda_n$.

The condition $\lambda_k \ll \lambda_{k+1}$ facilitates the design of a k-low-pass graph filter with a favorable ratio $\eta_k \ll 1$ and a small filter order P. For example, the order-1 graph filter $\mathcal{H}(\boldsymbol{L}) = \boldsymbol{I} - \lambda_n^{-1}\boldsymbol{L}$ is k-low-pass with the ratio $\eta_k = \frac{\lambda_n - \lambda_{k+1}}{\lambda_n - \lambda_k} = 1 - \frac{\lambda_k - \lambda_{k+1}}{\lambda_n - \lambda_k}$, where η_k is small if $\lambda_k \ll \lambda_{k+1}$.

An example of graph topologies favoring the condition $\lambda_k \ll \lambda_{k+1}$ is the stochastic block model (SBM) [92] for describing random graphs with k blocks/communities with nodes in $\mathcal{N}$ partitioned as $\mathcal{N}_1, \ldots, \mathcal{N}_k$. Consider a simplified SBM with k equal-sized blocks specified by a membership matrix $\boldsymbol{Z} \in \{0,1\}^{n \times k}$ such that $Z_{i\ell} = 1$ if and only if $i \in \mathcal{N}_\ell$; and a latent model $\boldsymbol{B} \in [0,1]^{k \times k}$ where $B_{j,\ell}$ is the probability of edges between nodes in block j and ℓ. Consider the homogeneous planted partition model (PPM) such that $\boldsymbol{B} = b\boldsymbol{1}\boldsymbol{1}^\top + a\boldsymbol{I}$ with $b, a > 0$. With the above specification, the adjacency matrix $\boldsymbol{A}$ is a symmetric binary matrix with independent entries satisfying $\mathbb{E}[\boldsymbol{A}] = \boldsymbol{Z}\boldsymbol{B}\boldsymbol{Z}^\top$. When the graph size grows to infinity ($n \to \infty$), the Laplacian matrix of an SBM-PPM graph converges almost surely to its expected value [92, Theorem 2.1]:

$$\boldsymbol{L} \xrightarrow{a.s.} \mathbb{E}[\boldsymbol{L}] = \frac{n(a+kb)}{k}\boldsymbol{I} - \boldsymbol{Z}\Big(b\boldsymbol{1}\boldsymbol{1}^\top + a\boldsymbol{I}\Big)\boldsymbol{Z}^\top. \tag{2.1.8}$$

From the above, it can be shown that $\lambda_{k+1} - \lambda_k = \frac{na}{k} \gg 1$ for $\mathbb{E}[\boldsymbol{L}]$, i.e., a favorable graph model for k-low-pass graph filters. Lastly, the bottom-k eigenvectors of the expected Laplacian associated with $\lambda_1, \ldots, \lambda_k$ can be collected into the matrix $\sqrt{\frac{k}{n}}\boldsymbol{Z}\boldsymbol{P}$, where $\boldsymbol{P}$ diagonalizes the matrix $\boldsymbol{B}$. In other words, the eigenvectors corresponding to the bottom-k eigenvalues of $\boldsymbol{L}$ will reveal the block structure.

In contrast, the Erdös-Rényi graphs have Laplacian matrices that do not generally satisfy $\lambda_k \ll \lambda_{k+1}$. In fact, asymptotically ($n \to \infty$) the empirical distribution of the eigenvalues of Laplacian matrices tends to the free convolution of the standard Gaussian distribution and the Wigners semi-circular law [93]. Such spectrum does not favor the design of a k-low-pass graph filter with $\eta_k \ll 1$, reflecting the fact that block structure or communities do not emerge in Erdös Rényi graphs.

2.1.2 Low-pass Graph-Temporal Filter

When the excitation to a graph filter is of time-varying nature, and the topology is fixed, consider a graph-temporal filter [94] with the impulse response:

$$\mathcal{H}(\boldsymbol{L}, t) := \textstyle\sum_{p=0}^{P-1} h_{p,t} \boldsymbol{L}^p, \tag{2.1.9}$$

such that the graph filter's output is given by the time-domain convolution $\boldsymbol{y}_t = \sum_{s=0}^{t} \mathcal{H}(\boldsymbol{L}, t-s)\boldsymbol{x}_s$. The filter is causal and $\boldsymbol{x}_s = \boldsymbol{0}$ for $s < 0$. By applying z-transform and the GFT to the graph signal process $\{\boldsymbol{x}_t\}_{t \geq 0}$, it is possible to obtain the z-GFT signal, $\widetilde{\boldsymbol{X}}(z)$, given by:

$$\boldsymbol{X}(z) = \textstyle\sum_{t=0}^{\infty} \boldsymbol{x}_t z^{-t}, \quad \widetilde{\boldsymbol{X}}(z) = \boldsymbol{U}^\top \boldsymbol{X}(z), \tag{2.1.10}$$

which represents $\{\boldsymbol{x}_t\}_{t \geq 0}$ in the joint z-graph frequency domain. Then, the input-output relation and graph-temporal joint transfer function are obtained as

$$\widetilde{\boldsymbol{Y}}(z) = \widetilde{\boldsymbol{h}}(z) \odot \widetilde{\boldsymbol{X}}(z), \tag{2.1.11}$$

$$\mathbb{H}(\lambda, z) := \sum_{t=0}^{\infty} \sum_{p=0}^{P-1} h_{p,t} \lambda^p z^{-t}, \tag{2.1.12}$$

respectively. A class of graph-temporal filters for modeling graph signal processes is the GF-ARMA (q, r) filter, whose input-output relation in time domain and z-GFT domain are described below, respectively:

$$\begin{aligned} &\boldsymbol{y}_t - \mathcal{A}_1(\boldsymbol{L})\boldsymbol{y}_{t-1} \cdots - \mathcal{A}_q(\boldsymbol{L})\boldsymbol{y}_{t-q} = \mathcal{B}_0(\boldsymbol{L})\boldsymbol{x}_t + \cdots + \mathcal{B}_r(\boldsymbol{L})\boldsymbol{x}_{t-r}, \\ &\widetilde{\boldsymbol{a}}(z) \odot \widetilde{\boldsymbol{Y}}(z) = \widetilde{\boldsymbol{b}}(z) \odot \widetilde{\boldsymbol{X}}(z), \end{aligned}$$

where the z-transform of the graph frequency responses of the graph filter taps $\{\mathcal{A}_s(\boldsymbol{L})\}_{s=1}^{q}$, $\{\mathcal{B}_s(\boldsymbol{L})\}_{s=0}^{r}$ for the GF-ARMA (q, r) filter are

$$\widetilde{\boldsymbol{a}}(z) = 1 - \sum_{s=1}^{q} \widetilde{\boldsymbol{a}}_s z^{-s}, \quad \widetilde{\boldsymbol{b}}(z) = \sum_{s=0}^{r} \widetilde{\boldsymbol{b}}_s z^{-s} \tag{2.1.13}$$

Note that the joint frequency response is given by

$$\mathbb{H}(\lambda_i, z) = \frac{[\widetilde{\boldsymbol{b}}(z)]_i}{[\widetilde{\boldsymbol{a}}(z)]_i}, \tag{2.1.14}$$

whose poles and zeros may vary depending on the graph frequencies $\lambda_1, \ldots, \lambda_n$. A relevant case is when $\mathcal{H}(\boldsymbol{L}, t)$ is a *low-pass graph-temporal filter*. Similar to Definition 3, $\mathcal{H}(\boldsymbol{L}, t)$ is said to be low-pass with a cutoff frequency (λ_k, ω_0) and ratio η_k if:

$$\eta_k = \frac{\max_{\lambda \in \{\lambda_{k+1}, \ldots, \lambda_n\}, \omega \in (\omega_0, 2\pi)} |\mathbb{H}(\lambda, e^{j\omega})|}{\min_{\lambda \in \{\lambda_1, \ldots, \lambda_k\}, \omega \in [0, \omega_0]} |\mathbb{H}(\lambda, e^{j\omega})|} < 1. \tag{2.1.15}$$

Graph signals filtered by a low-pass graph-temporal filter are also commonly found in applications, as illustrated next.

2.2 Models of Low-pass Graph Signals

It turns out that many physical and social processes are naturally characterized by low-pass graph filters. In this section, various such examples are presented briefly while showing that their generation processes can be represented as outputs from low-pass graph filters.

2.2.1 Diffusion Model

The first case pertains to observations from a diffusion process, whose variants are broadly applicable in network science. As an example, consider the heat diffusion model in [95]. In this example, the relevant graph is a proximity graph where each node $i \in \mathcal{N}$ is a location (e.g., cities), and if locations i, j are close to each other, then $(i,j) \in \mathcal{E}$. The graph is endowed with a symmetric weighted adjacency matrix encoding the distance between locations. The graph signal $\boldsymbol{y}_t \in \mathbb{R}^n$ encodes the temperature of n locations at time t, and let $\boldsymbol{x}_0 \in \mathbb{R}^n$ be the initial heat distribution. The temperature of a location is diffused to its neighbors. Let $\sigma > 0$ be a constant, then,

$$\boldsymbol{y}_t = e^{-t\sigma \boldsymbol{L}} \boldsymbol{x}_0 = \left(\boldsymbol{I} - t\sigma \boldsymbol{L} + \frac{(t\sigma)^2}{2} \boldsymbol{L}^2 - \cdots \right) \boldsymbol{x}_0 \tag{2.2.1}$$

where (2.2.1) is a discretization of the heat diffusion equation [95]. As $\boldsymbol{L}\boldsymbol{1} = \boldsymbol{0}$, the matrix exponential $e^{-t\sigma \boldsymbol{L}} = \boldsymbol{I} - t\sigma \boldsymbol{L} + \frac{(t\sigma)^2}{2} \boldsymbol{L}^2 - \cdots$ is row stochastic. The temperature at time t is thus a weighted average of neighboring locations' temperatures at $t = 0$, i.e., this is a diffusion dynamical process.

To understand (2.2.1) under the context of low-pass filtering, observe that $\boldsymbol{y}_t$ is a filtered graph signal with the excitation $\boldsymbol{x}_0$ and the graph filter $\mathcal{H}(\boldsymbol{L}) = e^{-t\sigma \boldsymbol{L}}$. Also verify that $\mathcal{H}(\boldsymbol{L})$ is k-low-pass with Definition 3 for any $k < n$. Note that the low-pass ratio η_k is:

$$\frac{e^{-t\sigma\lambda_{k+1}}}{e^{-t\sigma\lambda_k}} = e^{-t\sigma(\lambda_{k+1} - \lambda_k)}.$$

As $\lambda_{k+1} > \lambda_k$ and $t\sigma > 0$, note that $\mathcal{H}(\boldsymbol{L})$ is a k-low-pass graph filter for any $k = 1, \ldots, n-1$.

It is assumed that $\boldsymbol{x}_0$ is an impulse excitation affecting the system only at the initial time. In practice, the excitation signal may not be an impulse and the output graph signal $\boldsymbol{y}_t$ is expressed as the convolution $\boldsymbol{y}_t = \sum_{s=0}^{t} e^{-(t-s)\sigma \boldsymbol{L}} \boldsymbol{x}_s$. This corresponds to a low-pass graph-temporal filter with the joint transfer function $\mathbb{H}(\lambda, z) = (1 - e^{-\lambda\sigma} z^{-1})^{-1}$. Besides, the diffusion process is common in network science as similar models arise in contagion process and product adoption to name a few.

2.2.2 Opinion Dynamics

This example pertains to opinion data mined from social networks with the influence of external excitation [38, 96]. The relevant graph $\mathcal{G}$ is the social network graph where each node $i \in \mathcal{N}$ is an individual, and $\mathcal{E}$ is the set of friendships. Similar to the previous case study, this graph is endowed with a symmetric weighted adjacency matrix $\boldsymbol{A}$, where the weights measure the trust among pairs of individuals. Let $\alpha \in (0, \lambda_n^{-1})$, $\beta \in (0,1)$ be parameters of trust on others and susceptibility to external influence of an individual respectively. The evolution of opinions follows that of a combination of DeGroot's and Friedkin-Johnsen's model [97], which is a GF-AR(1) model:

$$\boldsymbol{y}_{t+1} = (1-\beta)\Big(\boldsymbol{I} - \alpha \boldsymbol{L}\Big)\boldsymbol{y}_t + \beta \boldsymbol{x}_t, \tag{2.2.2}$$

where $\boldsymbol{y}_t \in \mathbb{R}^n$ is a graph signal of the individuals' opinions at time t, and $\boldsymbol{x}_t \in \mathbb{R}^n$ is a graph signal of the external opinions perceived by the social network. Note that this also corresponds to a low-pass graph-temporal filter with the joint transfer function $\mathbb{H}(\lambda, z) = \beta[1 - (1-\beta)(1-\alpha\lambda)z^{-1}]^{-1}$.

To discuss the steady state of (2.2.2), assume that $\boldsymbol{x}_t \equiv \boldsymbol{x}$. Considering (2.2.2), observe that $\boldsymbol{y}_{t+1}$ is a convex combination of $\boldsymbol{x}$ and weighted average of the neighbors' opinions at time t that is formed by taking a weighted average of neighboring signals in $\boldsymbol{y}_t$ using a diffusion operator $\boldsymbol{I} - \alpha\boldsymbol{L}$. As $\beta > 0$, the recursion is stable, leading to the steady state (or equilibrium) opinions:

$$\boldsymbol{y} = \lim_{t \to \infty} \boldsymbol{y}_t = (\boldsymbol{I} + \widetilde{\alpha}\boldsymbol{L})^{-1}\,\boldsymbol{x} = \mathcal{H}(\boldsymbol{L})\boldsymbol{x}, \tag{2.2.3}$$

where $\widetilde{\alpha} = \beta(1-\alpha)/\alpha > 0$ and $\boldsymbol{y}$ is a filtered graph signal excited by $\boldsymbol{x}$.

The graph filter above is given by $\mathcal{H}(\boldsymbol{L}) = (\boldsymbol{I} + \widetilde{\alpha}\boldsymbol{L})^{-1}$. To verify that it is a k-low-pass graph filter with Definition 3, note that for any $k < n$, the low-pass ratio

η_k is

$$\frac{1+\widetilde{\alpha}\lambda_k}{1+\widetilde{\alpha}\lambda_{k+1}} = 1 - \widetilde{\alpha}\frac{\lambda_{k+1}-\lambda_k}{1+\widetilde{\alpha}\lambda_k}.$$

Again, observe that as $\lambda_{k+1} > \lambda_k$, the above graph filter is k-low-pass for any $k = 1, \ldots, n-1$. However, that this low-pass ratio may be undesirable with $\eta_k \approx 1$ when $\widetilde{\alpha} \ll 1$. Interestingly, a similar generative model as (2.2.2) is found in equilibrium problems such as quadratic games [98].

Two remarks are in order. First, social networks are typically directed, and this suggests using a non-symmetric shift operator as opposed to the symmetric Laplacian matrix, which is only used for simplicity of exposition. Second, many alternative models for social networks interactions are non-linear and linear GSP is insufficient in those contexts.

2.2.3 Finance Data

Financial systems such as stock market and hedge funds produce return reports periodically about their business performances. A collection of these reports can be studied as graph signals, where the relevant graph $\mathcal{G}$ consists of nodes $\mathcal{N}$ that are financial institutions, and edges $\mathcal{E}$ that are business ties between them. It has been studied [99] that business performances are correlated according to the business ties. Moreover, the returns are affected by a number of common factors [100]. Inspired by [99], [100, Ch. 12.2], let $\beta \in (0,1)$ be the strength of external influences, a reasonable model for the transient dynamics of the graph signal $\boldsymbol{y}_t$ of business performance measures is also a GF-AR(1):

$$\boldsymbol{y}_{t+1} = (1-\beta)\mathcal{H}(\boldsymbol{L})\boldsymbol{y}_t + \beta\boldsymbol{B}\boldsymbol{x}, \tag{2.2.4}$$

where $\mathcal{H}(\boldsymbol{L})$ is an unknown but low-pass graph filter, $\boldsymbol{B} \in \mathbb{R}^{n\times r}$ represents the factor model affecting financial institutions, and $\boldsymbol{x} \in \mathbb{R}^r$ is the excitation strength. The equilibrium of (2.2.4) is:

$$\boldsymbol{y} = \lim_{t\to\infty} \boldsymbol{y}_t = \left(\frac{1}{\beta}\boldsymbol{I} - \frac{\overline{\beta}}{\beta}\mathcal{H}(\boldsymbol{L}) \right)^{-1} \boldsymbol{B}\boldsymbol{x} \equiv \widetilde{\mathcal{H}}(\boldsymbol{L})\boldsymbol{B}\boldsymbol{x},$$

where $\overline{\beta} = 1 - \beta$. The excitation signal is $\boldsymbol{B}\boldsymbol{x}$ and the equilibrium $\boldsymbol{y}$ is the filter output. Suppose that $\mathcal{H}(\boldsymbol{L})$ is a k-low-pass graph filter with the frequency response satisfying $h(\lambda) \geq 0$, then for $\widetilde{\mathcal{H}}(\boldsymbol{L})$, the low-pass ratio η_k is evaluated as

$$1 - \frac{\overline{\beta}\Big\{ \min_{\ell=1,\ldots,k} h(\lambda_\ell) - \max_{\ell=k+1,\ldots,n} h(\lambda_\ell) \Big\}}{1 - \overline{\beta}\max_{\ell=k+1,\ldots,n} h(\lambda_\ell)}.$$

As $\min_{\ell=1,\ldots,k} h(\lambda_\ell) - \max_{\ell=k+1,\ldots,n} h(\lambda_\ell) > 0$ since $\mathcal{H}(\boldsymbol{L})$ is a k-low-pass graph filter itself, observe that $\widetilde{\mathcal{H}}(\boldsymbol{L})$ is again k-low-pass according to Definition 3.

For $\boldsymbol{y}$ to be a k-low-pass graph signal, one has to also assume that $\boldsymbol{B}\boldsymbol{x}$ is not high-pass (i.e., not orthogonal to a low-pass one). This is a mild assumption as the latent factor affecting financial institutions are either independent of the network, or are aligned with the communities. Above all, it is to be noted that (2.2.4) is an idealized model where determining the exact model is an open problem in economics, see [99, 100].

2.2.4 Power Systems

In the case of power systems, the relevant graph is the electrical transmission lines network. The nodes of this graph are referred to as *buses.* The graph signals considered are *complex-valued* voltage phasors. These voltage graph signals can be modeled as output of low-pass graph filter while considering the branch admittance matrix as the GSO. The detailed discussion of this model, which is a core contribution of this thesis is in Chapter 3.

2.3 Complex-valued Graph Signal Processing

Section 2.1 covered the basics of GSP. Certain concepts in GSP are to be considered differently if the graph signal is complex-valued which is the case with measurements from power systems. This section specifies GSP for complex-valued graph signals.

A complex-valued graph signal $\boldsymbol{x} \in \mathbb{C}^n$ is a vector whose ith entry $[\boldsymbol{x}]_i$ is associated to node $i \in \mathcal{N}$. As before, graph weighted Laplacian $\boldsymbol{L} \in \mathbb{C}^n$ is considered as the GSO, i.e.

$$[\boldsymbol{L}]_{i,j} = \begin{cases} \sum_{k \in \mathcal{N}_i} s_{i,k}, \; i = j \\ -s_{i,j}, \; i \neq j \end{cases} \tag{2.3.1}$$

The focus here is on *complex symmetric GSOs*, $\boldsymbol{L} = \boldsymbol{L}^\top$ as it is applicable to the GSO considered for the power grid. Consider the following eigenvalue decomposition of the complex symmetric GSO $\boldsymbol{L}$, given by Theorem 4.4.13 in [101] for diagonalizable complex symmetric matrices:

$$\boldsymbol{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top, \qquad \mathbf{U}^\top\mathbf{U} = \mathbf{U}\mathbf{U}^\top = \mathbb{I}. \tag{2.3.2}$$

Here $\boldsymbol{\Lambda}$ is the diagonal matrix with eigenvalues $\lambda_0, \lambda_1, \ldots \lambda_n$ on the principal diagonal and $\mathbf{U}$ are *complex orthogonal* eigenvectors. An equivalent concept of frequency domain in GSP is defined using eigenvalues and eigenvectors of the GSO.

Instead of using the graph quadratic form as in (2.1.1) with $p = 2$, smoothness of graph signal and the order of graph frequencies can be more generally defined via the total variation (TV) criterion [90,102] using the discrete p Dirichlet form $S_p(\boldsymbol{x})$ with $p = 1$ as in [103] as:

$$S_1(\boldsymbol{x}) = \|\boldsymbol{L}\boldsymbol{x}\|_1 \implies S_1(\boldsymbol{u}_i) = \|\boldsymbol{L}\boldsymbol{u}_i\|_1 = |\lambda_i| \, \|\boldsymbol{u}_i\|_1 \tag{2.3.3}$$

After normalizing the eigenvectors such that $\|\boldsymbol{u}_i\|_1 = 1\ \forall i$, it is clear that $S_1(\boldsymbol{u}_i) > S_1(\boldsymbol{u}_j) \implies |\lambda_i| > |\lambda_j|$. Hence, the *ascending* order of eigenvalues corresponds to increase in frequency, $|\lambda_0| = 0 \leq |\lambda_1| \leq |\lambda_2| \ldots |\lambda_{|\mathcal{N}|}|$. This ordering is not unique since two distinct complex eigenvalues can have the same magnitude.

The Graph Fourier Transform (GFT) basis is the complex orthogonal basis consisting of columns of $\mathbf{U}$ in (2.3.2). Hence, the GFT of a graph signal $\boldsymbol{x}$, $\tilde{\boldsymbol{x}}$ and the inverse GFT are given by $\tilde{\boldsymbol{x}} = \mathbf{U}^\top \boldsymbol{x}$ and $\boldsymbol{x} = \mathbf{U}\tilde{\boldsymbol{x}}$ respectively where $[\tilde{\boldsymbol{x}}]_m$ is the frequency component that corresponds to the m-th eigenvalue λ_m.

Note that for complex symmetric GSO, unlike Hermitian GSOs, unfortunately also Parseval theorem is not valid i.e.

$$\|\tilde{\boldsymbol{x}}\|_2^2 = \left\|\mathbf{U}^\top \boldsymbol{x}\right\|_2^2 \neq \|\boldsymbol{x}\|_2^2 \tag{2.3.4}$$

Chapter 3

GRID-GSP: A GRAPH SIGNAL PROCESSING FRAMEWORK FOR THE POWER GRID

Having described the relevant concepts in graph signal processing (GSP) in Chapter 2 and introducing power systems model as an example of data generated as outputs of low-pass graph filter in Subsection 2.2.4, in this chapter the Grid-GSP framework is introduced. Firstly, grid quantities and parameters of interest are described in Section 3.1. Secondly, GSO for the grid is defined and the low-pass graph-filter model that justifies voltage phasor measurements is introduced. Then, temporal dynamics of these measurements is also characterized. All of the above yields a GSP generative model for the voltage phasor measurements.

Section 3.2, lays the foundation for Grid-GSP mapping the physical laws to a spatio-temporal generative model for voltage signals. Through these lens, in Section 3.3, the chapter revisits algorithms and tools from GSP for PMU data pertaining sampling and reconstruction along with optimal placement of PMUs, interpolation of missing samples and network inference. The algorithms and methods are tested numerically in Section 5.5.

3.1 Measurements and Parameters of the Electric Grid

Consider a network of generators and non-generator/load buses that can be represented as a graph where vertex set is a union between set of generator, $\mathcal{N}_G$ and non-generator/ load buses, $\mathcal{N}_L$, $i \in \{\mathcal{N}_G \cup \mathcal{N}_L\} = \mathcal{N}$ and with undirected edge set $(i, j) \in \mathcal{E}$. Ohm's law for the network is described through a *branch admittance matrix*, $\boldsymbol{Y}$, is the complex-valued graph Laplacian matrix associated with the power grid [104]:

$$[\boldsymbol{Y}]_{i,j} = \begin{cases} \sum_{k\in\mathcal{N}_i} y_{i,k}, \ i=j \\ \\ -y_{i,j}, \ i\neq j \end{cases} \tag{3.1.1}$$

where $y_{i,j}$ is the admittance of the branch between nodes i and j if $(i,j)\in\mathcal{E}$. The branch admittance matrix $\boldsymbol{Y}$ is a complex symmetric matrix. Next, the nodes/buses are partitioned into generator and non-generators, so that:

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_{gg} & \boldsymbol{Y}_{g\ell} \\ \boldsymbol{Y}_{g\ell}^{\top} & \boldsymbol{Y}_{\ell\ell} \end{bmatrix}, \tag{3.1.2}$$

where $\boldsymbol{Y}_{gg}$ is the generators to generators network, $\boldsymbol{Y}_{g\ell}$ includes the portion connecting generators and loads and $\boldsymbol{Y}_{\ell\ell}$ corresponds to the section of the grid connecting the loads buses among themselves. To model thermal loss on the lines, one needs to introduce shunt elements at all generator buses by $\boldsymbol{y}_{sh}^{g}\in\mathbb{C}^{|\mathcal{N}_G|}$ and at all load buses by $\boldsymbol{y}_{sh}^{\ell}\in\mathbb{C}^{|\mathcal{N}_L|}$.

The state of the system, from which all other physical quantities of interest can be derived, are the voltage phasors at each of the buses. In the following it is assumed that a phasor measurement unit (PMU) [3] installed on node/bus $i\in\mathcal{N}$ provides a noisy measurement of voltage and current phasors at time t with voltage phasors denoted as $v_i(t)=|v_i(t)|e^{j\theta_i(t)}$. With some abuse of notation, PMU data is referred to as $v_i(t)$ as well. Let the vector of voltage phasors at time t be $\boldsymbol{v}_t\in\mathbb{C}^{|\mathcal{N}|}$. Let $\boldsymbol{v}_t$ be partitioned into voltages at generator and non-generator buses, let $\boldsymbol{i}_t^{g}\in\mathbb{C}^{|\mathcal{N}_G|}$ be the generator current and $\boldsymbol{i}_t^{\ell}\in\mathbb{C}^{|\mathcal{N}_L|}$ the load current. Ohm's law for a network is:

$$\left(\boldsymbol{Y}+\texttt{Diag}\left(\begin{bmatrix}\boldsymbol{y}_{sh}^{g}\\ \boldsymbol{y}_{sh}^{\ell}\end{bmatrix}\right)\right)\boldsymbol{v}_t=\boldsymbol{i}_t, \quad \text{where} \ \ \boldsymbol{v}_t=\begin{bmatrix}\boldsymbol{v}_t^{g}\\ \boldsymbol{v}_t^{\ell}\end{bmatrix}, \boldsymbol{i}_t=\begin{bmatrix}\boldsymbol{i}_t^{g}\\ \boldsymbol{i}_t^{\ell}\end{bmatrix} \tag{3.1.3}$$

To describe the operating conditions of the system it is useful to introduce a few more quantities. In power systems transient dynamic analysis the impact of generating units is modeled as an internal bus characterized by a generator impedance (or admittance) $\boldsymbol{y}_g \in \mathbb{C}^{|\mathcal{N}_G|}$ for $g \in \mathcal{N}_G$ connected to an ideal voltage source called *internal voltage* [104]; and denote its value at time t by $E_i(t) = |E_i(t)|e^{\delta_i(t)}, i \in \mathcal{N}_G$ and the corresponding vector as $\boldsymbol{e}_t \in \mathbb{C}^{|\mathcal{N}_G|}$. It holds that:

$$\boldsymbol{i}_t^g = \mathtt{Diag}(\boldsymbol{y}_g)\,(\boldsymbol{e}_t - \boldsymbol{v}_t^g) \tag{3.1.4}$$

The generators respond to the grid electric load. To model the generators response the approximation used is that loads buses $\ell \in \mathcal{N}_L$ are slowly varying admittances [16]. Let this admittance be denoted as $\boldsymbol{y}_\ell(t) \in \mathbb{C}^N$. Its value is given by the ratio of conjugate of the apparent load power $s_i^*(t)$ and the voltage amplitude squared $|v_i(t)|^2$ at a particular load bus [104],

$$[\boldsymbol{y}_\ell(t)]_i = s_i^*(t)/|v_i(t)|^2, \quad i \in \mathcal{N}_L. \tag{3.1.5}$$

3.2 Graph Signal Processing for the Grid

In order to derive the GSP based generative model for voltage phasors, the relevant GSO is firstly defined as follows:

Definition 4. *The* **graph shift operator** *(GSO) is a complex symmetric matrix equal to a diagonal perturbation of the branch admittance matrix with generator admittance values,*

$$\mathbf{S} \triangleq \boldsymbol{Y} + \mathtt{Diag}\left(\begin{bmatrix} \boldsymbol{y}_g + \boldsymbol{y}_{sh}^g \\ \boldsymbol{y}_{sh}^\ell \end{bmatrix}\right) \tag{3.2.1}$$

With the GSO $\mathbf{S}$ is defined as in (3.2.1),

$$\boldsymbol{v}_t = \mathcal{H}(\mathbf{S}) \begin{bmatrix} \mathtt{Diag}(\boldsymbol{y}_g)\boldsymbol{e}_t \\ \boldsymbol{i}_t^\ell \end{bmatrix} + \boldsymbol{\eta}_t \tag{3.2.2}$$

where $\boldsymbol{\eta}_t$ the measurement noise and $\mathcal{H}(\mathbf{S}) = \mathbf{S}^{-1}$. The graph filter $\mathcal{H}(\mathbf{S})$ can be written as a matrix polynomial using Cayley-Hamilton theorem and is therefore shift-invariant. It is a low-pass graph filter due to the inversion of GSO. From the definition of the GSO it follows that

Definition 5. *The grid* **Graph Fourier Transform** *(GFT) basis for voltage phasors is the orthogonal matrix* $\mathbf{U}$ *given by the eigenvalue decomposition of the GSO in Definition 4:*

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top, \quad |\lambda_{min}| > 0 \tag{3.2.3}$$

Here, the GSO $\mathbf{S}$ is a *complex-symmetric* matrix that has the same support as the electric-grid graph Laplacian as $\boldsymbol{Y}$ with the diagonal addition of generator admittances. Note that unlike the graph Laplacian, this GSO is *invertible*, $|\lambda_{\min}| > 0$. However, due to scale difference between additive elements in the diagonal and in the system admittance matrix $\boldsymbol{Y}$, the structure of generic power grids to be organized as *communities* and the system admittance matrix $\boldsymbol{Y}$ being sparse [105], the GSO $\mathbf{S}$ has a high condition number.

To explicitly visualize the underlying subspace in which voltage graph signal lies, consider $\boldsymbol{\Lambda}_{\mathcal{K}}$ be the diagonal matrix with entries $\lambda_i,\ i \in \mathcal{K} = \{1, \dots, k\}$. Define a low-pass filter with k frequency components and consequently the voltage phasor measurements as

$$\begin{aligned} &\mathcal{H}_k(\mathbf{S}) \triangleq \mathbf{U}\ \mathtt{Diag}\left(\tilde{\boldsymbol{h}}_k\right)\mathbf{U}^\top, \qquad \left[\tilde{\boldsymbol{h}}_k\right]_i = \begin{cases} \lambda_i^{-1}, & i \in \mathcal{K} \\ 0, & \text{else} \end{cases} \\ &\boldsymbol{v}_t \approx \mathcal{H}_k(\mathbf{S}) \begin{bmatrix} \mathtt{Diag}(\boldsymbol{y}_g)\boldsymbol{e}_t \\ \boldsymbol{i}_t^\ell \end{bmatrix} + \boldsymbol{\eta}_t, \end{aligned} \tag{3.2.4}$$

where $\mathcal{H}_k(\mathbf{S})$ will therefore represent the principal subspace of the voltage phasors whose dimensionality is the number of graph-frequencies $|\mathcal{K}|$. Therefore (3.2.4) defines the low-dimensional generative model for quasi-steady state voltage phasor measurements. The approximation error $\boldsymbol{\eta}_t$ captures modeling approximations.

While (3.2.2) is sufficient to establish the Grid-GSP model when loads are slowly varying the following proposition establishes a model that can be even more effective to capture the low-dimensional structure of the voltage phasors vector $\boldsymbol{v}_t$.

Proposition 1 (Grid-GSP generative model with low rank-excitation)**.** *In quasi-steady state, the voltage phasor vector* $\boldsymbol{v}_t$ *is approximately the output of a low-pass graph filter* $\mathcal{H}(\mathbf{S}_{ci}) \triangleq \mathbf{S}_{ci}^{-1}$ *with input,* $\mathtt{Diag}(\boldsymbol{y}_g)\boldsymbol{e}_t$*,*

$$\boldsymbol{v}_t = \mathcal{H}(\mathbf{S}_{ci}) \begin{bmatrix} \mathtt{Diag}(\boldsymbol{y}_g)\boldsymbol{e}_t \\ \mathbf{0} \end{bmatrix} + \boldsymbol{\eta}_t, \tag{3.2.5}$$

where the GSO $\mathbf{S}_{ci}$ *is defined as*

$$\mathbf{S}_{ci} = \mathbf{S} + \mathtt{Diag}\left(\begin{bmatrix} \mathbf{0} \\ \boldsymbol{y}_\ell \end{bmatrix}\right) \tag{3.2.6}$$

Proof. After partitioning the system into generator and non-generator buses, one can write [16]

$$\begin{bmatrix} \boldsymbol{v}_t^g \\ \boldsymbol{v}_t^\ell \end{bmatrix} \approx \left(\boldsymbol{Y} + \mathtt{Diag}\left(\begin{bmatrix} \boldsymbol{y}_{sh}^g + \boldsymbol{y}_g \\ \boldsymbol{y}_{sh}^\ell + \boldsymbol{y}_\ell \end{bmatrix}\right)\right)^{-1} \begin{bmatrix} \mathtt{Diag}(\boldsymbol{y}_g)\boldsymbol{e}_t \\ \mathbf{0} \end{bmatrix} + \boldsymbol{\eta}_t, \tag{3.2.7}$$

The result is a straightforward consequence of using the constant impedance model in (3.1.5) (with $\boldsymbol{y}_\ell(t) \approx \boldsymbol{y}_\ell$) to simplify Ohm's laws in (3.1.3). □

Proposition 1 implies that in quasi-steady state, the voltage phasor measurement $\boldsymbol{v}_t$, is approximately the output of a low-pass graph filter, $\mathcal{H}(\mathbf{S}_{\mathrm{ci}})$ with a low rank excitation. However, to fully leverage this approximation one would have to revise the definition of GSO and replace $\mathbf{S}$ with $\mathbf{S}_{\mathrm{ci}}$, to absorb the load as an impedance.

To provide insights on the temporal dynamics of the voltage phasors, it is necessary to capture the structure of the excitation term. As a matter of fact, $\boldsymbol{e}_t$ and $\boldsymbol{i}_t^{\ell}$, have different dynamics, as discussed in the subsequent subsections.

3.2.1 A GSP Model for Generator Dynamics

The excitation term corresponding to generator currents has elements as $[\boldsymbol{e}_t]_i$ coming from each generator $i \in G$. A non-linear dynamical model for the generators internal voltages, namely $\boldsymbol{e}_t \in \mathbb{C}^{\mathcal{N}_G}$ is illustrated. The model is inspired by the classical swing equations [106, 107] describing the coupled dynamics of the generators phase angles, $\delta_i(t),\ i \in G$ and the resulting variation in frequency, $\omega_i(t) \triangleq \dot{\delta}_i - \omega_0$ where $\omega_0 = 2\pi f_0$ with f_0 being the grid frequency (50 or 60 Hz).

The model, relies on two steps. First, the dynamics of a signal obtained through the following non-linear transformation of the internal generator voltages is modeled:

$$\boldsymbol{x}_t = (\texttt{Diag}(\boldsymbol{m}))^{\frac{1}{2}} \ln(\boldsymbol{e}_t) \quad \rightarrow \quad \boldsymbol{\delta}_t = \Im\{\ln(\boldsymbol{e}_t)\} \tag{3.2.8}$$

where the vector $\boldsymbol{m}$ entries are the so-called generators masses and $\boldsymbol{\delta}_t$ are the generators angles that appear in the swing equations. Second, like in the swing equations, Kron-reduction [106, 107] of the network in which generators are all adjacent, is utilized to describe the generators interactions. Let the Schur complement of block $\boldsymbol{B}$ of matrix $\boldsymbol{A}$ be denoted by $\mathfrak{Sh}(\boldsymbol{A}, \boldsymbol{B})$ To define this generator network and the corresponding GSO, first compute the Schur complement of block $\mathbf{S}_{\text{ci}}$ of the matrix $\boldsymbol{Y}_{\text{all}}$, defined as:

$$\boldsymbol{Y}_{\text{all}} = \begin{bmatrix} \texttt{Diag}\left(\boldsymbol{y}_g + \boldsymbol{y}_{sh}^g\right) & -\begin{bmatrix}\texttt{Diag}\left(\boldsymbol{y}_g + \boldsymbol{y}_{sh}^g\right) & \mathbf{0}\end{bmatrix} \\ -\begin{bmatrix}\texttt{Diag}\left(\boldsymbol{y}_g + \boldsymbol{y}_{sh}^g\right) & \mathbf{0}\end{bmatrix}^{\top} & \mathbf{S}_{\text{ci}} \end{bmatrix}$$

The Schur complement of the $\mathbf{S}_{\text{ci}}$ in $\boldsymbol{Y}_{\text{all}}$ has two contributions:

$$\mathfrak{Sh}(\boldsymbol{Y}_{\text{all}}, \mathbf{S}_{\text{ci}}) = j\boldsymbol{Y}_{\text{red}} + \boldsymbol{E}_{\text{red}} \tag{3.2.9}$$

where $\boldsymbol{E}_{\text{red}}$ is a real diagonal dominated matrix, and the imaginary part $\boldsymbol{Y}_{\text{red}}$ has the structure of a graph Laplacian. The proposed dynamical model for the graph signal $\boldsymbol{x}_t$ relies on the following definition for the GSO of the Kron-reduced generator-only graph:

Definition 6. *A GSO is defined for the Kron-reduced generator only network as*

$$\mathbf{S}_{red} = (\mathtt{Diag}(\boldsymbol{m}))^{-\frac{1}{2}}\boldsymbol{Y}_{red}(\mathtt{Diag}(\boldsymbol{m}))^{-\frac{1}{2}} \in \mathbb{R}^{|\mathcal{N}_G|} \tag{3.2.10}$$

with the following eigenvalue decomposition,

$$\mathbf{S}_{red} = \mathbf{U}_{red}\boldsymbol{\Lambda}_{red}\mathbf{U}_{red}^{\top} \tag{3.2.11}$$

and the orthonormal GFT basis being $\mathbf{U}_{red}$*.*

The GSP dynamics for the generator internal voltages $\boldsymbol{e}_t$ are as follows,

$$\boldsymbol{e}_t = \exp\left((\mathtt{Diag}(\boldsymbol{m}))^{-\frac{1}{2}}\boldsymbol{x}_t\right) \tag{3.2.12}$$

where $\boldsymbol{x}_t$ is a GF-AR (2) process, i.e. the z-GFT $\tilde{\boldsymbol{X}}(z)$ satisfies the following equation:

$$\mathtt{Diag}\left(\tilde{\boldsymbol{a}}(z)\right)\tilde{\boldsymbol{X}}(z) = \tilde{\boldsymbol{W}}(z) \tag{3.2.13}$$

$$\tilde{\boldsymbol{a}}(z) = 1 - \tilde{\boldsymbol{a}}_1 z^{-1} - \tilde{\boldsymbol{a}}_2 z^{-2} \tag{3.2.14}$$

To the best of our knowledge, the dynamics of the amplitudes of the generators internal voltages are typically ignored, but the swing equation for the generators angles are a key tool for power systems dynamical analysis:

$$\mathtt{Diag}\left(\boldsymbol{m}\right)\ddot{\boldsymbol{\delta}} + \mathtt{Diag}\left(\boldsymbol{d}\right)\dot{\boldsymbol{\delta}} = \overline{\boldsymbol{w}} - \mathbf{Y}_{\text{red}}\boldsymbol{\delta}, \tag{3.2.15}$$

where $\boldsymbol{m}$ are the generators masses, introduced previously, $\boldsymbol{d}$ are the damping coefficients of generators that are often neglected and $\overline{\boldsymbol{w}} - \boldsymbol{Y}_{\text{red}}\boldsymbol{\delta}$ is the imbalance between the electrical and mechanical power that triggers the change in generator angular velocity and acceleration. Note that $\Im\{\boldsymbol{x}\} = (\mathtt{Diag}\left(\boldsymbol{m}\right))^{\frac{1}{2}}\boldsymbol{\delta}$. By manipulating (3.2.15), the following is proved:

Proposition 2. *Let* $\boldsymbol{w}$ *be such that* $\Im\{\boldsymbol{w}\} = (\mathtt{Diag}(\boldsymbol{m}))^{-\frac{1}{2}}\overline{\boldsymbol{w}}$. *Using the approximation:*

$$\mathtt{Diag}\left(\boldsymbol{d}\right)\mathtt{Diag}\left(\boldsymbol{m}\right)^{-1} \approx \chi\mathbb{I},$$

the dynamics of $\Im\{\boldsymbol{x}_t\}$ *can be justified with an GF-AR (2) model with GSO* $\mathbf{S}_{red}$*:*

$$\begin{aligned} &\Im\{\boldsymbol{x}_t\} - \mathcal{A}_1(\mathbf{S}_{red})\Im\{\boldsymbol{x}_{t-1}\} - \mathcal{A}_2(\mathbf{S}_{red})\Im\{\boldsymbol{x}_{t-2}\} \approx \Im\{\boldsymbol{w}\}, \\ &\mathcal{A}_1(\mathbf{S}_{red}) := (2-\chi)\mathbb{I} - \mathbf{S}_{red}, \quad \mathcal{A}_2(\mathbf{S}_{red}) := (\chi-1)\mathbb{I} \end{aligned} \tag{3.2.16}$$

Proof. Simple algebra on (3.2.15) allows to recast the equations in the following form:

$$\Im\{\ddot{\boldsymbol{x}}\} + \chi\Im\{\dot{\boldsymbol{x}}\} = \mathtt{Diag}\left(\boldsymbol{m}\right)^{-\frac{1}{2}}\overline{\boldsymbol{w}} - \mathbf{S}_{\text{red}}\Im\{\boldsymbol{x}\}. \tag{3.2.17}$$

Assuming that the sampling rate is fast enough, and normalizing it to 1, the finite difference approximations for the derivatives are $\dot{\boldsymbol{x}} \approx \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$, $\ddot{\boldsymbol{x}} \approx \boldsymbol{x}_{t+1} - 2\boldsymbol{x}_t + \boldsymbol{x}_{t-1}$ and can be used to obtain GF-AR (2) equations for the samples $\Im\{\boldsymbol{x}_t\}$ in (3.2.16). □

The suggested model is simply extending the GF-AR (2) model to capture both the real and imaginary part of $\boldsymbol{x}_t$ and suggesting to search the $2|\mathcal{N}_{\mathcal{G}}|$ parameters to fit the model with $\tilde{\boldsymbol{a}}_1, \tilde{\boldsymbol{a}}_2$ rather than exploring a general MIMO filter response. For simplicity of representation, the dynamical equation for $\boldsymbol{x}_t$ is written in the GF domain,

$$\tilde{\boldsymbol{x}}_t = \mathtt{Diag}\left(\tilde{\boldsymbol{a}}_1\right)\tilde{\boldsymbol{x}}_{t-1} + \mathtt{Diag}\left(\tilde{\boldsymbol{a}}_2\right)\tilde{\boldsymbol{x}}_{t-2} + \tilde{\boldsymbol{w}}_t \tag{3.2.18}$$

such that the impulse response of the filter at graph-frequency $\lambda_{\text{red},i}$ is defined by $\left[\tilde{\boldsymbol{a}}_1\right]_i, \left[\tilde{\boldsymbol{a}}_2\right]_i$. Note that $\left[\tilde{\boldsymbol{a}}_1\right]_i, \left[\tilde{\boldsymbol{a}}_2\right]_i$ can be written as polynomials in graph frequency $\lambda_{\text{red},i}$,

$$\left[\tilde{\boldsymbol{a}}_1\right]_i = \textstyle\sum_{k=0}^{K_1-1} a_{k,1}\lambda_{\text{red},i}^k, \quad \left[\tilde{\boldsymbol{a}}_2\right]_i = \textstyle\sum_{k=0}^{K_2-1} a_{k,2}\lambda_{\text{red},i}^k, \tag{3.2.19}$$

which characterizes the poles of the AR system using graph frequencies. From (3.2.16), it is seen that a first order polynomial may suffice to characterize $\left[\tilde{\boldsymbol{a}}_1\right]_i, \left[\tilde{\boldsymbol{a}}_2\right]_i$ in most cases.

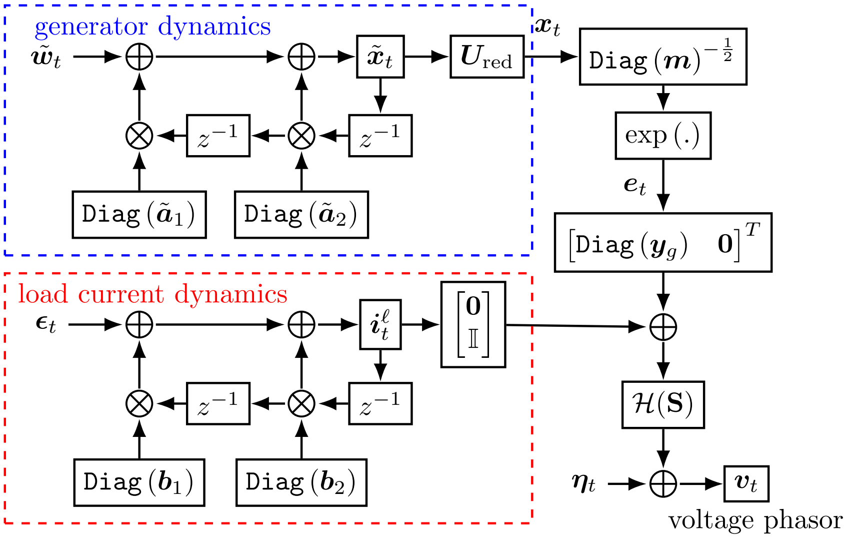


Figure 3.2.1: Block diagram showing generative model for voltage phasor measurements.

3.2.2 Load Dynamics

There are several papers in the literature that deal with load forecasting and modeling [108]. A simple autoregressive (AR)-2 model per node or load busis adopted to describe the dynamics of the load,

$$\boldsymbol{i}_t^\ell = \texttt{Diag}(\boldsymbol{b}_1)\, \boldsymbol{i}_{t-1}^\ell + \texttt{Diag}(\boldsymbol{b}_2)\, \boldsymbol{i}_{t-2}^\ell + \boldsymbol{\epsilon}_t \tag{3.2.20}$$

where parameters $\boldsymbol{b}_1, \boldsymbol{b}_2$ are estimated load data time series. The block diagram in Fig.3.2.1 summarizes the modeling efforts.

The unique nature of voltage phasor measurements allows to describe a similar model for any subset of measurements on a graph. This is discussed next.

3.2.3 Low-pass Property of Down-sampled Voltage Graph Signal

Let $\boldsymbol{v}_{\mathcal{M}}$ (time index t is ignored for simplicity) be the down-sampled voltage graph signal where $\mathcal{M} \subset \mathcal{N}$ is the set of node indices at which measurements are available. An interesting self-similarity/fractional property is observed for voltage graph signals which is summarized in the following proposition

Proposition 3. *Let* $\boldsymbol{v}_{\mathcal{M}}$ *be the graph signal down-sampled in the vertex-domain with* $|\mathcal{M}|$ *samples. Then, with the GSO defined with respect to the reduced-graph of* $\mathcal{M}$ *vertices as* $\mathbf{S}_{red,\mathcal{M}}$*, graph signal* $\boldsymbol{v}_{\mathcal{M}}$ *is the output of low-pass graph filter* $\mathcal{H}\left(\mathbf{S}_{red,\mathcal{M}}\right) \triangleq \mathbf{S}_{red,\mathcal{M}}^{-1}$

$$\boldsymbol{v}_{\mathcal{M}} = \mathcal{H}\left(\mathbf{S}_{red,\mathcal{M}}\right)\boldsymbol{\varphi} \tag{3.2.21}$$

where the GSO for the reduced-graph is given by Kron-reduction of $\mathbf{S}$*,* $\mathbf{S}_{red,\mathcal{M}} = \mathfrak{S}\mathfrak{h}\left(\mathbf{S}, \mathbf{S}_{\mathcal{M}^c\mathcal{M}^c}\right)$

Proof. The GSO can be written in a block form,

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{\mathcal{M}\mathcal{M}} & \mathbf{S}_{\mathcal{M}\mathcal{M}^c} \\ \mathbf{S}_{\mathcal{M}\mathcal{M}^c}^{\top} & \mathbf{S}_{\mathcal{M}^c\mathcal{M}^c} \end{bmatrix} \tag{3.2.22}$$

The Kron reduction gives the Schur complement of the block $\mathbf{S}_{\mathcal{M}^c\mathcal{M}^c}$ in the GSO $\mathbf{S}$ i.e.,

$$\mathbf{S}_{\text{red},\mathcal{M}} = \mathfrak{S}\mathfrak{h}\left(\mathbf{S}, \mathbf{S}_{\mathcal{M}^c\mathcal{M}^c}\right) = \mathbf{S}_{\mathcal{M}\mathcal{M}} - \mathbf{S}_{\mathcal{M}\mathcal{M}^c}\mathbf{S}_{\mathcal{M}^c\mathcal{M}^c}^{-1}\mathbf{S}_{\mathcal{M}\mathcal{M}^c}^{\top}$$

From Kron reduction [107], one can write:

$$\boldsymbol{v}_{\mathcal{M}} = \mathbf{S}_{\text{red},\mathcal{M}}^{-1}\overbrace{\left(\boldsymbol{i}_{\mathcal{M}} - \mathbf{S}_{\mathcal{M}\mathcal{M}^c}\mathbf{S}_{\mathcal{M}^c\mathcal{M}^c}^{-1}\boldsymbol{i}_{\mathcal{M}^c}\right)}^{\boldsymbol{\varphi}} \tag{3.2.23}$$

□

Note from (3.2.23) that the down-sampled version $\boldsymbol{v}_{\mathcal{M}}$ is still a *low-pass graph signal* with the GSO being the Kron-reduced matrix $\mathfrak{S}\mathfrak{h}\left(\mathbf{S}, \mathbf{S}_{\mathcal{M}^c\mathcal{M}^c}\right)$. This preservation of the low-pass property suggests that one can utilize all the methods for low-pass graph signals also for thw down-sampled versions of the voltage graph signal as well. This property has been illustrated empirically in several papers that highlight low-dimensional properties even with measurements from a subset of buses. Although the reduced-graph is denser compared to the original graph, it still helps to infer faults or events that occurred in a subset of nodes where sensors are not installed as long as correct placement strategies are devised i.e. that of choosing the subset $\mathcal{M}$. Work in [28] explored the optimal placement for fault localization in the under-sampled regime and also made connections with GSP theory.

3.3 Revisiting Algorithms from GSP for PMU Data

In this section some of the implications Grid-GSP has while understanding sampling, optimal placement of measurement devices in power systems, interpolation of missing samples and network inference are studied. The underlying generative model responsible for low-rank nature of data that has been established in the previous section helps explaining the success that many past works, such as [11, 109, 110], have attained in recovering missing PMU data using matrix completion methods. The low-pass nature of the voltage graph signals discussed in Proposition 1 provides the theoretical underpinning that support the arguments made in the literature.

3.3.1 Sampling and Recovery of Grid-Graph Signals

From the approximation in (3.2.4) it is seen that voltage graph signals have graph frequency content that drops as λ_k grows. This characteristic renders the signal approximately band-limited in the GFT domain [111] which means that there is a cut-off frequency λ_k such that frequency content corresponding to λ_{k+1} and higher is negligible. Let the GFT basis corresponding to the first dominant k graph frequecies be $\mathbf{U}_{\mathcal{K}}$. The *bandlimiting operator* is, $\boldsymbol{\mathcal{B}}_{\mathcal{K}} = \mathbf{U}_{\mathcal{K}}\mathbf{U}_{\mathcal{K}}^{\top} \in \mathbb{C}^{|\mathcal{N}|\times\mathcal{K}}$ and the low frequency component of $\boldsymbol{v}_t$ is:

$$\boldsymbol{\mathcal{B}}_{\mathcal{K}}\boldsymbol{v}_t = \mathbf{U}_{\mathcal{K}}\mathbf{U}_{\mathcal{K}}^{\top}\boldsymbol{v}_t \tag{3.3.1}$$

Similarly, a vertex limiting operator (with $|\mathcal{M}|$) vertices is $\boldsymbol{\mathcal{D}}_{\mathcal{M}} = \boldsymbol{P}_{\mathcal{M}}\boldsymbol{P}_{\mathcal{M}}^{\top}$ where $\boldsymbol{P}_{\mathcal{M}}$ has columns that are coordinate vectors such that each column chooses a vertex/node. When the voltage measurements on the electrical network are from a few nodes, $i \in \mathcal{M}$ at time t, it can be written as $[\boldsymbol{v}_t]_{\mathcal{M}} = \boldsymbol{P}_{\mathcal{M}}^{\top}\boldsymbol{v}_t$. For reconstruction, results in [111] dictate the necessary condition be that $|\mathcal{M}| \geq |\mathcal{K}|$. In the presence of noise or modeling error relative to the perfect band-limited definition, the optimal sampling pattern (i.e. the best placement for PMUs on the grid) to minimize reconstruction error is closely tied to the grid topology and the model mismatch relative to a strictly band-limited graph signal. An optimal placement of PMUs according meeting this criterion maximizes the smallest singular value, $\sigma_{\min}\left(\boldsymbol{\mathcal{D}}_{\mathcal{M}}\mathbf{U}_{\mathcal{K}}\right)$, i.e. choose rows of $\mathbf{U}_{\mathcal{K}}$ such that they are as uncorrelated as possible and the resulting matrix has the highest condition number [111]. Consider then the spatial sampling mask $\boldsymbol{\mathcal{D}}_{\mathcal{M}} = \texttt{Diag}(\mathbf{1}_{\mathcal{M}})$ that selects M locations.

3.3.1.1 Sampling

The optimal placement of M PMUs maximizes $\sigma_{\min}(\boldsymbol{\mathcal{D}}_{\mathcal{M}}\mathbf{U}_{\mathcal{K}})$ which amounts to choosing the rows of $\mathbf{U}_{\mathcal{K}}$ with the smallest possible coherence (as close as possible to being orthogonal). In [111] and references therein, a greedy method is employed to find M rows from $\mathbf{U}_{\mathcal{K}}$ so that the least singular value is maximized. The method is described in Algorithm 3.1.

Power systems topologies exhibit naturally a *community structure* that is reflected in the system admittance matrix $\boldsymbol{Y}$ [35] due to population density or clusters of loads. Therefore, the eigenvectors of $\boldsymbol{S}$ follow suit and reflect the community structure of the graph. Thus, choosing rows of $\mathbf{U}_{\mathcal{K}}$ to be uncorrelated is intuitively putting PMUs in different *graph-clusters* or communities. This fact was also discussed in [28] in the context of sensor placements for fault localization. The PMUs sampling rate

Algorithm 3.1: Optimal Placement Algorithm

```
Input: $M$, $\mathbf{U}$, $\mathcal{M} = \emptyset$ while $m < M$ do
    $i^* = \arg\max_{i \in \mathcal{M}^c} \sigma_{min}(\boldsymbol{\mathcal{D}}_{\{\mathcal{M},i\}} \mathbf{U}_{\mathcal{K}})$ ;
    $\mathcal{M} \leftarrow \mathcal{M} + \{i^*\}$ ;
    $m \leftarrow m + 1$ ;
Output: $\mathcal{M}$
```

in time exceeds the needs for reconstructions in a quasi-steady state conditions by a significant margin and it is designed to help detect sharp transients in the system.

3.3.1.2 Reconstruction

Voltage data samples are obtained down-sampling in space after the optimal placement of PMUs and also uniformly down-sampling in time. At time t when $|\mathcal{M}|$ samples, $[\boldsymbol{v}_t]_{\mathcal{M}}$ are available, the following model applies

$$[\boldsymbol{v}_t]_{\mathcal{M}} \approx \boldsymbol{P}_{\mathcal{M}}^{\top} \mathbf{U}_{\mathcal{K}} \tilde{\boldsymbol{v}}_t \tag{3.3.2}$$

where $\tilde{\boldsymbol{v}}_t$ is the GFT of graph signal $\boldsymbol{v}_t$ Therefore, reconstruction in spatial domain is done via GFT basis as

$$\hat{\boldsymbol{v}}_t = \mathbf{U}_{\mathcal{K}} \left(\boldsymbol{P}_{\mathcal{M}}^{\top} \mathbf{U}_{\mathcal{K}} \right)^{\dagger} [\boldsymbol{v}_t]_{\mathcal{M}} \tag{3.3.3}$$

Reconstruction in temporal domain can be done independently by *up-sampling*, i.e. via the windowed inverse Fourier transform of the up-sampled signal created from uniformly time-decimated data.

3.3.2 Interpolation of Missing Samples

When voltage measurements are missing or corrupted, denoising and interpolation of such data can be cast as a graph signal recovery problem by regularizing the total variation, (TV). Overall, the problem resembles time-vertex graph signal recovery [112]. Let the voltage phasor measurements matrix $\mathbf{V}$ be represented as

$$\mathbf{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_T \end{bmatrix}. \tag{3.3.4}$$

Let $\mathbb{P}_\Omega\left(\hat{\mathbf{V}}\right)$ be the set of available measurements that have samples in entries of set Ω and are noisy,

$$\begin{aligned} \min_{\mathbf{V}} \quad & \left\|\mathbb{P}_\Omega\left(\hat{\mathbf{V}}-\mathbf{V}\right)\right\|_F^2 + c_g \textstyle\sum_{t=1}^{T} \|\mathbf{S}\boldsymbol{v}_t\|_1 \\ & + c_t \textstyle\sum_{t=2}^{T} \|\boldsymbol{v}_t - \boldsymbol{v}_{t-1}\|_2^2 \end{aligned} \tag{3.3.5}$$

where the two regularizing terms measure the variation in the graph and time domain and c_g, c_t are the corresponding regularization constants. Importantly, one can use the GSO of the reduced graph, $\mathbf{S}_{\mathrm{red},\mathcal{M}}$ if only a subset of measurements on the grid, $\mathcal{M}$, are available and employ the same formulation as in (3.3.5) for interpolation of missing samples.

3.3.3 Network inference as Graph Laplacian Learning

Estimating $\mathbf{S}$ from voltage phasor can be accomplished along similar lines, by seeking the GSO that minimizes the total variation of the observed voltage phasors' signals, while imposing additional constraints on the GSO (complex-symmetry, dominant diagonal values and sparse off-diagonal entries). Also, $\mathbf{S}$ tends to have larger imaginary values than real especially on the diagonal. Therefore the problem becomes

$$\min_{\mathbf{S}} \quad \textstyle\sum_{t=1}^{T} \|\mathbf{S}\boldsymbol{v}_t\|_1 + \gamma \left\|\mathbf{S} - \mathtt{Diag}\left(\mathtt{Diag}\left(\mathbf{S}\right)\right)\right\|_F^2 + \sum_{t=1}^{T} \|\mathbf{S}\boldsymbol{v}_t - \boldsymbol{i}_t\|_2^2 \tag{3.3.6}$$

$$\text{subject to } \Re[\mathbf{S}]_{i,j} = \Re[\mathbf{S}]_{j,i}, \Im[\mathbf{S}]_{i,j} = \Im[\mathbf{S}]_{j,i}, i \neq j, \tag{3.3.7}$$

$$\Re\mathrm{Tr}\left(\mathbf{S}\right) = \alpha|\mathcal{N}|, \Im\mathrm{Tr}\left(\mathbf{S}\right) = \beta|\mathcal{N}| \tag{3.3.8}$$

Here $\alpha, \beta > 1$ control the amplitude of real and imaginary values on the diagonal. As before, the problem above can be recast with down-sampled voltage graph sig-

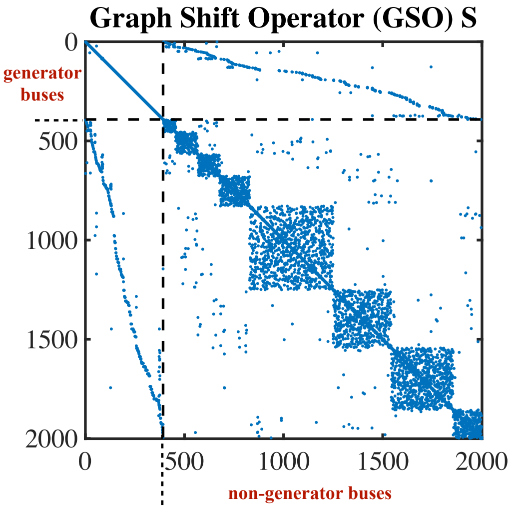


Figure 3.4.1: Support of the GSO $\mathbf{S}$ of the Network.

nals to infer the Kron-reduced GSO $\mathbf{S}_{\text{red},\mathcal{M}}$ with the approximation that the term $\mathbf{S}_{\mathcal{M}\mathcal{M}^c}\mathbf{S}^{-1}_{\mathcal{M}^c\mathcal{M}^c}\boldsymbol{i}_{\mathcal{M}^c}$ in (3.2.23) is treated as additive Gaussian noise.

3.4 Numerical Results

The numerical results in this section are mostly obtained using data from the synthetic ACTIVSg2000 case [113], a realistic model emulating the ERCOT system, which includes 2000 buses-with 432 generators and the rest non-generator buses. The ACTIVSg2000 case data include a realistic PMU data time series, in which 392 generators are dispatched to meet variable load demand. The sampling rate, as for real PMUs, is 30 samples per second. As all the system related parameters are known, it is easier to verify the proposed modeling strategy through the ACTIVSg2000 PMU data set. Fig. 3.4.1 shows the support of the GSO $\boldsymbol{S}$ or the graph Laplacian $\boldsymbol{Y}$ matrix when ordered into generator and non-generator buses. The block-diagonal structure is notable, and is the result of the population distribution in the state of Texas, which is concentrated in 8 metropolitan areas.

Grid-GSP model: In Fig. 3.4.2, magnitude of GFT of voltage graph signal $\boldsymbol{v}_t$ and

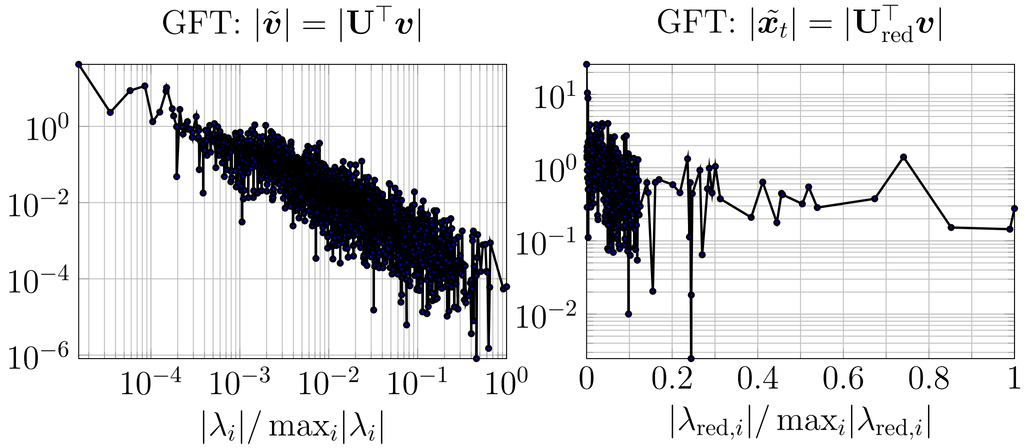


Figure 3.4.2: Magnitude of Graph Fourier Transform (GFT) for voltage graph signal, $|\tilde{\boldsymbol{v}}_t|$ (left) and input to generator-only network $|\tilde{\boldsymbol{x}}_t|$ (right) plotted with respect to normalized graph frequency.

the input $\boldsymbol{x}_t$ are plotted for a single time instant with respect to their corresponding normalized graph frequencies $|\lambda_i|/\max_i|\lambda_i|$ and shown in log-scale. From the linear decay, it is evident that the magnitude of GFT coefficients $|\tilde{\boldsymbol{v}}_t|$ corresponding to lower frequencies are more significant as compared to higher frequencies.

Similarly, the GFT of the exponent in the input, $\tilde{\boldsymbol{x}}_t = \mathbf{U}_{\text{red}}^\top \boldsymbol{x}_t$ with the generator GSO $\mathbf{S}_{\text{red}}$, is plotted with respect to the graph frequencies in Fig. 3.4.2. The decay in GFT coefficients with respect to frequency is less pronounced confirming that graph signal $\boldsymbol{x}_t$ is not necessarily low-pass and in general depends on the topology of the generator only network.

In Fig. 3.4.3, magnitude of GFT of the down-sampled voltage graph signal, $\tilde{\boldsymbol{v}}_{\mathcal{M}} = \mathbf{U}_{\text{red},\mathcal{M}}^\top \boldsymbol{v}_{\mathcal{M}}$ for $|\mathcal{M}| = 867$ and $|\mathcal{M}| = 261$ with two different down-sampling strategies: with PMUs placed at buses in few communities within the GSO $\mathbf{S}$ and the other being optimal placement for graph signal reconstruction (see algorithm 3.1). The placement strategy has an effect on the low-pass nature of the down-sampled signal. The steeper attenuation of GFT magnitude with placement strategy being community-wise is a result of loss in spatial-resolution.

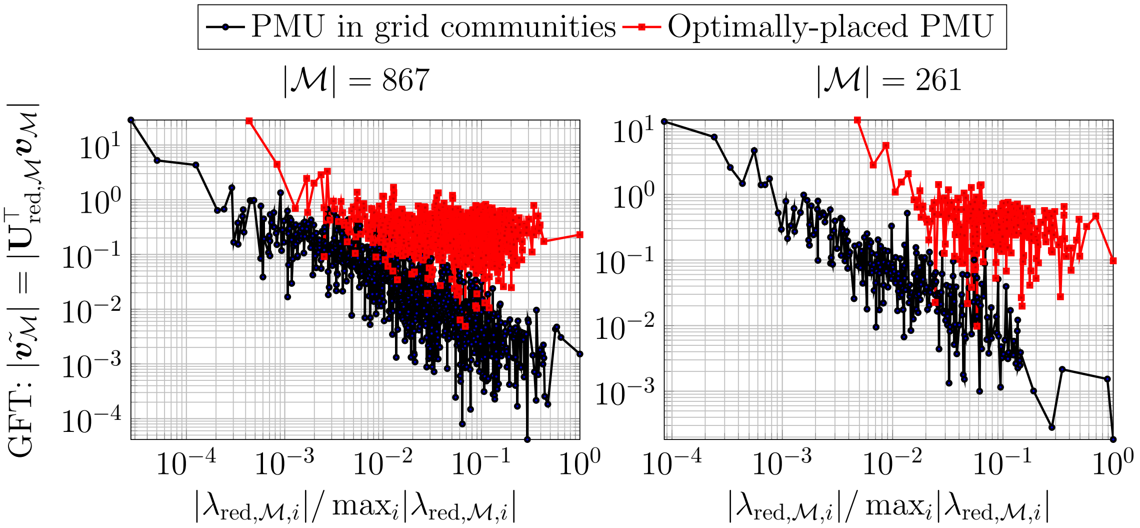


Figure 3.4.3: Magnitude of GFT for spatially down-sampled voltage graph signal, $|\tilde{\boldsymbol{v}}_{\mathcal{M}}|$ plotted with respect to normalized graph frequency.

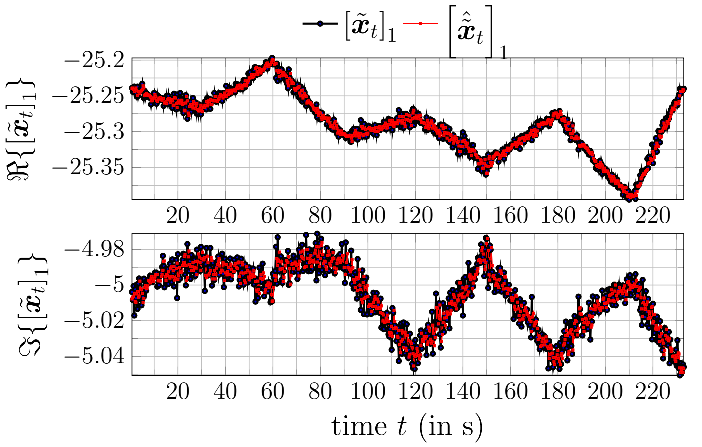


Figure 3.4.4: AR model fit to the GFT of $\boldsymbol{x}_t$. Component corresponding to smallest graph frequency, $\lambda_{\text{red},1}$, $[\tilde{\boldsymbol{x}}_t]_1$

To highlight the temporal variation in the GFT domain of input exponent, $\tilde{\boldsymbol{x}}_t$, a short time-series of the real and imaginary parts along with the fit of the AR model are shown in Fig. 3.4.4. As expected, the AR model fits well. Fig. 3.4.5 shows the similar AR-2 model fit to the load current at a bus that had the highest absolute value of load. To emphasize the temporal nature of the input, the 2-dimensional frequency response (in both graph and time domains) is plotted for the input $\tilde{\boldsymbol{x}}_t$ in Fig. 3.4.6. The figure provides evidence of the coupling between the graph frequencies and time

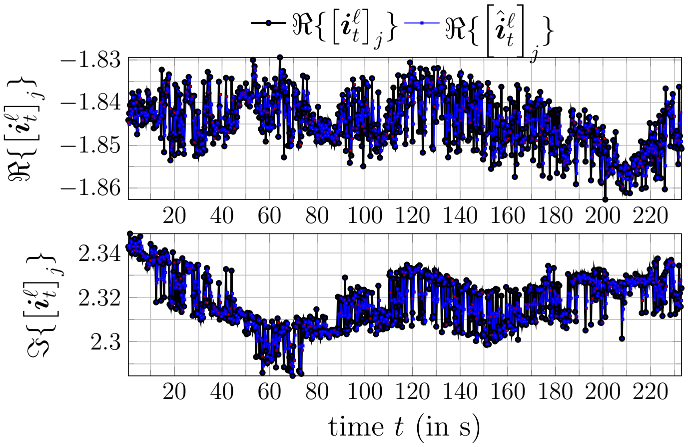


Figure 3.4.5: AR model for load bus, $j = 1312$, i.e. bus with highest absolute value of load.

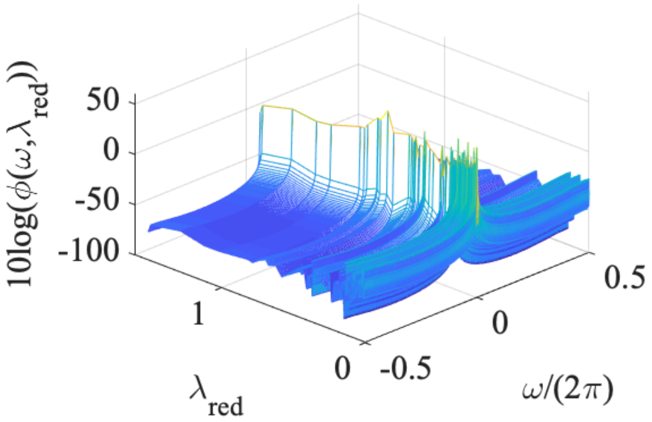


Figure 3.4.6: 2−D Frequency Response for $\tilde{\boldsymbol{x}}_t$

series Fourier power spectrum, and the variability of the temporal response depending on what GFT frequency mode is excited with Fourier spectra that are more or less concentrated towards low frequencies depending on the GFT mode. Hence, temporal dynamics can inform about what is happening in space (i.e. the trends are coupled).

Revising GSP Tools: Sampling and Optimal Placement Fig. 3.4.7 shows the placement of 100 PMUs super-imposed on the support of the ordered Grid-GSO, $\mathbf{S}$ when $|\mathcal{K}| = |\mathcal{M}| = 100$ graph frequency components are considered. Note the distribution of PMUs to different communities as well as on the generator buses as they belong to

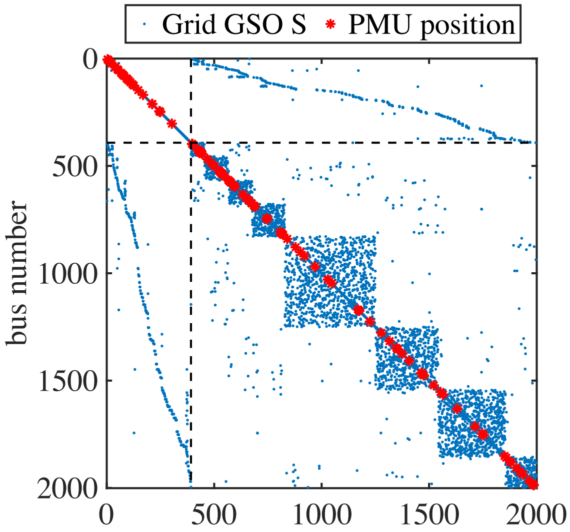


Figure 3.4.7: Optimal Placement of $|\mathcal{M}| = 100$ PMUs. $K = 100$, 2000 buses.

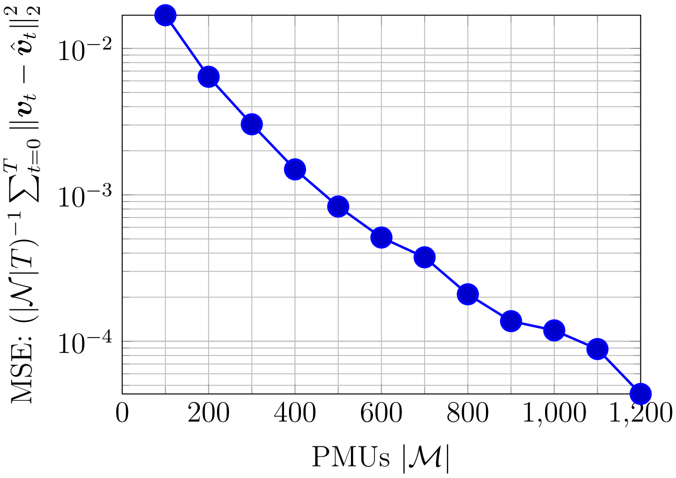
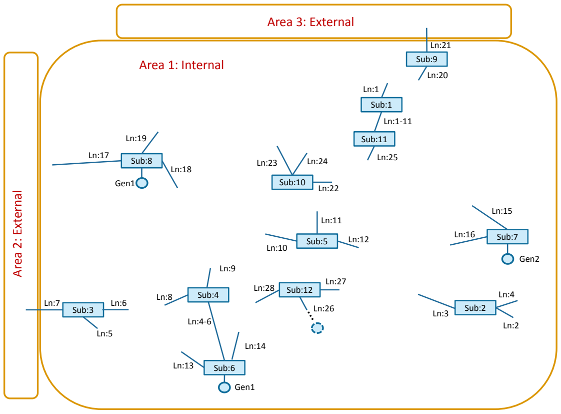
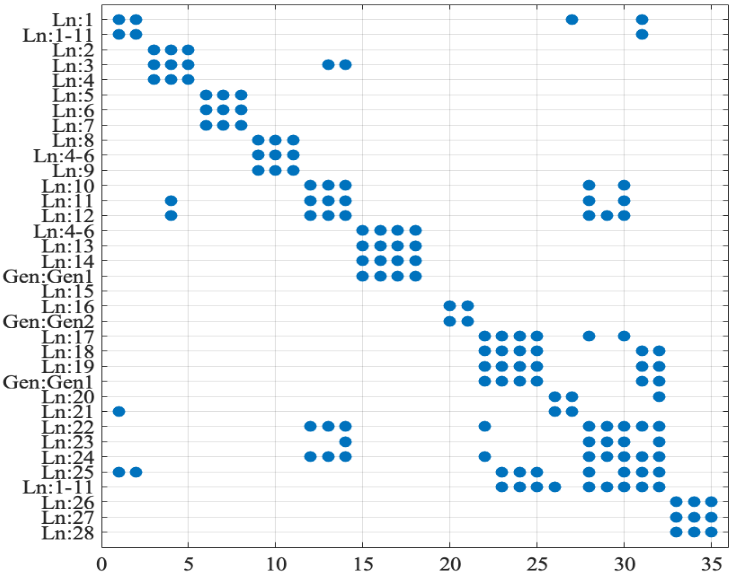


Figure 3.4.8: Reconstruction performance after optimal placement (algorithm 3.1) of $|\mathcal{M}|$ PMUs. Number of frequencies used: $|\mathcal{K}| = |\mathcal{M}|$. Total number of buses is 2000.



(a) ISO-NE Map



(b) Support of the Estimated GSO for the ISO-NE Test Case

Figure 3.4.9: (a) The map of PMUs placed in ISO-NE test case 3 [114], (b) the support of estimated GSO via (3.3.8). Note that the community structure exactly corresponds to groups of PMUs in the actual system.

different graph communities. Fig. 3.4.8 exhibits the performance of the GSP based reconstruction method on optimally placed $|\mathcal{M}|$ PMUs that provide down-sampled measurements, $\boldsymbol{v}_{\mathcal{M}}$. Even with just 5% of measurements (100 PMUs), the reconstruction error is extremely low.

To illustrate that the proposed modeling holds and algorithms work well also for real PMU data, in the next numerical experiments, a real-world dataset of measurements from 35 PMUs placed in ISO New-England grid (ISO-NE) [114] is used. The data corresponds to a period of 180 seconds when a large generator near Ln:2 and Ln:4 introduces oscillations in the system. PMU signals are decimated in time to have sampling frequency of 1 sample/s.

Network inference: As the underlying GSO is unknown, it is estimated via (3.3.8) with the goal of recovering the underlying reduced-GSO. Fig. 3.4.9 shows the support of the estimated GSO. Comparing it with the map of PMUs in Fig. 3.4.9a, it is seen that the block-diagonal nature of the recovered GSO captures the community structure highlighted in the map.

Interpolation of missing measurements: Once the GSO is estimated, consider the interpolation problem in (3.3.5) for the same ISO-NE dataset. Data is deleted at random and noise is added. The problem in (3.3.5) is solved to recover missing measurements. In Fig.3.4.10 voltage phasor angles $\boldsymbol{\theta}$ of the original, corrupted and recovered measurements are compared. Corrupted measurements have missing samples not just at random but also contiguous in time. The normalized MSE, $\left\|\mathbf{V}-\hat{\mathbf{V}}\right\|_F^2 / \|\mathbf{V}\|_F^2$ is the metric used to gauge the reconstruction performance. As a comparison, the AM-FIHT algorithm proposed in [115] is tested on the same data, which regularizes the reconstruction task assuming that the Hankel matrix formed with the columns of

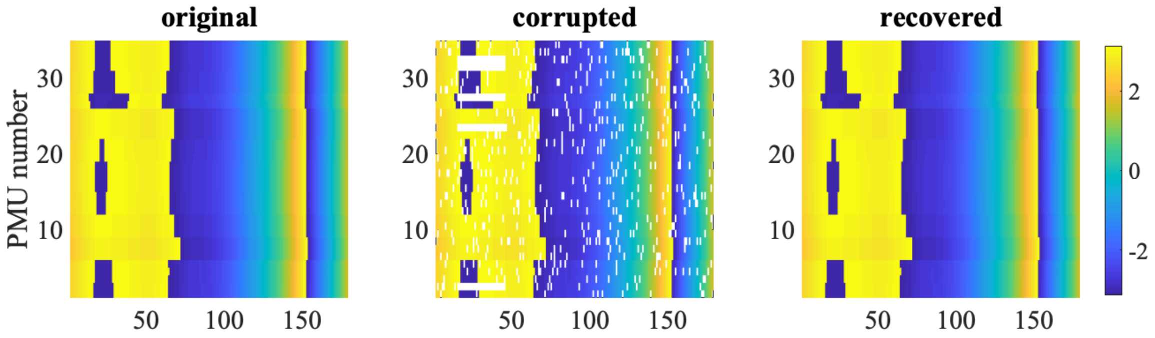


Figure 3.4.10: Interpolation of missing measurements for an ISO-NE case using GSO based regularization. Voltage phasor angles are shown. Note the contiguous missing of samples and our ability to interpolate. The relative noise level used is, $(|\mathcal{M}|T)\sigma^2/\left\|\mathbf{V}\right\|_F^2 = 10^{-4}$ Normalized MSE for this run is 6.22×10^{-4}.

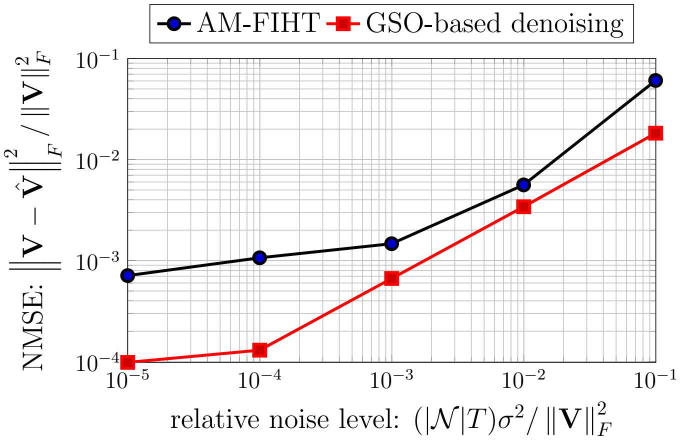


Figure 3.4.11: Comparison of AM-FIHT algorithm in [115] $(r = 10, n1 = 3, \beta = 0)$ and the proposed GSO based interpolation with 50% of missing measurements in the ISO-NE dataset.

$\mathbf{V}$, i.e. $\mathbb{H}(\mathbf{V})$, has low rank r,

$$\min_{\mathbf{V}} \left\|\mathbb{P}_\Omega\left(\hat{\mathbf{V}} - \mathbf{V}\right)\right\|_F^2 \quad \text{subject to} \quad \operatorname{rank}\left(\mathbb{H}(\mathbf{V})\right) = r \tag{3.4.1}$$

The plot comparing the two methods is shown in Fig. 3.4.11. As seen, the GSO based method outperforms the AM-FIHT for this dataset, indicating that the regularization using the GSO is more effective at capturing the low-rank nature of the data, compared to seeking an arbitrary low rank structure in the the Hankel matrix of the data.

Chapter 4

APPLICATIONS OF GRID-GSP

The goal of this chapter is to showcase the benefits of casting problems in the Grid-GSP framework introduced in Chapter 3 through four exemplary applications, namely blind community detection, fault localization in under-sampled regime, data compression and anomaly detection. In Section 4.1, a consequence of the low-pass nature of voltage measurements that is, the ability to identify communities in the electrical grid using blind community detection is studied. In Section 4.2, concepts pertaining to fault localization in under-sampled graph regime are explored. In Section 4.3, compression of PMU data by the use of Grid-GSP generative model for PMU data is discussed. In Section 4.4 the detection of false data injection attacks is described along with the method of isolating affected buses.

4.1 Identification of Community Structure in the Electrical Grid

In [38] it was established that the eigenvectors of the data covariance of graph signals can be used to unveil the community structure of the underlying graph, when signals are the output of a low-pass graph filter. However, the result does not hold if a GSO has edge weights with negative or complex-valued weights. In this case, the ratio-cut minimization formulation cannot rely on such Laplacian operators and the algorithm proposed based on spectral clustering returns meaningless results. But when the GSO has weights with a common phase, the technique of community detection makes sense. This is the case with nearly purely imaginary valued system admittance matrix. Hence, the voltage phasor measurements are good candidates to reveal the community structure of the power grid. The electric grid has an inherent community structure that is driven by the fact that population density is highly uneven, and its spatial distribution determines the graph density [116]. Assume that

Algorithm 4.1: Blind Community Detection for Grid Graph Signals

```
Input: Voltage phasors $\{\boldsymbol{v}_t\}_{t=1}^{T}$ and desired number of communities $K$.
    Compute sample covariance matrix $\boldsymbol{C}_v = T^{-1}\sum_{i=1}^{T}\boldsymbol{v}_t\boldsymbol{v}_t^H$.
    Find the $K$ eigenvectors corresponding to top $K$ eigenvalues. Denote them by $\hat{\mathbf{U}}_K \in \mathbb{C}^{|\mathcal{N}|\times K}$.
    Define a higher dimensional matrix $\hat{\mathbf{U}}_{2K} \in \mathbb{R}^{|\mathcal{N}|\times 2K}$ as $\hat{\mathbf{U}}_{2K} = \begin{bmatrix} \Re\{\hat{\mathbf{U}}_K\} & \Im\{\hat{\mathbf{U}}_K\} \end{bmatrix}$.
    Apply k-means clustering method that minimizes the objective
        $F = \min_{\mathcal{A}_1,\dots\mathcal{A}_K} \sum_{k=1}^{K}\sum_{i\in\mathcal{A}_k} \left\| \hat{\boldsymbol{u}}_i^{\text{row}} - \frac{1}{|\mathcal{A}_k|}\sum_{m\in\mathcal{A}_k}\hat{\boldsymbol{u}}_m^{\text{row}} \right\| \quad (4.1.1)$
    where $\hat{\boldsymbol{u}}_i^{\text{row}}$ is the $i^{\text{th}}$ row of $\hat{\mathbf{U}}_{2K}$.
Output: $K$ communities $\hat{\mathcal{A}}_1,\dots\hat{\mathcal{A}}_K$
```

there are K communities. The grid is also designed such that each community is served by a portfolio of generators, which means that there are more generators than communities, $K < |G|$. From (3.2.4), , PMU voltage phasor data $\boldsymbol{v}_t$ is the output of a low-pass graph filter $\mathcal{H}_k(\mathbf{S})$. Assuming that system admittance matrix is purely imaginary, the graph filter has the form $\boldsymbol{S}^{-1} = \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^{\top}$. The eigenvectors $\mathbf{U}$ are real valued and orthonormal and the connection between spectral clustering [117] and the blind community detection as in [38] is seamless. The community structure can be recovered using eigenvectors of [38], $\boldsymbol{C}_v = T^{-1}\sum_{i=1}^{T}\boldsymbol{v}_t\boldsymbol{v}_t^H$ where H is the complex conjugate operator. Each outer product yields

$$\boldsymbol{v}_t\boldsymbol{v}_t^H \approx \mathcal{H}(\mathbf{S}) \begin{bmatrix} \texttt{Diag}(\boldsymbol{y}_g)\boldsymbol{e}_t \\ \boldsymbol{i}_t^\ell \end{bmatrix} \begin{bmatrix} (\texttt{Diag}(\boldsymbol{y}_g)\boldsymbol{e}_t)^H & (\boldsymbol{i}_t^\ell)^H \end{bmatrix} (\mathcal{H}(\mathbf{S}))^H + \boldsymbol{C}_\eta \tag{4.1.2}$$

$$\approx \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T \underbrace{\begin{bmatrix} \texttt{Diag}(\boldsymbol{y}_g)\boldsymbol{e}_t \\ \boldsymbol{i}_t^\ell \end{bmatrix} \begin{bmatrix} (\texttt{Diag}(\boldsymbol{y}_g)\boldsymbol{e}_t)^H & (\boldsymbol{i}_t^\ell)^H \end{bmatrix}}_{=\boldsymbol{C}_{e,t}} \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T + \boldsymbol{C}_\eta \tag{4.1.3}$$

The rank of the excitation covariance matrix,

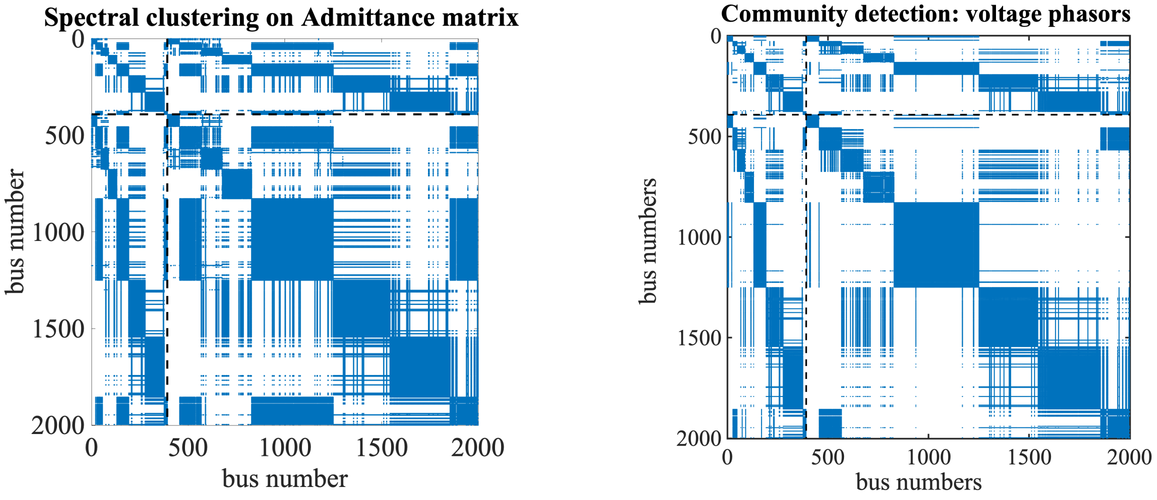
$$T^{-1}\textstyle\sum_{t=1}^{T} \boldsymbol{C}_{e,t} \tag{4.1.4}$$

is roughly the order of number of generators $|\mathcal{G}|$. Thus, the rank of $\boldsymbol{C}_v$ will be approximately K and therefore the top K eigenvectors of covariance matrix $\boldsymbol{C}_v$ can be used instead of the lowest K eigenvectors of graph Laplacian $\boldsymbol{S}$. Therefore, as in spectral clustering [118], a k-means clustering on the top K eigenvectors of $\boldsymbol{C}_v$ correspond to the smallest eigenvectors of $\boldsymbol{S}$ and can be used for community detection [38]. Clustering using k-means is performed in a higher dimension due to complex valued eigenvectors. Algorithm 6.1 specifies the steps undertaken for community detection.

Also, when the DC approximation holds, k-means clustering on the covariance matrix of phase angle measurements, $\boldsymbol{C}_\theta = T^{-1}\sum_{t=1}^{T} \boldsymbol{\theta}_t\boldsymbol{\theta}_t^T$ which has real-valued eigenvectors and non-negative eigenvalues, can also recover the community structure [35].

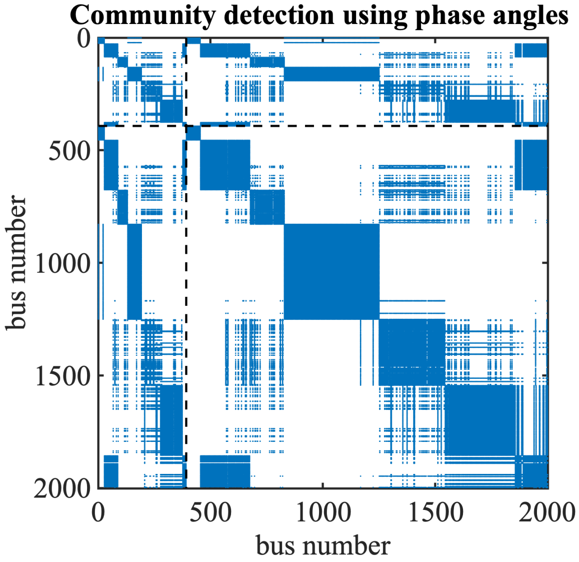
4.1.1 Numerical Results

As in Section 5.5, numerical results were obtained using data from Matpower testcase, ACTIVSg2000. From Fig.3.4.1, the community structure due to population cluster is evident. It is seen that there are $K = 8$ distinct clusters. As ground truth, spectral clustering is performed on the eigenvectors $\mathbf{U}$ of Laplacian $\boldsymbol{S}$ wherein the



(a) Spectral Clustering of Ordered Laplacian Matrix of the Network. The k-means optimal objective value $F^* = 1.8921$

(b) Community Detection Using Complex Voltage Phasors $\boldsymbol{v}_t$ of the Network. The k-means Optimal Objective value is $\tilde{F}_{\boldsymbol{y}} = 2.3369$



(c) Community detection using phase angles, $\boldsymbol{\theta}$ of the network. The k-means optimal objective value is $\tilde{F}_{\boldsymbol{\theta}} = 2.9233$

Figure 4.1.1: Spectral Clustering and Community Detection

goal was to recover K communities. The support of recovered adjacency matrix for the clusters is shown in Fig.4.1.1a. Then, algorithm 6.1 is utilized and clustered the eigenvectors of the data covariance matrix $\boldsymbol{C}_v$ where voltage phasor measurements were used. The corresponding support of adjacency matrix is given in Fig.4.1.1b. Also, to test the DC approximation, eigenvectors from the covariance matrix formed by phase angles $\boldsymbol{\theta}_t, \boldsymbol{C}_\theta$ were clustered and the results are shown in Fig.4.1.1c. For both of the above techniques for community detection, the k-means optimal objective values were $\tilde{F}_{\boldsymbol{y}} = 2.3369$ and $\tilde{F}_{\boldsymbol{\theta}} = 2.9233$ when using complex voltage phasors $\boldsymbol{v}_t$ and phase angle $\boldsymbol{\theta}$ respectively. The respective difference from optimal objective value for spectral clustering and the were found to be $\tilde{F}_{\boldsymbol{y}} - F^* = 0.4448$ and $\tilde{F}_{\boldsymbol{\theta}} - F^* = 1.0312$. This is indicative of the better performance using voltage phasor measurements although the DC approximation roughly holds for this dataset.

4.2 Fault Localization Using Under-sampled Grid Graph Signals

Line or current faults, switching events, voltage sags and faults at generators are associated with events that start with a localized spatio-temporal pattern, eventually smoothened by the low-pass response of the network. It is intuitive that to pin-point the location of such faults, one needs to be able to reconstruct sharp spatio-temporal derivatives, i.e. signals that have a wide graph-Fourier and time-Fourier transform based spectra. The question is to what extent one is able to localize faults in a system using a small number of PMU measurements, which could be sufficient in a quasi-steady state regime but are not sufficient if the signal has a broader graph-Fourier spectrum. The optimal placement of M PMUs in this case would not guarantee a good reconstruction but rather be aimed at detecting with the maximum possible resolution the fault in space and time. *Resolution* in localizing an event on a graph refers to the ability of localizing the correct ℓ-hop neighborhood, with ℓ as small as possible. Although there are both parametric and non-parametric approaches to localization such as in [119–121], the connection to topology was first made in [122] and elucidated upon in [28].

Assume that the measurements pre-fault and post-fault currents are given by

$$\text{Pre-fault: } \mathbf{I}_0 = \mathbf{Y}\boldsymbol{v}_0, \quad \text{Post-fault: } \mathbf{I}_F + \mathbf{I}_E = \mathbf{Y}\boldsymbol{v}_F \tag{4.2.1}$$

with $\mathbf{I}_E$ a sparse vector with 1 non-zero entry and equal to injected fault current at faulty bus. Subtracting pre-fault from post-fault,

$$(\mathbf{I}_F - \mathbf{I}_0) + \mathbf{I}_E = \boldsymbol{\delta}\mathbf{I} + \mathbf{I}_E = \boldsymbol{Y}(\boldsymbol{v}_F - \boldsymbol{v}_0) = \mathbf{Y}\boldsymbol{\delta v} \tag{4.2.2}$$

The current and voltages can be parsed as available and unavailable PMUs measurements:

$$\begin{pmatrix} \boldsymbol{\delta I}_a \\ \boldsymbol{\delta I}_u \end{pmatrix} + \begin{pmatrix} \boldsymbol{I}_E^a \\ \boldsymbol{I}_E^u \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_{aa} & \boldsymbol{Y}_{au} \\ \boldsymbol{Y}_{au}^T & \boldsymbol{Y}_{uu} \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta V}_a \\ \boldsymbol{\delta V}_u \end{pmatrix}. \tag{4.2.3}$$

Let us define the vector $\boldsymbol{z}$ obtained from the available measurements as follows:

$$\boldsymbol{z} = \overbrace{\left(\mathbb{I} \mid -(\boldsymbol{Y}_{aa} - \boldsymbol{Y}_{au}\boldsymbol{Y}_{uu}^{-1}\boldsymbol{Y}_{au}^T)\right)}^{\boldsymbol{H}} \begin{pmatrix} \boldsymbol{\delta I}_a \\ \boldsymbol{\delta V}_a \end{pmatrix} \tag{4.2.4}$$

Rearranging (4.2.3), the following measurement model for $\boldsymbol{z}$ can be obtained, tying the observation to the faults:

$$\boldsymbol{z} = (-\mathbb{I} \mid \overbrace{\boldsymbol{Y}_{au}\boldsymbol{Y}_{uu}^{-1}}^{\boldsymbol{C}}) \overbrace{\begin{pmatrix} \boldsymbol{I}_E^a \\ \boldsymbol{I}_E^u \end{pmatrix}}^{\boldsymbol{s}} + \overbrace{\boldsymbol{Y}_{au}\boldsymbol{Y}_{uu}^{-1}}^{\boldsymbol{C}} \overbrace{\boldsymbol{\delta I}_u}^{\boldsymbol{\epsilon}} \tag{4.2.5}$$

$$\boldsymbol{z} = [-\mathbb{I} \mid \boldsymbol{C}]\ \boldsymbol{s} + \boldsymbol{C\epsilon} \tag{4.2.6}$$

where $\boldsymbol{s}$ is sparse with fault current at non-zero entries and $\boldsymbol{\epsilon}$ is the noise term with complex normal distribution, $\boldsymbol{\epsilon} \sim \mathcal{CN}(\mathbf{0}, \sigma_\epsilon^2\mathbb{I})$. Let the singular value decomposition

(SVD) be $\boldsymbol{C} = \boldsymbol{U}_C \boldsymbol{\Sigma}_C \boldsymbol{W}_C^H$, whitening $\boldsymbol{\epsilon}$, i.e. multiplying $\boldsymbol{\Sigma}_C^{-1} \boldsymbol{U}_C^H$, gives:

$$\boldsymbol{d} = \boldsymbol{\Sigma}_C^{-1} \boldsymbol{U}_C^H \boldsymbol{z} = \boldsymbol{F}\boldsymbol{s} + \boldsymbol{\varepsilon} \tag{4.2.7}$$

where $\boldsymbol{\varepsilon} = \boldsymbol{W}_C^H \boldsymbol{\epsilon} \sim \mathcal{CN}(\mathbf{0}, \sigma_\epsilon^2 \mathbb{I})$, and $\boldsymbol{F} = \left[-\boldsymbol{\Sigma}_C^{-1} \boldsymbol{U}_C^H | \boldsymbol{W}_C^H \right]$. Model in (4.2.6) and (4.2.7) is a well-known linear model with a sparse input and additive Gaussian noise. However, remember that measurements $\boldsymbol{d}$ are obtained as a linear combination through $\boldsymbol{F}$ of graph signals at different nodes i.e. $\boldsymbol{s}$. Fault localization is then equivalent to recovering the support of $\boldsymbol{s}$. It is assumed that $\boldsymbol{s}$ is 1-sparse i.e. only 1 location at fault.

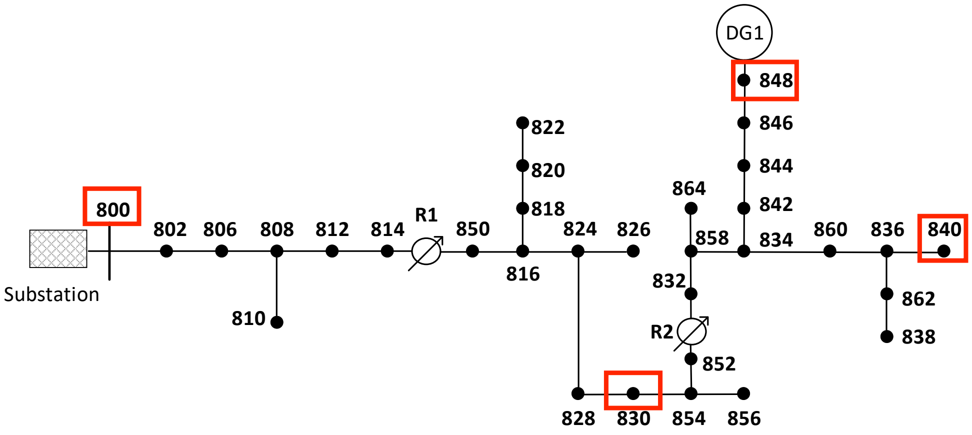
The sparsity of $\boldsymbol{s}$ underscores the problem with the under-sampled nature of the grid graph signals $\boldsymbol{d}$. Its excitation is no longer a bandlimited graph signal, all graph-frequency components are important which makes the recovery not possible when the number of PMUs is much less than number of nodes in the graph. It is then required to operate under this loss of resolution and place PMUs to obtain maximum resolution possible. Following [122], fault localization can be formulated as a multiple hypothesis testing problem based on (4.2.7) where the goal is to identify the fault location ℓ. With assumptions about the distribution of $\boldsymbol{s}$ for each hypothesis ℓ, the conditional distribution of $\boldsymbol{d}$ for hypothesis $\mathcal{H}_\ell$ is given by

$$\boldsymbol{d} | \mathcal{H}_\ell \ \sim \ \mathcal{CN}(\boldsymbol{F}\boldsymbol{\mu}_\ell, \sigma_\epsilon^2 \mathbb{I} + \boldsymbol{F}\boldsymbol{\Phi}_\ell \boldsymbol{F}^H) \tag{4.2.8}$$

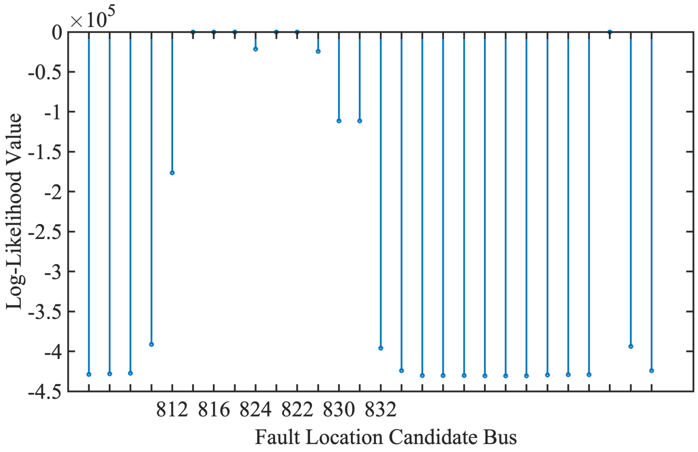
where $\boldsymbol{\mu}_\ell = \mathbb{E}[\boldsymbol{s}]$, $\boldsymbol{\Phi}_\ell = \mathbb{E}[\boldsymbol{s}\boldsymbol{s}^H | \mathcal{H}_\ell]$. The maximum-likelihood (ML) detector of the fault location is then

$$\ell^* = \underset{\ell}{\operatorname{argmax}} \ \lambda_\ell(\boldsymbol{d}) = \underset{\ell}{\operatorname{argmin}} \ (\boldsymbol{d} - \boldsymbol{m}_\ell)^H \boldsymbol{\Psi}_\ell^{-1} (\boldsymbol{d} - \boldsymbol{m}_\ell) + \ln(\pi^K |\boldsymbol{\Psi}_\ell|). \tag{4.2.9}$$

where $\boldsymbol{m}_\ell = \boldsymbol{F}\boldsymbol{\mu}_\ell$, $\boldsymbol{\Psi}_\ell = \boldsymbol{F}\boldsymbol{\Phi}_\ell \boldsymbol{F}^H$. If for a given fault at location ℓ^*, the value of log-likelihood $\lambda_{\ell^*}(\boldsymbol{d})$ is close to $\lambda_k(\boldsymbol{d})$ ($k \neq \ell^*$), the location of the fault can be mis-



(a) Reduced IEEE-34 Test Case with Added Generator. Boxes show optimal placement of PMUs.



(b) Log-likelihood value for fault at bus 820-Phase-A

Figure 4.2.1: Fault localization for IEEE-34 Bus Case [123]

identified. A cluster then can be defined as a set of nodes for which $\lambda_\ell(\boldsymbol{d}) \approx \lambda_{\ell^*}(\boldsymbol{d})$ under $\boldsymbol{y}|\mathcal{H}_{\ell^*}$. The distance between conditional distribution of data under different hypotheses is characterized by the Kullback-Leibler (KL) Divergence, $D_{KL}(f(\boldsymbol{d}|\mathcal{H}_\ell)\|f(\boldsymbol{d}|\mathcal{H}_k))$. In this case,

$$D_{KL}(f(\boldsymbol{d}|\mathcal{H}_\ell)\|f(\boldsymbol{d}|\mathcal{H}_k)) \propto (r_{\ell k})^{-1}, \quad r_{\ell k} = \frac{|\boldsymbol{f}_\ell^H \boldsymbol{f}_k|}{||\boldsymbol{f}_\ell||\ ||\boldsymbol{f}_k||} \tag{4.2.10}$$

where $\boldsymbol{f}_\ell$ is the ℓ^{th} column of $\boldsymbol{F}$. Higher the $r_{\ell k}$, more the possibility of misclassification between ℓ and k hypothesis. In other words, k, ℓ will belong to the same cluster. More small size clusters of vectors of $\boldsymbol{F}$ with high mutual correlation imply finer localization resolution. The mutual coherence of $\boldsymbol{F}$ [124] is:

$$\mu(\boldsymbol{F}) = \max_{k,\ell,k\neq\ell} r_{\ell k} \tag{4.2.11}$$

One can consider (4.2.7) as a sparse recovery problem. From the theory of recovery of sparse signals [124], it is well known that, when

$$\|\boldsymbol{x}\|_0 \leq \frac{1}{2}\left(1+\mu(\boldsymbol{F})^{-1}\right),\ \mu(\boldsymbol{F}) = \max_{k,\ell,k\neq\ell} r_{\ell k} \tag{4.2.12}$$

where $\mu(\boldsymbol{F})$ is the mutual coherence of matrix $\boldsymbol{F}$, then the ℓ_1 minimization of the support can recover $\boldsymbol{x}$. In general, the lower is the mutual coherence of a matrix, the better the recovery performance as it is similar to unitary matrices. To compute the mutual coherence, consider

$$\boldsymbol{F}^H\boldsymbol{F} = \begin{bmatrix} \boldsymbol{U}\boldsymbol{\Sigma}^{-2}\boldsymbol{U}^H & -\boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^H \\ -\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^H & \mathbb{I} \end{bmatrix}. \tag{4.2.13}$$

The magnitude of each entry of the matrix $\boldsymbol{F}^H\boldsymbol{F}$, with suitable normalization, corresponds to $r_{\ell k}$ except for the diagonal entries. Naturally, the magnitude depends on the spectrum $\boldsymbol{\Sigma}$ of the matrix $\boldsymbol{C} = \boldsymbol{Y}_{au}\boldsymbol{Y}_{uu}^{-1}$ which must be as *flat* as possible so as to lower the mutual coherence of the matrix $\boldsymbol{F}$. This suggests to minimize the so called *Shatten* infinity norm of the matrix $\boldsymbol{Y}_{au}\boldsymbol{Y}_{uu}^{-1}$, which is the infinity norm $\|\boldsymbol{\sigma}\|_\infty$ of the vector $\boldsymbol{\sigma}$ containing the non-zero singular values of $\boldsymbol{Y}_{au}\boldsymbol{Y}_{uu}^{-1}$. The design strategy, detailed in the next section, is aimed precisely at providing a flat spectrum for $\boldsymbol{C}$. In addition to interpretation using sparse recovery, one can also understand the loss of resolution and consequent best possible placement using sampling theory from graph signal processing [125]. Placing PMU sensors can be thought of as the problem of sampling a graph signal in a fixed spatial pattern. This is because voltage or current signals are graph signals that occur on the electrical network. Not having sufficient PMUs means not sampling from enough nodes in the graph. This undersampling leads to poor recovery or reconstruction. However, if the graph signal is sparse in the graph Fourier domain, then measurements of the order of the number of approximately non-zero components in the Fourier domain are enough for perfect reconstruction. For rigorous theory, please see [111]. The same theory can be applied to signals from the electrical grid like the measurement $\boldsymbol{d}$.

In [35] it was established that voltage phasor measurements are an output of a low-pass graph filter defined using the graph-shift operator of system admittance matrix $\boldsymbol{Y}$ which means that voltage signals are *sparse* or band-limited in the graph Fourier domain where graph Fourier transform is defined as $\mathbf{U}_Y^T\boldsymbol{a}$ where $\mathbf{U}_Y$ are the eigenvectors of $\boldsymbol{Y}$ and $\boldsymbol{a}$ is a graph signal. It means that measurements from a certain number of nodes say $\tilde{M}$ are needed for perfect reconstruction [125]. Since sparsity in graph Fourier domain is only approximate, optimal placement of $\tilde{M}$ measurement units on the network aims to minimize the reconstruction error. This criterion leads to the optimization problem of maximizing the singular value of $\boldsymbol{\mathcal{D}}_{\tilde{M}}\mathbf{U}_Y$ i.e. $\sigma_{min}(\boldsymbol{\mathcal{D}}_{\tilde{M}}\mathbf{U}_Y)$ where $\boldsymbol{\mathcal{D}}_{\tilde{M}}$ is a diagonal matrix with values from $\{0,1\}$ indicating 1 if the node is chosen for placement and it is the same size as $\boldsymbol{Y}$. Intuitively, this amounts to choosing the rows of $\mathbf{U}_Y$ with the smallest possible coherence or, in simpler terms, as close as possible to being orthogonal.

Similar intuition continues while looking at $\boldsymbol{C} = \boldsymbol{Y}_{au}\boldsymbol{Y}_{uu}^{-1}$. With $\mathbf{U}_Y$, it is choosing a representative row per ‘cluster’ in a graph. Similarly, in $\boldsymbol{C}$, choosing the rows of $\boldsymbol{Y}_{au}\boldsymbol{Y}_{uu}^{-1}$ to be uncorrelated is attained by having non overlapping support among of the rows in $\boldsymbol{Y}_{au}\boldsymbol{Y}_{uu}^{-1}$ and having $\boldsymbol{Y}_{uu}$ as close as possible to a block diagonal matrix, where the diagonal blocks as matched to the non zero portions of rows $\boldsymbol{Y}_{au}$. Therefore, it is possible to predict how to obtain good placements looking at the structure of the graph, and particularly of its natural clustering in sub-graphs with higher connectivity within themselves. In network science, these sub-graphs are often referred to as *communities*. This is because the spectral properties that are sought can be tied to selecting nodes in such a way that the sparsity patterns of the rows of $\boldsymbol{Y}_{au}$ and of $\boldsymbol{Y}_{uu}$ separate in clusters. Since there are few observable nodes, the rank of the matrix $\boldsymbol{Y}_{au}$ is limited by its number of rows. The algorithm performs best if the rows of $\boldsymbol{Y}_{au}\boldsymbol{Y}_{uu}^{-1}$ are as uncorrelated from each other as possible. Given the similarity of line parameter values, low correlation is mostly attained by having non overlapping support among of the rows in $\boldsymbol{Y}_{au}\boldsymbol{Y}_{uu}^{-1}$. As the next example illustrates, this can be attained by having $\boldsymbol{Y}_{uu}$ as close as possible to a block diagonal matrix, where the diagonal blocks are matched to the non zero portions of rows $\boldsymbol{Y}_{au}$. This is possible if the neighborhoods of the buses where the sensors are located, have the smallest intersection possible. For example, for the system in Fig. 4.2.2, if the PMUs are placed at bus 2 and 5 as shown, the admittance matrix is partitioned in the following desirable way:

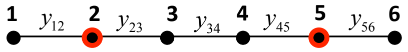


Figure 4.2.2: One-Line Diagram of a Sample Radial Network

$$\boldsymbol{Y} = \left(\begin{array}{cc|cc|cc} y_2 & 0 & -y_{12} & \multicolumn{1}{c}{-y_{23}} & 0 & 0 \\ 0 & y_5 & 0 & \multicolumn{1}{c}{0} & -y_{45} & -y_{56} \\ \hline -y_{12} & 0 & y_1 & 0 & 0 & 0 \\ -y_{23} & 0 & 0 & y_3 & -y_{34} & 0 \\ \cline{3-6} 0 & -y_{45} & 0 & -y_{34} & y_4 & 0 \\ 0 & -y_{56} & 0 & 0 & 0 & y_6 \end{array}\right) \qquad (4.2.14)$$

This design is clearly distancing the PMUs in the graph and dividing the network in neighborhoods, each associated with one of the sensors immediate and two hop neighbors where measurements are unavailable. The remaining ambiguity is confined to a connected set of buses that are topologically close to the PMU sensors. Hence, the optimum resolution for a certain grid is tied to the same intrinsic topological properties that are studied in graph clustering.

4.3 Compression of PMU Measurements

Due to their relatively high sampling rate and their wide deployment, PMU data have have called for compression. The performance of several off-the-shelf lossless encoding techniques applied to PMU data were investigated in [126]. In [127], a lossless compression called slack referenced encoding (SRE) method of PMU voltage measurements is introduced, by identifying a slack-bus and differentially encoding the difference between slack-bus measurements in time and all other buses. Similarly, in [128], phasor angle data is encoded in a lossless manner by preprocessing data using techniques from [127] and then using Golomb-Rice entropy encoding.

The idea of slow-variation with time in PMU data such as phase angles is used in [129] to firstly transform data into frequency domain (in time) and then using a 'reverse water-filling' technique to encode frequency components in the difference measurements. Many lossy compression techniques utilize the low-rank structure inherent in voltage phasor data [130–132]. Several other wavelet based compression algorithms exist in the literature as well [133].

The proposed compression algorithm leverages both (3.2.2) and (3.2.18). The measure of distortion used is the mean-squared error (MSE):

$$d(\boldsymbol{v}, \hat{\boldsymbol{v}}) \triangleq (|\mathcal{N}|T)^{-1} \sum_{t=0}^{\top} \|\boldsymbol{v}_t - \hat{\boldsymbol{v}}_t\|_2^2 \tag{4.3.1}$$

Since a temporal dynamical model for the evolution of voltage in time is used, differential encoding [134] is used to quantize the residuals in both generator and load dynamics, $\tilde{\boldsymbol{w}}_t$ and $\boldsymbol{\epsilon}_t$ respectively. The voltage at time t is:

```
Algorithm 4.2: Encoding Algorithm for Compression
Input: $\tilde{\boldsymbol{x}}_0, \tilde{\boldsymbol{x}}_1, \boldsymbol{i}^\ell_0, \boldsymbol{i}^\ell_1, \{\boldsymbol{v}_t\}_{t=2}^T$ for $t = 2 : T$ do
        States $\hat{\tilde{\boldsymbol{x}}}_t, \hat{\boldsymbol{i}}^\ell_t$ from (4.3.4) and (4.3.5) respectively. ;
        Voltage estimate:
            $\hat{\boldsymbol{v}}^0_t = \mathcal{H}(\mathbf{S}) \begin{bmatrix} \mathtt{Diag}(\boldsymbol{y}_g) \exp\left\{\mathtt{Diag}(\boldsymbol{m})^{-\frac{1}{2}} \mathbf{U}_{\text{red}} \hat{\tilde{\boldsymbol{x}}}_t\right\} \\ \hat{\boldsymbol{i}}^\ell_t \end{bmatrix} \qquad (4.3.2)$
    Compute GFT of modeling error: $\tilde{\boldsymbol{\xi}}_t = \mathbf{U}^\top (\boldsymbol{v}_t - \hat{\boldsymbol{v}}^0_t)$ ;
        Quantize: $\hat{\tilde{\boldsymbol{\xi}}}_t = \mathcal{Q}\left\{\tilde{\boldsymbol{\xi}}_t\right\}, \ \hat{\boldsymbol{v}}_t = \hat{\boldsymbol{v}}^0_t + \mathbf{U}\hat{\tilde{\boldsymbol{\xi}}}_t$;
        Update states,
            $\hat{\tilde{\boldsymbol{x}}}_t \leftarrow \mathbf{U}^\top_{\text{red}} \mathtt{Diag}(\boldsymbol{m})^{\frac{1}{2}} \ln(\hat{\boldsymbol{e}}_t), \hat{\boldsymbol{i}}^\ell_t \leftarrow [\mathbf{S}\hat{\boldsymbol{v}}_t]_{\mathcal{N}_L} \qquad (4.3.3)$
            where $\hat{\boldsymbol{e}}_t = (\mathtt{Diag}(\boldsymbol{y}_g))^{-1} [\mathbf{S}\hat{\boldsymbol{v}}_t]_{\mathcal{N}_G}$
    ;
Output: $\left\{\hat{\tilde{\boldsymbol{\xi}}}_t\right\}_{t=2}^T$
```

$$\boldsymbol{v}_t = \mathcal{H}(\mathbf{S}) \begin{bmatrix} \mathtt{Diag}\left(\boldsymbol{y}_g\right) \exp\left\{\mathtt{Diag}(\boldsymbol{m})^{-\frac{1}{2}} \mathbf{U}_{\text{red}} \left(\tilde{\boldsymbol{x}}^0_t + \tilde{\boldsymbol{w}}_t\right)\right\} \\ \boldsymbol{i}^{\ell,0}_t + \boldsymbol{\epsilon}_t \end{bmatrix}$$

$$\text{where} \quad \tilde{\boldsymbol{x}}^0_t = \mathtt{Diag}\left(\boldsymbol{a}_1\right) \tilde{\boldsymbol{x}}_{t-1} + \mathtt{Diag}\left(\boldsymbol{a}_2\right) \tilde{\boldsymbol{x}}_{t-2} \tag{4.3.4}$$

$$\boldsymbol{i}^{\ell,0}_t = \mathtt{Diag}\left(\boldsymbol{b}_1\right) \hat{\boldsymbol{i}}^\ell_{t-1} + \mathtt{Diag}\left(\boldsymbol{b}_2\right) \hat{\boldsymbol{i}}^\ell_{t-2} \tag{4.3.5}$$

Thus, $\boldsymbol{v}_t$ can be approximated as:

$$\boldsymbol{v}_t \approx \mathcal{H}(\mathbf{S}) \left(\begin{bmatrix} \mathtt{Diag}\left(\boldsymbol{y}_g\right) \exp\left\{\mathtt{Diag}(\boldsymbol{m})^{-\frac{1}{2}} \mathbf{U}_{\text{red}} \tilde{\boldsymbol{x}}^0_t\right\} \\ \boldsymbol{i}^{\ell,0}_t \end{bmatrix} + \boldsymbol{\xi}_t \right) \tag{4.3.6}$$

Note that,for the vector $\boldsymbol{\xi}_t$, the GFT $\mathbf{U}^\top \boldsymbol{\xi}_t$ has energy mostly in lower frequency components and is therefore an appropriate term to quantize using an optimal rate allocation. Specifically, bits are allocated to each component by setting a desired level of total distortion, applying the *reverse water-filling* result [135] which is optimum for a random vector $\mathbf{U}^\top \boldsymbol{\xi}_t$ whose entries are circularly symmetric complex independent Gaussian random variables and then quantize the components accordingly. The covariance matrix is not diagonal and ideally one would first whiten the vector $\mathbf{U}^\top \boldsymbol{\xi}_t$ and then quantize the individual components with bit-allocation akin to reverse water-filling. Since the statistics of $\mathbf{U}^\top \boldsymbol{\xi}_t$ are time-varying, one has to perform the whitening transform at each time instant which is a cumbersome operation. Therefore the assumption of a diagonal covariance matrix is made while sacrificing the benefit of modeling the underlying correlations among the random variables. Then, the quantized vector $\mathbf{U}^\top \boldsymbol{\xi}_t$ is used to *update* the state i.e. to estimate $\tilde{\boldsymbol{x}}_t$ and $\boldsymbol{i}_t^\ell$. Algorithms 4.2 and 4.3 describe the encoding and decoding algorithms respectively. Note that the proposed scheme of compression is sequential unlike others in literature.

Algorithm 4.3: Decoding Algorithm for Reconstruction

```
Input: $\tilde{\boldsymbol{x}}_0, \tilde{\boldsymbol{x}}_1, \boldsymbol{i}_0^\ell, \boldsymbol{i}_1^\ell, \left\{\hat{\hat{\boldsymbol{\xi}}}_t\right\}_{t=2}^T$ for $t = 2 : T$ do
    States $\hat{\tilde{\boldsymbol{x}}}_t, \hat{\boldsymbol{i}}_t^\ell$ from (4.3.4) and (4.3.5) respectively. ;
    Reconstruction of voltage from (4.3.2) $\hat{\boldsymbol{v}}_t = \hat{\boldsymbol{v}}_t^0 + \mathbf{U}\hat{\hat{\boldsymbol{\xi}}}_t$ ;
    Update states $\hat{\tilde{\boldsymbol{x}}}_t, \hat{\boldsymbol{i}}_t^\ell$ from (4.3.3) ;
Output: $\{\hat{\boldsymbol{v}}_t\}_{t=2}^T$
```

Several corrections can be made as data is collected in time such as the update of parameters $\tilde{\boldsymbol{a}}_1, \tilde{\boldsymbol{a}}_2, \boldsymbol{b}_1, \boldsymbol{b}_2$.

4.3.1 Numerical Results

: For voltage data compression, two other schemes are compared: scalar quantization and singular value thresholding (SVT) from [132]. Fig. 4.3.1 plots the em-

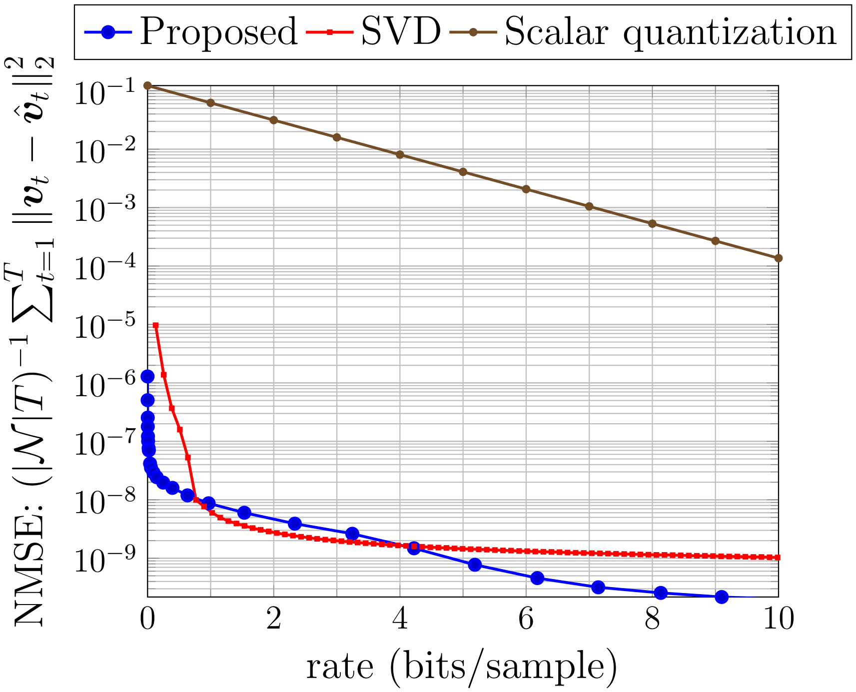


Figure 4.3.1: Empirical rate distortion (RD) curve for the proposed compression method compared with singular value thresholding and quantization.

pirical rate-distortion (RD) curve and shows the comparison between all 3 schemes. As expected, scalar quantization does poorly compared to the other schemes. The SVT scheme simply uses few of the largest singular vectors for data reconstruction. Considering that it is indicative of voltage graph signal lying in a low-dimensional subspace, it is not surprising that the SVT scheme does well. However, the SVT curve rate-distortion curve eventually saturates. Note that the performance of the proposed method are comparable to those of the SVT. However, the latter is a batch method, while the proposed method is sequential, which has important implication for the online communications of PMU data.

4.4 Detection of Anomalies and False Data Injection Attacks

Measurement systems, in general, are vulnerable to several types of attacks, including denial of service (DoS), data framing, spoofing attacks and ransomware. Thus, significant efforts have been devoted to maintain confidentiality and integrity of PMU data by: (i) securing the transfer of PMU data; (ii) protecting state estimation (SE) through bad data detection (BDD); and (iii) placing the sensors in such a way that the reconstruction is robust to data erasures. BDD techniques, dropout rejection, and stale-data replacement [136–139] are effective against random sensor errors and intermittent communications. But, a large body of prior research has shown that False Data Injection (FDI) attacks can bypass classical BDD mechanisms [140–154] and trigger incorrect decisions or hide line overflows and contingencies [155–164]). This specific vulnerability, which is the focus of this section, is not addressed by existing communication PMU and data standards. Some recent papers on PMU data integrity have proposed to leverage the low rank spatio-temporal nature of PMU data not only to help with erasures but also to strengthen conventional BDD mechanisms [165]. In [162] the authors have suggested an FDI attack strategy that can pass the aforementioned BDD approach in [165] by generating false samples that approximately preserve the original subspace structure.

In this section the problem of False Data Injection (FDI) attacks to electric power systems synchrophasors measurements is revisited through the lens of graph signal processing (GSP). First, a physics based model that supports the empirical evidence that Phasor Measurement Unit (PMU) data are low-pass graph signals is introduced. This insight is used to investigate how one can leverage such a structure to construct more effective bad data detection (BDD) algorithms that can detect FDI attack signatures through appropriate utilization of the features of the PMU graph-signal. The typical structure of the FDI attack model is as follows. The observer is interested in recovering $\boldsymbol{x}$ from a set of noisy observations:

$$\boldsymbol{z} = \mathfrak{H}(\boldsymbol{x}) + \boldsymbol{w}, \tag{4.4.1}$$

where $\mathfrak{H}(\cdot)$ represents the, possibly non-linear, system equations. We make the following assumptions the attacker: 1) the attacker has knowledge of $\mathfrak{H}(\cdot)$; 2) the attacker can only manipulate data from PMUs on buses $i \in \mathcal{M}$. By manipulating just the subset $\boldsymbol{z}_{\mathcal{M}}$ of the measurements in $\boldsymbol{z}$, for a sufficiently large set $\mathcal{M}$, the attacker can stage a stealth attack that ensures (4.4.1) is met by a signal $\tilde{\boldsymbol{x}} = \boldsymbol{x} + \delta\boldsymbol{x}$, where $\delta\boldsymbol{x}$ is the error relative to the ground truth $\boldsymbol{x}$. Let $\mathfrak{H}_{\mathcal{M}}(\cdot)$ be the functions that are associated with the compromised measurements and $\mathfrak{H}_{\mathcal{H}}(\cdot)$ the remaining (*honest*) ones. To ensure the perturbed measurements are consistent with (4.4.1), the FDI attack manipulates the measurements $\boldsymbol{z}_{\mathcal{M}}$ by adding a vector $\delta\boldsymbol{z}_{\mathcal{M}} = (\mathfrak{H}_{\mathcal{M}}(\boldsymbol{x} + \delta\boldsymbol{x}) - \mathfrak{H}_{\mathcal{M}}(\boldsymbol{x}))$ such that $(\mathfrak{H}_{\mathcal{H}}(\boldsymbol{x} + \delta\boldsymbol{x}) - \mathfrak{H}_{\mathcal{H}}(\boldsymbol{x})) = \boldsymbol{0}$. In this case a BDD based on the residual error of the model will not identify the presence of the attack. What that entails is particularly obvious in the case of a linear mapping $\mathfrak{H}(\boldsymbol{x}) = \boldsymbol{H}\boldsymbol{x}$ with

$$\delta\boldsymbol{z}_{\mathcal{M}} = \boldsymbol{H}_{\mathcal{M}}\delta\boldsymbol{x}, \quad \text{such that} \quad \boldsymbol{H}_{\mathcal{H}}\delta\boldsymbol{x} = \boldsymbol{0}. \tag{4.4.2}$$

This implies that the perturbation should be such that the vector $\delta\boldsymbol{x}$ must be in the null-space of $\boldsymbol{H}_{\mathcal{H}}$. However, the fact that $\boldsymbol{x}$ is a graph signal, adds the extra constraint $\delta\boldsymbol{x} = \mathcal{H}(\mathbf{L})\delta\boldsymbol{e}$. Hence, the stealth attack needs to achieve:

$$\boldsymbol{H}_{\mathcal{H}}\mathcal{H}(\mathbf{L})\delta\boldsymbol{e} = \boldsymbol{0} \tag{4.4.3}$$

which is impossible if $\boldsymbol{H}_{\mathcal{H}}\mathcal{H}(\mathbf{L})$ is full rank because the nullspace of $\boldsymbol{H}_{\mathcal{H}}\mathcal{H}(\mathbf{L})$ is empty and there exists no non-zero $\delta\boldsymbol{e}$. This, in a nutshell, is the benefit uncovering additional structure in $\boldsymbol{x}$. Next it is clarified how to leverage the model presented in Section 3.2 to enhance the Byzantine fault tolerance towards FDI attacks.

4.4.1 FDI Attacks on SCADA Measurements

Many studies of FDI attacks to grid measurements focus on the so-called DC power flow approximation that can be used in conjunction with power injection measurements. The model consists of a linear relationship between power injection samples collected by legacy SCADA systems and the bus voltage phase angles. More specifically, let $\boldsymbol{\theta} = \angle \boldsymbol{v}$, $\boldsymbol{B} = \Im[\boldsymbol{Y}]$ and $\boldsymbol{P} = \Re[\boldsymbol{S}]$ where $\boldsymbol{S} = \mathtt{Diag}(\boldsymbol{i}^*)\boldsymbol{v}$. The DC power flow model is $\boldsymbol{P} = \boldsymbol{B}\boldsymbol{\theta}$, and the measurements are the bus power injections $\boldsymbol{P}$. The stealth FDI attacks studied in the literature are analogous to the ones described above, with $\boldsymbol{P}$ and $\boldsymbol{\theta}$ playing the role of $\boldsymbol{z}$ and $\boldsymbol{x}$ respectively. Owing to the fact that difference in phase at two connected buses is small, by assuming the GSO of the phase angles signal is the susceptance matrix of the grid, the authors in [26] suggest the signature of an FDI attack is apparent in unexpected components in the Graph Fourier Transform (GFT) of the phase angles. Specifically, during an FDI attack, they stipulate that the phase angles vectors will exhibit a broader than normal graph spectrum, which they use as the signature to detect the attack. Next beyond the empirical arguments and justifications the case of PMU measurements which has a unique low pass, low rank structure is examined, that can be leveraged to design BDD algorithms that raise the bar of Byzantine fault tolerance to FDI attacks.

4.4.2 FDI Attacks on PMU Measurements

Assuming that access to PMU measurements of voltage and current from the buses they are installed on. Let $\mathcal{A}$ be the set of available measurements where PMUs are installed and $\mathcal{U}$ be set of unavailable ones. A measurement model can be written using ‘state’ $\boldsymbol{x}$ to be the voltage as

$$\underbrace{\begin{bmatrix} \hat{\boldsymbol{i}}_{\mathcal{A}} \\ \hat{\boldsymbol{v}}_{\mathcal{A}} \end{bmatrix}}_{\boldsymbol{z}} = \underbrace{\begin{bmatrix} \boldsymbol{Y}_{\mathcal{A}\mathcal{A}} & \boldsymbol{Y}_{\mathcal{A}\mathcal{U}} \\ \mathbb{I}_{|\mathcal{A}|} & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{H}} \underbrace{\begin{bmatrix} \boldsymbol{v}_{\mathcal{A}} \\ \boldsymbol{v}_{\mathcal{U}} \end{bmatrix}}_{\boldsymbol{x}} + \boldsymbol{w} \qquad (4.4.4)$$

The attacker follows the strategy outlined in (4.4.2) to manipulate both current and voltage on buses $i \in \mathcal{M} \subset \mathcal{A}$ by introducing a perturbation

$$\delta \boldsymbol{x}^T = \begin{bmatrix} \delta \boldsymbol{v}_{\mathcal{M}}^T & \mathbf{0}_{|\mathcal{H}|+|\mathcal{U}|}^T \end{bmatrix}, \text{ such that } \boldsymbol{Y}_{\mathcal{HM}} \delta \boldsymbol{v}_{\mathcal{M}} = \mathbf{0} \tag{4.4.5}$$

where $\mathcal{H}$ is the set of honest nodes. This requires special conditions and accurate attack placement, since $\boldsymbol{Y}_{\mathcal{HM}}$ is tall. Nonetheless, since the system admittance matrix $\boldsymbol{Y}$ is generally sparse [105], $\boldsymbol{Y}_{\mathcal{HM}}$ does not have full column-rank for a sufficient number of attackers $\mathcal{M}$ even when all the measurements are available with $\mathcal{A} = \mathcal{N}$.

4.4.3 Detection of Attack

This application is based on preliminary work in [36]. Note that, even though we cast the problem as that of FDI attacks detection, the idea can be easily extended to unveil sudden changes due to physical events (like fault-currents, or topology changes) that similarly excite high GF content. Our detection problem entails deciding between the hypotheses of attack $\mathcal{H}_1$ and no attack $\mathcal{H}_0$. To this end, we can leverage the low-dimensional generative model for the voltage graph signal that comes from (3.2.4), which imposes additional constraint on the perturbation along with that in (4.4.5). In short, for the attacker to be successful and undetected, she needs to have knowledge of system parameters and the graph filter with k frequency components $\mathcal{H}_k(\mathbf{S})$. However, since the attacker does not have all this knowledge, a typical FDI attack as studied in literature is launched using (4.4.5). Using the generative model in (3.2.4), we know that under normal operating conditions in quasi-steady state, the received data $\boldsymbol{z}$ under the no-attack and attack hypotheses $\mathcal{H}_0, \mathcal{H}_1$ respectively have the structure:

$$\boldsymbol{z} = \begin{cases} \mathcal{H}_0 : \boldsymbol{H} \mathcal{H}_k(\mathbf{S}) \begin{bmatrix} (\mathtt{Diag}(\boldsymbol{y}_g) \boldsymbol{e}_t)^\top & (\boldsymbol{i}_t^\ell)^\top \end{bmatrix}^\top + \boldsymbol{\varepsilon} \\ \mathcal{H}_1 : \boldsymbol{H} \mathcal{H}_k(\mathbf{S}) \begin{bmatrix} (\mathtt{Diag}(\boldsymbol{y}_g) \boldsymbol{e}_t)^\top & (\boldsymbol{i}_t^\ell)^\top \end{bmatrix}^\top + \boldsymbol{H} \delta \boldsymbol{v} + \boldsymbol{\varepsilon} \end{cases} \tag{4.4.6}$$

Therefore, we project the received $\boldsymbol{z}$ into the subspace orthogonal to columnspace of $\boldsymbol{H}\mathcal{H}_k(\mathbf{S})$ to get a test statistic, $d(\boldsymbol{z})$. The projector is:

$$\mathbf{\Pi}_{\perp \boldsymbol{H}\mathcal{H}_k(\mathbf{S})} \triangleq \mathbb{I} - [\boldsymbol{H}\mathcal{H}_k(\mathbf{S})]\,[\boldsymbol{H}\mathcal{H}_k(\mathbf{S})]^{\dagger} \tag{4.4.7}$$

and under the no attack hypothesis $\mathcal{H}_0$, energy in the orthogonal subspace is less than when there is an attack, $\mathcal{H}_1$. This can be converted to the following test,

$$d(\boldsymbol{z}) \triangleq \left\|\mathbf{\Pi}_{\perp \boldsymbol{H}\mathcal{H}_k(\mathbf{S})}\boldsymbol{z}\right\|_2^2 \quad \underset{H_0}{\overset{H_1}{\gtrless}} \quad \tau \tag{4.4.8}$$

where τ is a threshold that can be chosen based on an empirical receiver operator characteristics (ROC) curve. Note that, since $\mathcal{H}_k(\mathbf{S})$ is a low pass filter, the projector $\mathbf{\Pi}_{\perp \boldsymbol{H}\mathcal{H}_k(\mathbf{S})}$ in (4.4.7) is filtering high graph frequencies and the detection measures the energy on such frequencies as a signature for anomalies.

4.4.4 Isolation of Compromised Buses

Isolation of compromised buses or estimate of $\delta\boldsymbol{v}$ can also undertaken with a similar logic. Firstly, using the assumptions in the previous section we can solve the following regression problem to recover $\delta\boldsymbol{v}$, formulating a LASSO relaxation of the sparse support recovery problem:

$$\min_{\delta\boldsymbol{v}} \ \left\|\mathbf{\Pi}_{\perp \boldsymbol{H}\mathcal{H}_k(\mathbf{S})}(\boldsymbol{z}-\boldsymbol{H}\delta\boldsymbol{v})\right\|_2^2 \ \text{ subject to } \ \|\delta\boldsymbol{v}\|_1 \leq \mu \tag{4.4.9}$$

Constraint on the ℓ_1 norm is used to incorporate the prior knowledge that the attacker has access to a few measurement buses, $\mathcal{M} \ll \mathcal{N}$. Note that the performance of the algorithm is also dependent on the number of graph-frequency components i.e. k considered.

4.4.5 Numerical Results

Fig.4.4.1 shows the magnitude of the projection of the received measurement $\boldsymbol{z}$ on the orthogonal subspace $\mathbf{\Pi}_{\perp \boldsymbol{H}\mathcal{H}_k(\mathbf{S})}$. From Fig. 4.4.1 it is evident that when

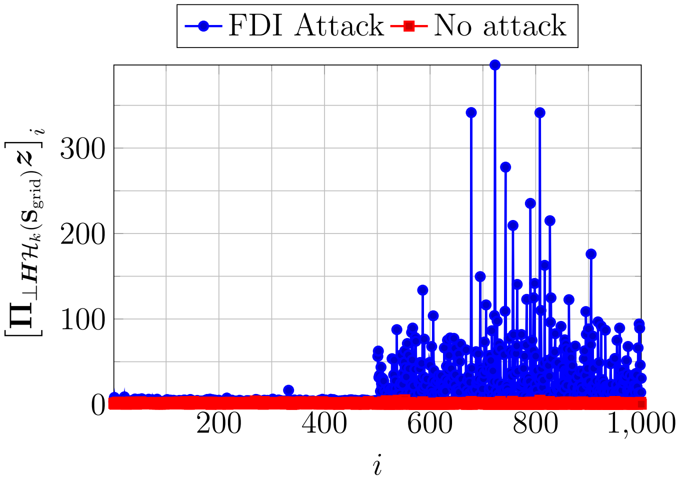


Figure 4.4.1: Components of projection of received measurement $\boldsymbol{z}$ onto the orthogonal subspace, $\mathbf{\Pi}_{\perp \boldsymbol{H}\mathcal{H}_k(\mathbf{S})}\boldsymbol{z}$, $|\mathcal{A}| = 500, |\mathcal{K}| = 200, |\mathcal{M}| = 250$.

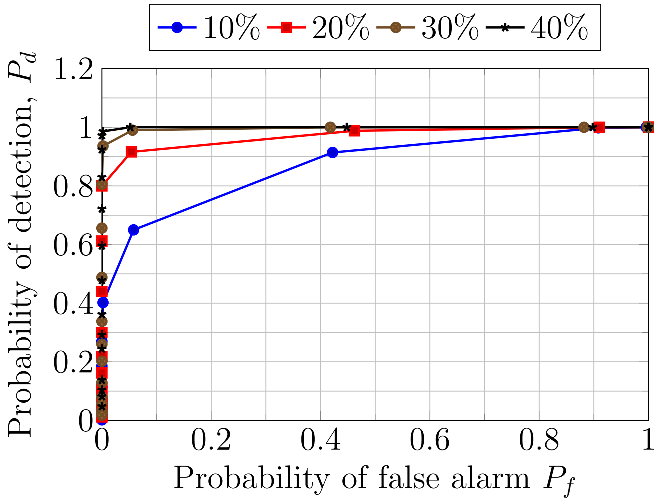


Figure 4.4.2: Empirical ROC curve for different $|\mathcal{M}|$ when a percentage of them are malicious, $|\mathcal{M}|/|\mathcal{A}| \times 100$ with $|\mathcal{A}| = 500$ (out of 2000) are available.

there is no attack, the magnitude of the projected component is orders of magnitude lower than when the measurements are under the FDI attack. This validates the idea of using high GFT frequency activity as an indicator of anomalies. Fig. 4.4.2 shows the empirical receiver operator characteristics (ROC) curve highlighting the detection performance of the proposed FDI attack detection scheme. The detection performance remains good, even when very few buses are attacked. Fig. 4.4.3 shows the reconstruction of magnitude of the attack vector $\delta\boldsymbol{v}$ when 500 measurements are

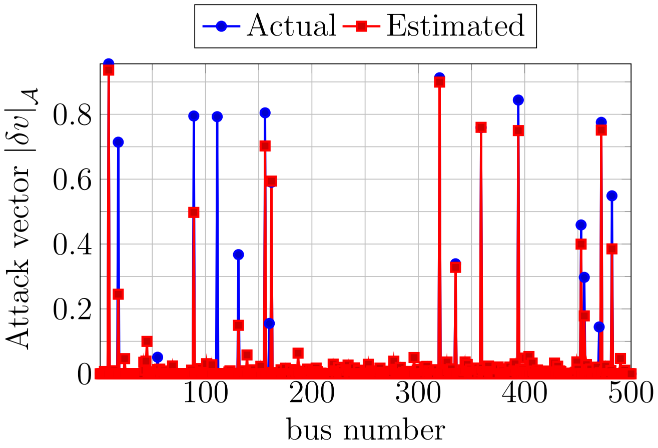


Figure 4.4.3: Reconstruction of Attack Vector

available and number of attacked buses $|\mathcal{M}| = 50$.

Chapter 5

MODEL FOR JOINT PROBABILISTIC FORECAST OF SOLAR PHOTOVOLTAIC POWER AND OUTDOOR TEMPERATURE

In this chapter, a stochastic model is proposed for a joint statistical description of solar photovoltaic (PV) power and outdoor temperature. The underlying correlation emerges from solar irradiance that is responsible in part for both the variability in solar PV power and temperature. The proposed model can be used to capture the uncertainty in solar PV power and its correlation with the electric power consumption of thermostatically controlled loads. First, a model for solar PV power that explicitly incorporates the stochasticity due to clouds via a regime-switching process between the three classes of *sunny*, *overcast* and *partly cloudy* is proposed. Then, the relationship between temperature and solar power is postulated using a second-order Volterra model. This joint modeling is leveraged to develop a joint probabilistic forecasting method for solar PV power and temperature. Real-world datasets that include solar PV power and temperature measurements in California are analyzed and the effectiveness of the joint model in providing probabilistic forecasts is verified. The proposed forecasting methodology outperforms several reference methods thus portraying that the underlying correlation between temperature and solar PV power is well defined and only requires a simple lower-complexity sampling space.

Prior work in [39] briefly described in Section 6.1 involved the development of a parametric model that was proven to efficiently capture the effect of clouds on solar PV power while providing a compact representation. In [41], the prior modeling technique is extended to fit a switching process to solar PV power, using which a solar power prediction algorithm to provide short-term probabilistic forecasts is designed.

The resulting low order model ensures reduced computational complexity for the proposed algorithm.

In Section 6.1, the deterministic modeling strategies for solar PV power and temperature from [39] and [40] respectively are reviewed. In Section 5.2, the joint stochastic models is detailed. Section 5.5 contains the numerical results by testing the forecasting methodology with real-world data and benchmarking with other reference methods.

5.1 Discrete-time Model for Solar PV Power and Temperature

In this section, the discrete time model for solar PV power output that was derived in detail in the authors' prior work in [39] and Volterra-type model relating temperature and solar irradiance developed in [40] are briefly reviewed as a precursor to the joint stochastic model. Instead of directly formulating a stochastic model, the deterministic model with parameters is first constructed. Then, the stochasticity in parameters is characterized and further leveraged in Section 5.2 to define stochastic models for solar power output and temperature.

5.1.1 Solar PV Power Model

It is hypothesized that the panel sums solar irradiation by weighting each contribution with a bi-dimensional gain function that handles the scaling factors to obtain total electrical power. The solar irradiation is attenuated by clouds modeled as a random mask that subtracts a percentage of the light coming from the patch of sky it covers at a certain time. The motion of the clouds over the panel can be approximated to be moving at a constant speed in a certain direction throughout the day. This assumption is reasonable considering the size of the panel relative to that of the displacement of the clouds. It is known that the solar irradiation has two major components [166], *direct beam component* and *diffuse beam component.* Each of these components is attenuated by the cloud coverage in different ways. Let the power produced by the PV panel be $w_d[n]$ on day d and time instant $n \in \{0, \ldots, N\}$, then:

$$w_d[n] = s_d[n] - (p_d^b[n] + p_d^{dif}[n]) + p_d^e[n] + \eta_d[n] \tag{5.1.1}$$

where $s_d[n]$ is the solar power if the dth day is sunny, $p_d^b[n]$ and $p_d^{dif}[n]$ are the components pertaining to direct and diffused beam component attenuation by the clouds respectively, $p_d^e[n]$ is attributed to edge of the cloud effect and $\eta_d[n]$ is Gaussian measurement and modeling noise.

Sunny days parametrization: Each cloudless day is modeled using a simple basis expansion model, whose expansion coefficients are periodically updated to reflect seasonal variations. Let $\mathcal{S}$ denote the set of sunny days. For $d \in \mathcal{S}$ the solar PV power samples are modeled as:

$$w_d[n] = s_d[n] = \sum_{q=0}^{Q} \mathrm{s}_{dq} b_q[n] + \eta_d[n] \tag{5.1.2}$$

where $\{b_q[n]\}$ are given basis functions; for example, one could use a set of non-overlapping cubic splines that cover three daylight periods delimited by two control points n_{k1}, n_{k2} chosen as time instants at which an abrupt change in the data (first and second order, i.e. C^1 and C^2 discontinuity) is observed. This particular basis is constructed using $Q = 9$ i.e. 10 functions derived from Bernstein polynomials [167], $B_{j,\nu}(t)$ of degree $\nu = 3$ as

$$b_{q=\nu i+j}[n] = B_{j,\nu}\left(t_i[n]\right), \; i = 0, 1, 2 \tag{5.1.3}$$

$$B_{j,\nu}(t) = \binom{\nu}{j} t^j (1-t)^{\nu-j} \mathrm{rect}(t), \; j = 0, 1, 2, 3, \tag{5.1.4}$$

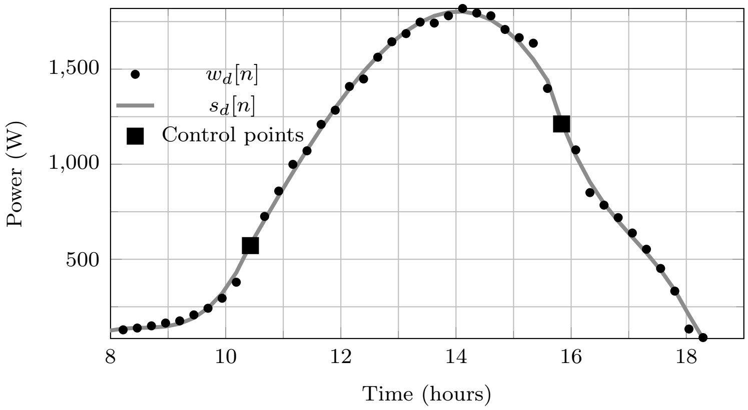


Figure 5.1.1: Plot Showing $s_d[n]$ and $w_d[n]$ for a Sunny Day on October 22, 2009

rect(t) denotes the rectangular function between $[0, 1)$ and

$$t_0[n] = n/n_{k1}, \qquad 0 \leq n < n_{k1} \tag{5.1.5}$$

$$t_1[n] = (n - n_{k1})/(n_{k2} - n_{k1}), \qquad n_{k1} \leq n < n_{k2} \tag{5.1.6}$$

$$t_2[n] = (n - n_{k2})/(N - n_{k2}), \qquad n_{k2} \leq n \leq N \tag{5.1.7}$$

Coefficients s_{dq} are found by linear regression with respect to power data $w_d[n]$. The residual error is minimal for sunny days, $w_d[n] = s_d[n]$. The approximated $s_d[n]$ for a sunny day is shown in Fig. 5.1.1. It highlights the very specific pattern obtained in October due to shading.

Cloudy days parametrization: By referring to the previous work in [39] where the expressions for $p_d^b[k]$ and $p_d^{dif}[n]$ were derived as,

$$p_d^b[n] \approx a_d^b[n]s_d[n], \quad p_d^{dif}[n] \approx \sum_q h[q]z_d[n-q] \tag{5.1.8}$$

and where $a_d^b[n]$ is the stochastic time series capturing the sudden power attenuations caused by clouds directly occluding the sun. The diffuse beam attenuation is smoother and, thus, is modeled as the convolution of a one-dimensional filter $h[k]$ with a stochastic input $z_d[n]$ that represents the cloud attenuation. Furthermore, to explain the increase of power even beyond the expected sunny day power $s_d[n]$, along the lines of direct beam attenuation, the term $p_d^e[n]$ is introduced to be present only when $w_d[n] > s_d[n]$,

$$p_d^e[n] \approx a_d^e[n] s_d[n], \tag{5.1.9}$$

This term captures the so called *edge of the cloud effect* that has been reported in literature [168, 169]. The edges of some clouds act like magnifying lenses when their paths intersect with that of the sun thereby boosting the power. It is important to note that the edge of cloud effect cannot occur simultaneously with cloud related attenuation and in general this term will be far sparser.

Complete model: From the cloudy day parametrization, the complete model for power on day d is written in vector form as

$$\boldsymbol{w}_d = \mathbf{s}_d - \mathbf{U}\left[\mathbf{S}_d \mathbf{a}_d^b + \mathcal{T}(\mathbf{h})\mathbf{z}_d\right] + \tilde{\mathbf{U}}\mathbf{S}_d \mathbf{a}_d^e + \boldsymbol{\eta}_d \tag{5.1.10}$$

where $\boldsymbol{w}_d, \mathbf{s}_d, \mathbf{a}_d^b, \mathbf{a}_d^b \in \mathbb{R}_+^N$, $\mathbf{z}_d \in \mathbb{R}_+^{(N+M-1)}$ and $\mathbf{U} = \mathrm{diag}(U(\mathbf{s}_d - \boldsymbol{w}_d))$, where $U(.)$ is the Heaviside step function operating element-wise, $\tilde{\mathbf{U}} = \mathbb{I} - \mathbf{U}$ and $\mathbb{I}$ is the identity matrix of size N. The convolution term in matrix-vector form with extended end conditions [170] as $\mathcal{T}(\mathbf{h})\mathbf{z}_d$ where, $\mathbf{h} \in \mathbb{R}_+^M$ and $\mathcal{T}(\mathbf{h}) \in \mathbb{R}_+^{N\times(N+M-1)}$ is the Toeplitz matrix with first column $\begin{bmatrix} h[M-1], & \mathbf{0}^{1\times N-1}\end{bmatrix}^T$ and first row $\begin{bmatrix} h[M-1], \ldots, h[0], & \mathbf{0}^{1\times N-1}\end{bmatrix}$. Cloud cover parameters $\mathbf{a}_d^b, \mathbf{a}_d^e$ are assumed to be sparse since direct cloud occlusions are rare events in time and $\mathbf{z}_d$ is regularized to avoid overfitting. Here, the estimation of the cloud coverage parameters is seen as a blind deconvolution problem that falls in the class of sparse dictionary learning problems [171, 172]; this can be solved, for example, by alternating between the estimation of the vectors $\mathbf{z}_d, \mathbf{a}_d^b, \mathbf{a}_d^e$ by sparse coding [173] and the estimation of filter $\mathbf{h}$ via least squares over multiple iterations. More specifically, as in a typical sparse coding problem formulation, estimates can be obtained by solving:

$$\begin{aligned}\min_{\mathbf{h},\mathbf{z}_d,\mathbf{a}_d^b,\mathbf{a}_d^e} \quad & \sum_d \left\| \mathbf{U}\left[\mathbf{s}_d - \boldsymbol{w}_d - \mathbf{S}_d\mathbf{a}_d^b - \mathcal{T}(\mathbf{h})\mathbf{z}_d\right] + \tilde{\mathbf{U}}\left[\mathbf{s}_d - \boldsymbol{w}_d + \mathbf{S}_d\mathbf{a}_d^e\right]\right\|_2^2 \\ & + \sum_d \lambda_1(\mathbf{1}^T\mathbf{a}_d^e) + \lambda_2(\mathbf{1}^T\mathbf{a}_d^b) + \lambda_3(\mathbf{1}^T\mathbf{z}_d) \qquad (5.1.11) \\ \text{subject to} \quad & \mathbf{a}_d^b \geq 0, \quad ,\mathbf{a}_d^e \geq 0, \quad \mathbf{z}_d \geq 0 \;\; \forall d, \quad \mathbf{h} \geq 0 \\ & \tilde{\mathbf{U}}\left[\mathbf{S}_d\mathbf{a}_d^b + \mathcal{T}(\mathbf{h})\mathbf{z}_d\right] = \mathbf{0}, \quad \mathbf{U}\mathbf{S}_d\mathbf{a}_d^e = \mathbf{0}\end{aligned}$$

The alternating algorithm is guaranteed to find only a locally optimal solution and it depends on the initialization [174] . Regularization constants are chosen such that $\lambda_1 \geq \lambda_2 \gg \lambda_3$ in order to force $\mathbf{a}_d^b$ and $\mathbf{a}_d^e$ to be more sparse than $\mathbf{z}_d$. Such a choice is made since sudden power attenuation due to direct occlusion of sun ($\mathbf{a}_d^b$) and the edge of the cloud effect ($\mathbf{a}_d^e$) are less likely events as compared to diffuse beam attenuation ($\mathbf{z}_d$). It was shown in [39] that the proposed model led to an excellent fit with the data. Even though the regression problem in [39] was solved in a completely deterministic fashion, such a model allowed the separation of the components and study a plausible stochastic model for them. This is explained in detail in the section 5.2.

5.1.2 Relationship Between Temperature and Solar PV Power

In [40], a Volterra model was derived to describe the relationship between temperature and solar irradiance. This model can be extended to solar PV power since the stochasticity in solar power is due to solar irradiance itself. Also, solar PV power is a linear transformation of solar irradiance [166]. Thus, the deriation in [40] can be directly applied to capture the dependency of PV power and temperatures.

Solar irradiance heats up landmass during the day that causes temperature increase in a certain geographical region. At night, heat absorbed by land is dissipated. This phenomenon is analogous to charging and discharging a capacitor in an RC circuit where voltage across the resistor can be thought of as temperature $\tau_d[n]$ at time instant n on day d. Current source is a function of solar irradiance and consequently solar PV power,

$$P_{in}^{w}[n] \approx \sum_{q=0}^{M_1-1} g[q]w_d[n-q]. \tag{5.1.12}$$

The resistor $R[n]$ and capacitor $C[n]$ themselves are quantities dependent on temperature [175] since they relate to thermal resistance of air and heat capacity [1] of the landmass respectively:

$$R[n] \approx \sum_{q=1}^{M_2} a[q]\tau[n-q],\ C[n] = \frac{C_0 + \sum_{q=1}^{M_1} b[q]\tau[n-q]}{(\tau[n]-\tau[n-1])}. \tag{5.1.13}$$

Here $g[q], a[q], b[q]$ are coefficients describing $P_{in}^{w}[n], R[n], C[n]$ respectively at a certain location. Wind and other weather related phenomena are considered as noise in the model. Fig. 5.1.2 depicts the equivalent circuit describing the physics of the system. Writing the first order linear difference equation for the analogous circuit

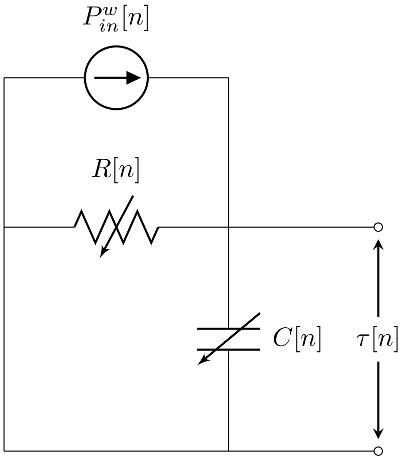


Figure 5.1.2: Equivalent thermal circuit to describe the effect of solar PV power $w_d[n]$ on temperature $\tau[n]$

---

[1] In [176], heat capacity is defined as "*the proportionality constant between the heat that the object absorbs or loses and the resulting temperature change of the object*". The heat absorbed by the surroundings to change the temperature from $\tau[n-1]$ to $\tau[n]$ can also be modeled as a linear relationship with temperature.

and performing further simplification as in [40] the model for temperature is:

$$\tau[n] \approx \sum_{q=1}^{M_2} \gamma[q]\tau[n-q] + \sum_{k=0}^{M_1-1}\sum_{p=1}^{M_1-k} \alpha[k]\tau[n-p]\tau[n-p-k]$$
$$+ \sum_{k=0}^{M_1-1}\sum_{p=1}^{M_1-k} \beta[k]w_d[n-p+1]\tau[n-p-k] + \epsilon[n] \tag{5.1.14}$$

where $\epsilon[n]$ is the modeling error. This model for temperature is interpreted as a Volterra series expansion [177, 178] up to second order with solar PV power as an input. In this way, the non-linearity in the system which is a result of temperature dependent thermal resistivity and heat capacity is captured. The parameters $\alpha[k], \beta[k], \gamma[q]$ can be estimated by solving a least-squares problem using training data of temperature and solar power output. Let a vector of parameters $\mathbf{x} \in \mathbb{R}^{M_2+2M_1}$ be defined as $\mathbf{x}^T = \begin{bmatrix} \boldsymbol{\alpha}^T & \boldsymbol{\beta}^T & \boldsymbol{\gamma}^T \end{bmatrix}$ where

$$[\boldsymbol{\alpha}]_k = \alpha[k], \;\; [\boldsymbol{\beta}]_k = \beta[k], [\boldsymbol{\gamma}]_q = \gamma[q] \tag{5.1.15}$$

For time instant n, define

$$\boldsymbol{q}_n^T = \begin{bmatrix} \boldsymbol{r}_{\tau\tau}^T & \boldsymbol{r}_{w\tau}^T & \boldsymbol{r}_{\tau}^T \end{bmatrix}, \quad \text{where}$$
$$[\boldsymbol{r}_{\tau\tau}]_k = \textstyle\sum_{p=1}^{M_1-k}\tau[n-p]\tau[n-p-k],$$
$$[\boldsymbol{r}_{w\tau}]_k = \textstyle\sum_{p=1}^{M_1-k}w_d[n-p+1]\tau[n-p-k],$$
$$k = 0, \dots M_1-1, \boldsymbol{r}_{\tau\tau} \in \mathbb{R}^{M_1}, \boldsymbol{r}_{w\tau} \in \mathbb{R}^{M_1}$$
$$[\boldsymbol{r}_{\tau}]_q = \tau[n-q], q = 1,\dots M_2, \boldsymbol{r}_{\tau} \in \mathbb{R}^{M_2}.$$

Let $\mathbf{Q} \triangleq \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_N \end{bmatrix}, \mathbf{Q} \in \mathbb{R}^{N\times M_2+2M_1}$Then, the estimate of $\hat{\mathbf{x}}$ is given by solving

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{Q}\mathbf{x} - \boldsymbol{\tau}\|_2^2, \tag{5.1.16}$$

where $\boldsymbol{\tau} \in \mathbb{R}^N, [\boldsymbol{\tau}]_n = \tau[n]$. If $\mathbf{Q}$ is full rank, $\hat{\mathbf{x}}$ can be written in terms of pseudo-inverse as $\hat{\mathbf{x}} = \mathbf{Q}^{\dagger}\boldsymbol{\tau}$.

5.1.3 Spatio-temporal Model for Solar Power and Temperature

After modeling solar PV power and temperature for a single location, it is desirable to exploit spatial correlation among solar PV panel installations that are in close proximity. For example, installations that belong to a single ZIP code which means that the geographical area is $6-7$ km in diameter share similar power patterns as shown in [179]. Let us consider multiple residential solar PV installations within a ZIP code. It can be assumed that all the locations under consideration are seeing the same cloud cover which means that the sparse parameters $\mathbf{a}_d^b, \mathbf{a}_d^e, \mathbf{z}_d$ are *approximately* same for all the locations on a day $d = 1, 2, \ldots D$. This assumption gives rise to the following model for solar PV power (in vector form) for locations indexed by $\hat{\ell} = 1, 2, \ldots, L$ similar to (5.1.10):

$$\mathbf{w}_d^{\hat{\ell}} = \mathbf{s}_d^{\hat{\ell}} - \mathbf{U}\left[\mathbf{S}_d^{\hat{\ell}}\mathbf{a}_d^b + \mathcal{T}(\mathbf{h}^{\hat{\ell}})\mathbf{z}_d\right] + \tilde{\mathbf{U}}\mathbf{S}_d^{\hat{\ell}}\mathbf{a}_d^e + \boldsymbol{\eta}_d^{\ell} \tag{5.1.17}$$

The goal is to estimate the sparse parameters $\mathbf{a}_d^b, \mathbf{a}_d^e, \mathbf{z}_d$ for each day d and filter $\mathbf{h}^{\hat{\ell}}$ for each location $\hat{\ell}$. Then, the regression problem can be written as in (5.1.11),

$$\begin{aligned} \min_{\mathbf{h}^{\hat{\ell}}, \forall\hat{\ell}, \mathbf{z}_d, \mathbf{a}_d^b, \mathbf{a}_d^e} \quad & \sum_{\hat{\ell}}\sum_{d}\left\|\mathbf{U}\left[\mathbf{s}_d^{\hat{\ell}} - \boldsymbol{w}_d^{\hat{\ell}} - \mathbf{S}_d^{\hat{\ell}}\mathbf{a}_d^b - \mathcal{T}(\mathbf{h}^{\hat{\ell}})\mathbf{z}_d\right] + \tilde{\mathbf{U}}\left[\mathbf{s}_d^{\hat{\ell}} - \boldsymbol{w}_d^{\hat{\ell}} + \mathbf{S}_d^{\hat{\ell}}\mathbf{a}_d^e\right]\right\|_2^2 \\ & + \sum_{d}\lambda_1(\mathbf{1}^T\mathbf{a}_d^e) + \lambda_2(\mathbf{1}^T\mathbf{a}_d^b) + \lambda_3(\mathbf{1}^T\mathbf{z}_d) \\ \text{subject to} \quad & \mathbf{a}_d^b \geq 0, \quad , \mathbf{a}_d^e \geq 0, \quad \mathbf{z}_d \geq 0 \ \ \forall d, \quad \mathbf{h}^{\hat{\ell}} \geq 0, \forall\hat{\ell} \\ & \tilde{\mathbf{U}}\left[\mathbf{S}_d^{\hat{\ell}}\mathbf{a}_d^b + \mathcal{T}(\mathbf{h}^{\hat{\ell}})\mathbf{z}_d\right] = \mathbf{0}, \ \mathbf{U}\mathbf{S}_d^{\hat{\ell}}\mathbf{a}_d^e = \mathbf{0} \end{aligned} \tag{5.1.18}$$

It can be solved similarly as the single location problem using alternating minimization. Outdoor temperature can be assumed to be almost constant for all locations in a ZIP code. As a proxy to solar irradiance, sum of solar power values from all locations $\hat{\ell}$ serves as the current source,

$$P_{in}^{w,\hat{\ell}}[n] \approx \sum_{q=0}^{M_1-1} g[q] \sum_{\hat{\ell}=1}^{L} w_d^{\hat{\ell}}[n-q]. \tag{5.1.19}$$

Thus, with different values for the parameters, temperature for an entire ZIP code, $\tilde{\tau}[n]$ can be modeled as in (5.1.14) with different parameters $\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}$:

$$\begin{aligned}\tilde{\tau}[n] \approx &\sum_{q=1}^{M_2} \tilde{\gamma}[q]\tilde{\tau}[n-q] + \sum_{k=0}^{M_1-1}\sum_{p=1}^{M_1-k} \tilde{\alpha}[k]\tilde{\tau}[n-p]\tilde{\tau}[n-p-k] \\ &+ \sum_{k=0}^{M_1-1}\sum_{p=1}^{M_1-k} \tilde{\beta}[k] \sum_{\hat{\ell}=1}^{L} w_d^{\hat{\ell}}[n-p+1]\tilde{\tau}[n-p-k] + \epsilon[n]\end{aligned} \tag{5.1.20}$$

5.2 Stochastic Models for Classification and Forecast of Solar Power Data

In spite of the fact that the switching regime of solar PV power is not intrinsically part of the model discussed in section 6.1 and that parameters were considered deterministic, as reported in [39], the results of the least-squares fit after solving (5.1.11) highlighted such nature for solar irradiation/power. The solar PV power produced in a period of time can be broadly classified as coming from *sunny, overcast* or *partly cloudy* models. The model switches between the three classes as shown in Fig. 5.2.1 due to weather changes. In this section, a stochastic model for each of the three classes is proposed. The first application of this model is to classify power data as coming from one of the three models to provide a probabilistic short-term forecast by assuming that the model persists.

5.2.0.1 Stochastic Model For Sunny Period

For sunny periods, it is hypothesized that the solar power is the deterministic solar power pattern i.e.,

$$w_d[n] = s_d[n] + \eta_d[n] \tag{5.2.1}$$

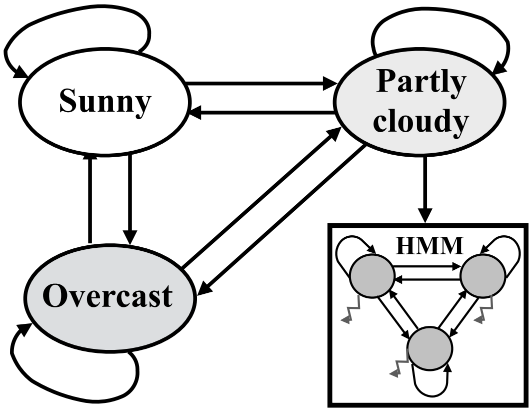


Figure 5.2.1: Block diagram highlighting the proposed switching process between stochastic models.

The modeling error is given by $\eta_d[n] \sim \mathcal{N}(0, \sigma_s^2)$ $\forall n$ and $d \in \mathcal{S}$. The variance σ_s^2 is estimated using the error values after fitting the sunny day pattern from Section. 6.1 to the sunny days, $d \in \mathcal{S}$.

5.2.0.2 Stochastic Model For Overcast Period

During overcast periods, the attenuation of solar power is mostly from the diffuse beam component [166] which is why $z_d[n]$ accounts for the relevant attenuation. Also, there is an average component in the overcast days for $z_d[n]$ that mimics a scaled version of sunny day pattern $s_d[n]$ [39]. Given the smoothness of this attenuation, the model postulated for the overcast period is:

$$w_d[n] \approx \alpha_d s_d[n] + \eta_n[k], \tag{5.2.2}$$

The parameter α_d is analogous to *clear sky index* defined as $w_d[n]/s_d[n]$ that many papers use to model solar PV power [180]. However, all the samples in the overcast period are used to estimate α_d unlike the determination of clear sky index. This leads to robustness with respect to noise. Noise $\eta_n[k]$ follows a truncated Gaussian distribution with mean 0 and noise variance σ_{oc}^2 for overcast period so as to limit the values of power $w_d[n]$ to lie in the interval $\mathcal{I}_{oc} \triangleq [0, s_d[n]]$. The distribution considered for $w_d[n]$ during overcast period is,

$$f_{oc}(w_d[n]) = \begin{cases} \dfrac{\left(2\pi\sigma_{oc}^2\right)^{-\frac{1}{2}} e^{-\left(w_d[n]-\alpha s_d[n]\right)^2/2\sigma_{oc}^2}}{\Phi\left(\frac{s_d[n]-\alpha_d s_d[n]}{\sigma_{oc}}\right) - \Phi\left(\frac{-\alpha s_d[n]}{\sigma_{oc}}\right)}, & w_d[n] \in \mathcal{I}_{oc} \\ 0 & \text{otherwise} \end{cases} \tag{5.2.3}$$

where $\Phi(.)$ denotes the CDF of a standard normal distribution.

5.2.0.3 Stochastic Model For Partly Cloudy Period

The model for the partly cloudy period includes all the three parameters in (5.1.10). However, a hidden Markov model (HMM) can capture the underlying on-off process that characterizes the sparse parameters in periods with fast moving clouds that cause sharp fluctuations in solar PV power. The observed solar PV power data $w_d[n]$ is modeled as coming from underlying hidden states that are Markovian in nature. Let the state/latent vector, $\mathbf{q}_n$ be a coordinate vector that enumerates all the possible support combinations of sparse parameters for time instant n, i.e. of the vector

$$\mathbf{x}_n = \begin{bmatrix} z_d[k-M+1] & \ldots & z_d[n] & a_d^b[k] & a_d^e[k] \end{bmatrix}^T \tag{5.2.4}$$

. The relationship between observations and latent state is:

$$w_d[n] = \alpha_d s_d[n] - \mathbf{P}\ \mathrm{diag}\left(\mathbf{\Phi}\mathbf{q}_n\right)\mathbf{x}_n \tag{5.2.5}$$

$$\mathbb{E}\left(\mathbf{q}_{n+1}|\mathbf{q}_n\right) = \mathbf{A}^T\mathbf{q}_n \tag{5.2.6}$$

$$\mathbf{P} = \begin{bmatrix} h[M-1] & \ldots & h[0] & s_d[n] & -s_d[n] \end{bmatrix} \tag{5.2.7}$$

Parameter α_d is included to model any constant attenuation over the time period considered unlike the other sparse parameters $\mathbf{x}_n$. Let the total number of states be N_s. Then, $\mathbf{A} \in \mathbb{R}^{N_s \times N_s}$ is the state transition matrix where $\mathbf{A}(i,j)$ is the probability of going from state i to state j. The state vector $\mathbf{q}_n \in \mathbb{R}^{(M+2)}$ is a binary vector taking values from the set of coordinate vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{N_s}\}$ where $\mathbf{e}_i \in \mathcal{R}^{N_s}$ has a 1 at position i and zero elsewhere. Each column in matrix $\mathbf{\Phi} \in \mathbb{R}^{(M+2)\times N_s}$ consists of a possible support of vector $\mathbf{x}_n$.

Certain assumptions are made to decrease the number of states. Firstly, the vector $\begin{bmatrix} z_n[k-M+1] & \dots & z_d[n] \end{bmatrix}$ is restricted to have $\ell < M$ non-zero entries. Secondly, $a_d^e[k]$ cannot co-exist with the other parameters due to the fact that edge of cloud effect is indicative of the absence of attenuation. As a simplification, it is also assumed that direct beam and diffuse beam attenuations do not occur together which means that the total number of states is:

$$N_s = \sum_{\tilde{\ell}=0}^{\ell} \begin{pmatrix} M \\ \tilde{\ell} \end{pmatrix} + 2 \tag{5.2.8}$$

Noise term can be ignored in the observation (5.2.5) since it is small in amplitude relative to the 'noisy' nature of the solar power data that is caused by the fast movement of clouds. The simplest case of choosing $\ell = 1$ and having $N_s = M + 3$ states is considered. All non-zero coefficients in $\mathbf{x}_n$ are hypothesized to come from independent exponential distributions with different parameters since they have different levels of sparsity. While in state i a certain $w_d[n]$ is observed:

$$w_d[n] = \begin{cases} \alpha_d s_d[n], \ i = 1 \\ \alpha_d s_d[n] - h[i-2] z_d[n-i+2], i = 2, \dots, M+1, \\ \alpha_d s_d[n] - s_d[n] a_d^b[n], \ i = M+2 \\ \alpha_d s_d[n] + s_d[n] a_d^e[n], \ i = N_s \end{cases} \tag{5.2.9}$$

where the following statistical model is postulated for the parameters

$$z_d[n] \sim \exp(\lambda_z), a_d^b[n] \sim \exp(\lambda_{a^b}), a_d^e[n] \sim \exp(\lambda_{a^e}) \tag{5.2.10}$$

The corresponding conditional probability distribution given the state i is denoted as

$f_i(w_d[n]) \triangleq f(w_d[n]|\mathbf{q}_n = \mathbf{e}_i)$ and is equal to,

$$f_i(w_d[n]) = \begin{cases} \delta(\alpha_d s_d[n] - w_d[n]),\ i = 1 \\ \frac{C_i\lambda_z}{h[i-2]}\ e^{-\frac{\lambda_z(\alpha_d s_d[n]-w_d[n])}{h[i-2]}}\ \ i=2,\cdots,M+1 \\ \frac{C_i\lambda_a^b}{s_d[n]}\ e^{-\frac{\lambda_a^b}{s_d[n]}(\alpha_d s_d[n]-w_d[n])},\ i= M+2 \\ \frac{\lambda_a^e}{s_d[n]}\ e^{-\frac{\lambda_a^e}{s_d[n]}(w_d[n]-\alpha_d s_d[n])},\ i = N_s \end{cases} \tag{5.2.11}$$

where C_i is the normalizing constant for the probability distribution given by

$$C_i^{-1} = \begin{cases} 1 - e^{-\lambda_z s_d[n]/h[i-2]},\ i = 2, 3, \ldots M + 1 \\ 1 - e^{-\lambda_a^b},\ i = M + 2 \end{cases} . \tag{5.2.12}$$

The normalization is done so that $w_d[n] \in [0, s_d[n]]$.

Learning the parameters of the HMM: Viterbi training

The models for sunny and the overcast periods are such that the only the mean of the process can be predicted in both cases. The mean during sunny period, $s_d[n]$ is assumed to be known. The mean for overcast periods can be determined by estimating the scaling α_d. Hence, the problem of learning the stochastic parameters of the model to perform predictions is non-trivial only during partly cloudy periods. To do so, it is assumed that the values of the parameters $\lambda_z, \lambda_a^b, \lambda_a^e$ of conditional probability distributions are known. It was seen that the algorithm is not very sensitive to the exact values of these parameters as long as they follow $\lambda_z \leq \lambda_{a^b} \leq \lambda_a^e$ which is consistent with the regularization performed in the regression problem. The probability of starting from a state i denoted by $\pi_i = 1/N_s$ is also assumed to be known. In order to learn the the state transition matrix $\mathbf{A}$, Viterbi training [181] or segmental k-means [182] approach was adopted. Let

$$\xi = \{\mathbf{A}(i,j)|i, j \in \{1, \cdots, N_s\}\} \tag{5.2.13}$$

be the set of unknown parameters to be estimated. Let

$$\mathbf{Q} = \begin{bmatrix}\mathbf{q}_1 & \dots & \mathbf{q}_N\end{bmatrix} \text{ and } \boldsymbol{w} = \begin{bmatrix}w_d[1] & \dots & w_d[N]\end{bmatrix}$$

denote the sequence of hidden states and solar power observations respectively. In the Viterbi training algorithm, instead of maximizing the likelihood over all possible state sequences, $\bar{\mathcal{Q}}$, the likelihood is maximized only over the most probable state sequence to find the estimates of parameters in ξ. The algorithm starts with an initial estimate for all the unknown parameters $\xi_0 = \{\mathbf{A}_0(i,j) | i,j \in \{1, \cdots, N_s\}\}$ and performs this maximization iteratively [182],

$$\hat{\xi}_k = \arg\max_{\xi} \left(\max_{\mathbf{Q}} f(\boldsymbol{w}, \mathbf{Q} | \hat{\xi}_{k-1}) \right) \tag{5.2.14}$$

where k is the iteration number and

$$f(\boldsymbol{w}, \mathbf{Q}|\xi) = f(\mathbf{q}_1) \prod_{n=1}^{N} f(w_d[n]|\mathbf{q}_n, \xi) \prod_{n=1}^{N-1} f(\mathbf{q}_{n+1}|\mathbf{q}_n, \xi) \tag{5.2.15}$$

The inner maximization is performed by using Viterbi algorithm [183]. As a result of this maximization,

$$\hat{\mathbf{Q}}_k = \arg\max_{\mathbf{Q}} f(\boldsymbol{w}, \mathbf{Q}|\hat{\xi}_{k-1}) = \hat{\mathbf{q}}_1^k, \hat{\mathbf{q}}_2^k, \dots, \hat{\mathbf{q}}_{\tilde{N}}^k, \tag{5.2.16}$$

is the most likely state sequence at iteration k which best describes the observed data. Then maximum-likelihood (ML) estimates, $\hat{\xi}_k$, are

$$\hat{\xi}_k = \arg\max_{\xi} f(\boldsymbol{w}, \hat{\mathbf{Q}}_m | \hat{\xi}_{k-1}) \tag{5.2.17}$$

Maximizing $\log f(\boldsymbol{w}, \hat{\mathbf{Q}}_k | \hat{\xi}_{k-1})$ with respect to $\mathbf{A}_k(i,j)$ under the constraint that $\mathbf{A}_k$ is a column-stochastic matrix gives

$$\hat{\mathbf{A}}_k(i,j) = \hat{N}_{ij} \Bigg/ \sum_{j=1}^{N_s} \hat{N}_{ij} \tag{5.2.18}$$

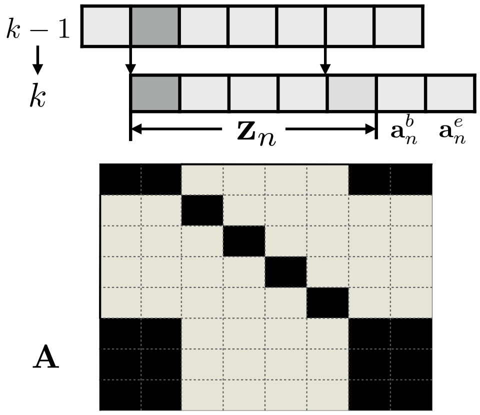


Figure 5.2.2: The specific way in which state transition from time instant $k-1$ to k takes places determines the structure of the state transition matrix $\mathbf{A}$

where $\hat{N}_{ij}$ is the number of times the transition from state i to state j occurs within the state sequence $\hat{\mathbf{Q}}_k$. Following from section 5.2.0.3 wherein the number of active coefficients at time instant n in $\mathbf{x}_n$ is restricted to 1, only a limited number of transitions from state i are possible. Also, since z_n is the input to a filter with memory M, it means that $M-1$ components need to be retained and shifted while a new one comes in. All of the above reasons give the state transition matrix $\mathbf{A}$ a sparse and specific structure as shown in Fig. 5.2.2 which is forced on $\mathbf{A}_0$ during the initialization . As a result, only $(M-1)+4\times 3$ entries of the matrix need to be estimated when $\ell = 1$ instead of $(N_s)^2$.

5.3 Regime Classification for Prediction

The premise for prediction is the persistence in the weather condition for the time horizon over which a forecast of solar power is provided. Therefore, the proposed prediction algorithm has two steps:

- Classification of the solar power from a given period as coming from one of the three classes of models: *sunny, overcast, partly cloudy*
- Assuming that this weather condition *persists* for the duration of the prediction

horizon and provide with a point forecast corresponding to the class decided in the classification step.

Such a scheme captures the inherent switching behavior that solar power exhibits, i.e. that of going from one model to another, while persisting for a certain duration in each of these. Note that the classification step can be skipped if prior knowledge in the form of weather prediction is available.

5.3.1 Classification Algorithm for Solar Power

The classification algorithm uses the stochastic models for the solar power data as detailed in 5.2. The underlying principle behind classification leverages the fact that it is easier to classify a segment of data as coming from *sunny* or *overcast* regime by computing the negative log-likelihood since they both follow Gaussian distributions with different mean and variance. If the negative log- likelihood is low, then the means are compared to distinguish between the two classes. If neither of the negative log-likelihoods is lower than a certain threshold, then it is classified to be *partly cloudy*. The thresholds for classification are determined empirically using training data. Let $\boldsymbol{w}_n^{-M_1} \triangleq \begin{bmatrix} w_d[n-M_1] & \dots & w_d[n] \end{bmatrix}$ be the M_1 solar power samples that have to be classified into one of the three regimes. Here, M_1 is the length of the filter $\beta[q]$ in (5.1.14). The algorithm begins by computing the error made by assuming *sunny* period and comparing with some threshold. If

$$\sum_{k=n-M_1}^{n} (w_d[k] - s_d[k])^2 \leq \tau_{sunny}, \tag{5.3.1}$$

then it is classified as a *sunny* period, where the threshold is chosen so that $\tau_{sunny} = \mu\sigma_s$, $\mu > 1$. If that is not the case, then the hypothesis that it is *overcast* is tested by calculating

$$\hat{\alpha}_n = \arg\min_{\alpha_n} \sum_{k=n-M_1}^{n} (w_d[k] - \alpha_n s_d[k])^2 \tag{5.3.2}$$

and checking if both the value of $\hat{\alpha}_n$ as well as the estimation error is less than some threshold,

$$\hat{\alpha}_n \leq \tau_\alpha, \text{ and } \sum_{k=n-M_1}^{n} (w_d[k] - \hat{\alpha}_n s_d[k])^2 \leq \tau_{overcast}. \tag{5.3.3}$$

If these conditions are also not satisfied, then the regime is classified as *partly cloudy*. If the decision is in favor of *partly cloudy*, the most likely state sequence $\mathbf{Q}_{M_1}$ that generated the power observations $\boldsymbol{w}_n^{-M_1}$ is determined using the Viterbi algorithm with state transition matrix $\hat{\mathbf{A}}$. Fig.5.3.1 shows the flowchart describing the classification algorithm.

5.3.2 *Prediction for Each Class of Models*

Based on the classification results on $\boldsymbol{w}_n^{-M_1}$, a probabilistic solar power forecast for m time steps ahead is provided. It is an m dimensional probability density function, $f\left(\boldsymbol{w}_n^{+m} | \boldsymbol{w}_n^{-M_1}\right)$ where $\boldsymbol{w}_n^{+m} \triangleq \begin{bmatrix} w_d[n+1] & \dots & w_d[n+m] \end{bmatrix}$.

5.3.2.1 Prediction Using Sunny Model

When the classification algorithm chooses the hypothesis that the present solar power data is from a *sunny* model, then:

$$f_s\left(\boldsymbol{w}_n^{+m} | \boldsymbol{w}_n^{-M_1}\right) = (2\pi\sigma_s)^{-\frac{m}{2}} e^{-\left\|\boldsymbol{w}_n^{+m} - \boldsymbol{s}_n^{+m}\right\|_2^2 \Big/ 2\sigma_s^2} \tag{5.3.4}$$

where $\boldsymbol{s}_n^{+m} \triangleq \begin{bmatrix} s_d[n+1] & \dots & s_d[n+m] \end{bmatrix}$ Note that the deterministic sequence of the sunny day solar power pattern is known beforehand and it is updated at a very slow pace on days that are classified as being sunny, to adjust for seasonal variations.

5.3.2.2 Prediction Using Overcast Model

If it is overcast, the predicted distribution is

$$f_{oc}\left(\boldsymbol{w}_n^{+m}|\boldsymbol{w}_n^{-M_1}\right) = \frac{(2\pi\sigma_{oc})^{-\frac{m}{2}}\, e^{-\left\|\boldsymbol{w}_n^{+m}-\hat{\alpha}_n\boldsymbol{s}_n^{+m}\right\|_2^2 \Big/ 2\sigma_{oc}^2}}{\prod_{i=1}^{m}\left(\Phi\left(\frac{s_d[n+i]-\hat{\alpha}_n s_d[n+i]}{\sigma_{oc}}\right)-\Phi\left(\frac{-\hat{\alpha}_n s_d[n+i]}{\sigma_{oc}}\right)\right)}, \tag{5.3.5}$$

when $w_d[n+i] \in [0, s_d[n+i]]\,, \forall i$ and $\hat{\alpha}_n$ is estimated using (5.3.2).

5.3.2.3 Prediction Using Partly Cloudy Model

Since solar PV power on a partly cloudy day has an underlying Markov model, the estimated state transition matrix $\hat{\mathbf{A}}$ and state at time n is used to compute a forecast distribution for m steps ahead as

$$f_{pc}(\boldsymbol{w}_n^{+m}|\boldsymbol{w}_n^{-M_1}) = \prod_{\ell=1}^{m}\sum_{j=1}^{N_s}\mathbf{A}^{\ell}(\hat{i},j)f_j(w_d[n]) \tag{5.3.6}$$

where $\hat{i}$ is the most probable state at time n. This is estimated by Viterbi algorithm [184, see section III.B] using solar power values $\boldsymbol{w}_n^{-M_1}$. Therefore, the predictive distribution of solar PV power can be summarized as

$$f\left(\boldsymbol{w}_n^{+m}\middle|\boldsymbol{w}_n^{-M_1}\right) = \begin{cases} f_s(\boldsymbol{w}_n^{+m}\big|\boldsymbol{w}_n^{-M_1}), \text{sunny} \\ f_{oc}(\boldsymbol{w}_n^{+m}\big|\boldsymbol{w}_n^{-M_1}), \text{overcast} \\ f_{pc}(\boldsymbol{w}_n^{+m}\big|\boldsymbol{w}_n^{-M_1}), \text{ partly cloudy} \end{cases} \tag{5.3.7}$$

where the expressions of $f_s(\boldsymbol{w}_n^{+m}\big|\boldsymbol{w}_n^{-M_1}), f_{oc}(\boldsymbol{w}_n^{+m}\big|\boldsymbol{w}_n^{-M_1})$ and $f_{pc}(\boldsymbol{w}_n^{+m}\big|\boldsymbol{w}_n^{-M_1})$ are in (5.3.4), (5.3.5) and (5.3.6) respectively.

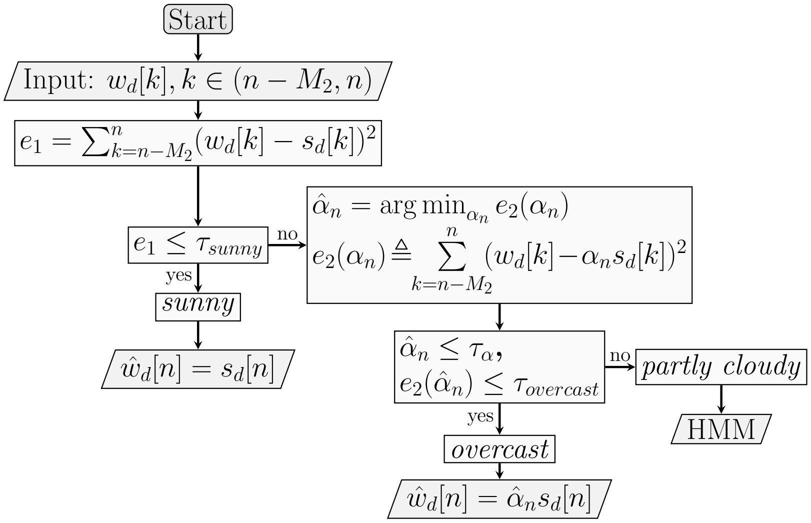


Figure 5.3.1: Flowchart of the Solar Power Prediction Algorithm

5.4 Joint Probabilistic Forecasts

Since the relationship between temperature and solar PV power was established in section.6.1, joint description of future values, i.e. probabilistic forecasts of temperature and solar power can be provided.

5.4.1 Conditional Distribution of Temperature Given Solar Power

After solving the problem in (5.1.16) in section 5.1.2, residual $\epsilon[n]$ from (5.1.14) was analyzed and found that it can be approximated to be drawn from a Laplacian distribution with parameter μ. This is illustrated in Fig. 5.4.1. To write the conditional distribution of temperature at time n given all the past values of temperature and solar PV power, define the vectors

$$\boldsymbol{\tau}_n^{-M_2} = \begin{bmatrix} \tau[n] & \tau[n-1] & \cdots & \tau[n-M_2] \end{bmatrix}^T$$

$$\boldsymbol{s}_n^{-M_1} = \begin{bmatrix} s_d[n] & s_d[n-1] & \cdots & s_d[n-M_1] \end{bmatrix}^T$$

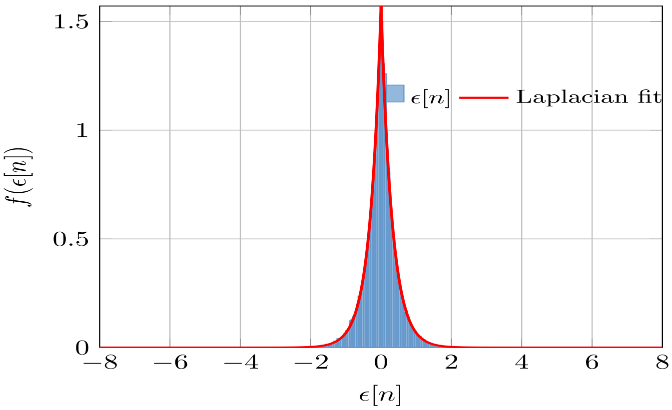


Figure 5.4.1: Laplacian Fit for $\epsilon[n]$ with $\mu = 3.1422$

Then the conditional distribution from (5.1.14) is

$$f\left(\tau[n]\middle|\boldsymbol{w}_n^{-M_1}, \boldsymbol{\tau}_{n-1}^{-M_2}\right) = \frac{\mu}{2} e^{-\mu\left|\tau[n] - g\left(\boldsymbol{w}_n^{-M_1}, \boldsymbol{\tau}_{n-1}^{-M_2}\right)\right|}, \tag{5.4.1}$$

$$\begin{aligned} g(\boldsymbol{w}_n^{-M_1}, \boldsymbol{\tau}_{n-1}^{-M_2}) = & \sum_{q=1}^{M_2} \gamma[q]\tau[n-q] \\ & + \sum_{k=0}^{M_1-1} \sum_{p=1}^{M_1-k} \alpha[k]\tau[n-p]\tau[n-p-k] \\ & + \sum_{k=0}^{M_1-1} \sum_{p=1}^{M_1-k} \beta[k] w_d[n-p+1]\tau[n-p-k] \end{aligned} \tag{5.4.2}$$

5.4.2 Joint Distribution and Probabilistic Forecast

From the conditional probability formula, the joint distribution of $\tau[n], w_d[n]$ at time n given $\boldsymbol{w}_{n-1}^{-M_1}$ and $\boldsymbol{\tau}_{n-1}^{-M_2}$ is

$$f\left(\tau[n], w[n]\middle|\boldsymbol{w}_{n-1}^{-M_1}, \boldsymbol{\tau}_{n-1}^{-M_2}\right) = \tag{5.4.3}$$

$$f\left(\tau[n]\middle|w[n], \boldsymbol{w}_{n-1}^{-M_1}, \boldsymbol{\tau}_{n-1}^{-M_2}\right) f\left(w_d[n]\right) \tag{5.4.4}$$

Let the vectors of m steps ahead forecasts of $\tau[n]$ be

$$\boldsymbol{\tau}_n^{+m} = \begin{bmatrix} \tau[n+m] & \cdots & \tau[n+1] \end{bmatrix}^T, \boldsymbol{\tau}_n^{+0} = \boldsymbol{\tau}_n^{-M_2}$$

Probabilistic forecasts i.e. the joint distribution which following (5.4.3) and the probability chain rule can be written as,

$$f\left(\boldsymbol{w}_n^{+m}, \boldsymbol{\tau}_n^{+m} \middle| \boldsymbol{w}_n^{-M_1}, \boldsymbol{\tau}_n^{-M_2}\right) = \tag{5.4.5}$$
$$\prod_{\ell=1}^{m} f\left(\tau[n+\ell] \middle| ,\boldsymbol{w}_n^{+m}, \boldsymbol{\tau}_n^{+\ell-1}, \boldsymbol{w}_n^{-M_1}, \boldsymbol{\tau}_n^{-M_2}\right) f\left(\boldsymbol{w}_n^{+m} \middle| \boldsymbol{w}_n^{-M_1}\right)$$

Each component in the product over ℓ is the conditional distribution of temperature $\tau[n+\ell]$ given current and all past values of solar power, $\boldsymbol{w}_n^{+\ell}, \boldsymbol{w}_n^{-M_1}$ and past values of temperature $\boldsymbol{\tau}_n^{+\ell-1}, \boldsymbol{\tau}_n^{-M_2}$. This can be written using (5.4.1). The probabilistic forecast of solar PV power for m steps ahead, $f\left(\boldsymbol{w}_n^{+m} \middle| \boldsymbol{w}_n^{-M_1}\right)$ is given by (5.3.7).

Spatio-temporal model for solar power and temperature: Since the stochastic parameters for solar power remain the same for a ZIP code based on the modeling assumption, installations within a ZIP code have solar power from the same regime for a certain period of time. Given the regime, the predicted distributions at different locations come from the same family of distributions with the parameters of the distribution different due to solar sunny day power pattern $s_d^{\hat{\ell}}$ and the filter $\boldsymbol{h}^{\hat{\ell}}$ which are not the same across locations. Analogous to the single location case, let $\boldsymbol{w}_{n,\hat{\ell}}^{-M_1}, \boldsymbol{w}_{n,\hat{\ell}}^{+m}$ be the power at $\hat{\ell}$-th location with sunny day solar power pattern $\boldsymbol{s}_{n,\hat{\ell}}^{-M_1}$. Then, the predictive distribution of solar PV power for all locations can be written as

$$f\left(\left\{\boldsymbol{w}_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^{L} \middle| \left\{\boldsymbol{w}_{n,\hat{\ell}}^{-M_1}\right\}_{\hat{\ell}=1}^{L}\right) = \tag{5.4.6}$$
$$\begin{cases} \prod_{\hat{\ell}=1}^{L} f_s(\boldsymbol{w}_{n,\hat{\ell}}^{+m} \big| \boldsymbol{w}_{n,\hat{\ell}}^{-M_1}), \text{sunny} \\ \prod_{\hat{\ell}=1}^{L} f_{oc}(\boldsymbol{w}_{n,\hat{\ell}}^{+m} \big| \boldsymbol{w}_{n,\hat{\ell}}^{-M_1}), \text{overcast} \\ \prod_{\hat{\ell}=1}^{L} f_{pc}(\boldsymbol{w}_{n,\hat{\ell}}^{+m} \big| \boldsymbol{w}_{n,\hat{\ell}}^{-M_1}), \text{ partly cloudy} \end{cases}$$

The conditional distribution of temperature from (5.1.20) given solar power at all locations is

$$f\left(\tilde{\tau}[n]\middle|\left\{\boldsymbol{w}_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^{L},\tilde{\boldsymbol{\tau}}_{n-1}^{-M_2}\right)=\frac{\tilde{\mu}}{2}e^{-\tilde{\mu}\left|\tilde{\tau}[n]-g\left(\left\{\boldsymbol{w}_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^{L},\tilde{\boldsymbol{\tau}}_{n-1}^{-M_2}\right)\right|},$$

$$\text{where } g\left(\left\{\boldsymbol{w}_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^{L},\tilde{\boldsymbol{\tau}}_{n-1}^{-M_2}\right)= \tag{5.4.7}$$
$$\sum_{q=1}^{M_2}\tilde{\gamma}[q]\tilde{\tau}[n-q]+\sum_{k=0}^{M_1-1}\sum_{p=1}^{M_1-k}\tilde{\alpha}[k]\tilde{\tau}[n-p]\tilde{\tau}[n-p-k]$$
$$+\sum_{k=0}^{M_1-1}\sum_{p=1}^{M_1-k}\tilde{\beta}[k]\sum_{\hat{\ell}=1}^{L}w_d^{\hat{\ell}}[n-p+1]\tilde{\tau}[n-p-k]$$

The joint probabilistic distribution for m steps ahead can be written as,

$$f\left(\left\{\boldsymbol{w}_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^{L},\tilde{\boldsymbol{\tau}}_{n}^{+m}\middle|\left\{\boldsymbol{w}_{n,\hat{\ell}}^{-M_1}\right\}_{\hat{\ell}=1}^{L},\tilde{\boldsymbol{\tau}}_{n}^{-M_2}\right)= \tag{5.4.8}$$
$$\prod_{\ell=1}^{m}f\left(\tilde{\tau}[n+\ell]\middle|,\left\{\boldsymbol{w}_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^{L},\tilde{\boldsymbol{\tau}}_{n}^{+\ell-1},\left\{\boldsymbol{w}_{n,\hat{\ell}}^{-M_1}\right\}_{\hat{\ell}=1}^{L},\tilde{\boldsymbol{\tau}}_{n}^{-M_2}\right)$$
$$f\left(\left\{\boldsymbol{w}_{n,\hat{\ell}}^{+m}\right\}_{\hat{\ell}=1}^{L}\middle|\left\{\boldsymbol{w}_{n,\hat{\ell}}^{-M_1}\right\}_{\hat{\ell}=1}^{L}\right)$$

5.5 Numerical Results

In this section, firstly the results for solar PV power forecasting alone are presented. A dataset of solar PV power from a rooftop installation in California is used for this purpose. Then, the results for joint forecasting are presented using datasets from multiple weather stations in California. In this dataset, solar irradiance is available and not solar power. Since the source of stochasticity remains the same, the effect of the algorithm does not change.

5.5.1 Solar PV Power Forecasting

The dataset used for this work was acquired from a rooftop panel installation in Antioch, California and was provided by SolarCity. This dataset was also used in authors' prior work in [39]. The format of this solar power data consisted of current (in A), voltage measurements (in V) and timestamps (in Hours) at the inverter approximately every 15 minutes recorded for a duration of two years. Each panel had a rating of 170 W and there were a total of 22 panels. Therefore, the nameplate rating of all panels combined was $170 \times 22 = 3740$ W.

5.5.1.1 Algorithm Settings

To solve (5.1.11), the alternating minimization algorithm is initialized with the filter $\boldsymbol{h}$ being a scaled Hamming window of length M, $h[q] = g \times (0.54 - 0.46 \cos(2\pi q/(M-1)))$. To address the scale ambiguity inherent in blind deconvolution problems, the scale g is chosen such that $h[q]$ and $s_d[n]$ have similar amplitudes. Prediction horizon was $m = 12$ i.e. 3 hours and a filter, $h[n]$, of length $M = 5$ was used for *partly cloudy* conditions. The algorithm started with $M_1 = 6$ samples (1.5 hour) for each day and predicted for the next m samples. Then, the window was moved by one sample.

5.5.1.2 Metrics For Evaluation

The results are presented using probabilistic forecast metrics. In the probabilistic setting at each m-step prediction there is a cumulative distribution function (CDF) $F_m(w[n]) \triangleq F\left(w[n+m] \middle| \boldsymbol{w}_n^{-M_1}\right)$ instead of a point forecast. The metrics used for evaluation are continuous rank probability score (CRPS) [185], reliability metric and score [186]. CRPS is defined for each m-step prediction as an average over all the samples,

$$\text{CRPS}(m) = (\tilde{P})^{-1} \sum_{n,d} \int_0^{\infty} (F_m(y) - u(y - w_d[n]))^2 \, dy \tag{5.5.1}$$

where $u(.)$ is the Heaviside step function and $\tilde{P}$ is the total number of samples. The CRPS evaluates to mean absolute error (MAE) when it is a point forecast.

Reliability of a probabilistic forecasting method is a useful metric in understanding the proximity of the estimated CDF to the actual CDF of the data. Let a probability interval (PI), I_m^b, be defined such that the interval covers the observed value $w_d[n]$ with probability $(1 - b)$. Then, to calculate the reliability, define

$$R_b(m) = (\tilde{P})^{-1} \sum_{n,d} \mathbb{I}_{\left\{ w_d[n] \in I_m^b \right\}} \tag{5.5.2}$$

as the estimated probability of coverage where $\mathbb{I}_{\{.\}}$ is an indicator function with value 1 if the observed sample belongs to the probability interval. Now, the probabilistic forecast is more reliable if the quantity ΔR_b is small,

$$\Delta R_b(m) \triangleq R_b(m) - (1 - b). \tag{5.5.3}$$

Another metric used for evaluation is the interval score [187]. This metric is helpful in determining the sharpness of the forecast probability interval by imposing a penalty when an observation is outside the interval, by a value proportional to the size of the interval. If the upper and lower bounds of the PI are denoted as U_m^b and L_m^b respectively then interval score is defined as

$$\text{Score}_b^m[k] = \begin{cases} D_m^b - 4\left(L_m^b - w_d[n]\right), w_d[n] < L_m^b \\ D_m^b, \ w_d[n] \in I_m^b \\ D_m^b - 4\left(w_d[n] - U_m\right), w_d[n] > U_m \end{cases} \tag{5.5.4}$$

$$\text{where} \ \ D_m^b \triangleq -2b\left(U_m - L_m^b\right) \tag{5.5.5}$$

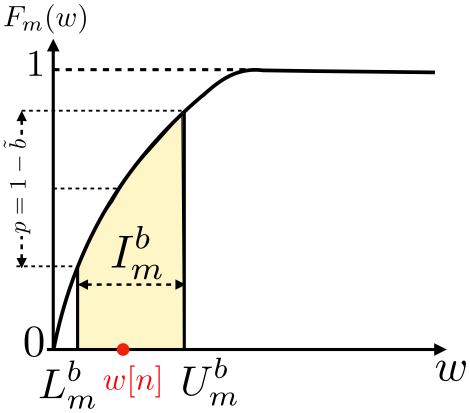


Figure 5.5.1: Figure highlighting predictive density and the corresponding probability interval.

The average interval score is,

$$\mathrm{Score}_b(m) = (\tilde{P})^{-1} \sum_{n,d} \mathrm{Score}_b^m[k] \tag{5.5.6}$$

Lower values of this score indicate sharper and more reliable forecasts. Fig.5.5.1 illustrates the predictive density $F_m(w)$ and the corresponding probability interval I_m^b. Performance of the prediction methods is also analyzed using average reliability and score defined as

$$R_b^{\mathrm{avg}} = \sum_{m=1}^{\chi} R_b(m)/\chi, \qquad \mathrm{Score}_b^{\mathrm{avg}} = \sum_{m=1}^{\chi} \mathrm{Score}_b(m)/\chi \tag{5.5.7}$$

As representative examples, Fig. 5.5.2, 5.5.3 and 5.5.4 show the actual and predicted power for different days with a multitude of weather conditions. This predicted power is one-step prediction, $m = 1$. For days that are entirely *overcast* or *sunny*, predictions have little error as can be seen in Fig. 5.5.2 and 5.5.3. These results highlight that the stochasticity of power in both these regimes is minimal leading to better predictions if the weather condition persists. However, there is higher error whenever there is a change in regime, for example going from *partly cloudy* condition to *overcast* around 12 PM as seen in Fig. 5.5.4. It can be attributed to the delay in detecting the change in model. This uncertainty cannot be avoided in days with sudden change in weather

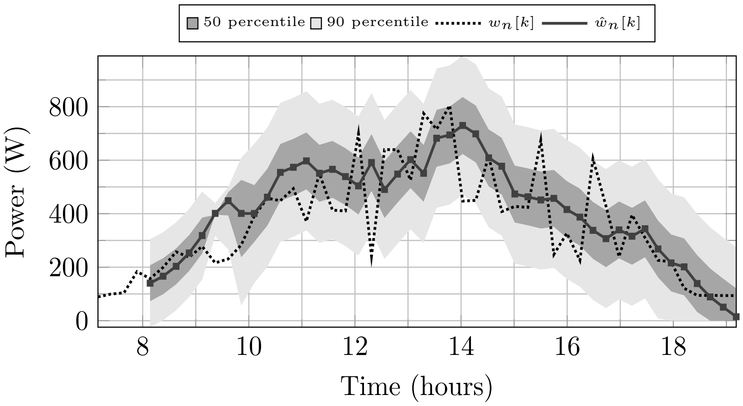


Figure 5.5.2: Plot of actual and predicted value with one-step prediction for day that is *overcast*

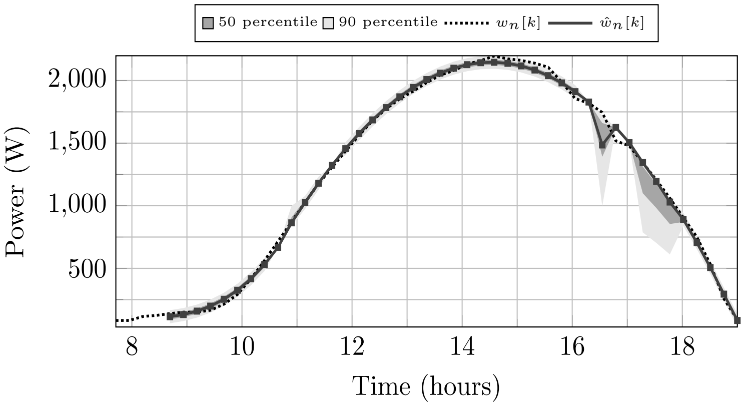


Figure 5.5.3: Plot of actual and predicted value with one-step prediction for day that is *sunny* day

unless there is some additional information in the form of cloud motion information or accurate weather forecasts. To summarize, prediction during *partly cloudy* conditions is prone to larger errors than during *overcast* or *sunny*. This is in accordance with the associated uncertainty in solar PV power for those periods.

5.5.1.3 Comparison With Reference Methods

Probabilistic smart persistence: It is assumed that the mean of predictive density for the next m steps is given as the continued fraction of clear sky component at the

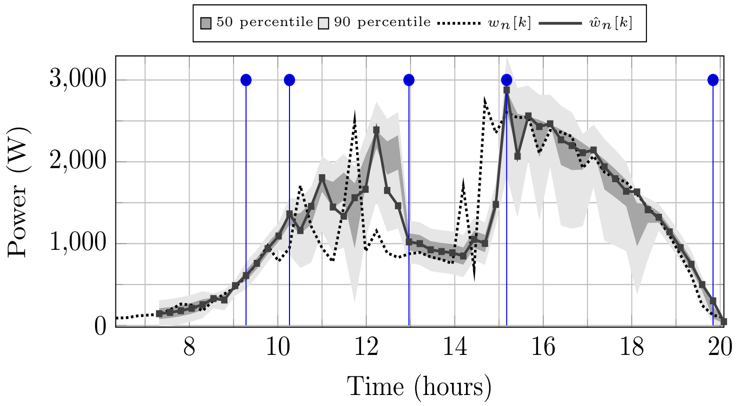


Figure 5.5.4: Plot of actual and predicted value with one-step prediction for day with variety of weather conditions. Time instants at which algorithm detected regime change is shown in blue.

current time step,

$$\hat{w}_d^{per}[n+m] = s_d[n+m] \left(w_d[n] \Big/ s_d[n] \right), m = 1, \ldots, \chi \tag{5.5.8}$$

Variance is given by the sample variance of the power values in $\boldsymbol{w}_n^{-M_1}$. The distribution is assumed to be Gaussian.

AR model: The power is expressed as the sum of a clear sky component and a stochastic component, and it is assumed that the stochastic component follows an autoregressive model:

$$\begin{aligned} w_d^{AR}[n+m] &= s_d[n+m] - x_d^{AR}[n+m], \\ x_d^{AR}[n+m] &= \sum_{i=1}^{M_{AR}} a[i] x_d^{AR}[n+m-i] + \epsilon_{AR}[n+m] \end{aligned} \tag{5.5.9}$$

AR model with regime switching: It is assumed that each of the classes *sunny, partly cloudy* and *overcast* have stochastic components with different coefficients for the AR model:

$$x_d^{AR}[n+m] = \tag{5.5.10}$$
$$\begin{cases} \sum_{i=1}^{M_{AR_s}} a_s[i]x_d^{AR}[n+m-i] + \epsilon_{AR_s}[n+m], \textit{sunny} \\ \sum_{i=1}^{M_{AR_{pc}}} a_{pc}[i]x_d^{AR}[n+m-i] + \epsilon_{AR_{pc}}[n+m], \textit{partly cloudy} \\ \sum_{i=1}^{M_{AR_{oc}}} a_{oc}[i]x_d^{AR}[n+m-i] + \epsilon_{AR_{oc}}[n+m], \;\textit{overcast} \end{cases}$$

Classification into different regimes for the AR model is done using the proposed approach in subsection 5.3.1. In essence, it is the closest method to the proposed one by design.

In both the AR model and the regime switching AR models, the point forecast value is the mean and the variance of Gaussian noise, ϵ_{AR} is estimated along with the coefficients.

Quantile regression (QR) Linear quantile regression (QR) [188] is used. b^{th} quantile of solar power, $Q_m(b)$ for predictive time step m, where $b = 0.01, 0.02, \ldots 0.99$ is expressed as:

$$Q_m(b) = x^{QR} s_d[n+m], x^{QR} = \boldsymbol{\beta}_{m,b}^T \boldsymbol{x}, m = 1, \ldots, \chi \tag{5.5.11}$$

$$\lfloor \boldsymbol{x} \rfloor_i = w_d[n - M_1 + i] \Big/ s_d[n - M_1 + i], i = 1, \ldots M_1 \tag{5.5.12}$$

where $\boldsymbol{\beta}_{m,b}$ are vectors of parameters that were estimated using the procedure in [188].

5.5.1.4 Results: Probabilistic Forecast Of Solar Power

All the simulations were performed using one year of training data and one year of testing data for validation. Most of the training for estimation of parameters of HMM is done apriori making computational time of the proposed method very short since

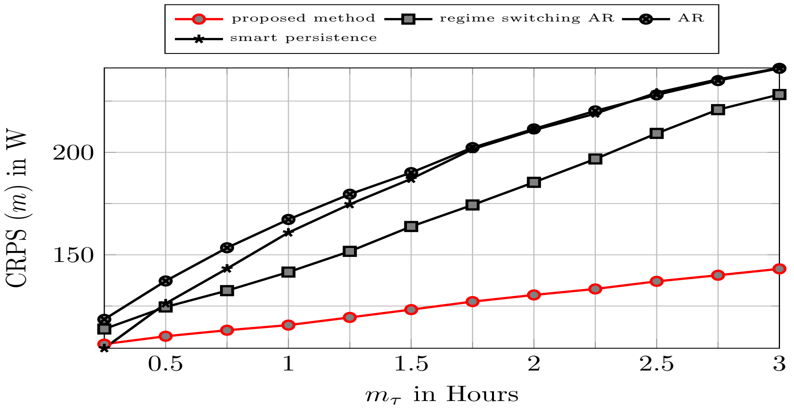


Figure 5.5.5: CRPS for m-step prediction using proposed model and other reference models.

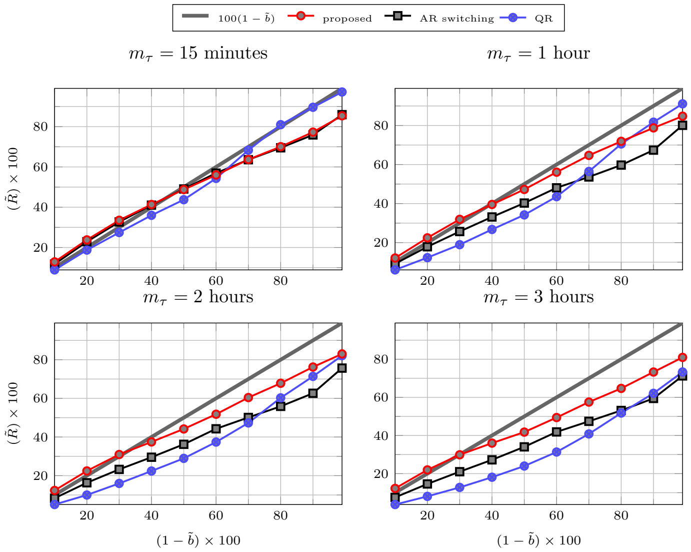


Figure 5.5.6: Reliability prediction horizons using proposed model and few better performing reference models.

the Viterbi algorithm, which is efficient [183] was used. The computational time specifically depends on the acquisition time of samples in a real-time setting.

Comparison with CRPS as the metric is shown in Fig.5.5.5. The AR models beat the smart persistence. For the QR method, CRPS cannot be calculated since the predictive density is not continuous. CRPS is inversely proportional to the parameters[2] [189] which makes larger values of parameters better.

---

[2] Closed form expressions for CRPS in (5.5.1) can be computed. For Gaussian densities, it is

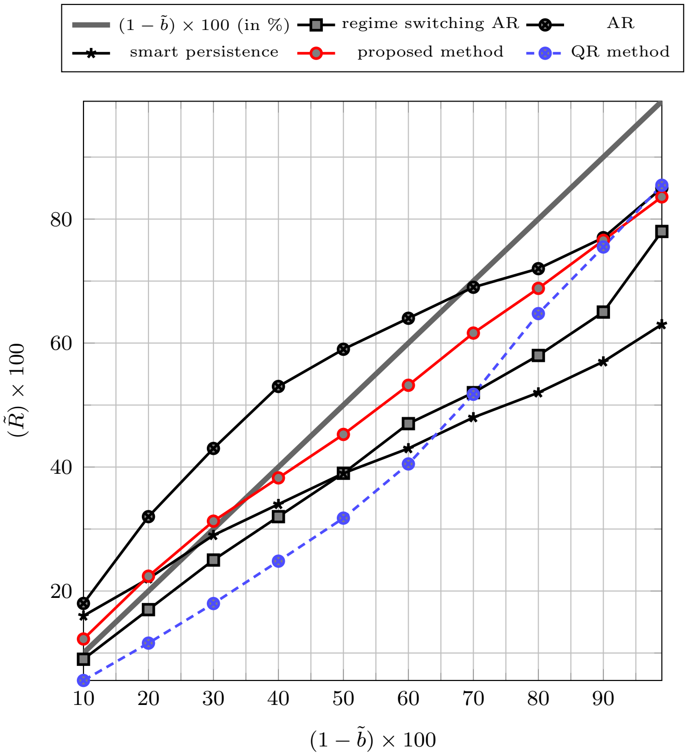


Figure 5.5.7: Average reliability (over all prediction horizons) using proposed model and other reference models.

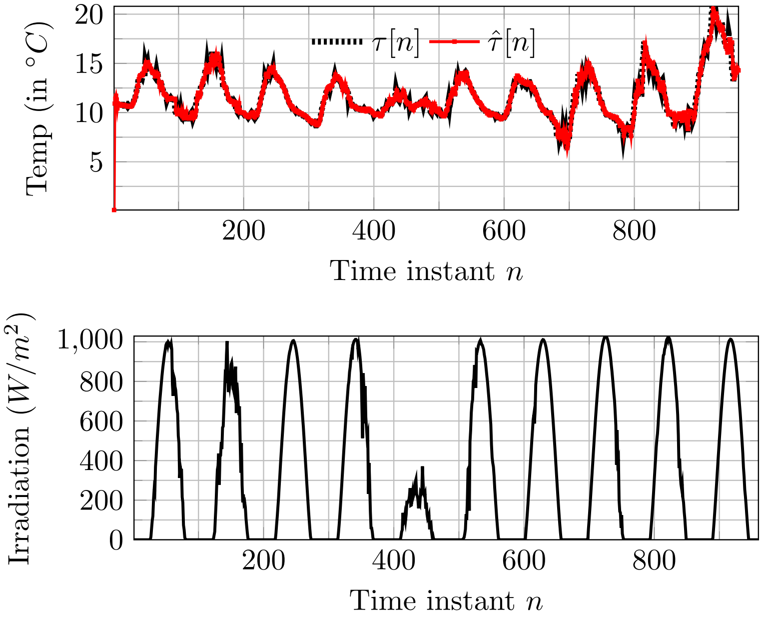


Figure 5.5.8: Estimate of temperature, $\hat{\tau}[n]$ and corresponding irradiance $w[n]$ for May $2-12, 2014$ at NOAA weather station Merced, CA.

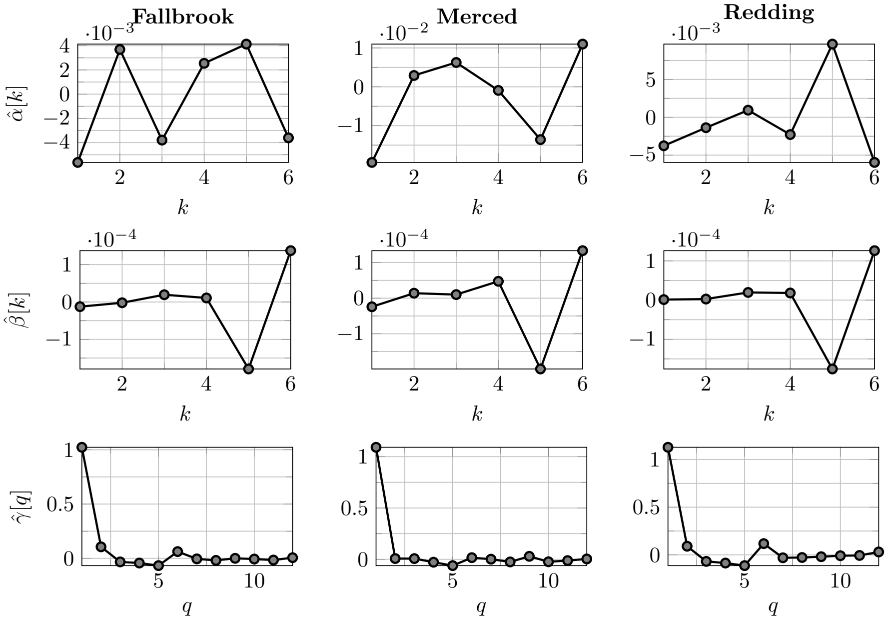


Figure 5.5.9: Estimates, $\hat{\alpha}[k], \hat{\beta}[k], \hat{\gamma}[q]$ for 3 NOAA weather stations in California.

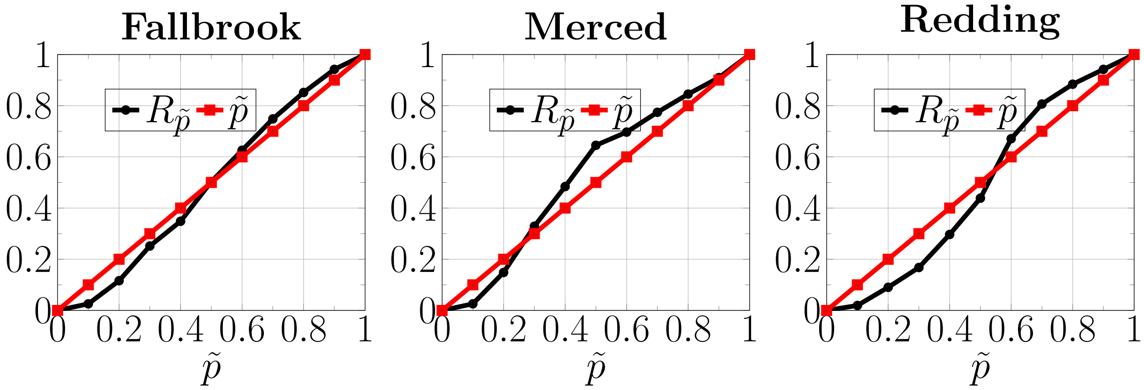


Figure 5.5.10: Reliability $R_{\tilde{p}}$ of joint probabilistic forecasts: 2 steps ahead for Fallbrook, Merced and Redding, CA.

Fig. 5.5.6 shows the reliability metric for prediction horizons $m_\tau = 0.25, 1, 2, 3$ hours only for certain better performing methods whereas Fig.5.5.7 shows the average reliability metric, R_b^{avg} and comparison with the curve $1 - b$ for all benchmarks. Both figures are indicative of the probability of the actual occurrences of power being well covered by the predicted intervals. The closer the metric to the curve $1 - b$, the more reliable the method. From Fig. 5.5.6, it is seen that at $m_\tau = 0.25$, quantile regression performs the best. However, as the prediction horizon increases, the performance of QR method deteriorates. This can be attributed to fewer samples of data available for training as the prediction horizon increases. AR method with regime-switching performs the closest to the proposed method. It even matches the proposed method for lower quantiles across all prediction horizons but does not fare well in higher quantiles. The key takeaway is that the proposed method consistently (in all prediction horizons) does better than the benchmark methods considered. The trend of under performance in higher quantiles by the proposed method can be attributed to the choice of the values of $\lambda_z, \lambda_a^b, \lambda_a^e$ and to classification error in the regime i.e. error due to assumption of persistence in regime of *sunny, partly cloudy, overcast*.

Using the same set of parameters, Table.5.1 shows the average interval score, $\text{Score}_b^{\text{avg}}$ normalized by the nameplate capacity. In terms of interval score, the proposed method outperforms all the other methods considered. This is indicative of the fact that the proposed method of forecasting is sharp [185].

5.5.2 Joint Probabilistic Forecast

Experiments were run on a dataset of solar irradiance (in W/m^2) and temperature (in $^\circ C$) obtained from 3 NOAA weather stations: Fallbrook, Redding and Merced

---

inversely proportional to σ. For exponential densities, it is inversely proportional to λ.

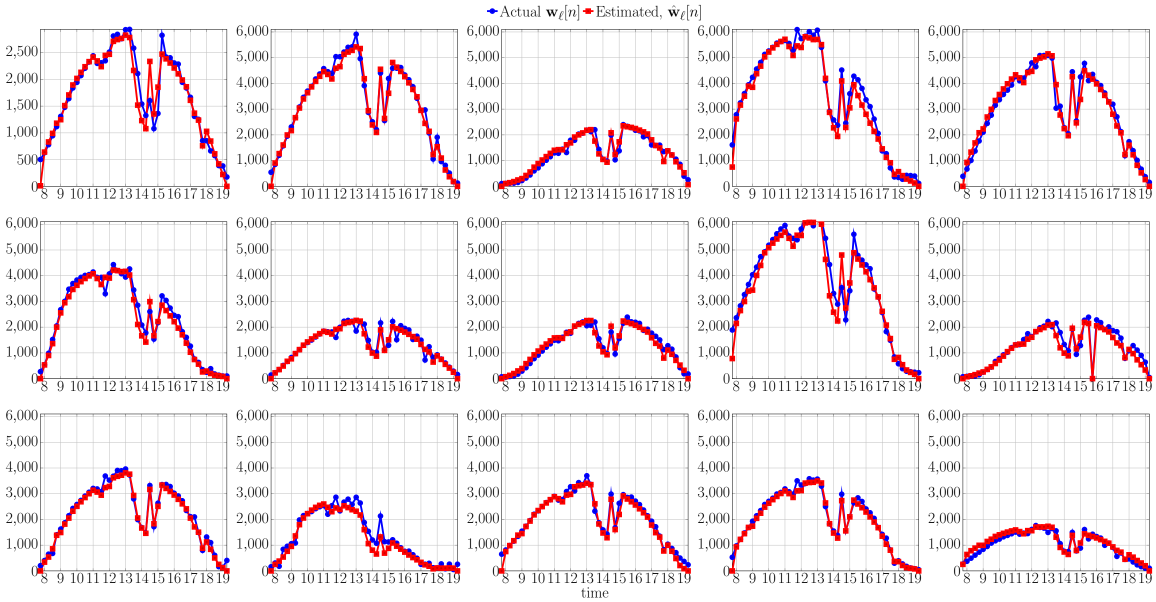


Figure 5.5.11: Fit of the Model for ZIP 94531, Antioch, CA.

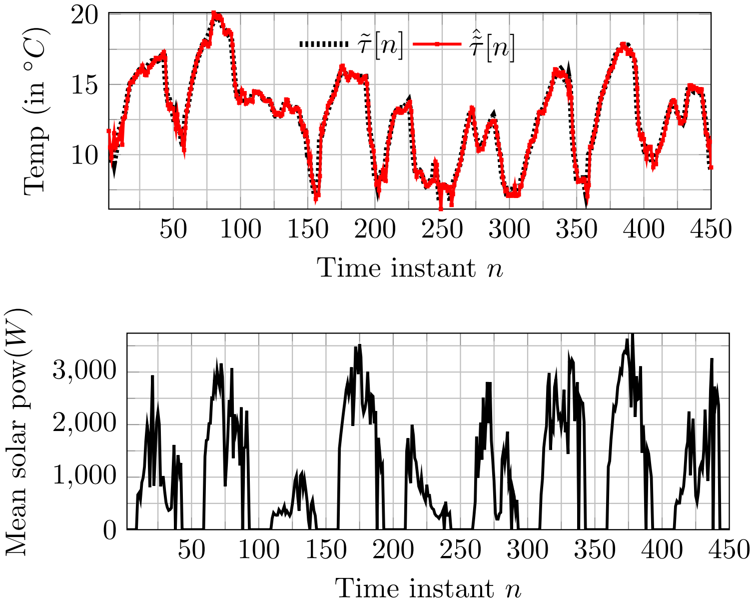


Figure 5.5.12: Estimate of temperature, $\hat{\tilde{\tau}}[n]$ and corresponding average solar power for $L = 15$ locations, $w[n] = L^{-1} \sum_{\hat{\ell}=1}^{L} w_d^{\hat{\ell}}[n]$ for March 26-April 3, 2014 in ZIP 94531, Antioch, CA.

Table 5.1: Table with Average Score Normalized by the Nameplate Capacity

$1-b$	**Proposed**	**QR**	**AR (switching)**	**Persistence**	**AR**
0.1	-0.214	-0.216	-0.229	-0.242	-0.223
0.2	-0.202	-0.209	-0.218	-0.232	-0.211
0.3	-0.196	-0.204	-0.212	-0.228	-0.208
0.4	-0.195	-0.202	-0.212	-0.231	-0.212
0.5	-0.199	-0.204	-0.218	-0.24	-0.224
0.6	-0.210	-0.215	-0.23	-0.256	-0.245
0.7	-0.232	-0.253	-0.25	-0.282	-0.276
0.8	-0.268	-0.349	-0.281	-0.32	-0.322
0.9	-0.334	-0.511	-0.331	-0.383	-0.394
0.99	-0.482	-0.994	-0.455	-0.503	-0.566

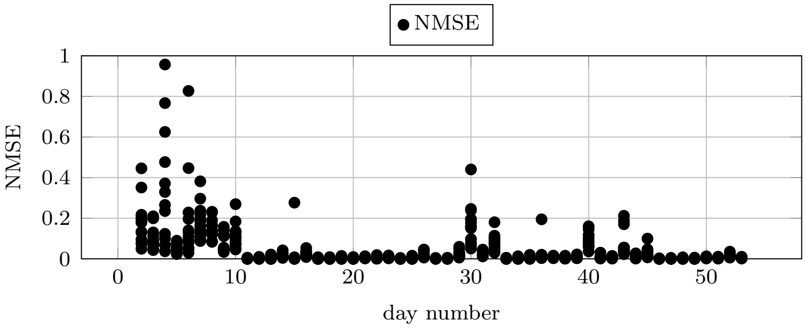


Figure 5.5.13: NMSE for 15 Locations in ZIP 94531, Antioch, CA

in California (CA). Data is sampled at 15-minute intervals. For estimating the parameters of temperature model from (5.1.14), $\alpha[k], \beta[k], \gamma[q]$, it was assumed that $M_2 = 12$ i.e. 3 hours and $M_1 = 6$ i.e. 1.5 hour. Temperature and irradiance data for 500 contiguous days were used for estimation of the parameter μ of the conditional distribution in (5.4.1).

As an example, the fit of the model to the temperature data along with the corresponding irradiance in Merced, CA is highlighted in Fig. 5.5.8 for a week in

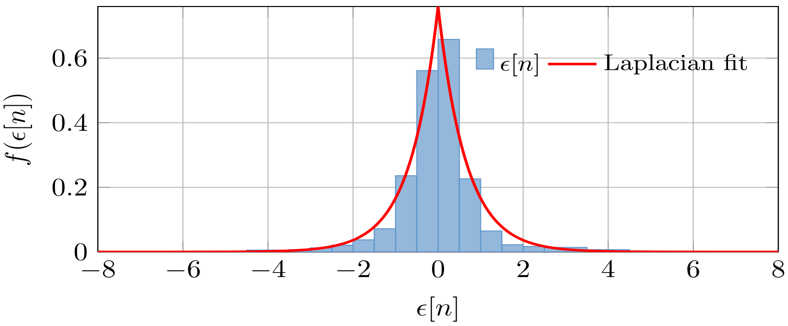


Figure 5.5.14: Laplacian Fit for $\epsilon[n]$ with $\mu = 1.5196$

May 2014. In Fig.5.5.9, the estimated parameters $\hat{\alpha}[k], \hat{\beta}[k], \hat{\gamma}[q]$ for all the weather stations are plotted.

To highlight the efficacy of the probabilistic forecasts, $m = 2$ step ahead forecast, i.e. 30 minutes ahead, was performed. The joint probability density takes the form of (5.4.5) with $m = 2$. This gives a 4-D probability distribution (2-D for temperature and 2-D for irradiance). The accuracy of the probabilistic forecast is evaluated in two ways: Continuous ranked probability score (CRPS) for one observation,

$$\begin{aligned}\text{CRPS}(\boldsymbol{w}_n^{+m}, \boldsymbol{\tau}_n^{+m}) = \int\limits_{\tilde{\boldsymbol{w}}_n^{+m}, \tilde{\boldsymbol{\tau}}_n^{+m}} \Bigg[F\left(\tilde{\boldsymbol{w}}_n^{+m}, \tilde{\boldsymbol{\tau}}_n^{+m} \middle| \boldsymbol{w}_d^{-M_1}, \boldsymbol{\tau}_n^{-M_2} \right) + \\ - u\left(\tilde{\boldsymbol{w}}_n^{+m} - \boldsymbol{w}_n^{+m}\right) u\left(\tilde{\boldsymbol{\tau}}_n^{+m} - \boldsymbol{\tau}_n^{+m}\right)\Bigg]^2 d\tilde{\boldsymbol{w}}_n^{+m} d\tilde{\boldsymbol{\tau}}_n^{+m}\end{aligned} \quad (5.5.13)$$

where $(\boldsymbol{w}_n^{+m}, \boldsymbol{\tau}_n^{+m})$ are the observed values and $F\left(\cdot\middle|\cdot\right)$ is the joint CDF (from (5.4.5)) and $u(\boldsymbol{a})$ is the Heaviside function acting on a vector. CRPS in the point forecast (single value) case reduces to mean absolute error (MAE) [185] after averaging over all observed samples $\boldsymbol{w}_n^{+m}, \boldsymbol{\tau}_n^{+m}$. Thus, the units of CRPS becomes $(W/m^2)^2(°C)^2$. Forecasting experiments for 30 days and for 15 minute time-intervals within each day were performed. To evaluate the metric of reliability for the joint distribution, let a probability interval (PI), $I_{\tilde{p}}$, cover the observed values $\xi = \{\boldsymbol{w}_n^{+m}, \boldsymbol{\tau}_n^{+m}\}, \xi \in \mathcal{P}$ with probability $\tilde{p}$. Then reliability $R_{\tilde{p}}$ is calculated as,

$$R_{\tilde{p}} = |\mathcal{P}|^{-1} \textstyle\sum_{\xi \in \mathcal{P}} \mathbb{I}_{\left(\xi \in I_{\tilde{p}}\right)} \tag{5.5.14}$$

where $\mathbb{I}$ is an indicator function with value 1 if the observed sample belongs to the probability interval obtained using the joint CDF. Fig.5.5.10 shows the reliability of prediction using the proposed forecasting technique for all three weather stations. Ideally, it should be close to the diagonal line $\tilde{p}$. The reliability of the proposed method closely matches the line as can be seen from the figure. Slight variation across different locations is observed. Broadly, it is demonstrative of the fact that the predictive distribution (CDF) provided is close to the observed CDF of real data. CRPS values are reported in the table 5.2.

Table 5.2: CRPS for Joint Forecast (in $(W/m^2)^2(°C)^2)$

Fallbrook	**Merced**	**Redding**
489.5	440.02	443.80

5.5.3 Spatio-temporal Model for Solar Power and Temperature

To illustrate the efficacy of the modeling approach for solar power and temperature from 15 installations within a ZIP code, data from SolarCity is used. Solar PV power from rooftop installations in Antioch, CA, with ZIP code 94531 is used along with hourly temperature information from the local weather station Concord Buchanan Field, CA available online [3]. The fit of the proposed model to solar power data for a single day for all locations is highlighted in Fig. 5.5.11. Fig. 5.5.12 shows the average solar power and the corresponding temperature. The corresponding normalized root-

---

[3] `https://www.ncdc.noaa.gov/cdo-web/datasets/LCD/stations/WBAN:23254/detail`

mean squared error (NMSE)

$$\mathrm{NMSE}_d^\ell = \sum_k (w_d^\ell[k] - \hat{w}_d^\ell[k])^2 \Bigg/ \sum_k (w_d[k])^2 \tag{5.5.15}$$

for each day across all locations is plotted in 5.5.13. For cloudy days in the beginning of the time-series, higher NMSE values are reported due to uncoordinated spikes among different locations when local effects are present. For temperature, the distribution of error is plotted in Fig. 5.5.14.

To evaluate the joint probabilistic forecasts, the CRPS measure is used with an alternative definition [189] for a multi-dimensional quantity. Let $\boldsymbol{X}$ be the multi-dimensional random variable,

$$\boldsymbol{X}^T \triangleq \left[\left(\boldsymbol{w}_{n,1}^{+m}\right)^T \quad \left(\boldsymbol{w}_{n,2}^{+m}\right)^T \quad \cdots \quad \left(\boldsymbol{w}_{L,1}^{+m}\right)^T \quad \left(\tilde{\boldsymbol{\tau}}_n^{+m}\right)^T \right]$$

with the distribution $f \triangleq f\left(\boldsymbol{X} \,\middle|\, \left\{\boldsymbol{w}_{n,\hat{\ell}}^{-M_1}\right\}_{\hat{\ell}=1}^{L}, \tilde{\boldsymbol{\tau}}_n^{-M_2} \right)$ and $\boldsymbol{X}'$ be also drawn from the same distribution. Let $\boldsymbol{y}$ correspond to a certain realization or the observed values of power and temperature. Then,

$$\mathrm{CRPS}\left(\left\{\boldsymbol{w}_{n,\ell}^{+m}\right\}_{\ell=1}^{L}, \tilde{\boldsymbol{\tau}}_n^{+m} \right) = \mathbb{E}_f \left\| \boldsymbol{X} - \boldsymbol{y} \right\| - \frac{1}{2} \mathbb{E}_f \left\| \boldsymbol{X} - \boldsymbol{X}' \right\|$$

This formulation is used in order to avoid computing the CDF since it is computationally intensive given the dimension of the distribution (16-D) even for a one-step ahead forecast. For the same reason of computational complexity, reliability measure is not calculated for the spatio-temporal solar power and temperature probabilistic forecast. The CRPS for one-step ahead forecast was found to be 5785 $(W/m^2)^{15}(^\circ C)$ for the entire ZIP code.

Chapter 6

FAULT DETECTION IN SOLAR PANELS

In this chapter, a distributed array processing algorithm to analyze the power output of solar photo-voltaic (PV) installations is presented while leveraging the low-rank structure inherent in the data to estimate possible faults. The proposed multi-agent algorithm requires near-neighbor communications only and is also capable of jointly estimating the common low rank cloud profile and local shading of panels. To illustrate the workings of the algorithm, experiments to detect shading faults in solar PV installations within a single ZIP code are performed. Additionally, a Bayesian lower bound on the shading parameter's mean squared estimation error (MSE) is also derived. The results are promising and show that one can successfully estimate the fraction of partial shading in solar installations that can usually go unnoticed. The problem can be thought of as an instance of decentralized dictionary learning [190] where the sparse coding step is replaced with a maximum aposteriori (MAP) estimation of shading parameter. The Bayesian framework makes it also possible to provide performance benchmarks. For this purpose an extension of Bayesian C̀ramer Rao lower bounds (BCRB) is derived for the partial shading parameter mean squared estimation error that, unlike the ones used in the context of Bayesian learning [191,192], applies to parameters with a finite support prior distribution that may not satisfy certain regularity conditions [193].

The chapter first introduces the distributed algorithm more generally in Section 6.1 and then presents the application of the introduced framework for partial shading estimation in solar PV panels in Section 6.2. Solar data from multiple residential

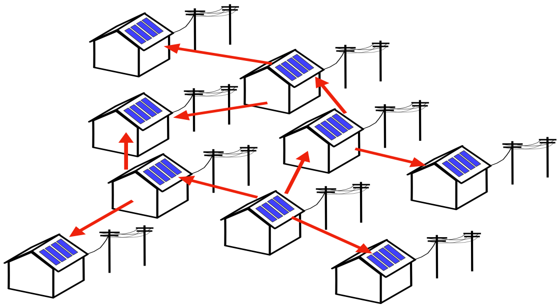


Figure 6.0.1: Schematic showing the communication links between different solar panel installations for distributed estimation.

rooftop PV panel installations within a ZIP code is used to test the efficacy of the proposed algorithm and results are presented in Section 6.3.

6.1 System Model

Let a discrete set of data at time indices $n = 1, 2, \ldots N$ from locations $\ell = 1, 2, \ldots L$ be available. Stacking them column-wise matrix, $\boldsymbol{Y} \in \mathbb{R}^{N \times L}$ is obtained. It is postulated that $\boldsymbol{Y}$ has a specific low-rank structure such that each column of the matrix is a scaled version of the other by a certain attenuation coefficient, thereby giving the following model:

$$\boldsymbol{Y} = \boldsymbol{x}(\boldsymbol{1} - \boldsymbol{\alpha})^T + \begin{bmatrix} \boldsymbol{\eta}_1 & \boldsymbol{\eta}_2 & \cdots & \boldsymbol{\eta}_L \end{bmatrix}, \boldsymbol{\eta}_\ell \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{6.1.1}$$

where $\boldsymbol{x} \in \mathbb{R}^N$ and belongs to a convex constraint set $\mathcal{C}$, $\boldsymbol{1}$ is a vector of ones with dimension L and $\boldsymbol{\alpha} \in \mathbb{R}^{L \times 1}$. It is assumed that noise covariance matrix $\boldsymbol{\Sigma}$ is diagonal and the scaling vector $\boldsymbol{\alpha}$ is sparse with the support of non-zero entries being finite. The goal is to estimate $\boldsymbol{x}, \boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}$ using $\boldsymbol{Y}$. The problem statement is,

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\Sigma}} \quad & \tfrac{1}{2} \textstyle\sum_{\ell=1}^{L} \left\| \boldsymbol{\Sigma}^{-\frac{1}{2}} \left[\boldsymbol{y}_\ell - (1-\alpha_\ell)\boldsymbol{x} \right] \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha} \right\|_1 \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{C}, \ \underline{\alpha} \leq \alpha_\ell \leq \overline{\alpha}, \ \ell = 1, 2, \ldots L \end{aligned} \tag{6.1.2}$$

In our setup, there is no central entity to estimate $\boldsymbol{x}$ and $\boldsymbol{\Sigma}$ so the agents have to do it in a decentralized fashion. From (6.1.2), the problem can be treated as a variant of *decentralized matrix factorization* problem where the rank of the matrix is exactly 1.

6.1.1 Solution Approach

An alternating minimization method is used since the cost is non-convex. Firstly, given $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$, maximum likelihood (ML) estimate $\hat{\boldsymbol{x}}_{ML}$, is computed in a distributed manner. Then, given $\hat{\boldsymbol{x}}_{ML}$ and $\boldsymbol{\Sigma}$, α_ℓ is updated at location ℓ with the maximum aposteriori (MAP) estimate of α_ℓ by imposing as a prior a truncated exponential distribution, with support between $\overline{\alpha}$ and $\underline{\alpha}$:

$$p(\alpha_\ell)=C^{-1}\lambda e^{-\lambda\alpha_\ell},\ C = e^{-\lambda\underline{\alpha}} - e^{-\lambda\overline{\alpha}},\ \underline{\alpha} \leq \alpha_\ell \leq \overline{\alpha}. \tag{6.1.3}$$

Lastly, given $\boldsymbol{x}$ and $\boldsymbol{\alpha}$, covariance matrix $\boldsymbol{\Sigma}$ is updated using residual error and average consensus based method.

Decentralized ML estimate : Let the network of sensors be described by the graph $G = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of vertices and $\mathcal{E}$ the set of edges/connections. Then, cost for each agent at ℓ is,

$$f_\ell(\boldsymbol{x}, \alpha_\ell) = \frac{1}{2}\left(\boldsymbol{y}_\ell - (1-\alpha_\ell)\boldsymbol{x}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}_\ell - (1-\alpha_\ell)\boldsymbol{x}\right)$$

and total cost is $F(\boldsymbol{x}, \boldsymbol{\alpha}) = \frac{1}{2}\sum_{\ell=1}^{L} f_\ell(\boldsymbol{x}, \alpha_\ell)$. Decentralized Frank-Wolfe [194] method is used(note that any distributed optimization algorithm is viable) to estimate $\boldsymbol{x}$. Assuming that the values of α_ℓ and $\boldsymbol{\Sigma}$ are fixed for this step, average consensus (AC) is employed to update the estimate of vector $\boldsymbol{x}$ as well as the gradient, $\sum_{\ell=1}^{L} \nabla_{\boldsymbol{x}} f_\ell(\boldsymbol{x}, \alpha_\ell)$. For AC, the Metropolis-Hasting update is used which provides the following weight matrix:

$$\boldsymbol{W} = \mathbb{I} - \boldsymbol{M}\mathbf{1} + \boldsymbol{M},\, [\boldsymbol{M}]_{ij} = \begin{cases} \left(\max\{d_i, d_j\}\right)^{-1}, (i,j) \in \mathcal{E} \\ 0, (i,j) \notin \mathcal{E} \end{cases}$$

where d_i, d_j are degrees of nodes i and j in the graph G and $\mathbb{I}$ is the identity matrix of size L. The steps of the decentralized Frank-Wolfe method are as follows:

1. Begin with an estimate of $\boldsymbol{x}_\ell(0)$ and $\alpha_\ell(0)$ at each ℓ . Let

$$\boldsymbol{X}(0) \triangleq \begin{bmatrix} \boldsymbol{x}_1(0) & \boldsymbol{x}_2(0) & \dots & \boldsymbol{x}_L(0) \end{bmatrix} \in \mathbb{R}^{N \times L} \tag{6.1.4}$$

2. A step of AC is carried out, $\boldsymbol{X}(t) = \boldsymbol{X}(t-1)\boldsymbol{W}$.

3. At sensor ℓ, the gradient is calculated:

$$\nabla_{\boldsymbol{x}} f_\ell(\boldsymbol{x}_\ell(t), \alpha_\ell) = (1-\alpha_\ell)\boldsymbol{\Sigma}^{-1} \left(\boldsymbol{y}_\ell - (1-\alpha_\ell)\boldsymbol{x}_\ell(t)\right).$$

   The full gradient at time t, $\nabla_{\boldsymbol{x}} F(\boldsymbol{x}(t), \boldsymbol{\alpha})$, is unavailable. Hence the local estimate of the gradient at ℓ is given by

$$\begin{aligned}\nabla_{\boldsymbol{x}} F_\ell^t(\boldsymbol{x}_\ell(t)) = \nabla_{\boldsymbol{x}} F_\ell^{t-1}(\boldsymbol{x}_\ell(t-1)) + \nabla_{\boldsymbol{x}} f_\ell(\boldsymbol{x}_\ell(t), \alpha_\ell) \\ - \nabla_{\boldsymbol{x}} f_\ell(\boldsymbol{x}_\ell(t-1), \alpha_\ell).\end{aligned} \tag{6.1.5}$$

   Then, the AC step is employed,

$$\nabla_{\boldsymbol{x}} F_\ell^{t+1}(\boldsymbol{x}_\ell(t)) = \textstyle\sum_j [\boldsymbol{W}]_{\ell,j} \nabla_{\boldsymbol{x}} F_j^t(\boldsymbol{x}_j(t)) \tag{6.1.6}$$

4. To perform the constrained optimization, Frank-Wolfe (FW) update [194] is used which involves projection to the constraint space as a linear optimization problem. General FW step is:

$$\boldsymbol{x}_\ell(t+1) \leftarrow (1-\tau_t)\boldsymbol{x}_\ell(t) + \tau_t \boldsymbol{b}_t^\ell \tag{6.1.7}$$

$$\text{where} \quad \boldsymbol{b}_t^\ell := \arg\min_{\boldsymbol{b} \in \mathcal{C}} \langle \boldsymbol{b}, \nabla_{\boldsymbol{x}} F_\ell^{t+1}(\boldsymbol{x}_\ell(t)) \rangle \tag{6.1.8}$$

where $\tau_t = 2/(t+1)$ is the step size. From [194], it is known that the decentralized optimization problem to estimate $\boldsymbol{x}$ converges to a stationary point since the original problem is non-convex (bi-convex) with respect to both $\boldsymbol{x}$ and $\boldsymbol{\alpha}$.

Updating α: Consider the second step of alternating minimization, i.e. that of estimating α_ℓ. From (6.1.1) (dropping time index t),

$$\boldsymbol{y}_\ell = (1-\alpha_\ell)\boldsymbol{x}_\ell + \boldsymbol{\eta}_\ell,\ \boldsymbol{\eta}_\ell \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}). \tag{6.1.9}$$

Introduce a modified scalar measurement, $\tilde{y}_\ell$, by multiplying $\boldsymbol{x}_\ell^T\boldsymbol{\Sigma}^{-1}$ on both sides of (6.1.9) as $\tilde{y}_\ell \triangleq 1 - \left(\boldsymbol{x}_\ell^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell \middle/ \boldsymbol{x}_\ell^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_\ell\right)$. Then,

$$\tilde{y}_\ell = \alpha_\ell + \tilde{\eta}_\ell,\ \tilde{\eta}_\ell \sim \mathcal{N}\left(0, \gamma_\ell^2\right),\ \gamma_\ell^2 = \left(\boldsymbol{x}_\ell^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_\ell\right)^{-1}. \tag{6.1.10}$$

With prior on α_ℓ as in (6.1.3), MAP estimate of α_ℓ at iteration t is

$$\hat{\alpha}_\ell(t) = \begin{cases} \underline{\alpha}, & \tilde{y}_\ell \le \underline{\alpha} + \lambda\gamma_\ell^2 \\ \tilde{y}_\ell - \lambda\gamma_\ell^2, & \underline{\alpha} + \lambda\gamma_\ell^2 \le \tilde{y}_\ell \le \overline{\alpha} + \lambda\gamma_\ell^2 \\ \overline{\alpha}, & \tilde{y}_\ell \ge \overline{\alpha} + \lambda\gamma_\ell^2 \end{cases} \tag{6.1.11}$$

Estimation of $\boldsymbol{\Sigma}$: It is assumed that covariance matrix $\boldsymbol{\Sigma}$ is diagonal:

$$\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\sigma})\mathbb{I},\ i = 1, 2, \ldots, N \tag{6.1.12}$$

Then, ML estimate of $[\sigma]_i$ at time t is,

$$[\sigma]_i(t) = L^{-1}\textstyle\sum_{\ell=1}^{L}\left([\boldsymbol{y}_\ell]_i - \alpha_\ell\left[\boldsymbol{x}_\ell(t)\right]_i\right)^2. \tag{6.1.13}$$

AC can be employed on the N variances by letting

$$[\boldsymbol{\sigma}_\ell]_i(t) = \left([\boldsymbol{y}_\ell]_i - \alpha_\ell\left[\boldsymbol{x}_\ell(t)\right]_i\right)^2,\ \ell = 1, 2, \ldots L \tag{6.1.14}$$

One iteration of AC is $\boldsymbol{Q}(t+1) = \boldsymbol{Q}(t)\boldsymbol{W}$ where

$$\boldsymbol{Q}(t) \triangleq \begin{bmatrix} \boldsymbol{\sigma}_1(t) & \boldsymbol{\sigma}_2(t) & \ldots & \boldsymbol{\sigma}_L(t) \end{bmatrix}. \tag{6.1.15}$$

Algorithm 6.1 describes the steps of the proposed distributed Bayesian estimation method. Next, a lower bound for estimation error of α_ℓ is presented given $\boldsymbol{x}$ and $\boldsymbol{\Sigma}$.

Theorem 1 (Bayesian CRB for finite support prior)**.** *A lower bound for the mean squared error of the MAP estimator in* (6.1.11) *given* $\boldsymbol{x}$ *and* $\boldsymbol{\Sigma}$ *is*

$$\mathbb{E}_{\alpha_\ell,\tilde{y}_\ell}\left[(\hat{\alpha}_\ell-\alpha_\ell)^2\middle|\boldsymbol{x},\boldsymbol{\Sigma}\right]\geq\frac{\gamma_\ell^2(1-\nu)^2}{\lambda^2\gamma_\ell^2+1} \tag{6.1.16}$$

$$\begin{aligned}\textit{where } \nu =&(p(\overline{\alpha})+p(\underline{\alpha}))\left[\gamma_\ell\phi(\lambda\gamma_\ell)+\lambda\gamma_\ell^2\Phi(\lambda\gamma_\ell)\right]+Rp(\underline{\alpha})\\&-p(\overline{\alpha})\left[\gamma_\ell\phi\left(\lambda\gamma_\ell-\frac{R}{\gamma_\ell}\right)+(\lambda\gamma_\ell^2-R)\Phi\left(\lambda\gamma_\ell-\frac{R}{\gamma_\ell}\right)\right]\\&-p(\underline{\alpha})\left[\gamma_\ell\phi\left(\lambda\gamma_\ell+\frac{R}{\gamma_\ell}\right)+(\lambda\gamma_\ell^2+R)\Phi\left(\lambda\gamma_\ell+\frac{R}{\gamma_\ell}\right)\right],\\\Phi(x)=&\int_{-\infty}^{x}e^{-\frac{z^2}{2}}dz,\ \phi(x)=(\sqrt{2\pi})^{-1}e^{-x^2/2},\ R=\overline{\alpha}-\underline{\alpha}.\end{aligned}$$

Proof given in the Appendix A. Since the bound applies to a case when $\boldsymbol{x}$ and $\boldsymbol{\Sigma}$ are known exactly, it is not likely to be tight when the estimation error in $\boldsymbol{x}$ and $\boldsymbol{\Sigma}$ are significant. This is seen in the numerical results.

6.2 Partial Shading Estimation in Solar Panels

The system model described in Section 6.1 can be applied to detect partial shading type of faults.

Model for small utility-scale farm: If the area covered by the panels in a farm is relatively small, one can assume that the cloud cover induced attenuation of solar power is the same across each panel j at a given time instant n on day d. Assuming that all the panels have uniform characteristics, the solar power data matrix $\boldsymbol{Y}_d\in\mathbb{R}^{N\times P}$ constructed using power measurements from panels $j=1,\ldots,P$ for times $n=1,2,\ldots N$ on a day d is

$$\boldsymbol{Y}_d=\boldsymbol{w}_d\times(\boldsymbol{1}-\boldsymbol{\alpha}_d)^T,\quad \boldsymbol{w}_d\in\mathbb{R}^N,\ \boldsymbol{\alpha}_d\in\mathbb{R}^P \tag{6.2.1}$$

Algorithm 6.1: Distributed Bayesian Estimation

```
Input: $\boldsymbol{x}_\ell(0), \alpha_\ell(0), \boldsymbol{\sigma}_\ell(0) \;\; \forall \ell, T, M$ and $\boldsymbol{W}$. ;
    while $iter \leq M$ do
        for $t = 0 : T - 1$ do
            $\boldsymbol{X}(t+1) \leftarrow \boldsymbol{X}(t)\boldsymbol{W}$ ;
            $\boldsymbol{x}_\ell(t+1) \leftarrow (1-\tau_t)\boldsymbol{x}_\ell(t+1) + \tau_t \boldsymbol{b}_t^\ell \;\forall \ell$ (6.1.8) ;
            Update $\hat{\alpha}_\ell(t) \;\forall \ell$ from (6.1.11) ;
        for $k = 1 : K$ do
            $[\boldsymbol{\sigma}_\ell]_i(1) = \left([\boldsymbol{y}_\ell]_i - \alpha_\ell(T)\,[\boldsymbol{x}_\ell(T)]_i\right)^2, \;\ell \in \mathcal{L}_i$;
            $\boldsymbol{Q}(k+1) = \boldsymbol{Q}(k)\boldsymbol{W}$ ;
        $\boldsymbol{x}_\ell(0) \leftarrow \boldsymbol{x}_\ell(T)$, $\alpha_\ell(0) \leftarrow \alpha_\ell(T)$ , $\boldsymbol{\sigma}_\ell(0) \leftarrow \boldsymbol{\sigma}_\ell(K), \;\; \forall \ell$ ;
        iter = iter+1 ;
Output: $\boldsymbol{x}_\ell, \alpha_\ell, \boldsymbol{\sigma}_\ell \;\; \forall\, \ell$
```

where $\boldsymbol{w}_d$ is the vector of power measurements on day d. Note that with the use of irradiance sensors, it is possible to obtain approximately the common power measurement $\boldsymbol{w}_d$ by converting irradiance to power [166]. If the panels are operating normally (without shading faults) then $\boldsymbol{\alpha} = \mathbf{0}$. If not, the support of $\boldsymbol{\alpha}$ corresponds to panels which are affected by shading/soiling. This means that if one had access to granular information such as power measurement from each panel or a string of panels, it is possible to localize the fault upto the level of a single panel.

Model for installations in a ZIP code: To detect faults among multiple installations within a ZIP code using solar power data $w_{\ell,d}[n]$ at location $\ell = 1, 2, \ldots L$, time instant $n = 1, 2, \ldots N$ and day d, normalization by the sunny day solar power pattern, $s_{\ell,d}[n]$, is necessary to assume the low-rank structure. In general this pattern changes slowly with the season and can be estimated locally for day d [39]. In the absence of faults the data matrix $\boldsymbol{Y}_d \in \mathbb{R}^{N\times L}$ has entries $[\boldsymbol{Y}_d]_{n,\ell} = w_{\ell,d}[n] \Big/ s_{\ell,d}[n]$. where $x_{\ell,d}[n] \triangleq w_{\ell,d}[n] \Big/ s_{\ell,d}[n]$ is interpreted as the attenuation due to cloud cover. This attenuation is approximately the same within a ZIP code. Further evidence is provided in Fig. 6.2.1 that shows the singular values of the matrix $\boldsymbol{Y}_d$. Due to partial shading fault in the system, further attenuation of power is observed. It then becomes important to distinguish the cause of attenuation which could be due to cloud cover and/or partial shading. Let the common attenuation vector for a given day be denoted as $\boldsymbol{x}_d \in \mathbb{R}^N$ and fraction of shading at each location be $\alpha_{\ell,d}$. Then, the problem of partial shading detection can be modeled similar to (6.1.1) where the measured attenuation at location ℓ, $\boldsymbol{y}_{\ell,d} \in \mathbb{R}^N$ is the product of cloud cover induced attenuation and partial shading fault,

$$\boldsymbol{y}_{\ell,d} = (1-\alpha_{\ell,d})\boldsymbol{x}_d + \boldsymbol{\eta}_{\ell,d}, \quad \boldsymbol{Y}_d = \boldsymbol{x}_d \times (\mathbf{1} - \boldsymbol{\alpha}_d) \tag{6.2.2}$$

Thus, the problem of fault detection is to estimate the common attenuation vector $\boldsymbol{x}_d$ in order to determine $\boldsymbol{\alpha}_d \in \mathbb{R}^L$. Note that the model for $\boldsymbol{Y}_d$ is same as in (6.1.1). It is assumed that there are fewer locations with faulty panels than the total number of locations. This promotes sparsity in $\boldsymbol{\alpha}_d$. Also, attenuation by definition is less than 1. Therefore, $\mathbf{0} \preceq \boldsymbol{\alpha}_d \preceq \mathbf{1}$ and $\mathbf{0} \preceq \boldsymbol{x}_d \preceq \mathbf{1}$.

The estimation algorithm starts with random initialization for all parameters. The FW update for $\boldsymbol{x}_d$ is from (6.1.7) and $\boldsymbol{b}_t^\ell$ is

$$\left[\boldsymbol{b}_t^\ell\right]_i = \begin{cases} 1, & \left[\nabla_{\boldsymbol{x}} F_\ell^{t+1}(\boldsymbol{x}_{\ell,d}(t))\right]_i < 0 \\ 0, & \left[\nabla_{\boldsymbol{x}} F_\ell^{t+1}(\boldsymbol{x}_{\ell,d}(t))\right]_i \geq 0 \end{cases} . \tag{6.2.3}$$

The update to $\alpha_{\ell,d}$ is from (6.1.11) where $\underline{\alpha} = 0,\ \overline{\alpha} = 1$.

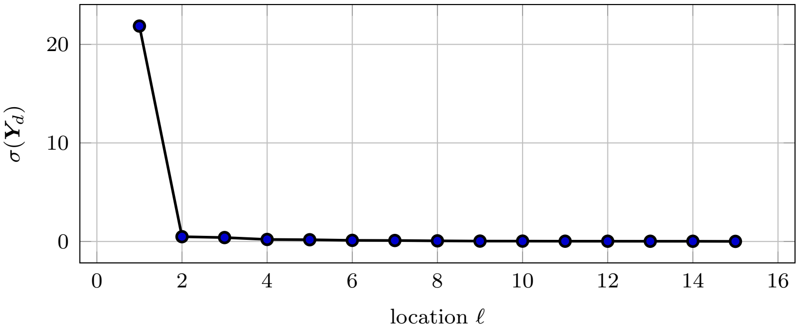


Figure 6.2.1: Singular values of measurement matrix $\boldsymbol{Y}_d$: low rank with one dominant singular value. Solar power data from PV panels at different locations within a ZIP code.

6.3 Numerical Results

Simulations on synthetic data: Synthetic data is generated according (6.1.1). Parameter α_ℓ is generated using (6.1.3) with $\underline{\alpha} = 0, \overline{\alpha} = 1$. The noise covariance matrix is simulated as $\boldsymbol{\Sigma} = \sigma^2\mathbb{I}$. Parameter $\boldsymbol{x}$ has each entry drawn at random, $[\boldsymbol{x}]_i \sim \mathcal{U}(0,1)$ where $\mathcal{U}(.)$ is uniform distribution. Graph $G = (\mathcal{V}, \mathcal{E})$ is constructed as a Erdos-Renyi graph with connectivity $p = 0.3$. Fig.6.2.2 shows lower bound on the estimation error. As seen, the bound is not tight. This is due to the non-convexity of the problem due to which it only converges to a stationary point [194] which compounds the error in estimating α_ℓ.

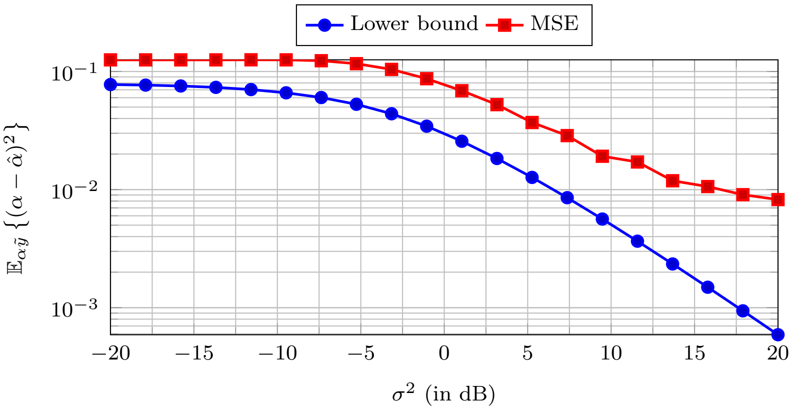


Figure 6.2.2: Comparison of lower bound with MSE for synthetic data when the prior of α_ℓ is constrained to be between 0 and 1 which is what one requires in the real-world example of fault detection in solar PV panels.

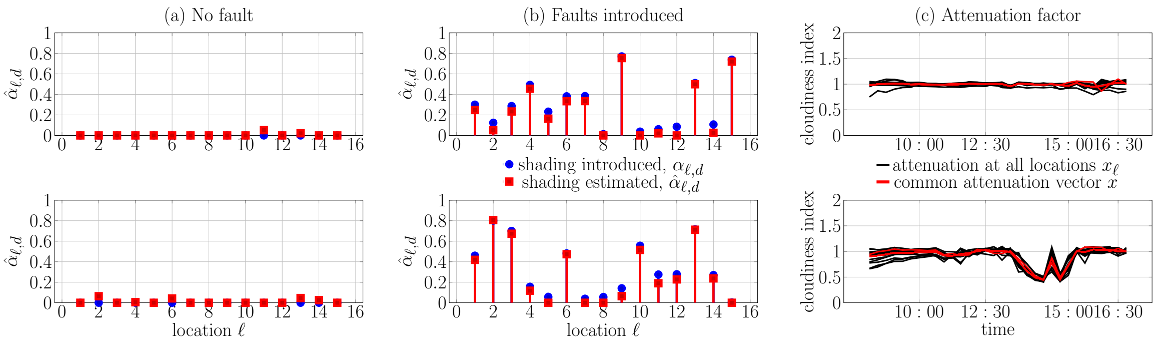


Figure 6.3.1: Results for April 29 and April 16, 2014 when (a) *no fault* introduced and (b) when faults are introduced according to (6.1.3) . In (c), attenuation vector for all locations, $\boldsymbol{x}_\ell$ is plotted prior to introduction of faults. The estimated $\boldsymbol{x}$ is plotted in red after the introduction of faults.

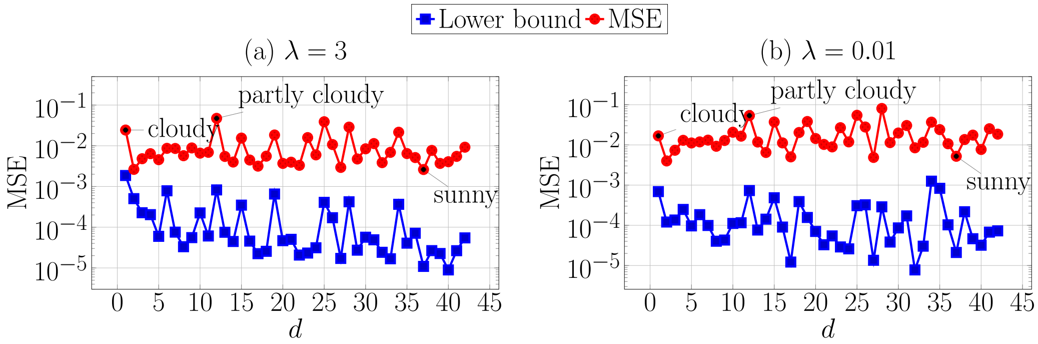


Figure 6.3.2: Comparison of lower bound with mean squared error for every day d for (a) $\lambda = 3$ and (b) $\lambda = 0.01$. Examples of cloudy, partly cloudy and sunny days are marked on both plots.

Fault detection for solar panels: As in the previous chapter, solar power data from SolarCity ordered by ZIP code is used. Faults (amount of partial shading) are generated using the prior distribution on $\alpha_{\ell,d}$ from (6.1.11) and introduced artificially in the data. The proposed estimation technique is applied to identify installations with partial shading faults within a ZIP code. The algorithm returns a value of $\alpha_{\ell,d}$ which is the fraction of soiling/ partial shading. Firstly, the results of estimation of α_ℓ are shown for two days-one sunny with little attenuation and other with varying cloud cover in Fig.6.3.1. As expected, the value of $\alpha_{\ell,d} \approx 0$, $\forall \ell$ for both the days when no fault is introduced. Then, the estimated partial shading $\tilde{\alpha}_\ell$ and the introduced shading fault are plotted. The estimated attenuation vector for the day is also plotted.

Lower bound on partial shading estimation error: Simulations are run to test the Bayesian lower bound in (6.1.16) while estimating the partial shading parameter $\alpha_{\ell,d}$. To compute the lower bound for each day, the common attenuation vector, $\boldsymbol{x}$ is the average: $\boldsymbol{x} = L^{-1}\boldsymbol{Y}\boldsymbol{1}^T$ assuming there is *no fault* in any panel. This only serves as a benchmark to compute the lower bound on mean squared error (MSE) of $\alpha_{\ell,d}, \ell = 1, 2, \ldots, L$ and not used in the process of estimating $\alpha_{\ell,d}$. In Fig. 6.3.2 different amount of faults are introduced with $\lambda = 0.01$ and $\lambda = 3$ for all days to compare the proximity of the lower-bound to MSE with changing weather conditions. As examples, cloudy, partly cloudy and sunny days are marked on the plot to show that the error on partly cloudy and cloudy days is higher than on sunny days in general. On partly cloudy days, the low rank structure is not exact due to the presence of uncoordinated spikes in power at different locations due to attenuation from local clouds.

REFERENCES

[1] A. G. Phadke, “Synchronized phasor measurements–a historical overview,” in *IEEE/PES Transmission and Distribution Conference and Exhibition*, vol. 1. IEEE, 2002, pp. 476–479.

[2] A. Phadke and J. Thorp, “History and Applications of Phasor Measurements,” in *2006 IEEE PES Power Systems Conference and Exposition*. IEEE, 2006, pp. 331–335.

[3] A. Monti, C. Muscas, and F. Ponci, *Phasor Measurement Units and Wide Area Monitoring Systems*. Academic Press, 2016.

[4] A. Rodriguez-Calvo, R. Cossent, and P. Frías, “Integration of PV and EVs in unbalanced residential LV networks and implications for the smart grid and advanced metering infrastructure deployment,” *International Journal of Electrical Power & Energy Systems*, vol. 91, pp. 121–134, 2017.

[5] Zhu, Xiaofan and Rabbat, Michael, “Approximating signals supported on graphs,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012.

[6] A. Sandryhaila and J. M. Moura, “Discrete Signal Processing on Graphs,” *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[7] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains,” *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[8] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, “Graph Signal Processing: Overview, Challenges, and Applications,” *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[9] P. Gao, M. Wang, S. G. Ghiocel, J. H. Chow, B. Fardanesh, and G. Stefopoulos, “Missing Data Recovery by Exploiting Low-dimensionality in Power System Synchrophasor Measurements,” *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1006–1013, 2016.

[10] P. Gao, M. Wang, S. G. Ghiocel, and J. H. Chow, “Modeless Reconstruction of Missing Synchrophasor Measurements,” in *2014 IEEE PES General Meeting Conference and Exposition*, 2014.

[11] M. Wang, "Data quality management of synchrophasor data in power systems by exploiting low-dimensional models," in *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, 2017.

[12] M. Jamei, A. Scaglione, C. Roberts, E. Stewart, S. Peisert, C. McParland, and A. McEachern, "Anomaly Detection Using Optimally-Placed PMU Sensors in Distribution Grids," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 3611–3622, July 2018.

[13] L. Xie, Y. Chen, and P. R. Kumar, "Dimensionality Reduction of Synchrophasor Data for Early Event Detection: Linearized Analysis," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2784–2794, 2014.

[14] W. Li, M. Wang, and J. H. Chow, "Real-Time Event Identification Through Low-Dimensional Subspace Characterization of High-Dimensional Synchrophasor Data," *IEEE Transactions on Power Systems*, vol. 33, no. 5, Sept 2018.

[15] J. Kim, L. Tong, and R. J. Thomas, "Subspace Methods for Data Attack on State Estimation: A Data Driven Approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1102–1114, March 2015.

[16] T. Ishizaki, A. Chakrabortty, and J.-I. Imura, "Graph-Theoretic Analysis of Power Systems," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 931–952, May 2018.

[17] K. Amare, V. A. Centeno, and A. Pal, "Unified PMU placement algorithm for power systems," in *2015 North American Power Symposium (NAPS)*, Oct 2015, pp. 1–6.

[18] B. Xu and A. Abur, "Observability analysis and measurement placement for systems with PMUs," in *IEEE PES Power Systems Conference and Exposition, 2004.*, vol. 2, Oct 2004, pp. 943–946.

[19] P. Du, Z. Huang, R. Diao, B. Lee, and K. K. Anderson, "PMU Placement for enhancing dynamic observability of a power grid," in *2010 IEEE Conference on Innovative Technologies for an Efficient and Reliable Electricity Supply*, Sept 2010, pp. 15–21.

[20] A. Pal, G. A. Sanchez-Ayala, V. A. Centeno, and J. S. Thorp, "A PMU Placement Scheme Ensuring Real-Time Monitoring of Critical Buses of the Network," *IEEE Transactions on Power Delivery*, vol. 29, no. 2, pp. 510–517, April 2014.

[21] Z. Wang, A. Scaglione, and R. Thomas, "Generating Statistically Correct Random Topologies for Testing Smart Grid Communication and Control Networks," *IEEE Transactions on Power Systems*, vol. 1, no. 1, pp. 28–39, 2010.

[22] Li, Xiao and Poor, H Vincent and Scaglione, Anna, "Blind topology identification for power systems," in *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2013, pp. 91–96.

[23] V. Kekatos, G. B. Giannakis, and R. Baldick, “Grid topology identification using electricity prices,” in *IEEE PES General Meeting Conference & Exposition*, 2014, pp. 1–5.

[24] J. Kim and L. Tong, “On topology attack of a smart grid,” in *2013 IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, Washington, DC, February 2013, pp. 1–6.

[25] J. Kim and L. Tong, “On topology attack of a smart grid: Undetectable attacks and countermeasures,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1294–1305, 2013.

[26] E. Drayer and T. Routtenberg, “Detection of False Data Injection Attacks in Power Systems with Graph Fourier Transform,” in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP). IEEE, 2018*, 2018, pp. 890–894.

[27] E. Drayer and T. Routtenberg, “Detection of False Data Injection Attacks in Smart Grids Based on Graph Signal Processing,” *IEEE Systems Journal*, August 2019.

[28] M. Jamei, R. Ramakrishna, T. Tesfay, R. Gentz, C. Roberts, A. Scaglione, and S. Peisert, “Phasor Measurement Units Optimal Placement and Performance Limits for Fault Localization,” *IEEE Journal on Selected Areas in Communications* , 2019.

[29] L. Fan, “Interarea Oscillations Revisited,” *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1585–1586, 2017.

[30] T. Huang, N. M. Freris, P. R. Kumar, and L. Xie, “Localization of Forced Oscillations in the Power Grid Under Resonance Conditions,” in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, 2018.

[31] L. Guo, C. Zhao, and S. H. Low, “Graph Laplacian Spectrum and Primary Frequency Regulation,” in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 158–165.

[32] Y. Weng, R. Negi, and M. D. Ilić, “Graphical model for state estimation in electric power systems,” in *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm).* IEEE, 2013, pp. 103–108.

[33] D. Deka, M. Chertkov, and S. Backhaus, “Topology Estimation Using Graphical Models in Multi-Phase Power Distribution Grids,” *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1663–1673, May 2020.

[34] K. Dvijotham, M. Chertkov, P. Van Hentenryck, M. Vuffray, and S. Misra, “Graphical models for optimal power flow,” *Constraints*, vol. 22, no. 1, pp. 24–49, 2017.

[35] R. Ramakrishna and A. Scaglione, "On Modeling Voltage Phasor Measurements as Graph Signals," in *IEEE Data Science Workshop (DSW) 2019*, June 2019, pp. 275–279.

[36] R. Ramakrishna and A. Scaglione, "Detection of False Data Injection Attack using Graph Signal Processing for the Power Grid," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP). IEEE*, 2019.

[37] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Autoregressive Moving Average Graph Filtering," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 274–288, 2017.

[38] H.-T. Wai, S. Segarra, A. E. Ozdaglar, A. Scaglione, and A. Jadbabaie, "Blind Community Detection From Low-Rank Excitations of a Graph Filter," *IEEE Transactions on Signal Processing*, vol. 68, pp. 436–451, 2020.

[39] R. Ramakrishna and A. Scaglione, " A Compressive Sensing framework for the analysis of Solar Photo-Voltaic Power," in *Conference Record of the Fiftieth Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 308–312.

[40] R. Ramakrishna, A. Bernstein, E. Dall'Anese, and A. Scaglione, "Joint Probabilistic Forecasts of Temperature and Solar Irradiance," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[41] R. Ramakrishna, A. Scaglione, V. Vittal, E. Dall'Anese, and A. Bernstein, "A Model for Joint Probabilistic Forecast of Solar Photovoltaic Power and Outdoor Temperature," *IEEE Transactions on Signal Processing*, vol. 67, no. 24, pp. 6368–6383, 2019.

[42] W. B. Powell and S. Meisel, "Tutorial on Stochastic Optimization in Energy—Part I: Modeling and Policies," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1459–1467, March 2016.

[43] M. Bazrafshan and N. Gatsis, "Decentralized Stochastic Optimal Power Flow in Radial Networks With Distributed Generation," *IEEE Transcations on Smart Grid*, vol. 8, no. 2, pp. 787–801, 2016.

[44] B. Defourny, D. Ernst, and L. Wehenkel, "Multistage Stochastic Programming: A Scenario Tree Based Approach to Planning under Uncertainty," in *Decision Theory Models for Applications in Artificial Intelligence: Concepts and Solutions*, L. E. Sucar, E. F. Morales, and J. Hoey, Eds. IGI Global, 2012, pp. 97–143.

[45] T. J. VanderWeele and I. Shpitser, "On the definition of a confounder," *The Annals of Statistics*, vol. 41, no. 1, pp. 196–220, 2013.

[46] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and Solar Power Forecasting for Smart Grid Energy Management," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 4, pp. 38–46, December 2015.

[47] H. C. Hottel, " A Simple Model for Estimating the Transmittance of Direct Solar Radiation Through Clear Atmospheres," *Solar Energy*, vol. 18, pp. 129–134, 1976.

[48] S. Pfenninger and I. Staffell, "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data," *Energy*, vol. 114, August 2016.

[49] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 2–10, March 2009.

[50] D. Larson, L. Nonnenmacher, and C. F. Coimbra, "Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest," *Renewable Energy*, vol. 91, pp. 11–20, January 2016.

[51] T. E. Hoff and R. Perez, "Quantifying PV power Output Variability," *Solar Energy*, vol. 84, no. 10, pp. 1782–1793, 2010.

[52] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, and B. Washom, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed," *Solar Energy*, vol. 85, 2011.

[53] J. Bosch and J. Kleissl, "Cloud motion vectors from a network of ground sensors in a solar power plant," *Solar Energy*, vol. 95, pp. 13–20, May 2013.

[54] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, March 2009.

[55] E. B. Iversena, J. M. Moralesa, J. K. Møllera, and H. Madsen, "Probabilistic Forecasts of Solar Irradiance by Stochastic Differential Equations," *Environmetrics*, vol. 25, no. 3, pp. 152–164, 2014.

[56] J. Boland, M. Korolkiewicz, M. Agrawal, and J. Huang, "Forecasting solar radiation on short time scales using a Coupled AutoRegressive and Dynamical System (CARDS) model," in *50th Annual Conference Australian Solar Energy Society*, December 2012.

[57] C. Yang, A. A. Thatte, and L. Xie, "Multitime-Scale Data-Driven Spatio-Temporal Forecast of Photovoltaic Generation," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 11, January 2015.

[58] X. G. Agoua, R. Girard, and G. Kariniotakis, "Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 538–546, 2017.

[59] M. Lave, J. Kleissl, and J. S. Stein, "A Wavelet-Based Variability Model (WVM) for Solar PV Power Plants," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, April 2013.

[60] M. Fan, V. Vittal, G. T. Heydt, and R. Ayyanar, "Preprocessing Uncertain Photovoltaic Data," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 1, pp. 351–352, January 2014.

[61] P. Shamsi, M. Marsousi, H. Xie, and W. Fries, "Dictionary learning for short-term prediction of solar PV production," in *2015 IEEE Power & Energy Society General Meeting*, July 2015.

[62] A. Mellit and S. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review," *Progress in Energy and Combustion Science*, vol. 34, no. 5, pp. 574–632, 2008.

[63] H. Wang, H. Yi, J. Peng, G. Wang, Y. Liu, H. Jiang, and W. Liu, "Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network," *Energy Conversion and Management*, vol. 153, 2017.

[64] M. D. Tabone and D. S. Callaway, "Modeling Variability and Uncertainty of Photovoltaic Generation: A Hidden State Spatial Statistical Approach," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 2965–2973, Nov 2015.

[65] Y. Z. Li, R. Luan, and J. C. Niu, "Forecast of power generation for grid-connected photovoltaic system based on grey model and Markov chain," in *IEEE Conference on Industrial Electronics and Applications*, June 2008.

[66] Z. Ren, W. Yan, X. Zhao, W. Li, and J. Yu, "Chronological Probability Model of Photovoltaic Generation," *IEEE Transactions on Power Systems*, vol. 29, no. 3, 2014.

[67] M. J. Sanjari and H. B. Gooi, "Probabilistic Forecast of PV Power Generation Based on Higher Order Markov Chain," *IEEE Transactions on Power Systems*, vol. 32, no. 4, July 2017.

[68] A. Bracale, G. Carpinelli, and P. D. Falco, "A Probabilistic Competitive Ensemble Method for Short-Term Photovoltaic Power Forecasting," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 551–560, April 2017.

[69] S. A. Fatemi, A. Kuh, and M. Fripp, "Parametric methods for probabilistic forecasting of solar irradiance," *Renewable Energy*, 2018.

[70] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," *International Journal of Forecasting*, vol. 32, no. 3, 2016.

[71] Y. Li, B. Zhao, Z. Zhao, R. Taylor, and R. Wang, "Performance study of a grid-connected photovoltaic powered central air conditioner in the south china climate," *Renewable Energy*, vol. 126, pp. 1113–1125, 2018.

[72] R. Perez, T. Hoff, C. Herig, and J. Shah, "Maximizing PV peak shaving with solar load control: validation of a web-based economic evaluation tool," *Solar Energy*, 2003.

[73] E. C. Kara, M. D. Tabone, J. S. MacDonald, D. S. Callaway, and S. Kiliccote, “Quantifying Flexibility of Residential Thermostatically Controlled Loads for Demand Response: A Data-driven Approach ,” Ernest Orlando Lawrence Berkeley National Laboratory, Technical Report, 2014.

[74] A. Arabali, M. Ghofrani, M. Etezadi-Amoli, and M. S. Fadali, “Stochastic Performance Assessment and Sizing for a Hybrid Power System of Solar/Wind/Energy Storage,” *IEEE Transactions on Sustainable Energy*, vol. 5, no. 2, Apr. 2014.

[75] Y. Fujimoto, H. Kikusato, S. Yoshizawa, S. Kawano, A. Yoshida, S. Wakao, N. Murata, Y. Amano, S.-i. Tanabe, and Y. Hayashi, “Distributed Energy Management for Comprehensive Utilization of Residential Photovoltaic Outputs,” *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1216–1227, 2016.

[76] R. Luthander, J. Widén, D. Nilsson, and J. Palm, “Photovoltaic self-consumption in buildings: A review,” *Applied Energy*, vol. 80, no. 94, 2015.

[77] J. D. Hamilton, “Regime switching models,” in *Macroeconometrics and time series analysis.* Springer, 2010.

[78] H. Braun, S. T. Buddha, V. Krishnan, A. Spanias, C. Tepedelenlioglu, T. Yeider, and T. Takehara, “Signal Processing for fault detection in photovoltaic arrays,” in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 1681–1684.

[79] S. Rao, D. Ramirez, H. Braun, J. Lee, C. Tepedelenlioglu, E. Kyriakides, D. Srinivasan, J. Frye, S. Koizumi, Y. Morimoto, and A. Spanias, “An 18 kW solar array research facility for fault detection experiments,” in *2016 18th Mediterranean Electrotechnical Conference (MELECON)*, April 2016, pp. 1–5.

[80] A. Scaglione, R. Pagliari, and H. Krim, “The Decentralized Estimation of the Sample Covariance,” in *Conference Record of the 42nd Asilomar Conference on Signals, Systems and Computers*, 2008.

[81] K. A. Kim, G.-S. Seo, B.-H. Cho, and P. T. Krein, “Photovoltaic Hot-Spot Detection for Solar Panel Substrings Using AC Parameter Characterization,” *IEEE Transcations on Power Electronics*, 2016.

[82] Y. Zhao, R. Ball, J. Mosesian, J.-F. de Palma, and B. Lehman, “Graph-Based Semi-supervised Learning for Fault Detection and Classification in Solar Photovoltaic Arrays,” *IEEE Transcations on Power Electronics*, 2015.

[83] M. N. Akram and S. Lotfifard, “Modeling and Health Monitoring of DC Side of Photovoltaic Array,” *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, 2015.

[84] S. Katoch, G. Muniraju, S. Rao, A. Spanias, P. Turaga, C. Tepedelenlioglu, M. Banavar, and D. Srinivasan, “Shading Prediction, Fault Detection and Consensus Estimation for Solar Array Control,” in *IEEE International Conference on Industrial Cyber-Physical Systems (ICPS)*, 2018.

[85] H. Braun, C. Tepedelenlioglu, A. Spanias, and M. Banavar, *Signal Processing for Solar Array Monitoring, Fault Detection, and Optimization, Synthesis Lectures on Power Electronics*, ser. ISBN 978-1608459483. Morgan & Claypool, 2012.

[86] Z. Yi and A. H. Etemadi, "Fault Detection for Photovoltaic Systems Based on Multi-Resolution Signal Decomposition and Fuzzy Inference Systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, 2017.

[87] R. Platon, J. Martel, N. Woodruff, and T. Y. Chau, "Online Fault Detection in PV Systems," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1200–1207, 2015.

[88] S. Rao, S. Katoch, V. Narayanaswamy, G. Muniraju, C. Tepedelenlioglu, A. Spanias, and P. Turaga, *Machine Learning for Solar Array Monitoring, Optimization, and Control*, J. Hudgins, Ed. Morgan & Claypool Publishers, 2020.

[89] J. Fan, S. Rao, G. Muniraju, C. Tepedelenlioglu, and A. Spanias, "Fault Classification in Photovoltaic Arrays Using Graph Signal Processing," in *IEEE International Conference on Industrial Cyber-Physical Systems (ICPS)*, June 2020.

[90] A. Sandryhaila and J. M. F. Moura, "Discrete Signal Processing on Graphs: Frequency Analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, June 2014.

[91] N. Tremblay, P. Gonçalves, and P. Borgnat, "Design of Graph Filters and Filterbanks," in *Cooperative and Graph Signal Processing.* Elsevier, 2018, pp. 299–324.

[92] K. Rohe, S. Chatterjee, B. Yu *et al.*, "Spectral clustering and the high-dimensional stochastic blockmodel," *The Annals of Statistics*, vol. 39, no. 4, pp. 1878–1915, 2011.

[93] X. Ding, T. Jiang *et al.*, "Spectral distributions of adjacency and Laplacian matrices of random graphs," *The Annals of Applied Probability*, vol. 20, no. 6, pp. 2086–2117, 2010.

[94] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Separable autoregressive moving average graph-temporal filters," in *2016 24th European Signal Processing Conference (EUSIPCO).* IEEE, 2016, pp. 200–204.

[95] D. Thanou, X. Dong, D. Kressner, and P. Frossard, "Learning Heat Diffusion Graphs," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 484–499, 2017.

[96] C. Ravazzi, R. Tempo, and F. Dabbene, "Learning Influence Structure in Sparse Social Networks," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 4, pp. 1976–1986, 2017.

[97] N. E. Friedkin, "A Formal Theory of Reflected Appraisals in the Evolution of Power," *Administrative Science Quarterly*, vol. 56, no. 4, pp. 501–529, 2011.

[98] O. Candogan, K. Bimpikis, and A. Ozdaglar, "Optimal Pricing in Networks with Externalities," *Operations Research*, vol. 60, no. 4, pp. 883–905, 2012.

[99] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon, "Econometric measures of connectedness and systemic risk in the finance and insurance sectors," *Journal of Financial Economics*, vol. 104, no. 3, pp. 535–559, 2012.

[100] R. N. Mantegna and H. E. Stanley, *Introduction to Econophysics: Correlations and Complexity in Finance.* Cambridge University Press, 1999.

[101] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge University Press, 1990.

[102] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way.* Academic press, 2009.

[103] R. Singh, A. Chakraborty, and B. Manoj, "Graph Fourier transform based on direcetd Laplacian," in *2016 International Conference on Signal Processing and Communications (SPCOM)*, 2016.

[104] J. D. Glover, M. S. Sarma, and T. J. Overbye, *Power system analysis and design.* Cengage Learning, 2008.

[105] N. Sato and W. F. Tinney, "Techniques for Exploiting the Sparsity or the Network Admittance Matrix," *IEEE Transactions on Power Apparatus and Systems*, vol. 82, no. 69, pp. 944–950, 1963.

[106] F. Dörfler and F. Bullo, "Kron Reduction of Graphs With Applications to Electrical Networks," *IEEE Transactions on Circuits and Systems-I: Regular Papers*, vol. 60, no. 1, 2013.

[107] F. Dorfler and F. Bullo, "Spectral Analysis of Synchronization in a Lossless Structure-Preserving Power Network Model," in *2010 First IEEE International Conference on Smart Grid Communications*, 2010.

[108] F. Gao, J. S. Thorp, A. Pal, and S. Gao, "Dynamic state prediction based on Auto-Regressive (AR) Model using PMU data," in *2012 IEEE Power and Energy Conference at Illinois.* IEEE, 2012, pp. 1–5.

[109] N. Dahal, R. L. King, and V. Madani, "Online dimension reduction of synchrophasor data," in *IEEE PES Transmission and Distribution Conference and Exposition (T & D)*, 2012.

[110] M. Liao, D. Shi, Z. Yu, Z. Yi, Z. Wang, and Y. Xiang, "An Alternating Direction Method of Multipliers Based Approach for PMU Data Recovery," *IEEE Transactions on Smart Grid*, 2018.

[111] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on Graphs: Uncertainty Principle and Sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, 2016.

[112] F. Grassi, A. Loukas, N. Perraudin, and B. Ricaud, “A Time-Vertex Signal Processing Framework: Scalable Processing and Meaningful Representations for Time-Series on Graphs,” *IEEE Transactions on Signal Processing*, vol. 66, no. 3, pp. 817–829, February 2018.

[113] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, “Grid Structural Characteristics as Validation Criteria for Synthetic Networks,” *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3258–3265, July 2017.

[114] S. Maslennikov, B. Wang, Q. Zhang, F. Ma, X. Luo, K. Sun, and E. Litvinov, “A Test Cases Library for Methods Locating the Sources of Sustained Oscillations,” in *IEEE PES General Meeting, Boston, MA*, 2016.

[115] S. Zhang, Y. Hao, M. Wang, and J. H. Chow, “Multi-Channel Hankel Matrix Completion through Nonconvex Optimization,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 4, pp. 617–632, 2018.

[116] A. B. Birchfield, T. Xu, and T. J. Overbye, “Power flow convergence and reactive power planning in the creation of large synthetic grids,” *IEEE Transactions on Power Systems*, vol. 33, no. 6, 2018.

[117] U. Von Luxburg, “A tutorial on spectral clustering,” *Statistics and computing, Springer*, vol. 17, no. 4, pp. 395–416, 2007.

[118] A. Y. Ng, M. I. Jordan, and Y. Weiss, “On spectral clustering: Analysis and an algorithm,” in *Advances in Neural Information Processing Systems*, 2002.

[119] S. M. Brahma, “Fault Location in Power Distribution System With Penetration of Distributed Generation,” *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1545–1553, 2011.

[120] I. Dzafic, R. A. Jabr, S. Henselmeyer, and T. Donlagic, “Fault Location in Distribution Networks Through Graph Marking,” *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1345–1353, 2018.

[121] H. Jiang, J. J. Zhang, W. Gao, and Z. Wu, “=Fault Detection, Identification, and Location in Smart Grid Based on Data-Driven Computational Methods,” *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2947–2956, 2014.

[122] M. Jamei, A. Scaglione, and S. Peisert, “Low-Resolution Fault Localization Using Phasor Measurement Units with Community Detection,” in *2018 IEEE International Conference on Smart Grid Communications*, October 2018.

[123] W. H. Kersting, “Radial distribution test feeders,” in *Power Engineering Society Winter Meeting, 2001. IEEE*, vol. 2. IEEE, 2001, pp. 908–912.

[124] D. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization,” *Proceedings of the National Academy of Sciences*, 2003.

[125] P.D.Lorenzo, S. Barbarossa, and P. Banelli, *Cooperative and Graph Signal Processing.* Elsevier, June 2018, ch. 9.

[126] P. Top and J. Breneman, "Compressing phasor measurement data," in *2013 IEEE Power Energy Society General Meeting*, July 2013, pp. 1–4.

[127] R. Klump, P. Agarwal, J. E. Tate, and H. Khurana, "Lossless compression of synchronized phasor measurements," in *IEEE PES General Meeting*, July 2010, pp. 1–7.

[128] J. E. Tate, "Preprocessing and Golomb-Rice encoding for lossless compression of phasor angle data," *IEEE Trans. on Smart Grid*, vol. 7, no. 2, pp. 718–729, March 2016.

[129] S. Kirti, Z. Wang, A. Scaglione, and R. Thomas, "On the Communication Architecture for Wide-Area Real-Time Monitoring in Power Networks," in *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, 2007.

[130] P. H. Gadde, M. Biswal, S. Brahma, and H. Cao, "Efficient compression of PMU data in WAMS," *IEEE Trans. on Smart Grid*, vol. 7, no. 5, pp. 2406–2413, Sept 2016.

[131] Y. Ge, A. J. Flueck, D. K. Kim, J. B. Ahn, J. D. Lee, and D. Y. Kwon, "Power system real-time event detection and associated data archival reduction based on synchrophasors," *IEEE Trans. on Smart Grid*, vol. 6, no. 4, pp. 2088–2097, July 2015.

[132] J. C. S. de Souza, T. M. L. Assis, and B. C. Pal, "Data Compression in Smart Distribution Systems via Singular Value Decomposition," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, January 2017.

[133] R. Mehra, V. Patel, F. Kazi, N. M. Singh, and S. R. Wagh, "Modes preserving wavelet based multi-scale PCA algorithm for compression of smart grid data," in *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Aug 2013, pp. 817–821.

[134] W. Weber, "Differential Encoding for Multiple Amplitude and Phase Shift Keying Systems," *IEEE Transactions on Communication*, vol. 26, no. 3, 1978.

[135] T. M. Cover and J. Thomas, *Elements of Information Theory.* John Wiley, 1991.

[136] K. D. Jones, A. Pal, and J. S. Thorp, "Methodology for performing synchrophasor data conditioning and validation," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1121–1130, May 2015.

[137] M. Wu and L. Xie, "Online detection of low-quality synchrophasor measurements: A data-driven approach," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2817–2827, July 2017.

[138] S. M. Blair, C. D. Booth, G. Williamson, A. Poralis, and V. Turnham, "Automatically detecting and correcting errors in power quality monitoring data," *IEEE Transactions on Power Delivery*, vol. 32, no. 2, pp. 1005–1013, April 2017.

[139] Y. Seyedi, H. Karimi, and J. M. Guerrero, "Centralized disturbance detection in smart microgrids with noisy and intermittent synchrophasor data," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2775–2783, Nov 2017.

[140] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on DC state estimation," in *Preprints of the First Workshop on Secure Control Systems, CPSWEEK*, vol. 2010, 2010.

[141] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, 2010, pp. 5991–5998.

[142] G. Dan and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, 2010, pp. 214–219.

[143] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks: characterizations and countermeasures,," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 232–237.

[144] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[145] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 645–658, 2011.

[146] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 326–333, 2011.

[147] Q. Yang, J. Yang, W. Yu, D. An, N. Zhang, and W. Zhao, "On false data-injection attacks against power system state estimation: Modeling and countermeasures," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 3, pp. 717–729, 2014.

[148] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 612–621, 2014.

[149] G. Hug and J. A. Giampapa, "Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1362–1370, 2012.

[150] L. Xie, Y. Mo, and B. Sinopoli, “False Data Injection Attacks in Electricity Markets,” in *2010 First IEEE International Conference on Smart Grid Communications*, 2010, pp. 226–231.

[151] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, “A review of false data injection attacks against modern power systems,” *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, July 2017.

[152] Y. Chen, L. Xie, and P. R. Kumar, “Power system event classification via dimensionality reduction of synchrophasor data,” in *2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2014, pp. 57–60.

[153] Z. Chu, J. Zhang, O. Kosut, and L. Sankar, “Vulnerability assessment of large-scale power systems to false data injection attacks,” [Online] arXiv:1705.04218, May 2017, submitted to *IEEE Trans. on Power Systems*.

[154] Z.-H. Yu and W.-L. Chin, “Blind false data injection attack using pca approximation method in smart grid,” *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1219–1226, 2015.

[155] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, “On malicious data attacks on power system state estimation,” in *International Universities Power Engineering Conference (UPEC)*, Aug. 2010, pp. 1–6.

[156] O. Kosut, “Malicious data attacks against dynamic state estimation in the presence of random noise,” in *First IEEE Global Conference on Signal and Information Processing*, Dec. 2013.

[157] J. Liang, O. Kosut, and L. Sankar, “Cyber-attacks on AC state estimation: Unobservability and physical consequences,” in *IEEE PES General Meeting*, Jul. 2014.

[158] J. Liang, L. Sankar, and O. Kosut, “Vulnerability analysis and consequences of false data injection attack on power system state estimation,” *IEEE Trans. on Power Systems*, vol. 31, no. 5, pp. 3864–3872, Sept 2016.

[159] J. Zhang, Z. Chu, L. Sankar, and O. Kosut, “False data injection attacks on power system state estimation with limited information,” in *IEEE Power and Energy Society General Meeting (PESGM)*, July 2016, pp. 1–5.

[160] A. Pinceti, L. Sankar, and O. Kosut, “Load redistribution attack detection using machine learning: A data-driven approach,” in *2018 IEEE Power and Energy Society General Meeting (PESGM)*, Aug. 2018.

[161] I. Esnaola, S. M. Perlaza, H. V. Poor, and O. Kosut, “Maximum Distortion Attacks in Electricity Grids,” *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 2007–2015, Jul. 2016.

[162] J. Zhang, Z. Chu, L. Sankar, and O. Kosut, "False data injection attacks on phasor measurements that bypass low-rank decomposition," in *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Oct. 2017.

[163] J. Zhang and L. Sankar, "Implementation of unobservable state-preserving topology attacks," in *North American Power Symposium (NAPS)*, Oct. 2015, pp. 1–6.

[164] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505–2516, Sept 2017.

[165] P. Gao, M. Wang, J. H. Chow, S. G. Ghiocel, B. Fardanesh, G. Stefopoulos, and M. P. Razanousky, "Identification of successive "unobservable " cyber data attacks in power systems through matrix decomposition," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5557–5570, Nov 2016.

[166] G. Masters, *Renewable and Efficient Electric Power Systems.* Wiley, 2004.

[167] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions.* Dover Publications, 1965.

[168] A. Kankiewicz, M. Sengupta, and D. Moon, "Observed impacts of transient clouds on utility-scale PV fields," in *ASES National Solar Conference*, 2010.

[169] A. E. Curtright and J. Apt, "The Character of Power Output from Utility-Scale Photovoltaic Systems," *Progress in Photovoltaics: Research and Applications*, vol. 16, pp. 241–247, September 2007.

[170] Alan V. Oppenheim and Ronald W. Schafer, *Discrete-Time Signal Processing.* Pearson, 2010.

[171] Ivana Tosic and Pascal Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, pp. 27–38, March 2011.

[172] B. Mailhé, S. Lesage, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Shift-invariant dictionary learning for sparse representations: Extending K-SVD," in *European Signal Processing Conference*, vol. 4, 2008.

[173] J. Tropp and S. Wright, "Computational Methods for Sparse Solution of Linear Inverse Problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, June 2010.

[174] Stephen J. Wright, "Coordinate Descent Algorithms," *Mathematical Programming*, vol. 151, no. 1, pp. 3–34, June 2015.

[175] W. G. Kannuluik and E. H. Carman, "The Temperature Dependence of the Thermal Conductivity of Air," *Australian Journal of Scientific Research, Series A: Physical Sciences*, vol. 4, 1951.

[176] Halliday and Resnick, *Fundamentals of Physics.* Wiley, 2010.

[177] G.-O. A. Glentis, P. Koukoulas, and N. Kalouptsidis, “Efficient Algorithms for Volterra System Identification,” *IEEE Transactions on Signal Processing*, vol. 47, no. 11, 1999.

[178] Taiho Koh and Edward J. Powers , “Second-Order Volterra Filtering and Its Application to Nonlinear System Identification ,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 6, 1985.

[179] R. Ramakrishna, A. Scaglione, A. Spanias, and C. Tepedelenlioglu, “Distributed Bayesian Estimation with Low-rank Data: Application to Solar Array Processing,” in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

[180] R. H. Inman, H. T. Pedro, and C. F. Coimbra, “Solar forecasting methods for renewable energy integration,” *Progress in Energy and Combustion Science*, 2013.

[181] F. Jelinek, “Continuous Speech Recognition by Statistical Methods,” *Proceedings of the IEEE*, vol. 64, no. 4, pp. 532–556, April 1976.

[182] B. H. Juang and L. Rabiner, “The Segmental K-Means Algorithm for Estimating Parameters of Hidden Markov Models,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 9, pp. 1639–1641, September 1990.

[183] G.D.Forney, “The Viterbi algorithm,” *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, March 1973.

[184] L. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[185] T. Gneiting, F. Balabdaoui, and A. E. Raftery, “Probabilistic forecasts, calibration and sharpness,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology*, vol. 69, no. 243-268, 2007.

[186] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, “Optimal Prediction Intervals of Wind Power Generation,” *IEEE Transactions on Power Systems*, vol. 29, no. 3, May 2014.

[187] R. L. Winkler, “A decision-theoretic approach to interval estimation,” *Journal of American Statistical Association*, vol. 67, no. 33, 1972.

[188] R. Koenker and K. F. Hallock, “Quantile regression,” *Journal of economic perspectives*, vol. 15, no. 4, pp. 143–156, 2001.

[189] E. P. Grimit, T. Gneiting, V. Berrocal, and N. A. Johnson, “The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification,” *Quarterly Journal of the Royal Meteorological Society*, vol. 132, no. 621C, 2006.

[190] H.-T. Wai, T.-H. Chang, and A. Scaglione, "A consensus-based decentralized algorithm for non-convex optimization with application to dictionary learning," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[191] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, 2000.

[192] R.Prasad and C.R.Murthy, "Cramér-Rao-Type Bounds for Sparse Bayesian Learning," *IEEE Transactions on Signal Processing*, 2013.

[193] R. D. Gill and B. Y. Levit, "Applications of the van Trees Inequality: A Bayesian Cramér-Rao Bound," *Bernoulli Society for Mathematical Statistics and Probability*, vol. 1, no. 1/2, pp. 59–79, Jun 1995.

[194] H.-T. Wai, J. Lafond, A. Scaglione, and E. Moulines, "Decentralized Frank–Wolfe Algorithm for Convex and Nonconvex Problems," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, Nov 2017.

[195] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, 1995.

[196] S. S. Rao, "Advanced Statistical Inference," Texas A & M University, Tech. Rep., April 2017.

APPENDIX A

DERIVATION OF LOWER BOUND

For the proof of Theorem 1, we first introduce few lemmas.

Let θ be the random parameter to be estimated with prior distribution $p(\theta)$. Let y be the observation or measurement and $p(y|\theta)$ be the conditional distribution of observation given the parameter. In this section, a general lower bound for any estimator $\hat{\theta}(y)$ is postulated.

The following regularity conditions are necessary for the lower bound to hold:

1. For the conditional distribution, $p(y|\theta)$, the following holds

$$\int_y \frac{\partial}{\partial\theta} p(y|\theta)dy = 0, \tag{A.0.1}$$

   and Fisher information [195],

$$J(\theta) = \mathbb{E}_{y|\theta}\left\{\left(\frac{\partial}{\partial\theta}\log p(y|\theta)\right)^2\right\} \tag{A.0.2}$$

   is defined.
2. Prior distribution $p(\theta)$ has bounded support i.e. $p(\theta) = 0$ if $\theta \notin [\underline{\theta}, \overline{\theta}]$ and $\underline{\theta}, \overline{\theta} < \infty$. In addition, $p(\theta)$ is differentiable in the interval $\theta \in [\underline{\theta}, \overline{\theta}]$, i.e. $\frac{\partial}{\partial\theta}p(\theta) \triangleq p'(\theta)$ is defined when $\theta \in [\underline{\theta}, \overline{\theta}]$.

Lemma 1 (A Bayesian Lower Bound). *If conditions 1 and 2 are satisfied then for an estimator $\hat{\theta}(y)$ with bias function*

$$\beta(\theta) \triangleq \int \left(\hat{\theta}(y) - \theta\right) p(y|\theta)dy, \tag{A.0.3}$$

a lower bound on the MSE is given by

$$\mathbb{E}_{\theta,y}\left[\left(\hat{\theta}(y) - \theta\right)^2\right] \geq \frac{\left(1 + \beta(\overline{\theta})p(\overline{\theta}) - \beta(\underline{\theta})p(\underline{\theta})\right)^2}{J_P + J_B} \tag{A.0.4}$$

$$\textit{where } J_P \triangleq \mathbb{E}_\theta[J(\theta)], \quad J_B \triangleq \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log p(\theta)\right)^2\right] \tag{A.0.5}$$

Proof: Lemma 1. From [193,196] the basic step is the application of Cauchy-Schwarz inequality to the following integral,

$$\zeta \triangleq \left[\int\left(\hat{\theta}(y) - \theta\right)\frac{\partial p(y|\theta)p(\theta)}{\partial\theta}d\theta dy\right]^2 \leq \mathbb{E}_{\theta,y}\left[\left(\hat{\theta}(y) - \theta\right)^2\right] (J_P + J_B) \tag{A.0.6}$$

Now, to evaluate ζ consider

$$\int\!\Big(\hat{\theta}(y)-\theta\Big)\frac{\partial}{\partial\theta}p(y|\theta)p(\theta)d\theta dy=\!\int\!\Big[\Big(\hat{\theta}(y)-\theta\Big)p(y|\theta)p(\theta)\Big]_{\underline{\theta}}^{\overline{\theta}}dy \\ \underbrace{-\int\frac{\partial}{\partial\theta}\Big(\hat{\theta}(y)-\theta\Big)p(y|\theta)p(\theta)d\theta dy}_{=-1} \tag{A.0.7}$$

First term in (A.0.7) is written using the bias function:

$$\int_y \Big[\Big(\hat{\theta}(y)-\theta\Big)\,p(y|\theta)p(\theta)\Big]_{\underline{\theta}}^{\overline{\theta}}dy = \beta(\overline{\theta})p(\overline{\theta}) - \beta(\underline{\theta})p(\underline{\theta}) \tag{A.0.8}$$

$$\text{So,} \qquad \zeta = \Big(1+\beta(\overline{\theta})p(\overline{\theta}) - \beta(\underline{\theta})p(\theta)\Big)^2 \tag{A.0.9}$$

$$\boxed{\mathbb{E}_{\theta,y}\left[\Big(\hat{\theta}(y)-\theta\Big)^2\right] \geq \frac{\Big(1+\beta(\overline{\theta})p(\overline{\theta}) - \beta(\underline{\theta})p(\theta)\Big)^2}{J_P+J_B}} \tag{A.0.10}$$

□

Now, as consistent with the problem, consider a truncated exponential prior

$$p(\theta) = (C_\lambda)^{-1}\lambda e^{-\lambda\theta},\ \underline{\theta}\leq\theta\leq\overline{\theta} \tag{A.0.11}$$

where $C_\lambda = e^{-\lambda\underline{\theta}} - e^{-\lambda\overline{\theta}}$. Compute the proposed lower bound for a MAP estimator when the observed value of θ is corrupted with independent addition of Gaussian noise ν,

$$y = \theta + \nu, \quad \nu\sim\mathcal{N}(0,\sigma^2). \tag{A.0.12}$$

Assume that only one sample y is available for estimation with y. The MAP estimator is

$$\hat{\theta}_{\text{MAP}}(y) = \begin{cases} \underline{\theta}, & y<\underline{\theta}+\lambda\sigma^2 \\ y-\lambda\sigma^2, & \lambda\sigma^2+\underline{\theta}\leq y\leq\lambda\sigma^2+\overline{\theta} \\ \overline{\theta}, & y>\lambda\sigma^2+\overline{\theta} \end{cases} \tag{A.0.13}$$

$$\text{Let } a(\theta) = \big(\underline{\theta}+\lambda\sigma^2-\theta\big)/\sigma,\ b(\theta) = \big(\overline{\theta}+\lambda\sigma^2-\theta\big)/\sigma \tag{A.0.14}$$

Lemma 2. *For the MAP estimator in* (A.0.13) *where the prior distribution on* θ *is given by* (A.0.11)*, the bias function is*

$$\beta_{MAP}(\theta) = \overline{\theta}-\theta + \big[\underline{\theta}-\theta+\lambda\sigma^2\big]\Phi\left(a(\theta)\right) + \big[\theta-\overline{\theta}-\lambda\sigma^2\big]\Phi\left(b(\theta)\right) \\ + \sigma\left[\phi\left(a(\theta)\right) - \phi\left(b(\theta)\right)\right] \tag{A.0.15}$$

a lower bound for the MAP estimator is

$$\mathbb{E}_{\theta,y}\left[\left(\hat{\theta}_{MAP}(y)-\theta\right)^2\right] \geq \frac{\sigma^2(1-\nu)^2}{\lambda^2\sigma^2+1} \tag{A.0.16}$$

$$\begin{aligned} \textit{where } \nu = &(p(\overline{\theta})+p(\underline{\theta}))\left[\sigma\phi(\lambda\sigma)+\lambda\sigma^2\Phi(\lambda\sigma)\right]+rp(\underline{\theta}) && \text{(A.0.17)}\\ &-p(\overline{\theta})\left[\sigma\phi\left(\lambda\sigma-\frac{r}{\sigma}\right)+(\lambda\sigma^2-r)\Phi\left(\lambda\sigma-\frac{r}{\sigma}\right)\right] \\ &-p(\underline{\theta})\left[\sigma\phi\left(\lambda\sigma+\frac{r}{\sigma}\right)+(\lambda\sigma^2+r)\Phi\left(\lambda\sigma+\frac{r}{\sigma}\right)\right] && \text{(A.0.18)} \end{aligned}$$

Proof. Bias function (A.0.15) can be computed by evaluating (A.0.3). Subsequent simplification after necessary substitution in (A.0.4) gives (A.0.18). □

Direct application of lemma 2 to the problem proves Theorem 1.