Robust Identification, Estimation, and Control of Electric Power Systems Using the Koopman Operator-Theoretic Framework

Marcos Netto

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical Engineering

Lamine Mili, Chair
Daniel Stilwell
Michael von Spakovsky
Vassilis Kekatos
Virgilio Centeno
Yoshihiko Susuki

December 10, 2018
Falls Church, Virginia

Keywords: Dynamical systems, Kalman filtering, Koopman mode decomposition, modal analysis, participation factors, robust estimation, power system stability and control.

Robust Identification, Estimation, and Control of Electric Power Systems Using the Koopman Operator-Theoretic Framework

Marcos Netto

(ABSTRACT)

The study of nonlinear dynamical systems via the spectrum of the Koopman operator has emerged as a paradigm shift, from the Poincaré's geometric picture that centers the attention on the *evolution of states*, to the Koopman operator's picture that focuses on the *evolution of observables*. The Koopman operator-theoretic framework rests on the idea of lifting the states of a nonlinear dynamical system to a higher dimensional space; these lifted states are referred to as the Koopman eigenfunctions. To determine the Koopman eigenfunctions, one performs a nonlinear transformation of the states by relying on the so-called observables, that is, scalar-valued functions of the states. In other words, one executes a change of coordinates from the state space to another set of coordinates, which are denominated Koopman canonical coordinates. The variables defined on these intrinsic coordinates will evolve linearly in time, despite the underlying system being nonlinear. Since the Koopman operator is linear, it is natural to exploit its spectral properties. In fact, the theory surrounding the spectral properties of linear operators has well-known implications in electric power systems. Examples include small-signal stability analysis and direct methods for transient stability analysis based on the Lyapunov function. From the applications' standpoint, this framework based on the Koopman operator is attractive because it is capable of revealing linear and nonlinear modes by only applying well-established tools that have been developed for linear systems. With the challenges associated with the high-dimensionality and increasing uncertainties in the power systems models, researchers and practitioners are seeking alternative modeling approaches capable of incorporating information from measurements. This is fueled by an increasing amount of data made available by the wide-scale deployment of measuring devices such as phasor measurement units and smart meters. Along these lines, the Koopman operator theory is a promising framework for the integration of data analysis into our mathematical knowledge and is bringing an exciting perspective to the community. The present dissertation reports on the application of the Koopman operator for identification, estimation, and control of electric power systems. A dynamic state estimator based on the Koopman operator has been developed and compares favorably against model-based approaches, in particular for centralized dynamic state estimation. Also, a data-driven method to compute participation factors for nonlinear systems based on Koopman mode decomposition has been developed; it generalizes the original definition of participation factors under certain conditions.

That this work received support from CAPES Foundation, Ministry of Education of Brazil, under grant BEX 13594/13-3.

Robust Identification, Estimation, and Control of Electric Power Systems Using the Koopman Operator-Theoretic Framework

Marcos Netto

(GENERAL AUDIENCE ABSTRACT)

Electric power systems are complex, large-scale, and given the bidirectional causality between economic growth and electricity consumption, they are constantly being expanded. In the U.S., some of the electric power grid facilities date back to the 1880s, and this aging system is operating at its capacity limits. In addition, the international pressure for sustainability is driving an unprecedented deployment of renewable energy sources into the grid. Unlike the case of other primary sources of electric energy such as coal and nuclear, the electricity generated from renewable energy sources is strongly influenced by the weather conditions, which are very challenging to forecast even for short periods of time. Within this context, the mathematical models that have aided engineers to design and operate electric power grids over the past decades are falling short when uncertainties are incorporated to the models of such high-dimensional systems. Consequently, researchers are investigating alternative data-driven approaches. This is not only motivated by the need to overcome the above challenges, but it is also fueled by the increasing amount of data produced by today's powerful computational resources and experimental apparatus. In power systems, a massive amount of data will be available thanks to the deployment of measuring devices called phasor measurement units. Along these lines, the Koopman operator theory is a promising framework for the integration of data analysis into our mathematical knowledge, and is bringing an exciting perspective on the treatment of high-dimensional systems that lie in the forefront of science and technology. In the research work reported in this dissertation, the Koopman operator theory has been exploited to seek for solutions to some of the challenges that are threatening the safe, reliable, and efficient operation of electric power systems.

That this work received support from CAPES Foundation, Ministry of Education of Brazil, under grant BEX 13594/13-3.

Dedication

To my parents,

Moacir Netto and Francisca Grisotto Netto.

Acknowledgments

First and foremost, I would like to extend my gratitude to my advisor, Dr. Lamine Mili. His guidance has shaped the way I approached this research topic. Further, I would like to express my most sincere appreciation to Dr. Yoshihiko Susuki for accepting the invitation to serve on my committee, and for providing me with the opportunity to visit his research group in Osaka. Our discussions on the Koopman operator have been enlightening and enjoyable. Furthermore, I am obliged to Professors Vassilis Kekatos, Virgilio Centeno, Daniel Stilwell, and Michael von Spakovsky, for kindly serving on my committee.

Dr. Robson C. Pires has been my Professor, mentor, and friend since I was an undergraduate student at the Federal University of Itajubá. He has awakened in me the desire to foster my academic endeavors, and I am wholeheartedly grateful.

I had the opportunity to discuss my research with Dr. Eyad H. Abed. I appreciate the time he invested in sharing with me his impeccable knowledge on participation factors.

I was also fortunate to interact with other colleagues working on the most diverse topics, encompassing a broad spectrum of knowledge. This has enriched my learning experience and made it a joyful process. I would like to acknowledge Ahmad A. Tbaileh, Chetan Mishra, Ibrahima Diagne, Jens D. Schoene, Junbo Zhao, Mohammad S. Hossan, Srivats Shukla, Venkat K. Krishnan, Yajun Wang, and Yijun Xu.

James (Jim) Murphy is undoubtedly my best friend at Virginia Tech. His faith in me and his continuous encouragement kept me going throughout this whole journey. Roxanne Paul is also a close friend and an adorable person. I am glad that they were around during these years that I have spent at Virginia Tech.

I found a place to call home with the Maher's. It was a privilege to share experiences with Anthony, Temy, Nicholas, and Natali.

Last, and most importantly, my family has been always the source of love and inspiration. They are always in my thoughts and in my heart.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Electric power systems are undergoing radical changes triggered by the advent of novel technologies in sensing and measuring, communications and networking, signal processing, power electronics, energy conversion and storage, and control.

In particular, the development of the phasor measurement unit (PMU) [1–3] represents a major advance to measuring technology. The data acquired from PMUs, hereafter referred to as *PMU measurements*, are significantly more informative than the data acquired from the legacy remote terminal units (RTUs) [4]. If compared to RTUs, the additional advantages offered by PMUs are the following:

1. Their reporting rate is at least one order of magnitude faster;
2. They provide measurements of voltage/current phase angles in addition to measurements of voltage/current magnitudes;
3. Measurements are time-tagged by civilian global positioning system (GPS) receivers.

Important advances of science and technology also led to the emergence of high-speed, large bandwidth communication networks. The expansion of these communication networks, along with the deployment of PMUs, is creating wide area measurement systems (WAMS) that offer the possibility of tracking the dynamics of the grid. Further, along with other grid-edge metering devices, such as smart meters [5,6] and micro-PMUs [7], they are giving rise to the era of *big data* in the electric power industry. This is fueling interest in data-driven methods that can enhance the resilience of power grids [8].

Simultaneously, and perhaps most importantly, we are witnessing an unprecedented increase of the share of renewable energy sources in the generation portfolio [9–12]. The intermittent and stochastic nature of renewable energy sources such as wind and solar is imposing new challenges to generation dispatch and unit commitment [13,14], reliability evaluation [15,16], and stability assessment [17, 18], to mention a few. The increase in the stochastic dynamics of the net load challenges the traditional deterministic, model-based methods applied in the legacy systems [19, 20].

Due to the exposed facts, there is a growing interest in exploiting fast sampled measurement data to advance power grid modeling, analysis, and control through data analytics [21–24]. Along these lines, the enhancement of the existing forecasting methods is an ongoing effort. In general, the model-based methods are being improved by two broad approaches, which are enhancing the visibility of various sources of uncertainties through advanced forecasting methods, and utilizing machine learning and data analytics to add situational awareness.

In the course of the aforementioned changes, the control centers of electric power systems shall keep playing an important role by centralizing intelligence, performing monitoring, and coordinating control actions to maintain the backbone of the grid. Concurrently, control actions shall be executed locally to the extent that it is possible, thereby providing the system with a hierarchical decentralized [25] structure. For example, while the control of inter-area oscillations [26] rely on wide-area observability and naturally requires centralization, the control of reactive reserves for voltage stability is an inherently localized problem and shall be addressed locally. In the work reported in this dissertation, we place ourselves at the *control center of an electric transmission system* from a functional standpoint, where a certain area is being monitored and controlled in real-time; this area is supposed to be completely observable by PMU measurements. Note that in the United States, Europe, and China, entire regions are already observable by PMU measurements. The next large-scale deployments of PMUs are expected to happen in Brazil and India.

In order to comprehend and address some of the challenges associated with this changing panorama, it is necessary to understand the basics of how the measurements processed at the control center translate into meaningful information. Hence, in what follows, we will dive into the details starting from the well-established concept of control centers of electric transmission systems. See Fig. 1.1. Then, we will gradually move toward the envisioned future where the system is fully observable by PMU measurements, and which constitutes this work's starting point. While discussing the details, we will highlight the challenges and opportunities, pinpoint the limits of the status quo, and motivate the research work reported in this dissertation. Recall that the control centers are considered the central nervous system of the electric power systems, from where real-time monitoring and control are accomplished with the aim of specific computer programs. In what follows, each one of these computer programs is referred to as an *application*. The energy management system (EMS) running in the control centers encompasses a suite of applications. The supervisory control and data acquisition (SCADA) system collects measurements from RTUs and intelligent electronic devices (IEDs) every second or so, including the status of circuit breakers and other switching equipment, tap position of transformers, voltage and current magnitudes, and real and reactive power. These measurements are stored in a real-time database and made available to the EMS.

In what follows, the major groups of applications that constitute an EMS [13, 27–29] are briefly described.

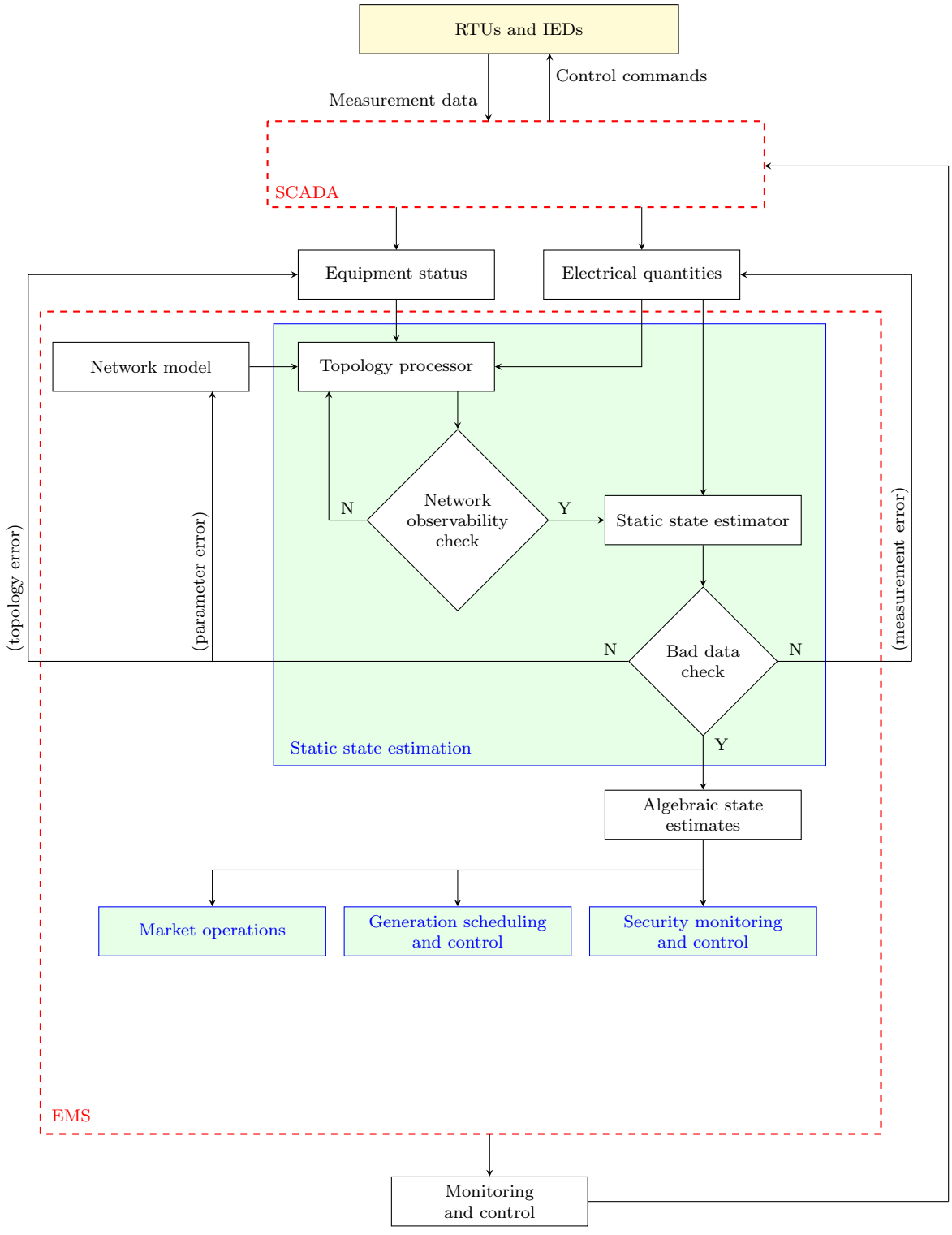


Figure 1.1: Functional schematic diagram of a SCADA/EMS.

Market operations

Applications developed for market operations are mainly devoted to business management. Examples include outage management, energy/reserve capacity market, and clearing market, to name a few. These applications constitute an important part of today's electric power exchange. More details can be found in [13, 25, 27–30] and references therein.

Generation scheduling and control

Generation scheduling and control applications essentially perform balancing between electric power demand and generation. They are also referred to as automatic generation control and include several applications, each of them with a specific and well-established role; *economic dispatch*, *unit commitment*, and *load frequency control* are typical examples. Other applications fall into this group, although they do not belong exclusively to it. For example, *short-term load forecasting* [19] is also of fundamental importance for security monitoring and control applications. Short-term load forecasting is typically carried out in intervals of several to tens of minutes and provides an envision of the close future, thereby allowing for preventive control actions to be taken in advance.

Note that only hydro-thermal generation are scheduled on legacy systems [13]; thus, the uncertainties are essentially associated with state estimation and load forecasting. This scenario is rapidly changing, as the deployment of renewables is adding uncertainties on the generation side. More details can be found in [13, 31, 32] and references therein.

Static state estimation

The topology processor and the state estimator are, together, "the front gate" of the control center; most of the applications running at the control center rely on their outcome [33, 34]. The *topology processor* retrieves data from the real-time database and processes it to determine the network topology. In simple words, it is responsible for creating the bus-branch model of the network, using the node-breaker network model existing in the EMS database. With the bus-branch model reflecting the topology of the network in real-time, along with redundant measurements of electrical quantities, the *state estimator* is able to determine the most likely state of the system.

As initially conceived by Schweppe et al. [35–37], the state estimator described at this point is based on an algebraic model of the electric power network. Therefore it is referred to as the *static state estimator* (SSE). It typically runs every minute or so, under the assumption that the system is in quasi-steady state. The SSE is only able to estimate the algebraic states of the network, i.e., voltage magnitudes and phase angles, and in some cases, tap position of transformers and the firing angle of power electronics converters. Formally, the SSE solves the following problem. For a fixed time $t \in \mathbb{R}$ with corresponding state $\boldsymbol{x}$, an element of the

state space $S \subset \mathbb{R}^n$,

$$\text{minimize} \qquad \sum_{i=1}^{m} \rho\left(r_{s_i}\right), \tag{1.1}$$

$$\text{subject to} \qquad \boldsymbol{y} = \boldsymbol{h}\left(\boldsymbol{x}\right) + \boldsymbol{r}_s, \tag{1.2}$$

where $\boldsymbol{y} \in \mathbb{R}^m$ denotes the system observations (or measurements), $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ is a vector-valued nonlinear function that relates the state, $\boldsymbol{x}$, to the observation, $\boldsymbol{y}$, $\boldsymbol{r}_s \in \mathbb{R}^m$ is a vector containing the standardized residues r_{s_i}, $i = 1, ..., m$.

The ρ-function in (1.1), also referred to as the loss function, depends on the choice of the estimator. For example, if the least squares estimator is adopted,

$$\rho\left(r_{s_i}\right) = r_{s_i}^2, \tag{1.3}$$

whereas if the *weighted* least squares estimator is adopted,

$$\rho\left(r_{s_i}\right) = \varpi_i r_{s_i}^2, \tag{1.4}$$

where ϖ_i denotes the i-th weight. Further, if the least absolute value (LAV) estimator [38–42] is chosen, then

$$\rho\left(r_{s_i}\right) = |r_{s_i}|, \tag{1.5}$$

where $|r_{s_i}|$ denotes the absolute value of r_{s_i}. The Huber estimator, widely used in the work reported in this dissertation, is defined as

$$\rho\left(r_{s_i}\right) = \begin{cases} \frac{1}{2} r_{s_i}^2, & |r_{s_i}| \leq \beta, \\ \beta |r_{s_i}| - \frac{1}{2}\beta^2, & \text{otherwise}, \end{cases} \tag{1.6}$$

where β is a parameter of choice. More specifically, a slightly modified version of (1.6) termed Schweppe-Huber estimator [43] is adopted. Formally:

$$\rho\left(r_{s_i}\right) = \begin{cases} \varpi_i^2 \left(\frac{1}{2} r_{s_i}^2\right), & |r_{s_i}| \leq \beta, \\ \varpi_i^2 \left(\beta |r_{s_i}| - \frac{1}{2}\beta^2\right), & \text{otherwise}. \end{cases} \tag{1.7}$$

In practice, the measured electrical quantities and statuses of switching equipment, as well as the parameters of the network model might be erroneous. Therefore, it is critical to detect, identify, and suppress the negative effect of outliers on the estimation process. This motivates the *robust static state estimators* [38–43]. The term robust comes from the fact that, up to a certain extent, they are able to suppress the undesired effect of the aforementioned errors in the estimation process, provided that there is enough redundancy. Note that, conceptually, the model itself is never questioned, which might not be reasonable when external network equivalents [44] are adopted, to mention an example. State estimation is a key component

of the work reported in this dissertation. It will be discussed in more details in section 1.1.1 and in the forthcoming chapters.

Security monitoring and control

Security monitoring applications essentially answer *what-if* types of questions. One typical application is the *steady-state contingency analysis*, where a list of credible contingencies is postulated, including generator and transmission line outages. Then, power flow is solved for each scenario in order to identify possible overloaded equipment and abnormal voltage levels along the system. Other applications responsible for enhancing *steady-state* and *dynamic* security monitoring might be executed as well.

If compared to off-line planning tools, real-time applications have stringent requirements. Firstly and foremost, the *computational speed* is critical as in real-time operation a delay in providing information can hide an imminent situation of risk, and the consequences might be severe. Secondly, the *robustness of the numerical solution* is fundamental as power system simulations are prone to divergence, especially in the case of very large networks; numerically ill-conditioned problems and model approximations are among the issues. In the planning department, these problems are not uncommon and can be handled by seasoned engineers. However, at the control center, there is not much time for analysis; an inappropriate outcome might mislead the operator to take a wrong decision. Lastly, *visualization* tools are very important; operators must be able to quickly interpret the situation and take actions in an assertive way. Single-line diagrams, nomograms [45], graphs and maps are normally adopted.

We shall mention that on an ordered time scale, security monitoring and control applications are a second line of defense. The first one is provided by the protective relays. Protective relays are mostly based on local measurements and oriented toward a single piece of equipment, e.g. generator, transformer, transmission line, busbar or feeder. They typically act in a few hundreds of milliseconds considering the entire time frame of actions, from the event detection to the completion of circuit breaker opening. Conversely, security monitoring and control applications are based on measurements collected throughout the system, available after being processed by the state estimator every minute or so. The time frame from state estimation, to the outcome of the security monitoring applications, to the decision making, to the control actions, is in the order of tens of minutes. There exists an information and control gap in the time comprehended between a few hundreds of milliseconds and tens of minutes. This gap, along with the lack of wide-area visibility of the system state, have been pinpointed as an important contributor to the missed opportunity in preventing cascading failures, such as the ones that led to the North American blackout and the Italian blackout in 2003 [29]. Security monitoring and control is also an important part of the work reported in this dissertation. It will be discussed in more details in Section 1.1.2 and in the forthcoming chapters.

To this point, we briefly looked into the conceptual role and available tools in the legacy control centers of electric transmission systems. Next, we look at the present situation with PMU measurements being incorporated.

1.1 Motivation

1.1.1 Power system dynamic state estimation

As aforementioned, the wide-scale deployment of PMUs is creating the opportunity to track the dynamics of power systems. Specifically, the dynamics of the synchronous generators and their controls, as well as the dynamics of the loads can be captured because they occur in the order of milliseconds. Conversely, the transients caused by switching and the dynamics of power electronics-based equipment occur in the order of microseconds to a few milliseconds and are not captured by PMUs.

PMU measurements are also susceptible to outliers, data dropouts, and impulsive noises. Therefore, the concept of robustness remains relevant when processing data at the control center. As before, a topology processor is required to determine the network topology, thereby allowing the state estimator to establish the most likely state of the system. In short, a *dynamic state estimator* (DSE) [46] is adopted in this case to estimate the system states and algebraic variables. Although the notion of a DSE is not recent [47], it is becoming realizable with the development of WAMS. In simple words, the DSE solves a prediction-correction problem, i.e.,

Prediction

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)), \tag{1.8}$$

where $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ is a vector-valued nonlinear function.

Correction

$$\text{minimize} \qquad \sum_{i=1}^{m} \rho(r_{s_i}), \tag{1.9}$$

$$\text{subject to} \qquad \boldsymbol{y} = \boldsymbol{h}(\hat{\boldsymbol{x}}(t)) + \boldsymbol{r}_s, \tag{1.10}$$

where $\hat{\boldsymbol{x}}(t)$ denotes the estimated value of $\boldsymbol{x}(t)$, obtained from the prediction step. The correction step of the DSE is similar to the SSE problem, apart from the fact that time is now explicitly taken into account.

The *static* and the *dynamic* state estimators are significantly different, as discussed next.

Firstly, their measurement scan rate is notably distinct as presented in Table 1.1. The scan rate of the DSE is at least 30 times faster as compared to the SSE, thus it has more stringent requirements on the communications network and it is more sensitive to time delays. Further, each k-th step of the prediction-correction process of the DSE must be numerically solved faster than its scan rate, thereby posing a strict requirement on the computational power.

Secondly, they differ on the required number of measurements of equipment status. Consider the example depicted in Fig. 1.2. If measurements of bus voltage magnitude and power

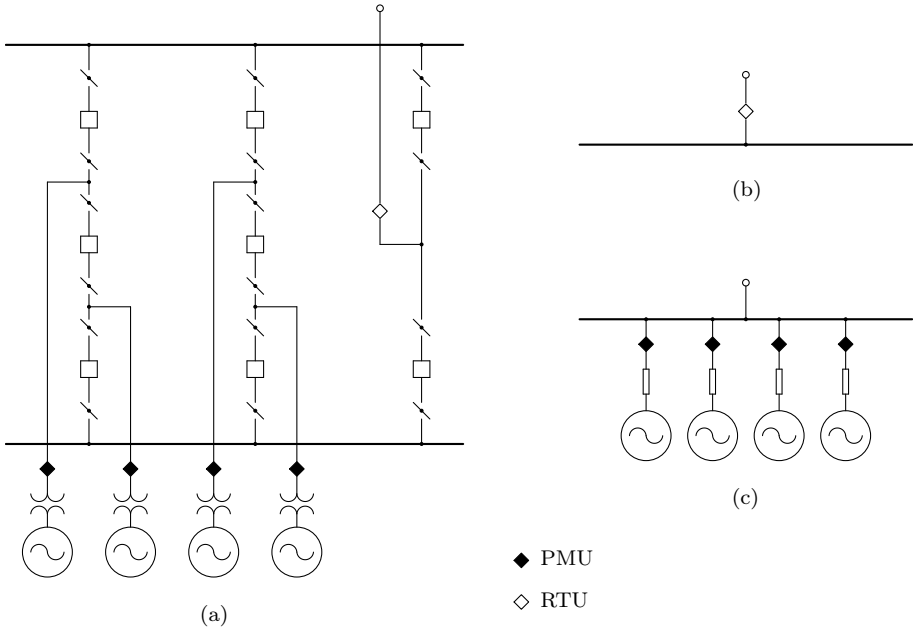


Figure 1.2: (a) Simplified node-breaker model of a typical power generation substation. (b) Bus-branch model of the static state estimator. (c) Bus-branch model of the dynamic state estimator.

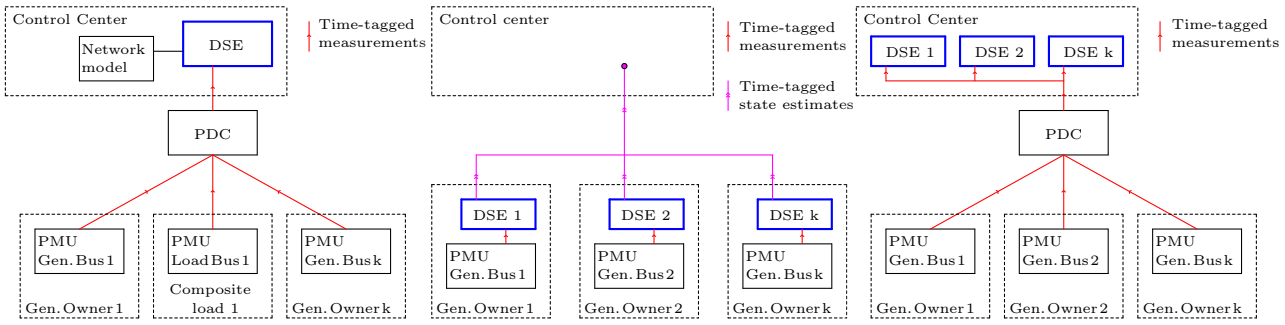
injection are collected at the point indicated in Fig. 1.2 as RTU, the status of the switching devices in the substation is irrelevant to the SSE. In other words, the entire substation reduces to a network node with a fixed voltage magnitude, to which an electric power is injected to. Of course, this fact alleviates the computational burden of the topology processor. Conversely, in the case of the DSE, measurements of all the switching devices in the substation are required, in addition to measurements at each generator terminal, as indicated in Fig. 1.2 as PMU.

Thirdly, they differ on the model. For example, in the SSE, the generators are simply modeled by (scalar) constant power injections in the equality constraints of the minimization problem. In the DSE, on the other hand, the generators and their controls are represented by a set of differential algebraic equations, thereby having more stringent requirements on the numerical solution. Further, for the DSE, the dimension of the model and of the set of parameters is much larger.

Table 1.1: Measurement scan rate of state estimators

	Instrumentation and measurement system	Scan rate (seconds)
Static state estimator	RTUs and IEDs	≥ 1
Dynamic state estimator	PMUs and phasor data concentrators (PDCs)	$\leq 1/30$

Lastly, as opposed to the SSE which is centralized, the DSE can either be centralized or



(a) Centralized DSE (b) Decentralized DSE (c) Centralized distributed DSE

Figure 1.3: Conceptual comparison among different dynamic state estimators.

decentralized. In practice, however, there are important challenges associated with both, centralized and decentralized DSEs. We will discuss this aspect in more details next.

Centralized DSE

Following the work of Modir and Schlueter [47], several DSEs [48–51] have been proposed based on the conceptual idea of transmitting all the PMU measurements to a centralized location, say the control center, where the DSE is supposedly installed. The assumption is that the system is observable by PMUs. This configuration is referred to as centralized DSE, see Fig. 1.3a. Due to the computational burden associated with the prediction step of centralized DSEs, researchers have been adopting the Kron-reduced network model (KNM) [52]. However, contrary to independent system operators (ISOs) and regional transmission organizations (RTOs) that typically have access to the KNM, this is a hurdle for local utilities because of limited data sharing across neighboring systems. Further, the KNM assumes that the electric loads are modeled as constant admittances, thereby not allowing for capturing the rich dynamics of composite electric loads [53]. Another important point that has been overlooked is that a topology processor is required to determine the KNM in real time; the existing topology processors embedded on the SSEs cannot supply this demand. Nevertheless, although of reduced dimension, the graph of the KNM is full [52] rather than sparse, hence sparsity techniques cannot be exploited. An alternative to the KNM is dynamic reduced order models [54], but unfortunately they come at the price of increasing the existing model uncertainties. The computational burden is another important barrier to the adoption of centralized DSEs. A robust centralized DSE [50, 51] will be presented in chapter 4.

Decentralized DSE

As an attempt to overcome the aforementioned challenges, decentralized DSEs have been investigated [55,56]. The idea is to individually estimate the states of each generator at their own location. See Fig. 1.3b. Because a decentralized DSE is designed for each generator (or equivalent generator), only a PMU at the generator terminal is required. The decentralized

DSE requires neither the model of the network nor the model of the electric loads, and it is computationally inexpensive because of a reduced number of involved state variables. Further, it offers the additional benefit of performing model calibration [57] without taking the generator offline, which is an appealing attribute from the system's reliability and economic standpoints. However, although a decentralized DSE allows for *local* stability assessment, control, and protection, for wide-area stability assessment and control, the DSE states will still need to be communicated to centralized processing, like any other architecture shown in Fig. 1.3; and, although the data transmissions in this case are not raw measurements but actual system states, it is still necessary to ensure data integrity, cybersecurity, and robustness against bad or missing information at the receiving end, i.e., at the control center. Additionally, in the event of loss of communication with a decentralized DSE, the states of the associated generator must be forecast. To circumvent some of these issues, Paul et al. [58] proposed a computationally distributed but physically centralized DSE, wherein the conceptual idea is to send all the measurements to a centralized location instead of performing the estimation separately for each generator; see Fig. 1.3c. Although ingenious, the approach in [58] relies on the reference signals of voltage and power that are accessible only to ISOs and RTOs, thereby precluding its adoption by transmission companies and alike utilities.

In fact, most of the previously discussed challenges are either directly or indirectly related to modeling. Given the promise that decentralized DSEs have in terms of reducing the computational burden on increasing proliferations of sensors and data, and given its favorable trade-off between local and wide-area assessments, a robust hierarchical decentralized DSE based on the Koopman operator-theoretic framework is advocated in chapter 5.

1.1.2 Power system monitoring and control

The bottleneck preventing the fulfilling of the aforementioned control gap is *time*. Next, we will elaborate to support this claim. The focus must certainly be on the *dynamic* security monitoring applications, also referred to as *dynamic security assessment* [59]. Steady-state security assessment will not be discussed, the interested reader is pointed to [31].

Dynamic security assessment

The applications based on time-domain simulations are among the most computationally intensive ones running in the EMS. Despite the fact that the computational power has significantly increased over the years, it is still challenging to perform real-time dynamic security assessment of very large networks. For example, analyzing post-fault stability of 5,000 different system operating points for a list of 1,000 contingencies requires 5 million dynamic simulations [60]. Different strategies have been used in order to handle (or reduce) the burden of carrying out this task, among which two should be highlighted. The first strategy focuses on the computational architecture, typically parallel and distributed processing [60, 61]. The second strategy focuses on dynamic reduced order models [54] that can be processed much faster. These strategies have enabled the computation of dynamic secu-

rity regions for very large systems; for details about the strategies being adopted in different countries, refer to [60, 62].

An approach to address the above challenges centers around the idea that power system transient stability can be understood as the separation of two generation areas. See [62, 63]. Although cascading effects might lead to further separations, it is claimed that it is of interest to avoid the initial separation, which is caused by a power imbalance in which generators of one area accelerate (or decelerate) with respect to the other. This leads to the concept of two coherent groups of generators, denominated *critical* and *noncritical* clusters. The critical cluster is defined to be the one with smaller inertia. If the critical cluster accelerates with respect to the noncritical cluster, it is said that it swings forward; if it decelerates then it swings backward. The importance of knowing whether the critical cluster swings forward (backward) lies on the fact that the corrective measure would be to reduce (increase) its generation. Although this assumption might be verified in some practical cases, it does not hold true in general. In fact, the results demonstrated in [64] offer a contradiction to such an assumption, since they show that the coherent groups of generators oscillating together are dependent on the disturbance location. In short, the researchers and practitioners working along these lines were essentially motivated by the possibility of adopting the equal-area criterion to assess rotor angle transient stability; and the reason why they would like to rely on such index is *time*. The equal-area criterion can be computed fast, even for very large systems.

Due to the inherent computational burden associated with dynamic security assessment, only *preventive* control actions can be performed, i.e., changing the operating point *before* a fault happens to ensure that the system can withstand any *a priori* listed credible disturbance. There is a lack of *corrective* real-time control schemes, i.e., *following* a disturbance. Although special protection schemes are designed to be corrective rather than preventive, they are *ad hoc* measures mainly based on protection relays. These are the result of limitations imposed in the past by the available technology [65].

Along these lines, control schemes based on WAMS might be more effective on improving the dynamic stability performance of power systems, particularly for inter-area oscillations. In order to be effectively damped, the modes of oscillation have to be both observable and controllable. The availability of wide-area measurements enables the choice of control inputs carrying important information about inter-area oscillations that cannot be extracted from local signals alone. In other words, the effectiveness of a conventional stabilizer using local measurements might be limited by the relative amplitude of low-frequency governor and local electromechanical modes. Using remote measurements as inputs, as opposed to local measurements only, increases the stabilizer's response to inter-area modes. If poor damping levels for the inter-area oscillatory modes are detected, control schemes might ultimately act tuning power systems stabilizers (PSS), static VAR compensators and high-voltage direct current links.

In fact, it is widely recognized by researchers and practitioners that most of the real-time

countermeasures/controls will require WAMS, with intensive use of communication networks. However, the computational burden and the uncertainties associated with model-based approaches will preclude their adoption for this problem. Now, given the availability of massive amounts of data, the question is *can we extract information and perform control actions without the use of models?* Or perhaps *can we extract the model itself from real-time measurements?* Aligned with this idea, the Koopman operator-theoretic framework has been shown to have interesting properties and potential for integration of data analysis into our mathematical knowledge of electric power systems.

1.2 Research objective and significance

The research objective is *to develop a robust, data-driven framework for system identification, dynamic state estimation, and control of electric power systems using the Koopman operator-theoretic framework.*

The significance of this study lies on the fact that the Koopman operator-theoretic framework might offer an alternative solution to the lack of *corrective* controls in electric power systems, which has contributed to the missed opportunity of preventing the occurrence of large scale blackouts. This framework aligns with and complements the vision of autonomous energy grids [66, 67].

1.3 Summary of achievements

- A robust extended Kalman filter was developed [50, 51], building upon the work of Gandhi and Mili [68]. The developed method was applied to dynamic state estimation of electric power systems, despite of being applicable to other dynamical systems.
- A robust, data-driven dynamic state estimator based on the Koopman operator-theoretic framework was developed [69, 70], following the work of Surana et al. [71–73].
- A robust filter that makes use of time correlation to identify and suppress outliers in magnitude and phase angle of complex-valued datasets was developed. This robust filter is an extension of the work of Tamburello and Mili [74–76], and it was applied in association with the Prony method to identify electromechanical modes of oscillation in electric power systems [77, 78].
- A hybrid framework combining model-based and data-driven methods for hierarchical decentralized, robust dynamic state estimation was originally proposed [79].
- A data-driven approach for the computation of modal participation factors for nonlinear systems based on the Koopman operator-theoretic framework was developed [80].

1.4 Organization of the dissertation

This dissertation proceeds as follows.

Chapter 2 provides an introduction to the Koopman operator theory, and to the numerical methods available to estimate the Koopman modes, eigenvalues, and eigenfunctions.

Chapter 3 gives a formal treatment of the Kalman filter (KF) from the perspective of statistics and estimation theory. A robust generalized-maximum likelihood Kalman filter is presented in sequence. Then, the extended Kalman filter (EKF) is briefly introduced, followed by the iterated extended Kalman filter, which provides a way of reducing the linearization errors inherent to the EKF.

Chapter 4 presents the model-based, centralized, robust DSE that relies on the developed robust extended Kalman filter. A discussion on the breakdown point of the estimator, as well as its computational efficiency is also provided.

Chapter 5 develops an alternative, robust DSE that is data-driven and relies on the Koopman operator-theoretic framework. Centralized and hierarchical decentralized configurations were investigated.

Chapter 6 proposes a robust filter based on the iteratively reweighted phase-phase correlator that is able to identify and suppress outliers in the magnitude and in the phase angle of complex-valued data-sets.

Chapter 7 introduces a novel data-driven method to compute modal participation factors for nonlinear systems based on the Koopman operator-theoretic framework, which is an important step toward control applications of the Koopman operator theory.

Chapter 8 presents the final conclusions and provide directions of future research.

Chapter 2

Koopman operator

The Koopman operator is at the core of the work reported in this dissertation. Its theory has been mostly developed within the fluid mechanics community. Perhaps the reason is that the study of fluid dynamics is often times an ill-posed problem in the sense of Hadamard. In particular, the dynamics of some fluid systems are so complex that there is no model available. Hence, researchers within the fluid mechanics community are used to the idea of extracting information from measurements to perform analysis and control; this is frequently their only option. Within the electric power systems community, models of certain fidelity are available and their performance is acceptable to a certain extent. However, due to the undergoing modernization of the electric grids that is taken place in the U.S., this scenario is changing fast and there is a growing interest on data-driven approaches. Yet, the Koopman operator is relatively less known within the electric power systems community. Thus, we start this chapter by presenting a brief historical development of this theory based on [81].

2.1 Historical evolution of the Koopman operator theory

The Koopman operator formalism originated in the early work of Bernard O. Koopman [82], in 1931 [83]. He introduced the linear transformation that we now call the Koopman operator, and realized that this transformation is unitary for Hamiltonian dynamical systems; the "U" notation comes from the unitary property. This observation by Koopman inspired John von Neumann [84] to give the first proof for a precise formulation of ergodic hypotheses, known as the mean ergodic theorem. In the next year, they wrote a paper together, in which they introduced the notion of the spectrum of a dynamical system, i.e. the spectrum of the associated Koopman operator, and noted the connection between chaotic behavior and the continuous part of the Koopman operator's spectrum [85]. Interestingly, during that period, other mathematicians were also studying nonlinear dynamical systems using the theory of *linear* differential equations via nonlinear transformations. For example, the

Swedish mathematician Torsten Carleman [86] was developing the theory of what we now call the Carleman linearization [87].

For several decades after the work of Koopman and von Neumann, the notion of Koopman operator was mostly limited to the study of measure-preserving systems. One could find it as the unitary operator in the proof of the mean ergodic theorem, on discussions on the spectrum of measure-preserving dynamical systems [88–90], or associated with Lie algebra [87]. It seldom appeared in other applied fields until it was brought back to the general scene of dynamical systems by Mezić and Banaszuk [91] and Mezić [92]. The first paper showed how to construct important objects like invariant sets in high-dimensional state spaces from data. It also emphasized the role of nontrivial eigenvalues of the Koopman operator to detect periodic trends of dynamics amidst chaotic data. The second paper discussed the spectral properties of the Koopman operator further, and introduced the notion of Koopman modes. Both papers also discussed the idea of applying the Koopman operator-theoretic framework to capture the regular components of data in systems with combination of chaotic and regular behavior.

In 2009, the idea of Koopman mode decomposition (KMD) was applied to a complex fluid dynamical system, namely, a jet in a cross flow [93]. This work showed the promise of KMD in capturing the dynamically relevant structures in the flow and their associated time scales. Unlike other decomposition techniques, KMD combines two advantageous properties: it makes a clear connection between the measurements in the physical domain and the dynamics in the state space, unlike proper orthogonal decomposition, and it is completely data-driven, unlike the global mode analysis. The work in [93] also showed that KMD can be computed through a numerical decomposition technique known as dynamic mode decomposition (DMD) [94]. A review of the Koopman theory applied to fluid mechanics can be found in [95].

In the recent years, the extent of KMD applications for data-driven analysis has enormously grown. In power systems, numerical results based on the analysis of the point spectrum of the Koopman operator were first demonstrated by Susuki and Mezić [64], with an application on coherency identification. Other applications have also been exploited, including the development of a precursor to rotor angle instability [96], a model-free technique for the assessment of transient stability [97], a strategy for power grid islanding [98], a dynamic state estimator [69,70,79], and a data-driven approach for the computation of modal participation factors for nonlinear systems [80].

2.2 Dynamical systems theory

At the time of Newton, mathematicians and physicists discovered that differential equations provide the precise tool to describe the laws of physics as applied to specific systems. Relatively simple systems can be described by a finite number of variables, say $\boldsymbol{x} = [x_1\ x_2\ ...\ x_n]^\top$,

called the *state*. The differential equation must then represent the physical laws governing the evolution of the system. It is required that the knowledge of this state, as initial condition, will be sufficient, together with the differential equation, to determine the future evolution of the system uniquely. The mathematical description of such a system is therefore given by a differential equation of the type

$$\dot{\boldsymbol{x}} = \boldsymbol{f}\left(t, \boldsymbol{x}\right). \tag{2.1}$$

The vector-valued function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$, representing the laws governing the evolution of the system, is assumed to be known. It must be such that, together with the initial conditions, $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$, determines the solution $\boldsymbol{x}(t)$ uniquely.

If the function $\boldsymbol{f}$ in (2.1) does not depend on t, i.e., is of the form

$$\dot{\boldsymbol{x}} = \boldsymbol{f}\left(\boldsymbol{x}\right), \tag{2.2}$$

then the system is said to be invariant in time and is called *autonomous*. For an autonomous system, if $\boldsymbol{x}(t)$ is a solution, then $\boldsymbol{x}(t - t_0)$ is also a solution for any t_0. In other words, the solutions are translation invariant.

Within the context of their mathematical study, autonomous systems are also called *dynamical systems* [99]. A dynamical system, in the abstract sense, consists of two things: a set of states through which we can describe the evolution of a system, and a rule for that evolution. Although this viewpoint is very general and might be applied to almost any system that evolves with time, often the fruitful and conclusive results are achievable when we pose some mathematical structure on the dynamical system. For example, we often assume that the set of states form a linear space with nice geometric properties, and that the rule of evolution has some order of regularity on that space. The prominent examples of such dynamical systems are amply found in physics, where we use differential equations to describe the change of physical variables with respect to time. In what follows, we specially focus on dynamical systems that can be represented as in (2.2). Formally,

$$\dot{\boldsymbol{x}} = \boldsymbol{f}\left(\boldsymbol{x}\right), \tag{2.3}$$

where the state $\boldsymbol{x} \in S \subset \mathbb{R}^n$, and $\boldsymbol{f} : S \to \mathbb{R}^n$ is a vector field on the state space. Occasionally, we will specify some regularity conditions for $\boldsymbol{f}$, like being smooth or differentiable. We will also consider dynamical systems given by the discrete time map

$$\boldsymbol{x}_{k+1} = \boldsymbol{T}\left(\boldsymbol{x}_k\right), \tag{2.4}$$

where $\boldsymbol{T} : S \to S$ is the map, and $k \in \mathbb{Z}$ is the discrete time index. Just like the continuous time system in (2.3), we might need to make some extra assumptions on $\boldsymbol{T}$. The discrete time representation of dynamical systems does not often show up in physical systems, but we can use it to represent continuous time systems, for example, through discrete time sampling. This representation has the benefit of being more practical because the data collected from dynamical systems often times comes in the form of discrete time samples.

The study of the dynamical systems in (2.3) and (2.4) has been dominated by the geometric viewpoint in much of the last century. In this viewpoint, originally due to Henri Poincaré [100], the qualitative properties of the solution curves in the state space are studied using geometric tools and the emphasis is put on the subsets of the state space that play an important role in the asymptotic behavior of the trajectories. We briefly describe some concepts from this theory here, but a more comprehensive exposition can be found in [101, 102].

Assuming that the solution to (2.3) exists, we define the solution map (also referred to as the flow map, or simply the flow) $F^t : S \to S$ to be the map that takes the initial state to the state at time t, i.e.,

$$F^t(\boldsymbol{x}_0) = \boldsymbol{x}_0 + \int_{\boldsymbol{x}_0, t'=0}^{t} \boldsymbol{f}(\boldsymbol{x}(t'))\, dt'. \tag{2.5}$$

The solution map satisfies the semi-group property, i.e., for every $t_2 > t_1 \geq 0$,

$$\begin{aligned} F^{t_2} \circ F^{t_1}(\boldsymbol{x}_0) &= F^{t_1}(\boldsymbol{x}_0) + \int_{F^{t_1}(\boldsymbol{x}_0), t'=0}^{t_2} \boldsymbol{f}(\boldsymbol{x}(t'))\, dt', \\ &= \int_{\boldsymbol{x}_0, t'=0}^{F^{t_1}(\boldsymbol{x}_0), t_1} \boldsymbol{f}(\boldsymbol{x}(t'))\, dt' + \int_{F^{t_1}(\boldsymbol{x}_0), t'=0}^{t_2} \boldsymbol{f}(\boldsymbol{x}(t'))\, dt', \\ &= \int_{\boldsymbol{x}_0, t'=0}^{t_1+t_2} \boldsymbol{f}(\boldsymbol{x}(t'))\, dt', \\ &= F^{t_1+t_2}(\boldsymbol{x}_0), \end{aligned} \tag{2.6}$$

where $\circ$ denotes the composition operator.

Some of the important geometric objects in the state space of continuous time dynamical systems are briefly introduced next.

2.2.1 Elementary concepts of dynamical systems

Phase space (or state space)

The phase space, also referred to as the state space, is a space wherein all possible states of the system are represented. Each one of the possible states corresponds to a unique point in the phase space.

Orbit (or trajectory)

An orbit is a set of points in the state space that are related by the function that defines the evolution of the dynamical system. An orbit is also a subset of the phase space, that is, different orbits do not intersect in the phase space. The set of all orbits is a partition of the phase space. The terms *orbit* and *trajectory* are often used interchangeably.

Fixed point (or equilibrium point)

A fixed point is any point $\boldsymbol{x}$ in the state space for which $\boldsymbol{f}(\boldsymbol{x}(t)) = \mathbf{0}$ or $F^t(\boldsymbol{x}) = \boldsymbol{x}$. In other words, a fixed point is a contant orbit. The fixed points correspond to the equilibria of physical systems. An important notion with regard to fixed points is the notion of stability, that is, whether the trajectories starting in some neighborhood of a fixed point stay in its neighborhood over time or not. Suppose that the system is slightly displaced from a fixed point. Then, it might move back to the fixed point, in which case the system is said *asymptotically stable*; or it might move away, in which case it is said *unstable*; or it might move in a neighborhood of the fixed point but not approach it, in which case it is said *stable* but not asymptotically stable. Fixed points are also called equilibrium points.

Periodic orbit

A periodic orbit is a special type of solution for a dynamical system, that is, a solution which repeats itself in time. A periodic orbit in the plane is termed a limit cycle.

Limit cycle

Limit cycles are (isolated) closed curves in the state space which correspond to the time-periodic solutions of (2.3). The generalized version of limit cycles are tori, Cartesian products of circles, which are associated with quasi-periodic motion.

Invariant set

An invariant set B in the state space satisfies $F^t(B) \subseteq B$ for all t, i.e., the trajectories starting in B remain in B. The invariant sets are important because we can isolate the study of the dynamics on them from the rest of the state space, and they include other important objects such as fixed points, limit cycles, attractors and invariant manifolds.

Attractor

An attractor is an attracting set with a dense orbit. An attracting set is an invariant subset of the state space to which many initial conditions converge. A dense orbit is a trajectory that comes arbitrarily close to any point on the attracting set. For example, a stable fixed point, or stable limit cycle, is an attractor; conversely, the union of two separate stable periodic orbits is an attracting set but not an attractor, because the trajectories on one periodic orbit cannot come arbitrarily close to the other periodic orbit. A more complicated example of attractor is the famous butterfly-shaped set in the chaotic Lorenz system which is called a strange attractor.

Attractors are the objects that determine the asymptotic, that is post-transient or long-term, dynamics of dissipative dynamical systems. In fact, the mere notion of dissipativity, which we can think of as shrinkage in the state space, is enough to guarantee the existence of an attractor in many systems [103]. In some cases, the state space contains more than one attractor, and the attractors divide the state space into basins of attraction; any point in the basin of attraction of an attractor will converge to it over infinite time.

Basin of attraction

The basin of attraction is a region of the phase space that surrounds an attractor. Any point inside this region will eventually converge to the attractor. In the case of a stable linear system, every point in the phase space is in the basin of attraction. Conversely, in the case of nonlinear systems, a point in the phase space might map to infinity, lie in a different basin of attraction and thus converge to a different attractor, or map to a non-attracting point or cycle where it will remain.

Hartman-Grobman Theorem

The asymptotic stability of fixed points of nonlinear systems can be established by relying on the Hartman-Grobman theorem [104]. The original Hartman-Grobman theorem for nonlinear systems is local [102], in the sense that we know a conjugacy exists for some neighborhood of the fixed point. The Koopman operator theory has been used to generalize the Hartman-Grobman theorem, thereby providing a linearizing coordinate transform that extended the conjugacy to the entire basin of attraction of a stable or unstable fixed point or periodic orbit [105, 106].

Bifurcation

Bifurcation is a change in the qualitative behavior of all the trajectories due to the changes in the vector field $\boldsymbol{f}$. For example, if we add some forcing term to the vector field $\boldsymbol{f}$, a stable fixed point might turn unstable or a limit cycle might appear out of the blue.

Having introduced some of the most elementary concepts of dynamical systems, we now discuss the classical approach to study dynamical systems.

2.2.2 Classical approach to the study of dynamical systems

In the classical approach, we start by discovering or building a model of the form of (2.3) or of (2.4). Sometimes, it is possible to find analytic or approximated solutions and use them to analyze the system dynamics, i.e., finding the attractors, invariant manifolds, imminent bifurcations, and so on. A lot of times, this is not possible and it is necessary to use various estimates or approximation techniques to evaluate the qualitative behavior of the system, for example, constructing Lyapunov functions to prove the stability of a fixed point. But most of the times, if we want a quantitative analysis or prediction, we have to employ numerical computation and then extract information by looking at a collection of trajectories in the state space.

This approach has contributed the most to the knowledge of dynamical and physical systems, but it is falling short in treating the high-dimensional systems that lie in the forefront of science and technology. A set of classic examples, which regularly arises in physics, is the set of systems that are governed by partial differential equations. In these systems, the state space is infinite-dimensional and the adopted numerical models might have up to billions

of degrees of freedom. Some examples of more recent interest include climate system of the earth, smart cars and buildings, electric power systems, and biological systems with interacting components like neural networks. The first problem with the traditional approach is that simulating the evolution of trajectories for these systems is just devastating due to the large size of the problem. Moreover, except for 2- or 3-dimensional systems, the geometric objects in the state space are difficult to realize and verify. The second problem is the uncertainty in the models or even the sheer lack of a model for simulation or analysis–in particular to electric power systems, model uncertainty is a timely topic due to the rapid growth of the number of renewable energy sources being deployed in the grid. As a result, the field of dynamical analysis has started shifting toward a less model-based and more data-driven perspective. This shift is not only motivated by the need to overcome the above problems, but it is also fueled by the increasing amount of data produced by today's powerful computational resources and experimental apparatus. In the next section, we introduce the Koopman operator theory, which is a promising framework for the integration of data analysis into the mathematical knowledge of dynamical systems.

2.3 Koopman operator

In the context of dynamical systems, we interpret the data as knowledge of some variable(s) related to the state of the system. A natural way to put this into the mathematical form is to assume that data is knowledge of variables which are functions of the state. We call these functions *observables* of the system. Let us discuss an example. The motion of the synchronous generators of an electric power system constitutes a dynamical system. For now, let us neglect the existence of the generator's controllers such as the governor and the excitation systems. One way to realize the state space is to think of it as the set of all generators' rotor acceleration that satisfy the Newton's second law for rotation. The generator rotors' speed changes with time according to a rule of evolution which is the swing equation given in Appendix A.1. Some examples of observables on this system are voltage magnitude/phase angle at given buses, and electric power injected into the system by the generators. In all these examples, the knowledge of the state, i.e. the generators' rotor angle and speed, uniquely determines the value of the observable. We see that this definition allows us to think of the data from most of the electric power systems measurements and simulations as values of observables. We also note that there are some type of data that don't fit the above definition as an observable of the system. For example, the generators' field current are not observables of the above system, since they cannot be determined by mere knowledge of the rotor angle and speed. In this case, we can alter our dynamical system to include that observable as well: we define a second dynamical system, given in Appendix A.2, which would explain all the above observables. We observe that determining the notion of observables on a system requires a careful consideration of the underlying processes that affects those observables.

In light of the above discussion, we can formulate the data-driven analysis of dynamical systems as follows: Given the knowledge of an observable in the form of time series generated by experiment or simulation, what can we say about the evolution of the state? The Koopman operator theory provides a solution framework by describing the precise relationship between the evolution of observables and the evolution of state.

2.3.1 Koopman operator for continuous time dynamical systems

Consider the continuous time dynamical system given in (2.3). We define $g : S \to \mathbb{R}$ to be a real-valued observable on this dynamical system. The collection of all such observables forms a linear vector space. We can define a one-parameter semi-group of Koopman operators, denoted by $\{U^t\}_{t\geq 0}$, where each element of this semi-group is given by

$$U^t g\left(\boldsymbol{x}\right) = g \circ F^t\left(\boldsymbol{x}\right) = g\left(F^t\left(\boldsymbol{x}\right)\right), \tag{2.7}$$

where $F^t\left(\boldsymbol{x}\right)$ is the solution map as defined in (2.5). The linearity of U^t follows from the linearity of the composition operator, i.e.,

$$U\left[g_1 + g_2\right]\left(\boldsymbol{x}\right) = \left[g_1 + g_2\right] \circ \boldsymbol{f}\left(\boldsymbol{x}\right) = g_1 \circ \boldsymbol{f}\left(\boldsymbol{x}\right) + g_2 \circ \boldsymbol{f}\left(\boldsymbol{x}\right) = Ug_1\left(\boldsymbol{x}\right) + Ug_2\left(\boldsymbol{x}\right). \tag{2.8}$$

The semi-group property of $\{U^t\}_{t\geq 0}$ follows from the semi-group property of the solution map for autonomous dynamical systems given in (2.6), i.e.,

$$U^{t_2}U^{t_1}g\left(\boldsymbol{x}\right) = U^{t_2}g \circ F^{t_1}\left(\boldsymbol{x}\right) = g \circ F^{t_2} \circ F^{t_1}\left(\boldsymbol{x}\right) = g \circ F^{t_1+t_2}\left(\boldsymbol{x}\right) = U^{t_1+t_2}g\left(\boldsymbol{x}\right). \tag{2.9}$$

2.3.2 Koopman operator for discrete time dynamical systems

Consider the discrete time dynamical system given in (2.4). The Koopman operator, U, given by

$$Ug\left(\boldsymbol{x}\right) = g \circ \boldsymbol{T}\left(\boldsymbol{x}\right) = g\left(\boldsymbol{T}\left(\boldsymbol{x}\right)\right), \tag{2.10}$$

is a linear transformation on the vector space formed by g. Again, the linearity of the Koopman operator follows from the linearity of the composition operation, i.e.,

$$U\left[g_1 + g_2\right]\left(\boldsymbol{x}\right) = \left[g_1 + g_2\right] \circ \boldsymbol{T}\left(\boldsymbol{x}\right) = g_1 \circ \boldsymbol{T}\left(\boldsymbol{x}\right) + g_2 \circ \boldsymbol{T}\left(\boldsymbol{x}\right) = Ug_1\left(\boldsymbol{x}\right) + Ug_2\left(\boldsymbol{x}\right). \tag{2.11}$$

2.3.3 Existence and uniqueness of the Koopman operator

In general, the Koopman operator exists by definition. In a discrete time setting, for instance, the Koopman operator associated with a discrete time map $\boldsymbol{T}$ is the operator U such that

$Ug = g \circ \boldsymbol{T}$. This operator always exists as long as $\boldsymbol{T}$ exists, and it is unique as long as $\boldsymbol{T}$ is unique.

In summary, we can think of the Koopman operator viewpoint as a lifting of the dynamics from the state space to the space of observables. The advantage of this lifting is that it provides a linear rule of evolution - given by Koopman operator - while the disadvantage is that the space of observables is infinite dimensional. In the next section, we discuss the spectral theory of the Koopman operator which leads to linear expansions for data generated by nonlinear dynamical systems.

2.4 Linear expansion in Koopman tuples

A somewhat naive but useful way of thinking about linear operators is to imagine them as infinite-dimensional matrices. Then, just like matrices, it is always useful to look at the their eigenvalues and eigenvectors since they give a better understanding of how a matrix acts on a vector space.

Now, let $\mathcal{F}$ be the space of observables, which are scalar-valued functions $\theta : S \to \mathbb{C}$ defined on the state space S. In what follows, we assume that the observables are continuous, i.e., $\mathcal{F} \subset \mathcal{C}^0(S)$.

2.4.1 Koopman linear expansion for continuous time dynamical systems

Consider the dynamical system given by (2.3). We call the pair (ϕ_i, λ_i) an eigenfunction-eigenvalue pair of the Koopman operator if they satisfy

$$U^t \phi_i = e^{\lambda_i t} \phi_i. \tag{2.12}$$

The set of all Koopman eigenvalues λ_i, $i = 1, 2, ...$, is referred to as the point spectrum of the Koopman operator [92]. Note that the Koopman operator might also have residual and continuous parts of spectrum.

An interesting property of the Koopman eigenfunctions which we will use later is that if (ϕ_i, λ_i) and (ϕ_j, λ_j) are two distinct eigenfunction-eigenvalue pairs, then

$$U^t (\phi_i \cdot \phi_j) = (\phi_i \cdot \phi_j) \circ F^t = \left(\phi_i \circ F^t\right) \cdot \left(\phi_j \circ F^t\right) = \left(U^t \phi_i\right) \cdot \left(U^t \phi_j\right) = e^{(\lambda_i + \lambda_j)t} (\phi_i \cdot \phi_j), \tag{2.13}$$

and $(\phi_i \cdot \phi_j, \lambda_i + \lambda_j)$ is also an eigenfunction-eigenvalue pair.

For now, let us assume that all the observables of the dynamical system lie in the linear span

of such Koopman eigenfunctions, that is,

$$g\left(\boldsymbol{x}\right)=\sum_{i=0}^{\infty}\upsilon_i\phi_i\left(\boldsymbol{x}\right), \tag{2.14}$$

where υ_i are the coefficients of the expansion, so-called Koopman modes [93]. Then we can describe the evolution of observables as

$$U^t g\left(\boldsymbol{x}\right)=\sum_{i=0}^{\infty}\upsilon_i e^{\lambda_i t}\phi_i\left(\boldsymbol{x}_0\right), \tag{2.15}$$

which says that the evolution of g has a linear expansion in terms of the Koopman eigenfunctions. If we fix the initial state $\boldsymbol{x}=\boldsymbol{x}_0$, we see that the signal generated by measuring g over a trajectory, which is given by $U^t g\left(\boldsymbol{x}_0\right)=g\circ F^t\left(\boldsymbol{x}_0\right)$ is the sum of (an infinite number of) sinusoids and exponentials.

2.4.2 Koopman linear expansion for discrete time dynamical systems

Now, consider the dynamical system given by (2.4). The Koopman eigenvalues and Koopman eigenfunctions are defined as follows:

$$U\phi_i=\mu_i\phi_i, \tag{2.16}$$

implying that (2.12) and (2.16) are equivalent with $\mu_i^k=e^{\lambda_i t}$ for fixed t.

Here, if (ϕ_i,μ_i) and (ϕ_j,μ_j) are two distinct eigenfunction-eigenvalue pairs, then

$$U\left(\phi_i\cdot\phi_j\right)=\left(\phi_i\cdot\phi_j\right)\circ\boldsymbol{T}=\left(\phi_i\circ\boldsymbol{T}\right)\cdot\left(\phi_j\circ\boldsymbol{T}\right)=\left(U\phi_i\right)\cdot\left(U\phi_j\right)=\left(\mu_i\cdot\mu_j\right)\left(\phi_i\cdot\phi_j\right), \tag{2.17}$$

and $(\phi_i\cdot\phi_j,\mu_i\cdot\mu_j)$ is also an eigenfunction-eigenvalue pair.

Now, using (2.4), we have that

$$\phi_i\left(\boldsymbol{x}_{k+1}\right)=\phi_i\left(\boldsymbol{T}\left(\boldsymbol{x}_k\right)\right)=U\phi_i\left(\boldsymbol{x}_k\right)=\mu_i\phi_i\left(\boldsymbol{x}_k\right), \tag{2.18}$$

and it follows that

$$\phi_i\left(\boldsymbol{x}_k\right)=\mu_i^k\phi_i\left(\boldsymbol{x}_0\right). \tag{2.19}$$

Now, if all the observables of (2.4) lie within the span of the Koopman eigenfunctions, then

$$g(\boldsymbol{x}_k)=\sum_{i=1}^{\infty}\phi_i(\boldsymbol{x}_k)\upsilon_i. \tag{2.20}$$

Substituting (2.19) into (2.20), we have that

$$g(\boldsymbol{x}_k)=\sum_{i=1}^{\infty}\mu_i^k\phi_i(\boldsymbol{x}_0)\upsilon_i. \tag{2.21}$$

2.4.3 Nonlinear systems with Koopman linear expansion

It turns out that the linear expansion in (2.15) holds for a large class of nonlinear systems, including the ones that have hyperbolic fixed points, limit cycles, and tori. For these systems, the spectrum of the Koopman operator consists of only eigenvalues, and their associated eigenfunctions span the space of observables. Now we consider some of these systems in more detail. We borrow these examples from [106], where more details on the regularity of the system and related proofs can be found.

Limit cycle

Limit cycling is a nonlinear property in the sense that there is no linear system, $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$, that can generate a limit cycle, i.e. an isolated periodic orbit. If a limit cycle has time period $T = 1/f$ where f is the frequency, then the signal generated by measuring $g(\boldsymbol{x})$ while $\boldsymbol{x}$ is moving around the limit cycle is going to be T-periodic. From Fourier analysis, we have

$$g(\boldsymbol{x}) = \sum_{i=0}^{\infty} \upsilon_i e^{\left(\sqrt{-1}\right) i(2\pi/T)t} = \sum_{i=0}^{\infty} \upsilon_i e^{\left(\sqrt{-1}\right) i(2\pi f)t} = \sum_{i=0}^{\infty} \upsilon_i e^{\left(\sqrt{-1}\right) i\omega t}, \tag{2.22}$$

where υ_i, $i = 0, 1, ...$, are the Fourier coefficients, $\sqrt{-1}$ is the imaginary unit, and $\omega = 2\pi f$ is the angular speed. We can construct the eigenfunctions by letting $\phi_i(x) = e^{\left(\sqrt{-1}\right) i\omega t}$, and eigenvalues by $\lambda_i = \left(\sqrt{-1}\right) i\omega$. It is easy to check that (ϕ_i, λ_i) satisfy (2.12), and the above equation is the Koopman linear expansion of g.

Hyperbolic fixed point

Consider a nonlinear system with a hyperbolic fixed point, that is, the linearization around the fixed points yields a matrix whose eigenvalues don't lie on the imaginary axis. There are a few well-known results in dynamical systems theory, such as the Hartman-Grobman theorem [102], which state that the nonlinear system is conjugate to a linear system of the same dimension in a neighborhood of the fixed point. To be more precise, they say that there is an invertible coordinate transformation $\underline{\boldsymbol{x}} = \mathcal{X}(\boldsymbol{x})$ such that the dynamics on $\underline{\boldsymbol{x}}$-coordinate is given by $\dot{\underline{\boldsymbol{x}}} = \boldsymbol{A}\underline{\boldsymbol{x}}$, with the solution $\underline{\boldsymbol{x}}(t) = e^{\boldsymbol{A}t}\underline{\boldsymbol{x}}(0)$ and such that

$$F^t(\boldsymbol{x}) = \mathcal{X}^{-1}\left(e^{\boldsymbol{A}t}\mathcal{X}(\boldsymbol{x})\right), \tag{2.23}$$

i.e., to solve the nonlinear system, we can lift it to the $\underline{\boldsymbol{x}}$-coordinate, solve the linear system, and then transform it back to the $\boldsymbol{x}$-coordinate. We first show the Koopman linear expansion for the linear systems, and then use the conjugacy to derive the expansion for the nonlinear system.

Let $\boldsymbol{v}_i$ and λ_i, $i = 1, ..., n$, denote, respectively, the eigenvectors and eigenvalues of $\boldsymbol{A}$. The Koopman eigenfunctions for the linear system are simply the eigen-coordinates, that is

$$\underline{\phi}_i(\underline{\boldsymbol{x}}) = \langle \underline{\boldsymbol{x}}, \boldsymbol{w}_i \rangle, \tag{2.24}$$

where $\langle \underline{\boldsymbol{x}}, \boldsymbol{w}_i \rangle$ denotes the inner product between $\underline{\boldsymbol{x}}$ and $\boldsymbol{w}_i$, and $\boldsymbol{w}_i$ are the normalized eigenvectors of the complex conjugate transpose of $\boldsymbol{A}$, denoted by $\boldsymbol{A}^*$. To see this note that

$$\begin{aligned} U^t \underline{\phi}_i (\underline{\boldsymbol{x}}) &= \langle U^t \underline{\boldsymbol{x}}, \boldsymbol{w}_i \rangle = \langle e^{\boldsymbol{A}t} \underline{\boldsymbol{x}}, \boldsymbol{w}_i \rangle = \langle \underline{\boldsymbol{x}}, e^{\boldsymbol{A}^* t} \boldsymbol{w}_i \rangle \\ &= \langle \underline{\boldsymbol{x}}, e^{\lambda_i^* t} \boldsymbol{w}_i \rangle = e^{\lambda_i t} \langle \underline{\boldsymbol{x}}, \boldsymbol{w}_i \rangle = e^{\lambda_i t} \underline{\phi}_i (\underline{\boldsymbol{x}}) . \end{aligned} \tag{2.25}$$

It is easy to show that $\phi_i (\boldsymbol{x}) = \underline{\phi}_i (\mathcal{X} (\boldsymbol{x}))$ are the eigenfunctions of the Koopman operator for the nonlinear system. Other Koopman eigenfunctions can be easily constructed using the algebraic structure noted in (2.13).

To find the Koopman expansion for the nonlinear system it is easier to further transform $\underline{\boldsymbol{x}}$ into a decoupled linear system. If the matrix $\boldsymbol{A}$ is diagonalizable and $\boldsymbol{V}$ is the matrix of its eigenvectors, then the state variables of the diagonal system are, not surprisingly, the Koopman eigenfunctions, i.e.,

$$\boldsymbol{z} = [z_1, ..., z_n]^\top = \boldsymbol{V}^{-1} \underline{\boldsymbol{x}} = \left[\underline{\phi}_1 (\underline{\boldsymbol{x}}) , ..., \underline{\phi}_n (\underline{\boldsymbol{x}}) \right]^\top = [\phi_1 (\boldsymbol{x}) , ..., \phi_n (\boldsymbol{x})]^\top , \tag{2.26}$$

where $\cdot^\top$ denotes the transpose of a vector.

Now, consider an observable of the nonlinear dynamical system

$$g (\boldsymbol{x}) = g \left(\mathcal{X}^{-1} (\underline{\boldsymbol{x}}) \right) = g \left(\mathcal{X}^{-1} (\boldsymbol{V} \boldsymbol{z}) \right) = \underline{g} (\boldsymbol{z}) , \tag{2.27}$$

where $\underline{g}$ is real analytic in $\boldsymbol{z}$, and so is $\underline{\boldsymbol{x}}$. The Taylor expansion of this observable in variable $\boldsymbol{z}$ reads

$$\begin{aligned} g (\boldsymbol{x}) = \underline{g} (\boldsymbol{z}) &= \sum_{\{k_1, ..., k_n\} \in \mathbb{N}^n} \alpha_{k_1, ..., k_n} z_1^{k_1} ... z_n^{k_n} \\ &= \sum_{\{k_1, ..., k_n\} \in \mathbb{N}^n} \alpha_{k_1, ..., k_n} \phi_1^{k_1} (\boldsymbol{x}) ... \phi_n^{k_n} (\boldsymbol{x}) . \end{aligned} \tag{2.28}$$

Using the algebraic property of the Koopman eigenfunctions in (2.13), we can write the Koopman linear expansion of g as

$$U^t g = \sum_{\{k_1, ..., k_n\} \in \mathbb{N}^n} \alpha_{k_1, ..., k_n} e^{(k_1 \lambda_1 + ... + k_n \lambda_n)} \phi_1^{k_1} ... \phi_n^{k_n} . \tag{2.29}$$

Basin of attraction of a (stable) limit cycle

Now consider the motion in the basin of attraction of a (stable) limit cycle. The Koopman linear expansion for observables on such system can be constructed by, roughly speaking, combining the above two examples. That is, observables are decomposed into Koopman eigenfunctions, and each Koopman eigenfunction is a product of a periodic component, corresponding to the limit cycling, and a linearly contracting component for the stable motion toward the limit cycle. The development of this expansion is lengthy and can be found in [106].

2.4.4 Spectrum of the Koopman operator for chaotic dynamical systems

The major class of dynamical systems for which the Koopman linear expansion does not hold is the class of chaotic dynamical systems. It turns out that for these systems, the eigenfunctions of the Koopman operator do not span the space of observables and we cannot decompose fluctuations of the system all into exponentials and sinusoids. In such cases the Koopman operator usually possesses a continuous spectrum. The continuous spectrum of the Koopman operator is similar to the power spectrum of a stationary stochastic process where the energy content is spread over a range of frequencies. In fact, if our dynamical system is measure-preserving (which is typically true for evolution on attractors) the spectral density of the Koopman operator coincides with the power spectral density of observable evolution. For more on this, and generally the connection between stochastic processes and Koopman representation of deterministic dynamics, see [107]. We also note that chaos in measure-preserving system is associated by continuous spectrum but continuous spectrum can also be seen in non-chaotic systems. See the cautionary tale in [106]. The continuous spectrum is further discussed in [92, 107, 108].

What is more interesting is that some systems might possess mixed spectra which is a combination of eigenvalues and continuous spectrum. For these systems the evolution of a generic observable is composed of two parts: one part associated with eigenvalues and eigenfunctions which evolves linearly in time and a fully chaotic part corresponding to continuous spectrum. As such, the linear expansion, and the Koopman modes that will be defined next, do not hold for part of the data. Examples of systems with mixed spectra are given in [91, 107].

2.5 Estimation of the Koopman tuples

To this point, we have considered a single observable, $g(\boldsymbol{x})$. However, the measurements acquired from dynamical systems are often times made available as multiple observables. For example, in electric power systems, we typically have access to the time series of power flows and bus voltages, or in the study of climate dynamics there are recordings of atmospheric temperature measured at different stations around the globe. We can easily integrate these multiplicity of time-series data into the Koopman operator-theoretic framework.

Consider a vector-valued observable $\boldsymbol{g} : S \to \mathbb{R}^m$, such that $\boldsymbol{g} = [g_1 \ ... \ g_m]^\top$, $g_j : S \to \mathbb{R}$, $1 \leq j \leq m$. If we apply the Koopman linear expansion (2.15) to each g_j, we can collect all those expansions into a vector-valued linear expansion for $\boldsymbol{g}$,

$$U^t \boldsymbol{g}(\boldsymbol{x}) = \sum_{i=0}^{\infty} \boldsymbol{v}_i e^{\lambda_i t} \phi_i(\boldsymbol{x}) . \tag{2.30}$$

The above expansion is the KMD of the observable $\boldsymbol{g}$, and $\boldsymbol{v}_i$ is called the Koopman mode

of observable $\boldsymbol{g}$ at the eigenvalue λ_i. The Koopman modes are in fact the projection of the observable onto the Koopman eigenfunctions. We can think of $\boldsymbol{v}_i$ as a structure, or shape, within the data that evolves as $e^{\lambda_i t}$ with time. Let us examine the concept of the Koopman modes in the examples mentioned above. In the context of power networks, we can associate the network instabilities with the Koopman eigenvalues that grow in time, that is $\Re\{\lambda_i\} > 0$, and as such, the entries of Koopman mode $\boldsymbol{v}_i$ give the relative amplitude of each node in unstable growth and hence predict which nodes are most susceptible to breakdown. In the example of climate time series, the Koopman modes of temperature recordings give us the spatial pattern, depending on the location of stations, of temperature change that is proportional to $e^{\lambda_i t}$, and therefore indicate the spots with extreme variations.

In some physical problems, we have a field of observables, i.e., an observable that assigns a physical field to each element of the state space. A prominent example is a fluid flow. The pressure field over a sub-domain of the flow, or the whole vorticity field, are two examples of field of observable defined on a flow, since the knowledge of the flow state, e.g. instantaneous velocity field, uniquely determines those fields. We can formalize the notion of a field of observable as a function $\boldsymbol{g} : (S, \Omega) \to \mathbb{R}$ where Ω is the flow domain and $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{z})$ determines the value of the field at point $\boldsymbol{z}$ in the flow domain when the flow is at state $\boldsymbol{x}$. The Koopman linear expansion for $\boldsymbol{g}$ would be

$$U^t \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{z}) = \sum_{i=0}^{\infty} \boldsymbol{v}_i(\boldsymbol{z})\, e^{\lambda_i t} \phi_i(\boldsymbol{x}). \tag{2.31}$$

where the Koopman mode $\boldsymbol{v}_i(\boldsymbol{z})$ is a fixed field by itself, and similar to the Koopman modes, determines a shape function on Ω which grows with the amplitude $e^{\lambda_i t}$ in time. In a fluid flow, the Koopman modes of vorticity, are steady vorticity fields, and the whole flow can be decomposed into such fields, with amplitudes that grow as $e^{\lambda_i t}$.

Now, for simplicity, suppose that we measure (or simulate) the evolution of states directly. In other words, suppose that the observable function is the identity map, i.e., $g(\boldsymbol{x}_k) = \boldsymbol{x}_k$. Then, in what follows, consider a sequential set of data vectors $\{\boldsymbol{x}_0, ..., \boldsymbol{x}_N\}$ coming from either measurements or simulations, where each $\boldsymbol{x} \in \mathbb{R}^n$. Assume that the data originate from linear dynamics, expressed as

$$\boldsymbol{x}_{k+1} = \boldsymbol{K}\boldsymbol{x}_k, \tag{2.32}$$

and that the matrix $\boldsymbol{K}$ is unknown. Nonetheless, even if the data originate from nonlinear dynamics, it is assumed that an operator $\boldsymbol{K}$ is able to approximate the underlying dynamics.

2.5.1 Modified Arnoldi method

Rowley et al. [93] showed that the so-called empirical Ritz values and empirical Ritz vectors approximate, respectively, the Koopman eigenvalues λ_i and factors $\phi_i(\boldsymbol{x}_0)\,\boldsymbol{v}_i$ in terms of a

finite truncation. The empirical Ritz values and vectors are computed in [93] using a modified version of the Arnoldi algorithm that does not require explicit knowledge of the underlying system matrix. This variant of the Arnoldi algorithm was put forth by Saad in [109].

Consider the sequential data generated by the linear dynamics (2.32). Define a residual vector $\boldsymbol{r}$, such that

$$\begin{aligned} \boldsymbol{r} &= \boldsymbol{x}_N - \sum_{i=0}^{N-1} c_i \boldsymbol{x}_i, \\ &= \boldsymbol{x}_N - \boldsymbol{X}\boldsymbol{c}, \end{aligned} \tag{2.33}$$

where $\boldsymbol{X} = [\boldsymbol{x}_0 \ ... \ \boldsymbol{x}_{N-1}]$, and $\boldsymbol{c} = [c_0 \ ... \ c_{N-1}]^\top$.

The constants c_i are chosen such that

$$\boldsymbol{r} \perp \text{span}\left\{\boldsymbol{x}_0, ..., \boldsymbol{x}_{N-1}\right\}. \tag{2.34}$$

Pre-multiplying (2.33) by $\boldsymbol{X}^\top$, and by virtue of (2.34), one has

$$\boldsymbol{X}^\top \boldsymbol{r} = \boldsymbol{X}^\top \boldsymbol{x}_N - \left(\boldsymbol{X}^\top \boldsymbol{X}\right)\boldsymbol{c} = \boldsymbol{b} - \boldsymbol{Y}\boldsymbol{c} = \boldsymbol{0}. \tag{2.35}$$

Because matrix $\boldsymbol{Y}$ is not full rank, it is impossible to determine a unique minimizer $\boldsymbol{c} \in \mathbb{R}^N$ of the norm $\|\boldsymbol{b} - \boldsymbol{Y}\boldsymbol{c}\|$. In [64], the authors proposed to minimize the norm by using the Moore-Penrose pseudoinverse of $\boldsymbol{Y}$, denoted by $\boldsymbol{Y}^\dagger$. Thus,

$$\boldsymbol{c} = \boldsymbol{Y}^\dagger \boldsymbol{b}. \tag{2.36}$$

Now, consider the Companion matrix

$$\boldsymbol{C} = \begin{bmatrix} 0 & 0 & ... & 0 & c_0 \\ 1 & 0 & & 0 & c_1 \\ 0 & 1 & & 0 & c_2 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & 1 & c_{N-1} \end{bmatrix}. \tag{2.37}$$

The empirical Ritz values are solutions to $\det\left(\boldsymbol{C} - \lambda \boldsymbol{I}\right) = 0$, where $\boldsymbol{I}$ is the identity matrix. Finally, define the Vandermonde matrix

$$\boldsymbol{V} = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & ... & \lambda_1^{N-1} \\ 1 & \lambda_2 & \lambda_2^2 & ... & \lambda_2^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_N & \lambda_N^2 & ... & \lambda_N^{N-1} \end{bmatrix}. \tag{2.38}$$

Vandermonde matrices and Companion matrices are closely related, in that $\boldsymbol{V}$ diagonalizes the companion matrix $\boldsymbol{C}$, as long as the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_m$ are distinct. The empirical Ritz vectors are defined as the columns of $\boldsymbol{X}\boldsymbol{V}^{-1}$. The empirical Ritz values are good approximations of the Koopman eigenvalues, λ_i. The empirical Ritz vectors, in turn, are approximations of the Koopman modes $\boldsymbol{\upsilon}_i$ scaled by constant values $\phi_i(\boldsymbol{x}_0)$. Nevertheless, they are computed for a finite sum as in (2.39) instead of an infinite sum.

$$\begin{cases} \boldsymbol{x}_k = \sum_{i=1}^{N} \lambda_i^k \boldsymbol{\upsilon}_i, & \text{for } k = 1, ..., N-1, \\ \boldsymbol{x}_N = \boldsymbol{r} + \sum_{i=1}^{N} \lambda_i^N \boldsymbol{\upsilon}_i, \end{cases} \tag{2.39}$$

where $\boldsymbol{r}$ is a residual vector that accounts for the approximation errors.

2.5.2 Dynamic mode decomposition

The original dynamic mode decomposition (DMD) algorithm by Schmid and Sesterhenn [94,110] was also formulated in terms of a Companion matrix. Then, Rowley et al. [93] made the connection to the modified Arnoldi algorithm and to Koopman operator theory. However, as argued by Tu et al. [111], the algorithm based on the singular value decomposition (SVD) is more numerically stable. We present it next.

Consider the sequential data generated by the linear dynamics (2.32). Arrange the data into matrices,

$$\boldsymbol{X} = [\boldsymbol{x}_0 \ ... \ \boldsymbol{x}_{N-1}], \qquad \boldsymbol{X}^+ = [\boldsymbol{x}_1 \ ... \ \boldsymbol{x}_N]. \tag{2.40}$$

Compute the reduced SVD of $\boldsymbol{X}$,

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*, \tag{2.41}$$

where $\boldsymbol{U}$ is an $n \times r$ real or complex matrix, $\boldsymbol{\Sigma}$ is an $r \times r$ diagonal matrix with non-negative real numbers on the diagonal, $\boldsymbol{V}$ is an $m \times r$ real or complex matrix, and r denotes the rank of $\boldsymbol{X}$.

Now, define the matrix

$$\widetilde{\boldsymbol{K}} := \boldsymbol{U}\boldsymbol{X}^+\boldsymbol{V}\boldsymbol{\Sigma}^{-1}. \tag{2.42}$$

Compute the eigenvalues and eigenvectors of $\widetilde{\boldsymbol{K}}$, denoted by

$$\widetilde{\boldsymbol{K}}\boldsymbol{w} = \lambda\boldsymbol{w}. \tag{2.43}$$

The DMD mode associated with the DMD eigenvalue λ is given by

$$\widetilde{\boldsymbol{\upsilon}} := \boldsymbol{U}\boldsymbol{w}. \tag{2.44}$$

Following Tu et al. [111], $\widetilde{\boldsymbol{\upsilon}}$ are referred to as the *projected DMD modes*.

2.5.3 Exact dynamic mode decomposition

The DMD algorithm is formulated under the assumption that the set of data vectors $\{\boldsymbol{x}_0, ..., \boldsymbol{x}_N\}$ is sequential and ordered. Also, the vectors should satisfy (2.32), at least approximately. Tu et al. [111] relaxed these restrictions on the data by considering data pairs $\{(\boldsymbol{x}_1, \boldsymbol{y}_1), ..., (\boldsymbol{x}_N, \boldsymbol{y}_N)\}$, as follows.

Define the data matrices

$$\boldsymbol{X} := [\boldsymbol{x}_1 \ ... \ \boldsymbol{x}_N], \qquad \boldsymbol{X}^+ := [\boldsymbol{y}_1 \ ... \ \boldsymbol{y}_N]. \tag{2.45}$$

From Table 2.1, note that the data matrices defined for the DMD algorithm are a special case of the ones defined for the exact DMD, with $\boldsymbol{y}_k = \boldsymbol{x}_{k+1}$.

Table 2.1: Comparison between data matrices

Data matrix	DMD	Exact DMD
$\boldsymbol{X}$	$[\boldsymbol{x}_0 \ ... \ \boldsymbol{x}_{N-1}]$	$[\boldsymbol{x}_1 \ ... \ \boldsymbol{x}_N]$
$\boldsymbol{X}^+$	$[\boldsymbol{x}_1 \ ... \ \boldsymbol{x}_N]$	$[\boldsymbol{y}_1 \ ... \ \boldsymbol{y}_N]$

For a dataset given by (2.45), define the operator

$$\boldsymbol{K} := \boldsymbol{X}^+ \boldsymbol{X}^\dagger. \tag{2.46}$$

The dynamic mode decomposition of the pair $(\boldsymbol{X}, \boldsymbol{X}^+)$ is given by the eigendecomposition of $\boldsymbol{K}$. That is, the DMD modes and eigenvalues are the eigenvectors and eigenvalues of $\boldsymbol{K}$. The exact DMD algorithm is as follows.

Arrange the data pairs $\{(\boldsymbol{x}_1, \boldsymbol{y}_1), ..., (\boldsymbol{x}_N, \boldsymbol{y}_N)\}$ into matrices $\boldsymbol{X}$ and $\boldsymbol{X}^+$, as in (2.45).

Compute the reduced SVD of $\boldsymbol{X}$,

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*. \tag{2.47}$$

Now, define the matrix

$$\widetilde{\boldsymbol{K}} := \boldsymbol{U}\boldsymbol{X}^+\boldsymbol{V}\boldsymbol{\Sigma}^{-1}. \tag{2.48}$$

Compute the eigenvalues and eigenvectors of $\widetilde{\boldsymbol{K}}$, denoted by

$$\widetilde{\boldsymbol{K}}\boldsymbol{w} = \lambda \boldsymbol{w}. \tag{2.49}$$

The DMD mode corresponding to the DMD eigenvalue λ is given by

$$\boldsymbol{v} := \frac{1}{\lambda}\boldsymbol{X}^+\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{w}. \tag{2.50}$$

The interested reader is referred to [111] for more details on the exact DMD.

2.5.4 Extended dynamic mode decomposition

Williams et al. [112] originally proposed the extended dynamic mode decomposition (EDMD). Instead of the original EDMD formulation, we rely on a slightly different formulation that has been proposed by Klus et al. [113]. The important difference is that Klus et al. utilize the left eigenvectors of the finite-dimensional approximation of the Koopman operator to perform a similarity transformation, whereas Williams et al. make use of the right eigenvectors. It turns out that this subtle detail is important for the participation factors developed in chapter 7. Thus, following Klus et al. [113], consider a set of snapshots of the system states, $\boldsymbol{x}_k$. Now, define the matrices

$$\boldsymbol{X} = [\boldsymbol{x}_1 \ ... \ \boldsymbol{x}_N] \quad \text{and} \quad \boldsymbol{X}^+ = [\boldsymbol{y}_1 \ ... \ \boldsymbol{y}_N], \tag{2.51}$$

$\boldsymbol{X}, \boldsymbol{X}^+ \in \mathbb{R}^{n \times N}$. Also, define a vector of observable functions

$$\boldsymbol{g}(\boldsymbol{x}_k) = [g_1(\boldsymbol{x}_k); \ ...; \ g_q(\boldsymbol{x}_k)]^\top, \tag{2.52}$$

where $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^q$, and the matrices of observables,

$$\boldsymbol{O}_{\boldsymbol{X}} = [\boldsymbol{g}(\boldsymbol{x}_1) \ ... \ \boldsymbol{g}(\boldsymbol{x}_N)], \quad \boldsymbol{O}_{\boldsymbol{X}^+} = [\boldsymbol{g}(\boldsymbol{y}_1) \ ... \ \boldsymbol{g}(\boldsymbol{y}_N)], \tag{2.53}$$

$\boldsymbol{O}_{\boldsymbol{X}}, \boldsymbol{O}_{\boldsymbol{X}^+} \in \mathbb{R}^{q \times N}$. A finite-dimensional approximation to the Koopman operator is estimated as follows:

$$\boldsymbol{K} = \boldsymbol{O}_{\boldsymbol{X}^+} \boldsymbol{O}_{\boldsymbol{X}}^\dagger, \tag{2.54}$$

where $\boldsymbol{K} \in \mathbb{R}^{q \times q}$. The eigenvalues of $\boldsymbol{K}$ are a finite-dimensional approximation to the Koopman eigenvalues, whereas the Koopman eigenfunctions, ϕ_i, are given by

$$\boldsymbol{\phi}(\boldsymbol{x}_k) = \boldsymbol{\Xi} \boldsymbol{g}(\boldsymbol{x}_k), \tag{2.55}$$

where $\boldsymbol{\Xi} = \left[\boldsymbol{\xi}_1^\top; \ ...; \ \boldsymbol{\xi}_q^\top\right]$ contains the left eigenvectors of $\boldsymbol{K}$, and $\boldsymbol{\phi}(\boldsymbol{x}_k) = [\phi_1(\boldsymbol{x}_k); \ ...; \ \phi_q(\boldsymbol{x}_k)]^\top$.

Finally, in order to obtain the Koopman modes for the full-state observable, i.e. $\boldsymbol{g}(\boldsymbol{x}_k) = \boldsymbol{x}_k$, let $\boldsymbol{B} \in \mathbb{R}^{n \times q}$ be a matrix defined such that

$$\boldsymbol{x}_k = \boldsymbol{B} \boldsymbol{g}(\boldsymbol{x}_k). \tag{2.56}$$

From (2.55), we have that $\boldsymbol{g}(\boldsymbol{x}_k) = \boldsymbol{\Xi}^{-1} \boldsymbol{\phi}(\boldsymbol{x}_k)$, and

$$\boldsymbol{x}_k = \boldsymbol{B} \boldsymbol{g}(\boldsymbol{x}_k) = \boldsymbol{B} \boldsymbol{\Xi}^{-1} \boldsymbol{\phi}(\boldsymbol{x}_k), \tag{2.57}$$

where $\boldsymbol{\Xi}^{-1}$ contains the right eigenvectors of $\boldsymbol{K}$.

Therefore the Koopman modes are the column vectors, $\boldsymbol{\upsilon}_i$, $i = 1, ..., q$, of $\boldsymbol{\Upsilon} = \boldsymbol{B} \boldsymbol{\Xi}^{-1}$, $\boldsymbol{\Upsilon} \in \mathbb{C}^{n \times q}$, and

$$\boldsymbol{x}_k = \sum_{i=1}^{q} \phi_i(\boldsymbol{x}_k) \boldsymbol{\upsilon}_i = \sum_{i=1}^{q} \phi_i(\boldsymbol{x}_0) \boldsymbol{\upsilon}_i \lambda_i^k. \tag{2.58}$$

The convergence of the EDMD to the Koopman operator was demonstrated by Korda and Mezić [114], and by Arbabi and Mezić [115].

2.5.5 Other numerical methods

The computation of the Koopman tuples (modes, eigenfunctions and eigenvalues) has seen a lot of major advancements. For post transient systems, the Koopman eigenvalues lie on the unit circle and Fourier analysis techniques can be used to find the Koopman spectrum and modes [91]. For dissipative systems, the Koopman spectral properties can be computed using a theoretical algorithm known as the Generalized Laplace Analysis [116]. In practice however, DMD remains the more popular technique for computation of Koopman spectrum from data. The works in [117] and [111] discussed the linear algebraic properties of the algorithm and suggested new variations for better performance and wider applications. Various important variants of the DMD algorithm have been proposed. Jovanović et al. [118] developed a sparsity-promoting variant of the DMD algorithm. Kutz et al. [119, 120] developed a DMD algorithm that allows for a hierarchical separation of multiresolution time-scale components. Proctor et al. [121] developed a DMD algorithm that is able disambiguate between the underlying dynamics and the effects of actuation. Due to constant growth in the size of the available data, new alterations or improvements of DMD are also devised to handle high-dimensional data sets [122, 123], different sampling techniques [111, 123, 124] and noise [125, 126].

2.6 Concluding remarks

The ultimate goal of many data analysis techniques is to provide information that can be used to predict and manipulate a system to our benefit. Application of the Koopman operator techniques to data-driven prediction and control are just being developed, with a few-year lag behind the above work. This lag is perhaps due to the need to account for the effect of input in the formalism, but promising results have already appeared in this line of research. The work in [127] showed an example of an optimal controller which was designed based on a finite-dimensional Koopman linear expansion of nonlinear dynamics. The works in [69–73, 79] have developed a framework to build state estimators for nonlinear systems based on Koopman expansions. More recent works, have shown successful examples of Koopman linear predictors for nonlinear systems [128], and optimal controllers of Hamiltonian systems designed based on Koopman eigenfunctions [129]. More recent applications include feedback control of fluid flows by using Koopman linear models computed from data in a model-predictive control framework [130, 131].

Chapter 3

Discrete time Kalman filtering

In the previous chapter, we introduced the Koopman operator theory for continuous time dynamical systems of the form

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}),$$

where $\boldsymbol{f} : S \to \mathbb{R}^n$ is a vector field on the state space S. We have also considered discrete time dynamical systems of the form

$$\boldsymbol{x}_{k+1} = \boldsymbol{T}(\boldsymbol{x}_k),$$

where $\boldsymbol{T} : S \to S$ is a discrete time map. Note that the discrete time map was denoted by a capital letter, $\boldsymbol{T}$. In this chapter, and in the remaining of this dissertation, $\boldsymbol{f}(\cdot)$ might either denote a vector field or a discrete time map. This choice is made in an effort to keep the notation of the present text consistent with the literature on Kalman filtering [132–134].

There are several ways to approach the Kalman filtering theory. In this chapter, we approach the Kalman filtering theory from the viewpoint of someone working in the field of estimation. This choice is motivated by our final goal, which is to introduce the reader to a relatively new and less known version of the Kalman filter that presents high statistical efficiency when the system and observation errors in the Kalman filter deviate from Gaussianity.

This chapter proceeds as follows. In section 3.1, we present the classical discrete time Kalman filter. In section 3.2, we introduce the notion of robust Kalman filter. This closes the part of this chapter dealing with linear systems. In section 3.3, we present the classical discrete time extended Kalman filter, which is probably the most popular version of the Kalman filter applied to nonlinear systems. Finally, in section 3.4, we present a slightly modified version of the extended Kalman filter, which is referred to as the *iterated* extended Kalman filter and has improved performance under high nonlinearities.

3.1 Kalman filter

Consider the discrete time linear control system,

$$\boldsymbol{x}_k = \boldsymbol{F}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{u}_{k-1} + \boldsymbol{w}_{k-1}, \tag{3.1}$$

where $\boldsymbol{x}_k \in \mathbb{R}^n$ is the state vector, $\boldsymbol{F}_k : \mathbb{R}^n \to \mathbb{R}^n$ is the state transition matrix, $\boldsymbol{u}_k \in \mathbb{R}^p$ is the input vector, and $\boldsymbol{w}_k \in \mathbb{R}^n$ is the system error vector that accounts for model and parameter uncertainties.

Furthermore, consider the observation equation,

$$\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{v}_k, \tag{3.2}$$

where $\boldsymbol{y}_k \in \mathbb{R}^m$ is the observation vector, $\boldsymbol{H}_k : \mathbb{R}^n \to \mathbb{R}^m$ is the observation matrix, and $\boldsymbol{v}_k \in \mathbb{R}^m$ is the observation error vector; in addition to model and parameter uncertainties, $\boldsymbol{v}_k$ also accounts for measurement noise.

Before proceeding, the following assumptions are in order.

Assumption 1. $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are independent Gaussian white processes with zero mean and known covariance matrices, $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$ respectively, i.e.,

$$E\left[\boldsymbol{w}_k\right] = \boldsymbol{0}\ \forall k \in \mathbb{Z}, \tag{3.3}$$
$$E\left[\boldsymbol{v}_k\right] = \boldsymbol{0}\ \forall k \in \mathbb{Z}, \tag{3.4}$$
$$E\left[\boldsymbol{w}_k\boldsymbol{w}_k^\top\right] = \boldsymbol{Q}_k, \tag{3.5}$$
$$E\left[\boldsymbol{v}_k\boldsymbol{v}_k^\top\right] = \boldsymbol{R}_k, \tag{3.6}$$
$$E\left[\boldsymbol{w}_k\boldsymbol{w}_j^\top\right] = E\left[\boldsymbol{v}_k\boldsymbol{v}_j^\top\right] = \boldsymbol{0}\ \forall k, j \in \mathbb{Z}, k \neq j, \tag{3.7}$$
$$E\left[\boldsymbol{w}_k\boldsymbol{v}_j^\top\right] = \boldsymbol{0}\ \forall k, j \in \mathbb{Z}, \tag{3.8}$$

where $E\left[\cdot\right]$ denotes the expectation operator.

Assumption 2. $\boldsymbol{u}_k$ is a known deterministic input vector for all $k \in \mathbb{Z}$.

Assumption 3. $\boldsymbol{F}_k$ and $\boldsymbol{H}_k$ are known deterministic matrices for all $k \in \mathbb{Z}$.

The Kalman filter is a prediction-correction process, as follows.

Prediction step

Suppose that an estimate of the state at time $k-1$, $\hat{\boldsymbol{x}}_{k-1|k-1}$, is available.

The predicted state at time k is given by

$$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{F}_{k-1}\hat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{u}_{k-1}. \tag{3.9}$$

The true state at time k is given by

$$\boldsymbol{x}_k = \boldsymbol{F}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{u}_{k-1} + \boldsymbol{w}_{k-1}, \tag{3.10}$$

where $\boldsymbol{x}_k$ and $\boldsymbol{x}_{k-1}$ are the true state at time k and $k-1$, respectively.

Define the following error vectors:

$$\text{Prediction error} := \boldsymbol{e}_{k|k-1} = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}. \tag{3.11}$$

$$\text{State estimate error} := \boldsymbol{e}_{k-1|k-1} = \boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1|k-1}. \tag{3.12}$$

Also, since $\boldsymbol{u}_k$ is a known deterministic input vector, define $\boldsymbol{u}_{k-1} := \boldsymbol{0}$ for simplicity.

Subtracting (3.9) from (3.10) we get

$$\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{F}_{k-1}\left(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1|k-1}\right) + \boldsymbol{w}_{k-1}, \tag{3.13}$$

and using (3.11), we have:

$$\boldsymbol{e}_{k|k-1} = \boldsymbol{F}_{k-1}\boldsymbol{e}_{k-1|k-1} + \boldsymbol{w}_{k-1}. \tag{3.14}$$

Assumption 4. The state estimation error at $k = 0$ is negligible, such that

$$E\left[\boldsymbol{e}_{0|0}\right] = \boldsymbol{0}. \tag{3.15}$$

Using (3.14) and (3.15), we have that $E\left[\boldsymbol{e}_{1|0}\right] = \boldsymbol{F}_0 E\left[\boldsymbol{e}_{0|0}\right] + E\left[\boldsymbol{w}_0\right] = \boldsymbol{0}$, and by induction, $E\left[\boldsymbol{e}_{k|k-1}\right] = \boldsymbol{0}$ for all $k \in \mathbb{Z}_{>0}$. We will show that $E\left[\boldsymbol{e}_{k|k}\right] = \boldsymbol{0}$ for all $k \in \mathbb{Z}_{>0}$.

Define the following covariance matrices:

$$\text{Prediction error covariance} := \boldsymbol{\Sigma}_{k|k-1} = \operatorname{cov}\left(\boldsymbol{e}_{k|k-1}\right) = E\left[\boldsymbol{e}_{k|k-1}\boldsymbol{e}_{k|k-1}^\top\right]. \tag{3.16}$$

$$\text{State estimate error covariance} := \boldsymbol{\Sigma}_{k-1|k-1} = \operatorname{cov}\left(\boldsymbol{e}_{k-1|k-1}\right) = E\left[\boldsymbol{e}_{k-1|k-1}\boldsymbol{e}_{k-1|k-1}^\top\right]. \tag{3.17}$$

Using (3.14), we can write

$$\begin{aligned}\boldsymbol{e}_{k|k-1}\boldsymbol{e}_{k|k-1}^\top &= \left(\boldsymbol{F}_{k-1}\boldsymbol{e}_{k-1|k-1} + \boldsymbol{w}_{k-1}\right)\left(\boldsymbol{e}_{k-1|k-1}^\top\boldsymbol{F}_{k-1}^\top + \boldsymbol{w}_{k-1}^\top\right)\\ &= \boldsymbol{F}_{k-1}\boldsymbol{e}_{k-1|k-1}\boldsymbol{e}_{k-1|k-1}^\top\boldsymbol{F}_{k-1}^\top + \boldsymbol{F}_{k-1}\boldsymbol{e}_{k-1|k-1}\boldsymbol{w}_{k-1}^\top\\ &\quad + \boldsymbol{w}_{k-1}\boldsymbol{e}_{k-1|k-1}^\top\boldsymbol{F}_{k-1}^\top + \boldsymbol{w}_{k-1}\boldsymbol{w}_{k-1}^\top.\end{aligned} \tag{3.18}$$

Applying the expectation operator in (3.18), we get

$$\begin{aligned}E\left[\boldsymbol{e}_{k|k-1}\boldsymbol{e}_{k|k-1}^\top\right] &= \boldsymbol{F}_{k-1}E\left[\boldsymbol{e}_{k-1|k-1}\boldsymbol{e}_{k-1|k-1}^\top\right]\boldsymbol{F}_{k-1}^\top + \boldsymbol{F}_{k-1}E\left[\boldsymbol{e}_{k-1|k-1}\boldsymbol{w}_{k-1}^\top\right]\\ &\quad + E\left[\boldsymbol{w}_{k-1}\boldsymbol{e}_{k-1|k-1}^\top\right]\boldsymbol{F}_{k-1}^\top + E\left[\boldsymbol{w}_{k-1}\boldsymbol{w}_{k-1}^\top\right].\end{aligned} \tag{3.19}$$

Assumption 5. $\boldsymbol{e}_{k|k}$ and $\boldsymbol{w}_k$ are white and independent of each other, i.e.,

$$E\left[\boldsymbol{e}_{k|k}\boldsymbol{w}_k^\top\right] = E\left[\boldsymbol{w}_k\boldsymbol{e}_{k|k}^\top\right] = \boldsymbol{0}\ \forall k \in \mathbb{Z}. \tag{3.20}$$

Substituting (3.20) in (3.19), and using (3.16) and (3.17), we have:

$$\boldsymbol{\Sigma}_{k|k-1} = \boldsymbol{F}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{F}_{k-1}^\top + \boldsymbol{Q}_{k-1}. \tag{3.21}$$

Correction (or filtering) step

At time k, we have the predicted state, $\hat{\boldsymbol{x}}_{k|k-1}$. From (3.11),

$$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{x}_k - \boldsymbol{e}_{k|k-1}. \tag{3.22}$$

In addition, we have the observation equation (3.2). Put (3.2) and (3.22) into a matrix form

$$\begin{bmatrix} \boldsymbol{y}_k \\ \hat{\boldsymbol{x}}_{k|k-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_k \\ \boldsymbol{I} \end{bmatrix} \boldsymbol{x}_k + \begin{bmatrix} \boldsymbol{v}_k \\ -\boldsymbol{e}_{k|k-1} \end{bmatrix}, \tag{3.23}$$

or in compact form,

$$\widetilde{\boldsymbol{y}}_k = \widetilde{\boldsymbol{H}}_k\boldsymbol{x}_k + \widetilde{\boldsymbol{v}}_k. \tag{3.24}$$

This is a linear regression model. The problem becomes one of estimating $\boldsymbol{x}_k$, in a certain sense, given $\widetilde{\boldsymbol{y}}_k$ and $\widetilde{\boldsymbol{H}}_k$.

Assumption 6. $\boldsymbol{e}_{k|k}$ and $\boldsymbol{v}_k$ are white and independent of each other, i.e.,

$$E\left[\boldsymbol{e}_{k|k-1}\boldsymbol{v}_k^\top\right] = E\left[\boldsymbol{v}_k\boldsymbol{e}_{k|k-1}^\top\right] = \boldsymbol{0}\ \forall k \in \mathbb{Z}. \tag{3.25}$$

In (3.24), we have that:

$$E\left[\widetilde{\boldsymbol{v}}_k\right] = E\begin{bmatrix} \boldsymbol{v}_k \\ -\boldsymbol{e}_{k|k-1} \end{bmatrix} = \boldsymbol{0}, \tag{3.26}$$

$$\operatorname{cov}\left(\widetilde{\boldsymbol{v}}_k\right) = E\left[\widetilde{\boldsymbol{v}}_k\widetilde{\boldsymbol{v}}_k^\top\right] = \widetilde{\boldsymbol{R}}_k. \tag{3.27}$$

Note that $\widetilde{\boldsymbol{R}}_k$ is not a diagonal matrix, as

$$\begin{aligned} \operatorname{cov}\left(\widetilde{\boldsymbol{v}}_k\right) = E\left[\widetilde{\boldsymbol{v}}_k\widetilde{\boldsymbol{v}}_k^\top\right] = \widetilde{\boldsymbol{R}}_k &= E\left[\begin{bmatrix} \boldsymbol{v}_k \\ -\boldsymbol{e}_{k|k-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{v}_k^\top & -\boldsymbol{e}_{k|k-1}^\top \end{bmatrix}\right] \\ &= \begin{bmatrix} \boldsymbol{R}_k & -E\left[\boldsymbol{v}_k\boldsymbol{e}_{k|k-1}^\top\right] \\ -E\left[\boldsymbol{e}_{k|k-1}\boldsymbol{v}_k^\top\right] & \boldsymbol{\Sigma}_{k|k-1} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{R}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{k|k-1} \end{bmatrix}, \end{aligned} \tag{3.28}$$

and $\boldsymbol{\Sigma}_{k|k-1}$ is typically a full matrix.

Apply the weighted least squares method, i.e.,

$$\text{minimize} \qquad J\left(\boldsymbol{x}_k\right) = \sum_{i=1}^{m} \varpi_i r_{s_i}^2, \tag{3.29}$$

to estimate $\boldsymbol{x}_k$ in the least squares sense. In matrix form,

$$\text{minimize} \qquad J\left(\boldsymbol{x}_k\right) = \left(\widetilde{\boldsymbol{y}}_k - \widetilde{\boldsymbol{H}}_k \boldsymbol{x}_k\right)^\top \widetilde{\boldsymbol{R}}_k^{-1} \left(\widetilde{\boldsymbol{y}}_k - \widetilde{\boldsymbol{H}}_k \boldsymbol{x}_k\right). \tag{3.30}$$

The necessary condition of optimality is

$$\frac{\partial J\left(\boldsymbol{x}_k\right)}{\partial \boldsymbol{x}_k} = -2\widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \left(\widetilde{\boldsymbol{y}}_k - \widetilde{\boldsymbol{H}}_k \boldsymbol{x}_k\right) = \boldsymbol{0}, \tag{3.31}$$

and solving for $\boldsymbol{x}_k$, we get

$$\hat{\boldsymbol{x}}_{k|k} = \left(\widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{H}}_k\right)^{-1} \widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{y}}_k. \tag{3.32}$$

Statistical analysis of $\boldsymbol{e}_{k|k}$

Substituting (3.24) into (3.32) yields

$$\hat{\boldsymbol{x}}_{k|k} = \left(\widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{H}}_k\right)^{-1} \widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \left(\widetilde{\boldsymbol{H}}_k \boldsymbol{x}_k + \widetilde{\boldsymbol{v}}_k\right) = \boldsymbol{x}_k + \left(\widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{H}}_k\right)^{-1} \widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{v}}_k. \tag{3.33}$$

Substituting (3.33) into (3.12), we have

$$\boldsymbol{e}_{k|k} = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k} = -\left(\widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{H}}_k\right)^{-1} \widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{v}}_k, \tag{3.34}$$

and $E\left[\boldsymbol{e}_{k|k}\right] = \boldsymbol{0}$ because $E\left[\widetilde{\boldsymbol{v}}_k\right] = \boldsymbol{0}$.

Now, as for the covariance,

$$\begin{aligned}
\operatorname{cov}\left(\boldsymbol{e}_{k|k}\right) = \boldsymbol{\Sigma}_{k|k} &= E\left[\boldsymbol{e}_{k|k}\boldsymbol{e}_{k|k}^\top\right] \\
&= E\left[\left(\widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{H}}_k\right)^{-1} \widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \left(\widetilde{\boldsymbol{v}}_k \widetilde{\boldsymbol{v}}_k^\top\right) \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{H}}_k \left(\widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{H}}_k\right)^{-1}\right] \\
&= E\left[\left(\widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{H}}_k\right)^{-1} \widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \left(\widetilde{\boldsymbol{R}}_k\right) \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{H}}_k \left(\widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{H}}_k\right)^{-1}\right], \\
&= \left(\widetilde{\boldsymbol{H}}_k^\top \widetilde{\boldsymbol{R}}_k^{-1} \widetilde{\boldsymbol{H}}_k\right)^{-1},
\end{aligned} \tag{3.35}$$

thereby attaining the Cramér-Rao lower bound under Gaussianity. Note that $\left(\widetilde{\boldsymbol{H}}_k^{\top}\widetilde{\boldsymbol{R}}_k^{-1}\widetilde{\boldsymbol{H}}_k\right)$ is the Fisher information matrix. Therefore, the Kalman filter is the optimal filter under Gaussian assumptions.

Substituting

$$\widetilde{\boldsymbol{H}}_k = \begin{bmatrix} \boldsymbol{H}_k \\ \boldsymbol{I} \end{bmatrix}, \quad \text{and} \quad \widetilde{\boldsymbol{R}}_k = \begin{bmatrix} \boldsymbol{R}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{k|k-1} \end{bmatrix}, \tag{3.36}$$

into (3.35) yields

$$\begin{aligned} \boldsymbol{\Sigma}_{k|k} &= \left(\begin{bmatrix} \boldsymbol{H}_k^{\top} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_k^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{k|k-1}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{H}_k \\ \boldsymbol{I} \end{bmatrix} \right)^{-1} \\ &= \left(\begin{bmatrix} \boldsymbol{H}_k^{\top}\boldsymbol{R}_k^{-1} & \boldsymbol{\Sigma}_{k|k-1}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{H}_k \\ \boldsymbol{I} \end{bmatrix} \right)^{-1} \\ &= \left(\boldsymbol{H}_k^{\top}\boldsymbol{R}_k^{-1}\boldsymbol{H}_k + \boldsymbol{\Sigma}_{k|k-1}^{-1} \right)^{-1}. \end{aligned} \tag{3.37}$$

Now, we will make use of the *matrix inversion lemma*, also known as the *Sherman-Morrison-Woodbury formula*. In simple words, this lemma is attractive because it allows to efficiently compute the inverse of very large matrices, under certain conditions. We state it without formalism and proof, as follows.

Lemma 3.1.1 *Assume that we know the inverse $\boldsymbol{A}_1^{-1}$ of a matrix $\boldsymbol{A}_1$ that is of very large dimension. Then, given a matrix $\boldsymbol{B} = \boldsymbol{A}_1 + \boldsymbol{A}_2\boldsymbol{A}_4^{-1}\boldsymbol{A}_3$ of the same dimension of $\boldsymbol{A}_1$, where $\boldsymbol{A}_4$ is of low dimension, possibly a scalar,*

$$\boldsymbol{B}^{-1} = \boldsymbol{A}_1^{-1} - \boldsymbol{A}_1^{-1}\boldsymbol{A}_2\left(\boldsymbol{A}_4 + \boldsymbol{A}_3\boldsymbol{A}_1^{-1}\boldsymbol{A}_2\right)^{-1}\boldsymbol{A}_3\boldsymbol{A}_1^{-1}, \tag{3.38}$$

where $\left(\boldsymbol{A}_4 + \boldsymbol{A}_3\boldsymbol{A}_1^{-1}\boldsymbol{A}_2\right)^{-1}$ is of the same dimension of $\boldsymbol{A}_4$.

Apply the matrix inversion lemma to (3.35). Define $\boldsymbol{B}^{-1} = \boldsymbol{\Sigma}_{k|k} = \left(\widetilde{\boldsymbol{H}}_k^{\top}\widetilde{\boldsymbol{R}}_k^{-1}\widetilde{\boldsymbol{H}}_k\right)^{-1}$, and put $\boldsymbol{A}_1 = \boldsymbol{\Sigma}_{k|k-1}^{-1}$, $\boldsymbol{A}_2 = \boldsymbol{H}_k^{\top}$, $\boldsymbol{A}_3 = \boldsymbol{H}_k$, and $\boldsymbol{A}_4 = \boldsymbol{R}_k$. Then,

$$\boldsymbol{\Sigma}_{k|k} = \boldsymbol{\Sigma}_{k|k-1} - \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\top}\left(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\top}\right)^{-1}\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}, \tag{3.39}$$

and

$$\boldsymbol{G}_k := \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\top}\left(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\top}\right)^{-1}, \tag{3.40}$$

is the Kalman filter gain. Thus,

$$\begin{aligned} \boldsymbol{\Sigma}_{k|k} &= \boldsymbol{\Sigma}_{k|k-1} - \boldsymbol{G}_k\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}, \\ &= \left(\boldsymbol{I} - \boldsymbol{G}_k\boldsymbol{H}_k\right)\boldsymbol{\Sigma}_{k|k-1}. \end{aligned} \tag{3.41}$$

As pointed out by Simon [132], the expression for $\boldsymbol{\Sigma}_{k|k}$ given by (3.41) might lead to numerical computing problems. Hence, we will find an alternative expression for $\boldsymbol{\Sigma}_{k|k}$.

By taking the inverse of both sides of (3.39), we have:

$$\boldsymbol{\Sigma}_{k|k}^{-1} = \left[\boldsymbol{\Sigma}_{k|k-1} - \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top \left(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right)^{-1}\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\right]^{-1}. \tag{3.42}$$

Apply the matrix inversion lemma to (3.42):

$$\begin{aligned}\boldsymbol{\Sigma}_{k|k}^{-1} &= \boldsymbol{\Sigma}_{k|k-1}^{-1} + \boldsymbol{\Sigma}_{k|k-1}^{-1}\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top \\ &\qquad \left[\left(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right) - \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{\Sigma}_{k|k-1}^{-1}\left(\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right)\right]^{-1}\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{\Sigma}_{k|k-1}^{-1} \\ &= \boldsymbol{\Sigma}_{k|k-1}^{-1} + \boldsymbol{H}_k^\top\boldsymbol{R}_k^{-1}\boldsymbol{H}_k. \end{aligned}\tag{3.43}$$

By taking the inverse of both sides of (3.43), we have:

$$\boldsymbol{\Sigma}_{k|k} = \left[\boldsymbol{\Sigma}_{k|k-1}^{-1} + \boldsymbol{H}_k^\top\boldsymbol{R}_k^{-1}\boldsymbol{H}_k\right]^{-1}. \tag{3.44}$$

Although (3.44) is more complicated in that it requires three matrix inversions, it might be computationally advantageous in some situations; see [132] pp.86-88. Also, (3.44) can be used to derive an equivalent expression for $\boldsymbol{G}$. Premultiplying the right-hand side of (3.40) by $\boldsymbol{\Sigma}_{k|k}\boldsymbol{\Sigma}_{k|k}^{-1}$, we have:

$$\boldsymbol{G}_k = \boldsymbol{\Sigma}_{k|k}\boldsymbol{\Sigma}_{k|k}^{-1}\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\left(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right)^{-1}. \tag{3.45}$$

Substituting (3.43) into (3.45), we have:

$$\begin{aligned}\boldsymbol{G}_k &= \boldsymbol{\Sigma}_{k|k}\left(\boldsymbol{\Sigma}_{k|k-1}^{-1} + \boldsymbol{H}_k^\top\boldsymbol{R}_k^{-1}\boldsymbol{H}_k\right)\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\left(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right)^{-1} \\ &= \boldsymbol{\Sigma}_{k|k}\left(\boldsymbol{H}_k^\top + \boldsymbol{H}_k^\top\boldsymbol{R}_k^{-1}\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right)\left(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right)^{-1} \\ &= \boldsymbol{\Sigma}_{k|k}\boldsymbol{H}_k^\top\left(\boldsymbol{I} + \boldsymbol{R}_k^{-1}\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right)\left(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right)^{-1} \\ &= \boldsymbol{\Sigma}_{k|k}\boldsymbol{H}_k^\top\boldsymbol{R}_k^{-1}\left(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right)\left(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right)^{-1} \\ &= \boldsymbol{\Sigma}_{k|k}\boldsymbol{H}_k^\top\boldsymbol{R}_k^{-1}. \end{aligned}\tag{3.46}$$

Now, using (3.41) and (3.46), we have:

$$\begin{aligned}\hat{\boldsymbol{x}}_{k|k} &= \left(\boldsymbol{I} - \boldsymbol{G}_k\boldsymbol{H}_k\right)\boldsymbol{\Sigma}_{k|k-1}\left[\boldsymbol{H}_k^\top\boldsymbol{R}_k^{-1}\boldsymbol{y}_k + \boldsymbol{\Sigma}_{k|k-1}^{-1}\hat{\boldsymbol{x}}_{k|k-1}\right] \\ &= \hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{G}_k\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1} + \left(\boldsymbol{I} - \boldsymbol{G}_k\boldsymbol{H}_k\right)\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\boldsymbol{R}_k^{-1}\boldsymbol{y}_k \\ &= \hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{G}_k\boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{\Sigma}_{k|k}\boldsymbol{H}_k^\top\boldsymbol{R}_k^{-1}\boldsymbol{y}_k \\ &= \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{G}_k\left(\boldsymbol{y}_k - \boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1}\right), \end{aligned}\tag{3.47}$$

where $\left(\boldsymbol{y}_k - \boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1}\right)$ is called the innovation vector. We remark that $\boldsymbol{G}_k$, $\boldsymbol{\Sigma}_{k|k}$, and $\boldsymbol{\Sigma}_{k|k}$ can be calculated offline.

The discrete time Kalman filter is summarized as follows.

Initialize the filter

$$\hat{\boldsymbol{x}}_{0|0} = E\left[\boldsymbol{x}_0\right], \tag{3.48}$$

$$\boldsymbol{\Sigma}_{0|0} = E\left[\left(\boldsymbol{x} - \hat{\boldsymbol{x}}_{0|0}\right)\left(\boldsymbol{x} - \hat{\boldsymbol{x}}_{0|0}\right)^\top\right]. \tag{3.49}$$

Then, for $k = 1, 2, ...,$

Prediction step

$$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{F}_{k-1}\hat{\boldsymbol{x}}_{k-1|k-1}, \tag{3.50}$$

$$\boldsymbol{\Sigma}_{k|k-1} = \boldsymbol{F}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{F}_{k-1}^\top + \boldsymbol{Q}_{k-1}. \tag{3.51}$$

Correction (or filtering) step

$$\boldsymbol{G}_k = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\left(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top\right)^{-1}, \tag{3.52}$$

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{G}_k\left(\boldsymbol{y}_k - \boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1}\right), \tag{3.53}$$

$$\boldsymbol{\Sigma}_{k|k} = \boldsymbol{\Sigma}_{k|k-1} - \boldsymbol{G}_k\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}. \tag{3.54}$$

In (3.50)–(3.54), $\hat{\boldsymbol{x}}_{k|k-1}$ and $\hat{\boldsymbol{x}}_{k|k}$ are the predicted and the filtered state, respectively; $\boldsymbol{\Sigma}_{k|k-1}$ and $\boldsymbol{\Sigma}_{k|k}$ are the predicted and the filtered state covariance matrix, respectively; $\boldsymbol{G}_k$ is the Kalman filter gain matrix; $\boldsymbol{y}_k$ is the measurement vector; and $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$ are the system and the observation error covariance matrix, respectively. The KF is the maximum-likelihood estimator under the aforementioned assumptions; however, its statistical efficiency strongly degrades under small deviations from those. The interested reader is referred to [132–134] for a comprehensive treatment of the Kalman filter theory.

3.2 Robust Kalman filter

The key idea of the generalized maximum-likelihood Kalman filter is to formulate the batch-mode regression (3.23) after performing the prediction step, i.e.,

$$\begin{bmatrix} \boldsymbol{y}_k \\ \hat{\boldsymbol{x}}_{k|k-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_k \\ \boldsymbol{I} \end{bmatrix}\boldsymbol{x}_k + \begin{bmatrix} \boldsymbol{v}_k \\ -\boldsymbol{e}_{k|k-1} \end{bmatrix}, \tag{3.55}$$

or in compact form (3.24),

$$\widetilde{\boldsymbol{y}}_k = \widetilde{\boldsymbol{H}}_k\boldsymbol{x}_k + \widetilde{\boldsymbol{v}}_k, \tag{3.56}$$

where $E\left[\widetilde{\boldsymbol{v}}_k\right] = \mathbf{0}$. The covariance of $\widetilde{\boldsymbol{v}}_k$ (3.28) is given by

$$\operatorname{cov}\left(\widetilde{\boldsymbol{v}}_k\right) = E\left[\widetilde{\boldsymbol{v}}_k\widetilde{\boldsymbol{v}}_k^\top\right] = \widetilde{\boldsymbol{R}}_k = \begin{bmatrix} \boldsymbol{R}_k & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{k|k-1} \end{bmatrix} = \boldsymbol{S}_k\boldsymbol{S}_k^\top, \tag{3.57}$$

where $\boldsymbol{S}_k$ is obtained from the Cholesky decomposition; its use is motivated by the fact that $\boldsymbol{\Sigma}_{k|k-1}$ is a non-diagonal matrix. Then, multiply (3.24) on the left by $\boldsymbol{S}_k^{-1}$ to perform a prewhitening,

$$\boldsymbol{S}_k^{-1}\widetilde{\boldsymbol{y}}_k = \boldsymbol{S}_k^{-1}\widetilde{\boldsymbol{H}}_k\boldsymbol{x}_k + \boldsymbol{S}_k^{-1}\widetilde{\boldsymbol{v}}_k, \tag{3.58}$$

or in a compact form,

$$\widetilde{\widetilde{\boldsymbol{y}}}_k = \widetilde{\widetilde{\boldsymbol{H}}}_k\boldsymbol{x}_k + \widetilde{\widetilde{\boldsymbol{v}}}_k, \tag{3.59}$$

where $\widetilde{\widetilde{\boldsymbol{v}}}_k = \boldsymbol{S}_k^{-1}\widetilde{\boldsymbol{v}}_k$,

$$E\left[\widetilde{\widetilde{\boldsymbol{v}}}_k\right] = \boldsymbol{S}_k^{-1}E\left[\widetilde{\boldsymbol{v}}_k\right] = \mathbf{0}, \tag{3.60}$$

$$\begin{aligned} \operatorname{cov}\left(\widetilde{\widetilde{\boldsymbol{v}}}_k\right) &= E\left[\widetilde{\widetilde{\boldsymbol{v}}}_k\widetilde{\widetilde{\boldsymbol{v}}}_k^\top\right] = E\left[\boldsymbol{S}_k^{-1}\widetilde{\boldsymbol{v}}_k\widetilde{\boldsymbol{v}}_k^\top\left(\boldsymbol{S}_k^{-1}\right)^\top\right] \\ &= \boldsymbol{S}_k^{-1}E\left[\widetilde{\boldsymbol{v}}_k\widetilde{\boldsymbol{v}}_k^\top\right]\left(\boldsymbol{S}_k^{-1}\right)^\top \\ &= \boldsymbol{S}_k^{-1}\boldsymbol{S}_k\boldsymbol{S}_k^\top\left(\boldsymbol{S}_k^\top\right)^{-1} \\ &= \boldsymbol{I}. \end{aligned} \tag{3.61}$$

The problem becomes one of estimating $\boldsymbol{x}_k$ in a certain sense, given $\widetilde{\widetilde{\boldsymbol{y}}}_k$ and $\widetilde{\widetilde{\boldsymbol{H}}}_k$. Now, the Schweppe-Huber generalized maximum-likelihood estimator, which has a convex cost function, is adopted and solved via the iterative reweighted least-squares algorithm. Since leverage points might be produced by $\widetilde{\widetilde{\boldsymbol{H}}}_k = \boldsymbol{S}_k^{-1}\widetilde{\boldsymbol{H}}_k$, the projection statistics estimator is applied on $\begin{bmatrix} \widetilde{\boldsymbol{y}}_k & \widetilde{\boldsymbol{y}}_{k-1} \end{bmatrix}$ before prewhitening (3.24). The p_s estimates are used to suppress the leverage points via the weight function

$$\varpi_i = \min\left(1,\, 1.5^2/p_{s_i}^2\right). \tag{3.62}$$

The problem becomes

$$\text{minimize} \qquad J\left(\boldsymbol{x}_k\right) = \sum_{i=1}^{m+n}\rho\left(r_{s_i}\right), \tag{3.63}$$

where

$$r_{s_i} = \frac{r_i}{s\cdot\varpi_i}, \tag{3.64}$$

r_i is the i-th element of $\boldsymbol{r}$,

$$\boldsymbol{r} = \widetilde{\widetilde{\boldsymbol{y}}}_k - \widetilde{\widetilde{\boldsymbol{H}}}_k\hat{\boldsymbol{x}}, \tag{3.65}$$

$s = 1.4826 \cdot b_m \cdot \text{median}_i \left| r_i \right|$ is a robust estimator of scale, b_m is a correction factor for unbiasedeness at the Gaussian distribution [43], and ϖ_i is computed as is (3.62). As in (1.7), the ρ-function associated with the Schweppe-Huber estimator is given by

$$\rho\left(r_{s_i}\right) = \begin{cases} \varpi_i^2 \left(\frac{1}{2} r_{s_i}^2\right), & |r_{s_i}| \leq \beta, \\ \varpi_i^2 \left(\beta |r_{s_i}| - \frac{1}{2}\beta^2\right), & \text{otherwise.} \end{cases} \tag{3.66}$$

where $\beta = 1.5$ yields a balanced trade-off between the statistical efficiency at the Gaussian distribution and under deviations from Gaussianity [68].

The necessary condition of optimality is

$$\frac{\partial J\left(\boldsymbol{x}_k\right)}{\partial \boldsymbol{x}_k} = \sum_{i=1}^{m+n} -\frac{\varpi_i \tilde{\tilde{\boldsymbol{h}}}_i}{s} \psi\left(r_{s_i}\right) = \boldsymbol{0}, \tag{3.67}$$

where $\tilde{\tilde{\boldsymbol{h}}}_i$ denotes the i-th column vector of $\tilde{\tilde{\boldsymbol{H}}}_k^{\top}$, and $\psi\left(r_{s_i}\right) = \partial \rho\left(r_{s_i}\right) / \partial r_{s_i}$ is referred to as the score function. The iteratively reweighted least squares algorithm is adopted to solve (3.67) numerically, given that it is a system of nonlinear equations. Upon convergence, the filtered state vector and covariance matrix are obtained as follows:

$$\hat{\boldsymbol{x}}_{k|k} = \left(\tilde{\tilde{\boldsymbol{H}}}_k^{\top} \boldsymbol{Q}_q \tilde{\tilde{\boldsymbol{H}}}_k\right)^{-1} \tilde{\tilde{\boldsymbol{H}}}_k^{\top} \boldsymbol{Q}_q \tilde{\tilde{\boldsymbol{y}}}_k, \tag{3.68}$$

$$\boldsymbol{\Sigma}_{k|k} = 1.0369 \left(\tilde{\tilde{\boldsymbol{H}}}_k^{\top} \tilde{\tilde{\boldsymbol{H}}}_k\right)^{-1} \left(\tilde{\tilde{\boldsymbol{H}}}_k^{\top} \boldsymbol{Q}_{\varpi} \tilde{\tilde{\boldsymbol{H}}}_k\right) \left(\tilde{\tilde{\boldsymbol{H}}}_k^{\top} \tilde{\tilde{\boldsymbol{H}}}_k\right)^{-1}, \tag{3.69}$$

where $\boldsymbol{Q}_q = \text{diag}\left(q\left(r_{s_i}\right)\right)$, $q\left(r_{s_i}\right) = \psi\left(r_{s_i}\right) / r_{s_i}$, and $\boldsymbol{Q}_{\varpi} = \text{diag}\left(\varpi_i^2\right)$.

The derivation of the expression of $\boldsymbol{\Sigma}_{k|k}$ requires the use of the influence function [135], and is a key component of the generalized maximum-likelihood Kalman filter. The interested reader is referred to [68] for a detailed derivation.

3.3 Extended Kalman filter

The extended Kalman filter (EKF) is a widely used nonlinear state estimation technique. The content of this section is mostly based on [132, 136].

Consider the system model

$$\boldsymbol{x}_k = \boldsymbol{f}_{k-1}\left(\boldsymbol{x}_{k-1}, \boldsymbol{u}_{k-1}\right) + \boldsymbol{w}_{k-1}, \tag{3.70}$$

$$\boldsymbol{y}_k = \boldsymbol{h}_k\left(\boldsymbol{x}_k\right) + \boldsymbol{v}_k. \tag{3.71}$$

As before, we rely on Assumptions 1 and 2. In addition, the following assumption is in order.

Assumption 7. $\boldsymbol{f}_k$ and $\boldsymbol{h}_k$ are known deterministic nonlinear vector-valued functions.

Since $\boldsymbol{u}_k$ is a known deterministic input vector, define $\boldsymbol{u}_{k-1} := \boldsymbol{0}$ for simplicity. Then, (3.70)-(3.71) can be written as

$$\boldsymbol{x}_k = \boldsymbol{f}\left(\boldsymbol{x}_{k-1}\right) + \boldsymbol{w}_{k-1}, \tag{3.72}$$
$$\boldsymbol{y}_k = \boldsymbol{h}\left(\boldsymbol{x}_k\right) + \boldsymbol{v}_k \tag{3.73}$$

Prediction step

We rely on the Taylor series to expand the state equation around $\boldsymbol{x}_{k-1} = \boldsymbol{x}_{k-1|k-1}$,

$$\begin{aligned}
\boldsymbol{x}_k &= \boldsymbol{f}_{k-1}\left(\hat{\boldsymbol{x}}_{k-1|k-1}\right) + \left.\frac{\partial \boldsymbol{f}_{k-1}}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{x}}_{k-1|k-1}} \left(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1|k-1}\right) + \boldsymbol{w}_{k-1} \\
&= \boldsymbol{f}_{k-1}\left(\hat{\boldsymbol{x}}_{k-1|k-1}\right) + \boldsymbol{F}_{k-1}\left(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1|k-1}\right) + \boldsymbol{w}_{k-1} \\
&= \boldsymbol{F}_{k-1}\boldsymbol{x}_{k-1} + \left[\boldsymbol{f}_{k-1}\left(\hat{\boldsymbol{x}}_{k-1|k-1}\right) - \boldsymbol{F}_{k-1}\hat{\boldsymbol{x}}_{k-1|k-1}\right] + \boldsymbol{w}_{k-1} \\
&= \boldsymbol{F}_{k-1}\boldsymbol{x}_{k-1} + \widetilde{\boldsymbol{f}}_{k-1} + \boldsymbol{w}_{k-1},
\end{aligned} \tag{3.74}$$

where $\widetilde{\boldsymbol{f}}_k = \boldsymbol{f}_k\left(\hat{\boldsymbol{x}}_{k|k}\right) - \boldsymbol{F}_k\hat{\boldsymbol{x}}_{k|k}$.

Correction (or filtering) step

Likewise, we use the Taylor series to expand the measurement equation around $\boldsymbol{x}_k = \hat{\boldsymbol{x}}_{k|k-1}$,

$$\begin{aligned}
\boldsymbol{y}_k &= \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right) + \left.\frac{\partial \boldsymbol{h}_k}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{x}}_{k|k-1}} \left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right) + \boldsymbol{v}_k \\
&= \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right) + \boldsymbol{H}_k\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right) + \boldsymbol{v}_k \\
&= \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right) - \boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{v}_k.
\end{aligned} \tag{3.75}$$

Equations (3.74) and (3.75) are linear. Thus, the standard Kalman filter can be used to estimate the state. Then, we have:

$$\boldsymbol{\Sigma}_{k|k-1} = \boldsymbol{F}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{F}_{k-1}^{\top} + \boldsymbol{Q}_{k-1}, \tag{3.76}$$
$$\boldsymbol{G}_k = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\top}\left(\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\top} + \boldsymbol{R}_k\right)^{-1}, \tag{3.77}$$
$$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{f}_{k-1}\left(\hat{\boldsymbol{x}}_{k-1|k-1}\right), \tag{3.78}$$
$$\begin{aligned}
\hat{\boldsymbol{x}}_{k|k} &= \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{G}_k\left[\boldsymbol{y}_k - \boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right) + \boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1}\right] \\
&= \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{G}_k\left[\boldsymbol{y}_k - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right)\right],
\end{aligned} \tag{3.79}$$
$$\boldsymbol{\Sigma}_{k|k} = \left(\boldsymbol{I} - \boldsymbol{G}_k\boldsymbol{H}_k\right)\boldsymbol{\Sigma}_{k|k-1}. \tag{3.80}$$

The extended Kalman filter is summarized as follows.

Initialize the filter

$$\hat{\boldsymbol{x}}_{0|0} = E\left[\boldsymbol{x}_0\right], \tag{3.81}$$

$$\boldsymbol{\Sigma}_{0|0} = E\left[\left(\boldsymbol{x} - \hat{\boldsymbol{x}}_{0|0}\right)\left(\boldsymbol{x} - \hat{\boldsymbol{x}}_{0|0}\right)^{\top}\right]. \tag{3.82}$$

Then, for $k = 1, 2, ...,$

Prediction step

Compute the partial derivative matrices and perform the time update of state estimate and estimation-error covariance matrix as follows:

$$\boldsymbol{F}_{k-1} = \left.\frac{\partial \boldsymbol{f}_{k-1}}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{x}}_{k-1|k-1}}, \tag{3.83}$$

$$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{f}_{k-1}\left(\hat{\boldsymbol{x}}_{k-1|k-1}\right), \tag{3.84}$$

$$\boldsymbol{\Sigma}_{k|k-1} = \boldsymbol{F}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{F}_{k-1}^{\top} + \boldsymbol{Q}_{k-1}. \tag{3.85}$$

Correction (or filtering) step

Compute the partial derivative matrices and perform the measurement update of the state estimate and estimation-error covariance, that is,

$$\boldsymbol{H}_k = \left.\frac{\partial \boldsymbol{h}_k}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{x}}_{k|k-1}}, \tag{3.86}$$

$$\boldsymbol{G}_k = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\top}\left(\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\top} + \boldsymbol{R}_k\right)^{-1}, \tag{3.87}$$

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{G}_k\left[\boldsymbol{y}_k - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right)\right], \tag{3.88}$$

$$\boldsymbol{\Sigma}_{k|k} = \left(\boldsymbol{I} - \boldsymbol{G}_k\boldsymbol{H}_k\right)\boldsymbol{\Sigma}_{k|k-1}. \tag{3.89}$$

The EKF method has been widely used as it presents good performance when the nonlinearities are not too strong. However, its main drawback resides in the linearization process and Jacobian matrix calculation because of the following:

- When the higher order terms of the Taylor series expansion are not negligible, the linearized approximation might be poor. In this case, if the time step is not set small enough, the filter might be unstable.
- The linearization process depends on the existence of the Jacobian matrix. However, the Jacobian might not exist in some cases, and then it is impossible to perform linearization during the filtering process.

Furthermore, it is well-known that the EKF is simply an extension of linear Kalman filter theory to nonlinear systems [136]. In turn, the Kalman filter is the optimal estimator when the noise is Gaussian (i.e., in this case the Kalman filter can be shown to be the least-squares

estimator). Although it is the best *linear* estimator when the noise is Gaussian, there are nonlinear filters that give better solutions [132]. Examples include the unscented Kalman filter, the particle filter, and the Ensemble Kalman Filter. A robust version of the extended Kalman filter is presented in chapter 4.

3.4 Iterated extended Kalman filter

The iterated extended Kalman filter (IEKF) provides a way to decrease the errors that are due to the linearization process in the EKF. The IEKF has overall better estimation performance as compared to the EKF, specially under high nonlinearities. But this improved performance comes at the cost of higher complexity and computational burden.

When deriving the discrete time EKF, we approximated $\boldsymbol{h}_k(\boldsymbol{x}_k)$ via an expansion in Taylor series around $\hat{\boldsymbol{x}}_{k|k-1}$ (3.75),

$$\boldsymbol{h}_k(\boldsymbol{x}_k) = \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right) + \left.\frac{\partial \boldsymbol{h}_k}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{x}}_{k|k-1}} \left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right) \tag{3.90}$$

$$= \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right) + \boldsymbol{H}_k\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right). \tag{3.91}$$

Based on this linearization, we have that

$$\boldsymbol{G}_k = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top \left(\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^\top + \boldsymbol{R}_k\right)^{-1}, \tag{3.92}$$

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{G}_k\left[\boldsymbol{y}_k - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right)\right], \tag{3.93}$$

$$\boldsymbol{\Sigma}_{k|k} = \left(\boldsymbol{I} - \boldsymbol{G}_k\boldsymbol{H}_k\right)\boldsymbol{\Sigma}_{k|k-1}. \tag{3.94}$$

The reason that we expanded $\boldsymbol{h}_k(\boldsymbol{x}_k)$ around $\hat{\boldsymbol{x}}_{k|k-1}$ was because that was our best estimate of $\boldsymbol{x}_k$ before the measurement at time k is taken into account. But after we implement the discrete EKF equations to obtain the *a posteriori* estimate $\hat{\boldsymbol{x}}_{k|k}$, we have a better estimate of $\boldsymbol{x}_k$. So we can reduce the linearization error by reformulating the Taylor series expansion of $\boldsymbol{h}_k(\boldsymbol{x}_k)$ around our new estimate. If we then use that new Taylor series expansion of $\boldsymbol{h}_k(\boldsymbol{x}_k)$ and recalculate the measurement-update equations, we should get a better a posteriori estimate of $\hat{\boldsymbol{x}}_{k|k}$. But then we can repeat the previous step; since we have an even better estimate of $\boldsymbol{x}_k$, we can again reformulate the expansion of $\boldsymbol{h}_k(\boldsymbol{x}_k)$ around this even better estimate to get an even better estimate. This process can be repeated as many times as desired, although for most problems the majority of the possible improvement is obtained by only re-linearizing one time.

In what follows, $\hat{\boldsymbol{x}}_{k|k,i}$ denotes the *a posteriori* estimate of $\boldsymbol{x}_k$ after i relinearizations. So $\hat{\boldsymbol{x}}_{k|k,0}$ is the *a posteriori* estimate that results from the use of the standard EKF. Likewise, $\boldsymbol{\Sigma}_{k|k,i}$ denotes the approximate estimation-error covariance of $\hat{\boldsymbol{x}}_{k|k,i}$, $\boldsymbol{G}_{k,i}$ denotes the Kalman gain that is used during the i-th relinearization step, and $\boldsymbol{H}_{k,i}$ denotes the matrix of partial

derivatives evaluated at $\boldsymbol{x}_k = \hat{\boldsymbol{x}}_{k|k,i}$. Using this notation, iterated EKF algorithm is described as follows.

At each time step k, initialize the iterated EKF estimates using the standard EKF estimates, i.e.,

$$\hat{\boldsymbol{x}}_{k|k,0} = \hat{\boldsymbol{x}}_{k|k}, \tag{3.95}$$
$$\boldsymbol{\Sigma}_{k|k,0} = \boldsymbol{\Sigma}_{k|k}. \tag{3.96}$$

Then, for $i = 0, 1, ...,$ evaluate:

$$\boldsymbol{H}_{k,i} = \frac{\partial \boldsymbol{h}_k}{\partial \boldsymbol{x}}\Big|_{\hat{\boldsymbol{x}}_{k|k,i}}, \tag{3.97}$$
$$\boldsymbol{G}_{k,i} = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_{k,i}^\top \left(\boldsymbol{H}_{k,i}\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_{k,i}^\top + \boldsymbol{R}_k\right)^{-1}, \tag{3.98}$$
$$\hat{\boldsymbol{x}}_{k|k,i+1} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{G}_{k,i}\left[\boldsymbol{y}_k - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right)\right], \tag{3.99}$$
$$\boldsymbol{\Sigma}_{k|k,i+1} = \left(\boldsymbol{I} - \boldsymbol{G}_{k,i}\boldsymbol{H}_{k,i}\right)\boldsymbol{\Sigma}_{k|k-1}. \tag{3.100}$$

This is done iteratively to improve the linearization. If only one iteration is considered, the iterated EKF reduces to the standard EKF. An additional modification to the above equations is necessary to obtain the iterated Kalman filter. Recall that in the derivation of the EKF, the expression for $\hat{\boldsymbol{x}}$ was obtained from the first-order Taylor series expansion of the measurement equation:

$$\begin{aligned}\boldsymbol{y_k} &= \boldsymbol{h}_k\left(\boldsymbol{x}_k\right) \\ &\approx \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right) + \boldsymbol{H}_k\big|_{\hat{\boldsymbol{x}}_{k|k-1}}\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right).\end{aligned} \tag{3.101}$$

The measurement-update equation for $\hat{\boldsymbol{x}}$ was obtained by evaluating the right side of (3.101) at the *a priori* estimate $\hat{\boldsymbol{x}}_{k|k-1}$, and subtracting it from $\boldsymbol{y_k}$, thereby finding the residual

$$\begin{aligned}\boldsymbol{r}_k &= \boldsymbol{y_k} - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right) - \boldsymbol{H}_k\big|_{\hat{\boldsymbol{x}}_{k|k-1}}\left(\hat{\boldsymbol{x}}_{k|k-1} - \hat{\boldsymbol{x}}_{k|k-1}\right) \\ &= \boldsymbol{y_k} - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k-1}\right).\end{aligned} \tag{3.102}$$

For the iterated EKF, one expands the measurement equation around $\hat{\boldsymbol{x}}_{k|k,i}$ as follows:

$$\boldsymbol{y_k} \approx \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k,i}\right) + \boldsymbol{H}_k\big|_{\hat{\boldsymbol{x}}_{k|k,i}}\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k,i}\right). \tag{3.103}$$

Then, the iterated EKF equation for $\hat{\boldsymbol{x}}$ is found by evaluating the right side of (3.103) at $\hat{\boldsymbol{x}}_{k|k-1}$, and subtracting it from $\boldsymbol{y_k}$ to get the correction term:

$$\boldsymbol{r}_k = \boldsymbol{y_k} - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k,i}\right) - \boldsymbol{H}_{k,i}\left(\hat{\boldsymbol{x}}_{k|k-1} - \hat{\boldsymbol{x}}_{k|k,i}\right). \tag{3.104}$$

This gives the iterated EKF update equation for $\hat{\boldsymbol{x}}$ as

$$\hat{\boldsymbol{x}}_{k|k,i+1} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{G}_{k,i}\left[\boldsymbol{y_k} - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k,i}\right) - \boldsymbol{H}_{k,i}\left(\hat{\boldsymbol{x}}_{k|k-1} - \hat{\boldsymbol{x}}_{k|k,i}\right)\right]. \tag{3.105}$$

The iterated extended Kalman filter is summarized as follows.

Initialize the filter

$$\hat{\boldsymbol{x}}_{0|0} = E\left[\boldsymbol{x}_0\right], \tag{3.106}$$

$$\boldsymbol{\Sigma}_{0|0} = E\left[\left(\boldsymbol{x} - \hat{\boldsymbol{x}}_{0|0}\right)\left(\boldsymbol{x} - \hat{\boldsymbol{x}}_{0|0}\right)^\top\right]. \tag{3.107}$$

Then, for $k = 1, 2, ...,$

Prediction step

Compute the partial derivative matrices and perform the time update of state estimate and estimation-error covariance matrix as follows:

$$\boldsymbol{F}_{k-1} = \left.\frac{\partial \boldsymbol{f}_{k-1}}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{x}}_{k-1|k-1}}, \tag{3.108}$$

$$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{f}_{k-1}\left(\hat{\boldsymbol{x}}_{k-1|k-1}\right), \tag{3.109}$$

$$\boldsymbol{\Sigma}_{k|k-1} = \boldsymbol{F}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{F}_{k-1}^\top + \boldsymbol{Q}_{k-1}. \tag{3.110}$$

Correction (or filtering) step

Initialize the correction step of the iterated EKF using the standard EKF estimate.

$$\hat{\boldsymbol{x}}_{k|k,0} = \hat{\boldsymbol{x}}_{k|k-1}, \tag{3.111}$$

$$\boldsymbol{\Sigma}_{k|k,0} = \boldsymbol{\Sigma}_{k|k-1}. \tag{3.112}$$

For $i = 1, 2, ..., N$, where N is the number of iterations, evaluate the following expressions:

$$\boldsymbol{H}_{k,i} = \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{x}}_{k|k,i}}, \tag{3.113}$$

$$\boldsymbol{G}_{k,i} = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_{k,i}^\top\left(\boldsymbol{H}_{k,i}\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_{k,i}^\top + \boldsymbol{R}_k\right)^{-1}, \tag{3.114}$$

$$\hat{\boldsymbol{x}}_{k|k,i+1} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{G}_{k,i}\left[\boldsymbol{y}_k - \boldsymbol{h}_k\left(\hat{\boldsymbol{x}}_{k|k,i}\right) - \boldsymbol{H}_{k,i}\left(\hat{\boldsymbol{x}}_{k|k-1} - \hat{\boldsymbol{x}}_{k|k,i}\right)\right], \tag{3.115}$$

$$\boldsymbol{\Sigma}_{k|k,i+1} = \left(\boldsymbol{I} - \boldsymbol{G}_{k,i}\boldsymbol{H}_{k,i}\right)\boldsymbol{\Sigma}_{k|k-1}. \tag{3.116}$$

Finally, the *a posteriori* state estimate and estimation-error covariance matrix of the iterated EKF are given by:

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k,N+1}, \tag{3.117}$$

$$\boldsymbol{\Sigma}_{k|k} = \boldsymbol{\Sigma}_{k|k,N+1}. \tag{3.118}$$

3.5 Concluding remarks

In this chapter, the classical as well as the robust version of the discrete-time Kalman filter were presented, followed by the extended Kalman filter and the *iterated* extended Kalman filter. There are several other Kalman filtering techniques, including the unscented Kalman filter, the ensemble Kalman filter, and the particle filter, to name a few.

Robust versions of the extended Kalman filter and of the *iterated* extended Kalman filter introduced in this chapter will be used in chapter 4 to develop robust, centralized, dynamic state estimators for power systems.

The robust Kalman filter introduced in this chapter will be used in chapter 5 to develop a robust, hierarchical decentralized dynamic state estimator. The later also relies on the Koopman operator-theoretic framework developed in chapter 2.

Chapter 4

Robust centralized dynamic state estimator

The Kalman filtering theory that was briefly presented in chapter 3 is a requirement in what follows. In this chapter, we present the first contribution of this dissertation, namely, the development of a robust version of the extended Kalman filter for power system dynamic state estimation [50, 51].

With the rapid deployment of PMUs along with high speed data communication links with large bandwidth, tracking in real-time the dynamics of a power system state is becoming possible. The natural tool to achieve this function is a dynamic state estimator (DSE). The DSE will provide a situational awareness [137] of the evolution of the system state and will pave the way to real-time wide-area monitoring and control [138]. The prerequisites are sufficient observability of the system, robust state estimation and a carefully maintained database. Among the various techniques being proposed in the literature for DSE, the ones based on the Kalman filter have received a great deal of attention. In particular, for nonlinear systems, the EKF is the most widely used technique. The reader is referred to [139] for a comprehensive comparative study. In [48], using the so-called single-machine infinite-bus system and the fourth-order generator model, the authors applied the EKF to estimate the rotor angles, rotor speeds and the direct and quadrature components of the generators internal voltage. In particular, a very important practical problem is addressed where the field voltage is not accessible to metering due to the presence of brushless excitation systems. However, one problem that is rarely addressed in the literature is the robustness of the state estimator to gross measurement errors, which might strongly bias the classical EKF.

As pointed out by Gandhi and Mili [68], outliers can be classified into three types, namely observation, innovation and structural outliers as shown in Table I. In power system dynamic state estimation, observation outliers might originate from either communication impulsive noise, or the loss of the PMUs' communications links, or an imperfect phasor synchronization in PMU measurements, or large biases caused by the saturation of metering current

transformers or by metering coupling capacitor voltage transformers, to name a few. As for the innovation outliers, they might be induced in several different ways. For instance, some of the generators are not well calibrated and their simulated outputs are not consistent with the measurements, like the 1996 blackout, where the model predicted a stable system state, while in reality it was highly unstable, resulting in the collapse of the system within minutes [140, 141]. Innovation outliers might also be induced by the approximations in the state prediction model or impulsive system process noise [142, 143]. By contrast, structural outliers might be caused by system parameter or topological errors. In this paper, we propose a robust GM-IEKF that can handle observation and innovation outliers but not structural outliers, which require a different formulation that is not addressed here.

Table 4.1: Definition of the three types of outlier

Type of outlier	Type of error	Affected elements
Observation outlier	Observation error, $\boldsymbol{v}_k$	$\boldsymbol{y}_k$
Innovation outlier	Process error, $\boldsymbol{w}_{k-1}$	$\hat{\boldsymbol{x}}_{k\|k-1}$
Structural outlier	Structural errors in $\boldsymbol{f}(\boldsymbol{x}_{k-1})$, $\boldsymbol{h}(\boldsymbol{x}_k)$	$\boldsymbol{y}_k$, $\hat{\boldsymbol{x}}_{k\|k-1}$, $\boldsymbol{\Sigma}_{k\|k-1}$, $\boldsymbol{\Sigma}_{k\|k}$

4.1 Robust extended Kalman filter

Our objective is to estimate the system state variables using the discrete time dynamical system model with additive error as defined in (3.72)-(3.73), expressed as

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_{k-1}, \tag{4.1}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) + \boldsymbol{v}_k. \tag{4.2}$$

The dynamical system model in (4.1)-(4.2) is general and can be applied to almost ant dynamical system. In what follows, we will rely on two power system models that can be put in the form of (4.1)-(4.2), and we will refer to them as *model A* and *model B*. The set of differential-algebraic equations of both models is given in Appendix A. Model A is known as the classical model or the swing equation; while it is useful for research at the initial stage, its degree of fidelity to real systems is very limited. This is the reason why we also consider model B, which is more detailed and de facto applied in the industry. The observation vector $\boldsymbol{y}$ is composed by measurements of the generators' terminal active and reactive power, and of the buses' voltage magnitude and angle, obtained or calculated from PMUs.

The proposed robust EKF method is developed in five main steps as follows. Firstly, the classical EKF recursive approach is converted into a batch-mode regression form such that the observations and the predictions are processed simultaneously, resulting in an enhanced data redundancy. This redundancy allows our estimator to suppress both innovation and observation outliers while exhibiting good tracking capabilities of the state dynamics. Secondly, the outliers in the data are identified by means of projection statistics. Thirdly, a

prewhitening of the data is performed to uncorrelate the prediction and the observation errors even in presence of outliers. Fourthly, a generalized maximum likelihood criterion based on the Schweppe-Huber convex ρ-function is minimized via the iteratively reweighted least squares (IRLS) algorithm. Finally, the error covariance matrices of the predictions and the state estimates are updated using expressions derived from the total influence function of the GM-estimator.

4.1.1 Batch-mode regression form

Given the filtered state vector $\hat{\boldsymbol{x}}_{k-1|k-1}$ at time step $k-1$, with covariance matrix $\boldsymbol{\Sigma}_{k-1|k-1}$, the state at the next time step k is predicted through (3.84) and (3.85). From (3.11), we have that the prediction error at time step k is given by

$$\boldsymbol{e}_{k|k-1} = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}. \tag{4.3}$$

Put (4.2) and (4.3) into a matrix form

$$\begin{bmatrix} \hat{\boldsymbol{x}}_{k|k-1} \\ \boldsymbol{y}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{h}\left(\boldsymbol{x}_k\right) \end{bmatrix} + \begin{bmatrix} -\boldsymbol{e}_k \\ \boldsymbol{v}_k \end{bmatrix}, \tag{4.4}$$

which can be rewritten in the following compact form:

$$\widetilde{\boldsymbol{y}}_k = \widetilde{\boldsymbol{h}}\left(\boldsymbol{x}_k\right) + \widetilde{\boldsymbol{v}}_k, \tag{4.5}$$

with the error covariance matrix given by

$$\operatorname{cov}\left(\widetilde{\boldsymbol{v}}_k\right) = E\left[\widetilde{\boldsymbol{v}}_k \widetilde{\boldsymbol{v}}_k^{\top}\right] = \widetilde{\boldsymbol{R}}_k = \begin{bmatrix} \boldsymbol{R}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{k|k-1} \end{bmatrix}, \tag{4.6}$$

and $\boldsymbol{\Sigma}_{k|k-1}$ is a full matrix. Hence, it is required to perform a prewhitening to uncorrelate the error of the predicted states. However, if we directly perform the prewhitening step, the outliers will corrupt the results [68]. Instead, we first detect and downweight the outliers in the derived batch-mode regression form by means of a statistical test applied to the projection statistics, which we present next.

4.1.2 Projection statistic

Following Mili et al. [43], consider a set of m data points, $\{\boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_m\}$, where each data point is an n-dimensional vector. The objective here is to estimate the center and the dispersion of this point cloud, respectively, by means of a multivariate location estimator and a covariance matrix.

Suppose that $\{\boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_m\}$ follows a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ denotes the mean value, and $\boldsymbol{\Sigma}$ denotes the covariance matrix. The probability distribution, in this case, is given by

$$f(\boldsymbol{\ell}) = (2\pi)^{-m/2} \left(\det\left(\boldsymbol{\Sigma}\right)\right)^{-1/2} e^{\frac{1}{2}(\boldsymbol{\ell}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\ell}-\boldsymbol{\mu})}. \tag{4.7}$$

Under the Gaussian model, the maximum likelihood estimate of $\boldsymbol{\mu}$ is the sample mean,

$$\bar{\boldsymbol{\ell}} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{\ell}_i, \tag{4.8}$$

and that of $\boldsymbol{\Sigma}$ is the sample covariance matrix. An unbiased estimator of the sample covariance matrix is given by

$$\boldsymbol{C} = \frac{1}{m-1}\sum_{i=1}^{m} \left(\boldsymbol{\ell}_i - \bar{\boldsymbol{\ell}}\right)\left(\boldsymbol{\ell}_i - \bar{\boldsymbol{\ell}}\right)^\top. \tag{4.9}$$

Both estimators $\bar{\boldsymbol{\ell}}$ and $\boldsymbol{C}$ are affine equivariant, that is, do not depend on the choice of the coordinate system and are statistically efficient at the Gaussian distribution.

In order to identify the outliers of the point cloud in n-dimensions, we need a measure of the distance of each point with respect to the bulk of the point cloud. The conventional measure is provided by the Mahalanobis distance, which makes use of the sample mean (4.8) and the sample covariance (4.9). The Mahalanobis distance of a point $\boldsymbol{\ell}_i$ is a scalar, defined as

$$m_{d_i} = \sqrt{\left(\boldsymbol{\ell}_i - \bar{\boldsymbol{\ell}}\right)^\top \boldsymbol{C}^{-1} \left(\boldsymbol{\ell}_i - \bar{\boldsymbol{\ell}}\right)}. \tag{4.10}$$

Suppose now that the vectors $\boldsymbol{\ell}_i$ are Gaussian with a distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In that case, the $m_{d_i}^2$ have approximately a chi-square distribution with n degrees of freedom, χ_n^2. Therefore, there is a probability of approximately $1-\alpha$ that a point $\boldsymbol{\ell}_i$ will fall inside the tolerance ellipsoid satisfying $m_{d_i}^2 \leq \chi_{n,1-\alpha}^2$. This provides the rationale of the classical method, which tags as outliers all data points whose Mahalanobis distances are larger than a cutoff value, say larger than $\left(\chi_{n,0.975}^2\right)^{1/2}$. Because the m_{d_i} are based on non-robust statistics, they are prone to the masking effect of multiple outliers, especially when the latter appear in clusters. In the latter case, the sample mean is attracted by the outliers and the sample covariance matrix is inflated, so that the tolerance ellipsoid covers in part or in totality the set of outliers, inducing the failure of m_{d_i} to reveal some or all of them.

The first affine equivariant multivariate estimator with a high breakdown point was independently proposed by Stahel [144] and Donoho [145]. See [146] pp. 238-239. It is motivated by the following expression of the Mahalanobis distances:

$$m_{d_i} = \max_{\|\boldsymbol{v}=1\|} \frac{\left|\boldsymbol{\ell}_i^\top \boldsymbol{v} - L\left(\boldsymbol{\ell}_1^\top \boldsymbol{v}, ..., \boldsymbol{\ell}_m^\top \boldsymbol{v}\right)\right|}{S\left(\boldsymbol{\ell}_1^\top \boldsymbol{v}, ..., \boldsymbol{\ell}_m^\top \boldsymbol{v}\right)}, \tag{4.11}$$

where L and S are the sample mean and the sample standard deviation of the projections of the data points $\boldsymbol{\ell}_i$ on the direction of the vector $\boldsymbol{v}$. Here, the equality (4.11) holds when all possible directions $\boldsymbol{v}$ are considered. A natural way to robustify (4.11) is to use robust estimators of location and scale for L and S, respectively. Gasko and Donoho [147, 148] suggest utilizing the sample median for L and the median-absolute-deviation (MAD) for S. Note that the MAD is aimed at symmetric distributions. Since in practice not all the directions can be investigated, Gasko and Donoho [147] advocate the use of the *projection* algorithm. The idea is to investigate only those directions $\boldsymbol{v}$ that originate from the coordinate-wise median $\boldsymbol{M}$ and pass through each of the data points $\boldsymbol{\ell}_i$. Formally, we have $\boldsymbol{v} = \boldsymbol{\ell}_i - \boldsymbol{M}$, where

$$\boldsymbol{M} = \left[\underset{j}{\text{median}}\,(\ell_{j1}) \;\; ... \underset{j}{\text{median}}\,(\ell_{jm})\right]^\top, \tag{4.12}$$

for $j = 1, ..., m$. The resulting distances are the so-called *projections statistics*. Formally, we have

$$p_{s_i} = \max_{\|\boldsymbol{v}\|=1} \frac{\left|\boldsymbol{\ell}_i^\top \boldsymbol{v} - \underset{j}{\text{median}}\left(\boldsymbol{\ell}_j^\top \boldsymbol{v}\right)\right|}{1.4826 \cdot \underset{k}{\text{median}}\left|\boldsymbol{\ell}_k^\top \boldsymbol{v} - \underset{j}{\text{median}}\left(\boldsymbol{\ell}_j^\top \boldsymbol{v}\right)\right|}. \tag{4.13}$$

Note that because the MAD is affine equivariant, $\boldsymbol{v}$ need not be of unit length. Following Donoho and Gasko, the projection statistic of a point $\boldsymbol{\ell}_i$ indicates how far the point is from the bulk of the data set in the worst one-dimensional projection. A two-dimensional idealization of the projection statistics is depicted in Fig. 4.1.

The p_s values are some kinds of robust distances of a collection of data points, $\boldsymbol{\ell}_i$, defined as realizations of a random vector. Outliers are data points that are distant from the bulk of the point cloud. To detect outliers, we propose to apply the p_s to a 2-dimensional matrix $\boldsymbol{Z}$ that contains serially correlated samples of the innovations and of the predicted state vector. Formally, we have

$$\boldsymbol{Z} = \begin{bmatrix} \hat{\boldsymbol{x}}_{k-1|k-2} & \hat{\boldsymbol{x}}_{k|k-1} \\ \boldsymbol{y}_{k-1} - \boldsymbol{h}\left(\hat{\boldsymbol{x}}_{k-1|k-2}\right) & \boldsymbol{y}_k - \boldsymbol{h}\left(\hat{\boldsymbol{x}}_{k|k-1}\right) \end{bmatrix}, \tag{4.14}$$

where $\boldsymbol{y}_{k-1} - \boldsymbol{h}\left(\hat{\boldsymbol{x}}_{k-1|k-2}\right)$ and $\boldsymbol{y}_k - \boldsymbol{h}\left(\hat{\boldsymbol{x}}_{k|k-1}\right)$ are the innovation vectors, while $\hat{\boldsymbol{x}}_{k-1|k-2}$ and $\hat{\boldsymbol{x}}_{k|k-1}$ are the predicted state vectors at time instants $(k-1)$ and k, respectively. We might also apply the p_s to higher dimensional samples, but we found that 2 dimensions are enough to identify outliers. The p_s values of the predictions and of the innovations are separately calculated because, as shown in Fig. 4.2, the values taken by the former are centered around one while those taken by the latter are centered around zero. Before calculating the p_s, the real powers produced by the generators are normalized using their own rated MVA values.

Once the p_s values are calculated, they are compared to a threshold to identify the outliers. Fig. 4.3 provides a typical example where the outliers stand far away from the bulk of

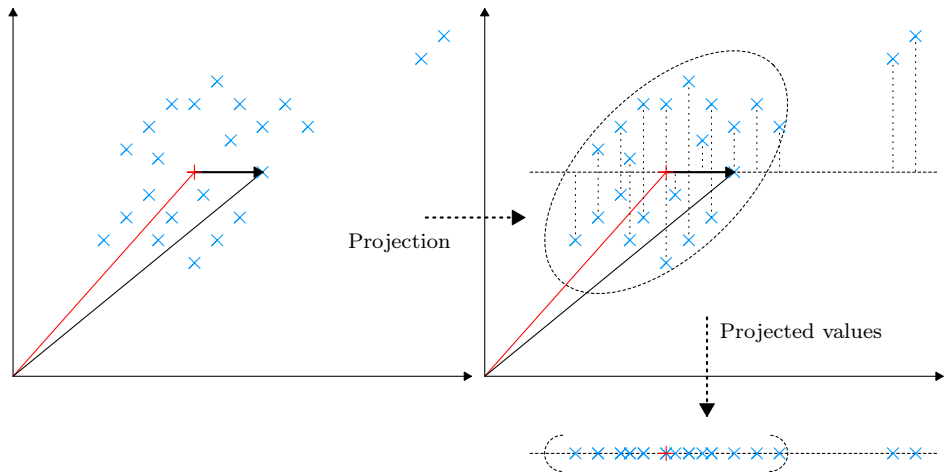


Figure 4.1: Two-dimensional idealization of the projection statistics. (left) Compute the median (+) of the point cloud (x). (right) Project the data points to all one-dimensional directions defined by the coordinate-wise median and that pass through the data points. At the end, assign to each point the maximum of the corresponding standardized projections [43].

innovations. The threshold is determined from the probability distribution of the p_s under the assumption that the data points follow a bivariate Gaussian probability distribution.

To this end, we conduct extensive Monte Carlo simulations and Q-Q plots. Consider two random variables $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ that are independent and identically distributed according to $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Then generate 1,000 realizations of these two random variables, apply the p_s to $[\boldsymbol{v}_1\, \boldsymbol{v}_2]$ and repeat the procedure 100 times. The sample medians and the interquantile ranges of the empirical p_s quantiles are finally plotted versus the corresponding quantiles of the chi-square probability density function with two degrees of freedom. The Q-Q plots displayed in Fig. 4.4 provide evidence that p_s follows that distribution. We pick the threshold ζ of the statistical test equal to $\chi^2_{\nu,\beta} = \chi^2_{2,0.975}$ at a significance level of 97.5%. The detected outliers, whose values satisfy $p_{s_i} > \zeta$, are downweighted via

$$\varpi_i = \min\left(1, d^2/p_{s_i}^2\right). \tag{4.15}$$

Choosing d=1.5 yields good statistical efficiency at the Gaussian distribution without increasing too much the bias induced by outliers [68].

4.1.3 Robust prewhitening

Apply the Cholesky decomposition to (4.6) in order to find $\widetilde{\boldsymbol{R}}_k = \boldsymbol{S}_k \boldsymbol{S}_k^\top$. Finally, the prewhitening of $\widetilde{\boldsymbol{v}}_k$ in (4.5) is performed by pre-multiplying that nonlinear regression model

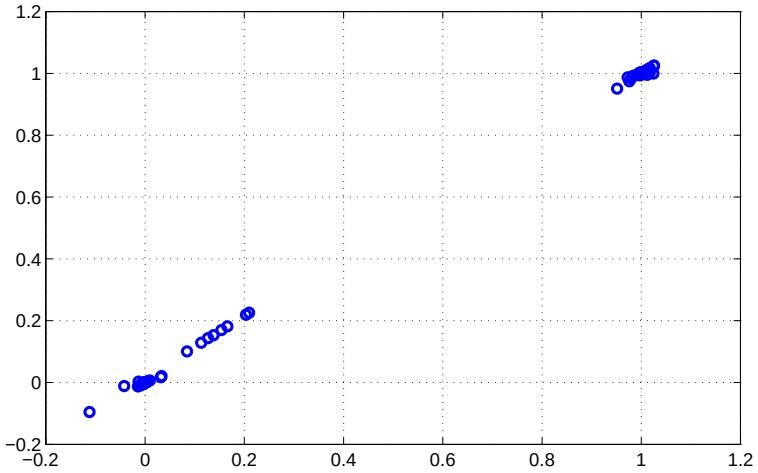


Figure 4.2: Scatter plot of the matrix $\boldsymbol{Z}$ without outliers

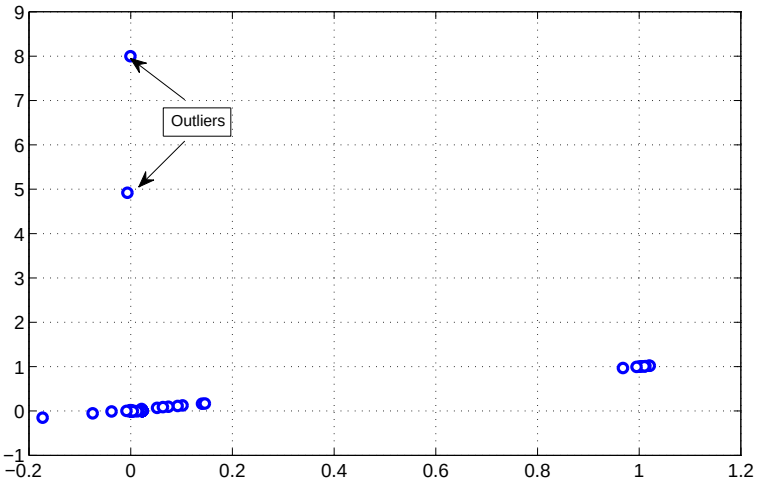


Figure 4.3: Scatter plot of the matrix $\boldsymbol{Z}$ with two outliers

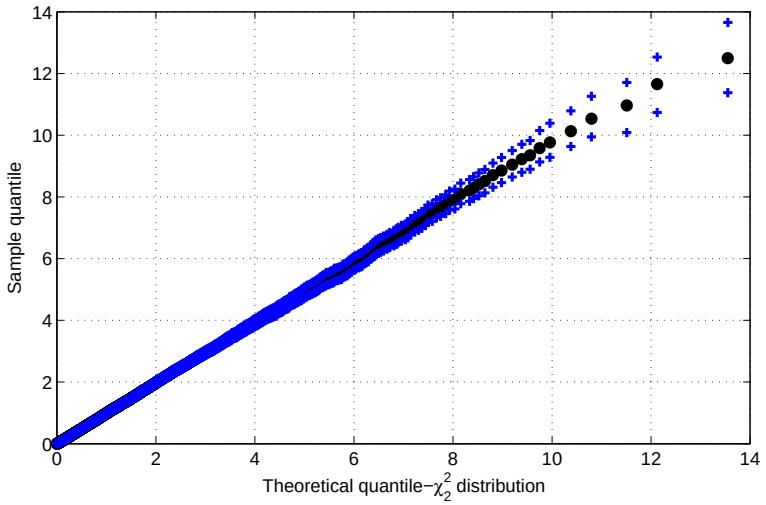


Figure 4.4: Q-Q plots of the sample quantiles of the p_s vs. the corresponding quantiles of the χ^2_2 distribution

by $\boldsymbol{S}_k^{-1}$, resulting in

$$\boldsymbol{S}_k^{-1}\widetilde{\boldsymbol{y}}_k = \boldsymbol{S}_k^{-1}\widetilde{\boldsymbol{h}}\left(\boldsymbol{x}_k\right) + \boldsymbol{S}_k^{-1}\widetilde{\boldsymbol{v}}_k, \tag{4.16}$$

which can be written in a compact form as

$$\widetilde{\widetilde{\boldsymbol{y}}}_k = \widetilde{\widetilde{\boldsymbol{h}}}\left(\boldsymbol{x}_k\right) + \widetilde{\widetilde{\boldsymbol{v}}}_k. \tag{4.17}$$

4.1.4 Robust filtering and solution

To bound the influence of outliers on the state estimates, we minimize an objective function given by

$$J\left(\boldsymbol{x}\right) = \sum_{i=1}^{m+n} \rho\left(r_{s_i}\right), \tag{4.18}$$

where the standardized residual

$$r_{s_i} = \frac{r_i}{s \cdot \varpi_i}, \tag{4.19}$$

r_i is the i-th element of $\boldsymbol{r}$,

$$\boldsymbol{r} = \widetilde{\widetilde{\boldsymbol{y}}}_k - \widetilde{\widetilde{\boldsymbol{H}}}_k\hat{\boldsymbol{x}}, \tag{4.20}$$

$s = 1.4826 \cdot b_m \cdot \text{median}_i\,|r_i|$ is a robust estimator of scale, b_m is a correction factor for unbiasedness at the Gaussian distribution [43], and ϖ_i is calculated as in (4.15). As in (1.7), the convex ρ-function associated with the Schweppe-Huber estimator is given by

$$\rho\left(r_{s_i}\right) = \begin{cases} \varpi_i^2\left(\frac{1}{2}r_{s_i}^2\right), & |r_{s_i}| \leq \beta, \\ \varpi_i^2\left(\beta|r_{s_i}| - \frac{1}{2}\beta^2\right), & \text{otherwise}, \end{cases} \tag{4.21}$$

where $\beta = 1.5$ leads to high statistical efficiency under Gaussian noise [68]. To minimize (4.18), one takes its partial derivative and sets it equal to zero, yielding

$$\frac{\partial J(\boldsymbol{x})}{\partial \boldsymbol{x}} = \sum_{i=1}^{m+n} -\frac{\varpi_i \tilde{\tilde{\boldsymbol{h}}}_i}{s} \psi(r_{s_i}) = \boldsymbol{0}, \tag{4.22}$$

where $\tilde{\tilde{\boldsymbol{h}}}_i$ denotes the i-th column vector of $\tilde{\tilde{\boldsymbol{H}}}^\top$, $\tilde{\tilde{\boldsymbol{H}}} = \partial \tilde{\tilde{\boldsymbol{h}}} / \partial \boldsymbol{x}|_{\boldsymbol{x}=\hat{\boldsymbol{x}}}$, and $\psi(r_{s_i}) = \partial \rho(r_{s_i}) / \partial r_{s_i}$. Then, by dividing and multiplying by r_{s_i} on both sides of (4.22), we get

$$\tilde{\tilde{\boldsymbol{H}}}^\top \boldsymbol{Q}_q \left(\tilde{\tilde{\boldsymbol{y}}} - \tilde{\tilde{\boldsymbol{h}}}(\boldsymbol{x}) \right) = \boldsymbol{0}, \tag{4.23}$$

where $\boldsymbol{Q}_q = \operatorname{diag}(q(r_{s_i}))$, and $q(r_{s_i}) = \psi(r_{s_i}) / r_{s_i}$.

Taking a first-order Taylor series expansion of $\tilde{\tilde{\boldsymbol{h}}}(\boldsymbol{x})$ around $\hat{\boldsymbol{x}}_{k|k}$ and using the IRLS algorithm [135], the state vector correction at the j-th iteration is calculated by

$$\Delta \hat{\boldsymbol{x}}_{k|k}^{(j+1)} = \left(\tilde{\tilde{\boldsymbol{H}}}^\top \boldsymbol{Q}_q^{(j)} \tilde{\tilde{\boldsymbol{H}}} \right)^{-1} \tilde{\tilde{\boldsymbol{H}}}^\top \boldsymbol{Q}_q^{(j)} \left(\tilde{\tilde{\boldsymbol{y}}} - \tilde{\tilde{\boldsymbol{h}}} \left(\boldsymbol{x}_{k|k}^{(j)} \right) \right), \tag{4.24}$$

where $\Delta \hat{\boldsymbol{x}}_{k|k}^{(j+1)} = \hat{\boldsymbol{x}}_{k|k}^{(j+1)} - \hat{\boldsymbol{x}}_{k|k}^{(j)}$, and $\tilde{\tilde{\boldsymbol{H}}}$ is evaluated at $\hat{\boldsymbol{x}}_{k|k}^{(j)}$. The algorithm converges when $\left\| \Delta \hat{\boldsymbol{x}}_{k|k}^{(j+1)} \right\|_\infty \leq 10^{-2}$.

4.1.5 Error covariance matrix

After the convergence of the iterative process, the estimation error covariance matrix $\boldsymbol{\Sigma}_{k|k}$ needs to be updated, thus allowing us to predict the state at the next time step. The updating of $\boldsymbol{\Sigma}_{k|k}$ in the GM-IEKF is derived as the covariance of the total influence function, $\boldsymbol{I_F}$ [68, 135, 149], yielding

$$\begin{aligned} \boldsymbol{\Sigma}_{k|k} &= E\left[\boldsymbol{I_F} \cdot \boldsymbol{I_F}^\top \right] \\ &= 1.0369 \left(\tilde{\tilde{\boldsymbol{H}}}_k^\top \tilde{\tilde{\boldsymbol{H}}}_k \right)^{-1} \left(\tilde{\tilde{\boldsymbol{H}}}_k^\top \boldsymbol{Q}_\varpi \tilde{\tilde{\boldsymbol{H}}}_k \right) \left(\tilde{\tilde{\boldsymbol{H}}}_k^\top \tilde{\tilde{\boldsymbol{H}}}_k \right)^{-1}. \end{aligned} \tag{4.25}$$

The expression (4.25) is derived in Appendix B. The flowchart of the robust extended Kalman filter is presented in Fig. 4.5.

4.2 Robust iterated extended Kalman filter

To circumvent first-order approximation errors of the EKF, which might be large under strong nonlinearities of the model, the IEKF [150] and the UKF have been proposed as alternative methods. Specifically, the IEKF linearizes the system nonlinear equations iteratively to

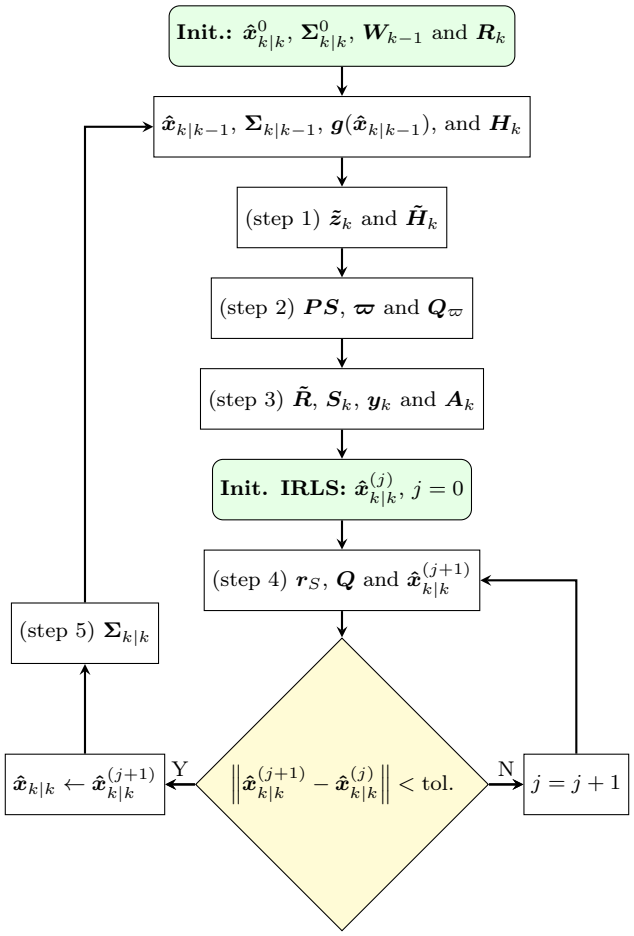


Figure 4.5: Flowchart of the robust extended Kalman filter.

compensate for the higher-order terms, whereas the UKF leverages the unscented transformation by deterministically providing sigma points to approximate the mean and covariance matrix of a random state vector, thus achieving better accuracy than the EKF [151, 152]. An UKF-based DSE using a fourth-order generator model is proposed in [49] to estimate the states of a single-machine infinite-bus power system. Along the same lines, a centralized UKF is developed in [153] for a multi-machine system, while a decentralized strategy that does not require transmission of local signals is advocated in [154], significantly increasing the computational efficiency. However, it has been demonstrated in [139] that the performance of either EKF or UKF is greatly degraded in the presence of observation outliers due to their lack of robustness. To mitigate this issue, a normalized innovation vector-based test is advocated in [154] to detect observation outliers despite the vulnerability of this test to innovation outliers. In [155], Rouhani and Abur developed a distributed two-stage robust UKF-based DSE using the LAV estimator that can handle observation outliers in PMU measurements. However, the authors do not address the vulnerability of the dynamic state estimator to innovation outliers that are induced by approximations in the state prediction

model or by impulsive system process noise.

Gandhi and Mili [68] proposed a robust Kalman filter for linear dynamical model. In this paper, we have extended that work to general nonlinear dynamic state estimation problems with several new features. The latter include the derivations of a new batch mode regression form to enhance data redundancy, a new outlier detection method based on projection statistics and applied to two time-sequence of the prediction and the innovation vectors, and the robust state covariance matrix of the proposed robust generalized maximum likelihood iterated EKF (GM-IEKF) method. These features allow our GM-IEKF to track power system dynamics more reliably and more rapidly than the conventional EKF, even in presence of observation and innovation outliers or non-Gaussian PMU noise.

In this section, we further justify why the robust IEKF is preferred to the robust EKF. When a non-iterative EKF is considered, the measurement function given by (4.2) is linearized using a first-order Taylor series expansion about the predicted state vector $\hat{\boldsymbol{x}}_{k|k-1}$, yielding

$$\boldsymbol{y}_k = \boldsymbol{h}\left(\hat{\boldsymbol{x}}_{k|k-1}\right) + \boldsymbol{H}_k\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right) + \boldsymbol{v}_k, \tag{4.26}$$

where $\boldsymbol{H}_k = \partial \boldsymbol{h}/\partial \boldsymbol{x}|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_{k|k-1}}$. Then the regression model in (4.4) changes to

$$\begin{bmatrix} \boldsymbol{y}_k - \boldsymbol{h}\left(\hat{\boldsymbol{x}}_{k|k-1}\right) + \boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1} \\ \hat{\boldsymbol{x}}_{k|k-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_k \\ \boldsymbol{I} \end{bmatrix} \boldsymbol{x}_k + \begin{bmatrix} \boldsymbol{v}_k \\ -\boldsymbol{e}_k \end{bmatrix}. \tag{4.27}$$

The p_s of the row vectors of the revised matrix $\boldsymbol{Z}$ are then calculated and a statistical test is applied to identify observation outliers. However, because innovation outliers occur in the predicted state $\hat{\boldsymbol{x}}_{k|k-1}$, their effects will spread out on the observation vector on the left side of (4.27) by matrix $\boldsymbol{H}_k$, making the p_s to break down. In contrast, only the innovation $\left[\boldsymbol{y}_k - \boldsymbol{h}\left(\hat{\boldsymbol{x}}_{k|k-1}\right) + \boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1}\right]$ corresponding to the outliers will be affected in (4.27). Thus, the p_s will not break down and both observation and innovation outliers are suppressed.

Comment 1: In most related literature, the state estimation error covariance matrix of an M-estimator is updated using the same equations as for the standard least-squares-based EKF, e.g., [156, 157]. Obviously, this will lead to a degraded performance of the estimator, including poor tracking capabilities and lower statistical efficiency. The correct asymptotic covariance matrix of that estimator should be utilized, which might be derived from (4.25).

Comment 2: In the proposed GM-IEKF, the matrix $\boldsymbol{Q}_q$ and the weights $\boldsymbol{\varpi}$ are used to bound the influence of residual and of position, guaranteeing a robust state estimation, while the matrix $\boldsymbol{Q}_{\varpi}$ contributes to a robust estimation of the state estimation error covariance matrix and to a robust prewhitening in the next time step.

The flowchart of the GM-IEKF is shown in Fig. 4.6.

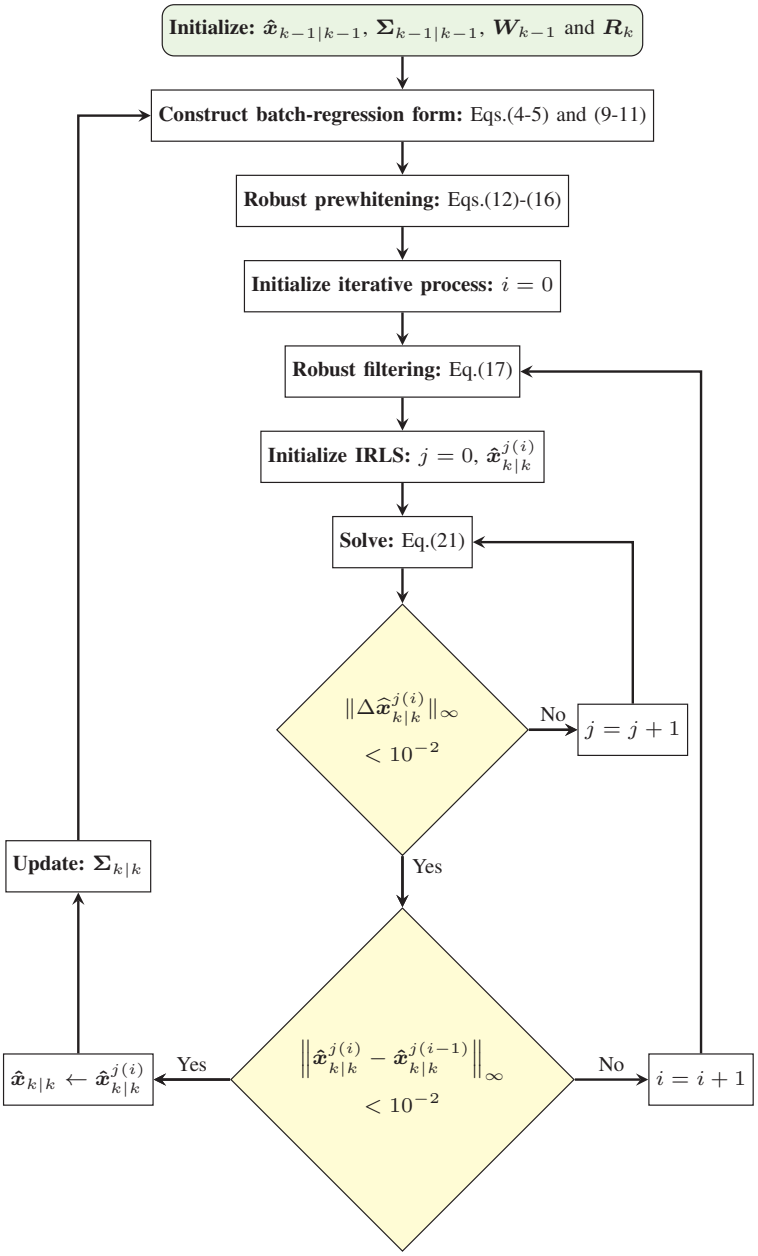


Figure 4.6: Flowchart of the robust *iterated* extended Kalman filter.

4.3 Numerical results

The performance of the proposed estimator is tested on the IEEE 39-bus test system, whose data can be found in [158]. The standard EKF and UKF are implemented for comparisons. In the simulations, the real and reactive power injections, and bus voltage phasors are assumed to be metered using PMUs; a random Gaussian noise with zero mean and standard deviation 10^{-2} is assumed for either system or measurement noise; the diagonal elements of the initial error covariance matrix of the UKF are set to 10^{-4}; the initial values of the state vector are arbitrarily chosen for all three estimators. The damping ratio D is set equal to 0.005 for all generators; the PMU measurements are assumed to be received at a rate of 48 samples per second; at t=0.5s Line 15-16 is switched off, which causes a large disturbance to the system; the time domain simulation results are assumed to be the true state values; the parameters for GM-IEKF are set as β=d=1.5, and the maximal number of iterations is 20.

4.3.1 Robust extended Kalman filter

Case 1: Comparison between EKF, GM-EKF and UKF under ideal conditions

In the simulations, we assume that $\boldsymbol{W}_{k-1} = \boldsymbol{R}_k = \text{diag}\,(10^{-4})$, which does not represent a stringent condition. The presence of outliers is not considered at this point. For comparison purpose, we also simulate the unscented Kalman filter (UKF), although no theoretical background has been previously presented; the reader is referred to [49, 159, 160] for details. The rotor speed versus time is displayed in Fig. 4.7. We pick the plot of Generator 4 as an example and omit the remaining plots since they do not bring any additional qualitative information. Clearly, it can be seen from Fig. 4.7 that all the three methods are able to track well system dynamics.

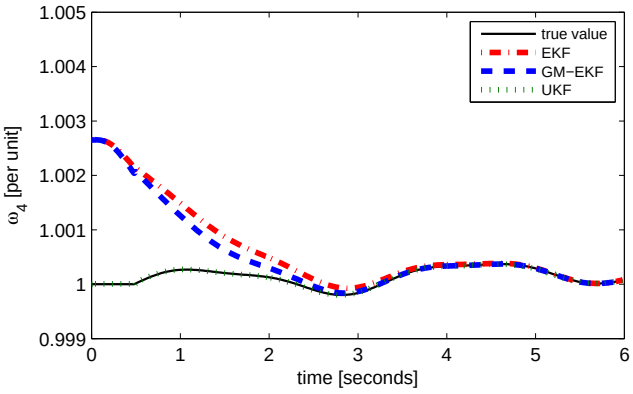


Figure 4.7: Rotor speed of Generator 4 after the applied disturbance at $t = 0.5s$.

Case 2: Momentary loss of communication link

The communication link with the PMU placed at Bus 34 where Generator 5 is connected is supposed to be lost from 4 to 6s. Therefore, the measurement set $\{P_5, Q_5, V_5, \theta_5\}$ becomes unavailable during this time frame; here, their values are set equal to zero for simulation purpose. Although such an event has a low probability of occurrence, it is of interest to investigate its effects on the estimation results. Fig. 4.8 and 4.9 illustrate the estimation of the state variables of Generator 5. As we can observe, the non-robust methods, namely the EKF and the UKF, are not capable of handling such condition. Note that although we are not showing the remaining plots, the estimation related to other generators are also strongly biased by the outliers. By contrast, the proposed GM-EKF downweights these outliers to the point to nearly suppress their effect on the state estimates.

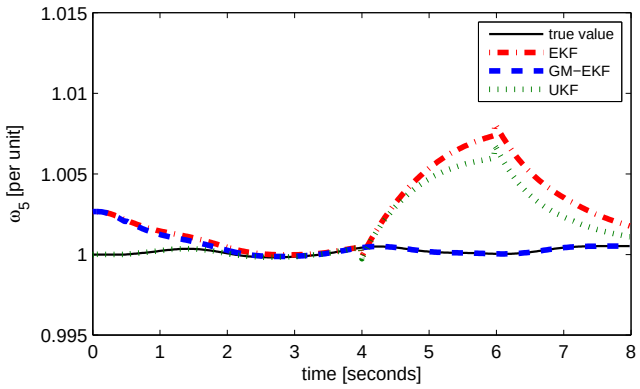


Figure 4.8: Rotor speed of Generator 5 with loss of PMU #5 from 4 to 6s.

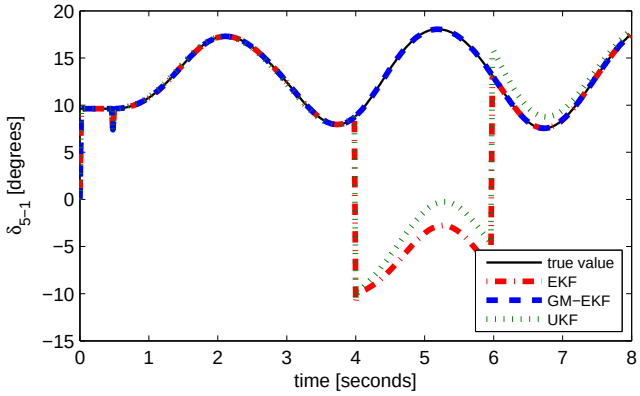


Figure 4.9: Rotor angle of Generator 5 with loss of PMU #5 from 4 to 6s.

Case 3: Measurements with bad data

In this case, we set $Q_7 = 10$ at $t = 4s$, which is a gross measurement error. Note that the reactive power injection of all the generators is under 2.2 per-unit. The simulation results are displayed in Figs. 4.10 and 4.11. We observe that the estimation results of the EKF and

the UKF are strongly affected by the magnitude of the outlier. As for the GM-EKF, the outlier is strongly downweighted, resulting in a good state estimate. The overall estimation error for the three simulated cases is presented in Table 4.2. For comparison, we present in Table 4.3 the computing times for all the simulated cases although the Matlab code has not been optimized. Here, we have run the algorithm 200 times for each case and calculated the average times.

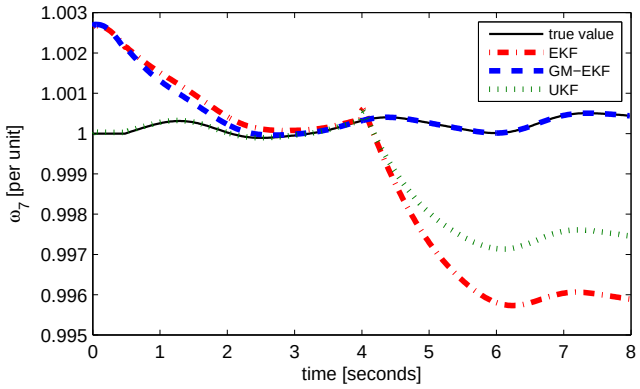


Figure 4.10: Rotor speed of Generator 7 with an outlier placed on Q_7 after $t = 4s$.

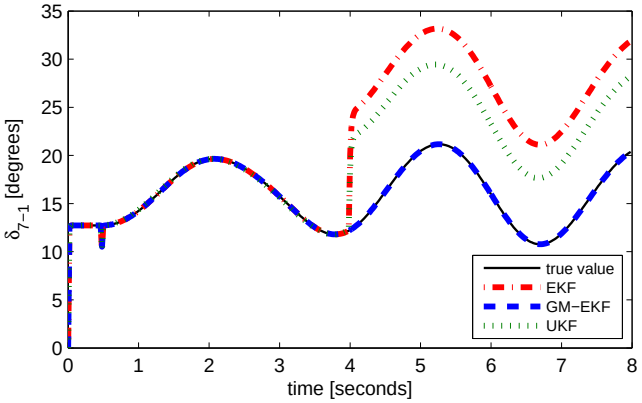


Figure 4.11: Rotor angle of Generator 7 with an outlier placed on Q_7 after $t = 4s$.

Table 4.2: Overall estimation errors.

Method	EKF	GM-EKF	UKF
Case 1	0.0169	0.0141	0.0009
Case 2	0.0235	0.0106	0.0101
Case 2	0.0275	0.0106	0.0098

Table 4.3: Computational time.

Method	EKF	GM-EKF	UKF
Case 1	$1.3482s$	$2.0398s$	$0.9266s$
Case 2	$1.7859s$	$2.6811s$	$1.2986s$
Case 2	$1.6956s$	$2.6662s$	$1.2328s$

4.3.2 Robust iterated extended Kalman filter

Case 1: Small process and measurement noise

We initialize each process noise with a very small variance of magnitude 10^{-4}. The error variance of the PMU measurements is set to 1%. Fig. 4.12 shows the tracking results of the three dynamic state estimators. We randomly pick Generator 5 as an example and omit the remaining plots since they do not bring any additional qualitative information. It can be seen from the figure that all three methods accurately track the system dynamic states. The GM-IEKF outperforms standard EKF in estimating both the rotor angle and the rotor speed. The UKF is the fastest method for tracking these state variables because it uses the *sigma points* to propagate the mean and the covariance matrix without relying on direct linearization, which improves the computational efficiency.

Case 2: Small measurement noise and large process noise

In order to investigate the impacts of system process noise on the tracking performance, the process noise is increased from 10^{-4} to 10^{-2} while the measurement variance is kept the same as that in Case 1. Fig. 4.13 presents the tracking results. It can be seen from this figure that UKF is greatly affected by system process noise regarding the rotor speed estimation, while GM-IEKF and EKF are slightly affected. In this case, GM-IEKF is able to bound the influence of the process noise, resulting in the best tracking performance among the three methods.

Case 3: Momentary loss of communication link

Some PMU devices might temporarily lose their communication links to the PDC due to device failure, cyber attacks, communication interruptions, to cite a few. In that case, the communication link with the PMU placed at Bus 34, where Generator 5 is connected, is assumed to be lost from t=4s to t=5s. Therefore, the measurement set $\{P_5, Q_5, V_5, \theta_5\}$ becomes unavailable during this time interval and their values are set equal to zero for simulation purpose. It is observed that UKF and EKF are not capable of tracking the trajectories of ω_5 and δ_{5-1}. By contrast, GM-IEKF exhibits good tracking capabilities by relying on the predicted state estimates.

Case 4: Occurrence of observation outliers

To investigate the effect of observation outliers on the EKF, UKF, and GM-IEKF, P_5 on

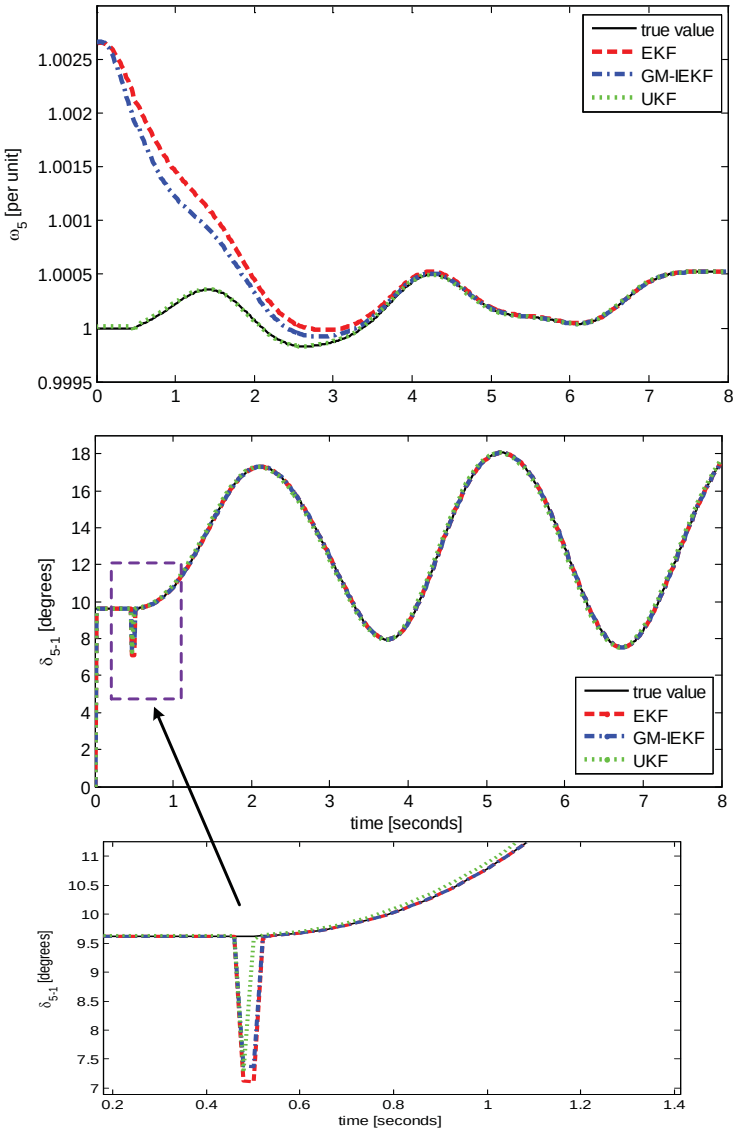


Figure 4.12: Performance comparisons of the three methods for estimating ω_5 and δ_{5-1} in Case 1: small process and measurement noise.

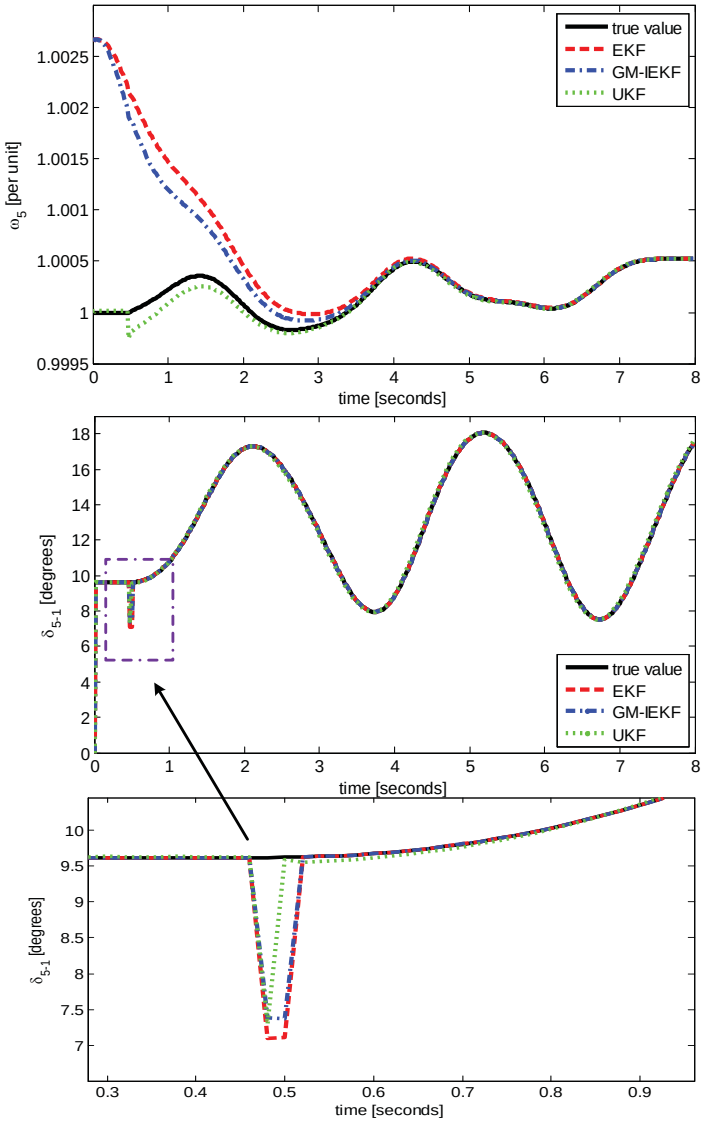


Figure 4.13: Performance comparisons of the three methods for estimating ω_5 and δ_{5-1} in Case 2, where process noise is increased from 10^{-4} to 10^{-2}.

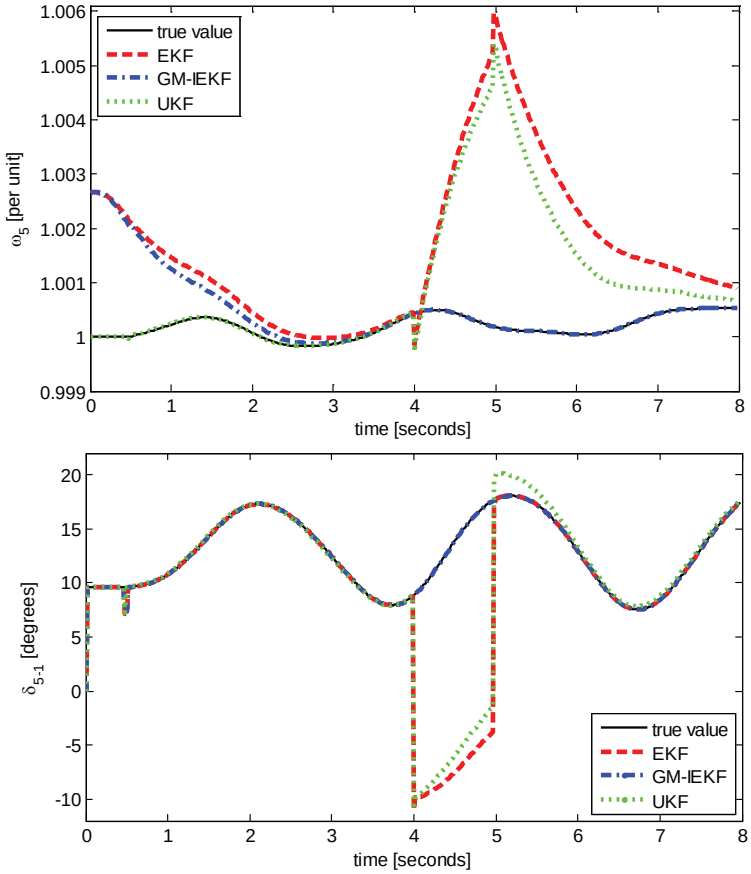


Figure 4.14: Performance comparisons of the three methods for estimating ω_5 and δ_{5-1} in Case 3, where the communication link with the PMU placed at the Generator 5 is lost from t=4s to t=5s.

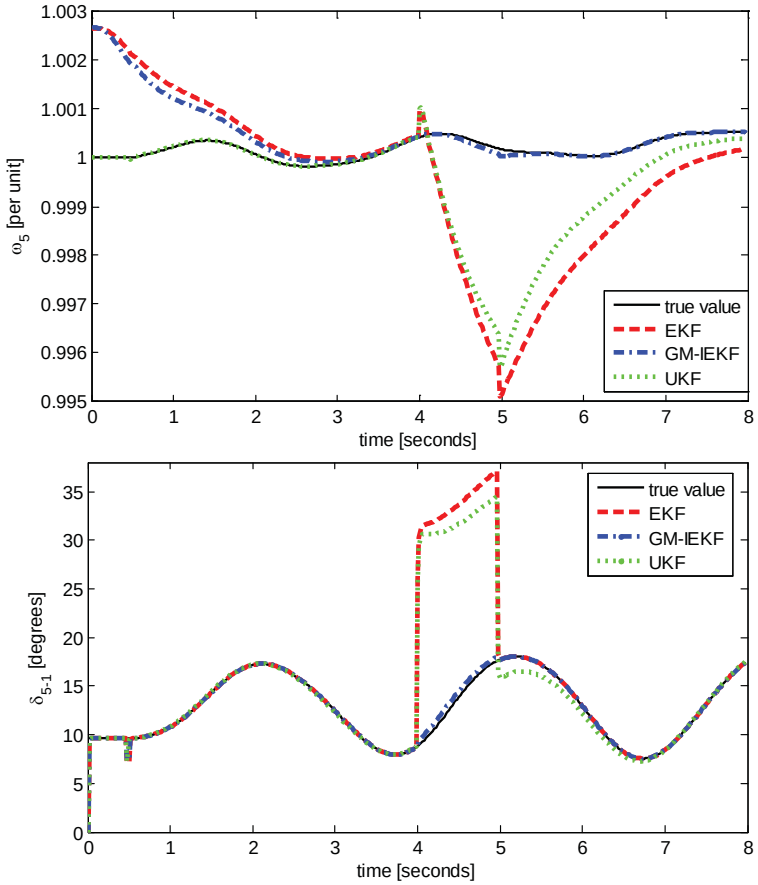


Figure 4.15: Performance comparisons of the three state estimation methods in Case 4, where one measurement coming from Bus 34 (P_5) is contaminated with gross errors from t=4s to t=5s.

Bus 34 is contaminated with gross errors from t=10s to t=14s by changing its value to 8 pu, simulating a faulty synchronization or an impulsive communication noise. Note that by applying the p_s to the Jacobian matrix $\overset{\approx}{\boldsymbol{H}}$, this outlier is flagged as bad leverage point. The test results are shown in Fig. 4.15. We note that when observation outlier occurs, both EKF and UKF deviate far away from the true system states since they are not robust to any outliers. However, GM-IEKF significantly reduces its influence, which is reflected by a very small weight $\varpi_i \approx 10^{-4}$, keeping good tracking performance.

Case 5: Occurrence of innovation outliers

As discussed before, the predictions might be unreliable due to the imperfect dynamical model or impulsive system process noise. To investigate the performances of the estimators under this condition, the predicted value of ω_5 is changed to 8 pu between t=4s and t=6s. Fig. 4.16 shows the simulation results. We observe that the UKF is greatly affected by innovation outliers, which makes its tracking trajectories unreliable. The EKF is less sensitive to the

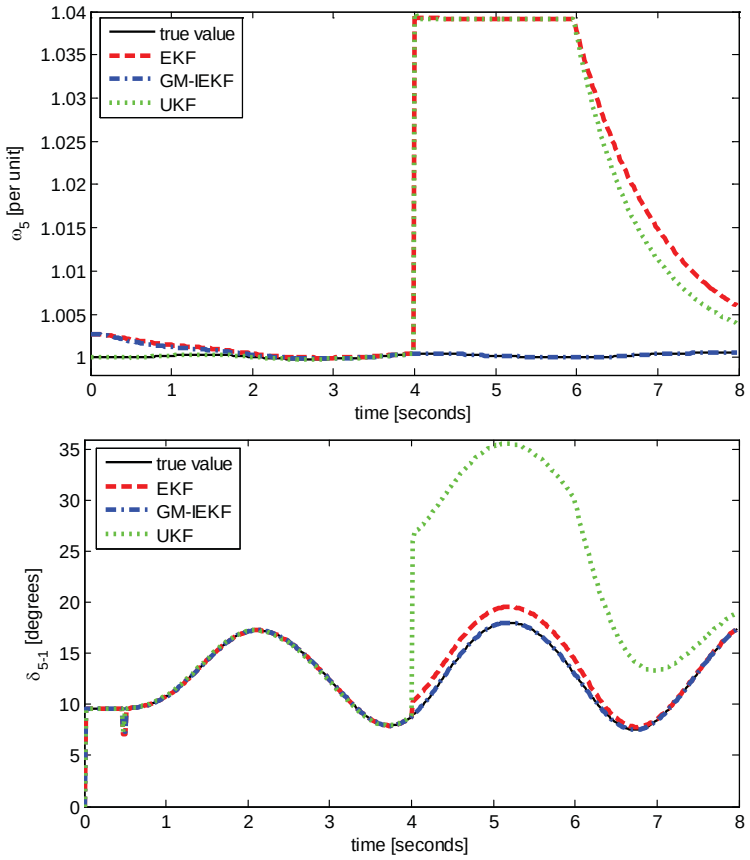


Figure 4.16: Performance comparisons of the three methods in Case 5, where the predicted value of ω_5 is changed from 0.2072 to 2 pu from t=4s to t=6s.

rotor angle estimation compared with the UKF, but its estimation for rotor speed is not acceptable. By contrast, the GM-IEKF effectively suppresses this innovation outlier, whose weight $\varpi_i \approx 2 \times 10^{-4}$, resulting in good states tracking.

Robustness to non-Gaussian noise

To demonstrate the robustness of the proposed method under thick-tailed distributions, non-Gaussian PMU noise is simulated using the models proposed in a recent PNNL report [143, 161]. Specifically, the bimodal Gaussian mixture model with zero mean, a variance of 10^{-4} and weights of 0.9 and 0.1 is assumed for the voltage magnitude measurement noise, whereas the Laplacian distribution with zero mean and a scale of 1 is assumed for the real and reactive power measurement noise. As an illustrative example, results for Generator 10 are displayed in Fig. 4.17. We observe that both EKF and UKF exhibit large oscillations, indicating inherent instabilities while our robust GM-IEKF yields a reasonable performance.

Now, let us describe and analyze the test results obtained for all the three methods in presence of observation or innovation outliers using the two-axis machine model. In this test,

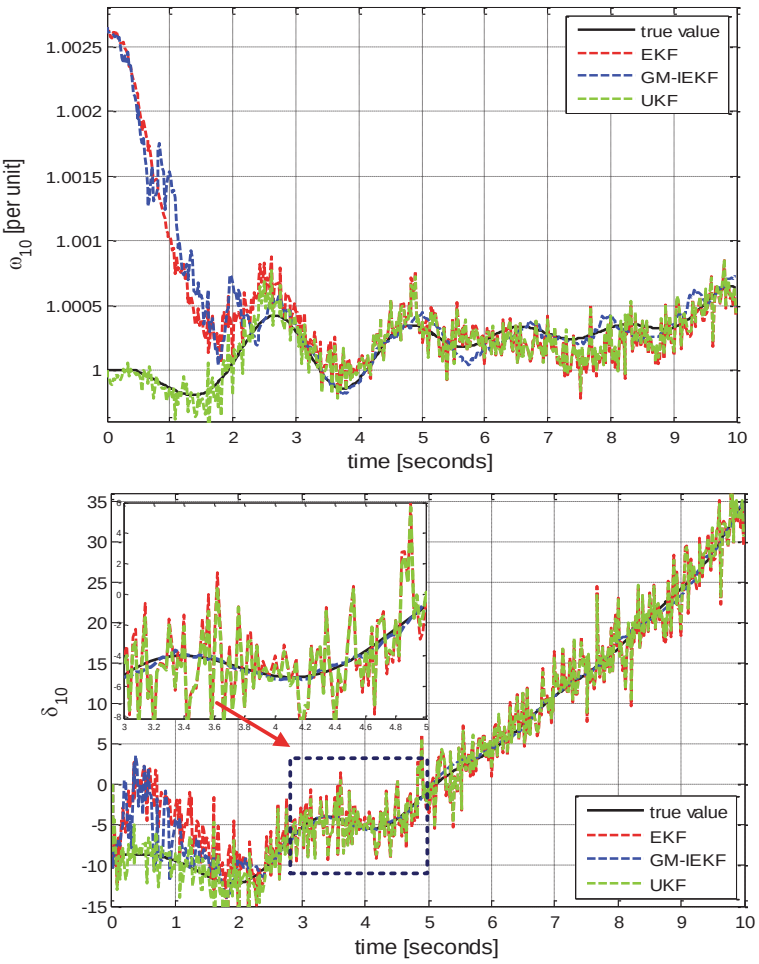


Figure 4.17: Performance comparisons of the three state estimation methods under non-Gaussian measurement noise.

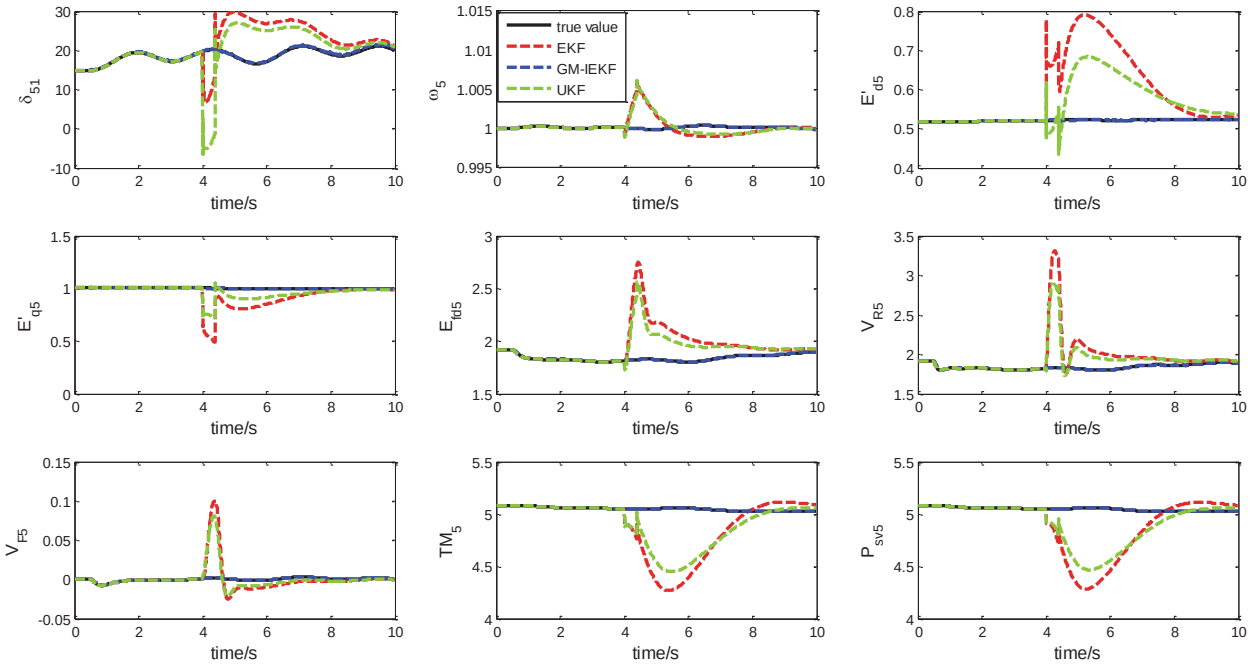


Figure 4.18: Performance comparisons of the three dynamic state estimation methods in presence of two observation outliers induced as follows: two measurement taken on Bus 34 (P_5 and Q_5) are contaminated with gross errors from t=4s to t=5s.

the measurement settings and fault condition are the same as those given in the previous simulations. The time step for time domain simulation is 0.008s. The estimated states of Generator 5 are provided as examples. All the parameter values of the generators are taken from [162]. Figs. 4.18 and 4.19 present the results of the three methods with two observation outliers, i.e., P_5 and Q_5 changed to 0 pu, and one innovation outlier, i.e., δ_6 changed to 5 pu. From these two figures, we observe that even with high-order generator model, the proposed GM-IEKF is able to effectively suppress both observation and innovation outliers. By contrast, EKF and UKF exhibit no robustness against any types of outliers, yielding unreliable estimation results.

4.4 Concluding remarks

4.4.1 Breakdown point

We demonstrate in the previous examples that the GM-IEKF can cope with observation and innovation outliers. The remaining question is how many outliers it can handle without giving unreliable estimation results, that is, what is its breakdown point? This concept provides a measure of the global robustness of an estimator, for the maximum (finite) possible bias

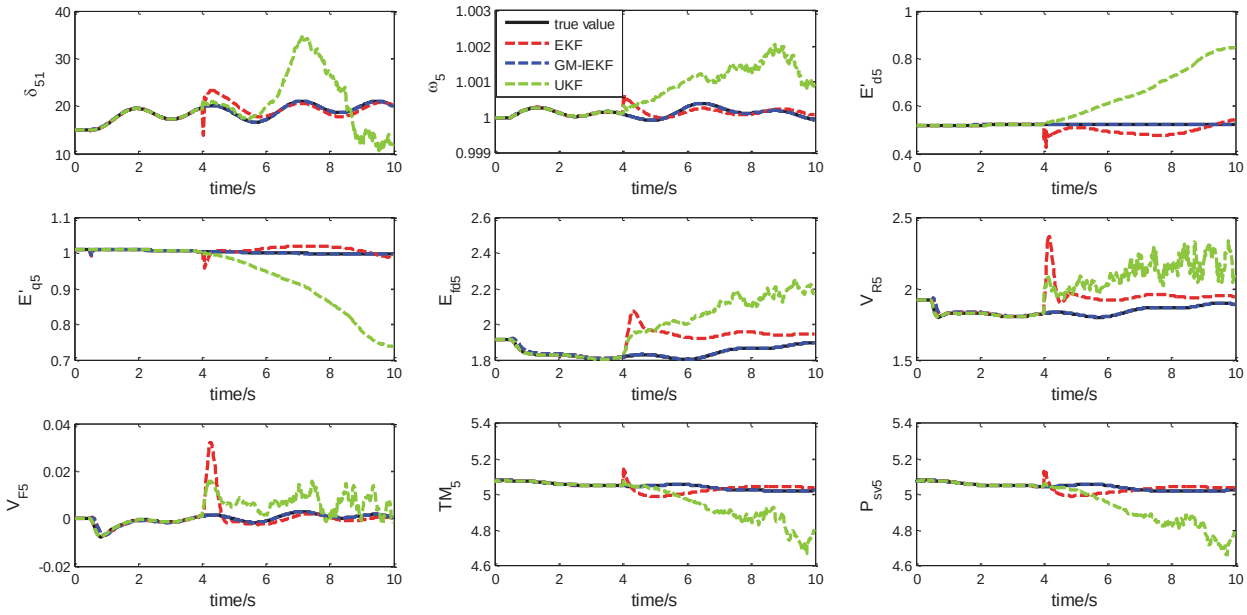


Figure 4.19: Performance comparisons of the three dynamic state estimation methods in presence of an innovation outlier induced as follows: the predicted rotor angle of Generator 5 is contaminated with gross errors from t=4s to t=5s.

induced by a given fraction of outliers. Formally, we have:

$$\varepsilon^* = \max\left(\varepsilon = \frac{N_f}{N_m}; \text{maximum bias}\right), \tag{4.28}$$

where N_f is the number of outliers; N_m is the total number of data points, which is the dimension of $\tilde{\tilde{\boldsymbol{y}}}_k$ in this paper. A definition of the finite-sample breakdown point of a robust estimator in nonlinear regression is given by Stromberg and Ruppert [163]. The expression of the maximum breakdown that any regression equivariant estimator might have in linear structured regression (e.g., in linearized power system state estimation model, which involves sparse Jacobian matrices) is derived by Mili and Coakley [164].

We carry out extensive simulations to determine the breakdown point of our GM-IEKF when applied to the IEEE 39-bus test system. By replacing a varying number of data points by outliers in the vector $\tilde{\tilde{\boldsymbol{y}}}_k$ of dimension $m_t = m + n$, it is observed that the GM-IEKF can handle at least 25% of outliers among the data set, be they innovation or observation outliers, the worst case being clustered ones. Note that the exact value of the breakdown point of the GM-IEKF in power systems still needs to be determined; it will be investigated in a future work.

4.4.2 Computational efficiency

To check whether the proposed GM-IEKF can keep up with PMUs at their scanned rate of 30 or 60 samples per second, its computational efficiency is analyzed and compared to that of the EKF and UKF under three different scenarios defined as follows: Scenario 1: no outliers occur; Scenario 2: with observation outliers; Scenario 3: with innovation outliers. Please note that all the three scenarios are tested with the two-axis model instead of the classical model of the synchronous generators. The testing environment consists of a PC configured with Intel(R) Xeon(R), CPU E5-1650, 3.5GHz and 32GB RAM. The average computing times of all the three methods in the three scenarios for every time sample are displayed in Table 4.4. From this table, we observe that EKF and UKF have similar computing times that are much lower than the PMU sampling periods of 33.3ms and 16.7ms for 30 sample/s and 60 samples/s, respectively. As for GM-IEKF, although its execution time is larger than that of EKF, it is also lower than the PMU sampling period, enabling it to track system state dynamics in real-time. Furthermore, unlike EKF and GM-IEKF, UKF might exhibit numerical instability in presence of innovation outliers.

Table 4.4: Average computing time of the three estimation methods in three scenarios for each time sample

Cases	EKF	UKF	GM-IEKF
Scenario 1	3.94ms	3.48ms	5.96ms
Scenario 2	3.95ms	3.49ms	5.99ms
Scenario 3	3.93ms	3.51ms or diverged	6.30ms

4.4.3 Centralized versus decentralized approach

The proposed GM-IEKF is a centralized DSE, but it can be extended to a decentralized framework, typically employed for the real-time monitoring and control of very large-scale power systems. For instance, we might adopt the same approach as the one proposed in [154] by treating the generator terminal voltage and current phasors as system input and output variables, which will be both assumed to be metered by PMUs. By doing so, the dynamic equations for one generation unit can be decoupled from those of the other units. Therefore, GM-IEKF can be applied to estimate the system dynamic state by processing only local PMU measurements. However, with the decentralized GM-IEKF, the measurements processed by a DSE are only taken on machine terminal buses, yielding a very low local measurement redundancy. If the terminal measurements are lost due to cyber attacks or communication failures, or include either observation or innovation outliers, the decentralized GM-IEKF might lose its tracking ability. On the other hand, with a centralized GM-IEKF, all available measurements are transferred to the control center with some communication bandwidth utilization, resulting in an enhanced measurement redundancy. In this case, outliers and the loss of measurements can be effectively handled by the centralized GM-IEKF, thereby

providing the control center with a complete set of state estimates, which enables it to take global and local control actions.

Let us now compare the centralized and decentralized scheme from the viewpoint of the communication bandwidth requirement. Obviously, with a decentralized GM-IEKF, only local controls are implemented. If coordinated control is deployed between different local DSEs, additional communication bandwidth is required, making the comparison with the communication cost of the centralized GM-IEKF difficult to assess. In summary, choosing between the two schemes depends on the applications and the communication infrastructures being used.

In this paper, a robust GM-IEKF method is proposed for tracking the dynamic states of a power system. The estimation problem is formulated as a nonlinear regression problem by processing both state predictions and observations in a batch mode. Furthermore, projection statistics and a GM-estimator are applied to PMU metered values for bounding the influence of observation and innovation outliers, including the loss of measurement communication links. Simulations carried out on the IEEE 39-bus test system provide results that demonstrate the robustness and the statistical efficiency of GM-IEKF.

Despite evident advantages of our GM-IEKF, which include fast computations and good robustness to observation and innovation outliers, it suffers from several weaknesses that call for further research and developments. One weakness is its vulnerability to system parameter and topology errors, which require a very high level of local measurement redundancy [165]. Obviously, this requirement is difficult to achieve in practice, mainly due to economic constraints. The second weakness is that the iterative EKF might produce unreliable state estimates under strong nonlinearities of the power system model. Dynamic estimators that circumvent this difficulty include the unscented transformation based filters, the ensemble Kalman filter, and the particle filter, among others. These estimators make use of sigma/ensemble points or particles to approximate the true distribution of random state variables, resulting in good state predictions and filtering when the assumptions are satisfied. As a future work, we plan to develop robust versions of these filters for handling outliers under strong system nonlinearities. Finally, the proposed GM-IEKF will be extended to a decentralized framework for the real-time monitoring and control of a very large-scale power system.

Chapter 5

Robust hierarchical decentralized dynamic state estimator

In the previous chapter, we presented a robust centralized *model-based* dynamic state estimator. In this chapter, we will first introduce the *data-driven* Koopman Kalman filter [71–73], followed by its robust version [69, 70]. The later, termed robust Koopman Kalman filter, is the core component of the developed hybrid framework that combines model-based and data-driven methods for hierarchical decentralized robust dynamic state estimation [79]. A two-level hierarchy is proposed, where the lower level consists of robust, model-based, decentralized DSEs. The state estimates sent from the lower level are received at the upper level, where they are filtered by a robust data-driven DSE after a principled sparse selection. This selection allows us to shrink the dimension of the problem at the upper level and hence significantly speed up the computational time. The proposed hybrid framework does not depend on the centralized infrastructure of the control centers; thus it can be completely embedded into the wide-area measurement systems. This feature will ultimately facilitate the placement of hierarchical decentralized control schemes at the phasor data concentrator locations. Also, the network model is not necessary; thus, a topology processor is not required. Finally, there is no assumption on the dynamics of the electric loads. The proposed framework is tested on the 2,000-bus synthetic Texas system, and shown to be capable of reconstructing the dynamic states of the generators with high accuracy, and of forecasting in the advent of missing data.

Researchers working on this field have been tackling the DSE design by resorting to either nonlinear observers or the KF and its nonlinear extensions. Typically, to make the design of a nonlinear observer tractable, the power systems are supposed to be lossless and the voltage magnitude at each bus is fixed to 1 p.u., which are rather restrictive assumptions [52,166,167]. It turns out that observers are typically limited to low-dimensional systems. Recently, the extended Kalman filter (EKF) [48, 50, 51] and the unscented Kalman filter (UKF) [49] have received a great deal of attention from researchers and practitioners. However, as we argued

in chapter 4, the model and parameter uncertainties and the computational burden preclude their adoption for large-scale power systems. For example, the level of uncertainty associated with composite load models is high, and it is expected to increase in the near future. Decentralized versions of the EKF [56] and of the UKF [55] that relieve the computational burden and lessen the communication requirements have also been proposed, but the model and parameter issues remain unsolved. In fact, the increasing uncertainties caused by the penetration of renewable energy sources in the grid pose a major challenge to power systems state estimation in general, be it *static* [168] or dynamic. Furthermore, both observers and KFs presume the availability of a dynamical model of the Kron-reduced network, a requirement that is not easily met by many power utilities. As an attempt to overcome these weaknesses, a model independent technique seems worth investigating.

We develop a two-level, hierarchical, decentralized, robust DSE by combining model-based and data-driven methods. The model-based decentralized DSEs at the lower level provide the necessary data for performing data-driven model identification at the upper level by using the KMD [93], which further allows for dynamic stability assessment and modal analysis of nonlinear dynamical systems [80]. The paper also illustrates the application of the proposed framework for large-scale systems by using compressed sensing [169] to find sparse state estimate selection. Because the use of all state estimates at the upper level could be prohibitive for high-dimensional systems, we rely on compressed sensing to find a sparse selection of state estimates, following the work of Brunton et al. [124, 170]. The proposed hybrid framework provides the opportunity to devise powerful tools by combining concepts from dynamical systems, estimation, and control theory. Firstly, the use of decentralized DSEs and KMD makes it independent of the network model; hence, it does not require a topology processor. Secondly, by virtue of the data-driven KMD, the method does not need to make assumptions about the underlying load model dynamics. Thirdly, and most notably, it can be completely embedded into the wide-area measurement systems instead of being an add-on to the energy management systems installed at the control centers. This attribute will ultimately facilitate the hierarchical control design of electric power systems [25] with the placement of control schemes at the PDC location, thereby exploiting the synergies between agile, low-latency, decentralized and holistic centralized monitoring and control architectures.

5.1 Robust data-driven Koopman Kalman filter

Surana and Banaszuk [71] and Surana [72, 73], proposed a Koopman operator-based KF (KKF). The main attractiveness of their approach is its data-driven feature and the possibility of tracking nonlinear dynamics through the conventional linear KF. The concept of a *model-free*, data-driven KF has appeared before in the literature; see for instance [171]. The key idea here is to discover both the transition and the observation matrices from recorded measurements. In fact, this notion of extracting models from data is an important problem in science and engineering [172] in general. To demonstrate the KKF, Surana

and Banaszuk [71] studied a canonical nonlinear dynamical system, which possess a finite-dimensional Koopman-invariant subspace [127] and has *a priori* known eigenvalues and eigenfunctions. It turns out that such a finite-dimensional subspace does not exist for an electric power system. This is due to the presence of sinusoidal nonlinearities that endow the system with multiple attractors [129], implying that from a theoretical standpoint, nonlinear dynamics of power systems are exactly represented by infinitely many Koopman modes. However, in practice, a finite-dimensional approximation is used instead. In Surana [73], a more in-depth analysis is performed that includes numerical results carried out on power systems under the assumption that the rotor angle of the synchronous generators is measured. Although it is not the industry practice, a project under development at Virginia Tech [173, 174] is demonstrating that synchronized measurements of the rotor angle is currently a feasible, cost-effective alternative.

A caveat of applying the KKF [71, 73] in power systems is that PMUs, and eventually synchronized measurements of the rotor angle, are not immune to failures; outliers, data dropouts, and impulsive noises occur ever so often. Thus, it is imperative to have a mechanism that is able to suppress their effect on the estimation [50, 51]. We therefore propose to develop a robust generalized maximum-likelihood KKF (GM-KKF) as a model-independent alternative approach for DSE. Apart from the obvious advantage of being robust to outliers, the GM-KKF presents high statistical efficiency when the system and observation errors in the Kalman filter deviate from Gaussianity. This is an important feature, which shall become clear after reading Section 5.1. Furthermore, the GM-KKF exhibits faster convergence rate as compared to the KKF thanks to a batch-mode regression formulation. In what follows, we compare the GM-KKF with the KKF proposed in [71, 73] to demonstrate the necessity of adopting a robust approach. To further evaluate its performance, we also compare the GM-KKF with the robust GM-EKF developed in [50]. Several aspects such as the rate of convergence and the computational speed are assessed.

Consider a discrete time-invariant nonlinear model defined as follows:

$$\boldsymbol{x}_k = \boldsymbol{f}\left(\boldsymbol{x}_{k-1}\right), \tag{5.1}$$

$$\boldsymbol{y}_k = \boldsymbol{h}\left(\boldsymbol{x}_k\right), \tag{5.2}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the system state vector, $\boldsymbol{y} \in \mathbb{R}^m$ is the system observation vector, and $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ and $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ are vector-valued nonlinear functions. Note that the *observables* mentioned in the Introduction section are not necessarily equal to the system observation vector, $\boldsymbol{y}$, defined by (5.2).

Our problem is to estimate the time evolution of the state vector, $\boldsymbol{x}$, in real-time for system monitoring, identification, and control. This problem can be tackled by using a nonlinear observer. There are many techniques to design a nonlinear observer. The interested reader is pointed to [71] and references therein. One design approach is to seek a transformation

$$\underline{\boldsymbol{x}} = \mathcal{X}\left(\boldsymbol{x}\right), \qquad \underline{\boldsymbol{y}} = \mathcal{Y}\left(\boldsymbol{y}\right), \tag{5.3}$$

such that (5.1) and (5.2) can be converted into a canonical observer form given by

$$\underline{\boldsymbol{x}}_k = \boldsymbol{F}_{k-1}\underline{\boldsymbol{x}}_{k-1} + \alpha\left(\boldsymbol{y}_k\right), \tag{5.4}$$

$$\underline{\boldsymbol{y}}_k = \boldsymbol{H}_k\underline{\boldsymbol{x}}_k + \beta\left(\boldsymbol{y}_k\right), \tag{5.5}$$

where $\boldsymbol{F} \in \mathbb{R}^{q\times q}$ is the state transition matrix, $\boldsymbol{H} \in \mathbb{R}^{p\times q}$ is the observation matrix, and $\alpha\left(\boldsymbol{y}\right)$, $\beta\left(\boldsymbol{y}\right)$ denote the output injection terms. For clarity of notation, we hereafter distinguish between the system's state space variables on their original coordinates, $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{y} \in \mathbb{R}^m$, and their counterpart mapped into canonical coordinates by the transformation (5.3), $\underline{\boldsymbol{x}} \in \mathbb{R}^q$, $\underline{\boldsymbol{y}} \in \mathbb{R}^p$.

Assuming that (5.3) exists, (5.4)–(5.5) can be put in the Kalman filter form. Interestingly, the nonlinear transformation based on the Koopman operator-theoretic framework does not yield to output injection terms, as it will be shown later. Thus, by neglecting the output injection terms in (5.4)–(5.5) and considering additive error terms, and based on the assumption that the transformation (5.3) exists, we cast the system (5.1)–(5.2) into the Kalman filter form (3.1)–(3.2). Now, our problem becomes one of finding the proper mapping (5.3). We will obtain (5.3) by resorting to the Koopman operator-theoretic framework, which we introduce next.

Assumption 8. $\boldsymbol{x}$ and $\boldsymbol{y}$ are spanned by the KEFs, i.e., $\mathcal{F}^q = \text{span}\left\{\phi_i\right\}_{i=1}^q$ is a subset of the KEFs of (5.1)–(5.2) such that $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{F}^q$.

From (2.58), we have that

$$\boldsymbol{x}_k = \sum_{i=1}^{q} \phi_i\left(\boldsymbol{x}_k\right)\boldsymbol{\upsilon}_i^{(\boldsymbol{x})}, \quad \boldsymbol{x}_{k+1} = \sum_{i=1}^{q} \phi_i\left(\boldsymbol{x}_k\right)\boldsymbol{\upsilon}_i^{(\boldsymbol{x})}\lambda_i, \tag{5.6}$$

$$\boldsymbol{y}_k = \sum_{i=1}^{q} \phi_i\left(\boldsymbol{x}_k\right)\boldsymbol{\upsilon}_i^{(\boldsymbol{y})}, \quad \boldsymbol{y}_{k+1} = \sum_{i=1}^{q} \phi_i\left(\boldsymbol{x}_k\right)\boldsymbol{\upsilon}_i^{(\boldsymbol{y})}\lambda_i. \tag{5.7}$$

where $\boldsymbol{\upsilon}_i^{(\boldsymbol{x})} \in \mathbb{C}^n$, and $\boldsymbol{\upsilon}_i^{(\boldsymbol{y})} \in \mathbb{C}^m$, $i = 1, ..., q$. In matrix form, (5.6) can be written as

$$\begin{bmatrix} x_{1,k} \\ \vdots \\ x_{n,k} \end{bmatrix} = \begin{bmatrix} \upsilon_{1,1}^{(\boldsymbol{x})} & \dots & \upsilon_{1,q}^{(\boldsymbol{x})} \\ \vdots & & \\ \upsilon_{n,1}^{(\boldsymbol{x})} & \dots & \upsilon_{n,q}^{(\boldsymbol{x})} \end{bmatrix} \begin{bmatrix} \phi_1\left(\boldsymbol{x}_k\right) \\ \vdots \\ \phi_q\left(\boldsymbol{x}_k\right) \end{bmatrix}, \tag{5.8}$$

$$\begin{bmatrix} x_{1,k+1} \\ \vdots \\ x_{n,k+1} \end{bmatrix} = \begin{bmatrix} \upsilon_{1,1}^{(\boldsymbol{x})} & \dots & \upsilon_{1,q}^{(\boldsymbol{x})} \\ \vdots & & \\ \upsilon_{n,1}^{(\boldsymbol{x})} & \dots & \upsilon_{n,q}^{(\boldsymbol{x})} \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_q \end{bmatrix} \begin{bmatrix} \phi_1\left(\boldsymbol{x}_k\right) \\ \vdots \\ \phi_q\left(\boldsymbol{x}_k\right) \end{bmatrix}. \tag{5.9}$$

Similarly, (5.7) can be written as

$$\begin{bmatrix} y_{1,k} \\ \vdots \\ y_{m,k} \end{bmatrix} = \begin{bmatrix} \upsilon_{1,1}^{(\boldsymbol{y})} & \dots & \upsilon_{1,q}^{(\boldsymbol{y})} \\ \vdots & & \\ \upsilon_{m,1}^{(\boldsymbol{y})} & \dots & \upsilon_{m,q}^{(\boldsymbol{y})} \end{bmatrix} \begin{bmatrix} \phi_1(\boldsymbol{x}_k) \\ \vdots \\ \phi_q(\boldsymbol{x}_k) \end{bmatrix}, \tag{5.10}$$

$$\begin{bmatrix} y_{1,k+1} \\ \vdots \\ y_{m,k+1} \end{bmatrix} = \begin{bmatrix} \upsilon_{1,1}^{(\boldsymbol{y})} & \dots & \upsilon_{1,q}^{(\boldsymbol{y})} \\ \vdots & & \\ \upsilon_{m,1}^{(\boldsymbol{y})} & \dots & \upsilon_{m,q}^{(\boldsymbol{y})} \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_q \end{bmatrix} \begin{bmatrix} \phi_1(\boldsymbol{x}_k) \\ \vdots \\ \phi_q(\boldsymbol{x}_k) \end{bmatrix}. \tag{5.11}$$

In what follows, we suppose that the Koopman tuples are ordered such that complex conjugate pairs appear adjacent to each other. Now, consider the first row of (5.9),

$$x_{1,k+1} = \upsilon_{1,1}^{(\boldsymbol{x})}\lambda_1\phi_1(\boldsymbol{x}_k) + \upsilon_{1,2}^{(\boldsymbol{x})}\lambda_2\phi_2(\boldsymbol{x}_k) + \upsilon_{1,3}^{(\boldsymbol{x})}\lambda_3\phi_3(\boldsymbol{x}_k) + ... + \upsilon_{1,q}^{(\boldsymbol{x})}\lambda_q\phi_q(\boldsymbol{x}_k). \tag{5.12}$$

Define $\upsilon_{l,i}^{(\boldsymbol{x})} = c_{l,i} + jd_{l,i}$, $\lambda_i = \sigma_i + j\omega_i$, and $\phi_i(\boldsymbol{x}_k) = u_i + jv_i$. Suppose that $\phi_1(\boldsymbol{x}_k)$ and $\phi_2(\boldsymbol{x}_k)$ is a complex conjugate pair of KEFs, whereas $\phi_3(\boldsymbol{x}_k)$ is real-valued, and neglect the Koopman tuples with index $i = 4, ..., q$. Thus, we have

$$x_{1,k+1} = (c_{1,1} + jd_{1,1})(\sigma_1 + j\omega_1)(u_1 + jv_1) + (c_{1,1} - jd_{1,1})(\sigma_1 - j\omega_1)(u_1 - jv_1) + c_{1,3}\sigma_3u_3. \tag{5.13}$$

After some algebraic manipulation, (5.13) becomes

$$x_{1,k+1} = (c_{1,1}\sigma_1 - d_{1,1}\omega_1)(2u_1) + (c_{1,1}\omega 1 + d_{1,1}\sigma_1)(-2v_1) + (c_{1,3}\sigma_3)(u_3). \tag{5.14}$$

or in matrix form, considering only the first element of $\boldsymbol{x}_{k+1}$, we have

$$x_{1,k+1} = \begin{bmatrix} c_{1,1} & d_{1,1} & c_{1,3} \end{bmatrix} \begin{bmatrix} \sigma_1 & \omega_1 & 0 \\ -\omega_1 & \sigma_1 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \begin{bmatrix} 2u_1 \\ -2v_1 \\ u_3 \end{bmatrix}, \tag{5.15}$$

and considering $\boldsymbol{x}_{k+1}$, it follows that

$$\boldsymbol{x}_{k+1} = \underbrace{\begin{bmatrix} \Re\left\{\boldsymbol{\upsilon}_1^{(\boldsymbol{x})}\right\} & \Im\left\{\boldsymbol{\upsilon}_1^{(\boldsymbol{x})}\right\} & \boldsymbol{\upsilon}_3^{(\boldsymbol{x})} \end{bmatrix}}_{\boldsymbol{\Upsilon}^{(\boldsymbol{x})}} \underbrace{\begin{bmatrix} \Re\{\lambda_1\} & \Im\{\lambda_1\} & 0 \\ -\Im\{\lambda_1\} & \Re\{\lambda_1\} & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{bmatrix} 2u_1 \\ -2v_1 \\ u_3 \end{bmatrix}}_{\underline{\boldsymbol{x}}_k}. \tag{5.16}$$

Therefore, a nonlinear change of coordinates is given by $\mathcal{T}_{\mathcal{K}} : \mathbb{R}^n \to \mathbb{R}^q$,

$$\mathcal{T}_{\mathcal{K}}(\boldsymbol{x}) = \underline{\boldsymbol{x}} = \left(\underline{x}_1, ..., \underline{x}_q\right)^\top, \tag{5.17}$$

where $\underline{x}_i = \phi_i$, if the i-th KEF is real; or $\underline{x}_i = 2\Re\{\phi_i\}$ and $\underline{x}_{i+1} = -2\Im\{\phi_i\}$ if the i- and $(i+1)$-th KEFs are complex conjugate pairs.

If the i-th KEF is real, the block diagonal $\Lambda_{i,i} = \lambda_i$, and the columns of $\boldsymbol{\Upsilon}^{(\boldsymbol{x})} \in \mathbb{R}^{n\times q}$ and $\boldsymbol{\Upsilon}^{(\boldsymbol{y})} \in \mathbb{R}^{m\times q}$ are, respectively, defined as $\boldsymbol{\Upsilon}^{(\boldsymbol{x})}_{:,i} = \boldsymbol{\upsilon}^{(\boldsymbol{x})}_i$ and $\boldsymbol{\Upsilon}^{(\boldsymbol{y})}_{:,i} = \boldsymbol{\upsilon}^{(\boldsymbol{y})}_i$.

If the i and $i+1$-th KEFs form a complex conjugate pair,

$$\begin{bmatrix} \Lambda_{i,i} & \Lambda_{i,i+1} \\ \Lambda_{i+1,i} & \Lambda_{i+1,i+1} \end{bmatrix} = \begin{bmatrix} \Re\{\lambda_i\} & \Im\{\lambda_i\} \\ -\Im\{\lambda_i\} & \Re\{\lambda_i\} \end{bmatrix}, \tag{5.18}$$

the columns of $\boldsymbol{\Upsilon}^{(\boldsymbol{x})}$ and $\boldsymbol{\Upsilon}^{(\boldsymbol{y})}$ are respectively defined as follows: $\boldsymbol{\Upsilon}^{(\boldsymbol{x})}_{:,i} = \Re\left\{\boldsymbol{\upsilon}^{(\boldsymbol{x})}_i\right\}$, $\boldsymbol{\Upsilon}^{(\boldsymbol{x})}_{:,i+1} = \Im\left\{\boldsymbol{\upsilon}^{(\boldsymbol{x})}_i\right\}$, $\boldsymbol{\Upsilon}^{(\boldsymbol{y})}_{:,i} = \Re\left\{\boldsymbol{\upsilon}^{(\boldsymbol{y})}_i\right\}$ and $\boldsymbol{\Upsilon}^{(\boldsymbol{y})}_{:,i+1} = \Im\left\{\boldsymbol{\upsilon}^{(\boldsymbol{y})}_i\right\}$.

Finally,

$$\underline{\boldsymbol{x}}_k = \boldsymbol{\Lambda}\underline{\boldsymbol{x}}_{k-1}, \tag{5.19}$$

$$\hat{\boldsymbol{x}}_k = \boldsymbol{\Upsilon}^{(\boldsymbol{x})}\underline{\boldsymbol{x}}_k, \tag{5.20}$$

$$\hat{\boldsymbol{y}}_k = \boldsymbol{h}\left(\hat{\boldsymbol{x}}_k\right) = \boldsymbol{\Upsilon}^{(\boldsymbol{y})}\underline{\boldsymbol{x}}_k. \tag{5.21}$$

Remark 3 (system dimension in canonical coordinates): The dimension of the vector defined by (5.17), $\underline{\boldsymbol{x}} \in \mathbb{R}^q$ is equal to the dimension of the vector of observable functions, $\boldsymbol{\gamma}(\boldsymbol{x}) \in \mathbb{R}^q$.

Remark 4 (state estimates): The state estimates, $\hat{\boldsymbol{x}}_k$, in (5.20) are mapped back to the state space through a linear transform. This is an important advantage as compared to other observers, which in general would require a nonlinear transform.

Remark 5 (covariance matrix): The filtered state covariance matrix, $\boldsymbol{\Sigma}_{k|k}$, can be recovered by $\boldsymbol{\Sigma}_{k|k} = \boldsymbol{\Upsilon}^{(\boldsymbol{x})}\underline{\boldsymbol{\Sigma}}_{k|k}\left(\boldsymbol{\Upsilon}^{(\boldsymbol{x})}\right)^\top$, where $\underline{\boldsymbol{\Sigma}}_{k|k}$ is the filtered covariance matrix in the Koopman canonical coordinates.

Remark 6 (system observability): Surana and Banaszuk [71] proved that if Assumption I holds true and the pair $\left(\boldsymbol{\Lambda}, \boldsymbol{\Upsilon}^{(\boldsymbol{y})}\right)$ is observable, then the system (5.1)–(5.2) is nonlinear observable. They also demonstrated through an example, that Assumption I is not a necessary condition. Further theoretical development with respect to system observability is required.

Remark 7 (modal observability): Interestingly, because we have access to the Koopman modes and the transition matrix, $\boldsymbol{\Lambda}$, a measure of modal observability [175, 176] is practically a byproduct of the KKF. This is an advantage since a system mode cannot be modified unless it is both observable and controllable. However, to define controllability, it is necessary to consider a input-output nonlinear system model [72]. The approach proposed by Hamdan and Nayfeh [177] is here recommended because it is scaling-independent.

5.1.1 Example: Canonical nonlinear dynamical system

Consider the autonomous nonlinear dynamical system expressed as [71]

$$\dot{\boldsymbol{x}} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \rho x_1 \\ \mu x_2 + (2\rho - \mu)\, c x_1^2 \end{bmatrix}, \tag{5.22}$$

$$y = h(\boldsymbol{x}) = x_1^2 + x_2. \tag{5.23}$$

Thus, by selecting the *observables*

$$\underline{\boldsymbol{x}} = \mathcal{T}_{\mathcal{K}}(\boldsymbol{x}) = \left[x_1;\, x_2 - c x_1^2;\, x_1^2\right]^{\top}, \tag{5.24}$$

we have

$$\begin{aligned} \dot{\underline{x}}_1 &= \dot{x}_1 = \rho x_1 = \rho \underline{x}_1, \\ \dot{\underline{x}}_2 &= \dot{x}_2 - 2c x_1 \dot{x}_1 = \mu x_2 + (2\rho - \mu)\, c x_1^2 - 2c x_1 \dot{x}_1 = \mu \underline{x}_2, \\ \dot{\underline{x}}_3 &= 2 x_1 \dot{x}_1 = 2\rho x_1^2 = 2\rho \underline{x}_3, \end{aligned} \tag{5.25}$$

and it follows that

$$\dot{\underline{\boldsymbol{x}}} = \begin{bmatrix} \rho & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & 2\rho \end{bmatrix} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \underline{x}_3 \end{bmatrix} = \boldsymbol{\Lambda} \underline{\boldsymbol{x}}, \tag{5.26}$$

$$\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & c \end{bmatrix} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \underline{x}_3 \end{bmatrix} = \boldsymbol{\Upsilon}^{(\boldsymbol{x})} \underline{\boldsymbol{x}}, \tag{5.27}$$

$$y = \begin{bmatrix} 0 & 1 & c+1 \end{bmatrix} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \underline{x}_3 \end{bmatrix} = \boldsymbol{\Upsilon}^{(y)} \underline{\boldsymbol{x}}. \tag{5.28}$$

As noted in [71], even when $\boldsymbol{x}_0$ follows a Gaussian distribution, $\underline{\boldsymbol{x}}_0$ will in general not follow a Gaussian distribution under the Koopman canonical transformation, $\mathcal{T}_{\mathcal{K}}$. Therefore, the adoption of the KF as presented in section 3 might result in severely biased estimates. Conversely, the robust Kalman filter introduced in section 3.2 has high statistical efficiency at the Gaussian distribution and, at the same time, provides reliable estimates in the presence of any deviation from Gaussianity, as long as the breakdown point of the estimator is not exceeded, which is about 35 percent.

Note that an improper estimation of the Koopman tuples would affect the transformation (5.17). Thus, we preprocess the data using a robust filter [77, 78], which will be presented in chapter 6.

5.2 Sparse selection of state estimates

Following Manohar et al. [170], suppose that a state $\boldsymbol{x}$ evolving according to the nonlinear dynamics (5.1) has a compact representation in a transform basis $\boldsymbol{\Psi}$. In a universal basis $\boldsymbol{\Psi} \in \mathbb{R}^{n\times n}$, $\boldsymbol{x}$ might have a sparse representation:

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{s}, \tag{5.29}$$

$\boldsymbol{s} \in \mathbb{R}^n$ is a sparse vector. In a tailored basis $\boldsymbol{\Psi}_r \in \mathbb{R}^{n\times r}$, such as a basis defined by a proper orthogonal decomposition, $\boldsymbol{x}$ might have a low-rank representation:

$$\boldsymbol{x} = \boldsymbol{\Psi}_r\boldsymbol{\ell}, \tag{5.30}$$

$\boldsymbol{\ell} \in \mathbb{R}^r$. We seek to find a matrix $\boldsymbol{C} \in \mathbb{R}^{p\times n}$ consisting of a small number ($p \ll n$) of optimized measurements:

$$\boldsymbol{b} = \boldsymbol{C}\boldsymbol{x}, \tag{5.31}$$

$\boldsymbol{b} \in \mathbb{R}^p$, which facilitates the accurate reconstruction of $\boldsymbol{s}$ or $\boldsymbol{\ell}$, and thus $\boldsymbol{x}$. Note that $\boldsymbol{C} = [\boldsymbol{e}_{\gamma_1}\ \boldsymbol{e}_{\gamma_2}\ ...\ \boldsymbol{e}_{\gamma_p}]^\top$, where $\boldsymbol{e}_{\gamma_i}$ is the unit vector with a unit entry at index γ_i and zeros elsewhere. Combining (5.29) and (5.31) yields:

$$\boldsymbol{b} = (\boldsymbol{C}\boldsymbol{\Psi})\,\boldsymbol{s} = \boldsymbol{\theta}\boldsymbol{s}. \tag{5.32}$$

Eq. (5.32) is referred to as the compressed sensing problem. Conversely, combining (5.30) and (5.31) yields:

$$\boldsymbol{b} = (\boldsymbol{C}\boldsymbol{\Psi}_r)\,\boldsymbol{\ell} = \boldsymbol{\theta}\boldsymbol{\ell}. \tag{5.33}$$

If $\boldsymbol{C}$ is properly structured such that $\boldsymbol{\theta}$ is well conditioned, it is possible to solve for the low-rank coefficients $\boldsymbol{\ell}$ given the measurements $\boldsymbol{b}$ in (5.33) as follows:

$$\hat{\boldsymbol{\ell}} = \begin{cases} \boldsymbol{\theta}^{-1}\boldsymbol{b} = \left(\boldsymbol{C}\boldsymbol{\Psi}_r\right)^{-1}\boldsymbol{b}, & p = r, \\ \boldsymbol{\theta}^{\dagger}\boldsymbol{b} = \left(\boldsymbol{C}\boldsymbol{\Psi}_r\right)^{\dagger}\boldsymbol{b}, & p > r, \end{cases} \tag{5.34}$$

$\boldsymbol{\theta}^\dagger$ denotes the Moore-Penrose pseudoinverse of $\boldsymbol{\theta}$. Thus, $\boldsymbol{x}$ can be estimated as $\hat{\boldsymbol{x}} = \boldsymbol{\Psi}_r\hat{\boldsymbol{\ell}}$. From (5.34), one seeks columns of $\boldsymbol{\Psi}_r$ corresponding to point sensor locations in the state space, $\boldsymbol{e}_{\gamma_i}$, that optimally condition the inversion of the matrix $\boldsymbol{\theta}$. The structure of the elements $\boldsymbol{e}_{\gamma_i}$ affect the condition number of $\boldsymbol{C}$ and consequently of $\boldsymbol{M}_{\boldsymbol{\gamma}} = \boldsymbol{\theta}^\top\boldsymbol{\theta}$. The condition number of the system might be indirectly bounded by optimizing the spectral content of $\boldsymbol{M}_{\boldsymbol{\gamma}}$ using its determinant, trace, or spectral radius. For example, we have:

$$\begin{aligned} \boldsymbol{\gamma}_* &= \underset{\boldsymbol{\gamma},|\boldsymbol{\gamma}|=p}{\arg\max}\ |\det \boldsymbol{M}_{\boldsymbol{\gamma}}| = \underset{\boldsymbol{\gamma}}{\arg\max}\ \prod_i |\lambda_i(\boldsymbol{M}_{\boldsymbol{\gamma}})| \\ &= \underset{\boldsymbol{\gamma}}{\arg\max}\ \prod_i \sigma_i(\boldsymbol{M}_{\boldsymbol{\gamma}}), \end{aligned} \tag{5.35}$$

where λ_i and σ_i are, respectively, the i-th eigenvalue and singular value of $\boldsymbol{M}_{\boldsymbol{\gamma}}$. The QR factorization with column pivoting decomposes a matrix $\boldsymbol{M}_{\boldsymbol{\gamma}} \in \mathbb{R}^{m \times n}$ into a unitary matrix $\boldsymbol{Q}$, an upper triangular matrix $\boldsymbol{R}$, and a column permutation matrix $\boldsymbol{C}$, that is:

$$\boldsymbol{M}_{\boldsymbol{\gamma}} \boldsymbol{C}^{\top} = \boldsymbol{Q}\boldsymbol{R}. \tag{5.36}$$

The key idea from [170] is, when applied to an appropriate basis, the QR pivoting procedure provides an approximate greedy solution method for the optimization in (5.35), also known as a submatrix volume maximization because the matrix volume is the absolute value of the determinant.

Sparse Selection of State Estimates

In [170], the QR pivoting procedure is proposed as a tool to optimize sensor placement, in particular for the reconstruction of high-dimensional states from point measurements given tailored bases. Instead of measurements, here, we have access to the state estimates received from the robust, model-based, decentralized DSEs. Our objective is to shrink the number of state estimates used by the robust KKF so as to speed up its processing.

Algorithm 1 Sparse selection of state estimates

procedure
  $\boldsymbol{\Psi}_r \leftarrow \text{svd}(\hat{\boldsymbol{x}})$
  if $p == r$ **then**
    $\boldsymbol{\gamma} \leftarrow pivot(\boldsymbol{\Psi}_r)$ ▷ $[Q, R, pivot] = \text{qr}(\boldsymbol{\Psi}_r)$
  else if $p > r$ **then**
    $\boldsymbol{\gamma} \leftarrow pivot(\boldsymbol{\Psi}_r \boldsymbol{\Psi}_r^{\top})$ ▷ $[Q, R, pivot] = \text{qr}(\boldsymbol{\Psi}_r \boldsymbol{\Psi}_r^{\top})$
  $\boldsymbol{b} \leftarrow \hat{\boldsymbol{x}}_{\boldsymbol{\gamma}}$ ▷ $\hat{\boldsymbol{x}}_{\boldsymbol{\gamma}} = \hat{\boldsymbol{x}}_{\gamma_1:\gamma_p}$

After $\boldsymbol{b}$ is determined as in Algorithm 1, we make use of (5.17) such that:

$$\mathcal{T}_{\mathcal{K}}(\boldsymbol{b}) = \underline{\boldsymbol{x}} = \left[\underline{x}_1, ..., \underline{x}_q\right]^{\top}. \tag{5.37}$$

5.3 Numerical results

5.3.1 Robust Koopman Kalman filter

The simulations are carried out on the IEEE 39-bus system, whose data can be found in [158]. At $t = 5$ seconds, the transmission line connecting Buses 15 and 16 is switched off, which results in a large disturbance. The values obtained from the time-domain simulation are supposedly the true values. We assume that PMUs are placed at each generator terminals, which provide measurements of voltage and current at 60 frames per second. A random

Gaussian noise, $\mathcal{N}(0, 0.01)$, is assumed for both system and measurement error such that the covariance matrices $\boldsymbol{W}_k = \boldsymbol{R}_k = \text{diag}\,(10^{-4})$. We performed tests with the classical and the two-axis model of synchronous generators; in the later case, the model of the excitation and the turbine-governor systems is also considered. The algebraic differential equations of both models are given in Appendix A.

The root mean squared error, defined as

$$\epsilon_k = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |x_{i,k} - \hat{x}_{i,k}|^2}, \quad k = 1, ..., \ell, \tag{5.38}$$

is used to evaluate the estimation accuracy.

On the choice of observables

Surana [73] adopted the exact dynamic mode decomposition proposed by Tu et al. [111] to compute approximations to the Koopman tuples, by making the observables $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{x} = \left[\boldsymbol{\delta}^\top; \boldsymbol{\omega}^\top\right]^\top$; it was assumed that real-time measurements of the rotor angle, $\boldsymbol{\delta}$, and rotor speed, $\boldsymbol{\omega}$, were (partially) available. For the simulations carried out on the IEEE 39-bus system, $[\delta_2, ..., \delta_8, \omega_2, ..., \omega_{10}]^\top$ were supposedly measured.

In this work, we apply instead the EDMD as proposed by Klus et al. [113]. Our choice of observables, $\boldsymbol{g}(\boldsymbol{x}) = \left[\boldsymbol{P}_e^\top; \boldsymbol{Q}_e^\top\right]^\top$, differs from the state vector, $\boldsymbol{x} = \left[\boldsymbol{\delta}^\top; \boldsymbol{\omega}^\top\right]^\top$; thanks to the strong coupling between the rotor angle, $\boldsymbol{\delta}$, and the real power, $\boldsymbol{P}_e$, we found that real-time measurements of the rotor angle are not necessary, although recommended. This is an important remark because it relaxes the requirement of real-time measurements of the rotor angle, thereby allowing the application of a hybrid approach where a model is simulated (offline) to find out the Koopman tuples, whereas only measurements of active and reactive power are processed in real-time.

Estimates versus GPS-synchronized measurements of the generators' rotor angle

Many attempts of estimating the rotor angle without having to place a sensor on the machine shaft to measure it directly have been reported; see for example [178–180]. After a careful review of the literature, we found that the proposed techniques either rely on excessively restrictive assumptions or rely on a machine model thereby being system dependent.

Interestingly, ongoing efforts are indicating that GPS-synchronized measurements of the rotor angle is a viable alternative, thanks to the advent of new technologies in sensing and communications [173, 174].

GM-KKF versus KKF

Here, a comparison between the KKF [71, 73] and the proposed GM-KKF is carried out. Two cases are analyzed; they are described next.

a) Single outlier in the data: The results are presented in Fig. 5.1a. It can be observed

that unlike the KKF, the GM-KKF is insensitive to the outlier. Furthermore, considering the same initial condition, the GM-KKF shows much faster convergence rate due to the batch-mode regression formulation in (3.23). This fact can be confirmed in Fig. 5.2, where the root mean squared error ϵ_k, $k = 1, ..., \ell$ (5.38) is plotted.

b) Multiple outliers in the data: The GM-KKF provides slight bias but bounded estimates in this case, whereas the KKF results are unbounded; in fact, the bias of the KKF estimates are unbounded since they might take values over $(-\infty, +\infty)$. The results are illustrated in Fig. 5.1b.

GM-KKF versus GM-EKF

A comparison between the GM-EKF [50] and the proposed GM-KKF is carried out. Two cases are analyzed as follows.

a) The model of the GM-EKF is perfect and its parameters are correct: The results displayed in Fig. 5.1c show that both the GM-EKF and the GM-KKF are able to suppress the influence of a single outlier placed at $t = 10$ seconds.

b) One parameter of the GM-EKF model is incorrect: One single diagonal element of the admittance matrix of the Kron reduced network model is modified by 10% to introduce a structural error into the GM-EKF. This procedure induces a structural outlier [51] whose effect on the estimation results is very difficult to suppress. Conversely, the GM-KKF is insensible to structural outliers as no model is assumed. Table 5.1 shows the overall mean squared estimation error, $\varepsilon = \sum_{k=1}^{\ell} \epsilon_k$, for this case, as well as for the previous case where a perfect model is assumed.

Table 5.1: The influence of model/parameter uncertainty in the estimation and the benefit of a model-free approach

	Overall Estimation Error, ε	
	GM-EKF [50]	GM-KKF
Perfect model	0.0141	0.0281
Single parameter error	0.0365	0.0281

Besides model independence, another advantage of the proposed GM-KKF over the GM-EKF is its computational speed. Indeed, the GM-KKF is about three times faster than the GM-EKF in the simulations carried out on the IEEE 39-bus test system; this advantage is expected to be more pronounced for larger systems. The computation time required to simulate one discrete time step of the KKF, the GM-EKF, and of the proposed GM-KKF, which are averaged over 500 runs, is presented in Table 5.2.

Probability distribution after transformation (5.17)

As stated in [71], even when $\boldsymbol{x}$ is normally distributed, $\underline{\boldsymbol{x}}$ will in general not be normally distributed under the transformation (5.17). From the simulations, we found that $\underline{\boldsymbol{x}}$ follows a Student's t-distribution with 20 degrees of freedom as depicted in Fig. 5.3. It is expected

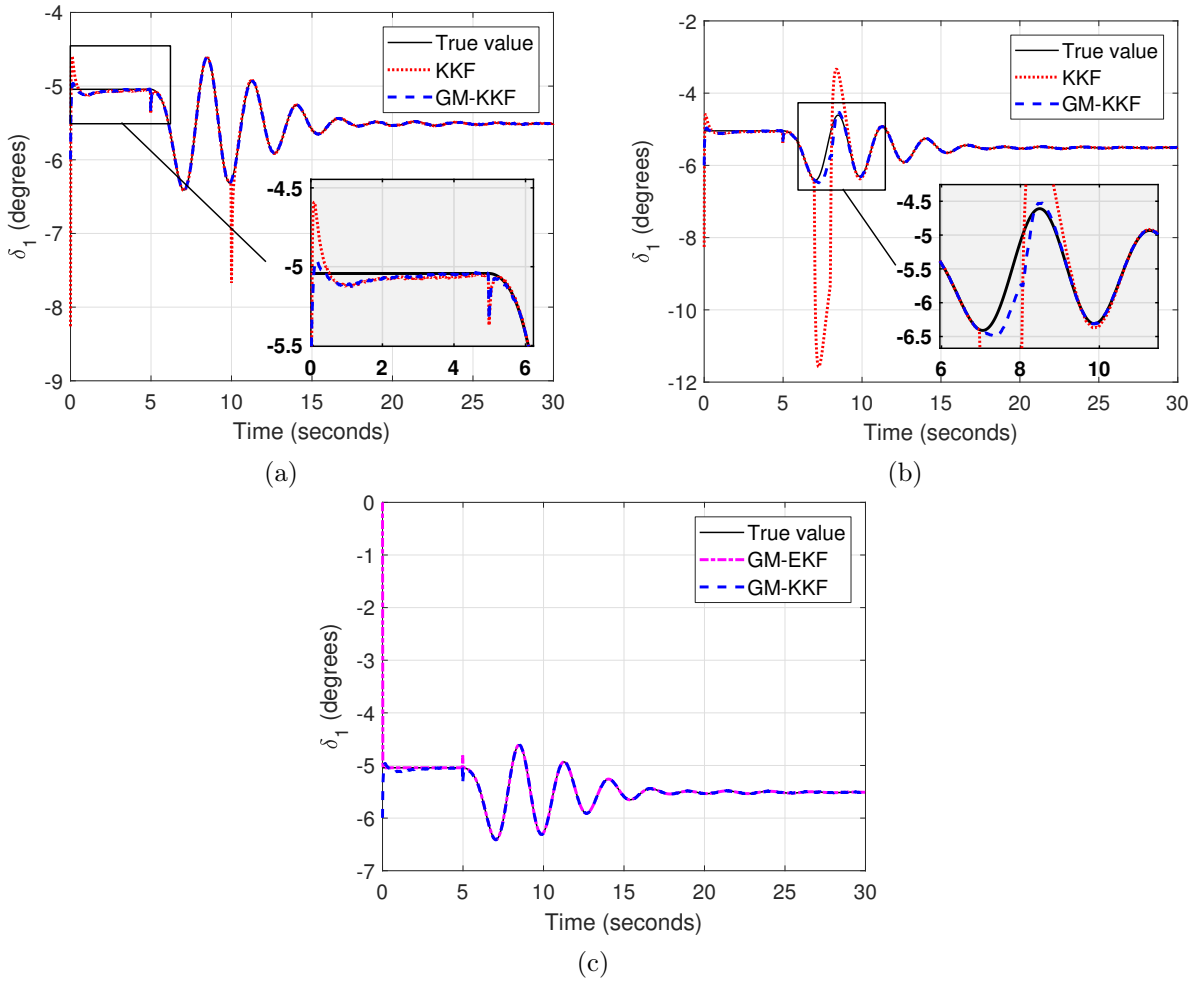


Figure 5.1: Comparison between the proposed GM-KKF and (a) the KKF as proposed in [71, 73] with one single outlier placed at $t = 10$ seconds. (b) the KKF as proposed in [71, 73] with multiple outliers placed from $t = 7$ to $t = 8$ seconds. (c) the GM-EKF as proposed in [50].

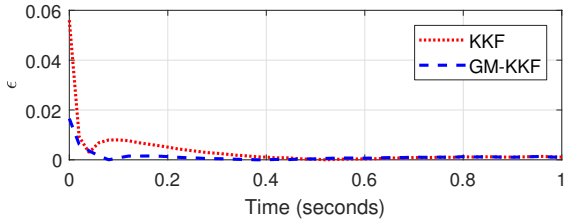


Figure 5.2: Comparison of the convergence rate between the KKF as proposed in [71, 73] and the GM-KKF proposed in here.

Table 5.2: Computation time of the KKF [71,73], GM-EKF [50], and the proposed GM-KKF per discrete time step

	KKF [71,73]	GM-EKF [50]	GM-KKF
Time (milliseconds)	0.1884	17.0335	6.1218

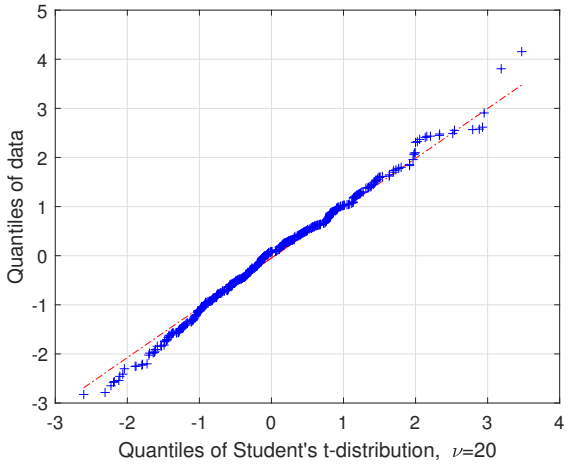


Figure 5.3: The Q-Q plot of data after transformation (5.17) indicates that it roughly follows a Student's t-distribution with 20 degrees of freedom.

that this probability distribution will vary in a case-by-case basis; this is not a problem for the GM-KKF since it exhibits a high efficiency under deviations from normality.

The case with multiple outliers in the data is also simulated using *Model B*. As observed from Fig. 5.4, the EKF exhibits a performance that severely degrade in presence of outliers, which is not the case for the GM-EKF and the GM-KKF, which are able suppress them.

5.3.2 Hierarchical decentralized dynamic state estimation

We carry out simulations on the 2,000-bus synthetic Texas system [181] comprising 544 synchronous and 87 non-synchronous generators; the latter are modeled as variable-speed wind generators with full converters and do not contribute to the electromechanical modes. A fault is applied to Bus 1017 and cleared after 10 milliseconds. Fig. 5.5 shows the rotor angle of the synchronous generators relative to the system average angle, with removed mean. We observe that 456 of 544 synchronous generators present coherent dynamics, whereas 88 of 544 synchronous generators do not contribute at all to the dynamics; see constant line at zero degrees in Fig. 5.5.

Upon application of the KMD explained in Section III, the 10 most important Koopman eigenvalues–i.e., the ones with the smaller damping coefficient–are shown in Fig. 5.6. They were estimated using 10 seconds of rotor speed estimates. The black crosses indicate results

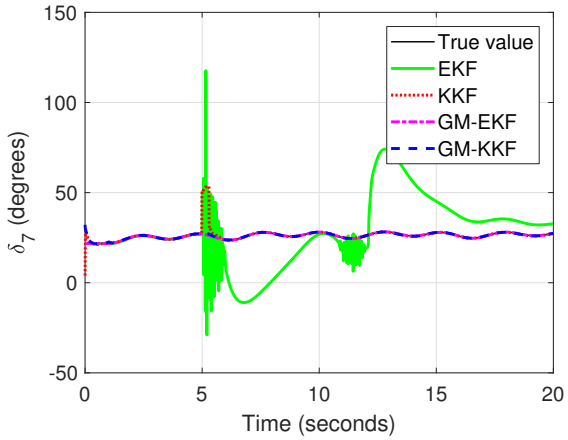


Figure 5.4: Comparison between the proposed GM-KKF and the EKF, KKF [73], and GM-EKF [50]. Outlier placed between $t = 5.0$ and $t = 5.3$ seconds.

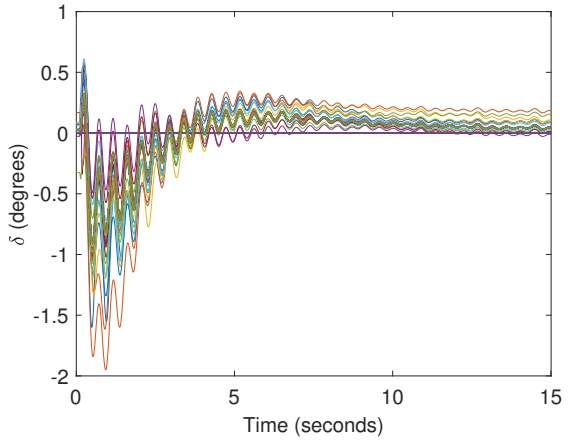


Figure 5.5: Rotor angle, δ, of the 544 synchronous generators relative to system average angle, after removing the mean.

obtained with all 544 estimates obtained from the decentralized DSE (assuming we have local PMU measurements), whereas the blue circles indicate results obtained with 240 principled selected estimates from the sparse state measurement selection technique explained in Section IV. The remaining 5 seconds of the time domain simulation are used for testing. We consider three different scenarios, as presented next.

No outliers

The results are shown in Fig. 5.7a. *True state*, *received state estimate*, and *filtered state estimate*, respectively, refer to the outcome of the time-domain simulation, the data received from the robust model-based decentralized DSEs, and the outcome of the robust KKF.

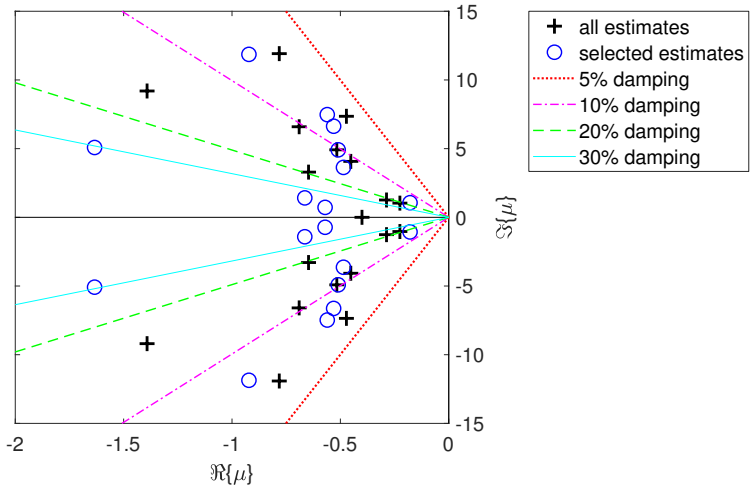


Figure 5.6: The Koopman modes calculated using estimates of the rotor speed of the synchronous generators.

Impulsive noise

The results are shown in Fig. 5.7b. We observe that the robust KKF is able to suppress the impulsive noise.

Loss of the communication link with a decentralized DSE

The results are shown in Fig. 5.7c, and demonstrate the forecasting capability of the robust KKF.

5.4 Concluding remarks

A robust Koopman operator-based Kalman filter has been developed and compared with its non-robust version as well as with a robust extended Kalman filter (GM-EKF). Examples have been provided to illustrate the importance of using a GM-KKF for dynamic state estimation because it is model independent and more computational efficient than the GM-EKF.

The proposed hybrid framework offers a balance between model-based and data-driven methods, and it has several important advantages compared to other methodologies. It is completely independent of the network model, does not attempt to model the dynamics of the loads, and, most importantly, does not depend on the infrastructure that is exclusively available at the control centers.

Although promising, the proposed approach requires further investigations. One important aspect is the system observability at the upper level, i.e., the sparse selection must be con-

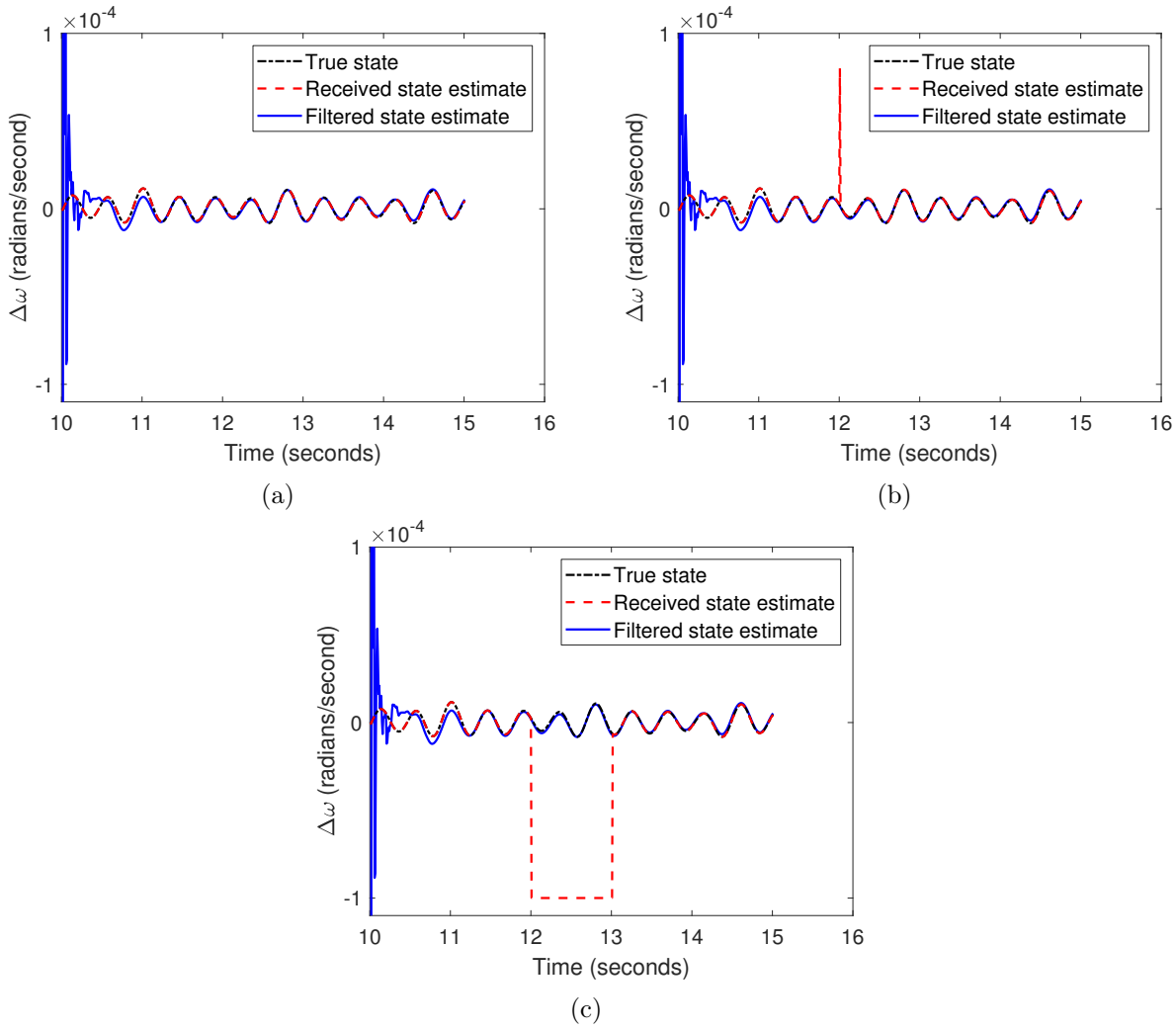


Figure 5.7: Rotor speed deviation, $\Delta\omega$, of the synchronous generator 4. (a) No outlier. (b) Impulsive noise at $t = 12$ seconds. (c) Loss of communication with the decentralized DSE between $t = 12$ and $t = 13$ seconds.

ditioned to a certain degree of redundancy. We will rely on the correspondence between the Koopman operator and the Lie derivatives to pursue this effort. In addition, the estimation of the KEFs [114] is an important open problem and must be addressed. Future work will also look into the applicability of using other observables from sensors in the upper level system identification by either partially or completely bypassing the lower level DSE in certain scenarios to gain increased agility.

Chapter 6

Robust data filtering

Practical experience has demonstrated that PMU measurements might exhibit impulsive noises, which are induced by: improper hardware wiring, unavailability of the GPS time reference, GPS synchronization error, and communication interferences [182], only to mention a few examples. Impulsive noises are sporadic noises with sudden short duration and large magnitudes, thereby deviating from the main pattern of the majority of the data points. Without a proper mechanism capable of suppressing the undesirable effects of the impulsive noises, any application relying on PMU measurements is vulnerable. This observation motivated us to develop the robust filter [77, 78] that is presented in this chapter. The developed filter is termed iteratively reweighted phase-phase correlator (IPPC).

Although we later adopted it within the Koopman Kalman filter that has been presented in the previous chapter, the IPPC was initially tested with the Prony method [183]. The Prony method is a measurement-based approach tailored to signals that arise following a transient event. It has been widely adopted by the electric power systems industry, since the work of Hauer et al. [184].

Different versions of the Prony method have been widely adopted, despite not being robust to outliers. This weakness has motivated recent work on the derivation of robust versions [185]. In [185], the minimum covariance determinant (MCD) proposed by Rousseeuw [146] is applied to robustly estimate the frequency and the associated damping factor of the assumed exponential model. However, MCD presents poor statistical efficiency under Gaussian noise and is computationally intensive for very large data sets. Also, the method proposed in [185] does not robustly estimate the amplitude and initial phase of the assumed exponential model, which together define the so-called residues. This is an important deficiency since the residues allow us to determine the mode shapes and to identify the coherent groups of generators, which is instrumental for dynamic model reduction. Furthermore, they are a necessary information for several procedures used to discriminate between dominant and spurious modes [186, 187]. Spurious modes are typically introduced by the estimation algorithms in order to improve their statistical efficiency while dealing with noisy measurements; but they

have no physical significance.

The robust multivariate Prony method presented in this chapter is able to robustly estimate the frequency and damping factor, as well as the amplitude and initial phase. The method consists of the following steps. Firstly, the voltage magnitude and voltage phase angle are combined to build a complex-valued series of data points. Then, using the IPPC proposed in [74–76] and extended in [77, 78] to handle outliers in the magnitude and in the phase angle of voltage phasor measurements, a robust estimator of the autocorrelation function is executed to identify and suppress the outliers throughout the iterative process. Finally, the multivariate Prony method is executed using the robustly estimated voltage angle measurements. The proposed method, which is termed R-Prony, presents good robustness to white impulsive noise while maintaining high statistical efficiency under Gaussian noise.

6.1 Iteratively reweighted phase-phase correlator

Tamburello and Mili [74] studied the statistical properties of the phase-phase correlator (PPC) [188], a non-linear estimator of correlation that exhibits robustness to impulsive noise in the magnitude of complex data samples. For several signal processing applications, such as radar and speech, the existence of an outlier in the phase argument is not of concern. This is not the case here since we make use of voltage phasors as input data, and the voltage phase angle measured by PMUs might be corrupted by outliers as well. To further decrease the PPC bias, Tamburello and Mili [75] proposed an iterative version of the PPC, the so-called IPPC, which again only exhibits robustness to outliers in the magnitude. We extended this estimator by making it robust to impulsive noise in the phase argument as well. In what follows, we first present the PPC and then we later explain its iterative version, the IPPC.

By using only the phase information, along with a scalar obtained from the Gaussian hypergeometric function, the PPC estimates the autocorrelation sequence of a zero-mean complex Gaussian process. It exploits the fact that the time correlation of a complex signal is contained completely in its phase. The PPC is phase-preserving and hard-limiting. In the complex domain, the hard-limiting function is the sign function, $y_k = \operatorname{sign}(x_k) = e^{j \arg(x_k)}$, where x_k is the data sample at time step k. At time lag τ, the PPC is simply defined as $R_{yy}(\tau) = E\left[y_k y_{k-\tau}^*\right]$, and an unbiased estimator is expressed as

$$\hat{R}_{yy}(\tau) = \frac{1}{N - |\tau|} \sum_{k=0}^{N-|\tau|-1} y_k y_{k-\tau}^*. \tag{6.1}$$

Recall that the unbiased sample autocorrelation estimator, which is asymptotically the maximum likelihood estimator of the correlation coefficient under Gaussianity, is defined as

$$\hat{\rho}_\tau = \hat{R}_{xx}(\tau) = \frac{1}{N - |\tau|} \sum_{k=0}^{N-|\tau|-1} x_k x_{k-\tau}^*. \tag{6.2}$$

An estimate of the autocorrelation coefficient at time lag τ based on the PPC is defined as

$$\hat{\varrho}_\tau = e^{j \arg\left(\hat{R}_{yy}(\tau)\right)} \left(f\left(\hat{R}_{yy}(\tau)\right)\right)^{-1}, \tag{6.3}$$

where

$$f\left(\hat{R}_{yy}(\tau)\right) = \left(\frac{\pi}{4}\right) \cdot \left|\hat{R}_{yy}(\tau)\right| \cdot {}_2F_1\left(\frac{1}{2}, \frac{1}{2};\ 2;\ \left|\hat{R}_{yy}(\tau)\right|^2\right), \tag{6.4}$$

and ${}_2F_1(\cdot)$ is the Gaussian Hypergeometric function,

$${}_2F_1(a, b;\ c;\ z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)} \frac{z^n}{n!}. \tag{6.5}$$

Note that from a practical viewpoint, we simply rely on a look-up table that has the values taken by the Gaussian hypergeometric function given by (6.5). Also, the PPC is only robust to outliers in the magnitude of the complex data samples, thanks to the use of the sign function. For detailed derivations and statistical properties of the PPC, the reader is referred to [74–76, 188]. Nonetheless, although (6.3) defines a robust estimator of the autocorrelation sequence, it is possible to define a robust PPC-based correlation estimator. Formally,

$$\hat{R}_{yv}(\tau) = \frac{1}{N - |\tau|} \sum_{k=0}^{N-|\tau|-1} y_k v^*_{k-\tau}, \tag{6.6}$$

where $\boldsymbol{y}$ and $\boldsymbol{v}$ are, in general, different time series.

The modified IPPC algorithm described in Table 6.1 is robust to outliers in the magnitude and in the phase argument. We take advantage of the procedure developed in [75] to firstly identify outliers through the use of robust Mahalanobis distances, and then suppress their influence in the estimation by means of a weighting function, introduced into the PPC by replacing the sample average in (6.1) with a weighted average,

$$\hat{R}_{yy-\varpi}(\tau) = \frac{\sum_{k=0}^{N-|\tau|-1} y_k y^*_{k-\tau} \varpi_k}{\sum_{k=1}^{N-|\tau|} \varpi_k}. \tag{6.7}$$

In summary, the outliers in the magnitude are removed through the use of the sign function, while the outliers in the phase argument are downweighted through an iterative process by a weighting function.

The vector of weights, $\boldsymbol{\varpi}$, is set to $\mathbf{1}$ at the initialization; $\boldsymbol{y}_i = [y_{\tau+i}, y_i]^\top$, $i = 0, ..., N-|\tau|-1$; $\hat{\zeta}_{\frac{1}{2}}(|\boldsymbol{y}|, \boldsymbol{\varpi})$ is the weighted median of $\boldsymbol{y}$; $\hat{\sigma}^2$ is a robust estimator of scale; $\hat{\boldsymbol{\Sigma}}$ is a robust covariance matrix; RMD stands for the robust Mahalanobis distance. For data samples with $N > 100$, Tamburello and Mili [75] determined that $\left(2 \cdot \text{RMD}^2\right)$ follows a Chi-Square distribution with 4 degrees of freedom; thus, values exceeding a threshold of 97.5 percentile,

Table 6.1: IPPC algorithm at iteration ν, modified from [75]

1) Calculate: $\hat{R}^{(\nu)}_{yy-\varpi}(\tau)$ using (6.7)

2) Calculate: $\hat{\sigma}^{2\,(\nu)} = \dfrac{2}{\ln(4)}\left(\hat{\zeta}_{\frac{1}{2}}\left(|\boldsymbol{y}|, \boldsymbol{\varpi}\right)\right)^2$

3) Calculate: $\hat{\boldsymbol{\Sigma}}^{(\nu)} = \hat{\sigma}^{2\,(\nu)} \begin{bmatrix} 1 & \hat{R}^{(\nu)}_{yy-\varpi}(\tau) \\ \hat{R}^{*\,(\nu)}_{yy-\varpi}(\tau) & 1 \end{bmatrix}$

4) Calculate: $\mathrm{RMD}_i^{(\nu)} = \sqrt{\boldsymbol{y}_i^H \left(\hat{\boldsymbol{\Sigma}}^{(\nu)}\right)^{-1} \boldsymbol{y}_i}$

5) Calculate: $\varpi_i^{(\nu)} = \min\left[1,\ \chi^2_{4,0.975} \Big/ \left(2\left(\mathrm{RMD}_i^{(\nu)}\right)^2\right)\right]$

6) Define: $y = y_i^{(\nu+1)}$ and $x = x_i^{(\nu+1)}$. Calculate:

a) $y = |y| e^{j\left[\arg(y) + \left(\arg\left(y_i^{(\nu)}\right) - \arg(y)\right)\cdot\left(1-\left(\varpi_{i+1}^{(\nu)}\right)^2\right)\right]}$

b) $\delta = \left[\arg(x) + \left(\arg\left(x_i^{(\nu)}\right) - \arg(x)\right)\cdot\left(1 - \left(\varpi_{i+1}^{(\nu)}\right)^2\right)\right]$

7) Stop if $\max\left(\left|\Delta\hat{\boldsymbol{\varrho}}^{(\nu)}\right|\right) < 10^{-2}$ or if $(k = k_{max})$, otherwise increment k and go to Step 1.

$\chi^2_{4,0.975}$, are flagged as outliers and properly downweighted through ϖ_i's. The algorithm converges if $\max\left(|\hat{\boldsymbol{\varrho}}^{(\nu)} - \hat{\boldsymbol{\varrho}}^{(\nu-1)}|\right) < 10^{-2}$.

Fig. 6.1 illustrates how the unbiased sample autocorrelation (6.2) is affected by a single outlier, while the robust IPPC-based autocorrelation (6.3) suppress the effect of the impulsive noise, providing results numerically very close to the true autocorrelation.

Note that the PPC, and consequently the IPPC, are derived under the assumption that the Gaussian complex-valued data sequences are wide-sense stationary [74, 188]. However, the Prony method is tailored to signals that arise following a transient event and are, in general, non-stationary. Furthermore, a studies carried out by using real PMU data revealed that the errors associated with measurements of the voltage phase angle roughly follow a Gaussian distribution. Also, the errors associated with measurements of the voltage magnitude follow a bimodal Gaussian mixture distribution. Thus, the IPPC is simply applied as a filter to suppress outliers in the measurements of voltage phase angle, before they are made available to the Prony method.

6.2 Robust Prony method

The single channel classical Prony method [183], followed by its extension to the multichannel version, is presented in this section. A generalized formulation in the complex domain is adopted.

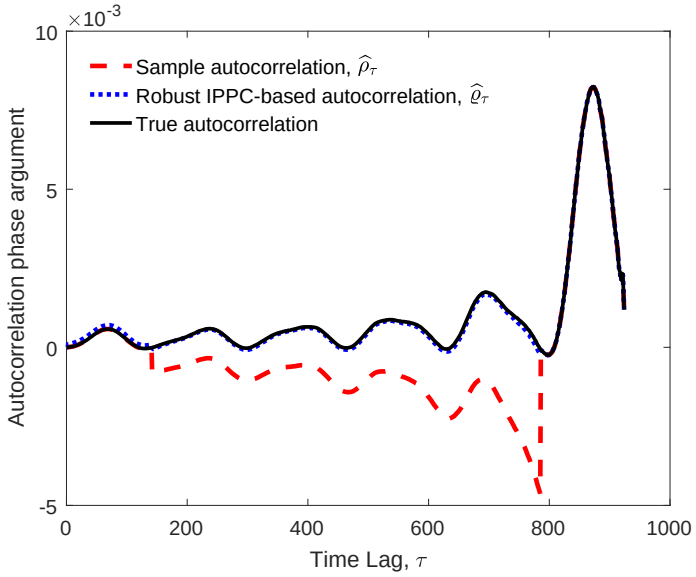


Figure 6.1: Autocorrelation values computed using: measurement with impulsive noise, IPPC-filtered measurement, true data sample without impulsive noise.

6.2.1 Single channel Prony method

Given a complex-valued data sequence $\mathbf{x} = [x_0 \ ... \ x_{N-1}]^\top$, the conventional Prony method fits an exponential model to the data in the least squares sense, and estimates $\hat{x}_k$, $k = 0, ..., N-1$. Formally, we have

$$\hat{x}_k = \sum_{i=1}^{p} A_i e^{(\alpha_i + j2\pi f_i)(k-1)T + j\phi_i} = \sum_{i=1}^{p} h_i z_i^{k-1}, \tag{6.8}$$

where the subscript i indicates a parameter associated with the i-th mode; A_i, α_i, f_i and ϕ_i are, respectively, amplitude, damping factor, frequency and initial phase; T is the sample interval in seconds; $h_i \triangleq A_i e^{j\phi_i}$ and $z_i \triangleq e^{(\alpha_i + j2\pi f_i)T}$ are the i-th residue and polynomial root, respectively. The Prony method can be divided in three steps as follows.

Estimation of the characteristic polynomial coefficients

Consider the linear prediction model,

$$\begin{bmatrix} x_p \\ x_{p+1} \\ \vdots \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} x_{p-1} & x_{p-2} & ... & x_0 \\ x_p & x_{p-1} & ... & x_1 \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-2} & x_{N-3} & ... & x_{N-p-1} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}, \tag{6.9}$$

in a compact form, $\mathbf{d} = \mathbf{D} \cdot \mathbf{a}$, where x_k, $k = 0, ..., N-1$, are complex-valued data sequences as previously defined, $\mathbf{a}$ is the vector of coefficients, and p is the polynomial order assumed to be known a priori. An estimate of $\mathbf{a}$ is obtained by solving (6.9) in the least squares sense

as $\hat{\mathbf{a}} = \left(\mathbf{D}^{\mathsf{H}}\mathbf{D}\right)^{-1}\mathbf{D}^{\mathsf{H}}\mathbf{d}$, where $\mathbf{D}^{\mathsf{H}}$ stands for the complex-conjugate transpose of $\mathbf{D}$, known as Hermitian transform.

Computation of the modes frequency, damping factor, and damping ratio

Using the previously estimated coefficients, the roots of the characteristic polynomial expressed as

$$z^p - \left(a_1 z^{p-1} + a_2 z^{p-2} + ... + a_p z^0\right) = 0, \tag{6.10}$$

are calculated, where each root z_i, $i = 1, ..., p$, is a discrete time approximation of its respective continuous time eigenvalue in the $\mathcal{Z}$-domain. It can be shown that the polynomial roots, $\boldsymbol{z}$, are equal to the eigenvalues of the Companion matrix, $\boldsymbol{C}$, defined as

$$\boldsymbol{C} = \begin{bmatrix} a_1 & a_2 & ... & a_{p-1} & a_p \\ 1 & 0 & ... & 0 & 0 \\ 0 & 1 & ... & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & 1 & 0 \end{bmatrix}. \tag{6.11}$$

The i-th mode frequency is defined as:

$$f_i = \frac{\tan^{-1}\left(\Im\left\{z_i\right\}/\Re\left\{z_i\right\}\right)}{2\pi T}. \tag{6.12}$$

The i-th mode damping factor is defined as:

$$\alpha_i = \frac{\ln|z_i|}{T}. \tag{6.13}$$

Finally, the i-th mode damping ratio is defined as:

$$\xi_i = \frac{\alpha_i}{\sqrt{\alpha_i^2 + \left(2\pi f_i\right)^2}}. \tag{6.14}$$

Computation of the modes amplitude and initial phase

Using the polynomial roots, z_i, the linear regression model given by

$$\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & ... & 1 \\ z_1^1 & z_2^1 & ... & z_p^1 \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{N-1} & z_2^{N-1} & ... & z_p^{N-1} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}, \tag{6.15}$$

is constructed; in a compact form, we have $\mathbf{x} = \mathbf{V}\mathbf{h}$, where $\mathbf{V}$ is a Vandermonde matrix. Then, the parameter vector $\mathbf{h}$ is estimated in the least squares sense as $\hat{\mathbf{h}} = \left(\mathbf{V}^{\mathsf{H}}\mathbf{V}\right)^{-1}\mathbf{V}^{\mathsf{H}}\mathbf{x}$.

The i-th mode amplitude is defined as:

$$A_i = |h_i|. \tag{6.16}$$

The i-th mode initial phase is defined as:

$$\phi_i = \tan^{-1}\left(\Im\left\{h_i\right\} / \Re\left\{h_i\right\}\right). \tag{6.17}$$

6.2.2 Multichannel Prony method

In the multichannel case, the measurements vector $\mathbf{x} \in \mathbb{C}^N$ becomes a matrix $\mathbf{X} \in \mathbb{C}^{N\times m}$, where m refers to the number of channels. As proposed in [189], (6.9) is rewritten as

$$\begin{bmatrix} \boldsymbol{x}_p \\ \boldsymbol{x}_{p+1} \\ \vdots \\ \boldsymbol{x}_{N-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_{p-1} & \boldsymbol{x}_{p-2} & \dots & \boldsymbol{x}_0 \\ \boldsymbol{x}_p & \boldsymbol{x}_{p-1} & \dots & \boldsymbol{x}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_{N-2} & \boldsymbol{x}_{N-3} & \dots & \boldsymbol{x}_{N-p-1} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}, \tag{6.18}$$

where each element $\boldsymbol{x}_k$ is now a vector $[x_{1,k} \dots x_{m,k}]$ with m elements. The improved accuracy brought about by the use of multiple measurements is traded off by an increase by a factor m of the dimension of the problem in (6.9) while (6.10) remains exactly the same as in the single channel case. The vector of residues $\hat{\mathbf{h}}$ is estimated for each channel individually; thus (6.15) is computed m times.

For simplicity, we henceforth refer to the conventional Prony method as LS-Prony. Note that because it makes use of the least squares estimator, LS-Prony is not robust to outliers. In fact, a single measurement x_i can strongly bias the parameter estimates $\hat{\mathbf{a}}$ and $\hat{\mathbf{h}}$.

A flowchart of the R-Prony method, including a detailed view of the IPPC algorithm is presented in Fig. 6.2. Note that step **S3** is performed off-line and kept in memory for speediness of the algorithm. Also, observe that each measurement channel is individually filtered using IPPC, which exploits the signal time correlation, while in the last stage a multichannel Prony approach is adopted thus increasing the measurement redundancy and improving the estimation performance.

6.3 Numerical results

We use three different data sets to assess the performance of the proposed R-Prony method. The first one is generated through time domain simulations on the two-area four-machine test system. Similarly, the second data set relies on time domain simulations carried out on the WECC 179-bus system, and the data is publicly available [190]. The last data set

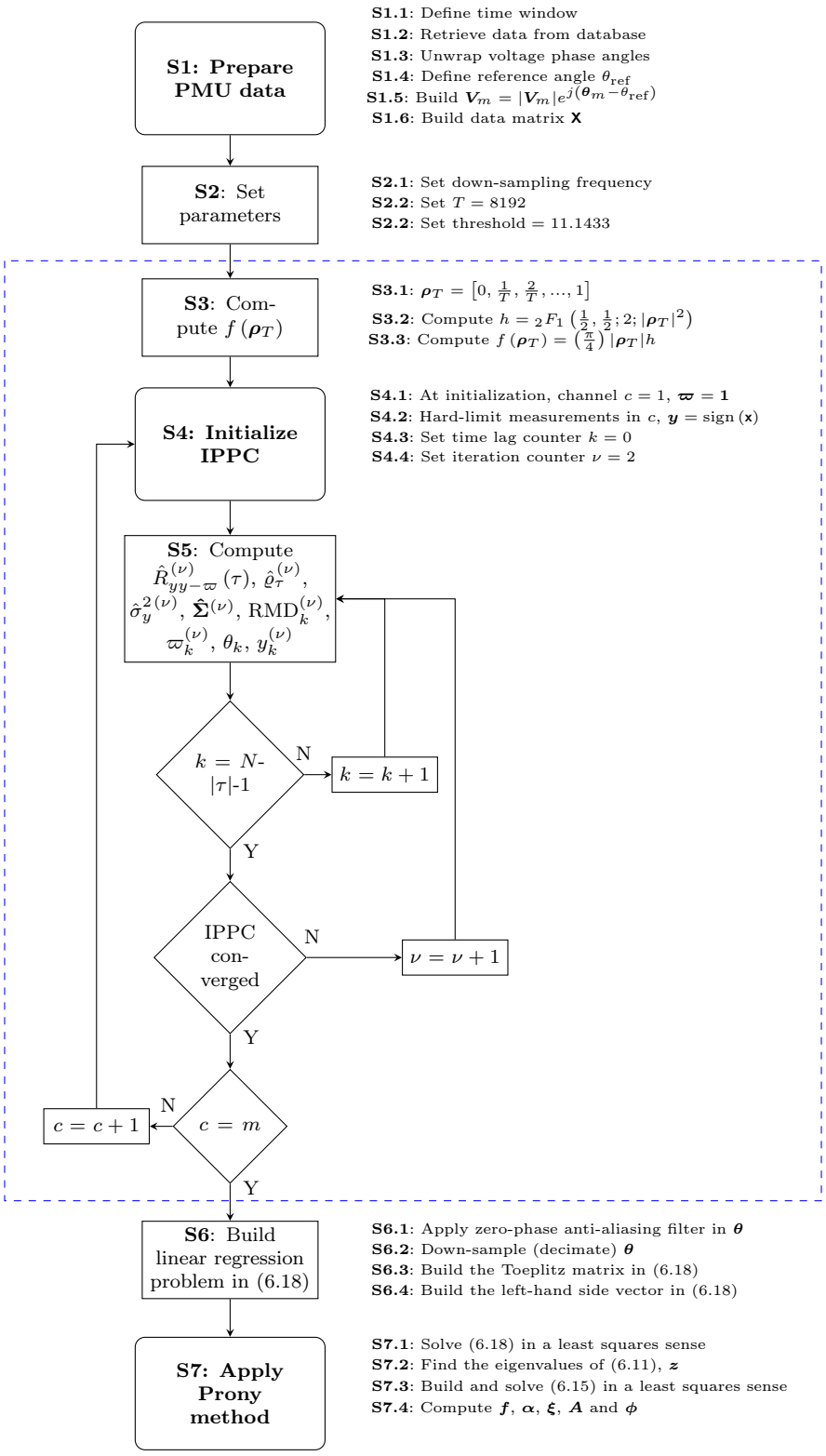


Figure 6.2: Flowchart of the R-Prony algorithm.

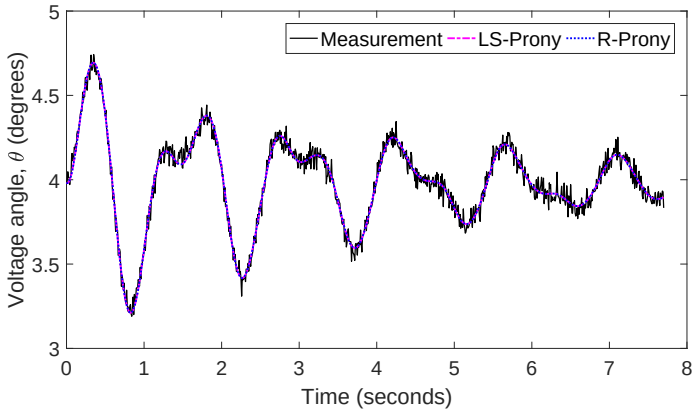


Figure 6.3: Comparison of the curve fitting results using LS-Prony and R-Prony for voltage angle at Bus 6 without impulsive noise.

comes from a low-voltage wide-area measurement system consisting of 24 PMUs installed nationwide in Brazilian university campuses [186].

Two-area four-machine test system

A fault is applied to Bus 3 and cleared after 80 milliseconds. Based on SSAT software [191], an inter-area mode with $f_1 = 0.7352$ Hz and $\xi_1 = 4.77\%$, and a local mode with $f_2 = 1.3602$ Hz and $\xi_2 = 2.58\%$, are found and assumed to be the true modes. A time window comprehended between 0.3 and 8.0 seconds after the fault clearing is selected as the data sample, to which a Gaussian noise is added making the signal-to-noise ratio (SNR) approximately equal to 40 dB. Recall that for a ringdown type of event, the SNR is typically high and 40 dB is considered a worst-case scenario. Bus voltage magnitudes, $|\boldsymbol{V}|$, and bus voltage angles, $\boldsymbol{\theta}$, are supposed to be measured by PMUs at all buses.

The voltage at Buses 4, 5 and 6 are selected to build the measurement matrix, $\mathbf{X}$. Table 6.2 shows the estimation results after applying LS-Prony and R-Prony to the data set without impulsive noise. From a practical standpoint the results are the same, with negligible numerical differences driven by noise. The curve fitting for Bus 6 is shown in Fig. 6.3, while results for Buses 4 and 5 are omitted since they do not bring additional information.

Table 6.2: Estimation of modes: $f_1 = 0.7352$ Hz, $\xi_1 = 4.77\%$, and $f_2 = 1.3602$ Hz, $\xi_2 = 2.58\%$

	LS-Prony	R-Prony
No impulsive noise	0.7233 Hz, 4.7889%	0.7235 Hz, 4.8566%
	1.3442 Hz, 2.4932%	1.3382 Hz, 2.3299%
With impulsive noise	**0.7060 Hz, 5.9394%**	0.7239 Hz, 4.8147%
	1.3929 Hz, 17.3674%	1.3518 Hz, 2.3276%

As illustrated in Fig. 6.4, a single outlier is now placed on each bus voltage angle measurement array. Fig. 6.5 shows Bus 6 voltage angle measurements, with focus on the outlier and

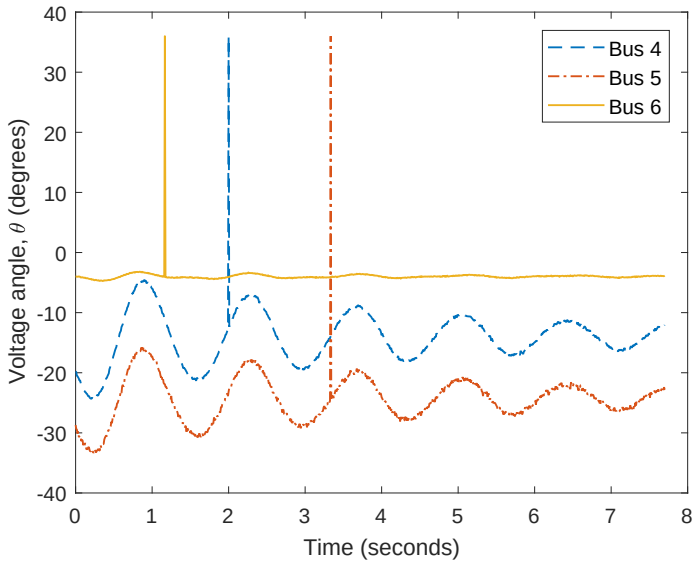


Figure 6.4: Voltage angle at Buses 4, 5 and 6 with an impulsive noise at $t = 2.00$, 3.33 and 1.17 seconds, respectively.

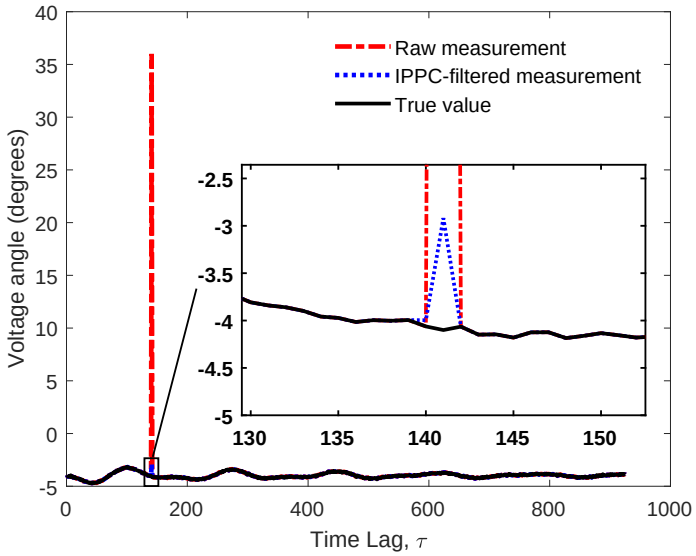


Figure 6.5: Voltage phase angle at Bus 6. The true signal without outlier is presented for comparison.

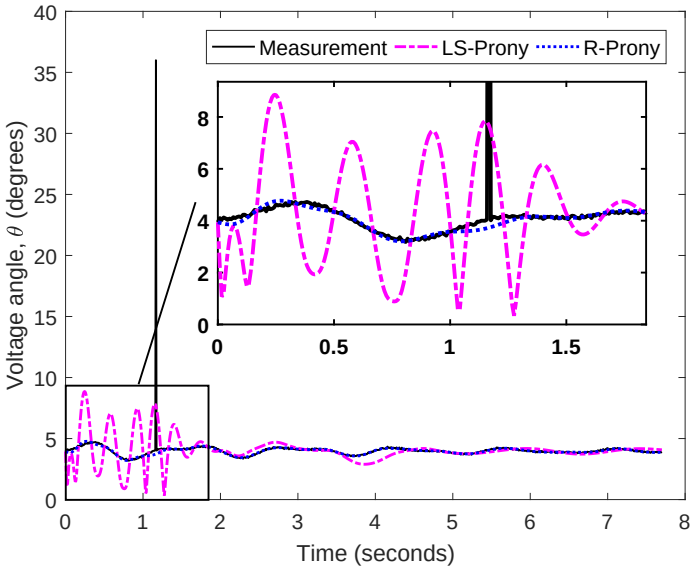


Figure 6.6: Comparison of the curve fitting results using LS-Prony and R-Prony for voltage angle at Bus 6 with impulsive noise.

how it is suppressed after an iteration of the IPPC algorithm. The signal without impulsive noise is included for comparison. The numerical results are also presented in Table 6.2. As one can see, LS-Prony results are strongly biased, specially in the damping ratio parameter. In fact, it is well-known in the literature that the estimation of the damping ratio is more challenging than the frequency estimation.

The signal reconstruction of Bus 6 voltage angle is presented in Fig. 6.6, while results for Buses 4 and 5 are omitted as before. The curve fitting provided by R-Prony indicates that the method is slightly affected by impulsive noises, i.e., its influence in the estimation process is bounded, which is clearly not the case for the LS-Prony method. Recall that a proper curve fitting along with good estimates of frequency and damping ratio necessarily imply good estimates of the residues. In conclusion, Fig. 6.6 not only illustrates that the R-Prony robustly estimates the frequency and the damping ratio, but also the amplitude and the initial phase of the exponential model.

WECC 179-bus system

In the test cases library [190], the third case is picked, which shows poorly damped natural electromechanical oscillations (ND3). For this case, there are two poorly damped inter-area modes, namely $f_1 = 0.46$ Hz, $\xi_1 = 2.22\%$ and $f_2 = 0.70$ Hz, $\xi_2 = 1.15\%$, and one unstable local mode, $f_3 = 1.63$ Hz, $\xi_3 = -0.54\%$. The complex voltage at Buses 33–35, 65–66, and 74 obtained from the time domain simulations are selected to build the measurement matrix, because they provide the highest possible observability for Mode 1. From the 40 seconds of available simulation data, a time window consisting of the last 20 seconds is chosen. As before, Gaussian noise is added to the time domain simulations making the

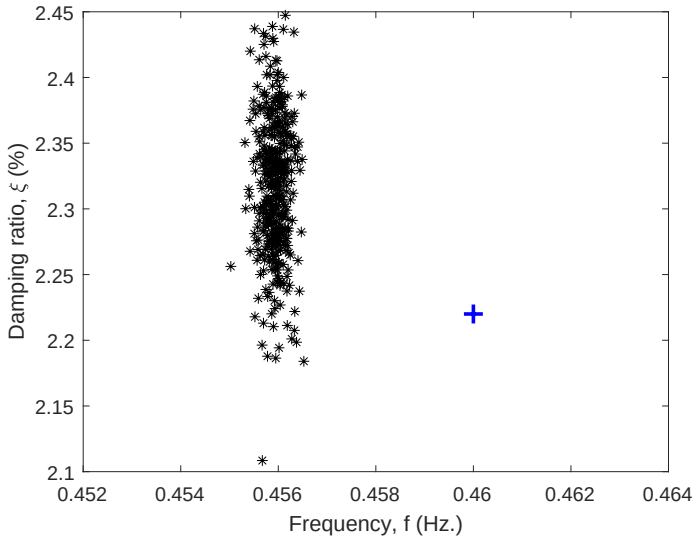


Figure 6.7: 400 estimates of the inter-area mode $f_1 = 0.46$ Hz, $\xi_1 = 2.22\%$, by Monte Carlo simulations. Case without impulsive noise, R-Prony method.

Table 6.3: Modal parameters estimation results, case without impulsive noise in the data set

	$E[f_1]$ (Hz)	$E[\xi_1]$ (%)	Number of Failures	Average CPU Time (seconds)
LS-Prony	0.4560	2.3156	0/400	0.1423
R-Prony	0.4559	2.3171	0/400	1.6736

SNR approximately equal to 40 dB. To verify the robustness of the R-Prony method for estimating Mode 1 from the given data set, Monte Carlo simulations are carried out 400 times; the results are presented in Fig. 6.7. We observe that there is no impulsive noise up to that point. The results of the LS-Prony method exhibit very similar dispersion from the true mode as the R-Prony method does, and therefore they are omitted. The numerical results for both the LS-Prony and the R-Prony are displayed in Table 6.3, where *Number of Failures* stands for how many times Mode 1 was not identified. Here, the higher average CPU time reveals the price that needs to be paid to gain robustness while there is no relevant difference between the results obtained from LS-Prony and the R-Prony.

The case with impulsive noise is now considered. One single outlier is injected in the voltage angle measurement referent to Bus 34. Considering the time window length of 20 seconds, the outlier is placed at $t = 8.3666$ seconds; this time position is kept constant throughout the Monte Carlo simulations. The magnitude of the impulsive noise, however, randomly varies between the interval $[-\pi, \pi]$.

Table 6.4 displays the numerical results. It is observed that the LS-Prony failed 106 times out of 400 Monte Carlo runs, and an average of the estimated frequency and damping ratio, for the remaining 294 runs, is presented as $E[f_1]$ and $E[\xi_1]$ respectively. The estimation of

Table 6.4: Modal parameters estimation results, case with random impulsive noise in the data set

	$E[f_1]$ (Hz)	$E[\xi_1]$ (%)	Number of Failures	Average CPU Time (seconds)
LS-Prony	0.4622	1.3525	106/400	0.1449
R-Prony	0.4565	2.8514	0/400	1.7013

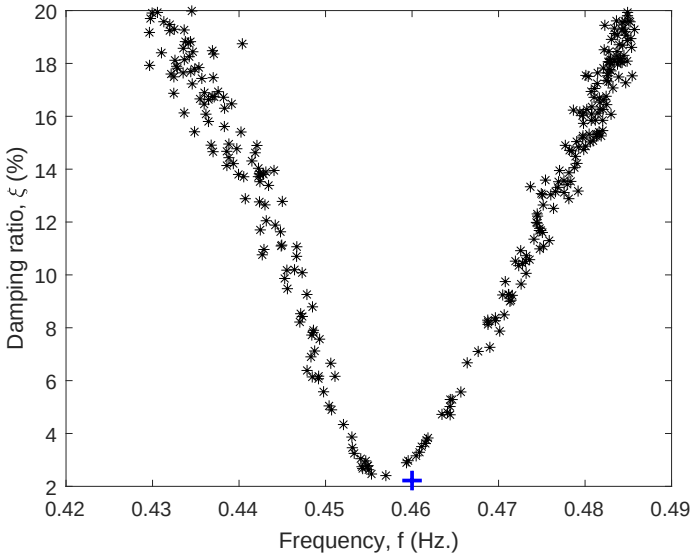


Figure 6.8: 400 estimates of the inter-area mode $f_1 = 0.46$ Hz, $\xi_1 = 2.22\%$, by Monte Carlo simulations. Case with impulsive noise, LS-Prony method.

the damping ratio is strongly biased, making it irrelevant as shown in Fig. 6.8.

The results obtained with R-Prony are depicted in Fig. 6.9, and also presented in Table 6.4. The method provides good results, but the average of the damping ratio estimates is slightly biased as compared to the results shown in Table 6.3. In fact, one can observe from Fig. 6.9 that the estimation of the damping ratio was far away from the true value in several occasions. Those few cases occurred because the magnitude of the impulsive noise was moderate, and thus the weights assigned by the IPPC algorithm were not small enough. This is expected to happen since the algorithm relies on some sort of Mahalanobis distance to identify and downweight outliers; measurements close to the boundaries of the confidence ellipsoid represents a challenge to the method.

Real PMU measurements from a perturbation occurred in the Brazilian interconnected transmission system

According to Brazilian power system operator [192], a momentary load curtailment of 3,200 MW occurred on October 26, 2016, caused by a voltage variation due to an equipment damage in the 345 kV sector of a substation located in the state of São Paulo. The total load

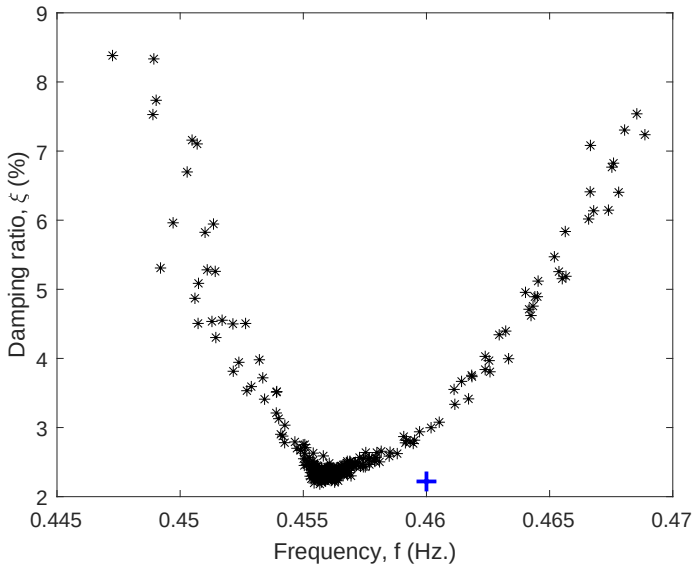


Figure 6.9: 400 estimates of the inter-area mode $f_1 = 0.46$ Hz, $\xi_1 = 2.22\%$, by Monte Carlo simulations. Case with impulsive noise, R-Prony method.

in the National interconnected transmission system was 74,000 MW. Fig. 6.10 shows the voltage angle measured in the Northern (UFAM, UFMA and UFPA) and Southern (UFRGS, UFSC, UNIPAMPA) regions. The voltage angle in the capital, Brasília, is taken as reference. The frequency of the North-South mode ranges between 0.20–0.40 Hz [186], depending on the system operating point. The estimates obtained with LS-Prony and R-Prony are presented in Table 6.5. Despite the fact that it is not possible to assess the estimation accuracy without having access to the system model, the estimated damping ratio is in agreement with operational practice in bulk interconnected transmission systems.

One can see from Fig. 6.10 that impulsive noise commonly occurs in real PMU data. Besides the aforementioned causes, impulsive noise might occur due to switching. In particular during large system disturbances, they might result from the action of special protection schemes.

Table 6.5: Estimation of Brazilian north-south mode

	LS-Prony	R-Prony
No impulsive noise	0.3027 Hz, 10.6872%	0.3016 Hz, 11.0612%

6.4 Concluding remarks

Recently, Susuki and Mezić [193] showed that the multichannel Prony method provides a finite approximation to the Koopman modes. This is an important result because it casts the Prony method on a solid theoretical ground. This is an interesting result because, since

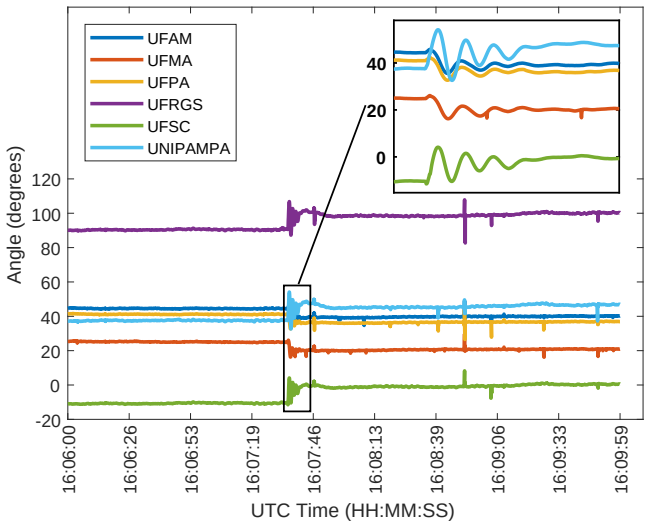


Figure 6.10: Voltage angle (phase A) measured by PMUs installed in the Northern and Southern regions in Brazil. The voltage angle in the capital, Brasília, is taken as reference.

Koopman mode decomposition can reveal both linear and nonlinear modes of oscillation, it allows us to carry out either small-signal or transient stability analysis without resorting to any model and without assuming any linearization [64].

As for the developed robust filter presented in this chapter, we foresee that it can be also applied to event detection based on PMU data, short-term load forecasting and forecast-aided state estimation, only to mention few examples. Also, it would be interesting to see a more comprehensive study of this filter from the perspective of signal processing.

Chapter 7

Modal participation factors for nonlinear systems

The participation factors are an important component of the so-called selective modal analysis, which is a comprehensive methodology for the modeling, analysis, and control of dynamical systems described by large-scale linear autonomous models, proposed by Pérez-Arriaga et al. [194, 195]. The participation factors are widely used in the power industry as they provide a measure of the relative contribution of modes to system states and vice versa. Applications include stability analysis [195–197], identification of coherent synchronous generators [198], dynamic model reduction [54], and placement of power system stabilizers [199]. Alternative [200, 201] and complementary [202] perspectives to the originally proposed linear participation factors also occur in the literature. For instance, Abed et al. [200, 201] advocate that the initial states are uncertain, thereby casting the problem of computing the participation factors as a stochastic one, as opposed to deterministic; then, by relying on the definition of the mathematical expectation and by assuming that the initial states follow a uniform probability distribution, a dichotomy between mode-in-state and state-in-mode participation factors is suggested. Despite the existence of different views, the participation factors are a well-accepted metric of the dynamic performance of linear systems; therefore, its extension to nonlinear models is of practical interest since it is known that the analysis of power systems through model linearization does not provide an accurate picture of the modal characteristics when the system is operating under stressed conditions [203].

An attempt to go beyond the linear paradigm was made by Lesieutre et al. [204] by applying a transformation from state variables to harmonic variables to gain insight about the state-in-mode participation factors at the Hopf bifurcation point. Although ingenious, their approach ultimately computed the participation factors of a transformed linear model associated with the stable limit cycle. A different approach was taken by Vittal et al. [203], with the intent of studying inter-area modes of oscillation in stressed power systems following large disturbances, when nonlinearities play an important role; their idea was to compute

participation factors by considering up to second-order terms in the Taylor series expansion of the system nonlinear model and then applying the method of Normal forms. The inclusion of third-order terms has also been exploited [205, 206]. Despite of the importance of this line of research [207–212], the shortcomings of the method of Normal forms are well-documented [213], and include a heavy computational burden that could make it prohibitive for large-scale systems even if second-order terms are considered only, and a highly nonlinear numerical problem that needs to be solved to retrieve the initial conditions. An alternative approach to the method of Normal forms based on the modal series method was proposed by Pariz et al. [214, 215], with the advantages of being valid under resonance conditions and not requiring nonlinear transformations. However, their approach is also restricted to polynomial nonlinearities, as is the case with the method of Normal forms. Furthermore, all the aforementioned approaches lack the state-in-mode participation factors.

In face of the exposed challenges, it has been suggested [213, 216] that the computation of the participation factors from measurements could either provide a solution to the aforementioned issues directly, or be complementary to model-based techniques such as the ones that rely on the Taylor series expansion of the power systems nonlinear model. The problem thus becomes one of estimating linear and *nonlinear* participation factors from measurements. In the present work, we propose to tackle this problem by relying on the Koopman operator-theoretic framework [217]. Recently, after the work of Mezić et al. [91–93], this approach based on the point spectrum of the Koopman operator, and henceforth referred to as the KMD, has gained momentum as a powerful data-driven tool to analyze nonlinear dynamical systems. The interested reader is pointed to [218] for a comprehensive review of KMD applied to power systems.

To approximate the Koopman eigenvalues, eigenfunctions, and modes, we adopt a slightly different formulation of the extended dynamic mode decomposition (EDMD) that has been proposed by Klus et al. [113], which we found suitable for the ongoing purpose. Then, we demonstrate how to compute linear and nonlinear participation factors from measurements. To the authors' best of knowledge, this is the first comprehensive application of the EDMD algorithm in power systems. Sako et al. [219] adopted the original algorithm as proposed by Williams et al. [112]. Netto and Mili [69, 70] adopted the algorithm proposed by Klus et al. [113], but did not consider nonlinear observables, which conversely are accounted for in the present work, thereby exploring the full potential of KMD. The nonlinear participation factors computed by the proposed approach are not restricted to any kind of nonlinearity, and both mode-in-state and state-in-mode nonlinear participation factors can be estimated. Provided that certain conditions are satisfied, it is shown that our approach generalizes the one proposed by Pérez-Arriaga et al. [194] to nonlinear dynamical systems, in the case of mode-in-state participation factors.

The paper proceeds as follows. Section II briefly revisits the formulation of the linear participation factors, and the nonlinear participation factors based on the method of Normal forms. Section III introduces the proposed data-driven technique to compute linear and nonlinear participation factors based on the KMD. Section IV discusses some interesting numerical

results. Conclusions and ongoing work are provided in Section V.

7.1 Preliminaries

Consider a power system dynamic model as a continuous time autonomous nonlinear system defined on n-dimensional Euclidean space $\mathbb{R}^n$ as follows:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}), \tag{7.1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the system state vector, and $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ is a vector-valued nonlinear function.

7.1.1 Participation factors for nonlinear systems based on the method of normal forms

By performing a Taylor series expansion of (7.1) around a s.e.p. and neglecting the third- and higher order terms, we get

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{o}_2, \tag{7.2}$$

where $\boldsymbol{o}_2$ denotes exclusively the second-order terms. The i-th state variable in (7.2) is given by

$$\dot{x}_i = \mathbf{a}_i^\top \boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{H}^{(i)} \boldsymbol{x}, \tag{7.3}$$

where $\boldsymbol{H}^{(i)} \in \mathbb{R}^{n\times n}$ is an Hessian matrix, $\boldsymbol{H}^{(i)} = \nabla^2 f_i(\boldsymbol{x}) = \frac{\partial^2 f_i(\boldsymbol{x})}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top}$, and $\mathbf{a}_i^\top$ stands for the i-th row of $\boldsymbol{A}$. Now, define the similarity transformation

$$\boldsymbol{y} = \boldsymbol{V}\boldsymbol{x}. \tag{7.4}$$

By substituting (7.4) into (7.2) and developing, we have

$$\dot{\boldsymbol{y}} = \boldsymbol{\Lambda}\boldsymbol{y} + \frac{1}{2}\boldsymbol{V} \begin{bmatrix} \boldsymbol{y}^\top \boldsymbol{U}^\top \boldsymbol{H}^{(1)} \boldsymbol{U}\boldsymbol{y} \\ \boldsymbol{y}^\top \boldsymbol{U}^\top \boldsymbol{H}^{(2)} \boldsymbol{U}\boldsymbol{y} \\ \vdots \\ \boldsymbol{y}^\top \boldsymbol{U}^\top \boldsymbol{H}^{(n)} \boldsymbol{U}\boldsymbol{y} \end{bmatrix}, \tag{7.5}$$

where the i-th variable in (7.5) is given by

$$\dot{y}_i = \lambda_i y_i + \sum_{j=1}^{n}\sum_{r=1}^{n} C_{jr}^{(i)} y_j y_r, \tag{7.6}$$

$\boldsymbol{C}^{(i)} = \frac{1}{2}\sum_{j=1}^{n} v_{ij}\left(\boldsymbol{U}^{\top}\boldsymbol{H}^{(j)}\boldsymbol{U}\right)$, and v_{ij} is the ij-th element of $\boldsymbol{V}$. Equation (7.5) is transformed into the Normal form as follows:

$$\boldsymbol{y} = \boldsymbol{z} + \alpha\left(\boldsymbol{z}\right), \tag{7.7}$$

where $\boldsymbol{z}$ is the vector of Normal form variables, and (7.7) can be expanded as

$$y_i = z_i + \sum_{j=1}^{n}\sum_{r=1}^{n} h_{jr}^{(i)} z_j z_r, \tag{7.8}$$

where $h_{jr}^{(i)} = \frac{C_{jr}^{(i)}}{\lambda_j+\lambda_r-\lambda_i}$, with $\lambda_j + \lambda_r - \lambda_i \neq 0$ and $i, j, r = 1, ..., n$. Finally, we have

$$\dot{\boldsymbol{z}} = \boldsymbol{\Lambda}\boldsymbol{z} + \boldsymbol{o}\left(3\right), \tag{7.9}$$

where $\boldsymbol{o}\left(3\right)$ denotes the neglected third- and higher order terms, and the second-order terms have been annihilated by (7.7). The solution to the original system (7.2) is given by

$$x_i\left(t\right) = \sum_{j=1}^{n} u_{ij} z_{j0} e^{\lambda_j t} + \sum_{j=1}^{n} u_{ij}\left[\sum_{r=1}^{n}\sum_{s=1}^{n} h_{rs}^{(j)} z_{r0} z_{s0} e^{(\lambda_r+\lambda_s)t}\right]. \tag{7.10}$$

For more details on the method of Normal forms, please refer to [220]. Apart from the linear contribution factors defined by (C.25), the nonlinear contribution factors in (7.10) are defined in [205] as follows:

$$\sigma_{2ij} = u_{ij} z_{j0}, \tag{7.11}$$

$$\sigma_{22rs}^{(i)} = z_{r0} z_{s0} \sum_{j=1}^{n} u_{ij} h_{rs}^{(j)}. \tag{7.12}$$

Similarly, apart from the linear participation factors defined in (C.26), the nonlinear participation factors derived from the method of Normal forms are given by

$$p_{2ij} = u_{ij}\left(v_{ji} + v_{2ji}^{(i)}\right), \tag{7.13}$$

$$p_{2rs}^{(i)} = r_{2rs}^{(i)}\left(v_{ri} + v_{2ri}^{(i)}\right)\left(v_{si} + v_{2si}^{(i)}\right), \tag{7.14}$$

$v_{2ji}^{(i)} = -\sum_{r=1}^{n}\sum_{s=r}^{n} h_{rs}^{(j)} v_{ri} v_{si}$ and $r_{2rs}^{(i)} = \sum_{j=1}^{n} u_{ij} h_{rs}^{(j)}$.

The method of Normal forms was introduced by Poincaré. Following [217], "a majority of methods from dynamical system analysis rely on Poincaré's geometric picture that focuses on dynamics of states. While this picture has fueled the field for a century, it has shown difficulties in handling high-dimensional, ill-described, and uncertain systems." The Koopman operator-based approach, presented in the next section, has emerged as an alternative framework based on the "dynamics of observables" picture.

7.2 Participation factors for nonlinear systems based on the Koopman operator-theoretic framework

In the previous Section, it is assumed that the reader is familiar with the continuous time formulation; this is the reason why we present the equations in that way. In the linear case (C.14), if a discrete time autonomous system of the form $\boldsymbol{x}_k = \boldsymbol{A}\,\boldsymbol{x}_{k-1}$ is assumed, we can show that

$$\boldsymbol{x}_k = \sum_{j=1}^{n} \left(\boldsymbol{v}_j^\top \boldsymbol{x}_0\right) \boldsymbol{u}_j \mu_j^k. \tag{7.15}$$

We adopt a discrete time formulation to introduce the Koopman operator-approach, motivated by the fact that the EDMD technique [112] is data-driven. Therefore, in what follows, consider a discrete time autonomous nonlinear system given by

$$\boldsymbol{x}_k = \boldsymbol{T}\left(\boldsymbol{x}_{k-1}\right), \tag{7.16}$$

where $\boldsymbol{x} \in S$, and $\boldsymbol{T} : S \to S$ is a vector-valued nonlinear function. The Koopman operator is a linear operator $\mathcal{K}$ that acts on scalar-valued functions defined on S in the following manner:

$$\mathcal{K}g\left(\boldsymbol{x}_k\right) = g\left(\boldsymbol{T}\left(\boldsymbol{x}_{k-1}\right)\right), \tag{7.17}$$

and $g : S \to \mathbb{R}$. The eigenvalues, μ_j, and eigenfunctions, ϕ_j, of $\mathcal{K}$ are defined as follows:

$$\mathcal{K}\phi_j\left(\boldsymbol{x}_k\right) = \mu_j \phi_j\left(\boldsymbol{x}_k\right), \quad j = 1, 2, ... \tag{7.18}$$

The set of all Koopman eigenvalues, μ_j, is called the point spectrum of the Koopman operator. Now, consider a vector-valued observable $\boldsymbol{g} : S \to \mathbb{C}^q$. If all the elements of $\boldsymbol{g}$ lie within the span of the Koopman eigenfunctions, ϕ_j, we have

$$\boldsymbol{g}\left(\boldsymbol{x}_k\right) = \sum_{j=1}^{\infty} \phi_j\left(\boldsymbol{x}_k\right) \boldsymbol{v}_j = \sum_{j=1}^{\infty} \phi_j\left(\boldsymbol{x}_0\right) \boldsymbol{v}_j \mu_j^k, \tag{7.19}$$

where $\boldsymbol{v}_j$ are the Koopman modes [217], and $(\mu_j, \phi_j, \boldsymbol{v}_j)$ are referred to as the Koopman tuples. As stated by Susuki and Mezić [64], "the real part of $\phi_j\left(\boldsymbol{x}_0\right)\boldsymbol{v}_j$ determines the initial amplitude of modal dynamics", and are in fact a nonlinear generalization of the linear contribution factors based on the Koopman operator, as defined before.

There are several algorithms in the literature that are able to approximate subsets of the Koopman tuples. The generalized Laplace analysis [64] can approximate the Koopman modes and eigenfunctions, but it requires *a priori* knowledge of the eigenvalues. The Arnoldi-like [64] and the DMD [94] algorithms provide approximations to the Koopman modes and eigenvalues, but not the eigenfunctions. Thus, these algorithms are not suitable to the present work. By contrast, the EDMD algorithm is able to approximate the Koopman modes, eigenvalues and eigenfunctions [112]. In what follows, we adopt the EDMD algorithm as proposed by Klus et al. [113], which we present next.

7.2.1 Mode-in-state participation factors for nonlinear systems

By making $\boldsymbol{x} = \boldsymbol{x}_0$ in (2.55) and plugging it into (2.58), the i-th state variable reduces to

$$x_{ik} = \sum_{j=1}^{q} \left(\boldsymbol{\xi}_j^\top \boldsymbol{\gamma}_0\right) \upsilon_{ij} \mu_j^k, \tag{7.20}$$

where $\boldsymbol{\gamma}_0 = \left[\gamma_1\left(\boldsymbol{x}_0\right); ...; \gamma_q\left(\boldsymbol{x}_0\right)\right]^\top$. Note that the expansion in (7.20) depends on the vector of observable functions of the initial state, $\boldsymbol{\gamma}_0$, as opposed to the linear case where the expansion is function of the initial state vector, $\boldsymbol{x}_0$, itself. As in [201], we view $\boldsymbol{\gamma}_0$ as uncertain and approach the problem of measuring modal participation by averaging the relative contribution over an uncertain set. Formally, we have

$$p_{ij} = E\left[\frac{\left(\boldsymbol{\xi}_j^\top \boldsymbol{\gamma}_0\right) \upsilon_{ij}}{\gamma_{i0}}\right], \tag{7.21}$$

where $\gamma_{i0} = \sum_{j=1}^{q} \left(\boldsymbol{\xi}_j^\top \boldsymbol{\gamma}_0\right) \upsilon_{ij}$ is the value of $x_{ik} = \gamma_i\left(\boldsymbol{x}_k\right)$ at $k = 0$.

Expanding (7.21), we have

$$\begin{aligned} p_{ij} &= E\left[\sum_{r=1}^{q} \frac{\left(\xi_{rj}\gamma_{r0}\right)\upsilon_{ij}}{\gamma_{i0}}\right] \\ &= E\left[\frac{\left(\xi_{ij}\gamma_{i0}\right)\upsilon_{ij}}{\gamma_{i0}}\right] + E\left[\sum_{r=1, r\neq i}^{q} \frac{\left(\xi_{rj}\gamma_{r0}\right)\upsilon_{ij}}{\gamma_{i0}}\right] \\ &= \xi_{ij}\upsilon_{ij} + \sum_{r=1, r\neq i}^{q} \xi_{rj}\upsilon_{ij} E\left[\frac{\gamma_{r0}}{\gamma_{i0}}\right]. \end{aligned} \tag{7.22}$$

By assuming that the observables $\gamma_1\left(\boldsymbol{x}_0\right), ..., \gamma_q\left(\boldsymbol{x}_0\right)$ are independent with zero mean, the second term in (7.22) vanishes and the KMD-based mode-in-state participation factors are given by

$$p_{ij} = \xi_{ij}\upsilon_{ij}, \tag{7.23}$$

where $i = 1, ..., n$ and $j = 1, ..., q$. Note that as opposed to the linear case, $q \geq n$ and the matrix of mode-in-state participation factors is in general not square.

7.2.2 State-in-mode participation factors for nonlinear systems

Suppose that $\boldsymbol{\gamma}\left(\boldsymbol{x}_{k-1}\right) = \widetilde{\boldsymbol{x}}_{k-1}$ and $\boldsymbol{\gamma}\left(\boldsymbol{y}_{k-1}\right) = \widetilde{\boldsymbol{x}}_k$. By making use of (2.54), we have

$$\widetilde{\boldsymbol{x}}_k = \boldsymbol{U}\widetilde{\boldsymbol{x}}_{k-1}. \tag{7.24}$$

Now, define a transformation

$$\boldsymbol{z}_k = \boldsymbol{\Xi}\widetilde{\boldsymbol{x}}_k. \tag{7.25}$$

Substituting (7.25) into (7.24), we have

$$\boldsymbol{z}_k = \boldsymbol{\Xi}\boldsymbol{U}\boldsymbol{\Xi}^{-1}\boldsymbol{z}_{k-1} = \boldsymbol{\Omega}\boldsymbol{z}_{k-1}, \tag{7.26}$$

where $\boldsymbol{\Omega} = \text{diag}\left(\mu_1, ..., \mu_q\right)$. Following Hashlamoun et al. [201], we assume that the observables $\gamma_1\left(\boldsymbol{x}_0\right), ..., \gamma_q\left(\boldsymbol{x}_0\right)$ are jointly uniformly distributed over the unit sphere in $\mathbb{R}^q$ centered at the origin. Due to space limitation, we present the final result directly; the derivation steps follow [201]. The state-in-mode participation factors based on KMD become

$$\pi_{ij} = \frac{\left(\Re\left\{\xi_{ij}\right\}\right)^2}{\left(\Re\left\{\boldsymbol{\xi}_j\right\}\right)^\top \Re\left\{\boldsymbol{\xi}_j\right\}}, \tag{7.27}$$

where $\Re\left\{\xi\right\}$ denotes the real part of ξ. An example using a canonical nonlinear dynamical system follows to elucidate the important points.

7.2.3 Example: A canonical nonlinear dynamical system

Consider the autonomous nonlinear dynamical system expressed as

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}\lambda_1\left(x_1 - x_2^2\right)\\ \lambda_2 x_2\end{bmatrix}, \tag{7.28}$$

with $\lambda_1 = -1$, $\lambda_2 = -0.05$, and $\boldsymbol{x}_0 = [-1; 2]^\top$. This system (7.28) has been studied by Brunton et al. in [172]. By selecting $w_1 = x_1$, $w_2 = x_2$, and $w_3 = x_2^2$, we have

$$\begin{aligned}
\dot{w}_1 &= \dot{x}_1 = \lambda_1\left(x_1 - x_2^2\right) = \lambda_1 w_1 - \lambda_1 w_3\\
\dot{w}_2 &= \dot{x}_2 = \lambda_2 x_2 = \lambda_2 w_2,\\
\dot{w}_3 &= 2x_2\dot{x}_2 = 2\lambda_2 x_2^2 = 2\lambda_2 w_3,
\end{aligned} \tag{7.29}$$

and

$$\begin{bmatrix}\dot{w}_1\\ \dot{w}_2\\ \dot{w}_3\end{bmatrix} = \begin{bmatrix}\lambda_1 & 0 & -\lambda_1\\ 0 & \lambda_2 & 0\\ 0 & 0 & 2\lambda_2\end{bmatrix}\begin{bmatrix}w_1\\ w_2\\ w_3\end{bmatrix}, \tag{7.30}$$

where $[w_1; w_2; w_3]^\top$ is the vector of observable functions. Note that through this particular choice of observable functions, the two-dimensional nonlinear system (7.28) is transformed into a three-dimensional linear system (7.30) without any linearization. Although such a finite-dimensional transformation only exists for certain classes of nonlinear dynamical systems [172], this example elucidates the key idea of the Koopman operator-theoretic approach.

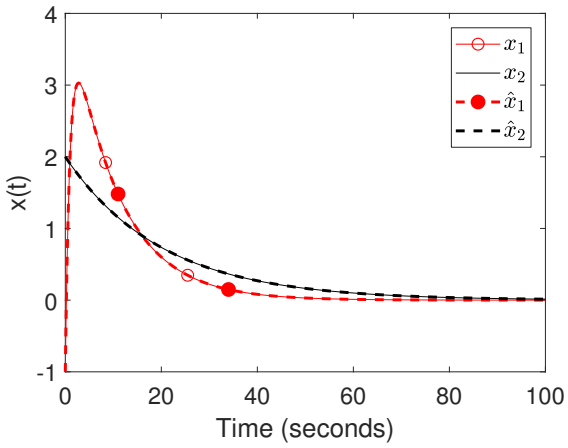


Figure 7.1: Reconstruction of the state variables using the estimated Koopman tuples $(\mu_j, \phi_j, \boldsymbol{v}_j)$, $j = 1, ..., q$.

We integrate (7.28) numerically with a time step of 0.01 seconds and use the results to build (2.51). Furthermore, we select $\boldsymbol{\gamma} = [x_1;\, x_2;\, x_2^2]^\top$ to build (2.53), with $\boldsymbol{\gamma}_0 = [-1;\, 2;\, 4]^\top$. By applying the EDMD, we get $\boldsymbol{\mu} = [-1;\, -0.05;\, -0.1]^\top$,

$$\boldsymbol{\Xi} = \begin{bmatrix} -1 & 0 & 1.1111 \\ 0 & 1 & 0 \\ 0 & 0 & 1.4948 \end{bmatrix}, \text{ and } \quad \boldsymbol{\Phi} = \begin{bmatrix} 1 & 0 & 0.7433 \\ 0 & 1 & 0 \end{bmatrix}. \tag{7.31}$$

We reconstruct the states evolution using (7.20) and verify that the estimation of the Koopman tuples via the EDMD algorithm is satisfactory, as presented in Fig. 7.1.

Then, from (7.23), we have that the

$$\text{mode-in-state participation factors} = \begin{bmatrix} 1 & 0 & 0.8259 \\ 0 & 1 & 0 \end{bmatrix}. \tag{7.32}$$

Note that the sum of the values of all the entries of a single row or column in (7.32) is not necessarily equal to 1. However, if only the first two columns in (7.32) related to the linear modes λ_1 and λ_2 are taken into consideration, the proposed matrix of linear and nonlinear participation factors reduces to the one proposed by Pérez-Arriaga et al. [194]. It is important to mention that the choice of the vector of observables (2.52) plays a key role. Furthermore, the nonlinear participation factors are not necessarily restricted to the unit interval; in fact, if for instance we choose $\lambda_1 = -1$ and $\lambda_2 = -0.4$, we get

$$\text{mode-in-state participation factors} = \begin{bmatrix} 1 & 0 & 4.9029 \\ 0 & 1 & 0 \end{bmatrix}, \tag{7.33}$$

and the participation factor associated with the nonlinear mode $2\lambda_2$ is equal to $4.9029 > 1$.

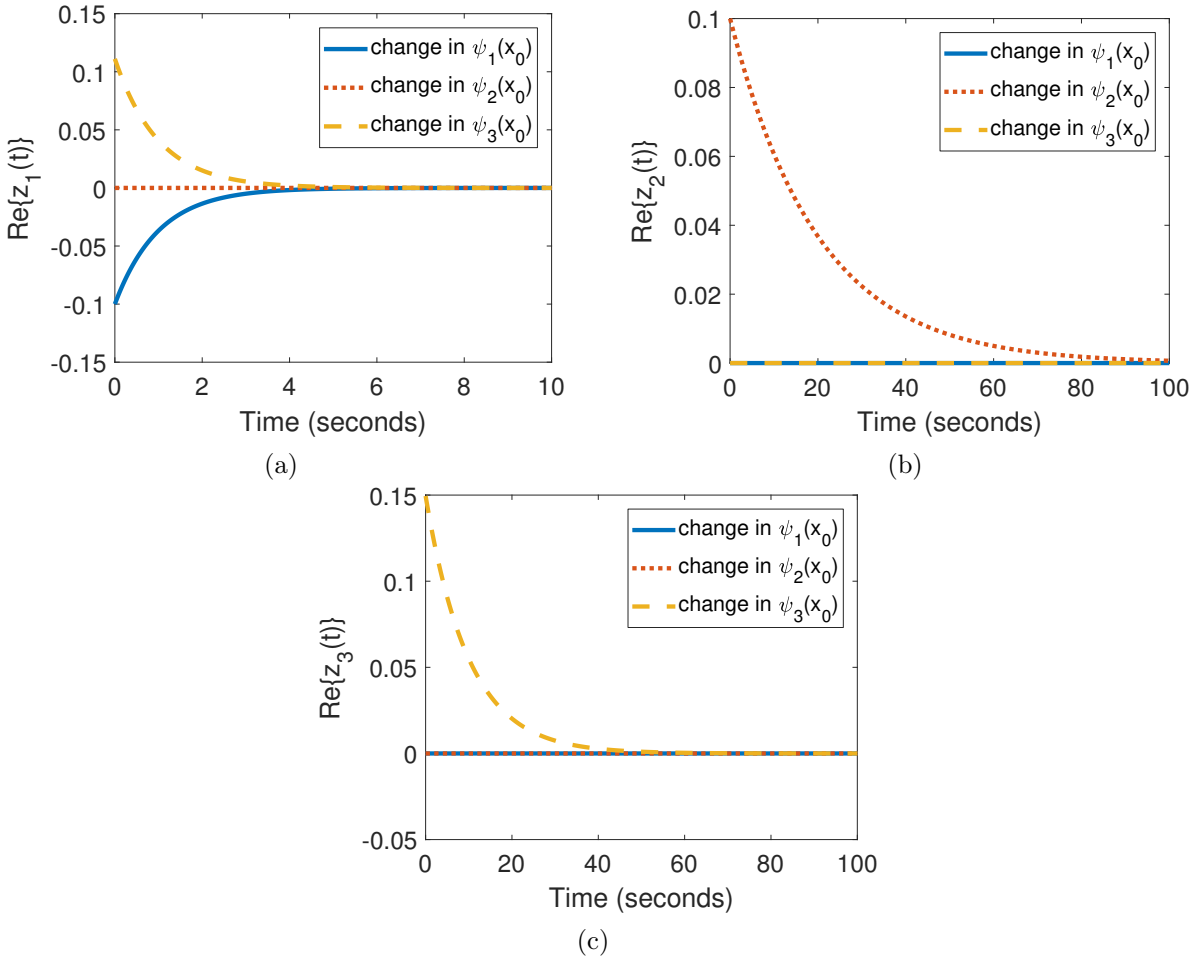


Figure 7.2: Evolution of the modal variables z_i, $i = 1, ..., q$, for the nonlinear dynamical system (7.28). (a) Evolution of mode 1, $z_1(t)$. (b) Evolution of mode 2, $z_2(t)$. (c) Evolution of mode 3, $z_3(t)$.

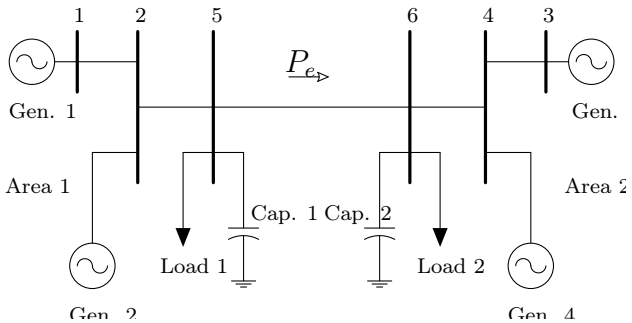


Figure 7.3: Two-Area Four-Machine System.

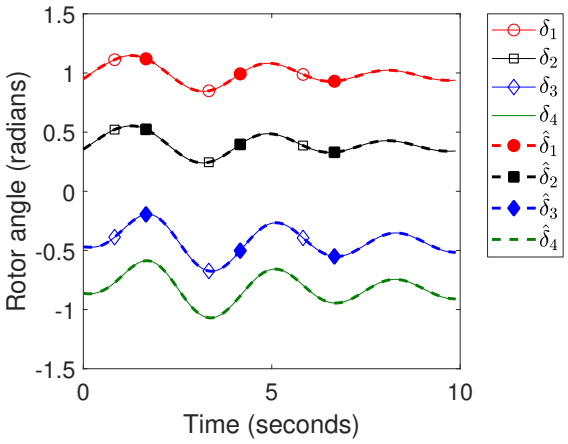


Figure 7.4: Reconstruction of the evolution of the generator's rotor angle using the estimated Koopman tuples $(\mu_j, \phi_j, \boldsymbol{v}_j)$, $j = 1, ..., q$.

Similarly, after computing the linear and nonlinear state-in-mode participation factors (7.27), we have

$$\boldsymbol{\Pi} = \begin{bmatrix} 0.4475 & 0 & 0 \\ 0 & 1 & 0 \\ 0.5525 & 0 & 1 \end{bmatrix}. \tag{7.34}$$

To demonstrate the results above, we compute the time evolution of each modal variable $z(t)$ for the following set of initial conditions: $\boldsymbol{\gamma}_0 = [0.1; 0; 0]^\top$, $\boldsymbol{\gamma}_0 = [0; 0.1; 0]^\top$, and $\boldsymbol{\gamma}_0 = [0; 0; 0.1]^\top$, which allows us to distinguish the influence of each observable function on each mode. The results depicted in Fig. 7.2 are in good agreement with $\boldsymbol{\Pi}$.

7.3 Numerical results in power systems

We perform simulations on the two-area four-machine system depicted in Fig. 7.3, whose power flow setpoint is presented in Appendix A.3; the remaining steady-state and dynamic

data is found in [213]. All the generators are modeled using the sub-transient model and are equipped with an AVR and a fast-response exciter, and the loads are modeled as constant admittances. The system operating condition is a highly stressed one, close to the point of voltage collapse, characterized by a tie line flow $P_e = 457$ MW from Area 1 to Area 2. A fault is applied to Bus 5 and cleared after 10 milliseconds with no line switching. We stress that we rely on time-domain simulation only to generate the data used by the EDMD algorithm, i.e., to emulate PMU measurements. The IEEE standard C37.118.1-2011 encourages that PMUs support reporting rates of 100 or 120 frames per second; in alignment with that, we assume a reporting rate of 120 frames per second.

From the small-signal stability analysis, the system presents an inter-area oscillatory (linear) mode, as well as two local oscillatory (linear) modes, as in Table 7.1; this information will be used later to analyze the obtained results.

Table 7.1: Linear electromechanical modes of oscillation from small-signal stability analysis

Mode	Eigenvalue	Freq. (Hz)	Damping Ratio (%)
Inter-area	$-0.0614 \pm j1.6522$	0.2629	0.0371
Local (Area 1)	$-1.2452 \pm j8.0146$	1.2756	0.1535
Local (Area 2)	$-1.7390 \pm j7.6730$	1.2212	0.2210

To compute approximations to the Koopman tuples via EDMD, we select the following vector of observable functions: $\boldsymbol{\gamma} = \left[\boldsymbol{\delta}^\top;\, \boldsymbol{\omega}^\top;\, \boldsymbol{E}_{fd}^\top;\, (\sin(\boldsymbol{\delta}))^\top;\, (\cos(\boldsymbol{\delta}))^\top;\, \boldsymbol{P}_e^\top\right]^\top$, where $\boldsymbol{\delta}$, $\boldsymbol{\omega}$, $\boldsymbol{E}_{fd}$, and $\boldsymbol{P}_e$ are vectors containing the generators' rotor angle, rotor speed deviation, field voltage, and real power injection, respectively. Although this particular set of observable functions led to good results in all of the extensive tests that we have performed, we note that the choice of observable functions for power systems, and in general [129,172], remains as an open problem and is out of the scope of the present work. We also remark that the generators' rotor angle is not directly measured in practice, and should be estimated via a dynamic state estimator [50, 51]. Likewise, brushless excitation systems are commonly found in practice and they do not provide access to measure the field voltage, which in this case shall be estimated as well.

After carrying out the EDMD, we reconstruct the evolution of the system states using (7.20). The results for the generators' rotor angle $\boldsymbol{\delta}$ are shown in Fig. 7.4, whereas other system state variables are omitted since they do not bring any additional information. To measure the estimation performance, we apply the Frobenius norm, defined as $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}\left|x_{ij} - \hat{x}_{ij}\right|^2}$, on the matrices containing the snapshots of the system states obtained from the time-domain simulation, $\boldsymbol{x}$, and from the EDMD, $\hat{\boldsymbol{x}}$, and found a value of 0.0952 that is satisfactory.

The eigenvalues obtained from the EDMD are presented in Table 7.2. Note that the pair of eigenvalues $\lambda_{4,5}$ is similar to the local (linear) mode of area 2. Likewise, $\lambda_{6,7}$ refer to the local

mode of area 1, and $\lambda_{19,20}$ to the inter-area mode. Note that we have 24 modes, as the system in Fig. 7.3 has 4 machines and we selected 6 observable functions. Furthermore, observe that because we adopt the center of angle reference frame and do not take any generator as reference, one eigenvalue is equal to zero, $\lambda_{16} = 0$. This is a well-known fact and helps on validating the estimation results.

Table 7.2: Modes frequency and damping ratio

	$\Re\{\lambda_i\}$	$\Im\{\lambda_i\}$	Freq. (Hz)	Damp. (%)
$\lambda_{1,2}$	-63.64	± 37.17	5.92	86.35
λ_3	-7.26	0.00	0.00	100.00
$\lambda_{4,5}$	-1.91	± 7.82	1.24	23.74
$\lambda_{6,7}$	-1.17	± 7.94	1.26	14.54
$\lambda_{8,9}$	-0.58	± 5.23	0.83	11.09
$\lambda_{10,11}$	-0.25	± 4.15	0.66	6.04
$\lambda_{12,13}$	-0.44	± 3.43	0.55	12.86
$\lambda_{14,15}$	-1.44	± 1.99	0.32	58.73
λ_{16}	0.00	0.00	0.00	100.00
$\lambda_{17,18}$	-0.15	± 1.95	0.31	7.73
$\lambda_{19,20}$	-0.03	± 1.47	0.23	1.95
$\lambda_{21,22}$	-0.52	± 0.94	0.15	48.70
$\lambda_{23,24}$	-0.32	± 0.10	0.02	95.12

Table 7.3 shows the mode-in-state participation factors computed using (7.23). Note that results are not in the unit interval as is the case for the model-based participation factors. In the model-based approach, if the eigenvalues are non-degenerate, each left eigenvector is orthogonal to all right eigenvectors except its corresponding one, and vice versa. This property does not hold for $\boldsymbol{\Xi}$, the matrix containing the left eigenvectors of $\boldsymbol{U}$, and $\boldsymbol{\Phi}$, the matrix containing the Koopman modes. We recommend to normalize the matrix containing the mode-in-state participation factors by row.

From Table 7.2, we can see that $\lambda_{1,2}$ is a control mode with frequency equal to 5.92 Hz. From Table 7.3, we observe that $\lambda_{1,2}$ has the highest participation on the state E_{fd2}, which is expected since generator 2 is electrically the closest to Bus 5, where we applied the fault. Although the participation factors are supposedly independent of the disturbance duration and location, the proposed technique is data-driven and rely on the most excited modes in the data set.

The modes $\lambda_{8,9}$, $\lambda_{14,15}$, and $\lambda_{17,18}$ respectively have the first, second and third highest participation on the states $\{\delta_1, \delta_2, \delta_3, \delta_4\}$. From Table 7.2, we observe that $f_{8,9} = 0.83$ Hz, $f_{14,15} = 0.32$ Hz, and $f_{17,18} = 0.31$ Hz, i.e., these are inter-area modes. Their frequency, however, differ from the linear inter-area mode in Table 7.1. We claim that $\lambda_{8,9}$, $\lambda_{14,15}$, and $\lambda_{17,18}$ are nonlinear modes not revealed by the linear analysis. The linear inter-area mode $\lambda_{19,20}$ appears immediately after the nonlinear inter-area modes with a significant participation. The linear local modes in Areas 1 and 2 show up in sequence, after the linear inter-area mode, with a high participation on the states $\{\delta_1, \delta_2\}$ and $\{\delta_3, \delta_4\}$ respectively.

Table 7.3: Mode-In-State Participation Factors Based on the Koopman Modes Decomposition

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}	λ_{19}	λ_{20}	λ_{21}	λ_{22}	λ_{23}	λ_{24}
δ_1	0.5	0.6	20.8	74.9	52.7	36.4	208.3	858.7	2.0	0.4	6.8	5.1	44.3	634.9	191.6	39.0	357.7	184.0	138.7	205.3	2.1	3.7	3.0	9.2
δ_2	0.5	0.6	25.9	53.9	37.9	86.4	494.3	877.8	2.0	0.3	5.0	4.0	34.8	645.3	194.8	15.2	372.7	191.8	145.6	215.5	2.1	3.7	3.5	10.4
δ_3	0.0	0.0	3.5	4.8	12.0	0.7	7.0	39.8	0.1	0.0	1.4	1.2	4.1	30.2	20.8	2.5	24.1	22.0	15.2	26.7	0.5	0.3	0.1	0.3
δ_4	0.0	0.0	0.6	1.0	2.7	1.1	7.9	28.0	0.0	0.0	0.4	0.4	2.2	7.5	2.4	0.5	9.7	4.5	2.5	3.7	0.0	0.0	0.3	0.4
ω_1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ω_2	0.0	0.0	0.0	0.0	0.1	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ω_3	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ω_4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.9	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Efd_1	45.5	44.8	17.8	6.4	6.9	3.8	10.5	15.8	0.0	0.0	0.3	0.3	4.9	15.8	3.1	1.5	16.4	7.7	1.6	2.7	0.0	0.0	0.0	0.2
Efd_2	383.3	349.7	50.6	24.6	22.9	35.0	98.6	17.3	0.0	0.1	2.4	2.2	30.4	58.0	2.4	1.5	87.5	37.8	2.7	3.4	0.0	0.1	0.2	0.9
Efd_3	19.3	17.6	4.2	4.1	3.8	5.6	15.7	65.8	0.0	0.0	1.8	2.0	28.4	59.1	2.5	0.8	75.2	32.5	2.8	3.6	0.0	0.1	0.1	0.4
Efd_4	26.2	22.4	4.7	4.8	11.0	5.7	15.9	46.5	0.0	0.0	2.2	2.6	26.6	51.0	5.9	4.8	73.7	30.0	1.9	4.1	0.1	0.1	0.2	0.9

Table 7.4: State-In-Mode Participation Factors Based on Koopman Modes Decomposition

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}	λ_{19}	λ_{20}	λ_{21}	λ_{22}	λ_{23}	λ_{24}
δ_1	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.00	0.00	0.05	0.05	0.05	0.05	0.19	0.19	0.08	0.07	0.07	0.01	0.01	0.25	0.25	0.04	0.04
δ_2	0.01	0.01	0.02	0.02	0.02	0.04	0.04	0.00	0.00	0.05	0.05	0.04	0.04	0.08	0.08	0.05	0.07	0.07	0.00	0.00	0.14	0.14	0.02	0.02
δ_3	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
δ_4	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00
ω_1	0.02	0.02	0.07	0.03	0.03	0.05	0.05	0.00	0.00	0.00	0.00	0.01	0.01	0.06	0.06	0.19	0.01	0.01	0.01	0.01	0.04	0.04	0.34	0.34
ω_2	0.01	0.01	0.00	0.03	0.03	0.00	0.00	0.01	0.01	0.04	0.04	0.03	0.03	0.26	0.26	0.02	0.03	0.03	0.01	0.01	0.13	0.13	0.01	0.01
ω_3	0.50	0.50	0.52	0.23	0.23	0.30	0.30	0.43	0.43	0.33	0.33	0.30	0.30	0.06	0.06	0.15	0.36	0.36	0.47	0.47	0.01	0.01	0.47	0.47
ω_4	0.43	0.43	0.30	0.60	0.60	0.45	0.45	0.55	0.55	0.43	0.43	0.49	0.49	0.02	0.02	0.39	0.32	0.32	0.41	0.41	0.03	0.03	0.02	0.02
E_{fd1}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{fd2}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{fd3}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E_{fd4}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\sin(\delta_1)$	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.05	0.05	0.02	0.02	0.02	0.00	0.00	0.06	0.06	0.01	0.01
$\sin(\delta_2)$	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.00	0.00	0.04	0.04	0.03	0.03	0.06	0.06	0.04	0.06	0.06	0.00	0.00	0.11	0.11	0.02	0.02
$\sin(\delta_3)$	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.00
$\sin(\delta_4)$	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
$\cos(\delta_1)$	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.00	0.00	0.04	0.04	0.04	0.04	0.13	0.13	0.06	0.04	0.04	0.01	0.01	0.18	0.18	0.04	0.04
$\cos(\delta_2)$	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.03	0.03	0.01	0.01
$\cos(\delta_3)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\cos(\delta_4)$	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
P_{gen1}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P_{gen2}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P_{gen3}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P_{gen4}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Similarly, the state-in-mode participation factors computed using (7.27) are shown in Table 7.4. Note that for the state-in-mode participation factors, the summation of any column is always equal to 1, which is a desirable property. Due to space limitations, the results in Table 7.4 are not discussed in details.

7.4 Concluding remarks

A novel data-driven technique that is able to compute linear and nonlinear participation factors based on the Koopman operator-theoretic framework has been proposed. Numerical simulations carried out on a canonical nonlinear dynamical system, and on the two-area four-machine system, demonstrated the performance of our approach.

The proposed technique is not restricted to operating conditions nearby the stable equilibrium point; in fact, it is applicable to any operating point including the ones off-attractor. Furthermore, the decomposition based on the Koopman modes is applicable to any dynamical system, and the proposed technique is not restricted to any particular kind of nonlinearity. To demonstrate the broadness of our technique, its performance under particular phenomena such as bifurcations will be evaluated and reported in future publications.

Chapter 8

Conclusions and future research

8.1 Conclusions

The potential application of the Koopman operator theory to electric power systems has been exploited with particular focus on the development of emergency controls. As it has been argued before, *time* is the main challenge that has to be overcome. A first step in this direction is the development of the robust Koopman Kalman filter, which allows us to perform robust nonlinear system identification, and most important, to rely on the linear Kalman filter to track nonlinear dynamics. The linear Kalman filter can be implemented very efficiently, thereby coping with the strict time requirement. Of course, a robust filter is certainly an *a priori* requirement for control. However, further theoretical and practical developments are required. From the theoretical standpoint, an investigation of the necessary and sufficient conditions for the convergence of the filter is needed. Likewise, a formal treatment of the required system observability is lacking. From the practical perspective, a comparison analysis between recorded measurements of the system response to an event, and numerical simulations using the robust Koopman Kalman filter is necessary to demonstrate the approach.

A second important step toward the design of emergency controls is the development of modal participation factors for nonlinear systems. The Koopman operator-theoretic framework extract modal information from measurements, whereas we ultimately act to control the system states. So, there has to be a measure of relative participation of modes in states, and states in modes; the modal participation factors provide this measure. The detail here is that we are able to compute the modal participation factors directly from measurements, are not restricted to operating conditions nearby the stable equilibrium point or to any particular kind of nonlinearity. However, further developments are also necessary in this topic. A formal treatment of this approach is lacking, as we essentially showed preliminary results.

8.2 Directions of future research

In this section, we will discuss the ongoing work and future directions of research. We start from the short-term developments that are already under course, and gradually move toward the big picture and main goals of this work.

Modal participation factors for nonlinear systems

A comparison between the results provided by the method of Normal forms and the results given by the Koopman operator theory is being prepared. This is important to demonstrate the fact that the proposed approach is not limited to any particular kind of nonlinearity. Also, we shall study cases with more rich nonlinear dynamics, with particular interest to bifurcations. Apart from the numerical tests, further theoretical development is also necessary to better understand the existing limitations on computing modal participation factors from recorded measurements.

Voltage stability evaluation using modal analysis

Following the seminal paper by Gao, Morison, and Kundur [221], Koopman mode decomposition offers a proper framework to the assessment of voltage stability using modal analysis. See also [222]. This is an interesting opportunity to apply the modal participation factors previously discussed, despite the fact that it will most certainly require further extensions such as defining the participation factors for algebraic variables as well.

The choice of observables for electric power systems

As discussed throughout the text, the choice of observables is key when computing the Koopman mode decomposition. However, the choice of observable functions for power systems, and in general, remains as an open problem. In the fluid dynamics and dynamical systems community, researchers are suggesting the use of algorithms that rely on large dictionaries of functions to discover the most suitable observables on the fly. These are "brute force" type of strategies that suit their applications where they do not have access to models. This is, however, not the case in electric power systems. Because we are restricted by *time*, and the possible options of observables are limited, it is of interest to find out what is the best compromise choice.

Koopman mode decomposition, model predictive control and uncertainty quantification using polynomial chaos

Initial results demonstrated that Koopman mode decomposition suits model predictive control [128, 223]. These results illuminated our way toward the application of Koopman mode decomposition to emergency control of electric power systems. Model predictive control fits naturally within energy management systems as it can act directly as feedback control or provide guidance to the operators by identifying the best compromise actions to be taken.

One of the first contributions to the application of model predictive control to emergency

control of electric power systems was made by Larsson, Hill and Olsson [224]. More recently, Hiskens et al. [225–228] have made several contributions to this topic. These works are either oriented toward quasi steady-state issues such as thermal limits of equipment, or long-term dynamics issues such as voltage control and automatic generation control. An application of model predictive control to protection schemes can be found in [229]. To the best of our knowledge, there is not an application focused on dynamic monitoring and control. One of the arguments is that "centralized model predictive control is impractical for control of large-scale, geographically expansive systems, such as power systems" [225]. The Koopman operator-theoretic framework holds the promise of putting that argument down, as we discussed throughout the text. We will investigate its application toward these lines.

Bibliography

[1] A. G. Phadke and J. S. Thorp, *Synchronized Phasor Measurements and Their Applications*. Springer, 2017.

[2] J. De La Ree, V. Centeno, J. Thorp, and A. Phadke, "Synchronized Phasor Measurement Applications in Power Systems," *Smart Grid, IEEE Transactions on*, vol. 1, no. 1, pp. 20–27, June 2010.

[3] M. Kezunovic, S. Meliopoulos, V. Venkatasubramanian, and V. Vittal, *Application of Time-Synchronized Measurements in Power System Transmission Networks*. Springer, 2014.

[4] M. S. Thomas and J. D. McDonald, *Power System SCADA and Smart Grids*. CRC Press, 2015.

[5] A. Vojdani, "Smart Integration," *IEEE Power and Energy Magazine*, vol. 6, no. 6, pp. 71–79, November 2008.

[6] S. Neumann, F. Wilhoit, M. Goodrich, and V. M. Balijepalli, "Everything's Talking to Each Other: Smart Meters Generate Big Data for Utilities and Customers," *IEEE Power and Energy Magazine*, vol. 14, no. 1, pp. 40–47, Jan 2016.

[7] A. von Meier, E. Stewart, A. McEachern, M. Andersen, and L. Mehrmanesh, "Precision Micro-Synchrophasors for Distribution Systems: A Summary of Applications," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2926–2936, Nov 2017.

[8] M. Henderson, "We Have the Data: Making It Work for Us [From the Editor]," *IEEE Power and Energy Magazine*, vol. 16, no. 3, pp. 4–100, May 2018.

[9] T. Mai, M. M. Hand, S. F. Baldwin, R. H. Wiser, G. L. Brinkman, P. Denholm, D. J. Arent, G. Porro, D. Sandor, D. J. Hostick, M. Milligan, E. A. DeMeo, and M. Bazilian, "Renewable Electricity Futures for the United States," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 2, pp. 372–378, April 2014.

[10] J. P. Chaves-Avila, K. Wurzburg, T. Gomez, and P. Linares, "The Green Impact: How Renewable Sources Are Changing EU Electricity Prices," *IEEE Power and Energy Magazine*, vol. 13, no. 4, pp. 29–40, July 2015.

[11] B. Kroposki, B. Johnson, Y. Zhang, V. Gevorgian, P. Denholm, B. M. Hodge, and B. Hannegan, "Achieving a 100% Renewable Grid: Operating Electric Power Systems with Extremely High Levels of Variable Renewable Energy," *IEEE Power and Energy Magazine*, vol. 15, no. 2, pp. 61–73, March 2017.

[12] R. Pires, M. Netto, and F. Becon-Lemos, "Analysis of Rotor Angle Stability in a real-operating Micro-grid in Southern Brazil," in *2012 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2012, pp. 1101–1106.

[13] A. J. Wood, B. F. Wollenberg, and G. B. Sheble, *Power generation, operation, and control*. John Wiley & Sons, 2012.

[14] H. Pandžić, Y. Dvorkin, T. Qiu, Y. Wang, and D. S. Kirschen, "Toward Cost-Efficient and Reliable Unit Commitment Under Uncertainty," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 970–982, March 2016.

[15] R. Billinton and R. N. Allan, *Reliability Evaluation of Power Systems*. Springer, 1996.

[16] A. M. L. da Silva, J. F. da Costa Castro, and R. Billinton, "Probabilistic Assessment of Spinning Reserve via Cross-Entropy Method Considering Renewable Sources and Transmission Restrictions," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4574–4582, July 2018.

[17] H. D. Nguyen and K. Turitsyn, "Robust Stability Assessment in the Presence of Load Dynamics Uncertainty," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1579–1594, March 2016.

[18] P. N. Papadopoulos and J. V. Milanović, "Probabilistic Framework for Transient Stability Assessment of Power Systems With High Penetration of Renewable Generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3078–3088, July 2017.

[19] G. Gross and F. D. Galiana, "Short-term load forecasting," *Proceedings of the IEEE*, vol. 75, no. 12, pp. 1558–1573, Dec 1987.

[20] D. W. Bunn, "Forecasting loads and prices in competitive power markets," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 163–169, Feb 2000.

[21] G. K. Venayagamoorthy, K. Rohrig, and I. Erlich, "One Step Ahead: Short-Term Wind Power Forecasting and Intelligent Predictive Control Based on Data Analytics," *IEEE Power and Energy Magazine*, vol. 10, no. 5, pp. 70–78, Sept 2012.

[22] A. Tuohy, J. Zack, S. E. Haupt, J. Sharp, M. Ahlstrom, S. Dise, E. Grimit, C. Mohrlen, M. Lange, M. G. Casado, J. Black, M. Marquis, and C. Collier, "Solar Forecasting: Methods, Challenges, and Performance," *IEEE Power and Energy Magazine*, vol. 13, no. 6, pp. 50–59, Nov 2015.

[23] C. Zhao and Y. Guan, "Data-Driven Stochastic Unit Commitment for Integrating Wind Generation," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2587–2596, July 2016.

[24] A. Bagheri, J. Wang, and C. Zhao, "Data-Driven Stochastic Transmission Expansion Planning," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3461–3470, Sep. 2017.

[25] M. Ilić and S. Liu, *Hierarchical Power Systems Control: Its Value in a Changing Industry*. Springer, 1996.

[26] G. Rogers, *Power System Oscillations*. Springer Science & Business Media, 2000.

[27] T. Dy-Liacco, "Control centers are here to stay," *Computer Applications in Power, IEEE*, vol. 15, no. 4, pp. 18–23, October 2002.

[28] F. Maghsoodlou, R. Masiello, and T. Ray, "Energy management systems," *Power and Energy Magazine, IEEE*, vol. 2, no. 5, pp. 49–57, Sept 2004.

[29] F. Wu, K. Moslehi, and A. Bose, "Power System Control Centers: Past, Present, and Future," *Proceedings of the IEEE*, vol. 93, no. 11, pp. 1890–1908, Nov 2005.

[30] J. Momoh and L. Mili, *Economic Market Design and Planning for Electric Power Systems*. Wiley, 2009.

[31] A. S. Debs, *Modern Power Systems Control and Operation*. Springer, 1988.

[32] H. Bevrani and T. Hiyama, *Intelligent Automatic Generation Control*. CRC Press, 2017.

[33] A. Monticelli, *State Estimation in Electric Power Systems: A Generalized Approach*. Springer, 1999.

[34] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Application*. CRC Press, 2004.

[35] F. C. Schweppe and J. Wildes, "Power System Static-State Estimation, Part I: Exact Model," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 120–125, Jan 1970.

[36] F. C. Schweppe and D. B. Rom, "Power System Static-State Estimation, Part II: Approximate Model," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 125–130, Jan 1970.

[37] F. C. Schweppe, "Power System Static-State Estimation, Part III: Implementation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 130–135, Jan 1970.

[38] M. R. Irving, R. C. Owen, and M. J. H. Sterling, "Power-system state estimation using linear programming," *Electrical Engineers, Proceedings of the Institution of*, vol. 125, no. 9, pp. 879–885, September 1978.

[39] W. W. Kotiuga and M. Vidyasagar, "Bad Data Rejection Properties of Weughted Least Absolute Value Techniques Applied to Static State Estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 4, pp. 844–853, April 1982.

[40] A. Abur, "A bad data identification method for linear programming state estimation," *IEEE Transactions on Power Systems*, vol. 5, no. 3, pp. 894–901, Aug 1990.

[41] A. Abur and M. K. Celik, "A fast algorithm for the weighted least absolute value state estimation [for power systems]," *IEEE Transactions on Power Systems*, vol. 6, no. 1, pp. 1–8, Feb 1991.

[42] M. K. Celik and A. Abur, "A robust WLAV state estimator using transformations," *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 106–113, Feb 1992.

[43] L. Mili, M. Cheniae, N. Vichare, and P. Rousseeuw, "Robust state estimation based on projection statistics [of power systems]," *Power Systems, IEEE Transactions on*, vol. 11, no. 2, pp. 1118–1127, May 1996.

[44] A. S. Debs, "Estimation of external network equivalents from internal system data," *IEEE Transactions on Power Apparatus and Systems*, vol. 94, no. 2, pp. 273–279, Mar 1975.

[45] J. D. McCalley, S. Wang, Q. L. Zhao, G. Z. Zhou, R. T. Treinen, and A. D. Papalexopoulos, "Security boundary visualization for systems operation," *IEEE Transactions on Power Systems*, vol. 12, no. 2, pp. 940–947, May 1997.

[46] J. Zhao, A. Gomez-Exposito, M. Netto, L. Mili, A. Abur, V. Terzija, I. Kamwa, B. C. Pal, A. K. Singh, J. Qi, Z. Huang, and A. P. S. Meliopoulos, "Power System Dynamic State Estimation: Motivations, Definitions, Methodologies and Future Work," *IEEE Transactions on Power Systems*, pp. 1–1, 2019.

[47] H. Modir and R. A. Schlueter, "A Dynamic State Estimator for Dynamic Security Assessment," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-100, no. 11, pp. 4644–4652, Nov 1981.

[48] E. Ghahremani and I. Kamwa, "Dynamic State Estimation in Power System by Applying the Extended Kalman Filter With Unknown Inputs to Phasor Measurements," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2556–2566, Nov 2011.

[49] E. Ghahremani and I. Kamwa, "Online State Estimation of a Synchronous Generator Using Unscented Kalman Filter From Phasor Measurements Units," *IEEE Transactions on Energy Conversion*, vol. 26, no. 4, pp. 1099–1108, Dec 2011.

[50] M. Netto, J. Zhao, and L. Mili, "A robust extended Kalman filter for power system dynamic state estimation using PMU measurements," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, July 2016, pp. 1–5.

[51] J. Zhao, M. Netto, and L. Mili, "A Robust Iterated Extended Kalman Filter for Power System Dynamic State Estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3205–3216, July 2017.

[52] F. Dorfler, F. Pasqualetti, and F. Bullo, "Continuous-Time Distributed Observers With Discrete Communication," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 296–304, April 2013.

[53] W. Price et al., "Load representation for dynamic performance analysis," *IEEE Transactions on Power Systems*, vol. 8, no. 2, pp. 472–482, May 1993.

[54] J. H. Chow, *Power System Coherency and Model Reduction.* Springer, 2013.

[55] A. K. Singh and B. C. Pal, "Decentralized Dynamic State Estimation in Power Systems Using Unscented Transformation," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 794–804, March 2014.

[56] E. Ghahremani and I. Kamwa, "Local and Wide-Area PMU-Based Decentralized Dynamic State Estimation in Multi-Machine Power Systems," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 547–562, Jan 2016.

[57] Z. Huang, P. Du, D. Kosterev, and B. Yang, "Application of extended Kalman filter techniques for dynamic model parameter calibration," in *IEEE Power and Energy Society General Meeting*, July 2009, pp. 1–8.

[58] A. Paul, I. Kamwa, and G. Joos, "Centralized Dynamic State Estimation Using a Federation of Extended Kalman Filters with Intermittent PMU Data from Generator Terminals," *IEEE Transactions on Power Systems*, 2018.

[59] A. A. Fouad, F. Aboytes, V. F. Carvalho, S. L. Corey, K. J. Dhir, and R. Vierra, "Dynamic security assessment practices in North America," *IEEE Transactions on Power Systems*, vol. 3, no. 3, pp. 1310–1321, Aug 1988.

[60] I. Konstantelos, G. Jamgotchian, S. H. Tindemans, P. Duchesne, S. Cole, C. Merckx, G. Strbac, and P. Panciatici, "Implementation of a Massively Parallel Dynamic Security Assessment Platform for Large-Scale Grids," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1417–1426, May 2017.

[61] M. Shahidehpour and Y. Wang, *Communications and Control in Electric Power Systems: Applications of parallel and distributed processing.* Wiley-Interscience, 2003.

[62] S. C. Savulescu, *Real-time stability in power systems: techniques for early detection of the risk of blackout.* Springer, 2014.

[63] M. Pavella, D. Ernst, and D. Ruiz-Vega, *Transient Stability of Power Systems: A Unified Approach to Assessment and Control*. Springer, 2000.

[64] Y. Susuki and I. Mezić, "Nonlinear Koopman Modes and Coherency Identification of Coupled Swing Dynamics," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 1894–1904, Nov 2011.

[65] Y. Makarov, V. Reshetov, A. Stroev, and N. Voropai, "Blackout Prevention in the United States, Europe, and Russia," *Proceedings of the IEEE*, vol. 93, no. 11, pp. 1942–1955, Nov 2005.

[66] B. Kroposki, "Basic Research Needs for Autonomous Energy Grids–Summary Report of the Workshop on Autonomous Energy Grids," NREL/TP-5D00-70428, Tech. Rep., 2017.

[67] B. Kroposki, E. Dall'Anese, A. Bernstein, Y. Zhang, and B.-M. Hodge, "Autonomous Energy Grids," in *51st Hawaii International Conference on System Sciences*, 2018, pp. 1–10.

[68] M. A. Gandhi and L. Mili, "Robust Kalman Filter Based on a Generalized Maximum-Likelihood-Type Estimator," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2509–2520, May 2010.

[69] M. Netto and L. Mili, "Robust Koopman Operator-based Kalman Filter for Power Systems Dynamic State Estimation," in *2018 IEEE Power Energy Society General Meeting (PESGM)*, Aug 2018, pp. 1–5.

[70] M. Netto and L. Mili, "A Robust Data-Driven Koopman Kalman Filter for Power Systems Dynamic State Estimation," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7228–7237, Nov 2018.

[71] A. Surana and A. Banaszuk, "Linear observer synthesis for nonlinear systems using Koopman Operator framework," *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 716 – 723, 2016, 10th IFAC Symposium on Nonlinear Control Systems NOLCOS 2016.

[72] A. Surana, "Koopman operator based observer synthesis for control-affine nonlinear systems," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec 2016, pp. 6492–6499.

[73] A. Surana, "Koopman Operator Framework for Time Series Modeling and Analysis," *Journal of Nonlinear Science*, Jan 2018.

[74] P. Tamburello and L. Mili, "Robustness Analysis of the Phase-Phase Correlator to White Impulsive Noise With Applications to Autoregressive Modeling," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 6053–6058, Nov 2012.

[75] P. Tamburello and L. Mili, "New Robust Estimators of Correlation and Weighted Basis Pursuit," *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 882–894, Feb 2015.

[76] P. M. Tamburello, "Iterative Memoryless Non-linear Estimators of Correlation for Complex-Valued Gaussian Processes That Exhibit Robustness to Impulsive Noise," Ph.D. dissertation, Virginia Polytechnic Institute and State University, 2015.

[77] M. Netto and L. Mili, "A Robust Prony Method for Power System Electromechanical Modes Identification," in *2017 IEEE Power and Energy Society General Meeting*, July 2017, pp. 1–5.

[78] M. Netto and L. Mili, "Robust Data Filtering for Estimating Electromechanical Modes of Oscillation via the Multichannel Prony Method," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4134–4143, July 2018.

[79] M. Netto, V. Krishnan, L. Mili, Y. Susuki, and Y. Zhang, "A Hybrid Framework Combining Model-Based and Data-Driven Methods for Hierarchical Decentralized Robust Dynamic State Estimation," in *2019 IEEE Power and Energy Society General Meeting*, Aug 2019, pp. 1–5.

[80] M. Netto, Y. Susuki, and L. Mili, "Data-Driven Participation Factors for Nonlinear Systems Based on Koopman Mode Decomposition," *IEEE Control Systems Letters*, vol. 3, no. 1, pp. 198–203, Jan 2019.

[81] H. Arbabi, "Introduction to Koopman operator theory of dynamical systems," 2018, unpublished.

[82] P. M. Morse, "In Memoriam: Bernard Osgood Koopman, 1900-1981," *Operations Research*, vol. 30, no. 3, pp. viii–427, 1982.

[83] B. O. Koopman, "Hamiltonian Systems and Transformation in Hilbert Space," *Proceedings of the National Academy of Sciences*, vol. 17, no. 5, pp. 315–318, 1931.

[84] P. R. Halmos, "The Legend of John Von Neumann," *The American Mathematical Monthly*, vol. 80, no. 4, pp. 382–394, 1973.

[85] B. O. Koopman and J. v. Neumann, "Dynamical Systems of Continuous Spectra," *Proceedings of the National Academy of Sciences*, vol. 18, no. 3, pp. 255–263, 1932.

[86] T. Carleman, "Application de la théorie des équations intégrales linéaires aux systèmes d'équations différentielles non linéaires," *Acta Math.*, vol. 59, pp. 63–87, 1932.

[87] K. Kowalski and W.-H. Steeb, *Nonlinear Dynamical Systems and Carleman Linearization*. World Scientific, 1991.

[88] A. Lasota and M. C. Mackey, *Probabilistic Properties of Deterministic Systems*. Cambridge University Press, 1985.

[89] R. Mañé, *Ergodic Theory and Differentiable Dynamics*. Springer-Verlag, 1987.

[90] K. E. Petersen, *Ergodic Theory*. Cambridge University Press, 1989, vol. 2.

[91] I. Mezić and A. Banaszuk, "Comparison of systems with complex behavior," *Physica D: Nonlinear Phenomena*, vol. 197, no. 1, pp. 101 – 133, 2004.

[92] I. Mezić, "Spectral Properties of Dynamical Systems, Model Reduction and Decompositions," *Nonlinear Dynamics*, vol. 41, no. 1, pp. 309–325, Aug 2005.

[93] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, "Spectral analysis of nonlinear flows," *Journal of Fluid Mechanics*, vol. 641, p. 115–127, 2009.

[94] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *Journal of Fluid Mechanics*, vol. 656, p. 5–28, 2010.

[95] I. Mezić, "Analysis of Fluid Flows via Spectral Properties of the Koopman Operator," *Annual Review of Fluid Mechanics*, vol. 45, no. 1, pp. 357–378, 2013.

[96] Y. Susuki and I. Mezić, "Nonlinear Koopman Modes and a Precursor to Power System Swing Instabilities," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1182–1191, Aug 2012.

[97] Y. Susuki and I. Mezić, "Nonlinear Koopman Modes and Power System Stability Assessment Without Models," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 899–907, March 2014.

[98] F. Raak, Y. Susuki, and T. Hikihara, "Data-Driven Partitioning of Power Networks Via Koopman Mode Analysis," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2799–2808, July 2016.

[99] E. O. Roxin, *Control Theory and its Applications*. Gordon and Breach Science Publishers, 1997.

[100] E. T. Bell, *Men of Mathematics*. Simon & Schuster, 1986.

[101] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer, 1983.

[102] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer, 2003.

[103] R. Temam, *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*. Springer-Verlag, 1997.

[104] H.-D. Chiang, *Direct Methods for Stability Analysis of Electric Power Systems*. John Wiley & Sons, 2011.

[105] Y. Lan and I. Mezić, "Linearization in the large of nonlinear systems and Koopman operator spectrum," *Physica D: Nonlinear Phenomena*, vol. 242, no. 1, pp. 42 – 53, 2013.

[106] I. Mezić, "Koopman Operator Spectrum and Data Analysis," *Preprint, arXiv:1702.07597v1*, 2017.

[107] H. Arbabi and I. Mezić, "Study of dynamics in post-transient flows using Koopman mode decomposition," *Phys. Rev. Fluids*, vol. 2, p. 124402, Dec 2017.

[108] M. Korda, M. Putinar, and I. Mezić, "Data-driven spectral analysis of the Koopman operator," *Preprint, arXiv:1710.06532v1*, 2017.

[109] Y. Saad, "Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices," *Linear Algebra and its Applications*, vol. 34, pp. 269–295, 1980.

[110] P. Schmid and J. Sesterhenn, "Dynamic Mode Decomposition of numerical and experimental data," http://meetings.aps.org/link/BAPS.2008.DFD.MR.7.

[111] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, "On dynamic mode decomposition: Theory and applications," *Journal of Computational Dynamics*, vol. 1, no. 2, pp. 391–421, 2014.

[112] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A Data–Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition," *Journal of Nonlinear Science*, vol. 25, no. 6, pp. 1307–1346, Dec 2015.

[113] S. Klus, P. Koltai, and C. Schütte, "On the numerical approximation of the Perron-Frobenius and Koopman operator," *Journal of Computational Dynamics*, vol. 3, no. 1, pp. 51–79, 2016.

[114] M. Korda and I. Mezić, "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator," *Journal of Nonlinear Science*, vol. 28, no. 2, pp. 687–710, Apr 2018.

[115] H. Arbabi and I. Mezić, "Ergodic Theory, Dynamic Mode Decomposition, and Computation of Spectral Properties of the Koopman Operator," *SIAM Journal on Applied Dynamical Systems*, vol. 16, no. 4, pp. 2096–2126, 2017.

[116] R. M. Mohr, "Spectral Properties of the Koopman Operator in the Analysis of Nonstationary Dynamical Systems," Ph.D. dissertation, University of California, Santa Barbara, 2014.

[117] K. K. Chen, J. H. Tu, and C. W. Rowley, "Variants of Dynamic Mode Decomposition: Boundary Condition, Koopman, and Fourier Analyses," *Journal of Nonlinear Science*, vol. 22, no. 6, pp. 887–915, Dec 2012.

[118] M. R. Jovanović, P. J. Schmid, and J. W. Nichols, "Sparsity-promoting dynamic mode decomposition," *Physics of Fluids*, vol. 26, no. 2, p. 024103, 2014.

[119] J. N. Kutz, X. Fu, S. L. Brunton, and N. B. Erichson, "Multi-resolution Dynamic Mode Decomposition for Foreground/Background Separation and Object Tracking," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, Dec 2015, pp. 921–929.

[120] J. Kutz, X. Fu, and S. Brunton, "Multiresolution Dynamic Mode Decomposition," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 2, pp. 713–735, 2016.

[121] J. L. Proctor, S. L. Brunton, and J. N. Kutz, "Dynamic Mode Decomposition with Control," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 1, pp. 142–161, 2016.

[122] M. S. Hemati, M. O. Williams, and C. W. Rowley, "Dynamic mode decomposition for large and streaming datasets," *Physics of Fluids*, vol. 26, no. 11, p. 111701, 2014.

[123] F. Guéniat, L. Mathelin, and L. R. Pastur, "A dynamic mode decomposition approach for large and arbitrarily sampled systems," *Physics of Fluids*, vol. 27, no. 2, p. 025113, 2015.

[124] S. L. Brunton, J. L. Proctor, J. H. Tu, and J. N. Kutz, "Compressed sensing and dynamic mode decomposition," *Journal of Computational Dynamics*, vol. 2, no. 2, pp. 165–191, 2015.

[125] S. T. M. Dawson, M. S. Hemati, M. O. Williams, and C. W. Rowley, "Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition," *Experiments in Fluids*, vol. 57, no. 3, p. 42, Feb 2016.

[126] M. S. Hemati, C. W. Rowley, E. A. Deem, and L. N. Cattafesta, "De-biasing the dynamic mode decomposition for applied Koopman spectral analysis of noisy datasets," *Theoretical and Computational Fluid Dynamics*, vol. 31, no. 4, pp. 349–368, Aug 2017.

[127] S. L. Brunton, B. W. Brunton, J. L. Proctor, and J. N. Kutz, "Koopman Invariant Subspaces and Finite Linear Representations of Nonlinear Dynamical Systems for Control," *PLOS ONE*, vol. 11, no. 2, pp. 1–19, 02 2016.

[128] M. Korda and I. Mezić, "Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control," *Preprint, arXiv:1611.03537v2*, 2017.

[129] E. Kaiser, J. N. Kutz, and S. L. Brunton, "Data-driven discovery of Koopman eigenfunctions for control," *Preprint, arXiv:1707.01146v1*, 2017.

[130] H. Arbabi, M. Korda, and I. Mezić, "A data-driven Koopman model predictive control framework for nonlinear flows," *Preprint, arXiv:1804.05291v2*, 2018.

[131] M. Korda, Y. Susuki, and I. Mezić, "Power grid transient stabilization using Koopman model predictive control," *Preprint, arXiv:1803.10744v1*, 2018.

[132] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. John Wiley & Sons, 2006.

[133] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice with MATLAB*. John Wiley & Sons, 2014.

[134] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering with Matlab Exercises*. John Wiley & Sons, 2012.

[135] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 1986, vol. 114.

[136] D. Simon, "Kalman Filtering," *Embedded Systems Programming*, vol. 14, no. 6, pp. 72–79, 2001.

[137] M. Panteli and D. S. Kirschen, "Situation awareness in power systems: Theory, challenges and applications," *Electric Power Systems Research*, vol. 122, pp. 140–151, 2015.

[138] C. Taylor, D. C. Erickson, K. Martin, R. Wilson, and V. Venkatasubramanian, "WACS-Wide-Area Stability and Voltage Control System: R D and Online Demonstration," *Proceedings of the IEEE*, vol. 93, no. 5, pp. 892–906, May 2005.

[139] N. Zhou, D. Meng, Z. Huang, and G. Welch, "Dynamic State Estimation of a Synchronous Machine Using PMU Data: A Comparative Study," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 450–460, Jan 2015.

[140] D. N. Kosterev, C. W. Taylor, and W. A. Mittelstadt, "Model validation for the August 10, 1996 WSCC system outage," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 967–979, Aug 1999.

[141] Z. Huang, P. Du, D. Kosterev, and S. Yang, "Generator dynamic model validation and parameter calibration using phasor measurements at the point of connection," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1939–1949, May 2013.

[142] N. Zhou, S. Lu, R. Singh, and M. A. Elizondo, "Calibration of reduced dynamic models of power systems using phasor measurement unit (PMU) Data," in *2011 North American Power Symposium*, Aug 2011, pp. 1–7.

[143] N. Zhou, Z. Huang, D. Meng, S. Elbert, S. Wang, and R. Diao, "Capturing Dynamics in the Power Grid: Formulation of Dynamic State Estimation through Data Assimilation," Technical Report PNNL-23213, Pacific Northwest National Laboratory, Tech. Rep., 2014.

[144] W. A. Stahel, "Robuste Schätzungen: infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen," Ph.D. dissertation, ETH Zürich, Switzerland, 1981.

[145] D. L. Donoho, "Breakdown Properties of Multivariate Location Estimators," qualifying paper, Harvard University, Boston, MA, 1982.

[146] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection.* John Wiley & Sons, 1987.

[147] M. Gasko and D. L. Donoho, "Influential Observation in Data Analysis," in *Proceedings of the Business and Economic Statistics Section*, 1982, pp. 104–110.

[148] D. L. Donoho and M. Gasko, "Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness," *Ann. Statist.*, vol. 20, no. 4, pp. 1803–1827, 12 1992.

[149] L. Thomas and L. Mili, "A Robust GM-Estimator for the Automated Detection of External Defects on Barked Hardwood Logs and Stems," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3568–3576, July 2007.

[150] L. Fan and Y. Wehbe, "Extended Kalman filtering based real-time dynamic state and parameter estimation using PMU data," *Electric Power Systems Research*, vol. 103, pp. 168 – 177, 2013.

[151] J. L. Crassidis, F. L. Markley, and Y. Cheng, "Survey of Nonlinear Attitude Estimation Methods," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 1, pp. 12–28, jan 2007.

[152] F. Gustafsson and G. Hendeby, "Some Relations Between Extended and Unscented Kalman Filters," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 545–555, Feb 2012.

[153] S. Wang, W. Gao, and A. P. S. Meliopoulos, "An Alternative Method for Power System Dynamic State Estimation Based on Unscented Transform," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 942–950, May 2012.

[154] A. K. Singh and B. C. Pal, "Decentralized Dynamic State Estimation in Power Systems Using Unscented Transformation," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 794–804, March 2014.

[155] A. Abur and A. Rouhani, "Linear Phasor Estimator Assisted Dynamic State Estimation," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2016.

[156] U. Hammes, E. Wolsztynski, and A. M. Zoubir, "Robust Tracking and Geolocation for Wireless Networks in NLOS Environments," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 889–901, Oct 2009.

[157] U. Hammes and A. M. Zoubir, "Robust MT Tracking Based on M-Estimation and Interacting Multiple Model Algorithm," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3398–3409, July 2011.

[158] A. Pai, *Energy function analysis for power system stability*. Springer Science & Business Media, 1989.

[159] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar 2004.

[160] G. Valverde and V. Terzija, "Unscented Kalman filter for power system dynamic state estimation," *Generation, Transmission Distribution, IET*, vol. 5, no. 1, pp. 29–37, Jan 2011.

[161] Z. Huang, N. Zhou, R. Diao, S. Wang, S. Elbert, D. Meng, and S. Lu, "Capturing real-time power system dynamics: Opportunities and challenges," in *2015 IEEE Power Energy Society General Meeting*, July 2015, pp. 1–5.

[162] R. Ramos, "Benchmark Systems for Small-Signal Stability Analysis and Control," IEEE PES, Tech. Rep., 2015.

[163] A. J. Stromberg and D. Ruppert, "Breakdown in nonlinear regression," *Journal of the American Statistical Association*, vol. 87, no. 420, pp. 991–997, 1992.

[164] L. Mili, C. W. Coakley *et al.*, "Robust estimation in structured linear regression," *The Annals of Statistics*, vol. 24, no. 6, pp. 2593–2607, 1996.

[165] L. Mili, G. Steeno, F. Dobraca, and D. French, "A robust estimation method for topology error identification," *IEEE Transactions on Power Systems*, vol. 14, no. 4, pp. 1469–1476, Nov 1999.

[166] J. Chang, G. N. Taranto, and J. H. Chow, "Dynamic state estimation using a nonlinear observer for optimal series-capacitor switching control," *International Journal of Electrical Power & Energy Systems*, vol. 19, no. 7, pp. 441 – 447, 1997.

[167] E. Scholtz, "Observer-based monitors and distributed wave controllers for electromechanical disturbances in power systems," Ph.D. dissertation, Massachusetts Institute of Technology, 2004.

[168] Y. Weng, R. Negi, C. Faloutsos, and M. D. Ilić, "Robust Data-Driven State Estimation for Smart Grid," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1956–1967, July 2017.

[169] E. J. Candès and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[170] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, "Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns," *IEEE Control Systems*, vol. 38, no. 3, pp. 63–86, June 2018.

[171] L. A. Aguirre, B. O. S. Teixeira, and L. A. B. Tôrres, "Using data-driven discrete-time models and the unscented Kalman filter to estimate unobserved variables of nonlinear systems," *Phys. Rev. E*, vol. 72, p. 026226, Aug 2005.

[172] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.

[173] D. C. Mazur, "Synchronized Rotor Angle Measurement of Synchronous Machines," Master's thesis, Virginia Polytechnic Institute and State University, 2012.

[174] J. Delport, "Synchronized Measurement of Machine Rotor Angle and Its Application," Master's thesis, Virginia Polytechnic Institute and State University, 2014.

[175] S. Chan, "Modal controllability and observability of power-system models," *International Journal of Electrical Power & Energy Systems*, vol. 6, no. 2, pp. 83 – 88, 1984.

[176] A. Hamdan and A. Elabdalla, "Geometric measures of modal controllability and observability of power system models," *Electric Power Systems Research*, vol. 15, no. 2, pp. 147 – 155, 1988.

[177] A. M. A. Hamdan and A. H. Nayfeh, "Measures of Modal Controllability and Observability for First- and Second-Order Linear Systems," *Journal of Guidance, Control, and Dynamics*, vol. 12, no. 3, pp. 421–428, May-June 1989.

[178] F. P. de Mello, "Measurement of synchronous machine rotor angle from analysis of zero sequence harmonic components of machine terminal voltage," *IEEE Transactions on Power Delivery*, vol. 9, no. 4, pp. 1770–1777, Oct 1994.

[179] A. D. Angel, P. Geurts, D. Ernst, M. Glavic, and L. Wehenkel, "Estimation of rotor angles of synchronous machines using artificial neural networks and local PMU-based quantities," *Neurocomputing*, vol. 70, no. 16, pp. 2668 – 2678, 2007, neural Network Applications in Electrical Engineering Selected papers from the 3rd International Work-Conference on Artificial Neural Networks (IWANN 2005).

[180] P. Tripathy, S. C. Srivastava, and S. N. Singh, "A Divide-by-Difference-Filter Based Algorithm for Estimation of Generator Rotor Angle Utilizing Synchrophasor Measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 6, pp. 1562–1570, June 2010.

[181] 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.

[182] Y. Liao and M. Kezunovic, "Online Optimal Transmission Line Parameter Estimation for Relaying Applications," *IEEE Transactions on Power Delivery*, vol. 24, no. 1, pp. 96–102, Jan 2009.

[183] P. R. Prony, "Essai expérimental et analytique: sur les lois de la dilatabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, à différentes températures," *Journal de l'École Polytechnique Floréal et Plairial*, vol. 1, no. 22, pp. 24–76, 1795.

[184] J. F. Hauer, C. J. Demeure, and L. L. Scharf, "Initial results in Prony analysis of power system response signals," *IEEE Transactions on Power Systems*, vol. 5, no. 1, pp. 80–89, Feb 1990.

[185] S. Rai, D. Lalani, S. K. Nayak, T. Jacob, and P. Tripathy, "Estimation of low-frequency modes in power system using robust modified Prony," *IET Generation, Transmission Distribution*, vol. 10, no. 6, pp. 1401–1409, 2016.

[186] R. B. Leandro, A. S. e Silva, I. C. Decker, and M. N. Agostini, "Identification of the Oscillation Modes of a Large Power System Using Ambient Data," *Journal of Control, Automation and Electrical Systems*, vol. 26, no. 4, pp. 441–453, 2015.

[187] D. J. Trudnowski, J. W. Pierre, N. Zhou, J. F. Hauer, and M. Parashar, "Performance of Three Mode-Meter Block-Processing Algorithms for Automated Dynamic Stability Assessment," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 680–690, May 2008.

[188] G. Jacovitti and A. Neri, "Estimation of the autocorrelation function of complex Gaussian stationary processes by amplitude clipped signals," *IEEE Transactions on Information Theory*, vol. 40, no. 1, pp. 239–245, Jan 1994.

[189] D. J. Trudnowski, J. M. Johnson, and J. F. Hauer, "Making Prony analysis more accurate using multiple signals," *IEEE Transactions on Power Systems*, vol. 14, no. 1, pp. 226–231, Feb 1999.

[190] 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 *2016 IEEE Power and Energy Society General Meeting (PESGM)*, July 2016, pp. 1–5.

[191] http://www.dsatools.com/ssat/.

[192] http://www.ons.org.br/.

[193] Y. Susuki and I. Mezić, "A prony approximation of Koopman Mode Decomposition," in *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 7022–7027.

[194] I. J. Pérez-Arriaga, G. C. Verghese, and F. C. Schweppe, "Selective Modal Analysis with Applications to Electric Power Systems, Part I: Heuristic Introduction," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 9, pp. 3117–3125, Sept 1982.

[195] G. C. Verghese, I. J. Pérez-Arriaga, and F. C. Schweppe, "Selective Modal Analysis With Applications to Electric Power Systems, Part II: The Dynamic Stability Problem," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 9, pp. 3126–3134, Sept 1982.

[196] F. L. Pagola, I. J. Perez-Arriaga, and G. C. Verghese, "On sensitivities, residues and participations: applications to oscillatory stability analysis and control," *IEEE Transactions on Power Systems*, vol. 4, no. 1, pp. 278–285, Feb 1989.

[197] L. C. P. da Silva, Y. Wang, V. F. da Costa, and W. Xu, "Assessment of generator impact on system power transfer capability using modal participation factors," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 149, no. 5, pp. 564–570, Sep 2002.

[198] I. Genc, H. Schättler, and J. Zaborszky, "Clustering the Bulk Power System with Applications Towards Hopf Bifurcation Related Oscillatory Instability," *Electric Power Components and Systems*, vol. 33, no. 2, pp. 181–198, 2005.

[199] Y. Y. Hsu and C. L. Chen, "Identification of optimum location for stabiliser applications using participation factors," *IEE Proceedings C - Generation, Transmission and Distribution*, vol. 134, no. 3, pp. 238–244, May 1987.

[200] E. H. Abed, D. Lindsay, and W. A. Hashlamoun, "On participation factors for linear systems," *Automatica*, vol. 36, no. 10, pp. 1489 – 1496, 2000.

[201] W. A. Hashlamoun, M. A. Hassouneh, and E. H. Abed, "New Results on Modal Participation Factors: Revealing a Previously Unknown Dichotomy," *IEEE Transactions on Automatic Control*, vol. 54, no. 7, pp. 1439–1449, July 2009.

[202] S. N. Vassilyev, I. B. Yadykin, A. B. Iskakov, D. E. Kataev, A. A. Grobovoy, and N. G. Kiryanova, "Participation factors and sub-Gramians in the selective modal analysis of electric power systems," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14 806–14 811, 2017.

[203] V. Vittal, N. Bhatia, and A. A. Fouad, "Analysis of the inter-area mode phenomenon in power systems following large disturbances," *IEEE Transactions on Power Systems*, vol. 6, no. 4, pp. 1515–1521, Nov 1991.

[204] B. C. Lesieutre, A. M. Stankovic, and J. R. Lacalle-Melero, "A study of state variable participation in nonlinear limit-cycle behavior," in *Proceedings of International Conference on Control Applications*, Sep 1995, pp. 79–84.

[205] S. K. Starrett, "Application of normal forms of vector fields to stressed power systems," Ph.D. dissertation, Iowa State University, 1994.

[206] T. Tian, X. Kestelyn, O. Thomas, A. Hiroyuki, and A. R. Messina, "An Accurate Third-order Normal Form Approximation for Power System Nonlinear Analysis," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2017.

[207] C.-M. Lin, V. Vittal, W. Kliemann, and A. A. Fouad, "Investigation of modal interaction and its effects on control performance in stressed power systems using normal forms of vector fields," *IEEE Transactions on Power Systems*, vol. 11, no. 2, pp. 781–787, May 1996.

[208] Y. X. Ni, V. Vittal, W. Kliemann, and A. A. Fouad, "Nonlinear modal interaction in HVDC/AC power systems with DC power modulation," *IEEE Transactions on Power Systems*, vol. 11, no. 4, pp. 2011–2017, Nov 1996.

[209] J. Thapar, V. Vittal, W. Kliemann, and A. A. Fouad, "Application of the normal form of vector fields to predict interarea separation in power systems," *IEEE Transactions on Power Systems*, vol. 12, no. 2, pp. 844–850, May 1997.

[210] S. Saha, A. A. Fouad, W. H. Kliemann, and V. Vittal, "Stability boundary approximation of a power system using the real normal form of vector fields," *IEEE Transactions on Power Systems*, vol. 12, no. 2, pp. 797–802, May 1997.

[211] S. K. Starrett and A. A. Fouad, "Nonlinear measures of mode-machine participation [transmission system stability]," *IEEE Transactions on Power Systems*, vol. 13, no. 2, pp. 389–394, May 1998.

[212] G. Jang, V. Vittal, and W. Kliemann, "Effect of nonlinear modal interaction on control performance: use of normal forms technique in control design. I. General theory and procedure," *IEEE Transactions on Power Systems*, vol. 13, no. 2, pp. 401–407, May 1998.

[213] J. J. Sanchez-Gasca, V. Vittal, M. J. Gibbard, A. R. Messina, D. J. Vowles, S. Liu, and U. D. Annakkage, "Inclusion of higher order terms for small-signal (modal) analysis: committee report-task force on assessing the need to include higher order terms for small-signal (modal) analysis," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1886–1904, Nov 2005.

[214] N. Pariz, H. M. Shanechi, and E. Vaahedi, "Explaining and validating stressed power systems behavior using modal series," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 778–785, May 2003.

[215] H. M. Shanechi, N. Pariz, and E. Vaahedi, "General nonlinear modal representation of large scale power systems," *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1103–1109, Aug 2003.

[216] B. Hamzi and E. H. Abed, "Local mode-in-state participation factors for nonlinear systems," in *53rd IEEE Conference on Decision and Control*, Dec 2014, pp. 43–48.

[217] M. Budišić, R. Mohr, and I. Mezić, "Applied Koopmanism," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 4, p. 047510, 2012.

[218] Y. Susuki, I. Mezić, F. Raak, and T. Hikihara, "Applied Koopman operator theory for power systems technology," *Nonlinear Theory and Its Applications, IEICE*, vol. 7, no. 4, pp. 430–459, 2016.

[219] K. Sako, Y. Susuki, and T. Hikihara, "An Analysis of Voltage Dynamics in Power System Based on Koopman Operator," in *Proc. Joint Convention of SICE Kansai Section and ISCIE*, January 2016, pp. 36–41 (in Japanese).

[220] A. H. Nayfeh, *The Method of Normal Forms*, 2nd ed. Wiley-VCH, 2010.

[221] B. Gao, G. K. Morison, and P. Kundur, "Voltage stability evaluation using modal analysis," *IEEE Transactions on Power Systems*, vol. 7, no. 4, pp. 1529–1542, Nov 1992.

[222] H. Pinto, N. Martins, X. Filho, A. Bianco, P. Gomes, and M. dos Santos, "Modal Analysis for Voltage Stability: Application at Base Case and Point of Collapse," in *Bulk Power System Voltage Phenomena - III Voltage Stability, Security & Control*, August 1994, pp. 1–14.

[223] E. Kaiser, J. N. Kutz, and S. L. Brunton, "Sparse identification of nonlinear dynamics for model predictive control in the low-data limit," *Preprint, arXiv:1711.05501v1*, 2017.

[224] M. Larsson, D. J. Hill, and G. Olsson, "Emergency voltage control using search and predictive control," *International Journal of Electrical Power & Energy Systems*, vol. 24, no. 2, pp. 121 – 130, 2002.

[225] A. N. Venkat, I. A. Hiskens, J. B. Rawlings, and S. J. Wright, "Distributed MPC Strategies With Application to Power System Automatic Generation Control," *IEEE Transactions on Control Systems Technology*, vol. 16, no. 6, pp. 1192–1206, Nov 2008.

[226] M. Almassalkhi and I. Hiskens, "Model-predictive cascade mitigation in electric power systems with storage and renewables, Part I: Theory and implementation," in *2015 IEEE Power Energy Society General Meeting*, July 2015, pp. 1–1.

[227] M. R. Almassalkhi and I. A. Hiskens, "Model-Predictive Cascade Mitigation in Electric Power Systems With Storage and Renewables–Part II: Case-Study," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 78–87, Jan 2015.

[228] J. A. Martin and I. A. Hiskens, "Corrective Model-Predictive Control in Large Electric Power Systems," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1651–1662, March 2017.

[229] L. Jin, R. Kumar, and N. Elia, "Model Predictive Control-Based Real-Time Power System Protection Schemes," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 988–998, May 2010.

[230] P. W. Sauer, M. A. Pai, and J. H. Chow, *Power System Dynamics and Stability: With Synchrophasor Measurement and Power System Toolbox 2e*. John Wiley & Sons, Ltd, 2017.

[231] C. R. Rao, *Linear statistical inference and its applications*. John Wiley & Sons, 1973, vol. 22.

[232] Z.-F. Fu and J. He, *Modal Analysis*. Butterworth-Heinemann, 2001.

[233] J. J. Sanchez-Gasca, "Task Force on identification of electromechanical modes in power systems," IEEE, Tech. Rep., 2012.

[234] J. Quintero, V. Vittal, G. T. Heydt, and H. Zhang, "The Impact of Increased Penetration of Converter Control-Based Generators on Power System Modes of Oscillation," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2248–2256, Sept 2014.

[235] J. Hauer, D. Trudnowski, G. Rogers, B. Mittelstadt, W. Litzenberger, and J. Johnson, "Keeping an eye on power system dynamics," *IEEE Computer Applications in Power*, vol. 10, no. 4, pp. 50–54, Oct 1997.

[236] V. M. Venkatasubramanian and Y. Li, "Analysis of 1996 Western American Electric Blackouts," in *Bulk Power System Dynamics and Control*, Aug 2004, pp. 685–721.

[237] J. C. H. Peng and N. K. C. Nair, "Enhancing Kalman Filter for Tracking Ringdown Electromechanical Oscillations," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 1042–1050, May 2012.

[238] M. Yazdanian, A. Mehrizi-Sani, and M. Mojiri, "Estimation of Electromechanical Oscillation Parameters Using an Extended Kalman Filter," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 2994–3002, Nov 2015.

[239] H. M. Khalid and J. C. H. Peng, "Tracking Electromechanical Oscillations: An Enhanced Maximum-Likelihood Based Approach," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1799–1808, May 2016.

[240] J. F. Hauer, C. J. Demeure, and L. L. Scharf, "Initial results in Prony analysis of power system response signals," *IEEE Transactions on Power Systems*, vol. 5, no. 1, pp. 80–89, Feb 1990.

[241] J. W. Pierre, D. J. Trudnowski, and M. K. Donnelly, "Initial results in electromechanical mode identification from ambient data," *IEEE Transactions on Power Systems*, vol. 12, no. 3, pp. 1245–1251, Aug 1997.

[242] P. Kundur, "Small-Signal Stability," in *Power System Stability and Control*. McGraw-Hill, 1994, ch. 12, pp. 699–825.

[243] S. K. Starret, V. Vittal, A. A. Fouad, and W. Kliemann, "A methodology for the analysis of nonlinear, interarea interactions between power system natural modes of oscillation utilizing normal forms," in *Proc. Int. Symp. Nonlinear Theory Application*, vol. 2, Dec. 1993, pp. 523–538.

List of Acronyms

DMD - dynamic mode decomposition
DSE - dynamic state estimator
EDMD - extended dynamic mode decomposition
EKF - extended Kalman filter
EMS - energy management system
GM - generalized maximum-likelihood
GPS - global positioning system
IED - intelligent electronic device
IEKF - iterated extended Kalman filter
IPPC - iteratively reweighted phase-phase correlator
IRLS - iteratively reweighted least squares
ISO - independent system operator
KEF - Koopman eigenfunction
KF - Kalman filter
KKF - Koopman Kalman filter
KMD - Koopman mode decomposition
KNM - Kron-reduced network model
LAV - least absolute value
MAD - median-absolute-deviation
MCD - minimum covariance determinant
PDC - phasor data concentrator
PMU - phasor measurement unit
PPC - phase-phase correlator
PSS - power systems stabilizer
RTU - remote terminal unit
RTO - regional transmission organization
SCADA - supervisory control and data acquisition
SNR - signal-to-noise ratio
SSE - static state estimator
SVD - singular value decomposition
UKF - unscented Kalman filter
WAMS - wide-area measurement system

Appendix A

Models and data

A.1 Model A

Differential equations:

$$\dot{\delta} = \omega - \omega_s, \tag{A.1}$$

$$\dot{\omega} = \frac{\omega_s}{2H}\left(T_m - T_e - D\left(\omega - \omega_s\right)\right). \tag{A.2}$$

Algebraic equation:

$$P_e = \left(\frac{E}{X'_d}\right)\left(V_i \sin\left(\delta\right)\cos\left(\theta_i\right) - V_i \cos\left(\delta\right)\sin\left(\theta_i\right)\right), \tag{A.3}$$

where the subscript i denotes the i-th generator bus; δ and ω denote the generator rotor angle and rotor speed, respectively; T_m is the scaled mechanical torque on the generator shaft; ω_s, H and D are the generator rated synchronous speed, scaled shaft inertia constant and damping torque coefficient, respectively; P_e is the generator real electrical power output and in per unit, the generator electric torque output $T_e = P_e$; $Ee^{j\delta}$ represents the voltage behind the transient reactance X'_d, and $V_i e^{j\theta_i}$ is the voltage at the i-th generator bus.

The synchronous generator model defined by (A.1)-(A.3) is also known as the classical generator model. Equations (A.1)-(A.2) can be expressed as a single second order differential equation, in which case it is called the swing equation.

A.2 Model B

Differential equations:

$$T'_{do}\dot{E}'_q = E_{fd} - (X_d - X'_d)\, I_d - E'_q, \tag{A.4}$$
$$T'_{qo}\dot{E}'_d = \left(X_q - X'_q\right) I_q - E'_d, \tag{A.5}$$
$$\dot{\delta} = \omega - \omega_s, \tag{A.6}$$
$$\dot{\omega} = \frac{\omega_s}{2H}\left(T_m - T_e - D\left(\omega - \omega_s\right)\right), \tag{A.7}$$
$$T_E\dot{E}_{fd} = V_R - \left(K_E + S_E\left(E_{fd}\right)\right) E_{fd}, \tag{A.8}$$
$$T_F\dot{V}_F = \frac{K_F}{T_E}\left(V_R - \left(K_E + S_E\left(E_{fd}\right)\right) E_{fd}\right) - V_F, \tag{A.9}$$
$$T_A\dot{V}_R = K_A\left(V_{ref} - V_F - V\right) - V_R, \tag{A.10}$$
$$T_{CH}\dot{T}_M = P_{SV} - T_m, \tag{A.11}$$
$$T_{SV}\dot{P}_{SV} = P_C - \frac{1}{R_D}\left(\frac{\omega}{\omega_s} - 1\right) - P_{SV}. \tag{A.12}$$

Algebraic equations:

$$V_d = V\sin\left(\delta - \theta\right), \tag{A.13}$$
$$V_q = V\cos\left(\delta - \theta\right), \tag{A.14}$$
$$I_d = \frac{E'_q - V_q}{X'_d}, \tag{A.15}$$
$$I_q = \frac{V_d - E'_d}{X'_q}, \tag{A.16}$$
$$P_e = V_q I_q + V_d I_d, \tag{A.17}$$
$$Q_e = V_q I_d - V_d I_q. \tag{A.18}$$
$$\tag{A.19}$$

where E'_d, E'_q and E_{fd} are the generator d- and q-axis transient internal voltages and scaled field voltage, respectively; V_F and V_R are the scaled output of the exciter stabilizing transformer and amplifier, respectively; P_{SV} is the steam valve position; T'_{do}, T'_{qo}, T_E, T_F, T_A, T_{CH} and T_{SV} are time constants; K_E, K_F and K_A are controller gains; V_{ref} and P_C are adjustable control inputs; X_d, X'_d, X_q and X'_q are generator parameters; R_D is the speed regulation; $S_E\left(E_{fd}\right)$ accounts for the saturation of the exciter iron; Q_e is the generator reactive electrical power output; I_d and I_q are the d- and q-axis currents, whereas V_d and V_q are the d- and q-axis voltages. For more details, please refer to [230].

Model B includes the two-axis synchronous generator model, the IEEE type DC1A excitation system model, and the IEEE TGOV1 steam turbine-governor system model.

A.3 Two-area four-machine system

The power flow setpoint utilized to submit the two-area four-machine system to a stressed operating condition is presented in Table A.1. The remaining steady-state and dynamic data can be found in [213].

Table A.1: Power flow setpoint

Bus Number	Bus Type	Voltage PU	Gen. MW	Load MW	Load Mvar	Shunt Mvar
1	Swing	1.0400	--	--	--	--
2	PV	1.0530	660	--	--	--
3	PV	1.0510	510	--	--	--
4	PV	1.0500	510	--	--	--
5	PQ	1.0063	--	900	250	255.1
6	PQ	1.0021	--	1400	250	254.3

Appendix B

Error covariance matrix of the robust extended Kalman filter

It has been shown by Hampel [135] that the $\boldsymbol{IF}$ can be utilized to measure the sensitivity of an estimator to an infinitesimal contamination and to derive the error covariance matrix of an estimator under the assumed probability distribution, e.g., the Gaussian distribution. This approach is applied here for the GM-IEKF. Consider the ϵ-contamination model $G = (1-\epsilon)\,\Phi + \epsilon\Delta_r$, where Φ is the target distribution and Δ_r is the probability mass at r, and let the cumulative probability distribution of the residual vector $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{\varphi}(\boldsymbol{x})$ calculated from (4.17) be $\Phi(\boldsymbol{r})$. The GM-estimator provides an estimate of the state by processing the redundant observation vector $\boldsymbol{y}$ and solving the following implicit equation:

$$\sum_{i=1}^{m+n} \boldsymbol{\lambda}_i(\boldsymbol{r}, \boldsymbol{x}) = \sum_{i=1}^{m+n} \varpi_i \frac{\partial \boldsymbol{\varphi}(\boldsymbol{x})}{\partial \boldsymbol{x}} \psi(r_{s_i}) = \boldsymbol{0}, \tag{B.1}$$

which, by virtue of the Glivenko-Cantelli theorem [231], asymptotically tends to

$$\int \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T})\, dG = \boldsymbol{0}. \tag{B.2}$$

Here, the GM-estimator $\hat{\boldsymbol{x}}_{k|k}$ at G has been put in a functional form, $\boldsymbol{T}(G)$. The asymptotic total influence function of $\boldsymbol{T}(G)$ is given by

$$\boldsymbol{I}_{\boldsymbol{F}}(\boldsymbol{r}, \Phi) = \frac{\partial \boldsymbol{T}(G)}{\partial \epsilon}\Big|_{\epsilon=0} = \lim_{\Delta\epsilon \to 0} \frac{\boldsymbol{T}((1-\epsilon)\,\Phi + \epsilon\Delta r) - \boldsymbol{T}(\Phi)}{\epsilon}. \tag{B.3}$$

Substituting G into (B.2) yields

$$\int \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T}(G))\, d\Phi + \epsilon \int \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T}(G))\, d(\Delta_r - \Phi) = \boldsymbol{0}. \tag{B.4}$$

Taking the differentiation with respect to ϵ and evaluating it at $\epsilon = 0$, assuming regularity conditions and Fisher consistency at Φ, given by $\int \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T}(\Phi))\, d\Phi = \mathbf{0}$, yields

$$\frac{\partial}{\partial \epsilon} \int \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T}(G))\, d\Phi \Big|_{\epsilon=0} + \int \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T}(G))\, d(\Delta_r) \Big|_{\epsilon=0} = \mathbf{0}. \tag{B.5}$$

After applying the sifting property of the Dirac impulse to the second term, we obtain

$$\frac{\partial}{\partial \epsilon} \int \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T}(G))\, d\Phi \Big|_{\epsilon=0} + \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T}(\Phi)) = \mathbf{0}. \tag{B.6}$$

Assuming $\boldsymbol{\lambda}(\cdot)$ is continuous and measurable and $\boldsymbol{\lambda}'(\cdot)$ is measurable, we can apply the interchangeability of differentiation and integration theorem to the first term in (B.6), yielding

$$\int \frac{\partial \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T}(G))}{\partial \epsilon}\Big|_{\boldsymbol{T}(\Phi)} \frac{\partial \boldsymbol{T}(G)}{\partial \epsilon}\Big|_{\epsilon=0} d\Phi + \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T}(\Phi)) = \mathbf{0}. \tag{B.7}$$

Thus, $\boldsymbol{I_F}(\boldsymbol{r}, \Phi)$ is expressed as

$$\begin{aligned} \boldsymbol{I_F}(\boldsymbol{r}, \Phi) &= \frac{\partial \boldsymbol{T}(G)}{\partial \epsilon}\Big|_{\epsilon=0} \\ &= -\left[\int \frac{\partial \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T}(G))}{\partial \epsilon}\Big|_{\boldsymbol{T}(\Phi)} d\Phi\right]^{-1} \boldsymbol{\lambda}(\boldsymbol{r}, \boldsymbol{T}(\Phi)). \end{aligned} \tag{B.8}$$

Taking the derivative of $\boldsymbol{\lambda}(\cdot)$ with respect to $\boldsymbol{x}$ and assuming that ϖ and s are independent of $\boldsymbol{x}$, we obtain

$$\frac{\partial \boldsymbol{\lambda}(\boldsymbol{r})}{\partial \boldsymbol{x}} = \varpi \psi(r_{s_i}) \left[\frac{\partial^2 \boldsymbol{\varphi}(\boldsymbol{x})}{\partial x_i \partial x_j}\right] + \varpi \left[\frac{\partial \psi(r_{s_i})}{\partial \boldsymbol{x}}\right] \left[\frac{\partial \boldsymbol{\varphi}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^\top. \tag{B.9}$$

By neglecting the second-order term on the right-hand side of (B.9), we get

$$\frac{\partial \boldsymbol{\lambda}(\boldsymbol{r})}{\partial \boldsymbol{x}} = \varpi \left[\frac{\partial \psi(r_{s_i})}{\partial \boldsymbol{x}}\right] \left[\frac{\partial \boldsymbol{\varphi}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^\top. \tag{B.10}$$

Applying the chain rule to the derivative of $\psi(r_{s_i})$ yields

$$\frac{\partial \boldsymbol{\lambda}(\boldsymbol{r})}{\partial \boldsymbol{x}} = -\frac{1}{s} \psi'(r_{s_i}) \left[\frac{\partial \boldsymbol{\varphi}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right] \left[\frac{\partial \boldsymbol{\varphi}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^\top. \tag{B.11}$$

By substituting (4.17), (B.1) and (B.11) into (B.8), we get

$$\boldsymbol{I_F}(\boldsymbol{r}, \Phi) = \left[\int \frac{1}{s} \psi'(r_{s_i})\, \boldsymbol{C}\boldsymbol{C}^\top \Big|_{T(\Phi)} d\Phi\right]^{-1} \varpi \boldsymbol{C} \psi(r_{s_i}). \tag{B.12}$$

Thus, the state estimation error covariance matrix for the GM-IEKF at time step k is given by

$$\begin{aligned} \boldsymbol{\Sigma}_{k|k} &= E\left[\boldsymbol{I_F} \cdot \boldsymbol{I}\boldsymbol{F}^\top\right] \\ &= \frac{E_\Phi\left[\psi^2(r_{s_i})\right]}{\left\{E_\Phi\left[\psi'(r_{s_i})\right]\right\}^2} \left(\boldsymbol{C}_k^\top \boldsymbol{C}_k\right)^{-1} \left(\boldsymbol{C}_k^\top \boldsymbol{Q}_\varpi \boldsymbol{C}_k\right) \left(\boldsymbol{C}_k^\top \boldsymbol{C}_k\right)^{-1}, \end{aligned} \tag{B.13}$$

where $\boldsymbol{Q}_\varpi = \operatorname{diag}(\varpi_i^2)$ and $\boldsymbol{C}_k = {}^{\partial \boldsymbol{\varphi}}/_{\partial \boldsymbol{x}}\Big|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_{k|k}}$.

Appendix C

Linear modal analysis of electric power systems

According to [232], "modal analysis is the process of determining the inherent dynamic characteristics of any system". The Koopman operator-theoretic framework, which is the central object of this dissertation, can be cast as a modal analysis tool in this sense. This "tool" based on the Koopman operator is nonetheless significantly more powerful than others that have been applied to electric power systems. In what follows, we will briefly touch upon this topic of modal analysis with particular interest in power systems oscillations. Then, we will introduce and discuss two classical techniques, one based on the linearization of the differential-algebraic equations, and another data-driven approach called Prony method.

Monitoring and controlling of oscillations is an important problem in modern power systems; see for example [233]. Among others, this is specially due to

- the presence of high-gain fast acting excitation systems and higher bandwidth speed governor systems, both well-known contributors to reducing the damping torque of synchronous generators;
- larger power transfers across the grid, which often decrease the damping of inter-area modes; and
- the increasing use of renewable energy with uncontrolled prime-mover sources, which adds to the challenges of monitoring oscillatory stability [234].

Although PSS control systems are usually applied to improve the damping of inter-area oscillations, they often have limited controllability of these modes, thereby mitigating the problem of oscillatory instability only to a certain extent. Therefore, the monitoring of the electromechanical modes of oscillation is a mission critical task in regional power system control centers. The earliest the detection of poorly damped oscillations occurs, the largest the time is available for taking control actions, ultimately reducing the risk of a blackout such as the one that occurred in the U.S. West Coast on August 10, 1996 [235, 236]. The

methods used to identify oscillations are typically classified as model- or measurement-based methods [233], although there exist hybrid techniques such as the ones based on the Kalman filters [237–239].

The model-based methods typically rely on eigenanalysis of the system matrix that is built after the linearization of the differential-algebraic equations governing the system, and is carried out around an equilibrium point. All the natural linear modes can be calculated for a given system topology and operating point, which makes this method suitable for off-line planning and control design. However, this method is unable to reveal the nonlinear modes of oscillation that occur when the system is temporarily pulled away far from equilibrium by a major disturbance, even when these modes become dominant. Furthermore, model uncertainties and parameters errors might induce biased results. Finally, this approach is incompatible with real-time applications because it is computationally intensive for very large systems, especially when the synchronous generators and their controls are modeled in details.

By contrast, measurement-based methods are not limited by system size, but their performance depends on the quality of the data samples that are available to the users. Furthermore, only the excited modes that are observable in the measurement set can be identified; the ones embedded in noisy signals or well damped might not be easily detected [233]. These methods have received recently a great deal of attention among practitioners and researchers in power systems due to the widespread deployment of PMUs. Measurement-based methods might be further classified as [233]:

- *linear ringdown methods*, typically used to process ringdown signals (the term ringdown refers to signals that arise following a transient event). One example of such a technique is the Prony method [240];
- *mode-meter methods*, which are tailored to ambient data, i.e., when the system is in a quasi steady-state condition with the primary excitation being the random variations caused by the switching of not large loads. One example here is the Yule Walker-based algorithms [241];
- *nonlinear and non-stationary methods*, group to which the method based on the Koopman operator might be included.

C.1 Linearization of the nonlinear differential-algebraic equations

The following theory is based on [26,242]. The dynamics of power systems might be described by a set of n first order nonlinear differential equations, such as:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}\left(\boldsymbol{x}, \boldsymbol{u}\right), \tag{C.1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the state vector, $\boldsymbol{u} \in \mathbb{R}^p$ is the input vector, and $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ is the vector-valued nonlinear system function.

In practice, we are often interested in output variables which can be observed on the system. In other words, we want to represent mathematically the variables which we can measure in the field. These are usually expressed as:

$$\boldsymbol{y} = \boldsymbol{h}\left(\boldsymbol{x}, \boldsymbol{u}\right), \tag{C.2}$$

where $\boldsymbol{y} \in \mathbb{R}^m$ is the vector of outputs, and $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ is the vector-valued nonlinear output function.

The state space representation of the system is therefore written as:

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{f}\left(\boldsymbol{x}, \boldsymbol{u}\right), \\ \boldsymbol{y} &= \boldsymbol{h}\left(\boldsymbol{x}, \boldsymbol{u}\right). \end{aligned} \tag{C.3}$$

Eq. (C.3) might be defined for any deterministic, continuous, nonlinear, time-invariant dynamical system. Now, assuming that the perturbations to which this system is submitted are sufficient small, its dynamic response can be modeled in terms of Taylor's series expansion, with second and higher order terms neglected. After some algebraic manipulation, we might find:

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{B}\,\boldsymbol{u}, \\ \boldsymbol{y} &= \boldsymbol{C}\,\boldsymbol{x} + \boldsymbol{D}\,\boldsymbol{u}, \end{aligned} \tag{C.4}$$

where $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is the state matrix, $\boldsymbol{B} \in \mathbb{R}^{n \times p}$ is the input matrix, $\boldsymbol{C} \in \mathbb{R}^{m \times n}$ is the output matrix, and $\boldsymbol{D} \in \mathbb{R}^{m \times p}$ is the feed-forward matrix. For more details, see for example [242].

Matrices $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ hold important information about the system. Some of their important properties, specially for control theory, were put forth by Kalman in the 1960s. For now, let us focus on the system state matrix $\boldsymbol{A}$, which is a Jacobian matrix. For small perturbations, the stability of the system is given by the eigenvalues of $\boldsymbol{A}$, which is referred to as small signal stability analysis. To compute its eigenvalues for large systems, QR-factorization and algorithms based on Krylov subspace have been recommended due to their numerical stability and ability to deal with very large matrices in an efficient way. Formally, one has

$$\boldsymbol{A}\boldsymbol{r}_i = \lambda_i \boldsymbol{r}_i, \qquad i = 1, ..., n, \tag{C.5}$$

where $\boldsymbol{r}_i$ is the *right eigenvector* associated with the i-th eigenvalue λ_i of the matrix $\boldsymbol{A}$. In a similar way, the row vector $\boldsymbol{\ell}_i^\top$ which satisfies

$$\boldsymbol{\ell}_i^\top \boldsymbol{A} = \lambda_i \boldsymbol{\ell}_i^\top, \qquad i = 1, ..., n, \tag{C.6}$$

is the *left eigenvector* associated with the i-th eigenvalue λ_i. Next, some important properties are highlighted:

- left and right eigenvectors associated with different eigenvalues are orthogonal. Formally, we have $\boldsymbol{\ell}_i^\top \boldsymbol{r}_j = 0$ for all $i \neq j$;
- left and right eigenvectors associated with the same eigenvalue imply $\boldsymbol{\ell}_i^\top \boldsymbol{r}_i = c_i$, where c_i is a non-zero constant, which is equal to 1 if the eigenvectors are normalized.

Finally, let us initiate the derivation to explicitly identify the modes of oscillation of the system. For this purpose, define the following modal matrices, all square of dimension n:

$$\boldsymbol{U} = [\boldsymbol{r}_1 \;\; \boldsymbol{r}_2 \;\; ... \;\; \boldsymbol{r}_n], \tag{C.7}$$

$$\boldsymbol{V} = [\boldsymbol{\ell}_1 \;\; \boldsymbol{\ell}_2 \;\; ... \;\; \boldsymbol{\ell}_n]^\top, \tag{C.8}$$

$$\boldsymbol{\Lambda} = \text{diag}\,(\lambda_1, \lambda_2, ..., \lambda_n). \tag{C.9}$$

The following relations are important. Note that, from this point forward, we are assuming that the eigenvectors are normalized.

$$\boldsymbol{A}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{\Lambda}, \tag{C.10}$$

$$\boldsymbol{V}\boldsymbol{U} = \boldsymbol{I}, \tag{C.11}$$

$$\boldsymbol{V} = \boldsymbol{U}^{-1}, \tag{C.12}$$

$$\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{\Lambda}. \tag{C.13}$$

Eq. (C.4) can be rewritten as $\dot{\boldsymbol{x}} = \boldsymbol{A}_{sys}\,\boldsymbol{x} + \boldsymbol{E}\,\boldsymbol{u}$, after some algebraic manipulations. Please refer to [230] for details. Now, assuming that no input is applied to the system,

$$\dot{\boldsymbol{x}} = \boldsymbol{A}_{sys}\,\boldsymbol{x}. \tag{C.14}$$

The relationship (C.14) shows that the rate of change of each state variable is a linear combination of all state variables. Indeed, for power systems, $\boldsymbol{A}_{sys}$ is sparse, implying that the rate of change of a state variable is a linear combination of a subset of the system state variables. Now, the question that arises is how to solve (C.14). Assuming that all the eigenvalues λ_i of $\boldsymbol{A}_{sys}$ are distinct, define the similarity transformation

$$\boldsymbol{x} = \boldsymbol{U}\boldsymbol{z}. \tag{C.15}$$

Substituting (C.15) into (C.14), one has

$$\begin{aligned} \boldsymbol{U}\dot{\boldsymbol{z}} &= \boldsymbol{A}_{sys}\boldsymbol{U}\boldsymbol{z} \\ \dot{\boldsymbol{z}} &= \boldsymbol{U}^{-1}\boldsymbol{A}_{sys}\boldsymbol{U}\boldsymbol{z} \\ \dot{\boldsymbol{z}} &= \boldsymbol{\Lambda}\boldsymbol{z}. \end{aligned} \tag{C.16}$$

Note that the matrix $\boldsymbol{\Lambda}$ is diagonal. Therefore, the states are now completely decoupled in terms of the new state vector $\boldsymbol{z}$, which is related to the original state vector $\boldsymbol{x}$ through

(C.15). The time-domain solution of system of first-order differential equations defined by (C.16) is given by

$$z_i(t) = z_i(0)\, e^{\lambda_i t} \qquad i = 1, ..., n. \tag{C.17}$$

Expanding (C.15) and using the relation expressed by (C.17) yields

$$\begin{aligned} \boldsymbol{x}(t) &= \boldsymbol{U}\boldsymbol{z}(t) \\ &= [\boldsymbol{r}_1 \;\; \boldsymbol{r}_2 \;\; ... \;\; \boldsymbol{r}_n] \begin{bmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_n(t) \end{bmatrix} \\ &= \sum_{i=1}^{n} \boldsymbol{r}_i z_i(t) \\ &= \sum_{i=1}^{n} \boldsymbol{r}_i z_i(0)\, e^{\lambda_i t}. \end{aligned} \tag{C.18}$$

Again from (C.15), we know that

$$\boldsymbol{z}(t) = \boldsymbol{U}^{-1}\boldsymbol{x}(t) = \boldsymbol{V}\;\boldsymbol{x}(t). \tag{C.19}$$

Therefore, for each element of the vector $\boldsymbol{z}$ we have

$$z_i(t) = \boldsymbol{\ell}_i^{\top}\boldsymbol{x}(t), \tag{C.20}$$

and for $t = 0$,

$$z_i(0) = \boldsymbol{\ell}_i^{\top}\boldsymbol{x}(0). \tag{C.21}$$

Substituting (C.21) into (C.18) we get

$$\boldsymbol{x}(t) = \sum_{j=1}^{n} \left(\boldsymbol{\ell}_j^{\top}\boldsymbol{x}(0)\right) \boldsymbol{r}_j e^{\lambda_j t}. \tag{C.22}$$

Finally, the time-domain response of the i-th state variable is given by

$$x_i(t) = \underbrace{\left(\ell_{i1}^{\top}x(0)\right) r_{i1} e^{\lambda_1 t}}_{\text{linear mode 1}} + \underbrace{\left(\ell_{i2}^{\top}x(0)\right) r_{i2} e^{\lambda_2 t}}_{\text{linear mode 2}} + ... + \underbrace{\left(\ell_{in}^{\top}x(0)\right) r_{in} e^{\lambda_n t}}_{\text{linear mode } n}. \tag{C.23}$$

Thus, the free motion time-domain response of the system is given by (C.23), which express $x_i(t)$ in terms of eigenvalues, right and left eigenvectors, and system initial condition.

C.2 Linear contribution and participation factors

The time evolution of the i-th state in (C.23) is written as

$$x_i(t) = \sum_{j=1}^{n} \left(\boldsymbol{\ell}_j^\top \boldsymbol{x}_0\right) r_{ij} e^{\lambda_j t} = \sum_{j=1}^{n} \sigma_{ij} e^{\lambda_j t}, \tag{C.24}$$

where $\boldsymbol{x}_0 = \boldsymbol{x}(0)$.

The contribution factor [211, 243] is defined as

$$\sigma_{ij} = \left(\boldsymbol{\ell}_j^\top \boldsymbol{x}_0\right) r_{ij}. \tag{C.25}$$

It is as a measure of the contribution of mode j to the oscillations of state i for a given disturbance. Note that the contribution factors are dependent on the initial condition $\boldsymbol{x}_0$, and therefore dependent on the location and duration of the disturbance, as well as the system topology and operating point. The nonlinear counterpart of (C.25) based on the Koopman operator-theoretic framework has been pinpointed by Susuki and Mezić in [64], although the terminology contribution factor has not been used. Whereas the contribution factors provide a valuable information, a measure of the system performance that depends only on the system topology and operating point, and not on the disturbance location or duration, is advantageous for capturing inherent system dynamics. By realizing this fact, Pérez-Arriaga et al. [194] take the contribution factors in (C.25) and select an initial condition $\boldsymbol{x}_0 = \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is the unit vector along the i-th coordinate axis, thereby defining the linear mode-in-state participation factors as a measure of the magnitude of the modal oscillations in a state when only that state is perturbed. Formally, we have

$$p_{ij} = \ell_{ij} r_{ij}. \tag{C.26}$$

Now, by substituting (C.15) into (C.14), $\dot{\boldsymbol{z}} = \boldsymbol{V}\boldsymbol{A}_{sys}\boldsymbol{U}\boldsymbol{z} = \boldsymbol{\Lambda}\boldsymbol{z}$, and the time evolution of the i-th modal variable $\boldsymbol{z}$ is given by

$$z_i = z_{i0} e^{\lambda_i t} = \boldsymbol{\ell}_i^\top \boldsymbol{x}_0 e^{\lambda_i t}. \tag{C.27}$$

Pérez-Arriaga et al. [194] select $\boldsymbol{x}_0 = \boldsymbol{r}_i$, which leads to

$$z_i = \left(\sum_{j=1}^{n} \ell_{ij} r_{ij}\right) e^{\lambda_i t}. \tag{C.28}$$

From (C.28), the linear state-in-mode participation factors are defined as a measure of the relative participation of the j-th state in the i-th mode. Formally, we have

$$p_{ij} = \ell_{ij} r_{ij}. \tag{C.29}$$

Note that as proposed by Pérez-Arriaga et al. [194], (C.26) and (C.29) are identical.