NONLINEAR KALMAN FILTERS FOR DYNAMIC ESTIMATION OVER COMPLEX NETWORKS

WANLI WANG

PhD

The Hong Kong Polytechnic University

2021

The Hong Kong Polytechnic University

Department of Electronic and Information Engineering

Nonlinear Kalman Filters for Dynamic Estimation over Complex Networks

Wanli Wang

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

June 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

_____(Signed)

Wanli Wang (Name of student)

Abstract

This thesis studies the applications of nonlinear Kalman filters in solving problems associated with the dynamic estimation of networked systems such as epidemic tracking over networks and monitoring the operation of power networks in real time. In particular, the epidemic spreading on networks is concerned with controlling morbidity. A compartmental model is in general utilized for describing epidemic transmission on networks. The compartmental model, however, is inadequate for describing the dynamics of epidemic spreading without considering measurements from transmission processes. A nonlinear Kalman filter can be utilized for solving this problem by considering the inherent dynamic model together with epidemic transmission processes. It is, however, non-trivial to choose appropriate nonlinear Kalman filters for epidemic tracking over various networks, such as the Erdös and Rényi (ER) network, the Newman and Watts (NW) network, and the Watts and Strogatz (WS) network. A guideline will be provided for choosing traditional nonlinear Kalman filters for studying epidemic spreading on commonly used complex networks. Specifically, epidemic spreading on networks is described by compartmental models, such as Susceptible-Infected-Recovered, Susceptible-Infected-Susceptible, and Susceptible-Infected-Recovered-Susceptible models. The dynamic study of epidemic spreading on various homogeneous networks is performed using nonlinear Kalman filters, including the extended Kalman filter (EKF), the unscented Kalman filter(UKF), and the cubature Kalman filter (CKF). Various traditional Kalman filters are compared in terms of accuracy, stability, and complexity. These traditional Kalman

filters are, however, based on the optimization of the minimum mean square error. As a result, these nonlinear Kalman filters may have degraded filtering precision when available measurements are corrupted by non-Gaussian noise. For solving this issue, a novel generalized correntropy sparse Gauss-Hermite quadrature filter is proposed by combining the generalized correntropy with the sparse Gauss-Hermite quadrature filter. Dependent on the Susceptible-Infected-Recovered-Susceptible compartmental model, the proposed generalized correntropy sparse Gauss-Hermite quadrature filter is applied to tracking epidemic spreading on homogeneous networks in the presence of non-Gaussian noise. In addition, the dynamic estimation of power systems is studied with the aim of enhancing the operation of power distribution infrastructure. Since the available measurements may be corrupted by non-Gaussian noise, a robust mixed p-norm square root unscented Kalman filter is proposed for estimating the state of power systems in the presence of non-Gaussian noise. The mixed p-norm square root unscented Kalman filter applies a mixed p-norm for weighting measurement errors for robustness improvement. Furthermore, unlike the generalized correntropy sparse Gauss-Hermite quadrature filter, the mixed p-norm square root unscented Kalman filter utilizes a piecewise function, i.e., multiple p-norms for handling varying measurements. As a result, the mixed p-norm square root unscented Kalman filter is more flexible in dealing with corrupted measurements but may be confronted with the choice of multiple parameters in comparison with the generalized correntropy sparse Gauss-Hermite quadrature filter. Simulation results demonstrate the efficiency of the p-norm square root unscented Kalman filter in the WSCC (Western System Coordinating Council) 3-machine system and the NPCC (Northeastern Power Coordinating Council) 48-machine system.

List of Publications

Journal papers

- **W. Wang**, C. K. Tse, and S. Wang, "Generalized correntropy sparse Gauss-Hermite quadrature filter for epidemic tracking on complex networks," *IEEE Transactions on Systems, Man and Cybernetics: Systems*, Feb. 2021, doi: 10.1109/TSMC.2021.3052538.

- **W. Wang**, C. K. Tse, and S. Wang, "Dynamic state estimation of power systems by p -norm nonlinear Kalman filter," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 5, pp. 1715–1728, May 2020.

- **W. Wang**, C. K. Tse, and S. Wang, "Performance comparison of nonlinear Kalman filters in epidemic tracking on networks," *IEEE Systems Journal*, vol. 14, no. 4, pp. 5475–5485, Dec. 2020.

- **W. Wang**, C. K. Tse, F. C. M. Lau, and S. Wang, "Smooth deep reinforcement learning for power control for spectrum sharing in cognitive radios," *IEEE Transactions on Wireless Communications*, second review.

- **W. Wang**, C. K. Tse, F. C. M. Lau, and S. Wang, "Square root unscented Kalman filter with modified measurement for dynamic state estimation of power systems," to be submitted.

Acknowledgments

In the course of writing this thesis, I have received a great deal of support and assistance. I would like to thank the following people, without whom I would not have been able to complete the writing of this thesis, and without whom I would not have finished my Ph.D. study.

First and foremost, I am extremely grateful to my supervisors, Prof. C. K. Michael Tse and Prof. Francis C. M. Lau for their support of my Ph.D. study and for their immense enthusiasm demonstrated in the course of this study. In addition, they have encouraged me to tackle significant and challenging research problems. Their insightful feedback has motivated me to sharpen my thinking and brought my work to a higher level. It is hard to imagine if I could finish my Ph.D. study without their help.

I have learnt a lot from them, including developing novel ideas and working hard to implement them. Our close collaboration has led to the realization of novel ideas in the important topic of dynamic estimation of nonlinear systems over complex networks.

It is my honor to be a member of the Nonlinear Circuits and Systems Group and to collaborate with other members of the group. The Nonlinear Circuits and Systems Group makes me feel being in a family as we share our experience and help each other. I appreciate their kindness and company during my Ph.D. study.

I would like to thank The Hong Kong Polytechnic University for providing me with a chance of studying in Hong Kong. I also appreciate the help from my friends at The Hong Kong Polytechnic University. They offered great help for my life.

I would like to thank my family for encouraging me in all of my pursuits and

inspiring me to follow my dreams. I am especially grateful to my parents, who support me and always believe in me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

A complex system consists of a large number of components interacting with each other. The emergence of network science has provided a network representation of a complex system. In network science, these components are modeled as nodes and the corresponding interactions among components are modeled as links, forming a network or graph [1].

Network science has originated from a branch of mathematics referred to as graph theory. In particular, the concept of graphs was first considered for solving a real-life problem which has been known as the Seven Bridges of Königsberg in 1736. Königsberg was a city in Prussia and contained four lands. These four lands were separated by a river named Pregel and connected by seven bridges. A practical question arouse as to whether there was a solution of walking through the city crossing each of the seven bridges once and only once. Euler proved that the problem could not be solved by formulating it in abstract mathematical terms. The four landmasses and seven bridges were abstracted as four nodes and seven links, respectively. As a result, the physical structure was transformed to a graph which allowed the solution to be derived formally, i.e., an even number of links connected to each node being necessary

and sufficient for touring the whole city crossing each bridge exactly once.

There exist well-established network or graph models for studying the structures and characteristics of real-world networks on the basis of network science. The commonly used network models include the Erdös and Rényi (ER) network [2, 3], the Newman and Watts (NW) network [4, 5], and the Watts and Strogatz (WS) network [3, 6]. Paul Erdös and Alfréd Rényi introduced an algorithm for generating random graphs in 1959 [2]. In particular, the Erdös and Rényi (ER) network is generated by adding links according to a connecting probability p sequentially [3]. The connecting probability p is applied for determining whether a pair of nodes is connected. For example, a pair of nodes in an ER network is connected when a randomly generated number is smaller than the pre-determined connecting probability p in the evolution of the ER network. Degree and the corresponding degree distribution are two important concepts [7]. In particular, degree d_j and corresponding degree distribution $P(d_j)$ denote the number of nodes connected with node j and the probability that a node has degree d_j, respectively. A Poisson node degree distribution has been observed for ER random networks. The ER random network is commonly used to represent a complex network due to the simplicity of the ER random network.

The ER random network, however, fails to analyze real-world systems since it is constructed by randomly adding links among nodes. This motivated the generation of special networks like small-world networks [3] and scale-free networks [8] in the late 1990s. In 1967, a small-world phenomenon was first found in the "six degrees of separation" experiment conducted by Stanley Milgram [2]. This experiment aimed to find the average shortest path length for the American social network. More concretely, this experiment considered two persons randomly invited from Sharon and Boston as two target points and two groups of volunteers from Kansas and Nebraska as start points, respectively. Each volunteer was asked to send a letter to friends until the letter was received by the target points. As shown in the experiment, the average path length was 5.2 hops. This means that the letter could be received from the starting

point to the target point via 5.2 intermediate friends on average. This was referred to as the small-world phenomenon in a social network. Watts and Strogatz in 1998 identified a small-world network by analogy with the small-world phenomenon [9]. The Watts and Strogatz's small-world network starts with a nearest-neighbor coupled network consisting of N nodes arranged in a ring where each node is adjacent to its neighbor nodes. A rewired procedure was then considered in the WS small-world network to rewire the edges of the network with a probability p [3]. It has been found that under an intermediate connection probability p, the rewired graph is a small-world network, which is highly clustered similar to a regular graph and has a small shortest path length, like a random graph. In addition, an alternative small-world model was proposed by Newman and Watts (NW). In the NW model, one does not break any connection between any two nearest neighbors, but instead, adds a connection between a pair of nodes with a probability p. The NW model reduces to the original nearest-neighbor coupled network with $p = 0$, while it becomes a globally coupled network with $p = 1$.

A power-law degree distribution in the World Wide Web was found by Albert *et al.* in 1999 [10]. In the World Wide Web, the Uniform Resource Locators (URLs) and hypertext links are formulated as nodes and links, respectively. As a result, the network can be characterized as a scale-free network, the degree distribution of which follows a power-law [11]. This means that a large majority of nodes have low degree and only a small number of nodes have relatively high degree. Barabási and Albert introduced a procedure for generating a scale-free network [8, 12].

Since then, a series of applications with respect to the dynamics of information flow over various networks have been studied, e.g., epidemic spreading on networks [13], dynamic analysis of power systems [14] and so on. Different methods like machine learning [15], dynamic model-based methods [16, 17] and nonlinear Kalman filtering [18] are utilized for studying the dynamics of information flow. In particular, machine learning-based approaches study dynamics by learning an unknown nonlinear

function. This method has the advantage of not requiring an epidemic transmission model on networks but requires a large number of samples for yielding an efficient training result.

In comparison with the machine learning-based approach, dynamic model-based method is dependent on a specific model for describing the dynamics of information flow. The dynamics of information flow are mainly categorized into the propagation dynamics and the dynamics of nonlinear systems. The epidemic spreading on networks and dynamic analysis of power systems are two typical applications for considering the propagation dynamics and the dynamics of nonlinear systems, respectively. In particular, the propagation dynamics focus on describing the transmission dynamics in applications such as the epidemic spreading in social networks [19], rumor propagation in computer networks [20, 21], and virus spreading in wireless sensor networks [22]. The propagation dynamic models are adjustable and a great deal of propagation dynamics models are presented for approximating the characteristics of transmission processes as precisely as possible. The control of epidemic spreading, for example, considers a commonly used compartmental model [19] for describing epidemic propagation on networks. The compartmental model includes Susceptible-Infected-Recovered (SIR) [23, 16], Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Recovered-Susceptible (SIRS) [17]. For the SIR model, the entire population is divided into different groups, i.e., S, I, and R which denote the relative sizes or proportions of the susceptible, infected, and recovered individuals. By contrast, the dynamics of nonlinear systems focus on describing the uncertainty of nonlinear systems. The dynamic model with respect to the dynamics of nonlinear systems is in general a well-established model according to specific characteristics. In the dynamic analysis of power systems [24, 25, 18, 26, 27], for example, a power system operates under normal conditions within a set of pre-designed system parameters, which is appropriate to the ratings of the power system elements. A well-established model is constructed according to the electrical characteristics of power systems. However, it

has been found inadequate for only using the propagation dynamics or the dynamics of nonlinear systems for the dynamic study of information flow. Like in the application of power control, for example, system parameters may remain unchanged at the pre-designed values. When the power system suffers from varying system faults such as short-circuit faults [28], these system parameters may diverge from the normal values. The traditional dynamic model of the power system has been found inadequate for studying its dynamics especially under fault conditions.

For solving this issue, machine learning is introduced in the dynamic model where practical data are used for learning the factors of the dynamic model for fitting the transmission process in practice [29]. This combined method is indeed beneficial for studying dynamics but requires a large number of samples due to the introduction of machine learning. In addition, it is difficult or even impossible to learn the parameters of the dynamic model effectively in a highly nonlinear model. By contrast, nonlinear Kalman filters [30, 31, 32] are effective for dynamic estimation with respect to information flow through a network. More concretely, nonlinear Kalman filters are used to perform dynamic estimation on the basis of a state space model consisting of a state function and a measurement function. The state function is used to describe the inherent dynamics of information flow without considering measurements. Therefore, the propagation dynamics or the dynamics of nonlinear systems demonstrated above can be used as the state function. In addition, a well-established measurement function is further introduced for providing available measurements for further correcting the predictive state obtained through the state function [33, 34]. In the application of epidemic tracking, for example, nonlinear Kalman filters have been shown to be able to capture the dynamics of epidemic spreading by considering the transmission model and measurements obtained from epidemic propagation simultaneously. Different from the aforementioned combined method, nonlinear Kalman filters study dynamics with no need of a large number of samples. This also means that nonlinear Kalman filters can perform epidemic tracking with stationary model parameters in comparison with

the aforementioned combined method which has to train model parameters all the time. In order to improve tracking accuracy, there also exist double nonlinear Kalman filters which estimate the state and model parameters alternatively [30]. In addition, the aforementioned combined method is a determined method which only provides estimation. By contrast, nonlinear Kalman filters are derived in the framework of the Bayesian theory which not only provides estimation but also the corresponding covariance. The estimation is effectively updated towards the expected value with the help of the corresponding error covariance. However, there are some cons of nonlinear Kalman filters in studying dynamics. For example, nonlinear Kalman filters with high accuracy are expected to be employed for studying the dynamics of information flow. However, these nonlinear Kalman filters in general have relatively high computational burden than the aforementioned combined method.

In general, nonlinear Kalman filters are categorized into global and local methods. Without specific assumption with respect to the probability density functions of the state variables, global approaches, e.g., the grid-based method [35], the particle filter [36, 37] and the Gaussian mixture filter [38], can achieve more accurate estimates at the cost of increasing computational complexity. The local approaches have been introduced to reduce the computational burden imposed by global approaches. The local approach is related on the application of aiming at minimizing the mean square error in the presence of Gaussian noise. The extended Kalman filter (EKF) is one of the most commonly used local methods by virtue of its simple first-order linearization of a nonlinear function [39, 40]. However, the extended Kalman filter cannot achieve the desirable filtering accuracy when the nonlinearity is severe. A second-order EKF [41, 42] is presented for addressing this issue. The high computational burden on calculating the Jacobian and Hessian matrices, however, greatly limits the application of the second-order EKF. Apart from the second-order EKF, various rules have also been proposed for approximating the Gaussian weighted integrals required in nonlinear Kalman filters. For example, the unscented Kalman filter (UKF)

calculates multiple integrals through a certain set of weighted sigma points using the unscented transformation (UT) [43, 44]. However, the UKF may be confronted with the choice of an optimal scaling parameter [28]. Nonlinear Kalman filters without the need of the scaling parameter have been proposed, e.g., the cubature Kalman filter (CKF) based on the spherical-radial rule [45, 46] and simplex spherical-radial cubature Kalman filter (SSRCKF) based on the simplex spherical-radial rule [47]. In comparison with the EKF employing the first-order linearization, UKF, CKF and SSRCKF achieve better filtering accuracy for nonlinear systems at the expense of increased computational burden. In addition, numerous robust Kalman filters like the Huber-based unscented Kalman filter (HUKF) and the maximum correntropy criterion-based unscented Kalman filter (MCCUKF) have also been proposed for improving robustness against outliers.

1.2 Motivation

Nonlinear Kalman filters have emerged as a powerful platform for dynamic estimation based on a state model and a measurement model. The state model is either classified as a propagation dynamic model or a nonlinear system model. The propagation dynamic model is constructed for matching the characteristics of transmission processes. By contrast, the nonlinear system model is generally developed for describing specific characteristics of the given deterministic nonlinear system. The epidemic tracking on networks and the dynamic analysis of power networks, for example, are two representative applications involving the propagation (or spreading) dynamics and the dynamics of nonlinear systems, and the representative problems will be considered in this thesis.

In the application of the epidemic tracking on networks, the estimation accuracy with respect to the epidemic tracking is determined by several factors including the type of the complex network model, the form of the state model describing the

epidemic spreading, and the type of nonlinear Kalman filters used. In particular, the profile of transmission or disease spreading on networks is greatly dependent on the type of networks, e.g., the Erdös and Rényi (ER) network, the Newman and Watts (NW) network, and the Watts and Strogatz (WS) network. The constructed state model is expected to be matched well with the transmission profile on networks. The state model in the application of epidemic tracking on networks usually adopts a compartmental model for studying the dynamic propagation of the epidemic spreading on networks. There exist numerous compartmental models, such as Susceptible-Infected-Recovered (SIR), Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Recovered-Susceptible (SIRS). Dependent on the chosen state model and collected measurements, various types of Kalman filters can be used for epidemic tracking on different networks. Commonly used nonlinear Kalman filters include the extended Kalman filter (EKF), the unscented Kalman filter (UKF) and the cubature Kalman filter (CKF). The EKF performs state estimation by applying a first-order linearization to a nonlinear function. In comparison with the EKF, the UKF uses the unscented transformation (UT) for capturing the statistical characteristics of a Gaussian variable through a nonlinear function. In the unscented Kalman filter, an adjusted scaling parameter is introduced for controlling the spreading of samples. The unscented Kalman filter reduces to the third-degree CKF with the appropriate choice of the adjusted scaling parameter. In addition, the fifth-degree CKF [48] is proposed for further improving the filtering performance in comparison with other nonlinear Kalman filters. It is non-trivial to choose appropriate nonlinear Kalman filters with multiple compartmental models for tracking epidemic spreading on different kinds of networks. A concerned question arises as to whether there is a guideline for choosing appropriate Kalman filters with different compartment models for epidemic tracking on various networks, which is considered as the first work in this thesis.

Nonlinear Kalman filters perform epidemic tracking on the basis of a state model and a set of collected measurements. Studies have demonstrated the filtering efficiency

of traditional nonlinear Kalman filters in tracking epidemic spreading when the measurements are corrupted by Gaussian noise. These nonlinear Kalman filters, however, cannot adequately handle measurements contaminated by non-Gaussian noise since these Kalman filters perform optimization based on the minimum mean square error criterion. Several attempts have been made to overcome this issue, like the Huber-based unscented Kalman filter (HUKF) [49] and maixmum correntropy criterion (MCC)-based unscented Kalman filter [50]. These nonlinear Kalman filters indeed achieve improved robustness against non-Gaussian noise in comparison with traditional Kalman filters, but have limited precision improvement. This motivates us to propose a more robust nonlinear Kalman filter for handling epidemic spreading in the presence of non-Gaussian noise. The generalized correntropy is a preferable choice for dealing with non-Gaussian noise in comparison with the maximum correntropy criterion in the framework of information theoretical learning [51]. In addition, the sparse Gauss-Hermite quadrature filter has desirable precision in comparison with other traditional Kalman filters. This motivates us to apply the generalized correntropy into the sparse Gauss-Hermite quadrature filter, yielding a generalized correntropy sparse Gauss-Hermite quadrature filter (GCSGHQF). The novel generalized correntropy sparse Gauss-Hermite quadrature filter considered as the second work is used to deal with tracking epidemic spreading well in the presence of non-Gaussian noise. In particular, different types of non-Gaussian noise such as the uniform noise and Laplace noise are all considered for better demonstrating the filtering precision of the novel generalized correntropy sparse Gauss-Hermite quadrature filter.

Unlike the use of a propagation dynamic model in epidemic tracking, the dynamics of nonlinear systems are analyzed by state models. Therefore, it is not necessary to study the choice of nonlinear Kalman filters with different state models in analyzing the dynamics of power systems as in epidemic tracking on networks. Similar to the epidemic tracking on networks, the dynamic estimation of power systems is

also confronted with one issue, i.e., the presence of measurements corrupted by non-Gaussian noise. More concretely, the power system is operating under normal conditions where system parameters are pre-designed. These system parameters are kept unchanged at the pre-designed values but may diverge from the normal values when the power system is confronted with system faults [28]. Nonlinear Kalman filters are required for dynamic estimation, aiming to reveal the dynamics of complex power systems. In the control of power systems, the measurements collected from the phasor measurement units (PMUs) [28, 52, 53, 54] stationed at generator buses in a power system may be corrupted by non-Gaussian noise. This leads us to develop a robust Kalman filter for handling contaminated measurements for efficiently improving filtering precision. As corrupted measurements may exhibit varying levels of sensitivity to non-Gaussian noise, a robust mixed p-norm square root unscented Kalman filter (PSRUKF) is considered for estimating the state of power systems in the presence of non-Gaussian noise. Different from the uniform noise and Laplace noise considered in our second work in this thesis, i.e., the application of epidemic tracking, a much more complex non-Gaussian noise called the alpha stable distribution noise is used for demonstrating the filtering precision of the mixed p-norm square root unscented Kalman filter.

The main objective of this thesis is to utilize Kalman filters for studying the dynamics of information flow over various networks. Nonlinear Kalman filters perform the state estimation on the basis of a state model and a set of collected measurements. On the one hand, the state model is classified either as a propagation dynamic model or a nonlinear system model. The epidemic spreading on networks and the dynamic analysis of power systems are representative applications which involve specific propagation dynamics and nonlinear system dynamics, respectively. On the other hand, the collected measurements may be contaminated by different types of non-Gaussian noise such as uniform noise, Laplace noise and alpha stable distribution noise. This thesis provides a guideline for choosing traditional nonlinear Kalman filters

with different compartmental models for tracking epidemic spreading in the presence of Gaussian noise. For handling the non-Gaussian noise case, a robust generalized correntropy sparse Gauss-Hermite quadrature filter (GCSGHQF) is then applied to capturing the dynamics of epidemic transmission in the presence of non-Gaussian noise. In addition to tracking epidemic transmission, the dynamics of power systems are another issue studied in depth in this thesis. A novel robust nonlinear Kalman filter, called mixed p-norm square root unscented Kalman filter (PSRUKF), is presented for dealing with the dynamic estimation of power systems when collected measurements are corrupted by non-Gaussian noise.

1.3 Thesis Organization

The remainder of this thesis is organized as follows.

Chapter 2 provides a literature review including a brief introduction of complex networks, nonlinear Kalman filters and robust nonlinear Kalman filters.

Chapter 3 considers epidemic tracking on networks using various Kalman filters on the basis of degree-based compartmental models in the presence of Gaussian noise. A comparison between nonlinear Kalman filters is made in terms of estimation accuracy and stability. The performance comparison provides a guideline for choosing appropriate nonlinear Kalman filters for studying the dynamics of epidemic transmission over complex networks. Several strategies are also proposed for reducing the computational burden caused by a high dimensional state vector for large networks.

Chapter 4 considers the dynamics of epidemic spreading on networks using a novel robust generalized correntropy sparse Gaussian Hermite Quadrature filter (GCSGHQF) that tackles the effects of non-Gaussian noise. The GCSGHQF uses the generalized Gaussian density (GGD) function for substituting the Gaussian kernel under the maximum correntropy criterion (MCC) so that the estimation accuracy and numerical stability can be improved in the presence of non-Gaussian noise. An extra

second-order statistical characteristic of the error is also considered in the GCSGHQF for further enhancing the numerical stability in the presence of non-Gaussian noise.

Chapter 5 proposes a novel robust mixed p-norm square root unscented Kalman filter (PSRUKF) for tackling abnormal measurements. The proposed mixed p-norm utilizes a piecewise function, i.e., multiple p-norms, for handling varying measurements. A threshold factor is utilized for determining the boundary of abnormal errors and then mixed p-norm is applied to dealing with varying measurement errors. The proposed mixed p-norm is a generalized p-norm defined over varying intervals, which eliminates abnormal errors. In addition, an extra second-order statistical characteristic of the error is also considered for reducing the negative effect caused by the inappropriate choice of the threshold parameter.

Chapter 6 concludes the thesis, reiterating the major findings of the project and providing some further thoughts for future work.

Chapter 2

Literature Review

This chapter provides a brief introduction of complex networks, nonlinear Kalman filters, and robust nonlinear Kalman filters. In the study of the information flow on various networks, nonlinear Kalman filters are essential for capturing flow dynamics and characteristics over complex networks by utilizing appropriate state models and measurement models simultaneously. Since practical measurements may be contaminated by non-Gaussian noise, robust nonlinear Kalman filters are expected to be useful for handling corrupted measurements for improving the estimation precision.

2.1 Brief Introduction to Complex Networks

This section briefly introduces the basic properties of complex networks and common complex network models. The average node degree, node degree distribution, shortest path length and clustering coefficient are common parameters for describing the structures and characteristics of complex networks. In addition, the Erdös and Rényi (ER) network, Newman and Watts (NW) network, Watts and Strogatz (WS) network, and Barabási and Albert (BA) network are common network models that possess some properties of real-world complex networks.

2.1.1 Measures of Network Topology

Some measures have been proposed for exploring and analyzing real-world complex networks. Commonly used measures include node degree, degree distribution, shortest path length, and clustering coefficient.

Node Degree

A network can be described as a graph $G = (N, E)$, where N and E denote the sets of nodes and links in the graph G [55]. Denote the size of N and E by $|N|$ and $|E|$, respectively. The degree is a simple measure for assessing the topological criticality of a node in a network. Since edges have no direction in an undirected network, the degree k_i of node i is defined as the number of edges incident on node i, i.e.,

$$k_j = \sum_i a_{ji} \tag{2.1}$$

where $a_{ji} \in \{0, 1\}$ is the j-th row i-th column element of the adjacency matrix $\mathbf{A}$ with $a_{ji} = 1$ denoting the connectivity of node j with node i. For a directed network, the out-degree and in-degree of a node refer to the number of outgoing edges from this node and the number of incoming edges to this node, respectively.

The node degree has an essential role in analyzing real-world networks. It is easy to understand that nodes with high degree acting as hubs are in general more important than those with low degree in facilitating the information flow on networks. For example, nodes with high degree may offer more pathways and hence accelerate the information flow through a complex network.

Average Node Degree

The average node degree $\langle k \rangle$ denotes the mean value of the degrees of all nodes of a complex network which reflects the connection density of a network. Considering the degrees of the nodes in a complex network $\{k_i\}_{i=1}^N$, we have the average node degree as

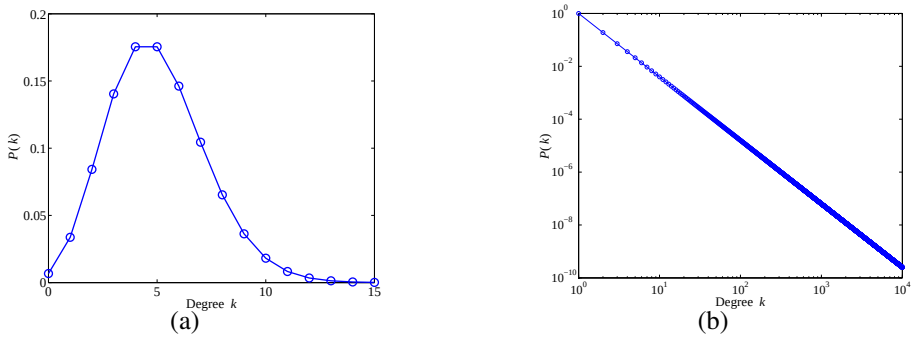


Figure 2.1: (a) Poisson distribution; (b) Power-law distribution.

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i, \tag{2.2}$$

with N denoting the total number of nodes in the network. A higher $\langle k \rangle$ means that the nodes are connected with each other more densely.

Node Degree Distribution

The node degree distribution is another statistical metric for describing the collective topological feature associated with node degrees. The node degree distribution $P(k)$ refers to the probability that a randomly chosen node has degree k, and is defined as

$$P(k) = \frac{N(k)}{N} \tag{2.3}$$

where $N(k)$ denotes the number of nodes having degree k.

The Poisson distribution and power-law distribution are two common types of node degree distribution, as shown in Fig. 2.1(a) and Fig. 2.1(b), respectively. More concretely, the Poisson distribution has the following form:

$$P(k) \sim e^{-\lambda} \lambda^k / k! \tag{2.4}$$

with λ representing the average node degree of the network. It can be seen from Fig. 2.1(a) that the Poisson distribution reaches a maximum when degree k is around the average node degree. The networks characterized by the Poisson distribution of

node degrees are in general homogeneous networks. Moreover, the power-law degree distribution is

$$P(k) \sim k^{-\gamma} \tag{2.5}$$

where γ is the degree exponent. A network can be interpreted as a heterogeneous network when the network has a power-law degree distribution as shown in Fig. 2.1(b). For complex networks with a power-law degree distribution, only a small fraction of nodes have a large number of edges while others have a few edges.

Shortest Path

In a network, the shortest path between nodes i and j is the path with the fewest number of edges between nodes i and j. The corresponding distance between nodes i and j, d_{ij}, is calculated as the number of edges along the shortest path. It is worth noting that the shortest path with a specific distance is not unique. The shortest path has been widely used in optimizing the efficiency of data transmission in a communication network and of delivering payloads in a transportation network. There has been research around the theme of finding the shortest path efficiently in computer science [56]. The largest distance between any pair of nodes in a complex network is termed as diameter.

The average shortest path length is considered as a measure of the separation between two nodes in a network. Consider a complex network with N nodes. Average shortest path length L is defined as the mean distance between all pairs of nodes, i.e.,

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}. \tag{2.6}$$

In general, the value of L is not large for online social networks, and equal to 4 and 3.5 for Facebook in the year 2012 [57] and the year 2016 [58], respectively. This in fact means that social distance has been shortened by the Internet.

Clustering Coefficient

The clustering coefficient is a measure of the extent to which the neighbors of a given node connect to each other in a network. Given a complex network with N nodes and E edges, the clustering coefficient Γ_i is defined as the ratio of the number of all edges connecting node i, i.e., E_i and the largest number of edges connected among the neighbors of node i.

$$\Gamma_i = \frac{2E_i}{k_i(k_i - 1)} \tag{2.7}$$

with k_i denoting the degree of node i.

The average clustering coefficient is computed by averaging clustering coefficients in a complex network, i.e.,

$$\Gamma = \frac{1}{N}\sum_{i=1}^{N}\Gamma_i. \tag{2.8}$$

In contrast to social networks which have a high clustering coefficient, hierarchical networks, like the Internet and power grid [59], have a rather low clustering coefficient.

2.1.2 Models of Network Topology

There are several types of complex network models including the Erdös and Rényi (ER) network, the Newman and Watts (NW) network, the Watts and Strogatz (WS) network, and the Barabási-Albert (BA) scale-free network. These network models are helpful for exploring the characteristics and structures of real-world networks.

ER Random Network

Erdös and Rényi in 1959 [2] first proposed a model for constructing a random network in two steps. First, N isolated nodes are fixed, and then each pair of nodes is connected according to a connection probability p. An ER random network with a large N has $pN(N-1)/2$ edges and an average node degree $\langle k\rangle = (N-1)p \approx Np$. In

addition, an ER random network has a Poission node degree distribution, i.e.,

$$P(k) = (\langle k \rangle)^k \frac{e^{-\langle k \rangle}}{k!}. \tag{2.9}$$

The ER network in general has a relatively small average path length $L_{\text{random}} \sim \ln(N)/\ln(\langle k \rangle)$ and a low clustering coefficient $\Gamma_{\text{random}} \sim \langle k \rangle / N$. The ER network is simple, but fails to describe a complex real-world network.

WS and NW Small-World Network Models

In 1998, Watts and Strogatz [9] proposed a network model for studying the small-world property of real-world networks:

(1) Begin with a nearest-neighbor coupled network consisting of N nodes arranged in a ring, where each node is adjacent to its neighbor nodes.

(2) Randomly rewire each edge of the network with probability p.

As a small-world network, the WS network has a short average path length and a relatively high clustering coefficient. These properties are found in most of real-world networks and distinguish them from random and regular networks. The statistical characteristics of the WS small-world network are well coincident with many large-scale social networks. For example, the friends of one individual are in reality also friends of each other. This means that many clusters exist in social networks. Since most people know each other through a few common friends, the average shortest path length of a social network is small.

Apart from the WS small-world network model, another constitution of small-world networks was proposed by Newman and Watts (NW) [5]. Different from the WS small-world network model, the NW small-world network is generated by adding a connection between a pair of nodes with probability p without breaking any connection between any two nearest neighbors. As a result, the NW model reduces to the original nearest-neighbor coupled network and globally coupled network when $p = 0$ and $p = 1$, respectively.

BA Scale-Free Network Model

Networks with a Poission degree distribution are not commonly found in the real-world. Instead, scale-free networks, where only a few nodes have high degree and most of the nodes have low degree, are more prevalent.

Barabási and Albert [12] proposed a model for constructing scale-free networks in two steps:

(1) One new node is added to the network with the connection of existing nodes in the network at each time step.

(2) The probability Π_i that a newly added edge between the new node and the existing node i will be connected is given by $\Pi_i = k_i / \sum_j k_j$, with k_i denoting the degree of node i .

This leads to a power-law degree distribution as the network grows. This property is found in many real-world networks, like the online social network Wikipedia [60], the protein-protein interaction network [61], and the citation network. New nodes in these networks are likely to link with nodes having more connections.

2.2 Nonlinear Kalman Filters

Nonlinear Kalman filters estimate the hidden state of a nonlinear dynamical system on the basis of a state-space model equipped with a state function $\boldsymbol{f}(\cdot)$ and measurement function $\boldsymbol{h}(\cdot)$ [62]. The dynamic state function $\boldsymbol{f}(\cdot)$ describes the inherent mechanism of a nonlinear system for yielding a predictive state. A well-established measurement function is constructed for providing observable measurements for correcting the predictive state [30].

Consider a discrete-time dynamical system [45]

$$\mathbf{x}_k = \boldsymbol{f}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1} \tag{2.10}$$

$$\mathbf{y}_k = \boldsymbol{h}(\mathbf{x}_k) + \mathbf{v}_k \tag{2.11}$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x \times 1}$ and $\mathbf{y}_k \in \mathbb{R}^{n_y \times 1}$ denote the state and measurement, respectively. The process noise $\mathbf{w}_{k-1}$ and the measurement noise $\mathbf{v}_k$ obey the Gaussian distribution, i.e., $\mathbf{w}_{k-1} \sim N(\mathbf{0}, \mathbf{Q}_{k-1})$, $\mathbf{v}_{k-1} \sim N(\mathbf{0}, \mathbf{R}_k)$, respectively. The core of nonlinear Kalman filters lies in the calculation of the predictive probability density $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ and corrected probability density $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ at the predictive step and the update step, respectively. Nonlinear Kalman filters are inferred by adopting the Bayesian rule [63] on the basis of the process function given in (2.10) and measurement model given in (2.11).

(1) Predictive Step

Assume that the probability density at discrete time $k-1$ is

$$p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{x}}_{k-1}, \mathbf{P}_{k-1}) \tag{2.12}$$

where $\hat{\mathbf{x}}_{k-1}$ and $\mathbf{P}_{k-1}$ represent the estimated state and corresponding error covariance at discrete time $k-1$, respectively. The predictive state $\hat{\mathbf{x}}_{k-1|k}$ is calculated by

$$\begin{aligned}\hat{\mathbf{x}}_{k|k-1} &= E[\boldsymbol{f}(\mathbf{x}_{k-1})|\mathbf{y}_{1:k-1}] \\ &= \int_{\mathbb{R}^{n_x}} \boldsymbol{f}(\mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1},\end{aligned} \tag{2.13}$$

and the corresponding estimation error is found by

$$\begin{aligned}\mathbf{P}_{k|k-1} &= E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}}|\mathbf{y}_{1:k-1}] \\ &= \int_{\mathbb{R}^{n_x}} \boldsymbol{f}(\mathbf{x}_{k-1})\, \boldsymbol{f}^{\mathrm{T}}(\mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\hat{\mathbf{x}}_{k-1}^{\mathrm{T}} \\ &\quad + \mathbf{Q}_{k-1}.\end{aligned} \tag{2.14}$$

(2) Update Step

A corrected probability density at discrete time k, $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, can be calculated upon receipt of the current measurement $\mathbf{y}_k$:

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \mathcal{N}(\hat{\mathbf{x}}_k, \mathbf{P}_k) \tag{2.15}$$

where $\hat{\mathbf{x}}_k$ and $\mathbf{P}_k$ represent the estimated state and corresponding error covariance at discrete time k, respectively. The conditional joint Gaussian probability distribution, denoted as $p(\bar{\mathbf{x}}_k|\mathbf{y}_{1:k-1})$, with joint variable $\bar{\mathbf{x}}_k = [\mathbf{x}_k^{\mathrm{T}} \ \ (\mathbf{y}_k)^{\mathrm{T}}]^{\mathrm{T}}$ can be found as

$$p(\bar{\mathbf{x}}_k|\mathbf{y}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{x}}^y_{k-1}, \mathbf{P}^y_{k-1}) \tag{2.16}$$

where $\hat{\mathbf{x}}^y_{k-1} = [\hat{\mathbf{x}}^{\mathrm{T}}_{k|k-1} \quad (\hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}}]^{\mathrm{T}}$ and covariance matrix $\mathbf{P}^y_{k-1}$ are given by

$$\mathbf{P}^y_{k-1} = \begin{bmatrix} \mathbf{P}_{k|k-1} & \mathbf{P}_{xy,k|k-1} \\ \mathbf{P}^{\mathrm{T}}_{xy,k|k-1} & \mathbf{P}_{yy,k|k-1} \end{bmatrix}. \tag{2.17}$$

Also, $\mathbf{P}_{yy,k-1|k}$ represents the error covariance of the predictive measurement, and

$$\begin{aligned} \mathbf{P}_{yy,k|k-1} &= E[(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}}|\mathbf{y}_{1:k-1}] \\ &= \int_{\mathbb{R}^{n_y}} \boldsymbol{h}(\mathbf{x}_k)(\boldsymbol{h}(\mathbf{x}_k))^{\mathrm{T}} p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) d\mathbf{x}_k - \hat{\mathbf{y}}_{k|k-1}(\hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}} \\ &\quad + \mathbf{R}_k. \end{aligned} \tag{2.18}$$

Furthermore, the cross covariance $\mathbf{P}_{xy,k|k-1}$ is calculated by

$$\begin{aligned} \mathbf{P}_{xy,k|k-1} &= E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}}|\mathbf{y}_{1:k-1}] \\ &= \int_{\mathbb{R}^{nx}} \boldsymbol{f}(\mathbf{x}_{k-1})(\boldsymbol{h}(\mathbf{x}_k))^{\mathrm{T}} p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) d\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}(\hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}}. \end{aligned} \tag{2.19}$$

The posterior probability density $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ can then be found upon the receipt of the current measurement $\mathbf{y}_k$, yielding the corrected state and the corresponding covariance of state error as

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1})) \tag{2.20}$$

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} + \mathbf{K}_k\mathbf{P}_{yy,k|k-1}\mathbf{K}_k^{\mathrm{T}}. \tag{2.21}$$

The Kalman gain $\mathbf{K}_k$ is set as

$$\mathbf{K}_k = \mathbf{P}_{xy,k|k-1}\mathbf{P}_{yy,k|k-1}^{-1} \tag{2.22}$$

where covariance matrices $\mathbf{P}_{yy,k|k-1}$ and $\mathbf{P}_{xy,k|k-1}$ are defined in (2.18) and (2.19), respectively.

Dependent on the method used in computing multiple integrals from (2.13) to (2.19), different nonlinear Kalman filters can be derived. Nonlinear Kalman filters are in general classified as global and local methods. Without making specific assumption with regard to the probability density functions of the state variable, global approaches like the grid-based method [35], the particle filter [36, 37] and the Gaussian mixture filter [38], achieve accurate estimation at the expense of imposing large computational burden.

By contrast, local approaches calculate multiple integrals using (2.13) to (2.19) by making approximation with respect to nonlinear functions in (2.10) and (2.11) or the probability density functions of the state variable. The extended Kalman filter (EKF) is one of the most commonly used local methods by virtue of the first-order linearization of a nonlinear function [39, 40], i.e.,

$$\boldsymbol{f}(\mathbf{x}_{k-1}) \approx \boldsymbol{f}(\hat{\mathbf{x}}_{k-1}) + \mathbf{F}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}) \tag{2.23}$$

$$\boldsymbol{h}(\mathbf{x}_k) \approx \boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{H}_k(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \tag{2.24}$$

where $\hat{\mathbf{x}}_{k-1}$ and $\hat{\mathbf{x}}_{k|k-1}$ represent the estimated state and predictive state at discrete time $k-1$, respectively. In (2.23) and (2.24), Jacobian matrices $\mathbf{F}_{k-1}$ and $\mathbf{H}_k$ are expressed as

$$\mathbf{F}_{k-1} = \frac{\partial \boldsymbol{f}\left(\mathbf{x}_{k-1}\right)}{\partial \mathbf{x}} \mid_{\mathbf{x}=\hat{\mathbf{x}}_{k-1}} \tag{2.25}$$

$$\mathbf{H}_k = \frac{\partial \boldsymbol{h}\left(\mathbf{x}_k\right)}{\partial \mathbf{x}} \mid_{\mathbf{x}=\hat{\mathbf{x}}_{k|k-1}} . \tag{2.26}$$

Substituting (2.23) and (2.24) into (2.10) and (2.11) yields

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \boldsymbol{f}\left(\hat{\mathbf{x}}_{k-1}\right) - \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{w}_{k-1} \tag{2.27}$$

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \boldsymbol{h}\left(\hat{\mathbf{x}}_{k|k-1}\right) - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} + \mathbf{v}_k. \tag{2.28}$$

The predictive state $\hat{\mathbf{x}}_{k|k-1}$ and measurement $\hat{\mathbf{y}}_{k|k-1}$ are therefore derived as

$$\hat{\mathbf{x}}_{k|k-1} = \boldsymbol{f}\left(\hat{\mathbf{x}}_{k-1}\right) \tag{2.29}$$

$$\hat{\mathbf{y}}_{k|k-1} = \boldsymbol{h}\left(\hat{\mathbf{x}}_{k|k-1}\right). \tag{2.30}$$

The corresponding estimated state is derived as [39, 40]

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}\right) \tag{2.31}$$

with the Kalman gain $\mathbf{K}_k$ defined by

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k^{-1}\right). \tag{2.32}$$

Algorithm 1 summarizes the operation of the extended Kalman filter.

Apart from the extended Kalman filter (EKF), there exist numerous Kalman filters calculating multiple integrals using (2.13) to (2.19) by making the Gaussian assumption of the probability density functions of the state variables. These nonlinear

Kalman filters include the unscented Kalman filter (UKF) and cubature Kalman filters (CKF) like the third-degree cubature Kalman filter (CKF3) and the fifth-order cubature Kalman filter (CKF5). In particular, the unscented Kalman filter has the ability of achieving balanced filtering performance in terms of estimation precision and computational complexity in comparison with the third-degree cubature Kalman filter and fifth-order cubature Kalman filter. The unscented Kalman filter utilizes an unscented transformation for computing multiple integrals where sampling points $\{\boldsymbol{\xi}_i\}_{i=0}^{2n_x}$ and corresponding weights $\{\omega_i^m, \omega_i^c\}_{i=0}^{2n_x}$ are considered for calculating the posterior probability density functions of the state, $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ and $p(\mathbf{x}_k|\mathbf{y}_{1:k})$. The sampling points $\{\boldsymbol{\xi}_i\}_{i=0}^{2n_x}$ are defined by

$$\begin{cases} \boldsymbol{\xi}_0 = \bar{\mathbf{x}} \\ \boldsymbol{\xi}_i = \bar{\mathbf{x}} + \left(\sqrt{(n_x+\lambda)\,\mathbf{P}_x}\right)_i, & i = 1, 2, \cdots, n_x \\ \boldsymbol{\xi}_i = \bar{\mathbf{x}} - \left(\sqrt{(n_x+\lambda)\,\mathbf{P}_x}\right)_{i-n_x}, & i = n_x+1, \cdots, 2n_x \end{cases} \tag{2.33}$$

with the error covariance $\mathbf{P}_x$ and $\left(\sqrt{(n_x+\lambda)\,\mathbf{P}_x}\right)_i$ denoting the i-th column vector of the matrix $\left(\sqrt{(n_x+\lambda)\,\mathbf{P}_x}\right)$. Here, ω_i^m and ω_i^c denote the weights for calculating the mean and covariance, respectively. The weights ω_i^m and ω_i^c are given symmetrically by

$$\begin{cases} w_0^m = \frac{\lambda}{n_x+\lambda} \\ w_0^c = \frac{\lambda}{n_x+\lambda} + 1 - \alpha^2 + \beta \\ w_i^m = w_i^c = \frac{1}{2(n_x+\lambda)}, \; i = 1, 2, \cdots, 2n_x. \end{cases} \tag{2.34}$$

Parameter λ is defined as $\lambda = \alpha^2(n_x + \kappa) - n_x$ where α determines the spreading range of sampling points at point $\bar{\mathbf{x}}$, and κ is a secondary scaling parameter. Parameter β in is utilized for incorporating the prior information of the distribution of $\bar{\mathbf{x}}$. The unscented Kalman filter reduces to the third-degree cubature Kalman filter with $\alpha = 1, \beta = 0$ and

Algorithm 1 Extended Kalman Filter

Prediction step

Compute the predicted mean $\hat{\mathbf{x}}_{k|k-1}$ and predicted error covariance $\mathbf{P}_{k|k-1}$

$$\hat{\mathbf{x}}_{k|k-1} = \boldsymbol{f}\left(\hat{\mathbf{x}}_{k-1}\right) \tag{2.37}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{P}_{k-1}\mathbf{F}_{k-1}^{\mathrm{T}} + \mathbf{Q}_{k-1} \tag{2.38}$$

with the Jacobi matrix $\mathbf{F}_{k-1}$

$$\mathbf{F}_{k-1} = \frac{\partial \boldsymbol{f}\left(\mathbf{x}_{k-1}\right)}{\partial \mathbf{x}} \mid_{\mathbf{x}=\hat{\mathbf{x}}_{k-1}} . \tag{2.39}$$

Update step

1: Compute the Kalman gain $\mathbf{K}_k$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k\right)^{-1} \tag{2.40}$$

with the Jacobi matrix $\mathbf{H}_k$

$$\mathbf{H}_k = \frac{\partial \boldsymbol{h}\left(\mathbf{x}_k\right)}{\partial \mathbf{x}} \mid_{\mathbf{x}=\hat{\mathbf{x}}_{k|k-1}} . \tag{2.41}$$

2: Compute the estimated mean $\hat{\mathbf{x}}_k$ and estimated error covariance $\mathbf{P}_k$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \boldsymbol{h}\left(\hat{\mathbf{x}}_{k|k-1}\right)\right) \tag{2.42}$$

$$\mathbf{P}_k = \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_{k|k-1}. \tag{2.43}$$

Output: $\hat{\mathbf{x}}_k$, $\mathbf{P}_k$

$\kappa = 0$. The third-degree cubature Kalman filter adopts following samples

$$\begin{cases} \boldsymbol{\xi}_i = \bar{\mathbf{x}} + \left(\sqrt{n_x\mathbf{P}_x}\right)_i, & i = 1, 2, \cdots, n_x \\ \boldsymbol{\xi}_i = \bar{\mathbf{x}} - \left(\sqrt{n_x\mathbf{P}_x}\right)_{i-n_x}, & i = n_x + 1, \cdots, 2n_x. \end{cases} \tag{2.35}$$

The corresponding weights are given by

$$w_i = \tfrac{1}{2n_x},\ i = 1, 2, \cdots, 2n_x. \tag{2.36}$$

The fifth-degree cubature Kalman filter is also proposed for further improving filtering performance at the cost of increasing computational burden. The fifth-degree cubature Kalman filter calculates the multiple integrals using (2.13) to (2.19) by incorporating

Algorithm 2 Nonlinear Kalman Filter Based on Deterministic Sampling

Prediction step

1: Calculate L sample points $\{\boldsymbol{\xi}_{i,k-1}\}_{i=1}^{L}$ and weights $\{\omega_i\}_{i=1}^{L}$ where $\boldsymbol{\xi}_{i,k-1}$ and ω_i represent the i-th sample point and i-th weight at the discrete time $k-1$ [43, 46].

2: Compute the predicted mean $\hat{\mathbf{x}}_{k|k-1}$ and predicted error covariance $\mathbf{P}_{k|k-1}$

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=1}^{L} \omega_i \boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}), \tag{2.44}$$

$$\mathbf{P}_{k|k-1} = \mathbf{P}^{s}_{k|k-1} + \mathbf{Q}_k \tag{2.45}$$

with the matrix

$$\mathbf{P}^{s}_{k|k-1} = \sum_{i=1}^{L} \omega_i (\boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k|k-1})(\boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}} \tag{2.46}$$

Update step

1: Calculate L sample points $\{\boldsymbol{\xi}_{i,k-1}\}_{i=1}^{L}$ and weights $\{\omega_i\}_{i=1}^{L}$ [43, 46].

2: Compute the estimated mean $\hat{\mathbf{x}}_k$ and the corresponding error covariance of the state $\mathbf{P}_k$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1})), \tag{2.47}$$

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{yy,k|k-1} \mathbf{K}_k^{T}. \tag{2.48}$$

The matrix $\mathbf{K}_k = \mathbf{P}_{xy,k|k-1}\mathbf{P}^{-1}_{yy,k|k-1}$ represents the Kalman gain with covariance matrix

$$\mathbf{P}_{xy,k|k-1} = \sum_{i=1}^{L} \omega_i (\boldsymbol{\xi}_{i,k|k-1} - \hat{\mathbf{x}}_{k|k-1})(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}} \tag{2.49}$$

$$\mathbf{P}_{yy,k|k-1} = \mathbf{R}_k + \sum_{i=1}^{L} \omega_i (\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}} \tag{2.50}$$

and predicted measurement $\hat{\mathbf{y}}_{k|k-1} = \sum_{i=1}^{L} \omega_i \boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1})$.

Output: $\hat{\mathbf{x}}_k$, $\mathbf{P}_k$

sampling points $\{\boldsymbol{\xi}_i\}_{i=0}^{2n_x^2}$

$$\begin{cases} \boldsymbol{\xi}_i = \bar{\mathbf{x}}, \\ \begin{cases} \boldsymbol{\xi}_{1+i} = \bar{\mathbf{x}} + \sqrt{(n_x+2)\,\mathbf{P}_x}\mathbf{p}^{+}, \\ \boldsymbol{\xi}_{\frac{n(n-1)}{2}1+i} = \bar{\mathbf{x}} - \sqrt{(n_x+2)\,\mathbf{P}_x}\mathbf{p}^{+}, \\ \boldsymbol{\xi}_{n(n-1)+1+i} = \bar{\mathbf{x}} + \sqrt{(n_x+2)\,\mathbf{P}_x}\mathbf{p}^{-}, \quad i = 1, 2, \cdots, \frac{n_x(n_x-1)}{2} \\ \boldsymbol{\xi}_{\frac{3n(n-1)}{2}1+i} = \bar{\mathbf{x}} - \sqrt{(n_x+2)\,\mathbf{P}_x}\mathbf{p}^{-}, \end{cases} \\ \begin{cases} \boldsymbol{\xi}_{2n(n-1)+1+i} = \bar{\mathbf{x}} + \sqrt{(n_x+2)\,\mathbf{P}_x}\mathbf{e}_i, \\ \boldsymbol{\xi}_{n(2n-1)+1+i} = \bar{\mathbf{x}} - \sqrt{(n_x+2)\,\mathbf{P}_x}\mathbf{e}_i, \quad i = 1, 2, \cdots, n_x \end{cases} \end{cases} \tag{2.51}$$

where vectors $\mathbf{p}_i^+$ and $\mathbf{p}_i^-$ are given by

$$\begin{aligned} \mathbf{p}_i^+ &= \sqrt{\tfrac{1}{2}}(\mathbf{e}_j + \mathbf{e}_k),\ j < k,\ j,k = 1,2,\cdots,n \\ \mathbf{p}_i^- &= \sqrt{\tfrac{1}{2}}(\mathbf{e}_j - \mathbf{e}_k),\ j < k,\ j,k = 1,2,\cdots,n \end{aligned} \tag{2.52}$$

with $\mathbf{e}_j$ denoting the unit vector with j-th element equal to one, and the others equal to zero. The corresponding weights $\{w_i\}_{i=0}^{2n_x^2}$ are given by

$$\begin{cases} w_0 = \frac{2}{n_x+2} \\ w_i = \frac{1}{(n_x+2)^2},\ i = 1,2,\cdots,2n_x(n_x-1). \\ w_i = \frac{4-n_x}{2(n_x+2)^2},\ i = 2n_x(n_x-1)+1, 2n_x(n_x-1)+2,\cdots,2n_x^2. \end{cases} \tag{2.53}$$

The operation of nonlinear Kalman filters based on deterministic sampling is summarized in **Algorithm** 2.

2.3 Robust Nonlinear Kalman Filters

Traditionally, nonlinear Kalman filters are based on minimizing the mean square error in the estimation. Thus, traditional nonlinear Kalman filters may fail to achieve adequate filtering precision in the presence of non-Gaussian noise. Since the available measurements are corrupted by non-Gaussian noise in different ways such as additive and multiplicative processes, robust nonlinear Kalman filters are expected to be derived for dealing with non-Gaussian noise. The additive manner for imposing non-Gaussian noise on measurements is mainly considered due to its universality.

There exist numerous robust nonlinear Kalman filters, e.g., robust Student's t-based Kalman filter (RSTKF) [64], particle filter (PF) [36, 37] and Gaussian sum filter (GSF) [38]. In particular, the Gaussian sum filter considers a set of Gaussian mixture distributions for computing the posterior distribution with respect to the state [65]. In the particle filter, a large number of random particles are applied

for calculating the posterior distribution [66]. The robust Student's t-based Kalman filter uses a Student's t-distribution to model the heavy-tailed non-Gaussian noise for yielding improved posterior estimates [64]. Furthermore, robust Kalman filters can be designed by transforming the state estimation into a multi-dimensional optimization problem. For example, a class of Huber-based Kalman filters [67, 68] has been developed for improving robustness against abnormal measurements. In [67], an improved Huber-based extended Kalman filter is derived on the basis of nonlinear regression theory. The Huber technique is also applied to the fixed-interval smoothing problem. Due to the improved filtering precision of the cubature Kalman filter, a robust Huber-based cubature kalman filter has been proposed for robust improvement [68]. As the generalization of the cubature Kalman filter, the unscented Kalman filter can be combined with the Huber technique for suppressing abnormal measurements efficiently [69]. These nonlinear Kalman filters all aim to solve the optimization over a piecewise Huber function [69, 70], i.e.,

$$L_h(e) = \begin{cases} \frac{1}{2}e^2, & |e| \leq \delta \\ \delta|e| - \frac{1}{2}\delta^2, & \text{otherwise} \end{cases} \tag{2.54}$$

with a positive threshold parameter δ for determining whether l_1-norm or l_2-norm is used for scaling error. Fig. 2.2 shows the Huber function $L_h(e)$ versus the error e with a positive threshold parameter $\delta = 1.345$. As shown in Fig. 2.2, the Huber function is used for scaling measurement errors with l_1-norm and l_2-norm. This is indeed beneficial for suppressing outliers efficiently, but also limits the improvement of filtering accuracy due to the limited use of l_1-norm and l_2-norm.

Recently, information theoretical learning (ITL) [71, 72, 73] is proposed for improving robustness in the presence of outliers for capturing a higher statistical characteristic of errors. In particular, the maximum correntropy criterion (MCC) [74, 75, 76, 77] is a widely used information theoretical learning approach due to its

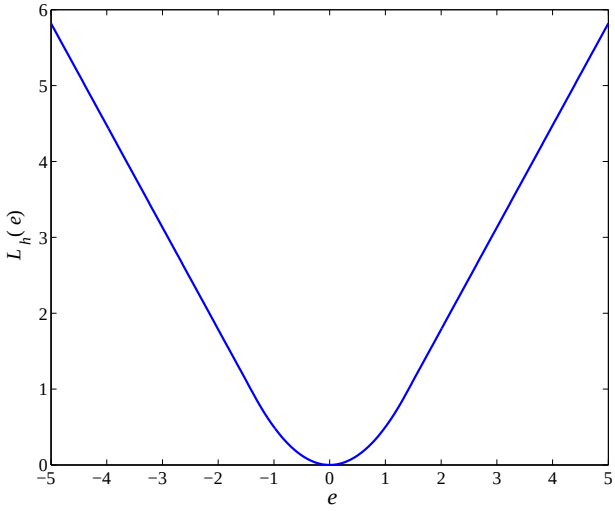


Figure 2.2: Huber function $L_h(e)$ versus error e.

simplicity. The correntropy of two arbitrary scalar random variables X and Y measures their similarity, and is defined by [78]

$$V(X,Y) = E_{XY}(\kappa_\sigma(X,Y)) = \int \kappa_\sigma(x,y)\, dF_{XY}(x,y) \tag{2.55}$$

where $E_{XY}(\cdot)$ denotes the expectation with distribution function F_{XY}, and $\kappa_\sigma(\cdot)$ is a kernel function with kernel bandwidth σ. The Gaussian kernel [79] is the most commonly used kernel function and is given by

$$\kappa_\sigma(x,y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{e^2}{2\sigma^2}\right) \tag{2.56}$$

where $\sigma > 0$ denotes the Gaussian kernel bandwidth and $e = x - y$ is the difference between x and y. Equation (2.55), however, cannot be calculated since the probability distribution function is unknown in practice. Thus, (2.55) is generally approximated by utilizing $\tilde{N}$ available samples $\{x_i, y_i\}_{i=1}^{\tilde{N}}$, i.e.,

$$\hat{V}(X,Y) \approx \frac{1}{\tilde{N}} \sum_{i=1}^{\tilde{N}} \kappa_\sigma(x_i - y_i). \tag{2.57}$$

Robust maximum correntropy-based Kalman filters are developed by combining the maximum correntropy criterion given in (2.57) with traditional Kalman filters [78, 80, 81, 82]. In [78], a maximum correntropy Kalman filter (MCKF) is derived by applying the maximum correntropy criterion to the linear Kalman filter for dealing with non-Gaussian noise. Recently, a maximum correntropy-based nonlinear Kalman filter, i.e., maximum correntropy unscented Kalman filter and its variant, i.e., iterated maximum correntropy unscented Kalman filters have been proposed [80, 82]. Due to the improved numerical stability of square root-based Kalman filters [45], a maximum correntropy square-root cubature Kalman filter is proposed for improving robustness as well as numerical stability [81]. These nonlinear Kalman filters all have the ability of dealing with abnormal measurements. In comparison with Kalman filters on the basis of the Huber technique, maximum correntropy-based Kalman filters handle corrupted measurements in a soft manner which is beneficial for retaining measurement information. Compared to the maximum correntropy criterion, generalized correntropy is more flexible and efficient in dealing with corrupted measurements. The generalized correntropy [83, 84] substitutes the Gaussian kernel in (2.56) by a generalized kernel function, i.e.,

$$G_{\alpha,\beta}(e) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left|\frac{e}{\beta}\right|^{\alpha}\right) \tag{2.58}$$

where $\Gamma(\cdot)$ is the gamma function; α (positive) and β denote the shape parameter and bandwidth parameter, respectively. According to (2.58), the Laplace distribution is obtained for $\alpha = 1$. In addition, (2.58) reduces to the Gaussian distribution when the shape parameter α is set to 2. The generalized correntropy density converges to a uniform density on $(-\beta, \beta)$ in a point-wise manner [73, 85]. Substituting (2.58) into (2.55) gives

$$V_{\alpha,\beta}(X, Y) = E_{XY}(G_{\alpha,\beta}(X - Y)). \tag{2.59}$$

Similar to (2.57), (2.59) can be approximated as

$$\hat{V}_{\alpha,\beta}(X,Y) = \frac{1}{\tilde{N}} \sum_{i=1}^{\tilde{N}} G_{\alpha,\beta}\left(x_i - y_i\right). \tag{2.60}$$

Some important properties of (2.58) can be identified in [73]. Robust generalized correntropy-based Kalman filters have already been presented in [83] and [84] for dealing with non-Gaussian noise. In [83], a generalized maximum correntropy Kalman filter has been proposed by using the generalized correntropy as the cost function in deriving a linear Kalman filter. In [84], the generalized correntropy is applied to the unscented Kalman filter (UKF) for estimating a power system's state with forecasting aid. The generalized correntropy-based Kalman filters exhibit superior filtering precision in comparison with traditional Kalman filters and maximum correntropy-based Kalman filters [83, 84].

Robust optimization-based nonlinear Kalman filters are in general derived by constructing a novel batch regression model and performing an optimization process. Consider the following state-space model [28]:

$$\mathbf{x}_k = \boldsymbol{f}\left(\mathbf{x}_{k-1}\right) + \mathbf{w}_{k-1} \tag{2.61}$$

$$\mathbf{y}_k = \boldsymbol{h}\left(\mathbf{x}_k\right) + \mathbf{v}_k. \tag{2.62}$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x \times 1}$ and $\mathbf{y}_k \in \mathbb{R}^{n_y \times 1}$ denote the state and measurement, respectively. The process noise $\mathbf{w}_{k-1}$ and the measurement noise $\mathbf{v}_k$ obey the Gaussian distribution, i.e., $\mathbf{w}_{k-1} \sim N(\mathbf{0}, \mathbf{Q}_{k-1})$, $\mathbf{v}_{k-1} \sim N(\mathbf{0}, \mathbf{R}_k)$, respectively. First, the following equation consisting of the predictive state error and measurement error is constructed [86], i.e.,

$$\begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_k \end{bmatrix} \mathbf{x}_k + \bar{\mathbf{v}}_k \tag{2.63}$$

where $\hat{\mathbf{x}}_{k|k-1}$ is the predictive state and $\mathbf{P}_{k|k-1}$ is the corresponding error covariance. The

measurement matrix $\mathbf{H}_k$ in (2.63) is expressed as [86]

$$\mathbf{H}_k = \left(\mathbf{P}_{k|k-1}^{-1}\mathbf{P}_{xy,k|k-1}\right)^T. \tag{2.64}$$

It should be noted that (2.63) utilizes a linearization method for handling non-linear filtering. Apart from linearization, a novel algorithm without approximating the measurement function has been proposed for handling nonlinear filtering [87]. In (2.63), $\mathbf{I} \in \mathbb{R}^n$ is the identity matrix, and $\bar{\mathbf{v}}_k$ takes the form of

$$\bar{\mathbf{v}}_k = \begin{bmatrix} -(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \\ \mathbf{v}_k \end{bmatrix} \tag{2.65}$$

where $\bar{\mathbf{v}}_k$ has a second-order statistical characteristic $E[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^T]$ given by

$$E[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^T] = \begin{bmatrix} \mathbf{P}_{k|k-1} & 0 \\ 0 & \mathbf{R}_k \end{bmatrix} \tag{2.66}$$

$$= \begin{bmatrix} \hat{\mathbf{P}}_{p,k|k-1}\hat{\mathbf{P}}_{p,k|k-1}^T & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{P}}_{v,k}\hat{\mathbf{P}}_{v,k}^T \end{bmatrix} \tag{2.67}$$

$$= \hat{\mathbf{P}}_k\hat{\mathbf{P}}_k^T \tag{2.68}$$

and $\hat{\mathbf{P}}_k$ takes the form of

$$\hat{\mathbf{P}}_k = \begin{bmatrix} \hat{\mathbf{P}}_{p,k|k-1} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{P}}_{v,k} \end{bmatrix}. \tag{2.69}$$

Applying Cholesky decomposition to $E[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^T]$ generates matrix $\hat{\mathbf{P}}_k$ in (2.68). A linear regression model can be developed by multiplying both sides of (2.63) by $\hat{\mathbf{P}}_k^{-1}$, i.e.,

$$\mathbf{\Gamma}_k = \mathbf{\Pi}_k\mathbf{x}_k + \mathbf{e}_k \tag{2.70}$$

where

$$\mathbf{\Gamma}_k = \hat{\mathbf{P}}_k^{-1} \begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} \end{bmatrix}, \mathbf{\Pi}_k = \hat{\mathbf{P}}_k^{-1} \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_k \end{bmatrix}. \tag{2.71}$$

The linear measurement function, i.e., $\boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1}) = \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}$ in (2.71), generates

$$\mathbf{\Gamma}_k = \hat{\mathbf{P}}_k^{-1} \begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k \end{bmatrix}. \tag{2.72}$$

The error $\mathbf{e}_k = \hat{\mathbf{P}}_k^{-1} \bar{\mathbf{v}}_k$ in (2.70) is white for $E[\mathbf{e}_k \mathbf{e}_k^T] = \mathbf{I}$. The optimal value of the state variable is then derived by optimizing a robust function such as the Huber function, maximum correntropy or generalized correntropy over the error in (2.70).

These robust optimization-based Kalman filters are a preferable choice for dealing with non-Gaussian noise. These optimization-based Kalman filters are derived by solving a multi-dimensional optimization problem. As a result, robust Kalman filters can be designed in a more flexible manner compared to other robust Kalman filters like the Student's t-based Kalman filter [64], the particle filter [66] and so on. In addition, these optimization-based Kalman filters in general incur acceptable computational burden in comparison with the original Kalman filters and other robust Kalman filters such as the particle filter [66] and the Gaussian sum filter [65]. Most importantly, these optimization-based Kalman filters all have greatly improved filtering precision in dealing with abnormal measurements in the presence of non-Gaussian noise in comparison with the original Kalman filters.

This thesis focuses on the dynamics of information flow on networks captured using nonlinear Kalman filters in the presence of Gaussian noise and non-Gaussian noise. Nonlinear Kalman filters are capable of capturing the dynamics of information flow by utilizing a state model and a set of available measurements. The state model can be either a propagation dynamic model or a nonlinear system model. In particular,

the propagation dynamic model focuses on describing the dynamics of information flow in applications such as epidemic spreading in human networks [19], rumor propagation in social networks [20, 21], and malware spreading in wireless sensor networks [22]. The propagation dynamics are not deterministic for approximating the characteristics of transmission processes. Consider the epidemic spreading on human networks, for example. Different compartmental models such as Susceptible-Infected-Recovered (SIR), Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Recovered-Susceptible (SIRS) are used for demonstrating the characteristics of epidemic transmission as accurately as possible. Unlike the propagation dynamic model, the nonlinear system model aims to describe the deterministic properties of nonlinear systems. The nonlinear system model is in general a well-established model according to specific characteristics like the electrical characteristics in power systems. The epidemic tracking and the dynamic analysis of power networks are representative applications which consider the propagation dynamics and the dynamics of nonlinear systems, respectively. In this thesis, application of nonlinear Kalman filters is considered for tackling the representative problems associated with epidemic spreading and nonlinear power networks, respectively.

Since the propagation dynamics are not deterministic for approximating the characteristics of transmission processes as precisely as possible, a practical question arises as to whether a guideline can be provided for choosing nonlinear Kalman filters for tracking information spreading on networks, which will be discussed in Chapter 3. Since the dynamic model of power systems is deterministic, the choice of Kalman filters with the dynamic model is not necessary and hence not considered in this thesis. In addition, nonlinear Kalman filters in the applications of epidemic tracking and the dynamic analysis of power systems are confronted with the common issue of practical measurements being contaminated by non-Gaussian noise, such as uniform noise, Laplace noise, and alpha-stable distributed noise. Various kinds of robust Kalman filters are expected to be provided for dealing with tracking epidemic spreading or the

dynamic analysis of power systems under non-Gaussian noise. Consider a sequence of measurement errors where measurements are corrupted by non-Gaussian noise. Different mechanisms within multiple intervals of measurement errors can be designed for dealing with measurement errors. This may be beneficial for improving the filtering precision of Kalman filters but also introduce extra parameters which need to be chosen appropriately. For example, the Huber-based Kalman filters introduce the threshold parameter in (2.54) for determining whether l_1-norm or l_2-norm is considered for weighting measurement errors. It is non-trivial to choose an appropriate threshold parameter in different applications. An alternative way is about only using single mechanism for dealing with all measurement errors. For example, maximum correntropy criterion-based Kalman filters only consider the maximum correntropy for handling measurement errors. In this thesis, single and multiple mechanisms are all considered for handling non-Gaussian noise. More concretely, a novel robust generalized correntropy sparse Gauss-Hermite quadrature filter (GCSGHQF) is presented in Chapter 4. In the generalized correntropy sparse Gauss-Hermite quadrature filter, a single mechanism, i.e., generalized correntropy based procedure, is applied in the sparse Gauss-Hermite quadrature filter for handling all measurement errors to improve robustness. Unlike the generalized correntropy sparse GaussHermite quadrature filter, the mixed p-norm square root unscented Kalman filter studied in Chapter 5 utilizes a piecewise function, i.e., multiple p-norms for handling varying measurement errors. As a result, the mixed p-norm square root unscented Kalman filter is flexible in dealing with corrupted measurements but may be confronted with the choice of multiple parameters in comparison with the generalized correntropy sparse Gauss-Hermite quadrature filter. These robust nonlinear Kalman filters are applied to tracking epidemic spreading and the dynamic analysis of power systems in the presence of non-Gaussian noise. Simulation results demonstrate the improved filtering precision of these novel robust Kalman filters in comparison with traditional Kalman filters.

Chapter 3

Performance Comparison of Nonlinear Kalman Filters in Epidemic Tracking on Networks

The problem of tracking epidemic spreading on networks is relevant to the control of morbidity. The transmission dynamics of an epidemic can be described by a simple compartmental model. Specifically, the estimation of epidemic spreading on networks can be achieved by a nonlinear Kalman filter, which is a tool for state estimation of nonlinear systems. In this section, epidemic spreading on networks is described by compartmental models, such as Susceptible-Infected-Susceptible, Susceptible-Infected-Recovered and Susceptible-Infected-Recovered-Susceptible models. The dynamics of epidemic spreading on homogeneous networks are estimated by several nonlinear Kalman filters, including extended Kalman filter, unscented Kalman filter, third-degree cubature Kalman filter and fifth-degree cubature Kalman filter. The performance comparison in terms of accuracy and stability forms a guideline of utilizing nonlinear Kalman filters for tracking epidemic spreading. The theoretical analysis has been validated through numerical experiments.

3.1 Introduction

Over the past decade, there has been renewed interest in the study of the dynamics of epidemic spreading on networks [19, 88]. The dynamics of epidemic transmission on different types of networks are different. Common types of networks include random networks, e.g., the Erdös and Rényi (ER) network [3, 89], and small-world networks, e.g., the Watts and Strogatz (WS) network [3, 90] and the Newman and Watts (NW) network [5]. The ER network and NW network are generated by adding links randomly. The WS network, however, can be obtained by removing and adding links randomly. In addition, networks can also be categorized into homogeneous and heterogeneous networks according to the degree distribution. For example, homogeneous networks [7], such as ER networks, NW networks and WS networks, possess a Poisson degree distribution, whereas heterogeneous networks like the scale-free network [8] exhibit a power-law degree distribution.

The dynamics of epidemic spreading on networks can be described by various models, such as agent models, time series models and compartmental models [19]. The compartmental model is generally utilized for modeling epidemic transmission for its simplicity and well understood mechanism. In compartmental models, the entire population is classified into different compartments according to the status of the individuals. A string of letters is used to indicate the structural information of the model. For example, Susceptible-Infected-Recovered (SIR) [16, 91] is a common compartmental model for modeling epidemic transmission. For the SIR model, the entire population is divided into three compartments or groups, i.e., S, I and R representing the relative sizes or proportions of the susceptible, infected and recovered individuals. In addition, Susceptible-Infected-Susceptible(SIS) and Susceptible-Infected-Recovered-Susceptible (SIRS) models [17] are also common models for studying the behavior of epidemic spreading. Since complex networks are generally degree-correlated, the degree distribution-based compartmental model [23]

is more relevant for epidemic spreading.

These modeling approaches can be adopted directly for forecasting the dynamic transmission of epidemic, i.e., finding the proportions of the susceptible, infected and recovered individuals as time elapses. However, forecast of dynamic propagation without considering the transmission process in real-world networks will fall short of the desired precision of epidemic tracking. Unlike traditional modeling methods, nonlinear Kalman filtering [30, 31, 32] aims to explore transmission dynamics based on the transmission model of epidemic dynamics and the available measurements. The measurements related to epidemic propagation over networks can be utilized for forecasting the transmission dynamics and further correcting the predictive transmission. In comparison with traditional modeling approaches, nonlinear Kalman filters therefore capture the dynamics of epidemic more accurately.

Commonly used nonlinear Kalman filters include the extended Kalman filter (EKF) [39, 40], unscented Kalman filter (UKF) [43] and cubature Kalman filter (CKF) [46, 45]. The EKF implements state estimation by utilizing the first-order linearization of a nonlinear function. In comparison with the EKF, the UKF utilizes the unscented transformation (UT) for capturing the statistical characteristics of a Gaussian variable through a nonlinear function. However, the UKF becomes unstable as the state dimension increases, degrading the filtering accuracy. In addition, the UKF introduces an adjusted scaling parameter for controlling the spreading of samples. Therefore, the CKF, including the third-degree CKF and the fifth-degree CKF [48], has been proposed without the requirement of introducing any scaling parameter. In addition, the third-degree CKF has a better stability in comparison with the UKF. The fifth-degree CKF achieves a higher precision at the expense of a higher complexity and being less stable than the third-degree CKF.

In this chapter, several nonlinear Kalman filters including the EKF, UKF and CKF are employed to capture the dynamic behavior of epidemic spreading on homogeneous networks. Specifically, the main contributions are as follows.

- Nonlinear Kalman filters are utilized for tracking epidemic spreading over complex networks on the basis of degree-based compartmental models. Since degree-based compartmental models involve a high dimensional state vector and incur heavy computational burden for large networks, several strategies are proposed for the efficient use of nonlinear Kalman filters in tracking epidemic transmission over large-scale complex networks.

- A comparison of various nonlinear Kalman filters is performed in terms of estimation accuracy and stability. The performance comparison provides a guideline for choosing appropriate nonlinear Kalman filters for tracking the transmission dynamics of epidemic over complex networks.

3.2 State Estimation of Epidemic Spreading on Networks

3.2.1 Review of Networks

A network can be described as a graph $G = (N, E)$, where N and E denote the sets of nodes and links in graph G [55]. Let $|N|$ and $|E|$ be the size of N and E, respectively. Common types of networks include ER networks [3], NW networks [4] and WS networks [5, 6]. The topological structure of an ER network is determined by the number of nodes $|N|$ and the connecting probability p among nodes, i.e., $(|N|, p)$ where the connecting probability p determines whether the nodes are connected. For example, the nodes in an ER network are connected when a randomly generated number is smaller than the pre-determined connecting probability p in the evolution of the ER network. In comparison with ER networks, NW and WS networks are generated based on the nearest-neighbor coupling network with $2K$ nearest-neighbor coupled nodes [92] and are therefore determined by $(|N|, K, p)$.

Different networks exhibit different properties such as the degree distribution [7]. Define a degree vector $\mathbf{d} = [d_1, d_2, \cdots, d_j, \cdots, d_{|N|}]$. Degree d_j of node j is the number of links incident on the node [7], i.e.,

$$d_j = \sum_i a_{ji} \tag{3.1}$$

where $a_{ji} \in \{0, 1\}$ is the j-th row i-th column element of the adjacency matrix $\mathbf{A}$ with $a_{ji} = 1$ denoting connectivity of node j with node i. The degree of graph G can also be represented by a collection of the types of degree $\mathbf{D} = [d_{p_1}, d_{p_2}, \cdots, d_{p_\varsigma}, \cdots, d_{p_{|\mathbf{D}|}}]$ with $|\mathbf{D}|$ denoting the size of the types of degree and d_{p_ς} being the ς-th type of degree. The set $\{d_{p_\varsigma}\}_{\varsigma=1}^{|\mathbf{D}|}$, is a subset of $\{d_j\}_{j=1}^{|N|}$. The degree distribution $P(d_{p_\varsigma})$ is defined as the probability that a node has degree d_{p_ς}. In addition, Fig. 3.1 shows the plot of $|\mathbf{D}|$ versus probability p in various networks. From Fig. 3.1, $|\mathbf{D}|$ is generally small for a small or large connecting probability and otherwise becomes large in ER and NW networks. However, $|\mathbf{D}|$ increases with p for the WS network.

3.2.2 State Space Model

The compartmental model [19] divides the population into different compartments. The most commonly used two-compartment model is the SIS model [17], and three-compartment models include the SIR [16] and SIRS [17] models. Define the proportions of the infected, susceptible and recovered individuals having degree d_{p_ς} at discrete time k by $I_{d_{p_\varsigma},k}$, $S_{d_{p_\varsigma},k}$ and $R_{d_{p_\varsigma},k}$, respectively. The SIS, SIR and SIRS models are represented by

$$S_{d_{p_\varsigma},k} = S_{d_{p_\varsigma},k-1} - \alpha d_{p_\varsigma} S_{d_{p_\varsigma},k-1} I^w_{d_{p_\varsigma},k-1} + \Delta S_{k-1} \tag{3.2}$$

$$I_{d_{p_\varsigma},k} = I_{d_{p_\varsigma},k-1} + \alpha d_{p_\varsigma} S_{d_{p_\varsigma},k-1} I^w_{d_{p_\varsigma},k-1} - \Delta I_{k-1} \tag{3.3}$$

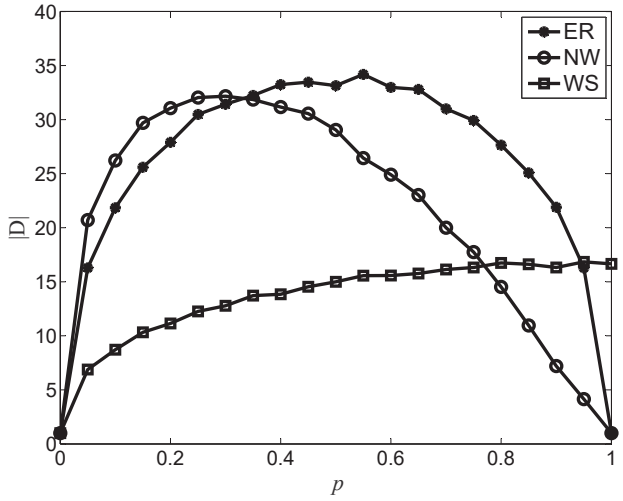


Figure 3.1: |**D**| versus connection probability p in networks.

with $\Delta S_{k-1} = \Delta I_{k-1} = \epsilon I_{d_{p_\varsigma},k-1}$ for the SIS model; $\Delta S_{k-1} = 0, \Delta I_{k-1} = \beta I_{d_{p_\varsigma},k-1}$ for the SIR model; $\Delta S_{k-1} = \gamma R_{d_{p_\varsigma},k-1}, \Delta I_{k-1} = \beta I_{d_{p_\varsigma},k-1}$ for the SIRS model. In (3.2),

$$I^w_{d_{p_\varsigma},k-1} = \sum_{m=1}^{|\mathbf{D}|} p\left(d_{p_m}|d_{p_\varsigma}\right) I_{d_{p_\varsigma},k-1} \tag{3.4}$$

where the scalar d_{p_ς} denotes the degree, ranging from d_{p_1} to $d_{p_{|\mathbf{D}|}}$, and $p(d_{p_m}|d_{p_\varsigma})$ is the conditional probability distribution that an edge departing from a node of degree d_{p_ς} is connected to a node of degree d_{p_m}. The SIS model as a closed-loop model only includes two basic events, i.e., infection and susceptibility, with α being the infected rate and ϵ being the transmission rate from the infected to the susceptible. Similarly, the SIRS model is also a closed-loop model with α being the infected rate, β being the removal or recovery rate and γ being the transmission rate from the recovered to the susceptible. However, the SIR model is an open-loop model with α being the infected rate and β being the removal or recovery rate.

Nonlinear Kalman filters perform state estimation based on a state-space model, i.e.,

$$\mathbf{x}_k = \boldsymbol{f}\left(\mathbf{x}_{k-1}\right) + \mathbf{w}_k \tag{3.5}$$

$$\mathbf{y}_k = \boldsymbol{h}(\mathbf{x}_k) + \mathbf{v}_k \tag{3.6}$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ and $\mathbf{y}_k \in \mathbb{R}^{n_y}$ denote the state and the measurement, respectively; the nonlinear functions $\boldsymbol{f}(\cdot)$ and $\boldsymbol{h}(\cdot)$ are the state function and measurement function, respectively. The process noise $\mathbf{w}_k$ and the measurement noise $\mathbf{v}_k$ are both uncorrelated zero-mean Gaussian white noise with covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively. A nonlinear Kalman filter calculates the predictive and estimated density functions of the state, $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ and $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, by prediction step and update step, respectively. Different methods for calculating the probability density function lead to different nonlinear Kalman filters, e.g., EKF, UKF and CKF. In comparison with the EKF utilizing the first-order linearization approximation of a nonlinear function, UKF and CKF are common nonlinear Kalman filters based on deterministic sampling. Nonlinear Kalman filters based on deterministic sampling are summarized in **Algorithm** 3 [43, 46].

To use nonlinear Kalman filters for tracking epidemic spreading, we define the state vector $\mathbf{x}_k = [(\mathbf{x}_k^1)^{\mathrm{T}}\ (\mathbf{x}_k^2)^{\mathrm{T}}]^{\mathrm{T}}$, where $\mathbf{x}_k^1$ and $\mathbf{x}_k^2$ are defined as according to (3.2) and (3.3)

$$\mathbf{x}_k^1 = \mathbf{S}_k = [S_{d_{p_1},k}, S_{d_{p_2},k}, \cdots, S_{d_{p_\varsigma},k}, \cdots, S_{d_{p_{|\mathbf{D}|}},k}]^{\mathrm{T}} \tag{3.7}$$

$$\mathbf{x}_k^2 = \mathbf{I}_k = [I_{d_{p_1},k}, I_{d_{p_2},k}, \cdots, I_{d_{p_\varsigma},k}, \cdots, I_{d_{p_{|\mathbf{D}|}},k}]^{\mathrm{T}} \tag{3.8}$$

where $I_{d_{p_\varsigma},k}$ and $S_{d_{p_\varsigma},k}$ represent the proportions of the infected and susceptible individuals having degree d_{p_ς} at discrete time k, respectively. The SIS, SIR and SIRS compartmental models in (3.2) and (3.3) can be rewritten as a state-space model by defining the state process function and measurement function, i.e.,

$$\mathbf{x}_k^1 = \mathbf{x}_{k-1}^1 - \alpha\mathbf{x}_{k-1}^1 \circ (\mathbf{P}\mathbf{x}_{k-1}^2) + \boldsymbol{\Pi}_{k-1} + \mathbf{w}_k^1 \tag{3.9}$$

$$\mathbf{x}_k^2 = \mathbf{x}_{k-1}^2 + \alpha\mathbf{x}_{k-1}^1 \circ (\mathbf{P}\mathbf{x}_{k-1}^2) - \boldsymbol{\Xi}_{k-1} + \mathbf{w}_k^2 \tag{3.10}$$

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \tag{3.11}$$

with $\mathbf{\Pi}_{k-1} = \epsilon\mathbf{x}_{k-1}^2$ and $\mathbf{\Xi}_{k-1} = \epsilon\mathbf{x}_{k-1}^2$ for the SIS model; $\mathbf{\Pi}_{k-1} = \mathbf{0}$ and $\mathbf{\Xi}_{k-1} = \beta\mathbf{x}_{k-1}^2$ for the SIR model; $\mathbf{\Pi}_{k-1} = \gamma(\bar{\mathbf{\Gamma}} - \mathbf{x}_{k-1}^1 - \mathbf{x}_{k-1}^2)$ and $\mathbf{\Xi}_{k-1} = \beta\mathbf{x}_{k-1}^2$ with $\bar{\mathbf{\Gamma}} = [1, 1, \cdots, 1]^{\mathrm{T}} \in \mathbb{R}^{|\mathbf{D}|\times 1}$ for the SIRS model. In (3.9), $\mathbf{P} = diag(\mathbf{D})\mathbf{P}_c$, where $diag(\mathbf{D})$ is a diagonal matrix with entries taken from $\mathbf{D}$. The i-th row j-th column element $p(d_{p_j}|d_{p_i})$ of $\mathbf{P}_c$ denotes the conditional probability that a node of degree d_{p_i} connects a node of degree d_{p_j}, and is dependent on the node with degree d_{p_i}. Also, $\tilde{\mathbf{C}} = \tilde{\mathbf{A}} \circ \tilde{\mathbf{B}}$ is the Hadamard product [93] and $\mathbf{w}_k = [(\mathbf{w}_k^1)^{\mathrm{T}} \quad (\mathbf{w}_k^2)^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{w}_k^1, \mathbf{w}_k^2 \in \mathbb{R}^{|\mathbf{D}|\times 1}$. The measurement matrix $\mathbf{H}$ is chosen by $\mathbf{H} = [\mathbf{\Gamma}, \mathbf{\Gamma}]$ with identity matrix $\mathbf{\Gamma} \in \mathbb{R}^{|\mathbf{D}|}$.

In (3.9) and (3.10), parameters like α, β, ϵ and γ are assumed to be stationary. Nonlinear Kalman filters not only consider epidemic spreading model but also take into account the transmission process in practice. As a result, nonlinear Kalman filters still perform satisfactorily even when stationary parameters are used. Apart from the work using stationary parameters, there are numerous studying using updated parameters like machine learning [15] and double nonlinear Kalman filters [30]. However, machine learning is limited by the requirement of a large number of samples of high quality. In addition, it is challenging for double nonlinear Kalman filters to yield an accurate state as well as exact parameters. In our work, stationary parameters are applied in nonlinear Kalman filters for tracking epidemic spreading. In addition, the connection probability p in Fig. 3.1 is used to describe the connectivity of nodes and is also stationary for a specific network.

Remark 1 *In ER and NW networks, the number of links $|E|$ increases with the connecting probability p. Unlike ER and NW networks, WS networks can be generated by removing and connecting randomly. Therefore, the values of* $\mathbf{P}$ *would likely increase as the connecting probability increases in ER and NW networks, and would generally decrease in the WS network. Morevoer, the increase of* $\mathbf{P}$ *may accelerate the infection process and degrade the numerical stability of nonlinear Kalman filters for performing epidemic tracking, thus deteriorating the filtering accuracy.*

Algorithm 3 Nonlinear Kalman Filter Based on Deterministic Sampling

Prediction step

1: Calculate L sample points $\{\boldsymbol{\xi}_{i,k-1}\}_{i=1}^{L}$ and weights $\{\omega_i\}_{i=1}^{L}$ where $\boldsymbol{\xi}_{i,k-1}$ and ω_i represent the i-th sample point and i-th weight at the discrete time $k-1$ [43, 46].

2: Compute the predicted mean $\hat{\mathbf{x}}_{k|k-1}$ and predicted error covariance $\hat{\mathbf{P}}_{k|k-1}$

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=1}^{L} \omega_i \boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}) \tag{3.12}$$

$$\hat{\mathbf{P}}_{k|k-1} = \hat{\mathbf{P}}^{s}_{k|k-1} + \mathbf{Q}_k \tag{3.13}$$

with the matrix $\hat{\mathbf{P}}^{s}_{k|k-1}$

$$\hat{\mathbf{P}}^{s}_{k|k-1} = \sum_{i=1}^{L} \omega_i (\boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k|k-1})(\boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}}. \tag{3.14}$$

Update step

1: Calculate L sample points $\{\boldsymbol{\xi}_{i,k-1}\}_{i=1}^{L}$ and weights $\{\omega_i\}_{i=1}^{L}$ [43, 46].

2: Compute the estimated mean $\hat{\mathbf{x}}_k$ and the corresponding error covariance of the state $\hat{\mathbf{P}}_k$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1})), \tag{3.15}$$

$$\hat{\mathbf{P}}_k = \hat{\mathbf{P}}_{k|k-1} - \mathbf{K}_k \hat{\mathbf{P}}_{yy,k|k-1} \mathbf{K}_k^{T}. \tag{3.16}$$

The matrix $\mathbf{K}_k = \hat{\mathbf{P}}_{xy,k|k-1} \hat{\mathbf{P}}^{-1}_{yy,k|k-1}$ represents the Kalman gain with covariance matrices $\hat{\mathbf{P}}_{xy,k|k-1}$ and $\hat{\mathbf{P}}_{yy,k|k-1}$

$$\hat{\mathbf{P}}_{xy,k|k-1} = \sum_{i=1}^{L} \omega_i (\boldsymbol{\xi}_{i,k|k-1} - \hat{\mathbf{x}}_{k|k-1})(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}} \tag{3.17}$$

$$\hat{\mathbf{P}}_{yy,k|k-1} = \mathbf{R}_k + \sum_{i=1}^{L} \omega_i (\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}} \tag{3.18}$$

and predicted measurement

$$\hat{\mathbf{y}}_{k|k-1} = \sum_{i=1}^{L} \omega_i \boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}). \tag{3.19}$$

Output: $\hat{\mathbf{x}}_k$, $\hat{\mathbf{P}}_k$

Remark 2 *The filtering accuracy of nonlinear Kalman filters can be affected by the dimension of the state [45]. In general, a high state dimension greatly degrades the filtering precision of nonlinear Kalman filters. Fig. 3.1 gives the size of the types of*

degree in ER, NW and WS networks, which is positively correlated with the dimension of the hidden state. Therefore, nonlinear Kalman filters perform epidemic tracking with a smaller estimation error when ER or NW networks are generated by a smaller or larger connecting probability and when WS networks are generated by a smaller connecting probability. In [43], an extra matrix $\Delta\mathbf{Q}$ *is added to the covariance matrix of estimation error for enhancing the stability at the expense of degrading the filtering performance.*

3.2.3 Observability of Discrete-Time Models

The observability property can be used for evaluating the existence of an estimator of a model [39]. More concretely, a system is observable at discrete time k if for the state at discrete time k, there is a finite $k_l > k$ so that there are sufficient amount of measurements for determining the state at discrete time k [94]. It is also easy to understand that an unobservable system is one where some states at discrete time k cannot be determined from measurements regardless of the number of measurements taken [95]. **Theorem 1** gives sufficient conditions for guaranteeing the observability of the model given in (3.5) and (3.6).

Theorem 1 *The nonlinear dynamical system given by (3.9)-(3.11) is locally weakly observable.*

Proof: According to the results of [39], the nonlinear dynamical system given by (3.5) and (3.6) is locally weakly observable if the nonlinear observability matrix at $\mathbf{x}_k \in \mathbb{R}^{n_x}$, i.e.,

$$U(\mathbf{x}_k) := \begin{pmatrix} \frac{\partial \boldsymbol{h}(\mathbf{x}_k)}{\partial \mathbf{x}} \\ \vdots \\ \frac{\partial \mathbf{h}\left(\mathbf{x}_{k+n_x-1}\right)}{\partial \mathbf{x}} \frac{\partial \boldsymbol{f}\left(\mathbf{x}_{k+n_x-2}\right)}{\partial \mathbf{x}} \cdots \frac{\partial \boldsymbol{f}(\mathbf{x}_k)}{\partial \mathbf{x}} \end{pmatrix} \tag{3.20}$$

has rank n_x. The derivative $\partial \boldsymbol{h}(\mathbf{x})/\partial \mathbf{x}|_{\mathbf{x}=\mathbf{x}_k}$ in (3.20) is calculated by $\partial \boldsymbol{h}(\mathbf{x})/\partial \mathbf{x}|_{\mathbf{x}=\mathbf{x}_k} = \mathbf{H}_k$. In addition, we have the derivative $\mathbf{F}_{k-1} = \partial \boldsymbol{f}(\mathbf{x})/\partial \mathbf{x}$ for SIS, SIR and SIRS models as

$$\mathbf{F}_{k-1} = \begin{bmatrix} \Delta\mathbf{F}_a - \alpha\boldsymbol{\Gamma} \circ \tilde{\mathbf{P}}_{\mathbf{I}} & \Delta\mathbf{F}_b - \alpha\tilde{\mathbf{P}}_{\mathbf{S}} \\ \alpha\boldsymbol{\Gamma} \circ \tilde{\mathbf{P}}_{\mathbf{I}} & \Delta\mathbf{F}_c + \alpha\tilde{\mathbf{P}}_{\mathbf{S}} \end{bmatrix} \tag{3.21}$$

with $\tilde{\mathbf{P}}_{\mathbf{I}}$ and $\tilde{\mathbf{P}}_{\mathbf{S}}$ defined by $\tilde{\mathbf{P}}_{\mathbf{I}} = \mathbf{P}[\mathbf{I}_{k-1} \cdots \mathbf{I}_{k-1}], \tilde{\mathbf{P}}_{\mathbf{S}} = \mathbf{P} \circ [\mathbf{S}_{k-1} \cdots \mathbf{S}_{k-1}]$. In (3.21), for the SIS model,

$$\Delta\mathbf{F}_a = \boldsymbol{\Gamma}, \Delta\mathbf{F}_b = \epsilon\boldsymbol{\Gamma}, \Delta\mathbf{F}_c = (1-\epsilon)\boldsymbol{\Gamma}. \tag{3.22}$$

As for the SIR model, we have

$$\Delta\mathbf{F}_a = \boldsymbol{\Gamma}, \Delta\mathbf{F}_b = \mathbf{0}, \Delta\mathbf{F}_c = (1-\beta)\boldsymbol{\Gamma}. \tag{3.23}$$

For the SIRS model,

$$\Delta\mathbf{F}_a = (1-\gamma)\boldsymbol{\Gamma}, \Delta\mathbf{F}_b = -\gamma\boldsymbol{\Gamma}, \Delta\mathbf{F}_c = (1-\beta)\boldsymbol{\Gamma}. \tag{3.24}$$

Therefore, we get $\text{rank}(U(\mathbf{x}_k)) = n_x = 2|\mathbf{D}|$. ■

Remark 3 *The nonlinear Kalman filter performs state estimation based on SIS, SIR and SIRS models. The SIS and SIRS models are closed-loop models, for which the proportions of the susceptible, infected and recovered individuals may not reduce to zero. However, the SIR compartmental model is an open-loop model, for which the proportions of the susceptible and infected people may become zero. Therefore, all entries of predictive covariance matrix $\hat{\mathbf{P}}_{k|k-1}$ in (3.13) may be close to zero under a small covariance matrix $\mathbf{Q}_k$. The estimated error covariance $\hat{\mathbf{P}}_k$ may thus be non-positive definite since the estimated error covariance $\hat{\mathbf{P}}_k$ is constructed by subtraction of predictive covariance matrix $\hat{\mathbf{P}}_{k|k-1}$ and $\hat{\mathbf{P}}^K_{k|k-1} = \mathbf{K}_k\hat{\mathbf{P}}_{yy,k|k-1}\mathbf{K}_k^T$ in (3.16). Since the*

sample points $\{\boldsymbol{\xi}_{i,k-1}\}_{i=1}^{L}$ *in* ***Algorithm*** *3 are generated by implementing the Cholesky decomposition on matrix* $\hat{\mathbf{P}}_k$*, the non-positive definite matrix* $\hat{\mathbf{P}}_k$ *is not beneficial for nonlinear Kalman filters choosing sample points* $\{\boldsymbol{\xi}_{i,k-1}\}_{i=1}^{L}$*. Therefore, SIS and SIRS models are more beneficial for nonlinear Kalman filters in performing state estimation comparing to the SIR model in terms of filtering stability.*

3.3 Accuracy, Complexity and Steady-State Analysis

3.3.1 Accuracy Analysis

Define the column vector $\mathbf{x} = [x_1, x_2, \cdots, x_j, \cdots, x_{n_x}]^{\mathrm{T}} \in \mathbb{R}^{n_x}$, $\mathbf{x} = \bar{\mathbf{x}} + \Delta\bar{\mathbf{x}} = \bar{\mathbf{x}} + \tilde{\mathbf{A}}\Delta\mathbf{x}$ with $\mathbf{x} \sim \mathbb{N}(\bar{\mathbf{x}}, \mathbf{P})$, $\Delta\bar{\mathbf{x}} \sim \mathbb{N}(\mathbf{0}, \mathbf{P})$, $\tilde{\mathbf{A}} = \sqrt{\mathbf{P}}$, $\Delta\bar{\mathbf{x}} = [\Delta\bar{x}_1, \Delta\bar{x}_2, \cdots, \Delta\bar{x}_j, \cdots, \Delta\bar{x}_{n_x}]^{\mathrm{T}}$. Expanding the function $f(\mathbf{x})$ at $\bar{\mathbf{x}}$ by Taylor series expansion, we get [43]

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + \sum_{j=1}^{\infty} \frac{D_{\Delta\bar{\mathbf{x}}}^{j} f}{j!} \tag{3.25}$$

where

$$D_{\Delta\bar{\mathbf{x}}}^{j} f = \left[\sum_{\bar{i}=1}^{n_x} \Delta\bar{x}_{\bar{i}} \sum_{\bar{j}=1}^{n} a_{\bar{i}\bar{j}} \frac{\partial}{\partial x_{\bar{j}}} \right]^{j} f \mid_{\mathbf{x}=\bar{\mathbf{x}}} \tag{3.26}$$

with $a_{\bar{i}\bar{j}}$ being the $\bar{i}$-row $\bar{j}$-column element of matrix $\tilde{\mathbf{A}}$. Defining $\nabla = \left[\frac{\partial}{\partial x_1} \quad \frac{\partial}{\partial x_2} \cdots \frac{\partial}{\partial x_{n_x}} \right]^{\mathrm{T}}$, we have

$$\begin{aligned} D_{\Delta\bar{\mathbf{x}}} f &= (\Delta\bar{\mathbf{x}})^{\mathrm{T}} \nabla f, \\ D_{\Delta\bar{\mathbf{x}}}^{2} f &= \left(\nabla^{\mathrm{T}} \Delta\bar{\mathbf{x}} \Delta\bar{\mathbf{x}}^{\mathrm{T}} \nabla \right) f. \end{aligned} \tag{3.27}$$

The expectation of (3.25) is given by [46]

$$f^m(\mathbf{x}) = f(\bar{\mathbf{x}}) + \frac{1}{2}\nabla^T \mathbf{P}_{\mathbf{x}} \nabla + E\bar{\mathbf{x}}_c + \sum_{j=2}^{\infty} \frac{\hat{\tau}}{(2j)!} \sum_{\hat{j}=1}^{n_x} \left(\sum_{\bar{j}=1}^{n_x} a_{\hat{j},\bar{j}} \frac{\partial}{\partial x_{\bar{j}}} \right)^{2j} f \mid_{\mathbf{x}=\bar{\mathbf{x}}} \tag{3.28}$$

where $\hat{\tau} = 1 \times 3 \times \cdots (2j-1)$ and $E\bar{\mathbf{x}}_c$ denotes the cross term. Also, L samples of the UKF, third-degree CKF and fifth-degree CKF take the form $\{\boldsymbol{\xi} : \bar{\mathbf{x}} + \Delta\bar{\boldsymbol{\xi}}_i = \bar{\mathbf{x}} + \bar{\mathbf{A}}\Delta\boldsymbol{\xi}_i\}_{i=1}^{L}$ where $\bar{\mathbf{A}} = \sqrt{\mathbf{P}_{\mathbf{x}}}$ and the mean and variance of $\Delta\bar{\boldsymbol{\xi}}_i$ are $\mathbf{0}$ and $\mathbf{P}_{\mathbf{x}}$, respectively [48]. Similar to (3.28), the UKF generates the mean of state prediction [46], i.e.,

$$\bar{\mathbf{x}}_{k+1|k}^{\mathrm{UKF}} = f(\bar{\mathbf{x}}) + \sum_{j=2}^{\infty} \frac{(n_x+\lambda)^{j-1}}{(2j)!} \sum_{\hat{j}=1}^{n_x} \left(\sum_{\bar{j}=1}^{n_x} a_{\hat{j}\bar{j}} \frac{\partial}{\partial \xi_{\bar{j}}} \right)^{2j} f \mid_{\boldsymbol{\xi}=\bar{\mathbf{x}}} + \frac{1}{2}\nabla^{\mathrm{T}} \mathbf{P}_{\mathbf{x}} \nabla + E\bar{\mathbf{x}}_u \tag{3.29}$$

where $E\bar{\mathbf{x}}_u$ is the cross term of state prediction of the UKF and the factor λ is an adjustable parameter for sampling points and weights [43]. Similarly, the third-degree CKF generates the mean of state prediction as

$$\bar{\mathbf{x}}_{k+1|k}^{\mathrm{CKF3}} = f(\bar{\mathbf{x}}) + \sum_{j=2}^{\infty} \frac{n_x^{j-1}}{(2j)!} \sum_{\hat{j}=1}^{n_x} \left(\sum_{\bar{j}=1}^{n_x} a_{\hat{j}\bar{j}} \frac{\partial}{\partial \xi_{\bar{j}}} \right)^{2j} f \mid_{\boldsymbol{\xi}=\bar{\mathbf{x}}} + \frac{1}{2}\nabla^{\mathrm{T}} \mathbf{P}_{\mathbf{x}} \nabla + E\bar{\mathbf{x}}_{c3} \tag{3.30}$$

where $E\bar{\mathbf{x}}_{c3}$ is the cross term of state prediction of the third-degree CKF. In addition, the mean of state prediction for the fifth-degree CKF is

$$\bar{\mathbf{x}}_{k+1|k}^{\mathrm{CKF5}} = f(\bar{\mathbf{x}}) + \frac{1}{2}\nabla^{\mathrm{T}} \mathbf{P}_{\mathbf{x}} \nabla + \hbar_b + \hbar_l \tag{3.31}$$

where $\hbar_b$ and $\hbar_l$ can be calculated by $\hbar_b = 2(n_x - 1)\tau + E\bar{\mathbf{x}}_{c51}$ and $\hbar_l = (4 - n_x)\tau + E\bar{\mathbf{x}}_{c52}$ where $\tau = \sum_{j=2}^{\infty} (n_x+2)^{j-2} \sum_{\hat{j}=1}^{n_x} \left(\sum_{\bar{j}=1}^{n_x} a_{\hat{j}\bar{j}} \frac{\partial}{\partial \xi_{\bar{j}}} \right)^{2j} f/(2j)!$. Therefore, the fifth-degree CKF generates the mean of state prediction as

$$\bar{\mathbf{x}}_{k+1|k}^{\mathrm{CKF5}} = f(\bar{\mathbf{x}}) + \frac{1}{2}\nabla^{\mathrm{T}} \mathbf{P}_{\mathbf{x}} \nabla + E\bar{\mathbf{x}}_{c5} + \sum_{j=2}^{\infty} \frac{(n_x+2)^{j-1}}{(2j)!} \sum_{\hat{j}=1}^{n_x} \left(\sum_{\bar{j}=1}^{n_x} a_{\hat{j}\bar{j}} \frac{\partial}{\partial \xi_{\bar{j}}} \right)^{2j} f \mid_{\boldsymbol{\xi}=\bar{\mathbf{x}}} \tag{3.32}$$

with the cross term $E\bar{\mathbf{x}}_{c5}$.

Remark 4 *From (3.28) to (3.32), the performances of UKF, third-order CKF, and fifth-order CKF are compared through Taylor series expansion in terms of the captured high-order statistical characteristic. For example, parameter λ in UKF is generally set as $n_x + \lambda = 3$ such that the UKF can capture the fourth-order statistical characteristic, i.e., $j = 2$ in (3.28). Therefore, in comparison with the third-degree CKF, UKF gives a higher filtering accuracy. As for the cubature rule, if $d_1 > d_2$, the d_1-degree cubature rule is more accurate than the d_2-degree cubature rule. Therefore, the fifth-degree CKF has a much higher filtering precision in comparison with the third-degree CKF. In addition, the fifth-degree CKF performs better than the third-degree CKF and UKF especially in the presence of high nonlinearities.*

3.3.2 Complexity Analysis

The computational complexity of nonlinear Kalman filters based on deterministic sampling is provided in Table 3.1. In Table 3.1, n_x and n_y represent the dimension of the state and measurement, respectively. Notations C_f and C_h denote the calculational burden of nonlinear functions f and h acting on one scalar, respectively. Nonlinear Kalman filters perform the dynamic estimation by two steps, i.e., prediction step and update step. In prediction step, L samples $\{\boldsymbol{\xi}_{i,k-1}\}_{i=1}^{L}$ and corresponding weight $\omega_i^{(c)}$ and $\omega_i^{(m)}$ at discrete time $k-1$ are used for calculating integrals by numerical method as shown in Algorithm 2 in Chapter 2. In update step, L samples $\{\boldsymbol{\xi}_{i,k|k-1}\}_{i=1}^{L}$ and corresponding weight $\omega_i^{(c)}$ and $\omega_i^{(m)}$ at discrete time $k-1$ are also required for computing integrals by numerical method in Algorithm 2 in Chapter 2. According to Table 3.1, complexities of nonlinear Kalman filters like UKF, third-degree CKF and fifth-degree CKF are all $O((L+n_x)n_x^2 + n_y^3 + n_x^2 n_y + n_y^2 n_x)$. Since the dimension of the measurement is lower than that of the state, these nonlinear Kalman filters all have complexity $O((L+n_x)n_x^2)$ in terms of scale.

Table 3.1: Computational Complexity of Nonlinear Kalman Filter Based on Deterministic Sampling

Computation	Operation	Cost of Operation	Overall Cost
Sample points and weights $\{\boldsymbol{\xi}_{i,k-1}, \omega_i^{(c)}, \omega_i^{(m)}\}_{i=1}^{L}$ in [43, 46]	Cholupdate decomposition	$O(n_x^3)$	$O(n_x^3)$
$\hat{\mathbf{x}}_{k\|k-1} = \sum_{i=1}^{L} \omega_i^{(m)} f(\boldsymbol{\xi}_{i,k-1})$	Evaluation of f	$L \times n_x \times C_f$	
	Scalar-vector products	$L \times n_x$	
	Sum of vectors	$(L-1) \times n_x$	$O(2Ln_x + Ln_xC_f - n_x)$
$\mathbf{P}_{k\|k-1} = \sum_{i=1}^{L} \omega_i^{(c)} (f(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k\|k-1})$ $(\boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k\|k-1})^{\mathrm{T}} + \mathbf{Q}_{k-1}$	Subtract of vectors $f(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k\|k-1}$	$L \times n_x$	
	Product of vectors $(f(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k\|k-1})(f(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k\|k-1})^{\mathrm{T}}$	$L \times n_x^2$	
	Scalar-Matrix multiplication $\omega_i^{(c)} (f(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k\|k-1})(f(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k\|k-1})^{\mathrm{T}}$	$L \times n_x^2$	
	Sum of matrices	$L \times n_x^2$	$O(3Ln_x^2 + Ln_x)$
Sample points and weights $\{\boldsymbol{\xi}_{i,k|k-1}, \omega_i^{(c)}, \omega_i^{(m)}\}_{i=1}^{L}$ in [43, 46]	Cholupdate decomposition	$O(n_x^3)$	$O(n_x^3)$
$\hat{\mathbf{y}}_{k|k-1} = \sum_{i=1}^{L} \omega_i^{(m)} h(\boldsymbol{\xi}_{i,k|k-1})$	Evaluation of h	$L \times n_x \times C_f$	
	Scalar-vector products	$L \times n_y$	
	Sum of vectors	$(L-1) \times n_y$	$O(2Ln_y + Ln_xC_f - n_y)$
$\mathbf{P}_{yy,k|k-1} = \sum_{i=1}^{L} \omega_i^{(c)} (\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})$ $(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}} + \mathbf{R}_k$	Subtract of vectors $\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1}$	$L \times n_y$	
	Product of vectors $(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}}$	$L \times n_y^2$	
	Scalar-Matrix multiplication $\omega_i^{(c)} (\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}}$	$L \times n_y^2$	
	Sum of matrices	$(L-1) \times n_y^2$	$O(3Ln_y^2 + Ln_y - n_y^2)$
$\mathbf{P}_{xy,k|k-1} = \sum_{i=1}^{L} \omega_i^{(c)} (\boldsymbol{\xi}_{i,k|k-1} - \hat{\mathbf{x}}_{k|k-1})$ $(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}}$	Subtract of vectors $\boldsymbol{\xi}_{i,k-1} - \hat{\mathbf{x}}_{k|k-1}$	$L \times n_x$	
	Product of vectors $(\boldsymbol{\xi}_{i,k-1} - \hat{\mathbf{x}}_{k|k-1})(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}}$	$L \times n_x \times n_y$	
	Scalar-Matrix multiplication $\omega_i^{(c)} (\boldsymbol{\xi}_{i,k-1} - \hat{\mathbf{x}}_{k|k-1})(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}}$	$L \times n_x \times n_y$	
	Sum of matrices	$(L-1) \times n_x \times n_y$	$O(3Ln_xn_y + Ln_x - n_xn_y)$
$\mathbf{K}_k = \mathbf{P}_{xy,k|k-1} \mathbf{P}_{yy,k|k-1}^{-1}$	Inverse of matrix $\mathbf{P}_{yy,k|k-1}^{-1}$	$O(n_y^3)$	
	Matrix-matrix multiplication $\mathbf{P}_{xy,k|k-1} \mathbf{P}_{yy,k|k-1}^{-1}$	$n_x \times n_y^2 + n_x \times (n_y^2 - n_y)$	$O(n_y^3 + n_xn_y^2)$
$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - h(\hat{\mathbf{x}}_{k|k-1}))$	Subtract of vectors $\mathbf{y}_k - h(\hat{\mathbf{x}}_{k|k-1})$	n_y	
	Matrix-vector multiplication $\mathbf{K}_k (\mathbf{y}_k - h(\hat{\mathbf{x}}_{k|k-1}))$	$2n_xn_y^2 - n_xn_y$	
	Sum of vectors $\hat{\mathbf{x}}_{k-1} + \mathbf{K}_k (\mathbf{y}_k - h(\hat{\mathbf{x}}_{k|k-1}))$	n_x	$2n_xn_y^2 - n_xn_y + n_x + n_y$
$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{yy,k|k-1} \mathbf{K}_k^T$	Multiplication of matrices $\mathbf{K}_k \mathbf{P}_{yy,k|k-1} \mathbf{K}_k^T$	$2n_yn_x^2 - n_xn_y + 2n_x^2n_y - n_x^2$	
	Subtract of matrices $\mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{yy,k|k-1} \mathbf{K}_k^{\mathrm{T}}$	n_x^2	$O(2n_yn_x^2 - n_xn_y + 2n_x^2n_y)$

However, it is necessary to note that the fifth-degree CKF uses $L = 2n_x^2 + 1$ sample points for calculating integrals in comparison with the third-degree CKF and UKF with $L = 2n_x$ and $L = 2n_x + 1$, respectively. This indicates that the fifth-degree CKF has the highest burden $O(n_x^4)$ in comparison with the third-degree CKF and UKF with complexity $O(n_x^3)$ in terms of scale. Although the third-degree CKF and UKF all have complexity $O(n_x^3)$, the third-degree CKF and UKF have different complexity in terms of computing flops. For example, UKF with $2n_x + 1$ samples has a slightly higher computational burden in comparison with the third-degree CKF with $2n_x$ sample points. By contrast, the extended Kalman filter calculates multiple integrals by the first-

order linearization of a nonlinear function instead of the deterministic sampling method given in Algorithm 1 in Chapter 2. The extended Kalman filter therefore has the lowest calculational complexity. In the extended Kalman filter, the calculations of the Kalman gain $\mathbf{K}_k$ and estimated error covariance $\mathbf{P}_k$ given in Algorithm 1 in Chapter 2 pose a significant computational burden $O(n_y^3 + n_y n_x^2 + n_x n_y^2)$. In consideration of $n_y < n_x$, the extended Kalman filter has complexity $O(n_y n_x^2)$ in terms of scale.

3.3.3 Steady-State State Estimation

This section presents the steady-state state estimation of nonlinear Kalman filters and gives the upper bound of state estimation error in terms of mean square error.

Defining the state estimation error $\tilde{\mathbf{x}}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$ and prediction error $\tilde{\mathbf{x}}_{k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}$, we have

$$\begin{aligned} \tilde{\mathbf{x}}_{k|k-1} &\approx \mathbf{F}_k \tilde{\mathbf{x}}_{k-1} + \mathbf{w}_k \\ &= \boldsymbol{\beta}_k \mathbf{F}_k \tilde{\mathbf{x}}_{k-1} + \mathbf{w}_k \end{aligned} \tag{3.33}$$

where $\boldsymbol{\beta}_k = \mathrm{diag}(\beta_{k,1}, \beta_{k,2}, \cdots, \beta_{k,n_x})$. To improve the stability, an extra matrix $\Delta\mathbf{Q}_k$ is included in the calculated covariance matrix $\hat{\mathbf{P}}_{k|k-1}$ [43], i.e.,

$$\begin{aligned} \hat{\mathbf{P}}_{k|k-1} &= E(\tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^{\mathrm{T}}) \\ &= \boldsymbol{\beta}_k \mathbf{F}_k \hat{\mathbf{P}}_{k-1} \mathbf{F}_k^{\mathrm{T}} \boldsymbol{\beta}_k + \bar{\mathbf{Q}}_k \end{aligned} \tag{3.34}$$

with the calculated covariance of estimation error at discrete time $k-1$, $\hat{\mathbf{P}}_{k-1}$. The matrix $\bar{\mathbf{Q}}_k$ is calculated by

$$\bar{\mathbf{Q}}_k = \mathbf{P}_{k|k-1}^e + \hat{\mathbf{P}}_{k|k-1} - \mathbf{P}_{k|k-1} + \mathbf{Q}_k + \Delta\mathbf{Q}_k \tag{3.35}$$

with the covariance of the state prediction error

$$\begin{aligned}\mathbf{P}_{k|k-1} &= \boldsymbol{\beta}_k \mathbf{F}_k \hat{\mathbf{P}}_{k-1} \mathbf{F}_k^{\mathrm{T}} \boldsymbol{\beta}_k + \mathbf{P}^e_{k|k-1} + \mathbf{Q}_k \\ \mathbf{P}^e_{k|k-1} &= E(\boldsymbol{\beta}_k \mathbf{F}_k \hat{\mathbf{P}}_{k-1} \mathbf{F}_k^{\mathrm{T}} \boldsymbol{\beta}_k) - \boldsymbol{\beta}_k \mathbf{F}_k \hat{\mathbf{P}}_{k-1} \mathbf{F}_k^{\mathrm{T}} \boldsymbol{\beta}_k. \end{aligned} \tag{3.36}$$

Theorem 2 *Suppose* $\tau_{\min}, \tau_{\max}, p_{\min}, \bar{p}_{\min}, q_{\min}$, $q_{\max}$, $\bar{q}_{\min}$, $\bar{q}_{\max}$, $r_{\min}$, $r_{\max}$ *are real positive numbers and* $\beta_{\min}$, $\beta_{\max}$, $\bar{f}_{\min}$, $\bar{f}_{\max}$, $h_{\min}$, $h_{\max}$ *are real numbers. The mean square estimation error* $\tilde{\mathbf{x}}_k$ *is bounded, provided the following conditions are fulfilled.*

$$\beta^2_{\min} \boldsymbol{\Gamma}_{n_x} \leq \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T \leq \beta^2_{\max} \boldsymbol{\Gamma}_{n_x} \tag{3.37}$$

$$\tau_{\min} \boldsymbol{\Gamma}_{n_x} \leq \mathbf{F}_k \hat{\mathbf{P}}_{k-1} \mathbf{F}_k^T \leq \tau_{\max} \boldsymbol{\Gamma}_{n_x} \tag{3.38}$$

$$\bar{f}^2_{\min} \boldsymbol{\Gamma}_{n_x} \leq \bar{\mathbf{F}}_k \bar{\mathbf{F}}_k^T \leq \bar{f}^2_{\max} \boldsymbol{\Gamma}_{n_x} \tag{3.39}$$

$$p_{\min} \boldsymbol{\Gamma}_{n_x} \leq \hat{\mathbf{P}}_k \leq \mathbf{P}^u \tag{3.40}$$

$$\bar{p}_{\min} \boldsymbol{\Gamma}_{n_x} \leq \hat{\mathbf{P}}_k^{-1} \leq \mathbf{P}^{+u}, 0 < \delta_{\min}(\hat{\mathbf{P}}_k^{-1}) \tag{3.41}$$

$$h^2_{\min} \boldsymbol{\Gamma}_{n_y} \leq \mathbf{H}_k \mathbf{H}_k^T \leq h^2_{\max} \boldsymbol{\Gamma}_{n_y} \tag{3.42}$$

$$\bar{q}_{\min} \boldsymbol{\Gamma}_{n_x} \leq \bar{\mathbf{Q}}_k \leq \bar{q}_{\max} \boldsymbol{\Gamma}_{n_x} \tag{3.43}$$

$$q_{\min} \boldsymbol{\Gamma}_{n_x} \leq \mathbf{Q}_k \leq q_{\max} \boldsymbol{\Gamma}_{n_x} \tag{3.44}$$

$$r_{\min} \boldsymbol{\Gamma}_{n_y} \leq \mathbf{R}_k \leq r_{\max} \boldsymbol{\Gamma}_{n_y} \tag{3.45}$$

where $\bar{\mathbf{F}}_k = \boldsymbol{\beta}_k \mathbf{F}_k$*;* $\hat{\mathbf{P}}_k$ *is bounded by* $\mathbf{P}^u$*;* $\delta_{min}(\cdot)$ *denotes the minimum eigenvalue; and the positive definite matrix* $\hat{\mathbf{P}}_k^{-1}$ *is bounded by* $\mathbf{P}^{+u}$*.*

Proof: Construct the Lyapunov function by

$$V_k(\tilde{\mathbf{x}}_k) = \tilde{\mathbf{x}}_k^T \hat{\mathbf{P}}_k^{-1} \tilde{\mathbf{x}}_k. \tag{3.46}$$

According to (3.41), the positive definite matrix $\hat{\mathbf{P}}_k^{-1}$ is upper-bounded by $\mathbf{P}^{+u}$.

Therefore, we can obtain the boundedness of the Lyapunov function $V_k(\tilde{\mathbf{x}}_k)$ as

$$(p_u^+)^{-1}\|\tilde{\mathbf{x}}_k\|^2 \leq V_k(\tilde{\mathbf{x}}_k) \leq \bar{p}_{\min}^{-1}\|\tilde{\mathbf{x}}_k\|^2 \tag{3.47}$$

where p^{+u} is a positive number.

Substituting $\tilde{\mathbf{x}}_k = \tilde{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\tilde{\mathbf{y}}_k$ in (3.15) into (3.46), we obtain

$$V_k(\tilde{\mathbf{x}}_k) = \left(\tilde{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\tilde{\mathbf{y}}_k\right)^T \hat{\mathbf{P}}_k^{-1}\left(\tilde{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\tilde{\mathbf{y}}_k\right) \tag{3.48}$$

where $\tilde{\mathbf{y}}_k = \mathbf{H}_k\tilde{\mathbf{x}}_{k|k-1} + \mathbf{v}_k$. Expanding (3.48), we obtain

$$V_k(\tilde{\mathbf{x}}_k) = \tilde{\mathbf{x}}_{k|k-1}^{\mathrm{T}}\hat{\mathbf{P}}_k^{-1}\tilde{\mathbf{x}}_{k|k-1} - \tilde{\mathbf{x}}_{k|k-1}^{\mathrm{T}}\hat{\mathbf{P}}_k^{-1}\mathbf{K}_k\tilde{\mathbf{y}}_k - \tilde{\mathbf{y}}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}\hat{\mathbf{P}}_k^{-1}\tilde{\mathbf{x}}_{k|k-1} + \tilde{\mathbf{y}}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}\hat{\mathbf{P}}_k^{-1}\mathbf{K}_k\tilde{\mathbf{y}}_k. \tag{3.49}$$

Take the expectation of $V_k(\tilde{\mathbf{x}}_k)$ conditional on $\tilde{\mathbf{x}}_{k-1}$ as

$$\begin{aligned} E(V_k(\tilde{\mathbf{x}}_k)|\tilde{\mathbf{x}}_{k-1}) = {} & E((\beta_k\mathbf{F}_k\tilde{\mathbf{x}}_{k-1})^{\mathrm{T}}(\beta_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}\mathbf{F}_k^T\beta_k + \bar{\mathbf{Q}}_k)^{-1}(\beta_k\mathbf{F}_k\tilde{\mathbf{x}}_{k-1}) + \mathbf{v}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\hat{\mathbf{P}}_k\mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{v}_k \\ & + \mathbf{w}_k^T\hat{\mathbf{P}}_{k|k-1}^{-1}\mathbf{w}_k - (\mathbf{H}_k\mathbf{w}_k)^{\mathrm{T}}\left(\mathbf{H}_k\hat{\mathbf{P}}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k\right)^{-1}(\mathbf{H}_k\mathbf{w}_k) \\ & - (\mathbf{H}_k\beta_k\mathbf{F}_k\tilde{\mathbf{x}}_{k-1})^{\mathrm{T}}(\mathbf{H}_k\hat{\mathbf{P}}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}\left(\mathbf{H}_k\beta_k\mathbf{F}_k\tilde{\mathbf{x}}_{k-1}\right)|\tilde{\mathbf{x}}_{k-1}). \end{aligned} \tag{3.50}$$

Based on matrix inequality [43]

$$\hat{\mathbf{A}}^{-1} > \hat{\mathbf{B}}(\hat{\mathbf{B}}^T\hat{\mathbf{A}}\hat{\mathbf{B}} + \hat{\mathbf{C}})^{-1}\hat{\mathbf{B}}^T \tag{3.51}$$

with $\hat{\mathbf{A}} \in \mathbb{R}^{m\times n}$, $\hat{\mathbf{B}} \in \mathbb{R}^{m\times n}$, $\hat{\mathbf{C}} \in \mathbb{R}^{m\times n}$ and $\hat{\mathbf{A}}, \hat{\mathbf{C}} > 0$, we obtain

$$\begin{aligned} \hat{\mathbf{P}}_{k|k-1}^{-1} &= (\beta_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}\mathbf{F}_k^T\beta_k + \bar{\mathbf{Q}}_k)^{-1} \\ &\leq (\beta_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}\mathbf{F}_k^T\beta_k)^{-1} \end{aligned} \tag{3.52}$$

by letting $\hat{\mathbf{A}} = \beta_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}\mathbf{F}_k^T\beta_k$, $\hat{\mathbf{B}} = \mathbf{I}$ and $\hat{\mathbf{C}} = \bar{\mathbf{Q}}_k$. Therefore, (3.50) can be rearranged

as

$$E[V_k(\tilde{\mathbf{x}}_k)|\tilde{\mathbf{x}}_{k-1}] - (1-\varrho_k)V_{k-1}(\tilde{\mathbf{x}}_{k-1}) \leq \theta \tag{3.53}$$

where $V_{k-1}(\tilde{\mathbf{x}}_{k-1}) = E(\boldsymbol{\beta}_k\mathbf{F}_k\tilde{\mathbf{x}}_{k-1})^{\mathrm{T}}(\boldsymbol{\beta}_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}\mathbf{F}_k^T\boldsymbol{\beta}_k)^{-1}(\boldsymbol{\beta}_k\mathbf{F}_k\tilde{\mathbf{x}}_{k-1}) = \tilde{\mathbf{x}}_{k-1}^{\mathrm{T}}\hat{\mathbf{P}}_{k-1}^{-1}\tilde{\mathbf{x}}_{k-1}^{\mathrm{T}}$, which means that $(\boldsymbol{\beta}_k\mathbf{F}_k)^{-1}$ exists. The scalar ϱ_k in (3.53) is defined by

$$\begin{aligned}\varrho_k = &(\mathbf{H}_k\boldsymbol{\beta}_k\mathbf{F}_k\tilde{\mathbf{x}}_{k-1})^{\mathrm{T}}\,(\mathbf{H}_k\boldsymbol{\beta}_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}(\mathbf{H}_k\boldsymbol{\beta}_k\mathbf{F}_k)^{\mathrm{T}} + \mathbf{H}_k\\ &\bar{\mathbf{Q}}_k\mathbf{H}_k^T + \mathbf{R}_k)^{-1}\,(\mathbf{H}_k\boldsymbol{\beta}_k\mathbf{F}_k\tilde{\mathbf{x}}_{k-1})\,/\tilde{\mathbf{x}}_{k-1}^T\hat{\mathbf{P}}_{k-1}^{-1}\tilde{\mathbf{x}}_{k-1}.\end{aligned} \tag{3.54}$$

Letting $\hat{\mathbf{A}} = \hat{\mathbf{P}}_{k-1}$, $\hat{\mathbf{B}} = (\mathbf{H}_k\boldsymbol{\beta}_k\mathbf{F}_k)^T$ and $\hat{\mathbf{C}} = \mathbf{H}_k\bar{\mathbf{Q}}_k\mathbf{H}_k^T + \mathbf{R}_k$, we have $\varrho_k < 1$. According to (3.37) to (3.45), ϱ_k is lower bounded by positive number

$$\varrho_{min} = p_{min}(h_{min}\bar{f}_{min})^2[(h_{max}\beta_{max})^2\tau_{max} + \bar{q}_{max}h_{max}^2 + r_{max}]^{-1}. \tag{3.55}$$

The scalar θ in (3.53) is given by

$$\begin{aligned}\theta &= E\{\mathbf{w}_k^T[\boldsymbol{\beta}_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}\mathbf{F}_k^T\boldsymbol{\beta}_k + \bar{\mathbf{Q}}_k]^{-1} - \mathbf{H}_k^T(\mathbf{H}_k(\boldsymbol{\beta}_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}\mathbf{F}_k^T\boldsymbol{\beta}_k + \bar{\mathbf{Q}}_k)\mathbf{H}_k^T + \mathbf{R}_k)^{-1}\mathbf{H}_k]\mathbf{w}_k\\ &\quad + \mathbf{v}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\hat{\mathbf{P}}_k\mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{v}_k|\tilde{\mathbf{x}}_{k-1}\}\\ &= E(tr(((\boldsymbol{\beta}_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}\mathbf{F}_k^T\boldsymbol{\beta}_k + \bar{\mathbf{Q}}_k)^{-1} - \mathbf{H}_k^T(\mathbf{H}_k(\boldsymbol{\beta}_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}\mathbf{F}_k^T\boldsymbol{\beta}_k + \bar{\mathbf{Q}}_k)\mathbf{H}_k^T + \mathbf{R}_k)^{-1}\mathbf{H}_k)\mathbf{w}_k\mathbf{w}_k^T)\\ &+ tr(\mathbf{R}_k^{-1}\mathbf{H}_k\hat{\mathbf{P}}_k\mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{v}_k\mathbf{v}_k^T)|\tilde{\mathbf{x}}_{k-1})\\ &= tr(\boldsymbol{\theta}_k\mathbf{Q}_k + \mathbf{R}_k^{-1}\mathbf{H}_k\hat{\mathbf{P}}_k\mathbf{H}_k^T)\end{aligned} \tag{3.56}$$

where the parameter $\boldsymbol{\theta}_k$ is defined as

$$\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}\mathbf{F}_k^T\boldsymbol{\beta}_k + \bar{\mathbf{Q}}_k)^{-1} - \mathbf{H}_k^T(\mathbf{H}_k(\boldsymbol{\beta}_k\mathbf{F}_k\hat{\mathbf{P}}_{k-1}\mathbf{F}_k^T\boldsymbol{\beta}_k + \bar{\mathbf{Q}}_k)\mathbf{H}_k^T + \mathbf{R}_k)^{-1}\mathbf{H}_k. \tag{3.57}$$

Similarly, applying the same matrix inequality by letting

$$\hat{\mathbf{B}} = \mathbf{H}_k^T, \hat{\mathbf{A}} = \beta_k \mathbf{F}_k \hat{\mathbf{P}}_{k-1} \mathbf{F}_k^T \beta_k + \bar{\mathbf{Q}}_k, \hat{\mathbf{C}} = \mathbf{R}_k, \tag{3.58}$$

we obtain $\boldsymbol{\theta}_k > 0$. According to (3.40), $\hat{\mathbf{P}}_k$ is upper bounded by $\mathbf{P}^u$, which gives the boundedness of $tr(\mathbf{R}_k^{-1} \mathbf{H}_k \hat{\mathbf{P}}_k \mathbf{H}_k^T)$ in (3.56) defined by p_h. Therefore, the upper bound of θ in (3.56) is

$$\theta \leq q_{max} \bar{q}_{min}^{-1} n_x + p_h = \tilde{\theta}. \tag{3.59}$$

According to [43], we obtain $E\{\|\tilde{\mathbf{x}}_\mathbf{k}\|^2\} \leq \tilde{\tau}$ by (3.37) through (3.45), where $\tilde{\tau}$ is defined by

$$\tilde{\tau} = \sum_{i=1}^{k-1} p_u^+ \tilde{\theta} (1 - \varrho_{min})^i + \frac{p_u^+ E\{\|\tilde{\mathbf{x}}_0\|^2\}}{p_{min}} (1 - \varrho_{min})^k. \tag{3.60}$$

■

It is necessary to note that **Theorem** 2 is a sufficient condition to ensure the stability of UKF in theory since parameters like β in (3.33) are not available.

Remark 5 *Inspecting $\mathbf{F}_{k-1}$ of the SIS compartmental model, the increasing infection rate α may not be beneficial for satisfying (3.38) and maintaining stability. A larger recovery rate may be beneficial for enhancing stability. Same with the SIS model, SIR and SIRS models are unstable for nonlinear Kalman filters performing nonlinear state estimation under a large infection rate.*

3.4 Simulation Results

This section compares the filtering performance of nonlinear Kalman filters for tracking epidemic spreading on ER, NW and WS networks with $N = 200$ and $K = 10$. When a small connecting probability p is chosen, the NW and WS networks reduce to

the nearest-neighbor coupling network, and very few nodes are connected in the ER network. In addition, a larger connecting probability makes the NW network a global neighbor coupling network. In our experiments, the connecting probability is chosen in the interval $p \in [0.05, 0.95]$.

The compartmental models used include SIS, SIR and SIRS models with $\varpi_{\text{SIS}} = (\alpha, \epsilon)$, $\varpi_{\text{SIR}} = (\alpha, \beta)$, $\varpi_{\text{SIRS}} = (\alpha, \beta, \gamma)$. The observations in (3.11) are generated synthetically for facilitating the study of epidemic tracking by nonlinear Kalman filters on different networks. The nonlinear Kalman filters including EKF, UKF, third-degree CKF denoted by CKF3 and fifth-degree CKF denoted by CKF5, are used to estimate the dynamics of epidemic spreading. The parameters of nonlinear Kalman filters are designed for balancing the filtering accuracy and stability.

In each experiment, 50 Monte Carlo simulations are run. The root mean square error (RMSE) is used to evaluate the filtering performance, i.e.,

$$\text{RMSE} = \frac{1}{\bar{N}} \sum_{k=1}^{\bar{N}} \sqrt{\sum_{j=\ell_1+1}^{\ell_1+|\mathbf{D}|} (\mathbf{x}_{j,k} \frac{\hat{\mathbf{N}}_{j-\ell_1}}{|N|} - \hat{\mathbf{x}}_{j,k} \frac{\hat{\mathbf{N}}_{j-\ell_1}}{|N|})^2} \tag{3.61}$$

where $\hat{\mathbf{N}} = \left[|N|_{\bar{d}_1}, |N|_{\bar{d}_2}, \cdots, |N|_{d_{p_\varsigma}}, \cdots, |N|_{\bar{d}_{|\mathbf{D}|}}\right]$ with $|N|_{d_{p_\varsigma}}$ denoting the number of nodes of degree d_{p_ς}, and $\bar{N}$ denotes the number of samples. Column vectors $\mathbf{x}_k = [\mathbf{x}_{1,k}, \mathbf{x}_{2,k}, \cdots, \mathbf{x}_{2|\mathbf{D}|,k}]^{\text{T}}$ and $\hat{\mathbf{x}}_k = [\hat{\mathbf{x}}_{1,k}, \hat{\mathbf{x}}_{2,k}, \cdots, \hat{\mathbf{x}}_{2|\mathbf{D}|,k}]^{\text{T}}$ denote the desired and estimated state, respectively. Scalar $\ell_1 = 0$ is for calculating the RMSE of the proportion of the susceptible individuals, denoted by RMSE_{S}, and $\ell_1 = |\mathbf{D}|$ is for calculating the RMSE of the proportion of the infected individuals, denoted by RMSE_{I}. The RMSE of the ratio of the infected and susceptible individuals is denoted by RMSE_{IS}.

3.4.1 Epidemic Spreading for SIR-Type Model

The SIR compartmental model is an open-loop model where the proportions of the susceptible and infected individuals converge to zero at steady state. Therefore, the

SIR model may be unstable in performing state estimation by nonlinear Kalman filters in comparison with SIS and SIRS models.

Figs. 3.2(a) to 3.4(c) show the RMSE of nonlinear Kalman filters based on the SIR model applied to tracking epidemic spreading in ER, NW and WS networks, respectively. The parameter sets for the SIR model in these networks are $\varpi_{\mathrm{ER,SIR}} = (0.2, 0.1)$, $\varpi_{\mathrm{NW,SIR}} = (0.07, 0.1)$, $\varpi_{\mathrm{WS,SIR}} = (0.25, 0.35)$, respectively. The state and measurement noises obey Gaussian distribution with zero mean and covariance set by $\mathbf{Q}_k = \{10^{-8}\mathbf{\Gamma}_{2|\mathbf{D}|}, 10^{-10}\mathbf{\Gamma}_{2|\mathbf{D}|}, 10^{-10}\mathbf{\Gamma}_{2|\mathbf{D}}\}$ and $\mathbf{R}_k = \{10^{-2}\mathbf{\Gamma}_{|\mathbf{D}|}, 10^{-5}\mathbf{\Gamma}_{|\mathbf{D}|}, 10^{-2}\mathbf{\Gamma}_{|\mathbf{D}}\}$ in these three networks. The values of $\Delta\mathbf{Q}$ for the UKF, CKF3, and CKF5 are set as $\{0, 0, 0.4\}$ in ER and NW networks and $\{10^{-10}, 10^{-10}, 10^{-2}\}$ in the WS network.

The mean of $\mathbf{P}$ in (3.9) for ER and NW networks increases with probability p. In contrast, $\mathbf{P}$ is large under a smaller p and becomes smaller with increasing p in the WS network. In ER and NW networks, when a small probability p is utilized for generating networks, a larger infection parameter may be required for achieving a better tracking function, degrading the stability of filtering since (3.38) is not satisfied. The UKF with poor stability therefore has a lower filtering accuracy when adopting a relatively smaller p. When adopting a larger probability, the UKF performs better than the CKF3 since a higher order statistical characteristic can be captured. Due to the larger $\Delta\mathbf{Q}$, the CKF5 cannot achieve the desired filtering performance in comparison with the UKF and CKF3. The EKF performs the worst filtering due to the use of linear approximation. In the WS network, the UKF has a lower filtering precision than the CKF3 for a small p and outperforms CKF3 when choosing a large p. In addition, the CKF5 performs worse than the UKF and CKF3 due to a larger $\Delta\mathbf{Q}$.

3.4.2 Epidemic Spreading for SIS-Type Model

Figs. 3.5(a) to 3.7(c) show the RMSE of nonlinear Kalman filters based on the SIS model in ER, NW and WS networks, respectively. The parameters for the SIS

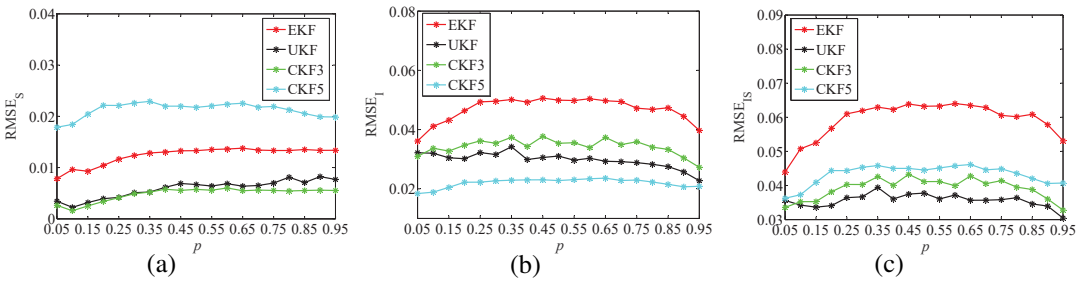


Figure 3.2: (a) $\mathrm{RMSE_S}$ of nonlinear Kalman filters applied to SIR model for ER network; (b) $\mathrm{RMSE_I}$ of nonlinear Kalman filters applied to SIR model for ER network; (c) $\mathrm{RMSE_{IS}}$ of nonlinear Kalman filters applied to SIR model for ER network.

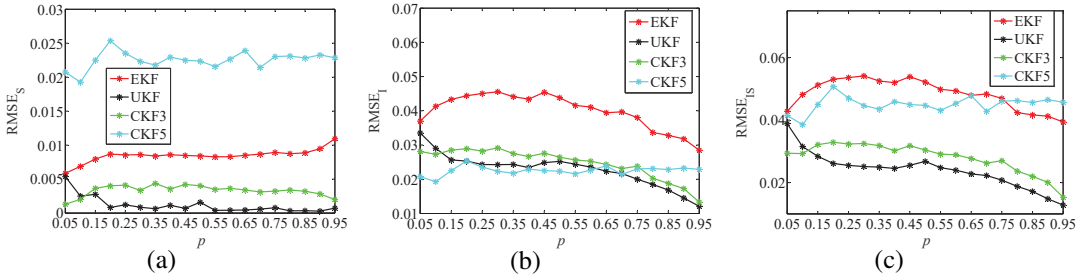


Figure 3.3: (a) $\mathrm{RMSE_S}$ of nonlinear Kalman filters applied to SIR model for NW network; (b) $\mathrm{RMSE_I}$ of nonlinear Kalman filters applied to SIR model for NW network; (c) $\mathrm{RMSE_{IS}}$ of nonlinear Kalman filters applied to SIR model for NW network.

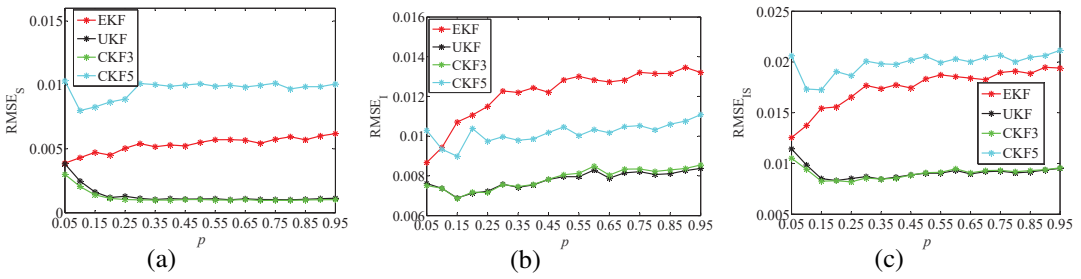


Figure 3.4: (a) $\mathrm{RMSE_S}$ of nonlinear Kalman filters applied to SIR model for WS network; (b) $\mathrm{RMSE_I}$ of nonlinear Kalman filters applied to SIR model for WS network; (c) $\mathrm{RMSE_{IS}}$ of nonlinear Kalman filters applied to SIR model for WS network.

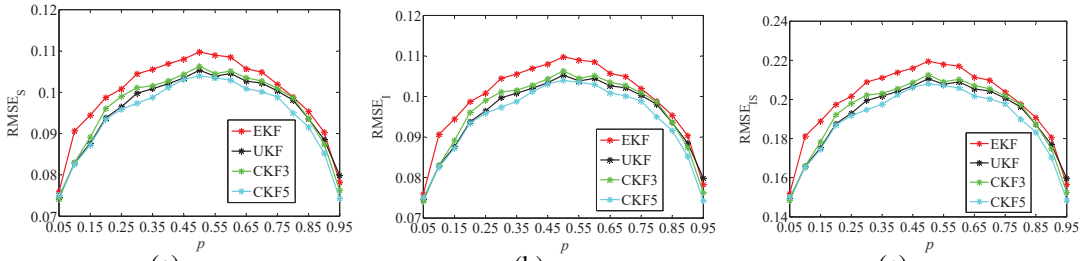


Figure 3.5: (a) $\mathrm{RMSE_S}$ of nonlinear Kalman filters applied to SIS model for ER network; (b) $\mathrm{RMSE_I}$ of nonlinear Kalman filters applied to SIS model for ER network; (c) $\mathrm{RMSE_{IS}}$ of nonlinear Kalman filters applied to SIS model for ER network.

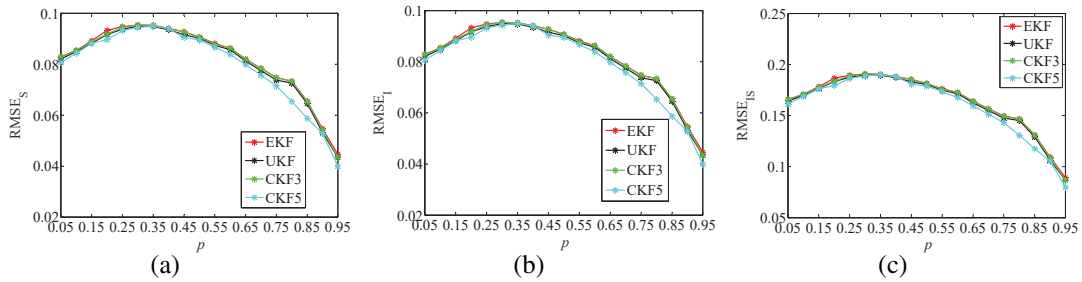


Figure 3.6: (a) $\mathrm{RMSE_S}$ of nonlinear Kalman filters applied to SIS model for NW network; (b) $\mathrm{RMSE_I}$ of nonlinear Kalman filters applied to SIS model for NW network; (c) $\mathrm{RMSE_{IS}}$ of nonlinear Kalman filters applied to SIS model for NW network.

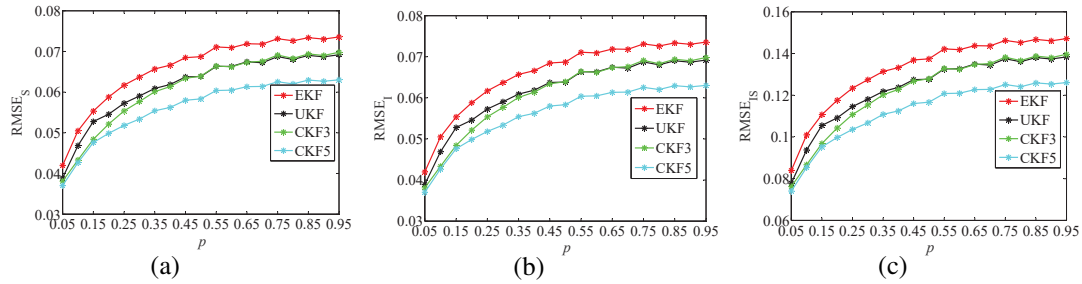


Figure 3.7: (a) $\mathrm{RMSE_S}$ of nonlinear Kalman filters applied to SIS model for WS network; (b) $\mathrm{RMSE_I}$ of nonlinear Kalman filters applied to SIS model for WS network; (c) $\mathrm{RMSE_{IS}}$ of nonlinear Kalman filters applied to SIS model for WS network.

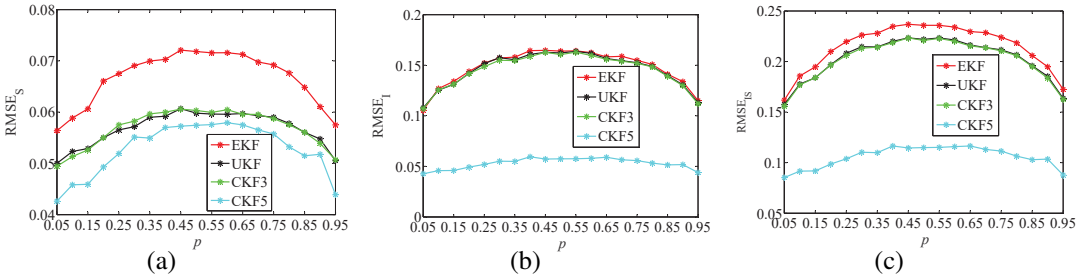


Figure 3.8: (a) $\mathrm{RMSE_S}$ of nonlinear Kalman filters applied to SIRS model for ER network; (b) $\mathrm{RMSE_I}$ of nonlinear Kalman filters applied to SIRS model for ER network; (c) $\mathrm{RMSE_{IS}}$ of nonlinear Kalman filters applied to SIRS model for ER network.

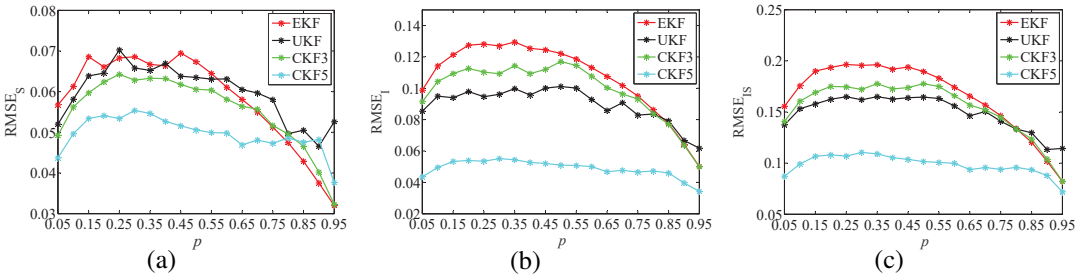


Figure 3.9: (a) $\mathrm{RMSE_S}$ of nonlinear Kalman filters applied to SIRS model for NW network; (b) $\mathrm{RMSE_I}$ of nonlinear Kalman filters applied to SIRS model for NW network; (c) $\mathrm{RMSE_{IS}}$ of nonlinear Kalman filters applied to SIRS model for NW network.

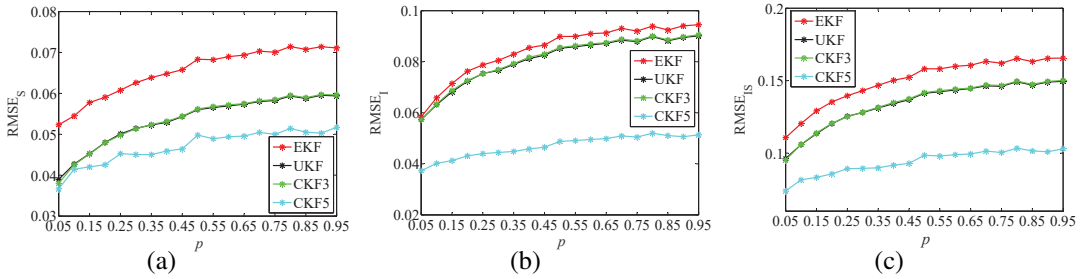


Figure 3.10: (a) $\mathrm{RMSE_S}$ of nonlinear Kalman filters applied to SIRS model for WS network; (b) $\mathrm{RMSE_I}$ of nonlinear Kalman filters applied to SIRS model for WS network; (c) $\mathrm{RMSE_{IS}}$ of nonlinear Kalman filters applied to SIRS model for WS network.

model are set by $\varpi_{\mathrm{ER,SIS}} = (0.1, 0.12)$, $\varpi_{\mathrm{NW,SIS}} = (0.07, 0.1)$, $\varpi_{\mathrm{WS,SIS}} = (0.14, 0.1)$, respectively. The state and measurement noises obey Gaussian distribution with covariance $\mathbf{Q}_k = \{10^{-2}\mathbf{\Gamma}_{2|\mathbf{D}|}, 10^{-4}\mathbf{\Gamma}_{2|\mathbf{D}|}, 10^{-3}\mathbf{\Gamma}_{2|\mathbf{D}}\}$ and $\mathbf{R}_k = \{10^{-3}\mathbf{\Gamma}_{|\mathbf{D}|}, 10^{-5}\mathbf{\Gamma}_{|\mathbf{D}|}, 10^{-4}\mathbf{\Gamma}_{|\mathbf{D}}\}$. The values of $\Delta\mathbf{Q}$ for the UKF, CKF3, and CKF5 are set as $\{10^{-2}, 10^{-2}, 0.25\}$, $\{10^{-10}, 10^{-10}, 10^{-2}\}$, $\{10^{-3}, 10^{-3}, 0.15\}$ in ER, NW and WS networks, respectively.

From Figs. 3.5(a) to 3.7(c), $\mathrm{RMSE_S}$, $\mathrm{RMSE_I}$ and $\mathrm{RMSE_{IS}}$ increase with the growing dimension shown in Fig. 3.1. Unlike the SIR compartmental model, the SIS model is a closed-loop model, which generates the non-zero ratio of the infected and susceptible individuals. In ER and NW networks, the CKF5 achieves the best filtering performance. In addition, with increasing the probability p in the ER network, the UKF obtains a higher accuracy for capturing the high order statistical characteristic than the CKF3. Due to the statistical characteristic of the $\mathbf{P}$, the UKF cannot obtain the desired accuracy under a larger p. In the NW network, the UKF and CKF3 achieve almost same performance. In the WS network, the precision of the UKF is lower than that of the CKF3 for a small probability p, but higher with increasing the probability p for the statistical characteristic of the $\mathbf{P}$. The CKF5 has the highest accuracy while the precision of the EKF is the lowest due to the lack of the second-order statistical characteristic.

3.4.3 Epidemic Spreading for SIRS-Type Model

Figs. 3.8(a) to 3.10(c) show the RMSE of nonlinear Kalman filters on the basis of the SIRS model applied to tracking epidemic transmission in ER, NW and WS networks, respectively. The parameter sets for the SIRS model in these networks are $\varpi_{\mathrm{ER,SIRS}} = (0.15, 0.1, 0.1)$, $\varpi_{\mathrm{NW,SIRS}} = (0.1, 0.1, 0.1)$, $\varpi_{\mathrm{WS,SIRS}} = (0.25, 0.35, 0.1)$, respectively. The state and measurement noises in these networks obey Gaussian distribution with zero mean and covariance set by $\mathbf{Q}_k = \{10^{-10}\mathbf{\Gamma}_{2|\mathbf{D}|}, 10^{-3}\mathbf{\Gamma}_{|\mathbf{D}|}, 10^{-20}\mathbf{\Gamma}_{2|\mathbf{D}}\}$ and $\mathbf{R}_k = \{10^{-4}\mathbf{\Gamma}_{|\mathbf{D}|}, 10^{-1}\mathbf{\Gamma}_{|\mathbf{D}|}, 10^{-1.5}\mathbf{\Gamma}_{|\mathbf{D}}\}$ in these three networks. The values of $\Delta\mathbf{Q}$ for

Table 3.2: Specific Choice of Kalman Filters

Networks	ER			NW			WS	
Interval	Υ_1	Υ_2	Υ_3	Υ_1	Υ_2	Υ_3	Υ_1	Υ_2
SIS	CKF5[1] CKF3[2] UKF[3] EKF[4]	CKF5[1] UKF[2] CKF3[3] EKF[4]	CKF5[1] CKF3[2] EKF[3] UKF[4]	CKF5[1] UKF[2] CKF3[3] EKF[4]	CKF5[1] UKF[2] CKF3[3] EKF[4]	CKF5[1] UKF[2] CKF3[3] EKF[4]	CKF5[1] CKF3[2] UKF[3] EKF[4]	CKF5[1] UKF[2] CKF3[3] EKF[4]
SIR	CKF3[1] UKF[2] CKF5[3] EKF[4]	UKF[1] CKF3[2] CKF5[3] EKF[4]	UKF[1] CKF3[2] CKF5[3] UKF[4]	CKF3[1] UKF[2] CKF5[3] EKF[4]	UKF[1] CKF3[2] CKF5[3] EKF[4]	UKF[1] CKF3[2] CKF5[3] EKF[4]	CKF3[1] UKF[2] EKF[3] CKF5[4]	UKF[1] CKF3[2] EKF[3] CKF5[4]
SIRS	CKF5[1] CKF3[2] UKF[3] EKF[4]	CKF5[1] CKF3[2] UKF[3] EKF[4]	CKF5[1] CKF3[2] UKF[3] UKF[4]	CKF5[1] UKF[2] CKF3[3] EKF[4]	CKF5[1] UKF[2] CKF3[3] EKF[4]	CKF5[1] CKF3[2] EKF[3] UKF[4]	CKF5[1] CKF3[2] UKF[3] EKF[4]	CKF5[1] UKF[2] CKF3[3] EKF[4]

Table 3.3: General Choice of Kalman Filters

Networks	ER			NW			WS	
Interval	Υ_1	Υ_2	Υ_3	Υ_1	Υ_2	Υ_3	Υ_1	Υ_2
SIS	CKF5[1] CKF3[2]	CKF5[1] UKF[2]	CKF5[1] CKF3[2]	CKF5[1] CKF3[2]	CKF5[1] UKF[2]	CKF5[1] CKF3[2]	CKF5[1] CKF3[2]	CKF5[1] UKF[2]
SIR	CKF3[1]	UKF[1]	CKF3[1]	CKF3[1]	UKF[1]	CKF3[1]	CKF3[1]	UKF[1]
SIRS	CKF5[1] CKF3[2]	CKF5[1] UKF[2]	CKF5[1] CKF3[2]	CKF5[1] CKF3[2]	CKF5[1] UKF[2]	CKF5[1] CKF3[2]	CKF5[1] CKF3[2]	CKF5[1] UKF[2]

the UKF, CKF3 and CKF5 are set as $\{10^{-10}, 10^{-10}, 0.45\}$, $\{10^{-3}, 10^{-3}, 0.55\}$ in ER and NW networks and $\{10^{-10}, 10^{-10}, 0.35\}$ in the WS network. Same as the SIS model, the SIRS model is a closed-loop model, which has the same conclusion as in the SIS model. However, when a larger probability p is chosen for generating the NW network, the UKF with poor stability has a higher estimation error than the other nonlinear Kalman filters due to the higher mean of $\mathbf{P}$ as p increases in the NW network.

3.5 Discussions

3.5.1 Choice of Nonlinear Kalman Filters

Based on the accuracy analysis, stability analysis and numerical experiments, filtering accuracy depends on the choice of the connecting probability p. For ER and NW networks, the connecting probability p is split into three intervals, i.e., Υ_1: $0 < p \leq p_1$; Υ_2: $p_1 < p \leq p_2$ and Υ_3: $p_2 < p < 1$. The connecting probability p for generating the

WS network is split into two intervals, i.e., Υ_1: $0 < p \leq p_3$ and Υ_2: $p_3 < p < 1$. The choice of parameters p_1, p_2 and p_3 depends not only on the types of networks but also the rates.

Table 3.2 compares nonlinear Kalman filters for tracking epidemic spreading on networks in terms of RMSE_{IS} as given in Section 3.4. Notation CKF5[1] denotes that the CKF5 has the highest filtering accuracy. The UKF[2] and CKF3[2] in an interval mean that the UKF and CKF3 can achieve a similar filtering precision. Since parameters p_1, p_2 and p_3 are not fixed, we provide a more general comparison in Table 3.3 for indicating high filtering accuracy at specific intervals. Combining Table 3.2 and Table 3.3, we may select nonlinear Kalman filters for tracking epidemic spreading as follows.

For epidemic spreading on ER or NW networks using SIRS and SIS models in intervals Υ_1 and Υ_3, the CKF5 is the best choice in terms of filtering accuracy. For reducing computational burden, the CKF3 is a favorable choice. In interval Υ_1, the mean of $\mathbf{P}$ is relatively small. Therefore, a larger infection parameter for performing nonlinear Kalman filters may be required for tracking epidemic at the expense of degrading the stability of the UKF. In addition, in interval Υ_3, the mean of distribution of $\mathbf{P}$ is generally large, which is not desirable for maintaining stability of the UKF. The UKF therefore cannot be utilized in these two intervals for guaranteeing the filtering accuracy. In the interval of Υ_2, the CKF5 achieves the highest accuracy. For reducing the complexity, the UKF is a better choice for capturing the high order statistical characteristic in comparison with the CKF3.

For epidemic spreading on ER or NW networks modeled by the SIR compartmental model, a larger matrix $\Delta\mathbf{Q}$ is needed for guaranteeing the stability of the CKF5. The CKF5 therefore cannot achieve a favorable filtering accuracy. For guaranteeing filtering accuracy and stability, the same conclusion as for ER or NW networks modeled by SIS and SIRS models can be drawn, i.e., the CKF3 is used for tracking epidemic in intervals Υ_1 and Υ_3 whereas the UKF is chosen for interval Υ_2.

For epidemic spreading on the WS network modeled by SIS and SIRS models, the

filtering accuracy of the CKF5 is the highest in intervals Υ_1 and Υ_2. Since the values of **P** are likely to decrease as the connecting probability increases in the WS network, the UKF is a better choice for tracking epidemic in interval Υ_2 in terms of complexity. The CKF3, however, achieves a higher filtering accuracy in interval Υ_1.

For epidemic spreading on the WS network modeled by the SIR compartmental model, the CKF5 and EKF perform a worse filtering function in comparison with the UKF and CKF3. In terms of reducing computational burden, the CKF3 can achieve a higher accuracy in interval Υ_1, and the UKF has a better filtering performance in interval Υ_2.

Based on the provided guideline, the proposed method can be applied to real data by several steps, namely, studying the type of the complex network, calculating the connection probability, choosing a compartmental model. More concretely, at first, it is necessary to study the type of the complex network such as ER network, NW network, or WS network. The connection probability of the studied network, then, needs to be calculated directly or to be found by using network measures like clustering coefficient [59]. After choosing a compartmental model, we can adopt different nonlinear Kalman filters from aspects of accuracy, complexity, or numerical stability for epidemic tracking over studied network with calculated connection probability.

3.5.2 Scaling of Networks

Nonlinear Kalman filters are used to estimate hidden states on the basis of a state-space model. The filtering accuracy of nonlinear Kalman filters would be degraded by the growing number of states [45]. Since the size of the types of degree **D** generally increases with the scale of the network, nonlinear Kalman filters may not perform epidemic tracking satisfactorily in large complex networks. Since there are numerous complex networks of considerable scale, we discuss here the community detection-based and sub-filter-based methods for applying nonlinear Kalman filters for tracking

epidemic spreading over large-scale complex networks.

1) The community detection-based nonlinear Kalman filters tackle the tracking problem by focusing on detected sub-networks of smaller scale called "communities". In general, a complex network can be represented by a graph consisting of dense communities [96, 97]. The nodes within a community are connected densely in comparison with those connecting across communities [97]. Recently, numerous methods have been proposed for detecting communities of information networks, biological networks and social networks in order to facilitate analysis of network structure [98]. Community detection is effectively a way to split a large-scale network into smaller interconnected communities. Consider a large-scale complex network C having N nodes. Based on the community detection-based method, complex network C can be segmented into Θ communities $\{C_\Theta\}_{\zeta=1}^{\Theta}$, where community C_Θ has size N_Θ, $N_\Theta \ll N$. A nonlinear Kalman filter can then be utilized for epidemic tracking over different communities. Moreover, community detection can be executed repeatedly if the network size of the community C_Θ, $N_\Theta \ll N$ is not small enough to permit manageable computation at the expense of degrading filtering precision.

2) The sub-filter-based nonlinear Kalman filters perform epidemic tracking by breaking a high dimensional state into several low dimensional states. The high dimensional state $\mathbf{x} \in \mathbb{R}^{2|\mathbf{D}|}$ is directly into S sub-states $\{\mathbf{x}_s\}_{s=1}^{S}$ with $\mathbf{x}_s \in \mathbb{R}^{2|\mathbf{D}_s|}$ and $|\mathbf{D}_s| < |\mathbf{D}|$. In due consideration of the correlations of the sub-states, we split the high-dimensional state into several low-dimensional states in terms of degree differences. Given a collection of the types of degree $\mathbf{D} = [d_{p_1}, d_{p_2}, \cdots, d_{p_\varsigma}, \cdots, d_{p_{|\mathbf{D}|}}]$, we divide collection $\mathbf{D}$ into S sub-collections $\{\mathbf{D}_s\}_{s=1}^{S}$, where the differences between the degrees in the s-th sub-collection are smaller than a pre-designed threshold parameter. Based on the state segmentation, S nonlinear Kalman filters are used simultaneously in estimating S sub-states $\{\mathbf{x}_s\}_{s=1}^{S}$ with $\mathbf{x}_s \in \mathbb{R}^{2|\mathbf{D}_s|}$ and $|\mathbf{D}_s| < |\mathbf{D}|$. Therefore, the sub-filter-based approach is applicable of studying dynamics of transmission over large-scale networks with lowered computational burden and improved numerical stability due to

the reduced dimension of the state variables.

3.5.3 Extended Applications

Table 3.2 and Table 3.3 provide a general guideline for using nonlinear Kalman filters for performing epidemic tracking on the basis of constructed state-space models given in (3.9), (3.10) and (3.11). In fact, these state-space models do not only describe the dynamics of epidemic spreading but also capture the information dissemination over real-world networks.

3.5.3.1 Rumor Propagation in Computer Networks

There has been renewed interest in rumor spreading over networks [20]. Rumors are fabricated messages or information lacking of reliable evidence, originating from subjective willingness instead of objectivity. On the Internet, rumors spread fast [21]. The study of rumor spreading over computer networks aims to identify the dynamic behavior of rumor spreading. Similar to the diffusion of epidemic, rumor spreading over computer networks can be captured by models of infectious diseases (3.2) and (3.3). Therefore, the guideline given in Table 3.2 and Table 3.3 is also applicable for choosing nonlinear Kalman filters for studying the dynamics of rumor spreading over computer networks.

3.5.3.2 Malware Spreading in Wireless Sensor Networks

A Wireless Sensor Network (WSN) [22] consists of sensors which transmit numerous observation data to control or processing centers. Wireless sensor nodes contain devices characterized by resource-restrained and low defense capabilities. The sensor nodes can therefore be attacked by emerging malicious software which aims to infect a small number of nodes initially. The neighboring nodes can then be infected via regular communications. In addition, the infected nodes can be recovered by some antivirus

programs installed at each sensor node. An epidemic model given in (3.2) and (3.3) can also be utilized for describing the transmission of communicable attack among wireless sensors. The modeling of attack transmission is a useful tool for exploring the dynamics of sensor attack in wireless sensor networks. Table 3.2 and Table 3.3 are applicable of choosing nonlinear Kalman filters for revealing the state of sensor nodes in wireless sensor networks.

3.6 Conclusion

This chapter studies the tracking of epidemic spreading on homogeneous networks by nonlinear Kalman filters. The epidemic transmission on networks can be described by compartmental models. The state estimation of epidemic transmission can be performed by various kinds of Kalman filters including the unscented Kalman filter, the third-degree cubature Kalman filter, the fifth-degree cubature Kalman filter, and the extended Kalman filter. The performance comparison in terms of accuracy and stability is provided for evaluating the filtering performance of various nonlinear Kalman filters. The connecting probability used for generating complex networks is split into several intervals. Specific and general choices of Kalman filters in each interval for tracking the dynamics of epidemic spreading are provided in terms of filtering precision and computational complexity. However, the scope of this study is limited to homogeneous networks such as random and small-world networks. Further work may assess performance comparison of nonlinear Kalman filters based on degree-based compartmental models for epidemic tracking over different types of networks.

Chapter 4

Generalized Correntropy Sparse Gauss-Hermite Quadrature Filter for Epidemic Tracking on Complex Networks

The dynamic estimation of epidemic spreading on networks is essential for controlling morbidity. Nonlinear Kalman filters, which are capable of estimating the hidden state of a nonlinear system, can be utilized for dynamic state estimation of epidemic spreading. Traditional nonlinear Kalman filters perform optimization using the minimum mean square error (MMSE) criterion. Since observable measurements are generally corrupted by non-Gaussian noise, maximum correntropy criterion-based nonlinear Kalman filters have been used for improving robustness against non-Gaussian noise. In comparison with the maximum correntropy criterion, generalized correntropy is more robust and flexible in the presence of non-Gaussian noise. In this work, epidemic spreading is described by a compartmental model, i.e., susceptible-infected-recovered-susceptible model. Based on the compartmental model, the generalized correntropy-based sparse Gauss-Hermite quadrature filter (GCSGHQF)

is proposed for epidemic tracking on homogeneous networks. The GCSGHQF approximates the Gaussian-weighted integrals by utilizing the sparse Gauss-Hermite quadrature rule. In addition, the generalized correntropy is applied to enhancing robustness in the presence of non-Gaussian noise. Simulation results show that the GCSGHQF exhibits a higher filtering accuracy and improves robustness compared to traditional nonlinear Kalman filters.

4.1 Introduction

As mentioned in Chapter 3, nonlinear Kalman filters [30] can be used to perform state estimation of epidemic transmission. A nonlinear Kalman filter aims to estimate the hidden state by calculating the posterior probability of the state. Commonly used nonlinear Kalman filters include the extended Kalman filter (EKF) [99, 100], unscented Kalman filter (UKF) [101, 102, 103], cubature Kalman filter (CKF) [46] and Gauss-Hermite quadrature filter (GHQF) [104]. The EKF calculates the posterior probability of the state by using a linear approximation, whereas the UKF, CKF and GHQF estimate the hidden state assuming a Gaussian distribution of the state. In comparison with the UKF and CKF, the GHQF achieves the highest filtering accuracy at the expense of increasing the computational burden, especially when the system dimension increases [104]. Recently, a sparse GHQF (SGHQF) [35] has been proposed for alleviating the computational load.

These nonlinear Kalman filters estimate the states by performing optimization in accordance with the minimum mean square error (MMSE) criterion, which is, however, sensitive to outliers. Recently, information theoretical learning (ITL) [51] has been utilized for improving the robustness in the presence of outliers for capturing a higher statistical characteristic of errors. In particular, the maximum correntropy criterion (MCC) is a widely used ITL approach due to its simplicity [78, 105]. Recently, MCC-based Kalman filters have been proposed for alleviating the sensitivity

to outliers. For example, the maximum correntropy Kalman filter (MCKF) [78] has been developed by applying the maximum correntropy criterion into the Kalman filter. Moreover, the use of the MCKF achieves a higher accuracy and stronger robustness against non-Gaussian noise. These advantages are also realized in the maximum correntropy sparse Gaussian Hermitian quadrature filter (MCSGHQF) [86] by introducing a correntropy measure to the sparse Gaussian Hermitian quadrature filter. In comparison with the MCC, the generalized correntropy [73] is more robust and flexible in the presence of non-Gaussian noise. In this chapter, the dynamics of epidemic transmission on networks are estimated by the proposed generalized correntropy sparse Gaussian Hermite Quadrature filter (GCSGHQF) for further improving the stability. Specifically, the main contributions are as follows.

- A robust nonlinear Kalman filter, referred to as generalized correntropy sparse Gaussian Hermite Quadrature filter (GCSGHQF), is proposed for dealing with non-Gaussian noises. In particular, the GCSGHQF utilizes the generalized Gaussian density (GGD) function in place of the Gaussian kernel in MCC for improving the robustness against non-Gaussian noises.

- An extra second-order statistical characteristic of the error is incorporated in the GCSGHQF for further enhancing the numerical stability in the presence of non-Gaussian noises.

- The proposed GCSGHQF is used for the state estimation of epidemic spreading in the presence of Gaussian and non-Gaussian noises. In addition, the GCSGHQF is also utilized for studying the behavior of epidemic spreading on commonly used complex networks like ER networks, NW and WS networks. Simulation results show the improved robustness of the GCSGHQF for tracking epidemic spreading.

The remainder of the chapter is organized as follows. In Section 4.2, we briefly review compartment models for describing epidemic spreading. Section 4.3 and

Section 4.4 introduce the generalized correntropy and the sparse Gauss-Hermite quadrature filter, respectively. In Section 4.5, the generalized correntropy sparse Gauss-Hermite quadrature filter is proposed for guaranteeing the estimation robustness under non-Gaussian noises. In Section 4.6, the robustness of the GCSGHQF is demonstrated for different scenarios through numerical experiments. A conclusion is given in Section 4.7.

4.2 Review of Compartmental Models

As mentioned previously, the SIS and SIRS models are closed-loop models which are suited for implementing nonlinear Kalman filters due to their numerical stability. In comparison with the SIS model, the SIRS model introduces an extra compartment of recovered individuals for closer resemblance with practical epidemic spreading. The SIRS compartmental model for describing epidemic transmission on a degree-correlated network is given by

$$s_{d_{\eta_j},k} = s_{d_{\eta_j},k-1} - \alpha d_{\eta_j} s_{d_{\eta_j},k-1} i^w_{d_{\eta_j},k-1} + \Delta s_{k-1} \tag{4.1}$$

$$i_{d_{\eta_j},k} = i_{d_{\eta_j},k-1} + \alpha d_{\eta_j} s_{d_{\eta_j},k-1} i^w_{d_{\eta_j},k-1} - \Delta i_{k-1} \tag{4.2}$$

with $\Delta s_{k-1}, \Delta i_{k-1}$ and $i^w_{d_{\eta_j},k-1}$ expressed by

$$\Delta s_{k-1} = \gamma r_{d_{\eta_j},k-1}, \Delta i_{k-1} = \beta i_{d_{\eta_j},k-1}, i^w_{d_{\eta_j},k-1} = \sum_{m=1}^{|\mathbf{D}|} p\left(d_{\eta_m} | d_{\eta_j}\right) i_{d_{\eta_j},k-1}. \tag{4.3}$$

Also, $s_{d_{\eta_j},k-1}$, $i_{d_{\eta_j},k-1}$ and $r_{d_{\eta_j},k-1}$ denote the fractions of susceptible, infected and recovered individuals associated with degree d_{η_j} at discrete time $k-1$. Parameters α, β and γ represent the infected rate, removal or recovery rate, and transmission rate, respectively.

The SIRS compartmental model given in (4.1) and (4.2) describes epidemic

spreading on degree-correlated networks. The networks, however, may have a large size of the types of degree $|\mathbf{D}|$, thus incurring a high computational cost for calculating the overall ratios of the susceptible, infected and recovered individuals. Unlike the SIRS model in (4.1) and (4.2), a simplified SIRS compartmental model, without considering the degree distribution, is defined by

$$s_k = s_{k-1} - \alpha s_{k-1} i_{k-1} + \gamma r_{k-1} \tag{4.4}$$

$$i_k = i_{k-1} + \alpha s_{k-1} i_{k-1} - \beta i_{k-1} \tag{4.5}$$

where s_{k-1}, i_{k-1} and r_{k-1} represent the fractions of susceptible, infected and recovered individuals at discrete time $k-1$. The simplified SIRS model without the requirement of finding the degree distribution is less computational intensive, but it neglects degree information when nonlinear Kalman filters perform state estimation.

Remark 6 *The SIRS compartmental model given by (4.1) and (4.2) is utilized for tracking the epidemic transmission by incorporating a prior information, i.e., degree distribution, which may better demonstrate epidemic spreading. As the size of the types of degree increases, however, the computational complexity increases significantly. In addition, the fraction of the infected individuals may be close or equal to zero since individuals with j-th type degree may be not infected. Similar conclusion can be obtained for the fractions of the susceptible and recovered individuals. Unlike the SIRS compartmental model in (4.1) and (4.2), the simplified SIRS model given by (4.4) and (4.5) is constructed without considering the degree distribution and therefore may not generate the fractions of individuals close or equal to zero, which is beneficial to numerical stability of nonlinear Kalman filters. In addition, the simplified SIRS model given by (4.4) and (4.5) has a significant advantage in terms of computational cost over the degree-based SIRS model given by (4.1) and (4.2).*

4.3 Generalized Correntropy

The correntropy of two arbitrary scalar random variables X and Y measures their similarity, and is defined by [78]

$$V(X,Y) = E_{XY}(\kappa_\sigma(X,Y)) = \int \kappa_\sigma(x,y)\, dF_{XY}(x,y) \tag{4.6}$$

where $E_{XY}(\cdot)$ denotes the expectation with distribution function F_{XY} and $\kappa_\sigma(\cdot)$ is a kernel function with kernel bandwidth σ. The Gaussian kernel [79] is the most commonly used kernel function and is given by

$$\kappa_\sigma(x,y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{e^2}{2\sigma^2}\right) \tag{4.7}$$

where $\sigma > 0$ denotes the Gaussian kernel bandwidth and $e = x - y$ is the difference between x and y.

Equation (4.6), however, cannot be calculated since the probability distribution function is unknown in practice. Thus, (4.6) is generally approximated by utilizing $\tilde{N}$ available samples $\{x_i, y_i\}_{i=1}^{\tilde{N}}$, i.e.,

$$\hat{V}(X,Y) \approx \frac{1}{\tilde{N}} \sum_{i=1}^{\tilde{N}} \kappa_\sigma(x_i - y_i). \tag{4.8}$$

The generalized correntropy substitutes the Gaussian kernel in (4.7) by a generalized kernel function, i.e., [84]

$$G_{\alpha,\beta}(e) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left|\frac{e}{\beta}\right|^\alpha\right) \tag{4.9}$$

where $\Gamma(\cdot)$ is the gamma function; α (positive) and β denote the shape parameter and bandwidth parameter, respectively. According to (4.9), the Laplace distribution can be obtained for $\alpha = 1$. In addition, (4.9) reduces to the Gaussian distribution when

the shape parameter α is set to 2. The generalized correntropy density converges to a uniform density on $(-\beta, \beta)$ in a point-wise manner [73, 85]. Equation (4.9) can also be simplified as

$$G_{\alpha,\sigma}(e) = \varsigma \exp\left(-\frac{|e|^{\alpha}}{2\sigma^2}\right) \tag{4.10}$$

where γ and σ are scaling factor and kernel width, respectively. Substituting (4.10) into (4.6) gives

$$V_{\alpha,\sigma}(X, Y) = E_{XY}(G_{\alpha,\sigma}(X - Y)). \tag{4.11}$$

Similar to (4.8), (4.11) can be approximated as

$$\hat{V}_{\alpha,\sigma}(X, Y) = \frac{1}{\tilde{N}} \sum_{i=1}^{\tilde{N}} G_{\alpha,\sigma}(x_i - y_i). \tag{4.12}$$

Some important properties of (4.12) can be found in [73].

4.4 Sparse Gauss-Hermite Quadrature Nonlinear Filter

4.4.1 State Space Model

Consider a nonlinear discrete-time dynamical system, with the state function $\boldsymbol{f}(\cdot)$ and measurement function $\boldsymbol{h}(\cdot)$ given by [86]

$$\mathbf{x}_k = \boldsymbol{f}(\mathbf{x}_{k-1}) + \mathbf{w}_k \tag{4.13}$$

$$\mathbf{y}_k = \boldsymbol{h}(\mathbf{x}_k) + \mathbf{v}_k \tag{4.14}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ and $\mathbf{y}_k \in \mathbb{R}^m$ represent the state vector and measurement vector, respectively; state noise $\mathbf{w}_k$ and measurement noise $\mathbf{v}_k$ are Gaussian distributed, i.e., $\mathbf{w}_k \sim \mathbb{N}(\mathbf{0}, \mathbf{Q}_k)$ and $\mathbf{v}_k \sim \mathbb{N}(\mathbf{0}, \mathbf{R}_k)$, respectively. Comparing (4.13) and (4.14) with (4.4) and (4.5), we define state variable $\mathbf{x}_k = [s_k \;\; i_k]^T$. The nonlinear process function $\boldsymbol{f}(\cdot)$ is defined by (4.4) and (4.5). The measurement function $\boldsymbol{h}(\cdot)$ takes a linear form with the measurement matrix $\mathbf{H} = [0 \;\; 1]$, generating a linear measurement function

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k. \tag{4.15}$$

A nonlinear Kalman filter aims to estimate the hidden state by a prediction step and an update step, which calculate the predictive and posterior density of the state at discrete time k, $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ and $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, respectively [47]. The *Chapman-Kolmogorov* equation [106] is generally utilized for calculating the predictive density of the state at discrete time k, $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$, i.e.,

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_{1:k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1} \\ &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1} \end{aligned} \tag{4.16}$$

where the disappearance of past measurements $\mathbf{y}_{1:k-1}$ is due to the Markov property of the sequence $\{\mathbf{x}_k : k = 1, 2, \ldots\}$ [106]. The Bayes' rule can be applied for calculating $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ at the update step [47], i.e.,

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) d\mathbf{x}_k}. \tag{4.17}$$

Moreover, to the independence of $p(\mathbf{y}_{1:k}) = \int p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) d\mathbf{x}_k$ on the state, we have

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k-1}). \tag{4.18}$$

Algorithm 4 Nonlinear Kalman Filter Based on Deterministic Sampling

Time update

1: Obtain sample points $\{\boldsymbol{\xi}_{i,k-1}\}_{i=1}^{\Theta}$ and the corresponding weights ω_i, e.g., the unscented transformation [103], cubature rule [46]. Herein, $i = 1, 2, \cdots, \Theta$, where Θ denotes the total number of the sample points.

2: Compute the predicted mean and the corresponding error covariance of state

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=1}^{\Theta} \omega_i \boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}), \tag{4.19}$$

$$\mathbf{P}_{k|k-1} = \mathbf{P}^{o}_{k|k-1} + \mathbf{Q}_k \tag{4.20}$$

where $\mathbf{P}^{o}_{k|k-1} = \sum_{i=1}^{\Theta} \omega_i (\boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k|k-1})(\boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}) - \hat{\mathbf{x}}_{k|k-1})^T$.

Measurement update

1: Obtain the sample points $\{\boldsymbol{\xi}_{i,k|k-1}\}_{i=1}^{\Theta}$ and the corresponding weights ω_i, $i = 1, 2, \cdots, \Theta$.

2: Compute the predicted mean and the corresponding error covariance of measurements

$$\hat{\mathbf{y}}_{k|k-1} = \sum_{i=1}^{\Theta} \omega_i \boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}), \tag{4.21}$$

$$\mathbf{P}_{yy,k|k-1} = \mathbf{P}^{o}_{yy,k|k-1} + \mathbf{R}_k \tag{4.22}$$

where matrix $\mathbf{P}^{o}_{yy,k|k-1} = \mathbf{R}_k + \sum_{i=1}^{\Theta} \omega_i (\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1})^T$.

3: Compute the cross covariance of state and measurement

$$\mathbf{P}_{xy,k|k-1} = \sum_{i=1}^{\Theta} \omega_i (\boldsymbol{\xi}_{i,k|k-1} - \hat{\mathbf{x}}_{k|k-1}) \left(\boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1}\right)^T. \tag{4.23}$$

4: Compute the Kalman gain

$$\mathbf{K}_k = \mathbf{P}_{xy,k|k-1} \mathbf{P}^{-1}_{yy,k|k-1}. \tag{4.24}$$

5: Compute the estimated mean and the corresponding error covariance of state

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1})), \tag{4.25}$$

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{yy,k|k-1} \mathbf{K}_k^T. \tag{4.26}$$

Therefore, a nonlinear Kalman filter estimates the hidden state by assuming a Gaussian distribution of the probability density functions, and computes the multi-dimensional

Algorithm 5 Points and Weights Based on Spare Grid Quadrature

FOR: $q = L - n : L - 1$
  Determine N_q^n
  For each element $\Xi = (l_1, \cdots, l_n)$ in N_q^n generates point $[u_{s_1}, u_{s_2}, \cdots, u_{s_n}]$ in $X_{l_1} \otimes X_{l_2} \cdots \otimes X_{l_n}$
    FOR each point in $X_{l_1} \otimes X_{l_2} \cdots \otimes X_{l_n}$
      IF the point is new, add it to the sample set, assign a new index i and calculate weight by $w_i = (-1)^{L-1-q} C_{n-1}^{L-1-q} \prod_{j=1}^{n} \bar{\omega}_{s_j}$
      ELSE update the old weight by $w_i = w_i + (-1)^{L-1-q} C_{n-1}^{L-1-q} \prod_{j=1}^{n} \bar{\omega}_{s_j}$
      END IF
    END FOR
  END FOR
END FOR

integrals in (4.16) and (4.17). **Algorithm** 4 summarizes the update procedure of the nonlinear Kalman filter using different methods for calculating the multi-dimensional integrals.

4.4.2 Sparse Gauss-Hermite Quadrature Rule

The Gauss-Hermite nonlinear filter uses the Gauss-Hermite rule [35] to approximate (4.16) and (4.17), incurring high computational cost [86]. In [35, 107, 108, 109], a sparse grid approach is utilized for alleviating the computational burden without degrading the filtering accuracy. Using the sparse grid method, the sparse Gauss-Hermite quadrature filter extends the univariate quadrature points to generate l multivariate quadrature points [35], i.e.,

$$\mathbf{u}_i = [u_{s_1}, u_{s_2}, \ldots, u_{s_n}]^T, \bar{\omega}_i = \prod_{j=1}^{n} \bar{w}_{s_j}, \quad 1 \leq i \leq l \tag{4.27}$$

where $\{u_{s_j}\}_{j=1}^{n}$ denotes the univariate quadrature points [86] and $\bar{w}_{s_j}$ represents the weight of the corresponding univariate quadrature point u_{s_j} [86]. The multivariate integrals shown in (4.16) and (4.17) are approximated by a linear combination of the lower-level tensor products of the univariate quadrature rule in the sparse-grid

approach on the basis of the Smolyak rule [78], given by

$$\begin{aligned}&\int_{\mathbb{R}^n} \boldsymbol{f}(\mathbf{x})\mathcal{N}(\mathbf{x};\mathbf{0},\mathbf{I}_n)d\mathbf{x} \approx I_{n,L}(\boldsymbol{f})\\ &= \sum_{q=L-n}^{L-1}(-1)^{L-1-q}C_{n-1}^{L-1-q}\sum_{\Xi\in N_q^n}(I_{l_1}\otimes\cdots\otimes I_{l_n})\boldsymbol{f}\end{aligned} \tag{4.28}$$

where $\mathbf{x} = [x(1), x(2), \cdots, x(p), \cdots, x(n)]^T \in \mathbb{R}^n$ and $I_{n,L}(\boldsymbol{f})$ approximates the n-dimensional integral of the function $\boldsymbol{f}$ with respect to the Gaussian distribution $\mathcal{N}(\mathbf{x};\mathbf{0},\mathbf{I}_n)$ with the accuracy level $L \in \mathbb{N}$, the set of natural numbers $\mathbb{N}$ [35] and the identity matrix $\mathbf{I}_n \in \mathbb{R}^n$. Also, I_{l_j} is the univariate quadrature rule with accuracy level of $l_j \in \Xi = (l_1, \cdots, l_n)$. The N_q^n is the set of accuracy level sequences and defined by

$$N_q^n = \begin{cases} \Xi : \sum_{j=1}^n l_j = n+q & \text{if } q \geq 0 \\ \varnothing & \text{else} \end{cases} \tag{4.29}$$

where $\varnothing$ is the null set and the auxiliary parameter q ranges from $L-n$ to $L-1$. In (4.28), $\otimes$ denotes tensor product [35]; C_{n-1}^{L-1-q} represents the binomial coefficient defined by $C_n^\nu = n!/\left[\nu!(n-\nu)!\right]$ with $n!$ denotes the factorial of n.

It should be noted that (4.28) can also take an extended form as [104]

$$\begin{aligned}I_{n,L}(\boldsymbol{f}) = &\sum_{q=L-n}^{L-1}\sum_{\Xi\in N_q^n}\sum_{u_{s_1}\in X_{l_1}}\cdots\sum_{u_{s_n}\in X_{l_n}}\boldsymbol{f}(u_{s_1},\ldots,u_{s_n})\\ &\left\{(-1)^{L-1-q}C_{n-1}^{L-1-q}\prod_{j=1}^n \bar{w}_{s_j}\right\}\end{aligned} \tag{4.30}$$

where X_{l_i} is the point set of the univariate quadrature rule I_{l_i} with accuracy level l_i. Thus, the set of sparse-grid points $X_{n,L}$ is given by [104]

$$X_{n,L} = \bigcup_{q=L-n}^{L-1}\bigcup_{\Xi\in N_q^n}(X_{l_1}\otimes X_{l_2}\otimes\cdots X_{l_n}) \tag{4.31}$$

where $\bigcup$ denotes the union of the quadrature point sets.

Algorithm 5 provides the detailed operation of points and weights on the basis of sparse grid quadrature [35].

4.5 Generalized Correntropy Sparse Gauss-Hermite Quadrature Filter

The generalized correntropy sparse Gauss-Hermite quadrature filter uses the generalized correntropy for further improving the robustness in the presence of outliers. To incorporate the generalized correntropy into the SGHQF, a new measurement update is constructed [86], i.e.,

$$\begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_k \end{bmatrix} \mathbf{x}_k + \bar{\mathbf{v}}_k \tag{4.32}$$

where $\mathbf{I} \in \mathbb{R}^n$ is the identity matrix, and

$$\bar{\mathbf{v}}_k = \begin{bmatrix} -(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \\ \mathbf{v}_k \end{bmatrix}. \tag{4.33}$$

In (4.33), $\bar{\mathbf{v}}_k$ has a second-order statistical characteristic $E[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^T]$ given by

$$E[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^T] = \begin{bmatrix} \mathbf{P}_{k|k-1} & 0 \\ 0 & \mathbf{R}_k \end{bmatrix} \tag{4.34}$$

$$= \begin{bmatrix} \tilde{\mathbf{P}}_{p,k|k-1}\tilde{\mathbf{P}}_{p,k|k-1}^T & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{P}}_{v,k}\tilde{\mathbf{P}}_{v,k}^T \end{bmatrix} \tag{4.35}$$

$$= \tilde{\mathbf{P}}_k\tilde{\mathbf{P}}_k^T \tag{4.36}$$

where $\tilde{\mathbf{P}}_k$ takes the form of

$$\tilde{\mathbf{P}}_k = \begin{bmatrix} \tilde{\mathbf{P}}_{p,k|k-1} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{P}}_{v,k} \end{bmatrix}. \tag{4.37}$$

Furthermore, the measurement matrix $\mathbf{H}_k$ in (4.32) is expressed as [86]

$$\mathbf{H}_k = \left(\mathbf{P}_{k|k-1}^{-1}\mathbf{P}_{xy,k|k-1}\right)^T. \tag{4.38}$$

Applying a Cholesky decomposition to $E[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^T]$ generates the matrix $\tilde{\mathbf{P}}_k$ in (4.36). A linear regression model can be developed by multiplying both sides of (4.32) by $\tilde{\mathbf{P}}_k^{-1}$, i.e.,

$$\mathbf{\Gamma}_k = \mathbf{\Pi}_k\mathbf{x}_k + \mathbf{e}_k \tag{4.39}$$

where

$$\mathbf{\Gamma}_k = \tilde{\mathbf{P}}_k^{-1}\begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} \end{bmatrix}, \mathbf{\Pi}_k = \tilde{\mathbf{P}}_k^{-1}\begin{bmatrix} \mathbf{I} \\ \mathbf{H}_k \end{bmatrix}. \tag{4.40}$$

The linear measurement function, i.e., $\boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1}) = \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}$ in (4.40), generates

$$\mathbf{\Gamma}_k = \tilde{\mathbf{P}}_k^{-1}\begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k \end{bmatrix}. \tag{4.41}$$

The error $\mathbf{e}_k = \tilde{\mathbf{P}}_k^{-1}\bar{\mathbf{v}}_k$ in (4.39) is white for $E[\mathbf{e}_k\mathbf{e}_k^T] = \mathbf{I}$.

For getting the optimal values of the state variables, the generalized correntropy-based cost function is considered as [84]

$$J(\mathbf{x}_k) = \frac{1}{m+n}\sum_{i=1}^{m+n} G_{\alpha,\sigma}(\Gamma_k(i) - \mathbf{\Pi}_k(i)\mathbf{x}_k) \tag{4.42}$$

where $\Gamma_k(i)$ is the i-th element of $\mathbf{\Gamma}_k \in \mathbb{R}^{n+m}$ and $\mathbf{\Pi}_k(i)$ is the i-th row of $\mathbf{\Pi}_k$ at discrete time k. In addition, the extra second-order statistical characteristic of error is incorporated into (4.42) for enhancing the numerical stability, giving

$$J\left(\mathbf{x}_k\right) = \sum_{i=1}^{m+n} G_{\alpha,\sigma}\left(\Gamma_k\left(i\right) - \mathbf{\Pi}_k\left(i\right)\mathbf{x}_k\right) - \frac{\lambda}{2}\left(\Gamma_k\left(i\right) - \mathbf{\Pi}_k\left(i\right)\mathbf{x}_k\right)^2 \tag{4.43}$$

where the parameter λ is considered for balancing the generalized correntropy and second-order statistical characteristic of error terms. Notations n and m represent the dimensions of the state and available measurement defined in (4.13) and (4.14). Therefore, the optimal estimate of $\mathbf{x}_k$ can be obtained as

$$\begin{aligned}\hat{\mathbf{x}}_k &= \arg\ \max_{\hat{\mathbf{x}}_k} J\left(\mathbf{x}_k\right) \\ &= \arg\max_{\hat{\mathbf{x}}_k} \sum_{i=1}^{m+n} G_{\alpha,\sigma}\left(e_k\left(i\right)\right) - \frac{\lambda}{2}(\Gamma_k\left(i\right) - \mathbf{\Pi}_k\left(i\right)\mathbf{x}_k)^2\end{aligned} \tag{4.44}$$

where error $e_k(i)$ denotes the i-th element of $\mathbf{e}_k$ at discrete time k, defined by

$$e_k(i) = \Gamma_k(i) - \mathbf{\Pi}_k(i)\mathbf{x}_k. \tag{4.45}$$

Differentiating (4.44) with respect to $\mathbf{x}_k$ gives

$$\begin{aligned}\frac{\partial J\left(\mathbf{x}_k\right)}{\partial \mathbf{x}_k}\,|_{\mathbf{x}_k=\hat{\mathbf{x}}_k} &= \sum_{i=1}^{m+n} G_{\alpha,\sigma}\left(e_k\left(i\right)\right)\left(\frac{\alpha|e_k\left(i\right)|^{\alpha-2}}{2\sigma^2} e_k\left(i\right)\mathbf{\Pi}_k^T\left(i\right)\right) + \lambda e_k\left(i\right)\mathbf{\Pi}_k^T\left(i\right)|_{\mathbf{x}_k=\hat{\mathbf{x}}_k} \\ &= 0.\end{aligned} \tag{4.46}$$

The optimal state estimation can therefore be written as

$$\begin{aligned}\hat{\mathbf{x}}_k = &\left(\sum_{i=1}^{m+n}\left(\tilde{G}_{\alpha,\sigma}\left(e_k\left(i\right)\right) + \bar{\lambda}\right)\mathbf{\Pi}_k^T\left(i\right)\mathbf{\Pi}_k\left(i\right)\right)^{-1} \\ &\left(\sum_{i=1}^{m+n}\left(\tilde{G}_{\alpha,\sigma}\left(e_k\left(i\right)\right) + \bar{\lambda}\right)\mathbf{\Pi}_k^T\left(i\right)\Gamma_k\left(i\right)\right)\end{aligned} \tag{4.47}$$

where $\tilde{G}_{\alpha,\sigma}(e_k(i)) = \exp\left(-|e_k(i)|^{\alpha}\Big/2\sigma^2\right)|e_k(i)|^{\alpha-2}$ and $\bar{\lambda} = 2\sigma^2\lambda/\alpha\varsigma = \sigma^2\tilde{\lambda}/\alpha$, $\tilde{\lambda} = 2\lambda/\varsigma$ with the scaling factor ς defined in (4.10).

Define a weighted matrix $\mathbf{\Lambda}_k \in \mathbb{R}^{(m+n)\times(m+n)}$ by

$$\mathbf{\Lambda}_k = \begin{bmatrix} \mathbf{\Lambda}_{x,k} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{y,k} \end{bmatrix} \tag{4.48}$$

where matrixes $\mathbf{\Lambda}_{x,k} \in \mathbb{R}^{n\times n}$ and $\mathbf{\Lambda}_{y,k} \in \mathbb{R}^{m\times m}$ are defined by

$$\mathbf{\Lambda}_{x,k} = diag(\tilde{G}_{\alpha,\sigma}(e_k(1)) + \bar{\lambda}, \cdots, \tilde{G}_{\alpha,\sigma}(e_k(n)) + \bar{\lambda}), \tag{4.49}$$

$$\mathbf{\Lambda}_{y,k} = diag(\tilde{G}_{\alpha,\sigma}(e_k(n+1)) + \bar{\lambda}, \cdots, \tilde{G}_{\alpha,\sigma}(e_k(n+m))) + \bar{\lambda}). \tag{4.50}$$

The $\mathbf{\Lambda}_{x,k}$ in (4.49) is a diagonal matrix with diagonal elements $\{\tilde{G}_{\alpha,\sigma}(e_k(j)) + \bar{\lambda}\}_{j=1}^{n}$. The diagonal matrix $\mathbf{\Lambda}_{y,k}$ in (4.50) has diagonal elements $\{\tilde{G}_{\alpha,\sigma}(e_k(j)) + \bar{\lambda}\}_{j=n+1}^{j=n+m}$. Equation (4.47) can therefore be simplified as

$$\begin{aligned} \hat{\mathbf{x}}_k &= \left(\mathbf{\Pi}_k^T\mathbf{\Lambda}_k\mathbf{\Pi}_k\right)^{-1}\mathbf{\Pi}_k^T\mathbf{\Lambda}_k\mathbf{\Gamma}_k \\ &= \mathbf{\Pi}_{C,k}^{-1}\mathbf{\Pi}_{D,k} \end{aligned} \tag{4.51}$$

where $\mathbf{\Pi}_{C,k} = \mathbf{\Pi}_k^T\mathbf{\Lambda}_k\mathbf{\Pi}_k$ and $\mathbf{\Pi}_{D,k} = \mathbf{\Pi}_k^T\mathbf{\Lambda}_k\mathbf{\Gamma}_k$ with $\mathbf{\Pi}_k$ and $\mathbf{\Gamma}_k$ defined in (4.40).

According to (4.35), $\mathbf{\Pi}_k$ and $\mathbf{\Gamma}_k$ can be further calculated as

$$\mathbf{\Gamma}_k = \begin{bmatrix} \tilde{\mathbf{P}}_{p,k|k-1}^{-1}\hat{\mathbf{x}}_{k|k-1} \\ \tilde{\mathbf{P}}_{v,k}^{-1}\mathbf{y}_k \end{bmatrix}, \mathbf{\Pi}_k = \begin{bmatrix} \tilde{\mathbf{P}}_{p,k|k-1}^{-1} \\ \tilde{\mathbf{P}}_{v,k}^{-1}\mathbf{H}_k \end{bmatrix}. \tag{4.52}$$

According to (4.33) and (4.52), $\mathbf{\Pi}_{C,k}$ in (4.51) can be expanded as

$$\mathbf{\Pi}_{C,k}^{-1} = (\mathbf{H}_k^T(\tilde{\mathbf{P}}_{v,k}^{-1})^T\mathbf{\Lambda}_{y,k}\tilde{\mathbf{P}}_{v,k}^{-1}\mathbf{H}_k + (\tilde{\mathbf{P}}_{p,k|k-1}^{-1})^T\mathbf{\Lambda}_{x,k}\tilde{\mathbf{P}}_{p,k|k-1}^{-1})^{-1} \tag{4.53}$$

with $\tilde{\mathbf{P}}_{p,k|k-1}$ and $\tilde{\mathbf{P}}_{v,k}$ given by (4.35). By using the matrix inverse lemma given in [79],

i.e.,

$$\left(\bar{\mathbf{A}}+\bar{\mathbf{B}}\bar{\mathbf{C}}\bar{\mathbf{D}}\right)^{-1}=\bar{\mathbf{A}}^{-1}-\bar{\mathbf{A}}^{-1}\bar{\mathbf{B}}\left(\bar{\mathbf{C}}^{-1}+\bar{\mathbf{D}}\bar{\mathbf{A}}^{-1}\bar{\mathbf{B}}\right)^{-1}\bar{\mathbf{D}}\bar{\mathbf{A}}^{-1} \tag{4.54}$$

and letting $\bar{\mathbf{A}}=\left(\tilde{\mathbf{P}}^{-1}_{p,k|k-1}\right)^{T}\mathbf{\Lambda}_{x,k}\tilde{\mathbf{P}}^{-1}_{p,k|k-1}$, $\bar{\mathbf{B}}=\mathbf{H}_k^T$, $\bar{\mathbf{C}}=\left(\tilde{\mathbf{P}}^{-1}_{v,k}\right)^{T}\mathbf{\Lambda}_{y,k}\tilde{\mathbf{P}}^{-1}_{v,k}$ and $\bar{\mathbf{D}}=\mathbf{H}_k$, (4.53) can be further expanded by

$$\mathbf{\Pi}^{-1}_{C,k}=\tilde{\mathbf{P}}_{p,k|k-1}\mathbf{\Lambda}^{-1}_{x,k}\tilde{\mathbf{P}}^{T}_{p,k|k-1}-\tilde{\mathbf{P}}_{p,k|k-1}\mathbf{\Lambda}^{-1}_{x,k}\tilde{\mathbf{P}}^{T}_{p,k|k-1}\mathbf{H}_k^T(\tilde{\mathbf{P}}_{v,k}\mathbf{\Lambda}^{-1}_{y,k}\tilde{\mathbf{P}}^{T}_{v,k}+\mathbf{H}_k\tilde{\mathbf{P}}_{p,k|k-1}\mathbf{\Lambda}^{-1}_{x,k}\tilde{\mathbf{P}}^{T}_{p,k|k-1}\mathbf{H}_k^T)^{-1}$$
$$\mathbf{H}_k\tilde{\mathbf{P}}_{p,k|k-1}\mathbf{\Lambda}^{-1}_{x,k}\tilde{\mathbf{P}}^{T}_{p,k|k-1}. \tag{4.55}$$

Likewise, $\mathbf{\Pi}_{D,k}$ in (4.51) can be derived by combining (4.33) and (4.52), i.e.,

$$\mathbf{\Pi}_k^T\mathbf{\Lambda}_k\mathbf{\Gamma}_k=\mathbf{H}_k^T\left(\tilde{\mathbf{P}}^{-1}_{v,k}\right)^{T}\mathbf{\Lambda}_{y,k}\tilde{\mathbf{P}}^{-1}_{v,k}\mathbf{y}_k+\left(\tilde{\mathbf{P}}^{-1}_{p,k|k-1}\right)^{T}\mathbf{\Lambda}_{x,k}\tilde{\mathbf{P}}^{-1}_{p,k|k-1}\hat{\mathbf{x}}_{k|k-1}. \tag{4.56}$$

From (4.53) to (4.56), the estimated state $\hat{\mathbf{x}}_k$ and corresponding covariance $\bar{\mathbf{P}}_k$ are represented by

$$\hat{\mathbf{x}}_k=\hat{\mathbf{x}}_{k|k-1}+\bar{\mathbf{K}}_k\left(\mathbf{y}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right) \tag{4.57}$$

$$\bar{\mathbf{P}}_k=(\mathbf{I}-\bar{\mathbf{K}}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}(\mathbf{I}-\bar{\mathbf{K}}_k\mathbf{H}_k)^T+\bar{\mathbf{K}}_k\mathbf{R}_k\bar{\mathbf{K}}_k^T \tag{4.58}$$

with the gain matrix $\bar{\mathbf{K}}_k$ defined by

$$\bar{\mathbf{K}}_k=\bar{\mathbf{P}}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{H}_k\bar{\mathbf{P}}_{k|k-1}\mathbf{H}_k^T+\bar{\mathbf{R}}_k\right)^{-1}. \tag{4.59}$$

The modified prediction error covariance matrices $\bar{\mathbf{P}}_{k|k-1}$ and $\bar{\mathbf{R}}_k$ are defined by

$$\bar{\mathbf{P}}_{k|k-1}=\tilde{\mathbf{P}}_{p,k|k-1}\mathbf{\Lambda}^{-1}_{x,k}\tilde{\mathbf{P}}^{T}_{p,k|k-1} \tag{4.60}$$

$$\bar{\mathbf{R}}_k=\tilde{\mathbf{P}}_{v,k}\mathbf{\Lambda}^{-1}_{y,k}\tilde{\mathbf{P}}^{T}_{v,k}. \tag{4.61}$$

From (4.60) and (4.61), matrix $\mathbf{\Lambda}_k$ plays the role of weighting the prediction covariance matrix $\mathbf{P}_{k|k-1}$ and covariance matrix of measurement noise $\mathbf{R}_k$. Due to the introduction of the matrix $\mathbf{\Lambda}_k$ in (4.60) and (4.61), the proposed method has the ability of suppressing abnormal measurements at non-Gaussian noise.

Remark 7 *The generalized correntropy with varying parameter α has the ability of suppressing abnormal errors. It is challenging to provide a specific guideline for choosing parameter α for dealing with non-Gaussian noise. However, a general guideline can be provided. In particular, a large value of α is suitable for dealing with measurement errors where measurements are greatly contaminated. Otherwise, a small parameter α is considered. In addition, a matrix inversion is calculated in (4.47) and a relatively larger error may cause numerical divergence. For solving this issue, a second-order statistical characteristic of errors is also introduced in (4.43) for enhancing the numerical stability.*

Remark 8 *The proposed generalized correntropy sparse Gauss-Hermite quadrature filter (GCSGHQF) is derived on the sparse Gauss-Hermite quadrature filter (SGHQF). The SGHQF is a simplified version of the traditional Gauss-Hermite quadrature filter but still achieves higher accuracy with appropriate parameters than methods in [110] like CKF and UKF [35]. In addition, the proposed GCSGHQF is robust by incorporating generalized correntropy and is suitable for dealing with various non-Gaussian noises by adjusting exponential term in (4.10). The approach in [111] also provides different degree of suppression with respect to varying magnitude measurements. However, the proposed method not only achieves this goal but also provides a soft way for dealing with non-Gaussian noises due to the use of exponential function in (4.10). This is especially advantageous in retaining measurement information for improving filtering precision.*

Table 4.1: RTAMSEs of Nonlinear Kalman Filter in the Presence of Gaussian and Non-Gaussian Noises

Algorithm	CKF3	SGHQF	GCSGHQF			
			$\alpha = 1$	$\alpha = 2$	$\alpha = 4$	$\alpha = 6$
$\text{RTAMSE}_s^G(10^{-1})$	0.62	0.60	0.60	0.60	0.60	0.60
$\text{RTAMSE}_s^U(10^{-1})$	0.75	0.71	0.68	0.70	0.68	0.60
$\text{RTAMSE}_s^L(10^{-1})$	0.38	0.36	0.29	0.35	0.33	0.34

4.6 Results

This section compares the filtering accuracy of nonlinear Kalman filters for state estimation of epidemic transmission on homogeneous networks, including the ER, NW and WS networks. The proposed generalized correntropy sparse Gauss-Hermite quadrature filter is compared with other nonlinear Kalman filters. In particular, the extended Kalman filter in Chapter 3 has the lowest filtering precision among all filters. Therefore, there is no need for making a comparison between extended Kalman filter and proposed filter. The unscented Kalman filter and fifth-degree cubature Kalman filter in Chapter 3 may be all confronted with the issue of numerical instability. Therefore, these two Kalman filters are also not considered here. The third-degree cubature Kalman filter has desirable filtering precision and numerical stability. As a result, only third-degree cubature Kalman filter in Chapter 3 is considered here. In addition, SGHQF and robust nonlinear Kalman filter like MCSGHQF ($\alpha = 2$ for GCSGHQF) are also compared with the proposed nonlinear Kalman filter. The proposed GCSGHQF is initially tested on synthetic data, which are generated by the compartmental model shown in (4.4) and (4.5) in the presence of Gaussian and non-Gaussian noises such as the uniformly distributed noise and Laplace distributed noise. In addition, the superior performance of the proposed GCSGHQF for state estimation on the ER, NW and WS networks is demonstrated.

In each experiment, 50 Monte Carlo simulations are run. Different performance metrics including the root mean square error (RMSE) [112] and root time averaged mean square error (RTAMSE) [113] are introduced to evaluate the filtering precision of nonlinear Kalman filters. For example, the RMSE of the fraction of susceptible individuals at discrete time k and the corresponding RTAMSE over all discrete time instants are defined by

$$\mathrm{RMSE}_s(k) = \sqrt{\frac{1}{M}\sum_{\ell=1}^{M}(s_k^\ell - \hat{s}_k^\ell)^2} \tag{4.62}$$

$$\mathrm{RTAMSE}_s = \sqrt{\frac{1}{M}\frac{1}{\bar{N}}\sum_{k=1}^{\bar{N}}\sum_{\ell=1}^{M}(s_k^\ell - \hat{s}_k^\ell)^2} \tag{4.63}$$

where $\bar{N}$ denotes the number of samples, and s_k^ℓ and $\hat{s}_k^\ell$ represent the desired state and estimated state at discrete time k at the ℓ-th Monte Carlo simulation, respectively. Similar definitions can be applied for the RMSE and RTAMSE of the fraction of the infected individuals. In consideration of the linear measurement matrix $\mathbf{H}$ shown in (4.15), we focus on $\mathrm{RMSE_s}$ and RTAMSE_s in the numerical examples.

4.6.1 Epidemic Spreading in Gaussian Noise

The estimation accuracy of nonlinear Kalman filters is first compared in the presence of Gaussian noise. The state noise follows the Gaussian distribution $\mathbb{N}(\mathbf{0}_2, 10^{-3}\mathbf{I}_2)$, where $\mathbf{I}_2 \in \mathbb{R}^2$ represents the identity matrix. The measurement noise follows the Gaussian distribution $\mathbb{N}(0, 0.05)$. The parameter set of the SIRS model, i.e., $\bar{\omega}_{\mathrm{SIRS}} = (\alpha, \beta, \gamma)$ with the infected rate α, removal or recovery rate β and transmission rate γ from the recovered to the susceptible, is set as $\bar{\omega}_{\mathrm{SIRS}} = (0.7, 0.1, 0.1)$. The parameter set $\bar{\omega}_{\mathrm{KF}} = (\alpha, \beta, \gamma)$ for nonlinear Kalman filters performing state estimation is set as $\bar{\omega}_{\mathrm{KF}} = (0.5, 0.073, 0.07)$. The kernel parameter σ in (4.47) is given by $\sigma = 100$. The regularization factor $\tilde{\lambda}$ in (4.47) is set as $\tilde{\lambda} = 3$ with $\lambda = 0.75$ and $\varsigma = 0.5$. Fig. 4.1(a)

shows the RMSE of s_k corresponding to the Gaussian distributed noise. The values of RTAMSE_s^G, RTAMSE_s^U and RTAMSE_s^L in Table 4.1 represent the RTAMSE of susceptible individuals in the presence of Gaussian and non-Gaussian noises including the uniform noise and Laplace noise. From Fig. 4.1(a) and Table 4.1, we see that the SGHQF achieves a higher precision than the CKF3. In addition, the GCSGHQF with $\alpha = 1, 2, 4$ and 6 achieves a similar filtering accuracy as the SGHQF.

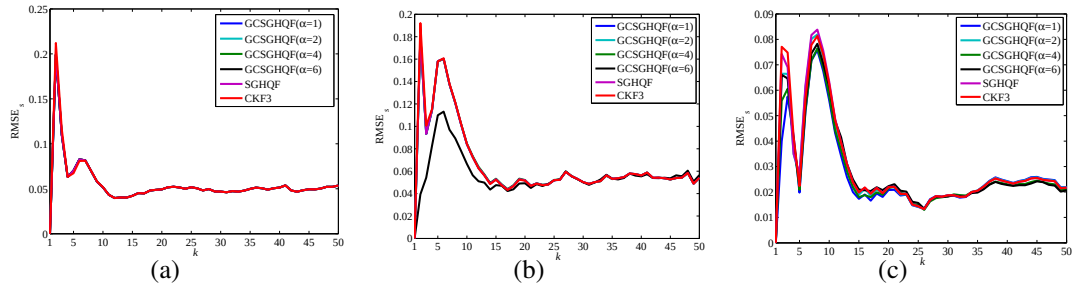


Figure 4.1: (a) RMSE_s in Gaussian noise; (b) RMSE_s in Uniform noise; (c) RMSE_s in Laplace noise.

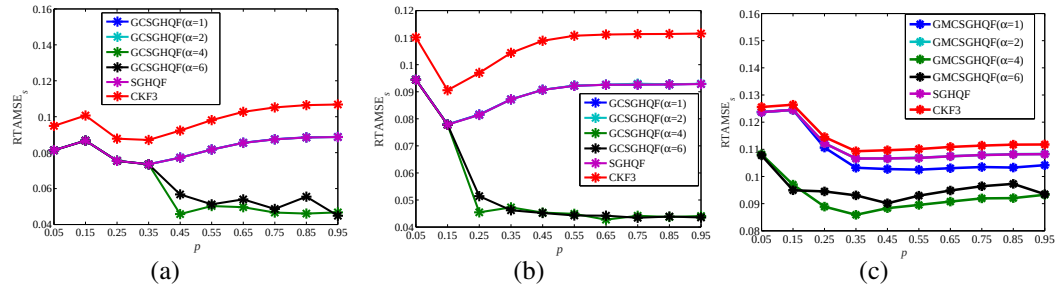


Figure 4.2: (a) RTAMSE_s in ER network; (b) RTAMSE_s in NW network; (c) RTAMSE_s in WS network.

4.6.2 Epidemic Spreading in Non-Gaussian Noise

In practice, measurements are generally affected by non-Gaussian noise rather than Gaussian noise. Therefore, we compare nonlinear Kalman filters for tracking epidemic transmission in the presence of non-Gaussian noises such as uniformly distributed noise and Laplace distributed noise.

4.6.2.1 Epidemic Spreading in Uniform Noise

Measurements, corrupted by uniformly distributed noise, are utilized for demonstrating the effectiveness of the proposed GCSGHQF in the presence of non-Gaussian noise. The state noise obeys Gaussian distribution $\mathbb{N}(\mathbf{0}_2, 10^{-3}\mathbf{I}_2)$. The measurement noise follows the uniform distribution $\mathbb{U}[-0.5, 0.5]$. The parameter sets $\bar{\omega}_{\text{SIRS}}$ and $\bar{\omega}_{\text{KF}}$ are given by $\bar{\omega}_{\text{SIRS}} = (0.9, 0.1, 0.1)$ and $\bar{\omega}_{\text{KF}} = (0.5, 0.07, 0.08)$, respectively. The kernel parameter is set as 10. The regularization factor $\tilde{\lambda}$ of the GCSGHQF with $\alpha = 2$ and 6 is set as 4×10^{-5} with $\lambda = 10^{-5}$ and $\varsigma = 0.5$, and that for the GCSGHQF with $\alpha = 1$ and 4 is set as $\tilde{\lambda} = 0.4$ with $\lambda = 0.1$ and $\gamma = 0.5$. The value of the RMSE of s_k in uniformly distributed noise is shown in Fig. 4.1(b). From Fig. 4.1(b) and Table 4.1, we can see that the CKF3 has a higher estimation error than other nonlinear Kalman filters. In addition, the GCSGHQF with $\alpha = 6$ can achieve the highest filtering precision.

4.6.2.2 Epidemic Spreading in Laplace Noise

The effectiveness of the proposed GCSGHQF can be demonstrated by measurements corrupted by the Laplace distributed noise. The state noise follows the Gaussian distribution $\mathbb{N}(\mathbf{0}_2, 10^{-4}\mathbf{I}_2)$. Moreover, the measurement noise follows the Laplace distribution $r_k \sim \mathit{Laplace}(u, b)$ with mean $u = 0$ and parameter $b = \delta/\sqrt{2}$ set by the variance $\delta^2 = 0.3$. The parameter sets for the SIRS compartmental model and nonlinear Kalman filters are $\bar{\omega}_{\text{SIRS}} = (0.7, 0.1, 0.1)$ and $\bar{\omega}_{\text{KF}} = (0.5, 0.07, 0.07)$, respectively. The kernel parameter and regularization factor are set as $\sigma = 0.7$ and $\tilde{\lambda} = 0.4$ with $\lambda = 0.1$ and $\varsigma = 0.5$, respectively. Fig. 4.1(c) shows the RMSE of s_k for the Laplace distributed noise. From Fig. 4.1(c) and Table 4.1, the CKF3 cannot achieve the desired filtering performance in comparison with other nonlinear Kalman filters. In addition, the GCSGHQF with $\alpha = 2$, 4 and 6 achieves almost the same filtering precision as the SGHQF. In contrast, the GCSGHQF with $\alpha = 1$ has the lowest estimation error.

4.6.3 Epidemic Spreading on Networks

The effectiveness of the proposed GCSGHQF for epidemic tracking on homogeneous ER, NW and WS networks is assessed for different values of connecting probability p. The number of nodes is 200 and connecting probability p ranges from 0 to 1. When a small connecting probability is adopted, the NW and WS networks reduce to a nearest neighbor coupled network, and very few nodes are connected in the ER network. By contrast, a larger connecting probability makes the NW network reduce to a global neighbor coupled network. In the numerical examples, the connecting probability p is chosen in the interval $p \in [0.05, 0.95]$. In addition, the parameter set $\bar{\omega}_{\text{SIRS}} = (\alpha, \beta, \gamma)$ of the SIRS model is given by $\bar{\omega}_{\text{SIRS}} = (0.7, 0.5, 0.2)$ for evolving epidemic transmission. For nonlinear Kalman filters, the parameter set for performing dynamic estimation $\bar{\omega}_{\text{KF}}$ is chosen to balance the numerical stability and filtering accuracy for the given connecting probability p. The kernel parameter σ and regularization factor $\tilde{\lambda}$ in (4.47) are chosen for achieving the best filtering performance for a given probability p.

4.6.3.1 Epidemic Spreading in ER Network

The process noise and measurement noise variances are set as $\mathbf{Q}_k = 10^{-1}\mathbf{I}_2$ and $R_k = 10^{-1}$. Fig. 4.2(a) plots the RMSE of s_k, i.e., RMSE_s. From Fig. 4.2(a), we can see that the SGHQF has a higher accuracy than the CKF3. The proposed GCSGHQF with $\alpha = 1$ and $\alpha = 2$ achieves a similar filtering accuracy as the SGHQF. In addition, the proposed GCSGHQF with $\alpha = 4$ and 6 achieves a higher estimation precision than the SGHQF and CKF3 when a relatively larger connecting probability p is adopted.

4.6.3.2 Epidemic Spreading in NW Network

Consider the NW network generated by the parameter set $(|N|, K, p)$ with $K = 2$. The covariances of state noise and measurement noise are $\mathbf{Q}_k = 10^{-1}\mathbf{I}_2$ and $R_k = 10^{-1}$. The

plot of the RMSE of s_k is shown in Fig. 4.2(b). Similar to Fig. 4.2(a), the proposed GCSGHQF with $\alpha = 2$ has similar filtering performance as the SGHQF. In addition, Fig. 4.2(b) demonstrates the superior filtering performance of the GCSGHQF with $\alpha = 4$ and 6 in comparison with the SGHQF and CKF3, especially when the connecting probability p is relatively large.

4.6.3.3 Epidemic Spreading in WS Network

The WS network is developed by the parameter set $(|N|, K, p)$ with $K = 2$. The process noise and measurement noise variances are set as $\mathbf{Q}_k = 10^{-5}\mathbf{I}_2$ and $R_k = 0.03$, respectively. The plot of the RMSE_s is shown in Fig. 4.2(c), from which we see that the CKF3 and SGHQF cannot achieve the desired filtering accuracy in terms of the RMSE_s. In addition, the GCSGHQF with $\alpha = 2$ has a similar estimation error as the SGHQF. The GCSGHQF with $\alpha = 1, 4$ and 6 gives a lower RMSE_s than other nonlinear Kalman filters.

4.7 Conclusion

The work described in this chapter aims to track epidemic transmission on networks by a novel generalized correntropy sparse Gauss-Hermite quadrature filter. The epidemic transmission is described by the susceptible-infected-recovered-susceptible compartmental model. The proposed generalized correntropy sparse Gauss- Hermite quadrature filter performs epidemic tracking by taking the susceptible-infected-recovered-susceptible model as a state process. In comparison with traditional Kalman filters based on the minimum mean square error criterion, the generalized correntropy sparse Gauss-Hermite quadrature filter uses the generalized correntropy and is therefore more robust in the presence of non-Gaussian noises. The effectiveness and superior accuracy of the proposed generalized correntropy sparse Gauss-Hermite quadrature filter for tracking epidemic spreading is demonstrated for selected

homogeneous networks.

Chapter 5

Dynamic Estimation of Power Systems by p-Norm Nonlinear Kalman Filter

The problem of dynamic state estimation of power systems is relevant to the monitoring of real-time operation of essential power distribution infrastructure. The nonlinear Kalman filter is utilized for dynamic state estimation of power systems based on the available measurements from phasor measurement units. However, these measurements are corrupted by non-Gaussian noise and exhibit varying levels of sensitivity to outliers, thereby degrading the estimation accuracy. This chapter proposes a robust mixed p-norm square root unscented Kalman filter for state estimation of power systems. Unlike traditional nonlinear Kalman filters which utilize the minimum mean square error criterion, the mixed p-norm square root unscented Kalman filter utilizes a mixed p-norm optimization for weighting the measurement errors to improve robustness against outliers and alleviate the filtering degradation caused by abnormal measurements. The performance of the p-norm square root unscented Kalman filter is demonstrated in the WSCC 3-machine system and the NPCC 48-machine system. Simulation results demonstrate that the p-norm square root unscented Kalman filter achieves excellent accuracy compared to other commonly used nonlinear Kalman filters.

5.1 Introduction

Recently, a considerable amount of effort has been devoted to the state estimation of power systems [24, 18]. The state estimation of power systems is implemented for the state tracking of generators based on the real-time data collected from observable measurements. The phasor measurement units (PMUs) stationed at generator buses can be used for providing measurements of voltage and current phasors [28]. Static and dynamic estimations are then performed to estimate the real-time states of power systems based on the phasor measurements. Static estimation [114, 115] has been applied to system monitoring by providing a set of static states of a power system. However, static states do not contain essential dynamic characteristics of practical systems, especially under fault conditions [28]. In general, the power system operates under normal conditions within a pre-designed set of system parameters, which is appropriate to the ratings of the power system elements. These system parameters may therefore remain at the pre-designed values. In the event of system faults such as three phase faults [28], these system parameters may deviate from the normal values. Therefore, dynamic state estimation [44] is more desirable for systems having high uncertainty and wide parameter variations.

Nonlinear Kalman filters for state estimation [116, 36, 86] include the extended Kalman filter (EKF) [102, 117, 118, 119], unscented Kalman filter (UKF) [103, 43, 120, 62] and cubature Kalman filter(CKF) [46, 48]. Moreover, the square root Kalman filters [30] propagate the square root of error covariance directly, enhancing the numerical stability and improving the filtering accuracy. Common square root Kalman filters include the third-degree square root cubature Kalman filter (SRCKF), fifth-degree square root cubature Kalman filter [45] and square root unscented Kalman filter (SRUKF) [121]. In comparison with third-degree and fifth-degree SRCKF, SRUKF can achieve a balance between filtering accuracy and computational burden, i.e., a higher filtering precision than the third-degree SRCKF and a lower computational

complexity than the fifth-degree SRCKF.

Traditionally, nonlinear Kalman filters are derived from the minimum mean square error (MMSE) criterion, and therefore fail to achieve adequate filtering precision using available measurements that are corrupted by outliers. Moreover, measurements are corrupted by outliers in different ways such as additive and multiplicative manners [122, 123], exhibiting varying effect on filtering precision. In this work, the additive manner for imposing outliers on measurement is considered due to its universality. Recently, numerous robust Kalman filters including the Huber-based unscented Kalman filter (HUKF) [49] and maximum correntropy criterion-based unscented Kalman filter (MCCUKF) [50] are proposed for enhancing the robustness against outliers. The commonly used robust nonlinear Kalman filters indeed exhibit robustness in the presence of outliers. However, these nonlinear Kalman filters still cannot achieve desirable filtering precision because the available measurements may exhibit varying levels of sensitivity to outliers. For example, the MCCUKF based on information theoretical learning (ITL) [51] utilizes a kernel width for scaling all estimation errors. The MCCUKF may therefore achieve undesirable filtering precision to abnormal measurements which exhibit varying levels of sensitivity to outliers. By contrast, the HUKF deals with measurements simply by the limited l_1-norm and l_2-norm of the error, which limits the improvement of filtering accuracy.

The goal of this chapter is to develop a robust nonlinear Kalman filter for the state estimation of the power system against large outliers. To be specific, the contributions include the following:

- A robust mixed p-norm is proposed for suppressing outliers. In comparison with traditional p-norm, the proposed mixed p-norm is characterized by stronger robustness and therefore deals with outliers more efficiently. In addition, the proposed mixed p-norm is deemed as a more generalized p-norm defined over varying intervals, therefore suppressing outliers flexibly. Based on the

appropriate choice of the mixed p-norm, a threshold parameter is utilized for determining the boundary of the abnormal errors. An extra second-order statistical characteristic of the error is utilized for alleviating the issue of the inappropriate choice of the threshold.

- Applying the mixed p-norm in the square root unscented Kalman filter generates a mixed p-norm square root unscented Kalman filter (PSRUKF) for improving the robustness and enhancing the filtering accuracy of the state estimation.

- The proposed PSRUKF is applied to performing the state estimation of power systems for dealing with the measurements corrupted by outliers. Simulation results show the superior robustness of the proposed PSRUKF in performing state estimation of power systems.

5.2 Review of Square Root Unscented Kalman Filters

Consider the discrete-time dynamical system which has a nonlinear state process $\boldsymbol{f}(\cdot)$ and measurement function $\boldsymbol{h}(\cdot)$ [45].

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

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

where $\mathbf{x}_k \in \mathbb{R}^{n_x \times 1}$, $\mathbf{u}_k \in \mathbb{R}^{n_u \times 1}$ and $\mathbf{y}_k \in \mathbb{R}^{n_y \times 1}$ denote the state, input and measurement, respectively. The process noise $\mathbf{w}_{k-1}$ and the measurement noise $\mathbf{v}_k$ obey the Gaussian distribution, i.e., $\mathbf{w}_{k-1} \sim N(\mathbf{0}, \mathbf{Q}_{k-1})$, $\mathbf{v}_{k-1} \sim N(\mathbf{0}, \mathbf{R}_k)$, respectively. A nonlinear Kalman filter is expected to calculate the posterior probability density function of the state, $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ in two steps, i.e., prediction step and update step [106]. The square root unscented Kalman filter propagates the square root of error covariance directly for avoiding refactorizing and improving numerical stability at each time.

Algorithm 6 Square Root Unscented Kalman Filter

Time update

1: Calculate predicted mean

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=0}^{2n_x} \omega_i^m \boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}) \tag{5.3}$$

with $\{\boldsymbol{\xi}_{i,k-1}\}_{i=0}^{2n_x}$ and $\{\omega_i^m\}_{i=0}^{2n_x}$ given by (5.11) and (5.12).

2: Calculate the predicted square root error covariance

$$\left[\hat{\mathbf{Q}}_k^p,\ \hat{\mathbf{L}}_k^p\right] \leftarrow \mathrm{qr}\left(\left[\boldsymbol{\Upsilon}_k^p \quad \sqrt{\mathbf{Q}_{k-1}}\right]^{\mathrm{T}}\right) \tag{5.4}$$

$$\tilde{\mathbf{L}}_k^p \leftarrow \mathrm{cholupdate}\left(\bar{\mathbf{L}}_k^p,\ \sqrt{|w_0^c|}\bar{\boldsymbol{f}}(\boldsymbol{\xi}_{0,k-1}),'\ s'\right) \tag{5.5}$$

where $s = \mathrm{sign}(w_0^c)$, $\bar{\mathbf{L}}_k^p$ is the upper triangular matrix of $\hat{\mathbf{L}}_k^p$ and $\boldsymbol{\Upsilon}_k^p = \sqrt{w_1^c}(\boldsymbol{f}(\boldsymbol{\xi}_{1:2n_x,k-1}) - \hat{\mathbf{x}}_{k|k-1})$, $\bar{\boldsymbol{f}}(\boldsymbol{\xi}_{0,k-1}) = \boldsymbol{f}(\boldsymbol{\xi}_{0,k-1}) - \hat{\mathbf{x}}_{k|k-1}$. Then, $\mathbf{S}_{\mathbf{x}}$ in (5.11) is substituted by the predicted square root error covariance $\tilde{\mathbf{L}}_k^s = (\tilde{\mathbf{L}}_k^p)^{\mathrm{T}}$, giving samples in measurement update.

Measurement update

1: Obtain predicted mean $\hat{\mathbf{y}}_{k|k-1} = \sum_{i=0}^{2n_x} \omega_i^m \boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1})$ with $\{\boldsymbol{\xi}_{i,k|k-1}\}_{i=0}^{2n_x}$ and $\{\omega_i^m\}_{i=0}^{2n_x}$ given by (5.11) and (5.12).

2: Find the measurement error $\tilde{\mathbf{y}}_{k|k-1} = \mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}$.

3: Obtain the covariance matrix $\tilde{\mathbf{L}}_k^m$:

$$\left[\tilde{\mathbf{Q}}_k^m,\ \hat{\mathbf{L}}_k^m\right] \leftarrow \mathrm{qr}\left(\left[\bar{\boldsymbol{\Upsilon}}_k^m \quad \sqrt{\mathbf{R}_k}\right]^{\mathrm{T}}\right) \tag{5.6}$$

$$\tilde{\mathbf{L}}_k^m \leftarrow \mathrm{cholupdate}\left(\bar{\mathbf{L}}_k^m,\ \sqrt{|w_0^c|}\bar{\boldsymbol{h}}(\boldsymbol{\xi}_{0,k-1}),'\ s'\right) \tag{5.7}$$

where $\bar{\mathbf{L}}_k^m$ in (5.7) is the upper triangular matrix of $\hat{\mathbf{L}}_k^m$, $\bar{\boldsymbol{\Upsilon}}_k^m = \sqrt{w_1^c}(\boldsymbol{h}(\boldsymbol{\xi}_{1:2n_x,k-1}) - \hat{\mathbf{y}}_{k|k-1})$, and $\bar{\boldsymbol{h}}(\boldsymbol{\xi}_{0,k-1}) = \boldsymbol{h}(\boldsymbol{\xi}_{0,k-1}) - \hat{\mathbf{y}}_{k|k-1}$.

4: Compute the cross covariance: $\mathbf{P}_{xy,k|k-1} = \sum_{i=0}^{2n_x} \omega_i^c \bar{\boldsymbol{\xi}}_{i,k-1} \left(\bar{\boldsymbol{h}}(\boldsymbol{\xi}_{i,k-1})\right)^{\mathrm{T}}$ with $\bar{\boldsymbol{\xi}}_{i,k|k-1} = \boldsymbol{\xi}_{i,k|k-1} - \hat{\mathbf{x}}_{k|k-1}$, and $\bar{\boldsymbol{h}}(\boldsymbol{\xi}_{i,k|k-1}) = \boldsymbol{h}(\boldsymbol{\xi}_{i,k|k-1}) - \hat{\mathbf{y}}_{k|k-1}$.

5: Calculate the Kalman gain:

$$\mathbf{K}_k = \mathbf{P}_{xy,k|k-1} \left[\tilde{\mathbf{L}}_k^m\right]^{-1} \left[(\tilde{\mathbf{L}}_k^m)^{\mathrm{T}}\right]^{-1}. \tag{5.8}$$

6: Compute the estimated mean and updated square root of error covariance:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k|k-1})) \tag{5.9}$$

$$\mathbf{L}_k = \mathrm{cholupdate}\left(\tilde{\mathbf{L}}_k^p, \mathbf{K}_k\tilde{\mathbf{L}}_k^o, \text{`}-\text{'}\right), \tag{5.10}$$

where $\tilde{\mathbf{L}}_k^o = (\tilde{\mathbf{L}}_k^m)^{\mathrm{T}}$, and let updated square root of error covariance $\tilde{\mathbf{L}}_k^u = \mathbf{L}_k^{\mathrm{T}}$ substitute $\mathbf{S}_{\mathbf{x}}$ in (5.11) for obtaining the sample points at the time step.

The square root unscented Kalman filter utilizes sampling points $\{\boldsymbol{\xi}_i\}_{i=0}^{2n_x}$ and corresponding weights $\{\omega_i^m, \omega_i^c\}_{i=0}^{2n_x}$ for calculating the posterior probability density function of the state, $p(\mathbf{x}_k|\mathbf{y}_{1:k})$. The sampling points $\{\boldsymbol{\xi}_i\}_{i=0}^{2n_x}$ are defined by

$$\begin{cases} \boldsymbol{\xi}_0 = \bar{\mathbf{x}} \\ \boldsymbol{\xi}_i = \bar{\mathbf{x}} + \left(\sqrt{(n_x+\lambda)}\mathbf{S}_{\mathbf{x}}\right)_i, & i = 1, 2, \cdots, n_x \\ \boldsymbol{\xi}_i = \bar{\mathbf{x}} - \left(\sqrt{(n_x+\lambda)}\mathbf{S}_{\mathbf{x}}\right)_{i-n_x}, & i = n_x+1, \cdots, 2n_x \end{cases} \tag{5.11}$$

where $\mathbf{S}_{\mathbf{x}}$ is the square root of the error covariance $\mathbf{P}_{\mathbf{x}}$ and $\left(\sqrt{(n_x+\lambda)}\mathbf{S}_{\mathbf{x}}\right)_i$ denotes the i-th column vector of the matrix $\left(\sqrt{(n_x+\lambda)}\mathbf{S}_{\mathbf{x}}\right)$. Also, ω_i^m and ω_i^c denote the weights for calculating the mean and covariance, respectively, i.e.,

$$\begin{cases} w_0^m = \frac{\lambda}{n_x+\lambda} \\ w_0^c = \frac{\lambda}{n_x+\lambda} + 1 - \alpha^2 + \beta \\ w_i^m = w_i^c = \frac{1}{2(n_x+\lambda)}, \ i = 1, 2, \cdots, 2n_x. \end{cases} \tag{5.12}$$

Parameter λ is defined as $\lambda = \alpha^2(n_x+\kappa) - n_x$ where α determines the spreading range of sampling points at point $\bar{\mathbf{x}}$ and κ is a secondary scaling parameter. Parameter β in (5.12) is utilized for incorporating the prior information of the distribution of $\bar{\mathbf{x}}$. The update procedures of the square root unscented Kalman filter (SRUKF) [121] are summarized in **Algorithm** 6.

Remark 9 *The unscented Kalman filter utilizes the unscented transformation for capturing the statistical characteristic of the variable through a nonlinear function. However, the UKF may make the error covariance non-positive definite since the first sampling point in (5.11) and (5.12) carries a large negative sampling weight for calculating the error covariance. For example, it has been proved that the UKF can capture a fourth-order statistical characteristic of the error by setting* $n_x + \lambda = 3$. *However, the instability becomes more serious as the dimension increases due to the negative* $\lambda = 3 - n_x$ *under* $n_x > 3$. *Therefore, an inappropriate choice of parameters*

may cause instability of the UKF. By contrast, SRUKF performs state estimation by propagating the square root of the error covariance directly for avoiding this issue.

5.3 Robust Mixed p-Norm Square Root Unscented Kalman Filter

The traditional square root unscented Kalman filter indeed achieves a desirable filtering accuracy under a Gaussian noise environment but exhibits weak robustness against outliers. The proposed p-norm square root unscented Kalman filter introduces a mixed p-norm optimization for weighting measurement errors to alleviate the filtering degradation in the presence of outliers.

Consider a linear measurement function in (5.2), i.e.,

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \tag{5.13}$$

where matrix $\mathbf{H}_k \in \mathbb{R}^{n_y \times n_x}$ represents a linear measurement matrix. Inspired by MCCUKF [50] and HUKF [49], the following linear regression model can be constructed for performing the p-norm optimization, i.e.,

$$\begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_k \end{bmatrix} \mathbf{x}_k + \bar{\mathbf{v}}_k \tag{5.14}$$

with the identity matrix $\mathbf{I} \in \mathbb{R}^{n_x \times n_x}$. The last term $\bar{\mathbf{v}}_k$ in (5.14) is given by

$$\bar{\mathbf{v}}_k = \begin{bmatrix} -(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \\ \mathbf{v}_k \end{bmatrix} \tag{5.15}$$

which is associated with covariance matrix $E[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^{\mathrm{T}}]$, i.e.,

$$E[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^{\mathrm{T}}] = \begin{bmatrix} \mathbf{P}_{k|k-1} & 0 \\ 0 & \mathbf{R}_k \end{bmatrix} \tag{5.16}$$

$$= \begin{bmatrix} \tilde{\mathbf{P}}_{w,k|k-1}\tilde{\mathbf{P}}_{w,k|k-1}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{P}}_{v,k}\tilde{\mathbf{P}}_{v,k}^{\mathrm{T}} \end{bmatrix}. \tag{5.17}$$

According to (5.17), $E[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^{\mathrm{T}}]$ can be rewritten as

$$E[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^{\mathrm{T}}] = \tilde{\mathbf{P}}_k\tilde{\mathbf{P}}_k^{\mathrm{T}} \tag{5.18}$$

with the matrix $\tilde{\mathbf{P}}_k$

$$\tilde{\mathbf{P}}_k = \begin{bmatrix} \tilde{\mathbf{P}}_{w,k|k-1} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{P}}_{v,k} \end{bmatrix} \tag{5.19}$$

where $\tilde{\mathbf{P}}_k$ can be obtained by Cholesky decomposition of $E[\bar{\mathbf{v}}_k\bar{\mathbf{v}}_k^{\mathrm{T}}]$. Multiplying both sides of (5.14) by $\tilde{\mathbf{P}}_k^{-1}$, we get

$$\mathbf{\Gamma}_k = \mathbf{\Pi}_k\mathbf{x}_k + \mathbf{e}_k \tag{5.20}$$

where

$$\mathbf{\Gamma}_k = \tilde{\mathbf{P}}_k^{-1}\begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k \end{bmatrix}, \mathbf{\Pi}_k = \tilde{\mathbf{P}}_k^{-1}\begin{bmatrix} \mathbf{I} \\ \mathbf{H}_k \end{bmatrix}. \tag{5.21}$$

The error $\mathbf{e}_k$ in (5.20) is given by

$$\mathbf{e}_k = \tilde{\mathbf{P}}_k^{-1}\bar{\mathbf{v}}_k. \tag{5.22}$$

The error $\mathbf{e}_k$ in (5.20) is white due to $E[\mathbf{e}_k\mathbf{e}_k^{\mathrm{T}}] = \mathbf{I}$.

The p-norm square root unscented Kalman filter aims to minimize

$$J(\mathbf{x}_k) = \sum_{i=1}^{n_x+n_y} \frac{\lambda_i}{p_i} J_i(e_k(i)) + \frac{1-\lambda_i}{2} e_k^2(i) \tag{5.23}$$

where factors $\{\lambda_i\}_{i=1}^{n_x+n_y}$ are incorporated into (5.23) for enhancing numerical stability. Here, $e_k(i)$ denotes the i-th element of $\mathbf{e}_k$ at discrete time k. Also, $e_k(i)$ is defined by $e_k(i) = \Gamma_k(i) - \mathbf{\Pi}_k^r(i)\mathbf{x}_k$ where $\Gamma_k(i)$ is the i-th element of $\mathbf{\Gamma}_k \in \mathbb{R}^{(n_x+n_y)\times 1}$ and $\mathbf{\Pi}_k^r(i)$ is the i-th row of $\mathbf{\Pi}_k$ at discrete time k in (5.21). Since measurements exhibit varying levels of sensitivity to outliers, it is more reasonable and flexible to optimize (5.23) in a mixed p-norm. The $J_i(e_k(i)) = |e_k(i)|^{p_i}$ in (5.23) is defined by a mixed p-norm of the error and chosen by

$$J_i(e_k(i)) = \begin{cases} |e_k(i)|^{\tilde{p}_1}, & |e_k(i)| < \vartheta \\ |e_k(i)|^{\tilde{p}_2}, & \vartheta \le |e_k(i)| < 1 \\ |e_k(i)|^{\tilde{p}_3}, & |e_k(i)| \ge \vartheta \text{ and } |e_k(i)| \ge 1 \end{cases} \tag{5.24}$$

with a positive threshold ϑ and $\tilde{p}_1 = 2$. The newly defined p-norm satisfies the properties of a norm [124]. Parameters $\tilde{p}_2$ and $\tilde{p}_3$ are positive numbers larger and smaller than 2, respectively. As given in (5.24), the p_i-order term of the $J(\mathbf{x}_k)$, i.e., $J_i(e_k(i))$, is a piece-wise function of the state estimation error $|e_k(i)|$. Then, (5.24) actually reduces to a piece-wise function defined over two intervals by setting $\vartheta \ge 1$, i.e.,

$$J_i(e_k(i)) = \begin{cases} |e_k(i)|^{\tilde{p}_1}, & |e_k(i)| < \vartheta \\ |e_k(i)|^{\tilde{p}_3}, & |e_k(i)| \ge \vartheta \end{cases} \tag{5.25}$$

where parameters $\tilde{p}_1$ and $\tilde{p}_3$ are the same as those given in (5.24). Equation (5.23) can

be expanded as

$$J(\mathbf{x}_k) = \sum_{i=1}^{n_x+n_y} \frac{\lambda_i}{p_i} |\Gamma_k(i) - \mathbf{\Pi}_k^r(i)\mathbf{x}_k|^{p_i} + \frac{(1-\lambda_i)}{2} \left|\Gamma_k(i) - \mathbf{\Pi}_k^r(i)\mathbf{x}_k\right|^2. \tag{5.26}$$

Therefore, the optimal estimate of $\mathbf{x}_k$ can be obtained by

$$\hat{\mathbf{x}}_k = \arg \min_{\hat{\mathbf{x}}_k} J(\mathbf{x}_k). \tag{5.27}$$

Taking the partial derivative of (5.26) with respect to $\mathbf{x}_k$, we obtain

$$\begin{aligned} \left.\frac{\partial J(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k=\hat{\mathbf{x}}_k} &= \sum_{i=1}^{n_x+n_y} \lambda_i |e_k(i)|^{p_i-2} e_k(i) (\mathbf{\Pi}_k^r(i))^{\mathrm{T}} + (1-\lambda_i) e_k(i) (\mathbf{\Pi}_k^r(i))^{\mathrm{T}} \,|_{\mathbf{x}_k=\hat{\mathbf{x}}_k} \\ &= 0. \end{aligned} \tag{5.28}$$

After simplifying, the optimal state estimation can be derived as follows:

$$\begin{aligned} \hat{\mathbf{x}}_k = &\left(\sum_{i=1}^{n_x+n_y} \left(\lambda_i |e_k(i)|^{p_i-2} + (1-\lambda_i) \right) (\mathbf{\Pi}_k^r(i))^{\mathrm{T}} \mathbf{\Pi}_k^r(i) \right)^{-1} \\ &\left(\sum_{i=1}^{n_x+n_y} \left(\lambda_i |e_k(i)|^{p_i-2} + (1-\lambda_i) \right) \right) (\mathbf{\Pi}_k^r(i))^{\mathrm{T}} \Gamma_k(i) \Bigg). \end{aligned} \tag{5.29}$$

Define a weighted matrix $\mathbf{C}_k \in \mathbb{R}^{(n_x+n_y)\times(n_x+n_y)}$ by

$$\mathbf{C}_k = \begin{bmatrix} \mathbf{C}_{x,k} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{y,k} \end{bmatrix} \tag{5.30}$$

where matrices $\mathbf{C}_{x,k} \in \mathbb{R}^{n_x \times n_x}$ and $\mathbf{C}_{y,k} \in \mathbb{R}^{n_y \times n_y}$ are expressed as

$$\mathbf{C}_{x,k} = diag(\lambda_1 |e_k(1)|^{p_1-2} + \tilde{\lambda}_1, \cdots, \lambda_{n_x} |e_k(n_x)|^{p_{n_x}-2} + \tilde{\lambda}_{n_x}), \tag{5.31}$$

$$\begin{aligned} \mathbf{C}_{y,k} = &diag((\lambda_{n_x+1} |e_k(n_x+1)|^{p_{n_x+1}-2}) + \tilde{\lambda}_{n_x+1}, \\ &\cdots, \lambda_{n_x+n_y} \left| e_k\left(n_x+n_y\right) \right|^{p_{n_x+n_y}-2} + \tilde{\lambda}_{n_x+n_y}) \end{aligned} \tag{5.32}$$

where $\tilde{\lambda}_i = (1 - \lambda_i)$. The $\mathbf{C}_{x,k}$ in (5.31) is a diagonal matrix with entries $\{\lambda_j |e_k(j)|^{p_j - 2} + \tilde{\lambda}_j\}_{j=1}^{n_x}$. The diagonal matrix $\mathbf{C}_{y,k}$ in (5.32) has entries $\{\lambda_j |e_k(j)|^{p_j - 2} + \tilde{\lambda}_j\}_{j=n_x+1}^{j=n_x+n_y}$. Equation (5.29) can thus be rewritten as

$$\hat{\mathbf{x}}_k = \mathbf{\Pi}_{C,k}^{-1} \mathbf{\Pi}_{D,k} \tag{5.33}$$

with $\mathbf{\Pi}_{C,k} = \mathbf{\Pi}_k^{\mathrm{T}} \mathbf{C}_k \mathbf{\Pi}_k$ and $\mathbf{\Pi}_{D,k} = \mathbf{\Pi}_k^{\mathrm{T}} \mathbf{C}_k \mathbf{\Gamma}_k$. According to (5.17) and (5.21), matrix $\mathbf{\Pi}_{C,k}^{-1}$ can be rewritten by

$$\mathbf{\Pi}_{C,k}^{-1} = (\mathbf{H}_k^{\mathrm{T}} (\tilde{\mathbf{P}}_{v,k}^{-1})^{\mathrm{T}} \mathbf{C}_{y,k} \tilde{\mathbf{P}}_{v,k}^{-1} \mathbf{H}_k + (\tilde{\mathbf{P}}_{w,k|k-1}^{-1})^{\mathrm{T}} \mathbf{C}_{x,k} \tilde{\mathbf{P}}_{w,k|k-1}^{-1})^{-1}. \tag{5.34}$$

By using the matrix inverse lemma [79], i.e.,

$$\left(\bar{\mathbf{A}} + \bar{\mathbf{B}}\bar{\mathbf{C}}\bar{\mathbf{D}}\right)^{-1} = \bar{\mathbf{A}}^{-1} - \bar{\mathbf{A}}^{-1}\bar{\mathbf{B}}\left(\bar{\mathbf{C}}^{-1} + \bar{\mathbf{D}}\bar{\mathbf{A}}^{-1}\bar{\mathbf{B}}\right)^{-1}\bar{\mathbf{D}}\bar{\mathbf{A}}^{-1}$$

and letting $\bar{\mathbf{A}} = \left(\tilde{\mathbf{P}}_{w,k|k-1}^{-1}\right)^{\mathrm{T}} \mathbf{C}_{x,k} \tilde{\mathbf{P}}_{w,k|k-1}^{-1}$, $\bar{\mathbf{B}} = \mathbf{H}_k^{\mathrm{T}}$, $\bar{\mathbf{C}} = \left(\tilde{\mathbf{P}}_{v,k}^{-1}\right)^{\mathrm{T}} \mathbf{C}_{y,k} \tilde{\mathbf{P}}_{v,k}^{-1}$ and $\bar{\mathbf{D}} = \mathbf{H}_k$, (5.34) can be further expanded as

$$\begin{aligned} \mathbf{\Pi}_{C,k}^{-1} &= \tilde{\mathbf{P}}_{w,k|k-1} \mathbf{C}_{x,k}^{-1} \tilde{\mathbf{P}}_{w,k|k-1}^{\mathrm{T}} - \tilde{\mathbf{P}}_{w,k|k-1} \mathbf{C}_{x,k}^{-1} \tilde{\mathbf{P}}_{w,k|k-1}^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} \\ &(\tilde{\mathbf{P}}_{v,k} \mathbf{C}_{y,k}^{-1} \tilde{\mathbf{P}}_{v,k}^{\mathrm{T}} + \mathbf{H}_k \tilde{\mathbf{P}}_{w,k|k-1} \mathbf{C}_{x,k}^{-1} \tilde{\mathbf{P}}_{w,k|k-1}^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}})^{-1} \mathbf{H}_k \tilde{\mathbf{P}}_{w,k|k-1} \mathbf{C}_{x,k}^{-1} \tilde{\mathbf{P}}_{w,k|k-1}^{\mathrm{T}}. \end{aligned} \tag{5.35}$$

Similar to (5.34), $\mathbf{\Pi}_{D,k}$ can be expanded to

$$\mathbf{\Pi}_k^{\mathrm{T}} \mathbf{C}_k \mathbf{\Gamma}_k = \mathbf{H}_k^{\mathrm{T}} \left(\tilde{\mathbf{P}}_{v,k}^{-1}\right)^{\mathrm{T}} \mathbf{C}_{y,k} \tilde{\mathbf{P}}_{v,k}^{-1} \mathbf{y}_k + \left(\tilde{\mathbf{P}}_{w,k|k-1}^{-1}\right)^{\mathrm{T}} \mathbf{C}_{x,k} \tilde{\mathbf{P}}_{w,k|k-1}^{-1} \hat{\mathbf{x}}_{k|k-1}. \tag{5.36}$$

From (5.34) to (5.36), equation (5.33) can be given in a recursive form, i.e.,

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \bar{\mathbf{K}}_k (\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \tag{5.37}$$

where $\bar{\mathbf{K}}_k$ takes the form of

$$\bar{\mathbf{K}}_k = \bar{\mathbf{P}}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}_k\bar{\mathbf{P}}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} + \bar{\mathbf{R}}_k\right)^{-1}. \tag{5.38}$$

The prediction error covariance matrices $\bar{\mathbf{P}}_{k|k-1}$ and $\bar{\mathbf{R}}_k$ are expressed as

$$\bar{\mathbf{P}}_{k|k-1} = \tilde{\mathbf{P}}_{w,k|k-1}\mathbf{C}_{x,k}^{-1}\tilde{\mathbf{P}}_{w,k|k-1}^{\mathrm{T}} \tag{5.39}$$

$$\bar{\mathbf{R}}_k = \tilde{\mathbf{P}}_{v,k}\mathbf{C}_{y,k}^{-1}\tilde{\mathbf{P}}_{v,k}^{\mathrm{T}}. \tag{5.40}$$

Also, matrix $\mathbf{C}_k$ is utilized for weighting the prediction covariance matrix $\mathbf{P}_{k|k-1}$ and covariance matrix of measurement noise $\mathbf{R}_k$.

Consider a nonlinear measurement function in (5.2). The nonlinear regression model can be constructed by

$$\begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{h}(\mathbf{x}_k) \end{bmatrix} + \begin{bmatrix} \delta\mathbf{x}_k \\ \mathbf{v}_k \end{bmatrix} \tag{5.41}$$

with $\delta\mathbf{x}_k = \hat{\mathbf{x}}_{k|k-1} - \mathbf{x}_k$. Define $\tilde{\boldsymbol{\Gamma}}_k$, $\breve{\boldsymbol{\Pi}}_k$ and $\breve{\mathbf{e}}_k$ as

$$\tilde{\boldsymbol{\Gamma}}_k = \breve{\mathbf{P}}_k^{-1}\begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{y}_k \end{bmatrix}, \breve{\boldsymbol{\Pi}}_k = \breve{\mathbf{P}}_k^{-1}\begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{h}(\mathbf{x}_k) \end{bmatrix}, \breve{\mathbf{e}}_k = \breve{\mathbf{P}}_k^{-1}\begin{bmatrix} \delta\mathbf{x}_k \\ \mathbf{v}_k \end{bmatrix} \tag{5.42}$$

where $\breve{\mathbf{P}}_k$ can be obtained by applying Cholesky decomposition as

$$\mathbf{P}_{\mathrm{R},k} = \begin{bmatrix} \mathbf{P}_{k|k-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_k \end{bmatrix}. \tag{5.43}$$

Therefore, we have

$$\tilde{\boldsymbol{\Gamma}}_k = \breve{\boldsymbol{\Pi}}_k + \breve{\mathbf{e}}_k. \tag{5.44}$$

Similar to (5.23), the nonlinear p-norm square root unscented Kalman filter aims to minimize

$$\check{J}(\mathbf{x}_k) = \sum_{i=1}^{n_x+n_y} \frac{\lambda_i}{p_i}|\check{e}_k(i)|^{p_i} + \frac{(1-\lambda_i)}{2}|\check{e}_k(i)|^2. \tag{5.45}$$

Define $\phi(\check{e}_k(i)) = \partial\check{J}(\mathbf{x}_k)/\partial\mathbf{x}_k$ and minimize (5.45) to yield

$$\frac{\partial\check{J}(\mathbf{x}_k)}{\partial\mathbf{x}_k}\,|_{\mathbf{x}_k=\hat{\mathbf{x}}_k} = \sum_{i=1}^{n_x+n_y} \phi(\check{e}_k(i))\frac{\partial\check{e}_k(i)}{\partial\mathbf{x}_k}\,|_{\mathbf{x}_k=\hat{\mathbf{x}}_k} \tag{5.46}$$

where $\phi(\check{e}_k(i)) = \lambda_i|\check{e}_k(i)|^{p_i-2}\check{e}_k(i) + (1-\lambda_i)\check{e}_k(i)$. Also, define $\psi(\check{e}_k(i)) = \phi(\check{e}_k(i))/\check{e}_k(i)$, i.e.,

$$\psi(\check{e}_k(i)) = \lambda_i|\check{e}_k(i)|^{p_i-2} + (1-\lambda_i) \tag{5.47}$$

and $\tilde{\mathbf{S}}_{k+1}$ as

$$\tilde{\mathbf{S}}_{k+1} = \check{\mathbf{P}}_k\check{\mathbf{\Gamma}}_k^{-1}\check{\mathbf{P}}_k^{\mathrm{T}} \tag{5.48}$$

where $\check{\mathbf{\Gamma}}_k = diag[\psi(\check{e}_k(i))]$. Since $\delta\mathbf{x}_k$ is set to zero, the prediction error of state remains unchanged, i.e.,

$$\tilde{\mathbf{S}}_{k+1}(1:n_x, 1:n_x) = \mathbf{P}_{k|k-1}. \tag{5.49}$$

Denote $\tilde{\mathbf{R}}_{k+1}$ as the updated measurement covariance and

$$\tilde{\mathbf{R}}_k = \tilde{\mathbf{S}}_{k+1}(n_x+1:n_x+n_y, n_x+1:n_x+n_y). \tag{5.50}$$

Replacing $\mathbf{R}_k$ with $\tilde{\mathbf{R}}_k$ in (5.6) in **Algorithm** 6 and performing square root UKF measurement update will give the measurement update process for PSRUKF.

Remark 10 *Table 5.2 shows the computational complexity of PSRUKF in estimating the hidden state* $\mathbf{x}_k$ *using available measurements* $\mathbf{y}_{1:k}$ *at discrete time k with* $\mathbf{x}_k \in \mathbb{R}^{n_x \times 1}$, $\mathbf{y}_k \in \mathbb{R}^{n_y \times 1}$. *From Table 5.2, PSRUKF introduces an extra computational burden on calculating matrix* $\tilde{\mathbf{R}}_k$ *at discrete time k in comparison to SRUKF. Since matrix* $\tilde{\mathbf{R}}_k \in \mathbb{R}^{n_y \times n_y}$ *at discrete time k needs to be updated, we only focus on matrices* $\breve{\mathbf{P}}_k(n_x{+}1 : n_x{+}n_y, n_x{+}1 : n_x{+}n_y) \in \mathbb{R}^{n_y \times n_y}$ *and* $\breve{\mathbf{\Gamma}}_k^{-1}(n_x{+}1 : n_x{+}n_y, n_x{+}1 : n_x{+}n_y) \in \mathbb{R}^{n_y \times n_y}$ *in calculating matrix* $\tilde{\mathbf{R}}_k$. *The additional computational complexity of updating matrix* $\tilde{\mathbf{R}}_k$ *is almost the same as the burden of performing the Cholesky factorization in SRUKF, which is acceptable for performing state estimation. Moreover, PSRUKF has a different computational burden in calculating the weighted matrix* $\breve{\mathbf{\Gamma}}_k^{-1}$ *in comparison with the maximum correntropy criterion unscented Kalman filter (MCCUKF) or MCCSRUKF. Different from the diagonal entries of* $\breve{\mathbf{\Gamma}}_k^{-1}$ *in MCCUKF or MCCSRUKF, the diagonal elements of* $\breve{\mathbf{\Gamma}}_k^{-1}$ *in PSRUKF and HSRUKF all consist of the p-norm of the error, which has almost the same burden. In contrast, MCCUKF or MCCSRUKF has a relatively large computational burden due to the introduction of diagonal entries constructed by the exponent function of the error.*

Remark 11 *Robust nonlinear Kalman filters in Chapter4 and Chapter5 have no advantage in comparison with traditional filters in Chapter3 at Gaussian noise and are only considered at non-Gaussian noises. From aspect of filtering precision, the generalized correntropy sparse Gauss-Hermite quadrature filter in Chapter4 has high accuracy but not well numerical stability due to the implementation of Cholesky decomposition. To increase the covariance of state noise is helpful for alleviating this issue but may degrade filtering precision. As for calculational complexity, the generalized correntropy sparse Gauss-Hermite quadrature filter in Chapter4 has higher complexity than the mixed p-norm square root unscented Kalman filter in Chapter5. As for the number of parameters, the generalized correntropy sparse Gauss-Hermite quadrature filter has fewer parameters including kernel parameter* σ,

balanced parameter λ and exponential term α than the mixed p-norm square root unscented Kalman filter including parameters p_1, p_2, p_3, balanced parameter λ and threshold factor ϑ. As a result, the generalized correntropy sparse Gauss-Hermite quadrature filter is chosen if high accuracy or few adjustable parameters are required. If high efficiency or numerical stability is in requirement, the mixed p-norm square root unscented Kalman filter is a preferable choice.

Remark 12 *In the proposed method, abnormal observations may be caused by system faults or non-Gaussian noises from natural environment or measuring device like the Phasor measurement unit (PMU). In our work, the proposed method suppresses abnormal measurement errors by the proposed mixed p-norm without making any distinction among these causes. The abnormal measurements with different causes all pose negative effect on the update step of nonlinear Kalman filters. By adopting appropriate parameters of the proposed mixed p-norm, a weakened updated step can be used for alleviating the effect of abnormal errors due to varying different causes. It is therefore not necessary to make any distinction among these causes.*

5.4 Choice of Parameters and Robustness of Mixed p-Norm Kalman Filter

The proposed mixed p-norm square root unscented Kalman filter (PSRUKF) utilizes a mixed p-norm for weighting varying levels of the error for improving the robustness against outliers. Apart from the mixed p-norm parameters including $\tilde{p}_1$, $\tilde{p}_2$ and $\tilde{p}_3$ in (5.24), robustness or stability is also closely related to other parameters including the threshold parameter ϑ in (5.24) and (5.25) and the balanced parameter λ_i in (5.45). In this section, the robustness of PSRUKF is demonstrated in respect of measurement update in the process of state estimation. In addition, the choice of these parameters is discussed for improving robustness and filtering precision in comparison

with traditional nonlinear Kalman filters.

5.4.1 Robustness of Mixed p-Norm Square Root Unscented Kalman Filter

The PSRUKF improves the stability of the state estimation by alleviating the effect of the measurement update when measurements are corrupted by outliers. Based on the choice of the p-norm in (5.24) and (5.25), PSRUKF aims to design a relatively small value of $\psi(\check{e}_k(i))$ in (5.47) in the presence of outliers. Due to varying levels of sensitivity of measurements to outliers, the corresponding estimation errors exhibit varying orders of magnitude. According to the magnitudes of the errors $\vartheta \leq |\check{e}_k(i)| < 1$ and $|\check{e}_k(i)| \geq 1$, $|\check{e}_k(i)| \geq \vartheta$ in (5.24) and (5.25), PSRUKF utilizes a mixed p-norm as $\tilde{p}_2 > 2$ for $\vartheta \leq |\check{e}_k(i)| < 1$ and $\tilde{p}_3 < 2$ for $|\check{e}_k(i)| \geq 1$ and $|\check{e}_k(i)| \geq \vartheta$ so that a relatively small value of $\psi(\check{e}_k(i))$ in the presence of outliers can be obtained.

Consider the measurement $\mathbf{y}_k = [y_k(1), y_k(2), \cdots, y_k(n_y)]$ at discrete time k and the j-th measurement $y_k(j)$ being corrupted by outliers. Since the diagonal matrix $\tilde{\mathbf{R}}_k$ in (5.50) is constructed by a function of matrix $\check{\mathbf{\Gamma}}_k^{-1}$ in (5.48) which consists of $\psi(\check{e}_k(i))$, the j-th diagonal element of $\tilde{\mathbf{R}}_k$ in (5.50) is large on the basis of the choice of the mixed p-norm. From matrix factorization of $\tilde{\mathbf{R}}_k$ in (5.6) and (5.7), the factorization matrix $\tilde{\mathbf{S}}_k^-$ in (5.6) and (5.7) is found to be crucial to suppress outliers. The factorization matrix $\tilde{\mathbf{S}}_k^-$ has a large absolute value of the j-th diagonal element. Therefore, the absolute values of the j-column elements of the gain matrix $\mathbf{K}_k$ are small since the Kalman gain is constructed by a function of the inverse of matrix $\tilde{\mathbf{S}}_k^-$. The PSRUKF utilizes the measurement update for further correcting the predictive state in the form of an update increment, which is constructed by multiplication of the Kalman gain $\mathbf{K}_k$ and measurement errors in (5.9).

Therefore, the proposed PSRUKF estimates the hidden state by a time update and a weakened measurement update, resulting in improved robustness against outliers.

The simulation results given in Section 5.6 are used to demonstrate the robustness of PSRUKF.

5.4.2 Choice of Mixed p-Norm Parameters

The choice of the mixed p-norm parameters is essential for improving the robustness against outliers and filtering precision. Since a second-order statistical characteristic of the error is suitable for dealing with normal measurements, $\tilde{p}_1 = 2$ in (5.24) and (5.45) is utilized for weighting the estimation error under the condition of $|\check{e}_k(i)| < \vartheta$. By contrast, mixed p-norm parameters $\tilde{p}_2 > 2$ and $\tilde{p}_3 < 2$ in (5.24) and (5.45) are generally adopted for dealing with abnormal errors corresponding to the measurements corrupted by outliers. In comparison with parameter $\tilde{p}_2$ utilized for suppressing the abnormal error $\vartheta \leq |\check{e}_k(i)| < 1$, parameter $\tilde{p}_3$ is much more effective in dealing with large error $|\check{e}_k(i)| \geq \vartheta$ and $|\check{e}_k(i)| \geq 1$.

The PSRUKF resembles SRUKF when $\tilde{p}_2 = 2$ and $\tilde{p}_3 = 2$, which exhibits poor robustness against outliers. Therefore, PSRUKF may not achieve the desirable filtering precision when $\tilde{p}_2$ or $\tilde{p}_3$ approaches 2. In PSRUKF, a small $\tilde{p}_2 > 2$ cannot reduce $\psi(\check{e}_k(i))$ in (5.47) efficiently and therefore may not be beneficial for improving the precision. By contrast, a large $\tilde{p}_2 > 2$ may make $\psi(\check{e}_k(i))$ in (5.47) close to zero, which is not beneficial for the inverse operation of matrix $\check{\mathbf{\Gamma}}_k$ in (5.48). The second-order statistical characteristic of the error in (5.23) and (5.45) can solve this problem. In fact, parameter $\tilde{p}_2 > 2$ (not close to 2) is helpful for PSRUKF to generate a higher estimation precision than traditional nonlinear Kalman filters.

The PSRUKF indeed exhibits a stronger robustness against outliers than SRUKF when $1 < \tilde{p}_3 < 2$ since the choice of $1 < \tilde{p}_3 < 2$ can be utilized for generating a weakened measurement update, as shown in Section 5.4.1. However, the choice of $1 < \tilde{p}_3 < 2$ is still not recommended since the measurement update may not be weakened efficiently under $1 < \tilde{p}_3 < 2$. Considering that HSRUKF utilizes a piecewise

Huber function with the l_1-norm of the error dealing with abnormal measurements, a positive number $\tilde{p}_3 \leq 1$ is adopted for guaranteeing the filtering robustness.

Therefore, PSRUKF generally adopts $\tilde{p}_1 = 2$ for dealing with error $|\check{e}_k(i)| < \vartheta$ in (5.24) and (5.45). By contrast, $\tilde{p}_2 > 2$ ($\tilde{p}_2$ not close to 2) can be utilized for the hidden state estimation within a wide range. Although PSRUKF can achieve an improved robustness against outliers than the SRUKF under $\tilde{p}_3 < 2$, we aim to adopt $\tilde{p}_3 \leq 1$ for guaranteeing robustness against outliers in comparison with traditional nonlinear Kalman filters. In addition, the choice of mixed p-norm parameters including parameters $\tilde{p}_1$, $\tilde{p}_2$ and $\tilde{p}_3$ is also dependent on the sensitivity of measurements to outliers and the varying magnitudes of outliers.

5.4.3 Choice of Balanced Parameters

The balanced parameters $\{\lambda_i\}_{i=n_x+1}^{n_x+n_y}$ in (5.23) and (5.45) play an essential role in improving filtering accuracy and numerical stability. A small threshold parameter ϑ in (5.24) and (5.25) means that the normal estimation error is weighted by the p-norm with $p \neq 2$, which does not help improve filtering accuracy. Thus, the second-order statistical characteristic of the error given in (5.23) and (5.45) can be used for alleviating the effect of the inappropriate choice of the threshold parameter. In addition, $\{\lambda_i\}_{i=n_x+1}^{n_x+n_y}$ in (5.23) and (5.45) can also be beneficial for operating the inverse of the matrix $\check{\mathbf{\Gamma}}_k$ in (5.48) since a large parameter $\tilde{p}_2 > 2$ in (5.24) and (5.25) may make $\check{\mathbf{\Gamma}}_k$ in (5.48) close to zero. Generally, $\{\lambda_i\}_{i=n_x+1}^{n_x+n_y}$ can be chosen by trials, which is complicated. Therefore, an adaptive scenario associated with factors $\{\lambda_i\}_{i=n_x+1}^{n_x+n_y}$, i.e., $\lambda_i = \exp(-e_k^2(i)/(2\varrho^2))$ or $\lambda_i = \exp(-\check{e}_k^2(i)/(2\varrho^2))$ for a linear or nonlinear measurement function in (5.13) and (5.1), respectively, is used with ϱ as the kernel width.

The kernel width ϱ therefore plays an essential role in estimation precision and numerical stability. When a small kernel size is utilized, the mixed p-norm in (5.45) may be alleviated or even completely eliminated, reducing to the traditional

SRUKF. Therefore, PSRUKF has an undesirable filtering accuracy under a small kernel size. In addition, the second-order statistical characteristic of the error in (5.45) is alleviated under a large kernel size, which is not beneficial for addressing the issue of inappropriate choice of the threshold parameter ϑ and the calculation of the inverse of $\breve{\mathbf{\Gamma}}_k$ in (5.48). The choice of the kernel width ϱ is further demonstrated in Section 5.6.

5.4.4 Choice of Threshold Parameter

The threshold parameter ϑ in (5.24) and (5.25) generally indicates whether the error is abnormal or the measurement is corrupted by outliers. An inappropriate choice of threshold parameter ϑ degrades the filtering precision and stability. For example, a large threshold may create abnormal errors weighted by the l_2-norm, degrading the filtering robustness against outliers. Moreover, a small threshold means that the normal estimation error corresponding to measurements not corrupted by outliers is dealt with using the p-norm with $p \neq 2$, which is not the desirable choice of the p-norm and may degrade the filtering accuracy. However, it is hard to accurately find the boundary of the abnormal error and therefore impossible to find the exact threshold parameter. In (5.23) and (5.45), an extra second-order statistical characteristic of the error is introduced for improving the filtering precision under a small threshold ϑ. Although the threshold parameter is set to be small, the measurement error in accordance with measurement not corrupted by outliers can also be handled by the l_2-norm. Therefore, the threshold parameter ϑ can be chosen as a relatively small value so that good filtering accuracy can be guaranteed. Moreover, the extra second-order statistical characteristic of the error is also beneficial for calculating the inverse of matrix $\breve{\mathbf{\Gamma}}_k$ in (5.48), which is constructed by a nonlinear function of the error $\psi(\breve{e}_k(i))$ shown in (5.47). The choice of the threshold ϑ is demonstrated in Section 5.6.

5.5 Dynamic State Estimation of Power Systems

This section discusses how nonlinear Kalman filters are applied in the state estimation of power systems. The generator and measurement model described in [125] is utilized for multi-machine systems and allows varying order generator models.

Let Θ_2, Θ_4, and Θ_p denote the sets of generators described by second-order models, fourth-order models and models where corrupted measurements are obtained from PMUs. Variables and constants related with power systems are defined in Table 5.1. For generator $\eta \in \Theta_4$, the fast sub-transient dynamics and saturation effects are neglected. This leads to the generator model expressed by a set of fourth-order differential equations in the *d-q* domain [14].

$$\dot{\delta}_\eta = \omega_\eta - \omega_0 \tag{5.51}$$

$$\dot{\omega}_\eta = \frac{\omega_0}{2H_\eta}\left(T_{m\eta} - T_{e\eta} - \frac{K_{D\eta}}{\omega_0}\left(\omega_\eta - \omega_0\right)\right) \tag{5.52}$$

$$\dot{e}^{'}_{\mathrm{q}\eta} = \frac{1}{T^{'}_{d_{0\eta}}}\left(E_{fd\eta} - e^{'}_{q\eta} - \left(x_{d\eta} - x^{'}_{d\eta}\right) i_{d\eta}\right) \tag{5.53}$$

$$\dot{e}^{'}_{\mathrm{d}\eta} = \frac{1}{T^{'}_{q_{0\eta}}}\left(-e^{'}_{d\eta} + \left(x_{q\eta} - x^{'}_{q\eta}\right) i_{q\eta}\right). \tag{5.54}$$

For generator $\eta \in \Theta_2$, the second-order model is derived from the fourth-order model directly by removing equations (5.53) and (5.54). For generator $\eta \in \Theta_p$, terminal voltage phasor $E_{t\eta} = e_{\mathrm{R}\eta} + je_{\mathrm{I}\eta}$ and terminal current phasor $I_{t\eta} = i_{\mathrm{R}\eta} + ji_{\mathrm{I}\eta}$ are used as outputs.

Comparing (5.1)–(5.2) with (5.51)–(5.54), we obtain the general state equations in state space form as

$$\dot{\mathbf{x}} = \boldsymbol{f}_c(\mathbf{x}, \mathbf{u}) \tag{5.55}$$

$$\mathbf{y} = \boldsymbol{h}_c(\mathbf{x}, \mathbf{u}) \tag{5.56}$$

with $\boldsymbol{f}_c$ and $\boldsymbol{h}_c$ representing the column vectors of continuous state transmission

function and measurement function. The state vector $\mathbf{x}$, input vector $\mathbf{u}$ and observable output $\mathbf{y}$ in (5.55) and (5.56) are given by

$$\mathbf{x} = \left[\boldsymbol{\delta}^{\mathrm{T}} \;\; \boldsymbol{\omega}^{\mathrm{T}} \;\; \mathbf{e}_q^{'\mathrm{T}} \;\; \mathbf{e}_d^{'\mathrm{T}}\right]^{\mathrm{T}} \tag{5.57}$$

$$\mathbf{u} = \left[\mathbf{T}_m^{\mathrm{T}} \;\; \mathbf{E}_{fd}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{5.58}$$

$$\mathbf{y} = \left[\mathbf{e}_{\mathrm{R}}^{\mathrm{T}} \;\; \mathbf{e}_{\mathrm{I}}^{\mathrm{T}} \;\; \boldsymbol{i}_{\mathrm{R}}^{\mathrm{T}} \;\; \boldsymbol{i}_{\mathrm{I}}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{5.59}$$

where the input $\mathbf{u} = \left[\mathbf{T}_m^{\mathrm{T}} \;\; \mathbf{E}_{fd}^{\mathrm{T}}\right]^{\mathrm{T}}$ is considered as a known constant. Variables $i_{q\eta}$, $i_{d\eta}$, and $T_{e\eta}$ in (5.52), (5.53) and (5.54) are expressed as functions of the state $\mathbf{x}$:

$$\boldsymbol{\Psi}_{\mathrm{R}\eta} = e_{d\eta}^{'} \sin\delta_\eta + e_{q\eta}^{'} \cos\delta_\eta \tag{5.60}$$

$$\boldsymbol{\Psi}_{\mathrm{I}\eta} = e_{q\eta}^{'} \sin\delta_\eta - e_{d\eta}^{'} \cos\delta_\eta \tag{5.61}$$

$$I_{t\eta} = \bar{\mathbf{Y}}_\eta \left(\boldsymbol{\Psi}_{\mathrm{R}} + j\boldsymbol{\Psi}_{\mathrm{I}}\right) \tag{5.62}$$

$$i_{\mathrm{R}\eta} = \mathrm{Re}(I_{t\eta}) \tag{5.63}$$

$$i_{\mathrm{I}\eta} = \mathrm{Im}(I_{t\eta}) \tag{5.64}$$

$$i_{q\eta} = \frac{S_B}{S_{N\eta}} \left(i_{\mathrm{I}\eta} \sin\delta_\eta + i_{\mathrm{R}\eta} \cos\delta_\eta\right) \tag{5.65}$$

$$i_{d\eta} = \frac{S_B}{S_{N\eta}} \left(i_{\mathrm{R}\eta} \sin\delta_\eta - i_{\mathrm{I}\eta} \cos\delta_\eta\right) \tag{5.66}$$

$$e_{q\eta} = e_{q\eta}^{'} - x_{d\eta}^{'} i_{d\eta} \tag{5.67}$$

$$e_{d\eta} = e_{d\eta}^{'} + x_{q\eta}^{'} i_{q\eta} \tag{5.68}$$

$$P_{e\eta} = e_{q\eta} i_{q\eta} + e_{d\eta} i_{d\eta} \tag{5.69}$$

$$T_{e\eta} = \frac{S_B}{S_{N_\eta}} P_{e\eta} \tag{5.70}$$

where $\bar{\mathbf{Y}}_\eta$ is the η-th row of the admittance matrix of the reduced network consisting of generators $\bar{\mathbf{Y}}$ [28]. Since outputs $i_{\mathrm{R}\eta}$ and $i_{\mathrm{I}\eta}$ are the function of state $\mathbf{x}$, outputs $e_{\mathrm{R}\eta}$

Table 5.1: Variables and Constants

Symbols	Definitions
δ	Rotor angle in rad
ω	Rotor speed in rad/s
ω_0	Rated rotor speed in rad/s
E_{fd}	Internal field voltage in pu
E_t	Terminal voltage phasor
e_d, e_q	Terminal voltage in d and q axes in pu
$e_d^{'}, e_q^{'}$	Transient voltage in d and q axes in pu
$e_{\mathrm{R}}, e_{\mathrm{I}}$	Real and imaginary parts of the terminal voltage phasor
x_d, x_q	Synchronous reactance in d and q axes in pu
$x_d^{'}, x_q^{'}$	Transient reactance in d and q axes in pu
I_t	Terminal current phasor
i_d, i_q	Currents in d and q axes in pu
$i_{\mathrm{R}}, i_{\mathrm{I}}$	Real and imaginary parts of the terminal current phasor in pu
T_m	Mechanical torque in pu
T_e	Electric air-gap torque in pu
$T_{q_0}^{'}, T_{d_0}^{'}$	Open-circuit time constants in q and d axes in second
S_B, S_N	System and generator base MVA
P_e	Electrical active output power in pu
K_D	Damping factor in pu
$\boldsymbol{\Psi}$	Voltage source vector
$\boldsymbol{\Psi}_{\mathrm{R}}, \boldsymbol{\Psi}_{\mathrm{I}}$	Column vectors of all generators' real and imaginary parts of the voltage source on system reference frame

and $e_{\mathrm{I}\eta}$ are also expressed as functions of state $\mathbf{x}$, i.e.,

$$e_{\mathrm{R}\eta} = e_{d\eta} \sin \delta_\eta + e_{q\eta} \cos \delta_\eta \tag{5.71}$$

$$e_{\mathrm{I}\eta} = e_{q\eta} \sin \delta_\eta - e_{d\eta} \cos \delta_\eta. \tag{5.72}$$

The discrete state space form can be obtained by discretizing the continuous models given in (5.51)–(5.54) as [125]

$$\mathbf{x}_k = \boldsymbol{f}\left(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}\right) + \boldsymbol{\omega}_{k-1} \tag{5.73}$$

Table 5.2: Computational Complexity of p-Norm Square Root Unscented Kalman Filter

Computation	Operation	Cost of Operation	Overall Cost
Sample points $\boldsymbol{\xi}_i, i = 0, \cdots, 2n_x$ in (5.11)	Scalar-Matrix product $\left(\sqrt{(n_x+\lambda)}\mathbf{S}_k\right)$	$n_x \times n_x$	$O(3n_x^2)$
	Sum of vectors	$2n_x \times n_x$	
$\boldsymbol{\chi}_{i,k\|k-1} = \boldsymbol{f}(\boldsymbol{\xi}_{i,k-1}), i = 0, \cdots, 2n_x$	Evaluation of f	$n_x C_f$	$O(n_x C_f)$
$\hat{\mathbf{x}}_{k\|k-1} = \sum_{i=0}^{2n_x} \omega_i^m \boldsymbol{\chi}_{i,k\|k-1}$	Scalar-vector products	$(2n_x+1) \times n_x$	$O(4n_x^2)$
	Sum of vectors	$2n_x \times n_x$	
$\boldsymbol{\Upsilon}_k = \sqrt{w_1^c}\,(\boldsymbol{\chi}_{1:2n_x,k\|k-1} - \hat{\mathbf{x}}_{k\|k-1})$	Subtract of matrices $\boldsymbol{\chi}_{1:2n_x,k\|k-1} - \hat{\mathbf{x}}_{k\|k-1}$	$2n_x \times n_x$	$O(4n_x^2)$
	Scalar-matrix product $\sqrt{w_1^c}\,(\boldsymbol{\chi}_{1:2n_x,k\|k-1} - \hat{\mathbf{x}}_{k\|k-1})$	$2n_x \times n_x$	
$\mathrm{qr}\left(\begin{bmatrix}\boldsymbol{\Upsilon}_k & \sqrt{\mathbf{Q}_{k-1}}\end{bmatrix}^{\mathrm{T}}\right)$	QR decomposition	$O(6n_x^3)$	$O(6n_x^3)$
$\sqrt{\lvert w_0^c\rvert}\,(\boldsymbol{\chi}_{0,k\|k-1} - \hat{\mathbf{x}}_{k\|k-1})$	Subtract of vectors $\boldsymbol{\chi}_{0,k\|k-1}$ and $\hat{\mathbf{x}}_{k\|k-1}$	$n_x \times 1$	$O(2n_x)$
	Scalar-vector product $\sqrt{\lvert w_0^c\rvert}\,(\boldsymbol{\chi}_{0,k\|k-1} - \hat{\mathbf{x}}_{k\|k-1})$	$n_x \times 1$	
$\mathrm{cholupdate}\left(\mathbf{S}_k^-,\ \sqrt{\lvert w_0^c\rvert}\,(\boldsymbol{\chi}_{0,k\|k-1} - \hat{\mathbf{x}}_{k\|k-1})\right)$	Cholupdate decomposition	$O(n_x^3)$	$O(n_x^3)$
Sample points $\boldsymbol{\xi}_i, i = 0, \cdots, 2n_x$ in (5.11)	Scalar-matrix product	$n_x \times n_x$	$O(3n_x^2)$
	Sum of vectors	$2n_x \times n_x$	
$\boldsymbol{\chi}_{i,k\|k-1} = \boldsymbol{h}(\boldsymbol{\xi}_{i,k-1}), i = 0, \cdots, 2n_x$	Evaluations of h	$n_x C_h$	$O(n_x C_h)$
$\hat{\mathbf{y}}_{k\|k-1} = \sum_{i=0}^{2n_x} \omega_i^m \boldsymbol{\chi}_{i,k\|k-1}$	Scalar-vector products	$(2n_x+1) \times n_y$	$O(4n_x n_y)$
	Sum of vectors	$2n_x \times n_y$	
$\bar{\boldsymbol{\Upsilon}}_k = \sqrt{w_1^c}\,(\boldsymbol{\chi}_{1:2n_x,k\|k-1} - \hat{\mathbf{y}}_{k\|k-1})$	Subtract of matrices $\boldsymbol{\chi}_{1:2n_x,k\|k-1} - \hat{\mathbf{y}}_{k\|k-1}$	$2n_x \times n_y$	$O(4n_x n_y)$
	Scalar-matrix product $\sqrt{w_1^c}\,(\boldsymbol{\chi}_{1:2n_x,k\|k-1} - \hat{\mathbf{y}}_{k\|k-1})$	$2n_x \times n_y$	
$\tilde{\mathbf{R}}_k$	Calculation of the matrix $\breve{\mathbf{P}}_k$	$O(n_y^3)$	$O(n_y^3)$
	Matrix-matrix product $\breve{\mathbf{e}}_k = \breve{\mathbf{P}}_k^{-1}\,[\delta\mathbf{x}_k \quad \mathbf{v}_k]^T$	$n_y^3 + n_y^2 + n_y(n_y-1)$	
	Inverse of diagonal matrix $\breve{\boldsymbol{\Gamma}}^{-1}$	n_y	
	Matrix-matrix product $\breve{\mathbf{P}}_k \breve{\boldsymbol{\Gamma}}_k^{-1} \breve{\mathbf{P}}_k^{\mathrm{T}}$	$n_y^2 + n_y^3 + n_y^2(n_y-1)$	
$\mathrm{qr}\left(\begin{bmatrix}\bar{\boldsymbol{\Upsilon}}_k & \sqrt{\tilde{\mathbf{R}}_k}\end{bmatrix}^{\mathrm{T}}\right)$	QR decomposition	$O(4n_x n_y^2 + 2n_y^3)$	$O(4n_x n_y^2 + 2n_y^3)$
$\sqrt{\lvert w_0^c\rvert}\,(\boldsymbol{\chi}_{0,k\|k-1} - \hat{\mathbf{y}}_{k\|k-1})$	Subtract of vectors$\boldsymbol{\chi}_{0,k\|k-1}$ and $\hat{\mathbf{y}}_{k\|k-1}$	$n_y \times 1$	$O(2n_y)$
	Scalar-vector product $\sqrt{\lvert w_0^c\rvert}\,(\boldsymbol{\chi}_{0,k\|k-1} - \hat{\mathbf{y}}_{k\|k-1})$	$n_y \times 1$	
$\mathrm{cholupdate}\left(\tilde{\mathbf{S}}_k^-,\ \sqrt{\lvert w_0^c\rvert}\,(\boldsymbol{\chi}_{0,k\|k-1} - \hat{\mathbf{y}}_{k\|k-1})\right)$	Cholupdate decomposition	$O(n_y^3)$	$O(n_y^3)$
$\mathbf{P}_{xy,k\|k-1} = \sum_{i=0}^{2n_x} \omega_i^c (\boldsymbol{\chi}_{i,k\|k-1} - \hat{\mathbf{x}}_{k\|k-1})$ $(\boldsymbol{h}(\boldsymbol{\chi}_{i,k\|k-1}) - \hat{\mathbf{y}}_{k\|k-1})^{\mathrm{T}}$	Vector-vector product	$(2n_x+1) \times n_x \times n_y$	$O(6n_x^2 n_y)$
	Scalar-matrix product	$(2n_x+1) \times n_x \times n_y$	
	Sum of matrix	$2n_x \times n_x \times n_y$	
$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k\|k-1} + \mathbf{K}_k(\mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k\|k-1}))$	Subtraction of vector $\mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k\|k-1})$	n_y	$O(2n_x n_y)$
	Vector-matrix product $\mathbf{K}_k(\mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k\|k-1}))$	$2n_x n_y - n_x$	
	Sum of vectors $\hat{\mathbf{x}}_{k\|k-1} + \mathbf{K}_k(\mathbf{y}_k - \boldsymbol{h}(\hat{\mathbf{x}}_{k\|k-1}))$	n_x	
$\mathbf{U} = \mathbf{K}_k \tilde{\mathbf{S}}_k^-$	Matrix-matrix product	$2n_x n_y^2 - n_x n_y$	$O(2n_x n_y^2)$
$\mathrm{cholupdate}\left(\mathbf{S}_k^-, \mathbf{U}, '-'\right)$	Cholupdate decomposition	$O(n_x^3)$	$O(n_x^3)$

$$\mathbf{y}_k = \boldsymbol{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) + \mathbf{v}_k. \tag{5.74}$$

The modified Euler method [126] can be utilized for generating the state transition functions $\boldsymbol{f}$, i.e.,

$$\tilde{\mathbf{x}}_k = \mathbf{x}_{k-1} + \boldsymbol{f}_c\left(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}\right)\Delta t \tag{5.75}$$

$$\tilde{\boldsymbol{f}} = \frac{\boldsymbol{f}_c\left(\tilde{\mathbf{x}}_k, \mathbf{u}_k\right) + \boldsymbol{f}_c\left(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}\right)}{2} \tag{5.76}$$

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \tilde{\boldsymbol{f}}\Delta t \tag{5.77}$$

where $\boldsymbol{f}_c$ is the column vector of continuous state transition functions constructed by (5.51)-(5.54).

The measurement function $\boldsymbol{h}$ in (5.74) is constructed through (5.63)-(5.64) and (5.71)-(5.72). Based on the constructed state space model, nonlinear Kalman filters can perform state estimation of power systems. The measurements from the power system in (5.59), i.e., $\mathbf{y} = \left[\mathbf{e}_{\mathrm{R}}^{\mathrm{T}} \;\; \mathbf{e}_{\mathrm{I}}^{\mathrm{T}} \;\; \boldsymbol{i}_{\mathrm{R}}^{\mathrm{T}} \;\; \boldsymbol{i}_{\mathrm{I}}^{\mathrm{T}}\right]^{\mathrm{T}}$, generally exhibit varying levels of sensitivity to outliers, thus degrading the filtering accuracy of nonlinear Kalman filters. Therefore, the proposed PSRUKF incorporating a mixed p-norm exhibits strong robustness against outliers in comparison with traditional nonlinear Kalman filters.

Remark 13 *A detailed model of the synchronous generator can be 5th-order [14], 6th-order [14], 10th-order [127] or even up to 14th-order [128]. However, the 4th-order model from (5.51) to (5.54) is the most commonly used model. From Table 5.2, the computational burden of the proposed mixed p-norm square root unscented Kalman filter (PSRUKF) is positively related to the dimension of the state. Therefore, the complex high order model imposes a high computational complexity in the process of state estimation. In addition, the complex high order model also needs to be characterized by more parameters [129, 130]. Therefore, the 4th-order model is*

generally utilized for state estimation using nonlinear Kalman filters in comparison with complex high order model like 5th-order [14], 6th-order [14], 10th-order [127] or even up to 14th-order [128] models.

5.6 Simulation Results

This section compares the filtering accuracy of the proposed p-norm square root unscented Kalman filter (PSRUKF) with traditional nonlinear Kalman filters including the square root unscented Kalman filter (SRUKF), Huber-based square root unscented Kalman filter (HSRUKF) and maximum correntropy criterion-based square root unscented Kalman filter (MCCSRUKF) for state estimation of power systems. The state estimation is executed by these nonlinear Kalman filters on the basis of the discrete state-space model described in Section 5.5, which is generated at a sampling rate of 120 samples per second. The observable measurements are obtained from the PMUs installed at different terminal buses of generators and the PMU sampling rate is set at 60 frames per second for sampling the measurements.

These nonlinear Kalman filters are tested on the WSCC 3-machine 9-bus system [28, 131, 132] and the NPCC 48-machine 140-bus system [133, 134], extracted from the Power System Toolbox [135, 136]. The efficiency of PSRUKF is demonstrated in the presence of alpha stable distribution noise. The alpha stable distribution noise as a non-Gaussian noise is represented by the following characteristic function [137, 138, 139]

$$f(t) = \exp\left\{j\delta t - \gamma|t|^{\sigma}\left[1 + j\rho \mathrm{sgn}\left(t\right) S\left(t, \sigma\right)\right]\right\} \tag{5.78}$$

where parameter set $V = (\sigma, \rho, \gamma, \delta)$ includes the characteristic factor $\sigma \in (0, 2]$, $\rho \in (-1, 1)$ measuring asymmetry, dispersion parameter $\gamma > 0$, location parameter $\delta \in$

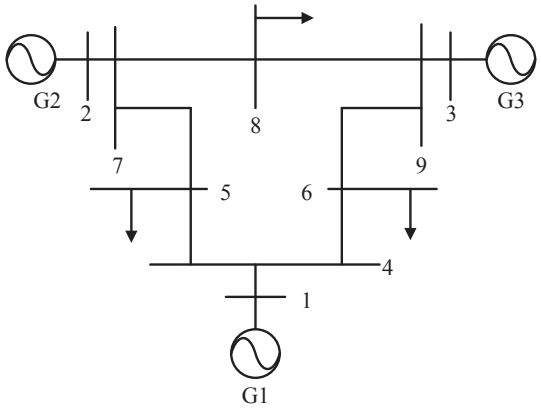


Figure 5.1: WSCC 3-machine 9-bus system.

$(-\infty, \infty)$, and

$$S(t,\sigma) = \begin{cases} \tan\frac{\sigma\pi}{2}, & \sigma \neq 1 \\ \frac{2}{\pi}\log|t|, & \sigma = 1. \end{cases} \tag{5.79}$$

In each experiment, 50 Monte Carlo simulations are run. The root mean square error (RMSE) is introduced for evaluating the filtering accuracy of nonlinear Kalman filters, which is defined by

$$\text{RMSE} = \frac{1}{\bar{N}}\sum_{k=1}^{\bar{N}}\sqrt{\sum_{j=\ell_1}^{\ell_2}(\mathbf{x}_k(j)-\hat{\mathbf{x}}_k(j))^2} \tag{5.80}$$

where $\mathbf{x}_k = [\mathbf{x}_k(1), \mathbf{x}_k(2), \cdots, \mathbf{x}_k(n_x)]^{\mathrm{T}}$ and $\hat{\mathbf{x}}_k = [\hat{\mathbf{x}}_k(1), \hat{\mathbf{x}}_k(2), \cdots, \hat{\mathbf{x}}_k(n_x)]^{\mathrm{T}}$ represent the desirable and estimated states, respectively. Notation $\bar{N}$ represents the number of samples. Scalars $\ell_1 = 1$, $\ell_2 = |\Theta_2|$ are utilized for calculating the RMSE of state estimation of $\boldsymbol{\delta} = [\delta_1, \delta_2, \cdots, \delta_\eta, \cdots, \delta_{|\Theta_2|}]$, where $|\Theta_2|$ is the cardinality of set Θ_2. Scalars $\ell_1 = |\Theta_2| + 1$, $\ell_2 = 2|\Theta_2|$ are for calculating state estimation error of $\boldsymbol{\omega} = [\omega_1, \omega_2, \cdots, \omega_\eta, \cdots, \omega_{|\Theta_2|}]$. In addition, $\ell_1 = 2|\Theta_2| + 1$, $\ell_2 = 2|\Theta_2| + |\Theta_4|$ are utilized for calculating the estimation error of $\mathbf{e}_q^{'} = [e_{q,1}^{'}, e_{q,2}^{'}, \cdots, e_{q,\eta}^{'}, \cdots, e_{q,|\Theta_4|}^{'}]$. The RMSE of state estimation of $\mathbf{e}_d^{'} = [e_{d,1}^{'}, e_{d,2}^{'}, \cdots, e_{d,\eta}^{'}, \cdots, e_{d,|\Theta_4|}^{'}]$ is calculated by setting $\ell_1 = 2|\Theta_2| + |\Theta_4| + 1$, $\ell_2 = 2|\Theta_2| + 2|\Theta_4|$.

5.6.1 WSCC 3-Machine System

The WSCC 3-machine system is utilized for testing the efficiency of the proposed p-norm square root unscented Kalman filter. It is assumed that generators are modelled by a second-order model in the WSCC 3-machine system associated with $\Theta_2 = \{1, 2, 3\}$, $\Theta_4 = \emptyset$, i.e., null set, and $\Theta_p = \{3\}$, generating state variable $\mathbf{x}_k \in \mathbb{R}^{6\times 1}$ and measurements $\mathbf{y}_k \in \mathbb{R}^{4\times 1}$.

A three-phase fault is applied at bus 8 of line $8-9$ in Fig. 5.1 and is cleared at the near and remote end after 0.05 and 0.1 second. The initial estimated mean of the system state is set to be the pre-contingency state. The dynamic state estimation is implemented on the post-contingency system. The initial estimation error covariance $\mathbf{P}_0$ is set as in [28], i.e.,

$$\mathbf{P}_0 = \begin{bmatrix} \varsigma_\delta^2 \mathbf{I}_{|\Theta_2|}^{|\Theta_2|} & \mathbf{0}_{|\Theta_2|}^{|\Theta_2|} & \mathbf{0}_{|\Theta_4|}^{|\Theta_2|} & \mathbf{0}_{|\Theta_4|}^{|\Theta_2|} \\ \mathbf{0}_{|\Theta_2|}^{|\Theta_2|} & \varsigma_w^2 \mathbf{I}_{|\Theta_2|}^{|\Theta_2|} & \mathbf{0}_{|\Theta_4|}^{|\Theta_2|} & \mathbf{0}_{|\Theta_4|}^{|\Theta_2|} \\ \mathbf{0}_{|\Theta_2|}^{|\Theta_4|} & \mathbf{0}_{|\Theta_2|}^{|\Theta_4|} & \varsigma_{e'_q}^2 \mathbf{I}_{|\Theta_4|}^{|\Theta_4|} & \mathbf{0}_{|\Theta_4|}^{|\Theta_4|} \\ \mathbf{0}_{|\Theta_2|}^{|\Theta_4|} & \mathbf{0}_{|\Theta_2|}^{|\Theta_4|} & \mathbf{0}_{|\Theta_4|}^{|\Theta_4|} & \varsigma_{e'_d}^2 \mathbf{I}_{|\Theta_4|}^{|\Theta_4|} \end{bmatrix} \tag{5.81}$$

with $\mathbf{I}_{|\Theta_2|}^{|\Theta_2|}$ representing the identity matrix $\mathbf{I} \in \mathbb{R}^{|\Theta_2|\times|\Theta_2|}$ and $\mathbf{0}_{|\Theta_4|}^{|\Theta_2|}$ being the zero matrix $\mathbf{0} \in \mathbb{R}^{|\Theta_2|\times|\Theta_4|}$. The parameters ς_δ, ς_w, $\varsigma_{e'_q}$ and $\varsigma_{e'_d}$ are set by $\varsigma_\delta = 0.5\pi/180$, $\varsigma_w = 10^{-3}\omega_0$ and $\varsigma_{e'_q} = \varsigma_{e'_d} = 10^{-3}$. The state noise follows the Gaussian distribution with the covariance set as a diagonal matrix and entries being the square of 10% of the largest state changes [136]. The measurements are corrupted by Gaussian noise with statistical characteristic $\mathbb{N}(\mathbf{0}_4, 10^{-2}\mathbf{I}_4)$, where $\mathbf{I}_4 \in \mathbb{R}^{4\times 4}$ represents the identity matrix. For better demonstration of the robustness of PSRUKF, the alpha stable distribution noises with $V = (0.3, 0, 0.5, 0)$, $V = (0.4, 0, 0.5, 0)$ and $V = (0.5, 0, 0.5, 0)$ are added on the first three dimensional measurements of $\{\mathbf{y}_k : \mathbf{y}_k = [y_k(1), y_k(2), y_k(3), y_k(4)]^T\}_{k=1}^{\bar{N}}$, i.e., $\{y_k(1), y_k(2), y_k(3)\}_{k=1}^{\bar{N}}$. Since the kernel width ϱ in (5.45) can alleviate the filtering degradation caused by an inappropriate threshold, Figs. 5.2(a) to 5.2(c) show the effect

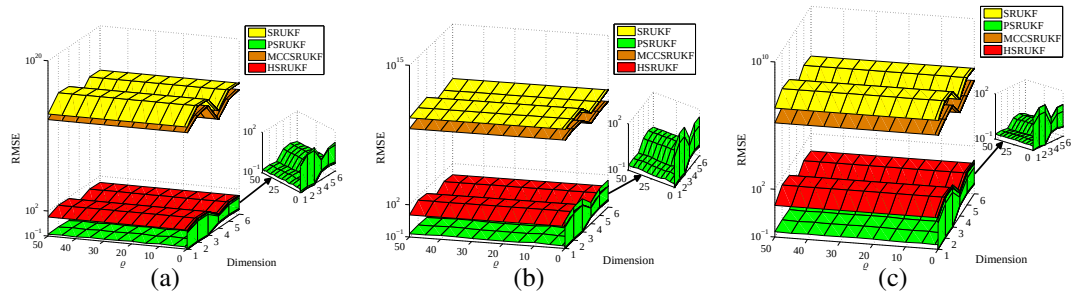


Figure 5.2: (a) RMSE of nonlinear Kalman filters with kernel width $\varrho \in [0, 50]$ under the alpha stable distribution $V = (0.3, 0, 0.5, 0)$; (b) RMSE of nonlinear Kalman filters with kernel width $\varrho \in [0, 50]$ under the alpha stable distribution $V = (0.4, 0, 0.5, 0)$; (c) RMSE of nonlinear Kalman filters with kernel width $\varrho \in [0, 50]$ under the alpha stable distribution $V = (0.5, 0, 0.5, 0)$.

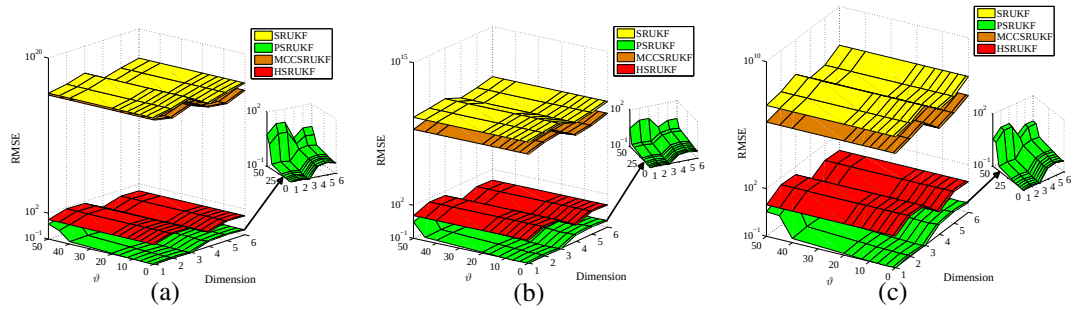


Figure 5.3: (a) RMSE of nonlinear Kalman filters with threshold $\vartheta \in [0, 50]$ under the alpha stable distribution $V = (0.3, 0, 0.5, 0)$; (b) RMSE of nonlinear Kalman filters withthreshold $\vartheta \in [0, 50]$ under the alpha stable distribution $V = (0.4, 0, 0.5, 0)$; (c) RMSE of nonlinear Kalman filters with threshold $\vartheta \in [0, 50]$ under the alpha stable distribution $V = (0.5, 0, 0.5, 0)$.

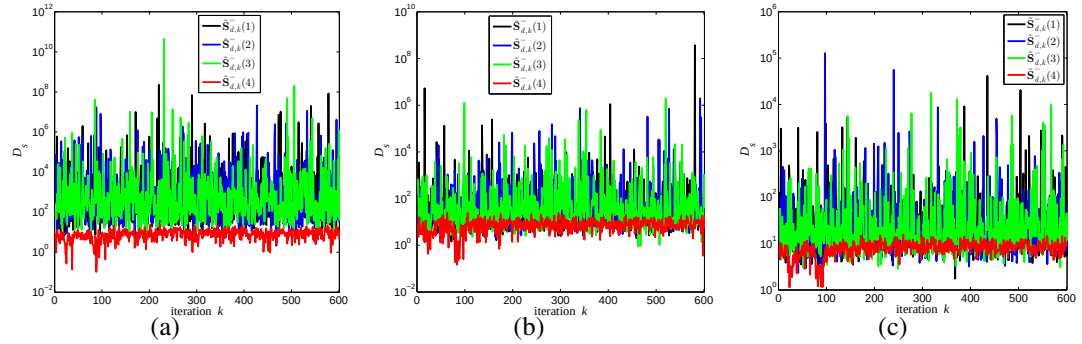
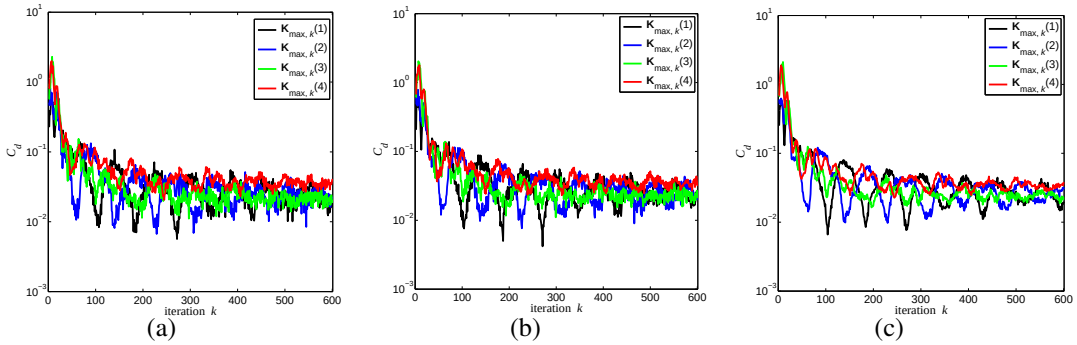


Figure 5.4: (a) D_s under the alpha stable distribution $V = (0.3, 0, 0.5, 0)$; (b) D_s under the alpha stable distribution $V = (0.4, 0, 0.5, 0)$; (c) D_s under the alpha stable distribution $V = (0.5, 0, 0.5, 0)$;



(a) (b) (c)

Figure 5.5: (a) C_d under the alpha stable distribution $V = (0.3, 0, 0.5, 0)$; (b) C_d under the alpha stable distribution $V = (0.4, 0, 0.5, 0)$; (c) C_d under the alpha stable distribution $V = (0.5, 0, 0.5, 0)$;

of the kernel width ϱ on estimating the hidden state $\mathbf{x}_k \in \mathbb{R}^6$ under alpha stable distribution noises. All nonlinear Kalman filters perform state estimation with $\kappa = 0$, $\alpha = 0.5$ and $\beta = 2$. For PSRUKF, we set $\vartheta = 0.3$, $\tilde{p}_1 = 2$, $\tilde{p}_2 = 15$ and $\tilde{p}_3 = 0.5$. In addition, the kernel parameter ϱ is in the range of $\varrho \in [0, 50]$. From Figs. 5.2(a) to 5.2(c), both PSRUKF and HSRUKF exhibit a higher filtering precision than traditional SRUKF and MCCSRUKF. Moreover, PSRUKF utilizes a mixed p-norm for dealing with measurements corrupted by outliers and achieves a higher filtering precision than HSRUKF. When a small kernel size ϱ is utilized, the mixed p-norm in (5.45) is alleviated, which approaches traditional SRUKF. Therefore, PSRUKF achieves undesirable filtering accuracy under a small kernel size ϱ. In addition, a large kernel size ϱ also affects the accuracy of the state estimation since the second-order statistical characteristic of the error in (5.45) is alleviated. From Figs. 5.2(a) to 5.2(c), PSRUKF performs state estimation with a higher filtering accuracy than nonlinear Kalman filters within a wide range of kernel parameter ϱ.

The threshold ϑ in (5.24) and (5.25) plays an essential role in dealing with outliers. Figs. 5.3(a) through 5.3(c) study the effect of the threshold parameter ϑ on estimating the hidden state $\mathbf{x}_k \in \mathbb{R}^6$ in the presence of alpha stable distribution noises. All nonlinear Kalman filters perform state estimation by setting $\kappa = 0$, $\alpha = 0.5$ and $\beta = 2$. For PSRUKF, the kernel width ϱ and parameters $\tilde{p}_1$, $\tilde{p}_2$ and $\tilde{p}_3$ in (5.23) and (5.24)

are set as $\varrho = 0.001$, $\tilde{p}_1 = 2$, $\tilde{p}_2 = 15$ and $\tilde{p}_3 = 0.5$. In addition, threshold ϑ in (5.24) and (5.25) is set in the range of $\vartheta \in [0, 50]$. From Fig. 5.3(a) to 5.3(c), both SRUKF and MCCSRUKF implement state estimation with a poor filtering precision. In addition, HSRUKF estimates the hidden state of the power system at a higher filtering precision than SRUKF and MCCSRUKF. By contrast, PSRUKF achieves the highest precision in comparison with SRUKF, MCCSRUKF and HSRUKF. Since PSRUKF may approach SRUKF under a large threshold in (5.24) and (5.25), PSRUKF has a poor filtering accuracy under a large threshold ϑ. Although PSRUKF generates an increasing RMSE under a large threshold parameter, PSRUKF still performs better than other nonlinear Kalman filters. When threshold ϑ in (5.24) and (5.25) is set very small, the measurement error corresponding to the measurement not corrupted by outliers may be weighted by a p-norm with $p \neq 2$, which is not the desirable choice and degrades the filtering accuracy. Since a second-order statistical characteristic of the error in (5.45) is introduced for avoiding this issue, PSRUKF can still achieve a desirable filtering precision with a small threshold ϑ.

As shown in Section 5.4.1, PSRUKF improves robustness against outliers by adaptively calculating $\tilde{\mathbf{R}}_k$ in (5.50). Since the alpha stable distribution noises are added on the first three dimensional measurements, i.e., $\{y_k(1), y_k(2), y_k(3)\}_{k=1}^{\tilde{N}}$, the first three diagonal elements of $\tilde{\mathbf{R}}_k$ in (5.50) are large based on the choice of the p-norm in (5.24) and (5.25). The matrix decomposition in (5.7) generates the factorization matrix $\tilde{\mathbf{S}}_k^-$, and the absolute values of the first three diagonal elements are large. This means that the absolute values of the first three columns' elements of gain matrix $\mathbf{K}_k$ are small because gain matrix $\mathbf{K}_k$ is constructed by the inverse of $\tilde{\mathbf{S}}_k^-$. Therefore, PSRUKF performs state estimation by utilizing a time update and weakened measurement update in (5.9). Define the absolute values of the diagonal elements of $\tilde{\mathbf{S}}_k^- \in \mathbb{R}^4$ as C_d and maximum absolute values at each column of Kalman gain $\mathbf{K}_k$ as D_s. Figs. 5.4(a) to 5.5(c) show D_s and C_d with parameter settings extracted from the experiments shown in Figs. 5.2(a) to 5.3(c), i.e., kernel width $\varrho = 0.001$,

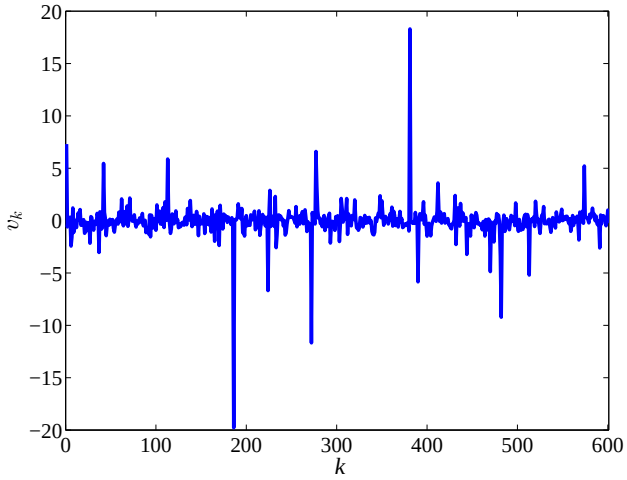


Figure 5.6: A sequence of alpha-stable noise with $V = (1.4, 0, 0.4, 0)$.

threshold parameter $\vartheta = 0.3$, $\tilde{p}_1 = 2$, $\tilde{p}_2 = 15$ and $\tilde{p}_3 = 0.5$ in the presence of alpha stable distribution noises. It can be seen from Figs. 5.4(a) to 5.4(c) that the absolute values of the first three diagonal elements of $\tilde{\mathbf{S}}_k^-$, i.e., $\tilde{\mathbf{S}}_{d,k}^-(1)$, $\tilde{\mathbf{S}}_{d,k}^-(2)$ and $\tilde{\mathbf{S}}_{d,k}^-(3)$, are generally larger than that of the last diagonal entry, i.e., $\tilde{\mathbf{S}}_{d,k}^-(4)$. Define the maximum absolute value of the jth column elements of the Kalman gain at discrete time k as $\mathbf{K}_{\max,k}(j)$. From Figs. 5.5(a) to 5.5(c), the values of $\mathbf{K}_{\max,k}(1)$, $\mathbf{K}_{\max,k}(2)$ and $\mathbf{K}_{\max,k}(3)$ are generally smaller than $\mathbf{K}_{\max,k}(4)$. The PSRUKF uses a mixed p-norm for dealing with measurements corrupted by outliers so that a relatively small error $\psi(\check{e}_k(i))$ in (5.47) can be obtained. The measurements not corrupted by outliers occasionally generate a small error $\psi(\check{e}_k(i))$ in (5.47). Therefore, $\mathbf{K}_{\max,k}(1)$, $\mathbf{K}_{\max,k}(2)$ and $\mathbf{K}_{\max,k}(3)$ may occasionally exhibit relatively larger values than the $\mathbf{K}_{\max,k}(4)$. From Figs. 5.4(a) to 5.5(c), PSRUKF improves the robustness of the state estimation by utilizing a time update and weakened measurement update where the first three dimensional measurements have less effect on state estimation.

5.6.2 NPCC 48-Machine System

The NPCC system [136] is a 48-machine system including 27 generators modelled by a fourth-order model and 21 generators modelled by a second-order classical model, i.e.,

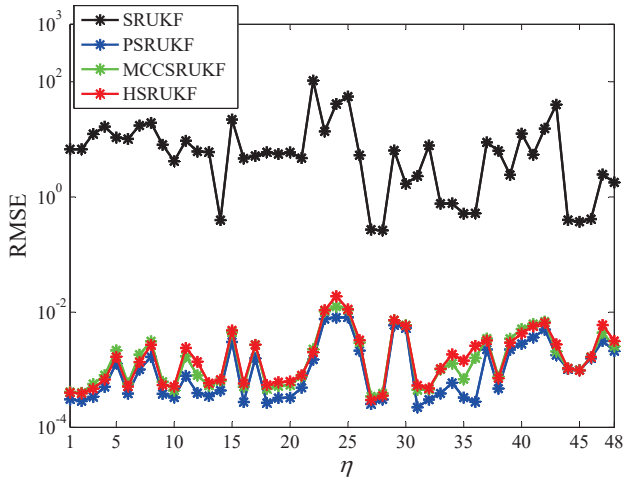


Figure 5.7: RMSE of state estimation of the rotor angle $\{\delta_\eta\}_{\eta=1}^{48}$.

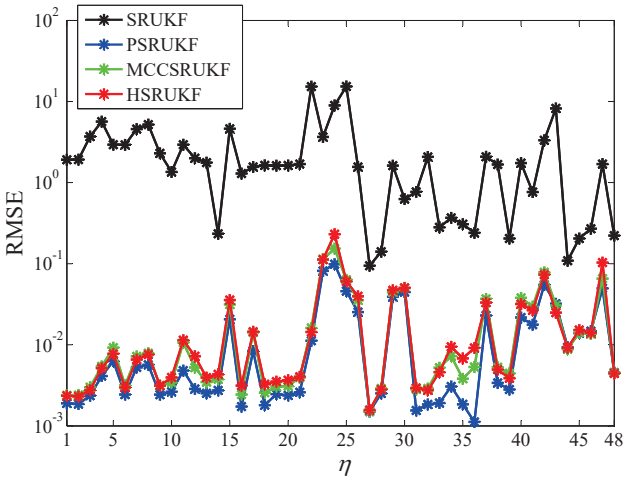


Figure 5.8: RMSE of state estimation of the rotor speed $\{\omega_\eta\}_{\eta=1}^{48}$.

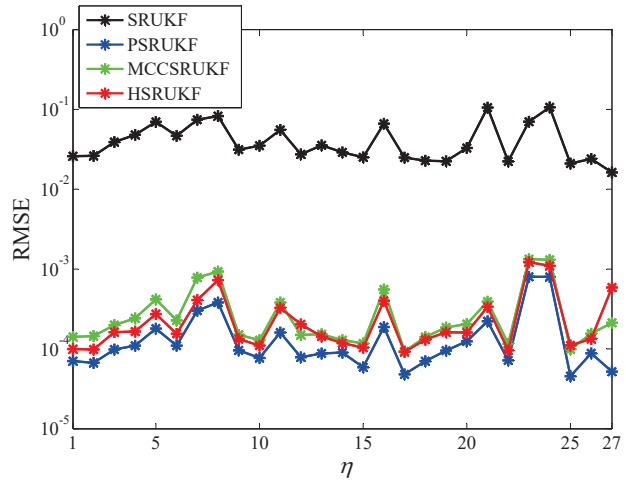


Figure 5.9: RMSE of state estimation of the transient voltage $\{e'_{q\eta}\}_{\eta=1}^{27}$ in the q-axis.

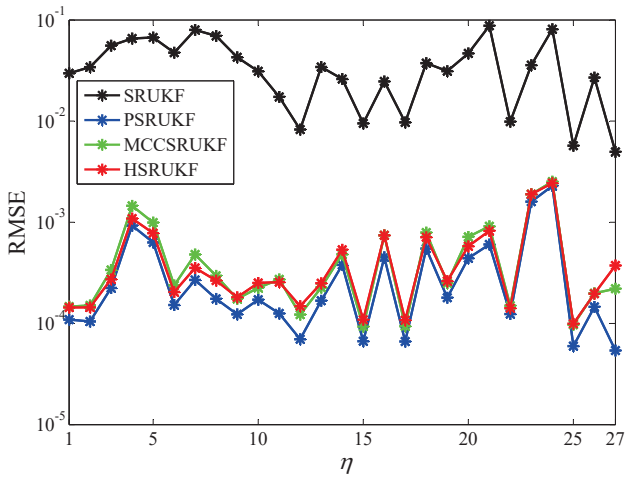


Figure 5.10: RMSE of state estimation of the transient voltage $\{e'_{d\eta}\}_{\eta=1}^{27}$ in the d-axis.

$|\Theta_2| = 21$ and $|\Theta_4| = 27$. The measurements are obtained from the generators associated with index set $\Theta_p = \{1, 2, 3, 4, 6, 9, 10, 12, 13, 14, 16, 18, 19, 20, 21, 27, 28, 31, 32, 35, 36, 38, 44, 45\}$, resulting in state variable $\mathbf{x}_k \in \mathbb{R}^{150\times 1}$ and measurement $\mathbf{y}_k \in \mathbb{R}^{96\times 1}$ [28].

Dynamic state estimation is performed when a three-phase fault is applied at the bus of one of the 50 branches with the highest line flow as in [28] and is cleared at the near and remote end after 0.05 and 0.1 second. Similar to that for the WSCC 3-machine system, the initial estimated mean of the system state is set by the pre-contingency state. The dynamic state estimation is performed on the post-contingency system. The initial estimation error covariance $\mathbf{P}_0$ is set similarly as that for the WSCC 3-machine system. The state estimation is implemented by nonlinear Kalman filters parameterized by $\kappa = 0$, $\alpha = 0.5$ and $\beta = 2$, respectively. The parameters of the proposed PSRUKF, ϑ and $\{\tilde{p}_i\}_{i=1}^{3}$ in (5.24) are set by $\vartheta = 0.05$, $\tilde{p}_1 = 2$, $\tilde{p}_2 = 3.5$ and $\tilde{p}_3 = 0.01$. The kernel parameter ϱ is set as $\varrho = 10^{-4}$ for achieving a better filtering precision.

The state noise obeys the Gaussian distribution, with covariance set as a diagonal matrix and entries being the square of 1% of the largest state changes. The measurements are corrupted by Gaussian noise with statistical characteristic

$\mathbb{N}(\mathbf{0}_{96}, 10^{-4}\mathbf{I}_{96})$, where $\mathbf{I}_{96} \in \mathbb{R}^{96\times96}$ represents the identity matrix. In addition, alpha stable distribution noise set by $(1.4, 0, 0.4, 0)$ in Fig. 5.6 is added randomly on $\tau = 50$ measurements.

Figs. 5.7 to 5.10 compare the filtering accuracy of several nonlinear Kalman filters for $\boldsymbol{\delta} \in \mathbb{R}^{48\times1}$, $\boldsymbol{\omega} \in \mathbb{R}^{48\times1}$, $\mathbf{e}_q^{'} \in \mathbb{R}^{27\times1}$ and $\mathbf{e}_d^{'} \in \mathbb{R}^{27\times1}$. From Figs. 5.7-5.10, SRUKF is sensitive to outliers and therefore gives the worst filtering accuracy. By contrast, both HSRUKF and MCCSRUKF exhibit robustness against outliers by utilizing a piecewise Huber function and information theoretical learning, respectively. The PSRUKF achieves the highest precision by using a mixed p-norm for weighting measurement errors efficiently. The PSRUKF achieves the highest precision by using a mixed p-norm for weighting measurement errors efficiently.

5.7 Conclusion

State estimation of power systems is essential for tracking the inherent dynamics due to various uncertainties. The nonlinear Kalman filter is often applied to dynamic state estimation of power systems by utilizing the available measurements obtained from phasor measurement units. Since measurements are corrupted by outliers and exhibit varying levels of sensitivity to outliers, a p-norm square root unscented Kalman filter is proposed for weighting measurement errors so that outliers can be handled efficiently. The performance of the p-norm square root unscented Kalman filter is demonstrated in different scenarios through numerical examples. Result shows superior filtering accuracy of the p-norm square root unscented Kalman filter in comparison with traditional square root unscented Kalman filters and robust nonlinear Kalman filters.

Chapter 6

Conclusions and Suggestions for Future Work

In this chapter, we summarize the main contributions of the thesis and discuss some potential research directions in the future.

6.1 Main Contributions of Thesis

This thesis aims to study the dynamics of information flow on networks by using nonlinear Kalman filters in the presence of Gaussian and non-Gaussian noise. Nonlinear Kalman filters capture the dynamics of information flow on the basis of a state model and the available measurements simultaneously. The state model can be used to describe the propagation dynamics or the dynamics of nonlinear systems. The epidemic tracking and the dynamic analysis of power systems are representative applications which consider the propagation dynamics and the dynamics of nonlinear systems, respectively. This thesis specifically considers the applications of nonlinear Kalman filters to the study of the dynamics of epidemic spreading and power networks.

The estimation of the dynamic profiles of epidemic spreading on networks is important for controlling morbidity. Nonlinear Kalman filters have emerged as a

powerful platform for studying the dynamics of epidemic spreading on networks. However, it may be hard to find appropriate nonlinear Kalman filters for epidemic tracking over various kinds of complex networks. A guideline is therefore provided for selecting appropriate Kalman filtering methods for tracking epidemic spreading on some common complex networks in the presence of Gaussian noise. In addition, traditional Kalman filters perform optimization based on the minimum mean square error criterion. Therefore, traditional nonlinear Kalman filters may be sensitive to abnormal measurements. Considering that available measurements may be corrupted by non-Gaussian noise, a novel nonlinear Kalman filter called generalized correntropy sparse Gauss-Hermite quadrature filter (GCSGHQF) has been derived by combining the information theoretical learning and the sparse Gauss-Hermite quadrature filter for handling non-Gaussian noise. Apart from the application of epidemic spreading over complex networks, the estimation of the dynamics of power systems is another important issue relevant to the monitoring of the operation of power distribution infrastructure. Similar to the available measurements in epidemic spreading on networks, measurements collected from power systems may be also contaminated by non-Gaussian noise. A robust mixed p-norm square root unscented Kalman filter is presented for estimating the state of power systems in the presence of non-Gaussian noise. Unlike the generalized correntropy sparse Gauss-Hermite quadrature filter (GCSGHQF) used in the application of epidemic tracking, the mixed p-norm square root unscented Kalman filter uses a piecewise function, i.e., multiple p-norms for dealing with measurements corrupted by non-Gaussian noise in some power systems including the WSCC (Western System Coordinating Council) 3-machine system and the NPCC (Northeastern Power Coordinating Council) 48-machine system.

The main contributions of this thesis can be summarized as follows:

1. The performance comparison in terms of accuracy and stability forms a guideline for utilizing nonlinear Kalman filters for tracking epidemic spreading in the presence of Gaussian noise.

A compartmental model is in general used for describing epidemic transmission on networks. Moreover, a nonlinear Kalman filter is a more preferable choice for the purpose of epidemic tracking by using an inherent dynamic model like the compartmental model together with epidemic transmission in practice. It is non-trivial to choose appropriate nonlinear Kalman filters for epidemic tracking over various networks. This motivates us to provide a guideline where appropriate nonlinear Kalman filters together with compartmental models are chosen for performing epidemic tracking on different kinds of networks like the Erdös and Rényi (ER) network, the Newman and Watts (NW) network, and the Watts and Strogatz (WS) network. A comparison between nonlinear Kalman filters has been performed in terms of estimation precision and stability. The performance comparison provides a guideline for choosing appropriate nonlinear Kalman filters for studying the dynamics of epidemic spreading over complex networks. Since compartmental models may yield a high dimensional state and pose a heavy computational burden, several strategies are provided for solving this issue especially for large-scale complex networks.

2. A novel generalized correntropy sparse Gauss-Hermite quadrature filter (GCSGHQF) is proposed for epidemic tracking on homogeneous networks in the presence of non-Gaussian noise.

A guideline has been provided for choosing appropriate nonlinear Kalman filters for epidemic tracking in the presence of Gaussian noise. However, these traditional Kalman filters are sensitive to abnormal measurements. A robust nonlinear Kalman filter, referred to as generalized correntropy sparse Gaussian Hermite Quadrature filter (GCSGHQF), has been proposed for dealing with non-Gaussian noise. The generalized correntropy sparse Gaussian Hermite Quadrature filter improves robustness by applying generalized correntropy into the sparse Gauss-Hermite quadrature filter. The generalized correntropy in the framework of information theoretical learning (ITL) is a preferable choice for retaining error information which is beneficial for handling non-Gaussian noise. An extra second-order statistical

characteristic of the error is also considered for further improving numerical stability. The generalized correntropy sparse Gauss-Hermite quadrature filter is utilized for exploring the behavior of epidemic spreading on complex networks such as the Erdös and Rényi (ER) network, the Newman and Watts (NW) network, and the Watts and Strogatz (WS) network.

3. A robust mixed p-norm square root unscented Kalman filter is proposed for the state estimation of power systems in the presence of non-Gaussian noise.

A robust mixed p-norm square root unscented Kalman filter (PSRUKF) has been developed by applying the mixed p-norm into the square root unscented Kalman filter for improving robustness in the presence of non-Gaussian noise. In particular, the proposed mixed p-norm is regarded as a more generalized p-norm over multiple intervals, which deals with measurements corrupted by non-Gaussian noise efficiently. A threshold parameter is used for determining the boundary of abnormal measurement errors and then the mixed p-norm is applied to handling measurement errors. Also, an extra second-order statistical characteristic of the error is used for reducing the effect of the inappropriate choice of the threshold parameter. The proposed PSRUKF has been applied for the state estimation of power systems including the WSCC (Western System Coordinating Council) 3-machine system and the NPCC (Northeastern Power Coordinating Council) 48-machine system at non-Gaussian noise.

6.2 Suggestions for Future Work

In this section, we present some possible research topics that deserve further in depth study.

6.2.1 Robust Clustering for Community Detection of Complex Networks

Nonlinear Kalman filters are indeed a powerful platform for tracking information flow through a complex network. Nonlinear Kalman filters, however, may be confronted with large computational burden along with the increasing scale of complex networks. As demonstrated in Section 3.5.2, we can perform community detection before nonlinear Kalman filters perform state estimation. Then, state estimation is implemented within each community with reduced computational effort.

Complex networks can be represented by graphs, in which the links represent some relationship among nodes. Complex networks are commonly organized in dense subgraphs referred to as communities [140]. The dense subgraph is an essential concept in understanding and exploring complex networks since the community is often used for explaining the functional characteristics of complex networks. The nodes of intracommunity connections are connected densely in comparison with those of intercommunity connections. Recently, there has been renewed interest in detecting the communities of complex networks such as biological networks, social networks and information networks for further analyzing network structures [98, 141, 142, 143]. Community detection can be interpreted as a clustering problem in which standard clustering techniques are applied to clustering available samples according to similarity, generating community structures in complex networks. Community detection has been implemented by numerous clustering algorithms of machine learning such as nonnegative matrix factorization (NMF) [144]. Community detection is commonly executed on an adjacency matrix taking a binary form with ”1” representing node connectivity and ”0” for node disconnectivity. Traditional community detection aims to find community structures without considering the binary characteristic of the adjacency matrix. Therefore, a class of weighted nonnegative matrix factorization including weighted least square errors-based nonnegative matrix

factorization (WLSENMF) and weighted KL divergence-based nonnegative matrix factorization (WKLNMF) can be considered by introducing a weighted matrix exponentially in accordance with the numerical characteristic of the adjacency matrix. The introduced weighted matrix compels the product of two nonnegative matrices to be updated for resembling the adjacency matrix. In addition, the prior information about communities into which some available samples are clustered can be known in advance. Clustering accuracy can thus be improved by incorporating the prior information [145] into the clustering task, constituting a semi-supervised community detection of complex networks. For further improving the accuracy of community detection, the use of prior information about the community to which nodes belong is also recommended for WLSESNMF and WKLNMF, generating semi-supervised weighted least square errors-based nonnegative matrix factorization (SWLSENMF) and semi-supervised weighted KL divergence-based nonnegative matrix factorization (SWKLNMF). By using SWLSENMF and SWKLNMF for community detection, nonlinear Kalman filters perform state estimation within each community for reducing the computational burden. Dynamic estimation of a large network can be performed by integrating the estimations on all communities.

6.2.2 Deep Reinforcement Learning-Based Nonlinear Kalman Filter

As demonstrated in (4.61) in Section 4.4.2, the robustness of nonlinear Kalman filters is greatly affected by the covariance matrix of the measurement noise. Nonlinear Kalman filters have desirable filtering precision if the covariance matrix of the measurement noise can be intelligently adjusted.

Deep reinforcement learning [146] is expected to achieve an appropriate choice of the covariance of the measurement noise via an agent that continuously interacts with the environment. The traditional deep Q-network [147] in the framework of

deep reinforcement learning utilizes a deep neural network for learning a nonlinear function which maps the state or observation to accumulated rewards conditional upon the current state and the agent action represented by the Q-value. In a deep Q-network, the target Q-value and the estimated Q-value are calculated by the deep neural network directly. The state or observation is in general corrupted by noise. The deep neural network may be sensitive to noise and also greatly affected by unsatisfactorily trained network parameters, giving undesirable Q-values.

It may therefore be beneficial to combine the deep Q-network with kernel-based adaptive filters. Kernel methods aim to solve nonlinear filtering in a reproducing kernel Hilbert space (RKHS) [148]. Commonly used kernel adaptive filters (KAFs) may employ the kernel least mean square algorithm (KLMS) [149], kernel affine projection algorithm (KAPA) [150] and kernel recursive least squares (KRLS) algorithm [151]. Among these algorithms, KLMS delivers desirable filtering performance with the lowest computational cost, which can be considered. However, choosing an appropriate step size is often non-trivial in KLMS. A small step size may not achieve a desirable smoothing efficiency. In contrast, a large step size is beneficial for smoothing but may cause over-smoothing. Therefore, a weighting procedure that uses past Q-values can be further considered. In the weighting process, each sample within the experience pool is equipped with a sub-pool which stores its corresponding past Q-values. These past Q-values are considered for yielding the newest smooth output without over-smoothing. A novel smooth deep Q-network (SDQN) can be developed by incorporating the KLMS algorithm and a weighting procedure used in the deep Q-network for smoothing the outputs produced by the deep neural network.

6.2.3 Quantized Nonlinear Kalman Filter

As shown in Section 4.4.1, the core of nonlinear Kalman filters lies in the calculation of the predictive probability density and corrected probability density at the predictive

step and the update step, respectively. Nonlinear Kalman filters may be confronted with high computational complexity in solving the state estimation of complex networks with large scale. Apart from strategies used for reducing computational burden in Section 3.5.2, novel nonlinear Kalman filters are also expected to be proposed for solving this issue. For example, a quantized nonlinear Kalman filter can be utilized for alleviating the calculational burden [149]. More concretely, several quantization areas can be constructed at first. Each quantization area includes the predictive states, corrected states and measurements where the Euclidean distance of any pair of predictive states and measurements is smaller than a pre-designed threshold parameter. The corrected state at discrete time k can be directly calculated by averaging all corrected states within one specific quantization area. As a result, the computational complexity can be reduced greatly by neglecting the corrected step of nonlinear Kalman filters.

Bibliography

[1] A. L. Barabási, *Network Science*. UK: Cambridge University Press, 2016.

[2] P. Erdös and A. Rényi, "On random graphs," *Publicationes Mathematicae*, vol. 6, no. 26, pp. 290–297, 1959.

[3] A. Dadlani, M. S. Kumar, S. Murugan, and K. Kim, "System dynamics of a refined epidemic model for infection propagation over complex networks," *IEEE Systems Journal*, vol. 10, no. 4, pp. 1316–1325, Dec. 2016.

[4] O. Erkaymaz and M. Ozer, "Impact of small-world network topology on the conventional artificial neural network for the diagnosis of diabetes," *Chaos, Solitons and Fractals*, vol. 83, pp. 178–185, Feb. 2016.

[5] G. Zhou, C. Li, T. Li, Y. Yang, C. Wang, F. He, and J. Sun, "Outer synchronization investigation between WS and NW small-world networks with different node numbers," *Physica A: Statistical Mechanics and its Applications*, vol. 457, pp. 506–513, Sep. 2016.

[6] A. G. Soriano-Sánchez and C. Posadas-Castillo, "Smart pattern to generate small-world networks," *Chaos, Solitons and Fractals*, vol. 114, pp. 415–422, Sep. 2018.

[7] R. Pastor-Satorras and A. Vespignani, "Immunization of complex networks," *Physical Review E*, vol. 65, no. 3, p. 036104, 2002.

[8] G. Timár, S. N. Dorogovtsev, and J. F. F. Mendes, "Scale-free networks with exponent one," *Physical Review E*, vol. 94, no. 2, p. 022302, 2016.

[9] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[10] R. Albert, H. Jeong, and A.-L. Barabási, "Diameter of the world-wide web," *Nature*, vol. 401, no. 6749, pp. 130–131, Sept. 1999.

[11] B. A. Huberman and L. A. Adamic, "Growth dynamics of the world-wide web," *Nature*, vol. 401, no. 6749, pp. 131–131, Sept. 1999.

[12] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.

[13] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, p. 3200, Apr. 2001.

[14] J. Machowski, J. W. Bialek, and J. R. Bumby, *Power System Dynamics*, 2nd ed. Wiley, 2008.

[15] M. D. Philemon, Z. Ismail, and J. Dare, "A review of epidemic forecasting using artificial neural networks," *International Journal of Epidemiologic Research*, vol. 6, no. 3, pp. 132–143, 2019.

[16] Y. Kryftis, G. Mastorakis, C. X. Mavromoustakis, J. M. Batalla, J. J. P. C. Rodrigues, and C. Dobre, "Resource usage prediction models for optimal multimedia content provision," *IEEE Systems Journal*, vol. 11, no. 4, pp. 2852–2863, Dec. 2017.

[17] A. Korobeinikov and G. C. Wake, "Lyapunov functions and global stability for SIR, SIRS and SIS epidemiological models," *Applied Mathematics Letters*, vol. 15, no. 8, pp. 955–960, Nov. 2002.

[18] 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.

[19] E. O. Nsoesie, J. S. Brownstein, N. Ramakrishnan, and M. V. Marathe, "A systematic review of studies on forecasting the dynamics of influenza outbreaks," *Influenza and Other Respiratory Viruses*, vol. 8, no. 3, pp. 309–316, May 2014.

[20] J. Wang, L. Zhao, and R. Huang, "SIRaRu rumor spreading model in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 398, no. 1, pp. 43–55, Mar. 2014.

[21] Y. Moreno, M. Nekovee, and A. F. Pacheco, "Dynamics of rumor spreading in complex networks," *Physical Review E*, vol. 69, no. 6, p. 066130, Jun. 2004.

[22] H. K. Deva Sarma, R. Mall, and A. Kar, "E^2R^2: Energy-efficient and reliable routing for mobile wireless sensor networks," *IEEE Systems Journal*, vol. 10, no. 2, pp. 604–616, Jun. 2016.

[23] Y. Wang, J. Cao, A. Alofi, A. AL-Mazrooei, and A. Elaiw, "Revisiting node-based sir models in complex networks with degree correlations," *Physica A: Statistical Mechanics and its Applications*, vol. 437, pp. 75–88, 2015.

[24] K. R. Shih and S. J. Huang, "Application of a robust algorithm for dynamic state estimation of a power system," *IEEE Transactions on Power Systems*, vol. 17, no. 1, pp. 141–147, Feb. 2002.

[25] J. Zhao *et al.*, "Power system dynamic state estimation: Motivations, definitions, methodologies, and future work," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3188–3198, Jul. 2019.

[26] W. S. Rosenthal, A. M. Tartakovsky, and Z. Huang, "Ensemble Kalman filter for dynamic state estimation of power grids stochastically driven by time-correlated mechanical input power," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 3701–3710, Jul. 2018.

[27] 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, Jul. 2017.

[28] J. Qi, K. Sun, J. Wang, and H. Liu, "Dynamic state estimation for multi-machine power system by unscented Kalman filter with enhanced numerical stability," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1184–1196, Mar. 2018.

[29] Y. C. Chen, P. E. Lu, C. S. Chang, and T. H. Liu, "A time-dependent SIR model for COVID-19 with undetectable infected persons," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 3279–3294, 2020.

[30] S. Haykin, *Kalman Filtering and Neural Networks*. John Wiley and Sons, New York, 2001.

[31] A. S. Musleh, H. M. Khalid, S. M. Muyeen, and A. Al-Durra, "A prediction algorithm to enhance grid resilience toward cyber attacks in WAMCS applications," *IEEE Systems Journal*, vol. 13, no. 1, pp. 710–719, Mar. 2019.

[32] P. T. Jardine, S. N. Givigi, S. Yousefi, and M. J. Korenberg, "Adaptive MPC using a dual fast orthogonal Kalman filter: application to quadcopter altitude control," *IEEE Systems Journal*, vol. 13, no. 1, pp. 973–981, Mar. 2019.

[33] G. Anagnostou and B. C. Pal, "Derivative-free Kalman filtering based approaches to dynamic state estimation for power systems with unknown inputs," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 116–130, Jan. 2018.

[34] A. Rouhani and A. Abur, "Constrained iterated unscented Kalman filter for dynamic state and parameter estimation," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2404–2414, May 2018.

[35] B. Jia, M. Xin, and Y. Cheng, "Sparse-grid quadrature nonlinear filtering," *Automatica*, vol. 48, no. 2, pp. 327–341, Feb. 2012.

[36] J. Y. Yu, M. J. Coates, M. G. Rabbat, and S. Blouin, "A distributed particle filter for bearings-only tracking on spherical surfaces," *IEEE Signal Processing Letters*, vol. 23, no. 3, pp. 326–330, Mar. 2016.

[37] H. Zhou, Z. Deng, Y. Xia, and M. Fu, "A new sampling method in particle filter based on Pearson correlation coefficient," *Neurocomputing*, vol. 216, no. 5, pp. 208–215, Dec. 2016.

[38] M. Raitoharju, S. Ali-Löytty, and R. Piché, "Binomial Gaussian mixture filter," *EURASIP Journal on Advances in Signal Processing*, vol. 36, pp. 1–18, 2015.

[39] F. Alonge, T. Cangemi, F. D'Ippolito, A. Fagiolini, and A. Sferlazza, "Convergence analysis of extended Kalman filter for sensorless control of induction motor," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2341–2352, Apr. 2015.

[40] D. DeVon, T. Holzer, and S. Sarkani, "Innovation-based fusion of multiple satellite positioning systems for minimizing uncertainty," *IEEE Systems Journal*, vol. 13, no. 1, pp. 928-939, Mar. 2019.

[41] F. Daowang, L. Teng, and H. Z. Tao, "Square-root second-order extended Kalman filter and its application in target motion analysis," *IET Radar, Sonar and Navigation*, vol. 4, no. 3, pp. 329–335, Jun. 2010.

[42] W. Mei, G. Shan, and C. Wang, "Practical development of the second-order extended Kalman filter for very long range radar tracking," *Signal Processing*, vol. 91, no. 5, pp. 1240–1248, 2011.

[43] K. Xiong, H. Y. Zhang, and C. W. Chan, "Performance evaluation of UKF-based nonlinear filteirng," *Automatic*, vol. 42, no. 2, pp. 261–270, Feb. 2006.

[44] G. Valverde and V. Terzija, "Unscented Kalman filter for power system dynamic state estimation," *IET Generation, Transmission and Distribution*, vol. 5, no. 1, pp. 29–37, Jan. 2011.

[45] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.

[46] B. Xu, P. Zhang, H. Wen, and X. Wu, "Stochastic stability and performance analysis of cubature Kalman filter," *Neurocomputing*, vol. 186, no. 19, pp. 218–227, Apr. 2016.

[47] S. Wang, J. Feng, and C. K. Tse, "Spherical simplex-radial cubature Kalman filter," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 43–46, Jan. 2014.

[48] B. Jia, M. Xin, and Y. Cheng, "High-degree cubature Kalman filter," *Automatica*, vol. 49, no. 2, pp. 510–518, Feb. 2013.

[49] B. Zhu, L. Chang, J. Xu, F. Zha, and J. Li, "Huber-based adaptive unscented Kalman filter with non-gaussian measurement noise," *Circuits, Systems, and Signal Processing*, vol. 37, no. 9, pp. 3842–3861, Jan. 2018.

[50] X. Liu, B. Chen, B. Xu, Z. Wu, and P. Honeine, "Maximum correntropy unscented filter," *International Journal of Systems Science*, vol. 48, no. 8, pp. 1607–1615, Jan. 2017.

[51] J. C. Príncipe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. New York, NY, USA: Springer Science and Business Media, 2010.

[52] C. Muscas, P. A. Pegoraro, S. Sulis, M. Pau, F. Ponci, and A. Monti, "New Kalman filter approach exploiting frequency knowledge for accurate PMU-based power system state estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6713–6722, Sept. 2020.

[53] 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*, vol. 33, no. 6, pp. 6109–6119, Nov. 2018.

[54] L. Hu, Z. Wang, I. Rahman, and X. Liu, "A constrained optimization approach to dynamic state estimation for power systems including PMU and missing measurements," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 2, pp. 703–710, Mar. 2016.

[55] S. Wang and J. Liu, "Constructing robust community structure against edge-based attacks," *IEEE Systems Journal*, vol. 13, no. 1, pp. 582–592, Mar. 2019.

[56] R. K. Ahuja, K. Mehlhorn, J. Orlin, and R. E. Tarjan, "Faster algorithms for the shortest path problem," *Journal of the ACM*, vol. 37, no. 2, pp. 213–223, Apr. 1990.

[57] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna, "Four degrees of separation," *Proceedings of the 4th Annual ACM Web Science Conference*, pp. 33–42, Jun. 2012.

[58] S. Edunov, C. Diuk, I. O. Filiz, S. Bhagat, and M. Burke, "Three and a half degrees of separation," *Research at Facebook*, 2016.

[59] S.-H. Yook, H. Jeong, and A.-L. Barabási, "Modeling the internets large-scale topology," *Proceedings of the National Academy of Sciences*, vol. 99, no. 21, pp. 13382–13386, Oct. 2002.

[60] A. Capocci, V. D. P. Servedio, F. Colaiori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli, "Preferential attachment in the growth of social networks: The internet encyclopedia wikipedia," *Physical Review E*, vol. 74, no. 3, p. 036116, Sept. 2006.

[61] E. Eisenberg and E. Y. Levanon, "Preferential attachment in the protein network evolution," *Physical Review Letters*, vol. 91, no. 13, p. 138701, Sept. 2003.

[62] M. Lorenz, R. Ritter, J. Anders, and M. Ortmanns, "Estimation of non-idealities in sigma-delta modulators for test and correction using unscented Kalman filters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 5, pp. 1240–1249, May 2015.

[63] B. Saha, K. Goebel, S. Poll, and J. Christophersen, "Prognostics methods for battery health monitoring using a Bayesian framework," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 2, pp. 291–296, Feb. 2009.

[64] Y. Huang and Y. Zhang, "A new process uncertainty robust Student's t based Kalman filter for SINS/GPS integration," *IEEE Access*, vol. 5, pp. 14391–14404, Jul. 2017.

[65] D. L. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, Aug. 1972.

[66] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[67] C. D. Karlgaard, "Nonlinear regression Huber-Kalman filtering and fixed-interval smoothing," *Journal of Guidance, Control, and Dynamics*, vol. 38, no. 2, pp. 322–330, 2015.

[68] T. Chien-Hao, S. F. Lin, and J. Dah-Jing, "Robust huber-based cubature kalman filter for gps navigation processing," *Journal of Navigation*, vol. 70, no. 3, pp. 527–546, 2016.

[69] L. Chang, B. Hu, G. Chang, and A. Li, "Huber-based novel robust unscented Kalman filter," *IET Science, Measurement and Technology*, vol. 6, no. 6, pp. 502–509, Nov. 2012.

[70] D. Peng, T. Zhou, J. Folkesson, and C. Chao, "Robust particle filter based on Huber function for underwater terrain-aided navigation," *IET Radar, Sonar and Navigation*, vol. 13, no. 11, pp. 1867–1875, 2019.

[71] C. Li, P. Shen, Y. Liu, and Z. Zhang, "Diffusion information theoretic learning for distributed estimation over network," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 4011–4024, Aug. 2013.

[72] W. Ma, B. Chen, J. Duan, and H. Zhao, "Diffusion maximum correntropy criterion algorithms for robust distributed estimation," *Digital Signal Processing*, vol. 58, pp. 10–19, 2016.

[73] B. Chen, L. Xing, and H. Zhao, "Generalized correntropy for robust adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3376–3387, Jul. 2016.

[74] B. Chen, J. Wang, H. Zhao, N. Zheng, and J. C. Príncipe, "Convergence of a fixed-point algorithm under maximum correntropy criterion," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1723–1727, Oct. 2015.

[75] F. Huang, J. Zhang, and S. Zhang, "Adaptive filtering under a variable kernel width maximum correntropy criterion," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 64, no. 10, pp. 1247–1251, Oct. 2017.

[76] Y. Feng, X. Huang, L. Shi, Y. Yang, and J. A. Suykens, "Learning with the maximum correntropy criterion induced losses for regression," *Journal of Machine Learning Research*, vol. 16, pp. 993–1034, Jan. 2015.

[77] W. Ma, H. Qu, G. Gui, L. Xu, J. Zhao, and B. Chen, "Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments," *Journal of the Franklin Institute*, vol. 352, no. 7, pp. 2708–2727, Jul. 2015.

[78] B. Chen, X. Liu, H. Zhao, and J. C. Príncipe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, Feb. 2017.

[79] J. Kivinen, A. J. Smola, and R. C. Williamson, "Online learning with kernels," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2165–2176, Aug. 2004.

[80] X. Liu, H. Qu, J. Zhao, P. Yue, and M. Wang, "Maximum correntropy unscented Kalman filter for spacecraft relative state estimation," *Sensors*, vol. 16, no. 9, p. 1530, 2016.

[81] X. Liu, H. Qu, J. H. Zhao, and P. C. Yue, "Maximum correntropy square-root cubature Kalman filter with application to SINS/GPS integrated systems," *ISA Transactions*, vol. 80, pp. 195–202, 2018.

[82] G. Wang, Y. Zhang, and X. Wang, "Iterated maximum correntropy unscented Kalman filters for non-Gaussian systems," *Signal Processing*, vol. 163, pp. 87–94, May 2019.

[83] F. Ma, J. He, and X. Zhang, "Robust Kalman filter algorithm based on generalized correntropy for ultra-wideband ranging in industrial environment," *IEEE Access*, vol. 7, pp. 27490–27500, 2019.

[84] W. Ma, J. Qiu, X. Liu, G. Xiao, J. Duan, and B. Chen, "Unscented Kalman filter with generalized correntropy loss for robust power system forecasting-aided state estimation," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 11, pp. 6091–6100, Nov. 2019.

[85] L. Chen, H. Qu, and J. Zhao, "Generalized correntropy based deep learning in presence of non-Gaussian noises," *Neurocomputing*, vol. 278, pp. 41–50, Feb. 2018.

[86] W. Qin, X. Wang, and N. Cui, "Maximum correntropy sparse Gauss-Hermite quadrature filter and its application in tracking ballistic missile," *IET Radar, Sonar and Navigation*, vol. 11, no. 9, pp. 1388–1396, Sep. 2017.

[87] L. Chang, B. Hu, G. Chang, and A. Li, "Huber-based novel robust unscented Kalman filter," *IET Science, Measurement and Technology*, vol. 6, no. 6, pp. 502–509, Nov. 2012.

[88] M. E. J. Newman, "The spread of epidemic disease on networks," *Physical Review E*, vol. 66, no. 1, p. 016128, Jul. 2002.

[89] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, pp. 268–276, 2001.

[90] X. Wang and G. Chen, "Complex network: Small-world, scale-free and beyond," *IEEE Circuits and Systems Magazine*, vol. 3, pp. 6–20, Jan. 2003.

[91] A. Korobeinikov, "Lyapunov functions and global stability for SIR and SIRS epidemiological models with non-linear transmission," *Bulletin of Mathematical biology*, vol. 68, pp. 615–626, 2006.

[92] H. Serrano-Guerrero, C. Cruz-Hernández, R. M. López-Gutiérrez, L. Cardoza-Avendaño, and R. A. ChávezPérez, "Chaotic synchronization in nearest-neighbor

coupled networks of 3D CNNs," *Journal of Applied Research and Technology*, vol. 11, no. 1, pp. 26–41, Feb. 2013.

[93] L. Sun, B. Zheng, J. Zhou, and H. Yan, "Some inequalities for the Hadamard product of tensors," *Linear and Multilinear Algebra*, vol. 66, no. 6, pp. 1199–1214, 2018.

[94] G. Feng, W. Wu, and J. Wang, "Observability analysis of a matrix Kalman filter-based navigation system using visual/inertial/magnetic sensors," *Sensors*, vol. 12, pp. 8877–8894, 2012.

[95] B. Southall, B. F. Buxton, and J. A. Marchant, "Controllability and observability: Tools for Kalman filter design," *Proceedings of British Machine Vision Conference*, vol. 1, pp. 164–173, 1998.

[96] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, pp. 175–308, Feb. 2006.

[97] C. Liu, J. Liu, and Z. Jiang, "A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2274–2287, Dec. 2014.

[98] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.

[99] X. Liu, L. Li, Z. Li, T. Fernando, and H. H. C. Iu, "Stochastic stability condition for the extended kalman filter with intermittent observations," *IEEE Transactions On Circuits and Systems II: Express Briefs*, vol. 64, no. 3, pp. 334–338, Mar. 2017.

[100] W. Li, G. Wei, D. Ding, Y. Liu, and F. E. Alsaadi, "A new look at boundedness of error covariance of Kalman filtering," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 2, pp. 309–314, Feb. 2018.

[101] K. Jiang, H. Zhang, H. R. Karimi, J. Lin, and L. Song, "Simultaneous input and state estimation for integrated motor-transmission systems in a controller area network environment via an adaptive unscented Kalman filter," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 4, pp. 1570–1579, Apr. 2020.

[102] 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.

[103] W. Li, S. Sun, Y. Jia, and J. Du, "Robust unscented Kalman filter with adaptation of process and measurement noise covariances," *Digital Signal Processing*, vol. 48, pp. 93–103, Jan. 2016.

[104] B. Jia, M. Xin, and Y. Cheng, "Sparse Gauss-Hermite quadrature filter with application to spacecraft attitude estimation," *Journal of Guidance, Control, and Dynamics*, vol. 34, no. 2, pp. 367–379, Mar. 2011.

[105] X. Liu, Z. Ren, H. Lyu, Z. Jiang, P. Ren, and B. Chen, "Linear and nonlinear regression-based maximum correntropy extended Kalman filtering," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, early access, Jun. 2019, doi: 10.1109/TSMC.2019.2917712.

[106] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge: Cambridge University Press, 2013.

[107] B. Jia and X. Ming, "Vision-based spacecraft relative navigation using sparse-grid quadrature filter," *IEEE Transactions on Control Systems Technology* , vol. 21, no. 5, pp. 1595–1606, Sep. 2013.

[108] B. Jia, M. Xin, and Y. Cheng, "Relations between sparse-grid quadrature rule and spherical-radial cubature rule in nonlinear Gaussian estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 1, pp. 199-204, Jan. 2015.

[109] B. Jia and M. Xin, "Sparse-grid quadrature H_∞ filter for discrete-time systems with uncertain noise statistics," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 3, pp. 1626–1636, Jul. 2013.

[110] W. Wang, S. Wang, and C. K. Tse, "Dynamic state estimation of power systems by p-norm nonlinear Kalman filter," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 5, pp. 1715–1728, May 2020.

[111] W. Wang, S. Wang, and C. K. Tse, "Performance comparison of nonlinear Kalman filters in epidemic tracking on networks," *IEEE Systems Journal*, vol. 14, no. 4, pp. 5475–5485, Dec. 2020.

[112] Y. Huang, Y. Zhang, P. Shi, Z. Wu, J. Qian, and J. A. Chambers, "Robust Kalman filters based on Gaussian scale mixture distributions with application to target tracking," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 10, pp. 2082–2096, Oct. 2019.

[113] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2003.

[114] G. He, S. Dong, J. Qi, and Y. Wang, "Robust state estimator based on maximum normal measurement rate," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2058–2065, Nov. 2011.

[115] J. Qi, G. He, S. Mei, and Z. Gu, "A review of power system robust state estimation," *Advanced Technology of Electrical Engineering and Energy*, vol. 30, no. 3, pp. 59–64, Jul. 2011.

[116] J. H. Kotecha and P. M. Djuric, "Gaussian sum particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2602–2612, Oct. 2003.

[117] O. Hugues-Salas and K. A. Shore, "An extended Kalman filtering approach to nonlinear time-delay systems: application to chaotic secure communications," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 9, pp. 2520–2530, Sep. 2010.

[118] Z. Zhu and H. Leung, "Adative blind equalization for chaotic communication systems using extended-Kalman filter," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 8, pp. 979–989, Aug. 2001.

[119] D. M. Walker and M. Small, "Detecting unstable fixed points using Kalman filters with constraints," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, no. 12, pp. 2818–2827, Dec. 2006.

[120] M. B. Luca, S. Azou, G. Burel, and A. Serbanescu, "On exact Kalman filtering of polynomial systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, no. 6, pp. 1329–1340, Jun. 2006.

[121] R. van der Merwe and E. Wan, "The square-root unscented Kalman filter for state and parameter estimation," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3461–3464, May 2001.

[122] Y. Li, H. R. Karimi, M. Zhong, S. X. Ding, and S. Liu, "Fault detection for linear discrete time-varying systems with multiplicative noise: The finite-horizon

case," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 10, pp. 3492–3505, Oct. 2018.

[123] Y. Li, S. Liu, M. Zhong, and S. X. Ding, "State estimation for stochastic discrete-time systems with multiplicative noises and unknown inputs over fading channels," *Applied Mathematics and Computation*, vol. 320, pp. 116–130, Mar. 2018.

[124] C. C. Pugh, *Real Mathematical Analysis*. New York: SpringerVerlag, 2002.

[125] J. Qi, K. Sun, and W. Kang, "Optimal PMU placement for power system dynamic state estimation by using empirical observability gramian," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 2041–2054, Jul. 2015.

[126] P. Kunder, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill, 1994.

[127] J. Qi, A. F. Taha, and J. Wang,"Comparing Kalman filters and observers for power system dynamic state estimation with model uncertainty and malicious cyber attacks," *IEEE Access*, vol. 6, pp.77155–77168, Dec. 2018.

[128] Y. Yu, Z. Wang, and C. Lu, "A joint filter approach for reliable power system state estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 1, pp. 87–94, Jan. 2018.

[129] 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.

[130] M. Ariff, B. Pal, and A. Singh, "Estimating dynamic model parameters for adaptive protection and control in power system," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 829–839, Mar. 2015.

[131] Y. Li, J. Li, J. Qi, and L. Chen, "Robust cubature kalman filter for dynamic state estimation of synchronous machines under unknown measurement noise statistics," *IEEE Access*, vol. 7, pp. 29139–29148, 2019.

[132] A. Srinivasan, P. Venkatesh, B. Dineshkumar, and N. Ramkumar, "Dynamic available transfer capability determination in power system restructuring environment using support vector regression," *International Journal of Electrical Power and Energy Systems*, vol. 69, pp. 123–130, Jul. 2015.

[133] Y. Zhou, H. Sun, Q. Guo, B. Xu, J. Wu, and L. Hao, "Data driven method for transient stability prediction of power systems considering incomplete measurements," *Proceedings of the IEEE Energy Internet and Energy System Integration (EI2)*, Beijing, China, pp. 1–6, Nov. 2017.

[134] Z. Miao and L. Fan, "Achieving economic operation and secondary frequency regulation simultaneously through local feedback control," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 85–93, Jan. 2017.

[135] S. Nugroho, A. Taha, and J. Qi, "Robust dynamic state estimation of synchronous machines with asymptotic state estimation error performance guarantees," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1923–1935, May 2020.

[136] J. Chow and G. Rogers, User manual for power system toolbox, Version 3.0, 1991–2008.

[137] W. Ma, H. Qu, G. Gui, L. Xu, J. Zhao, and B. Chen, "Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments," *Journal of the Franklin Institute*, vol. 352, no. 7, pp. 2708–2727, Jul. 2015.

[138] L. Li, T.-S. Qiu, and D.-R. Song, "Parameter estimation based on fractional power spectrum under alpha-stable distribution noise environment in wideband bistatic MIMO radar system," *International Journal of Electronics and Communications*, vol. 67, no. 11, pp. 947–954, Nov. 2013.

[139] M.H. Bibalan, H. Amindavar, and M. Amirmazlaghani, "Characteristic function based parameter estimation of skewed alpha-stable distribution: an analytical approach," *Signal Processing*, vol. 130, pp. 323–336, Jan. 2017.

[140] P. Chopade and J. Zhan, "Structural and functional analytics for community detection in large-scale complex networks," *Journal of Big Data*, vol. 2, no. 1, pp. 1–28, Jul. 2015.

[141] P. M. Zadeh and Z. Kobti, "A multi-population cultural algorithm for community detection in social networks," *Procedia Computer Science*, vol. 52, pp. 342–349, Jan. 2015.

[142] B. Amiri, L. Hossain, J. W. Crawford, and R. T. Wigand, "Community detection in complex networks: Multi–objective enhanced firefly algorithm," *Knowledge-Based Systems*, vol. 46, pp. 1–11, Jul. 2013.

[143] C. Pizzuti and S. E. Rombo, "Algorithms and tools for protein-protein interaction networks clustering, with a special focus on population-based stochastic methods," *Bioinformatics*, vol. 30, no. 10, pp. 1343–1352, May 2014.

[144] D. Cai, X. He, J. Han, and T. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[145] L. Yang, X.-C. Cao, D. Jin, X. Wang, and D. Meng, "A unified semisupervised community detection framework using latent space graph regularization," *IEEE Transactions on Cybernetics*, vol. 45, no. 11, pp. 2585–2598, Nov. 2015.

[146] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 2, pp. 257–265, Jun. 2018.

[147] X. Li, J. Fang, W. Cheng, H. Duan, Z. Chen, and H. Li, "Intelligent power control for spectrum sharing in cognitive radios: A deep reinforcement learning approach," *IEEE Access*, vol. 6, pp. 25463–25473, Apr. 2018.

[148] J. W. Xu, A. R. C. Paiva, I. Park, and J. C. Príncipe, "A reproducing kernel Hilbert space framework for information-theoretic learning," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 5891–5902, Dec. 2008.

[149] B. Chen, S. Zhao, P. Zhu, and J. C. Príncipe, "Quantized kernel least mean square algorithm," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 1, pp. 22–32, Jan. 2012.

[150] S. Wang, J. Feng, and C. K. Tse, "Kernel affine projection sign algorithms for combating impulse interference," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 60, no. 11, pp. 811–815, Nov. 2013.

[151] B. Chen, S. Zhao, P. Zhu, and J. C. Príncipe, "Quantized kernel recursive least squares algorithm," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 9, pp. 1484–1491, Sep. 2013.