State Estimation—Beyond Gaussian Filtering

Haozhan Meng
mzhan@uno.edu

State Estimation—Beyond Gaussian Filtering

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Engineering and Applied Science

by

Haozhan Meng

B.S. Xian Jiaotong University, 2010
M.S. University of New Orleans, 2019

May, 2022

Acknowledgements

I would like to express my sincerest appreciation to those without whom this dissertation would not have been possible.

To my parents, my aunt, my cousin, and my girlfriend, who continuously give me emotional and financial support throughout my research.

To my advisor, Prof. X. Rong Li, who is erudite in research, profound in critical thinking, and guides me to enjoy the beauty of the research world.

To Prof. Vesselin Jilkov and Prof. Huimin Chen, who provide me help and guidance during my research.

To other committee members, Prof. Linxiong Li and Prof. Xiaochuan Yu, for their efforts to improve the quality of my dissertation.

To my colleagues, Dr. Yu Liu, Dr. Reza Rezaie and Ms. Xiaolei Bian, without whom, my life during the Ph.D. study at University of New Orleans would become much harder.

Last but not the least, I want to thank all my wonderful friends at University of New Orleans, and share all my accomplishment with them.

This dissertation is dedicated to my parents.

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Nomenclature

$\mathbb{R}$	Set of real numbers
$\mathbb{R}^n$	n-dimension Euclidean space
$\mathcal{N}(x; u, C)$	Gaussian random variable x with mean u and covariance C
$\Phi_n(x; u, \Sigma)$	cdf of Gaussian random variable $\mathcal{N}(x; u, C)$
$z^n = \{z_1, \ldots, z_n\}$	A finite sequence (column vector) of length n starting at 1
$L(x \mid z)$	Likelihood function of x given data z
$\mathrm{vec}(\cdot)$	Conversion of a matrix to a column vector by consecutively stacking its columns.
$\mathrm{diag}(\cdot)$	Block diagonal operator
I	Identity matrix of appropriate dimension
$\otimes$	Kronecker product
$A^{\otimes k}$	$A \otimes \cdots \otimes A$ (k times).

Acronyms

All mathematical symbols used below is valid throughout this dissertation.

LMMSE	Linear minimum mean-squares error
MMSE	Minimum mean-squares error
PDF	Probability density function
CDF	Cumulative distribution function

GHQ	Gaussian-Hermite quadrature
SGHQ	Sparse-grid Gaussian-Hermite quadrature
SSR	Skew-symmetric representation
SGF	Skew-Gaussian filter
FIM	Fisher information matrix
CRLB	Cramer-Rao lower bound
FOSG	First-order skew-Gaussian
CGF	Cumulant generating function

Abstract

This dissertation considers the state estimation problems with *symmetric* Gaussian/*asymmetric* skew-Gaussian assumption under linear/nonlinear systems. It consists of three parts. The first part proposes a new recursive finite-dimensional exact density filter based on the linear skew-Gaussian system. The second part adopts a skew-symmetric representation (SSR) of distribution for nonlinear skew-Gaussian estimation. The third part gives an optimized Gauss-Hermite quadrature (GHQ) rule for numerical integration with respect to Gaussian integrals, and applies it to nonlinear Gaussian filters.

We first develop a linear system model driven by skew-Gaussian process and present the exact filter for the posterior density with fixed dimensional recursive representation, i.e., the skew-Gaussian filter (SGF). The SGF not only has an analytical recursion of a small dimension akin to the Kalman filter, but also possesses an efficiency comparable to the Kalman filter. The minimum mean-square error (MMSE) estimator based on our proposed skew-Gaussian filter is demonstrated via a simulation study.

Next, we propose a skew-symmetric presentation of the posterior density to handle the discrete-time filtering problem for a nonlinear system driven by non-Gaussian process. The skew-symmetric representation of distributions, which has a product form of a symmetric pdf (known as the base pdf) times a perturbation function (known as the skewing function), is employed in this dissertation. Based on a first-order skew-symmetric representation of Gaussian distribution, we propose the first-order skew-Gaussian filter (FOSGF) and demonstrate it by applications to the radar tracking problem.

For the filtering problem where Gaussian integrals are adopted in the state update, we propose an new set of Gauss-Hermite quadrature rules using an optimized proposal density. The optimized GHQ rule, proposed in this dissertation, finds an optimized way to improve GHQ-based Gaussian integration when the integrand is not close to a polynomial by transforming it to one approximated by a polynomial. The solution is formulated as a nonlinear least-squares problem with linear constraints. Several numerical examples based on the optimized GHQ rule are studied and compared with the traditional methods.

Keywords: Bayesian estimation, exact filtering, nonlinear filter, skew-Gaussian filtering, first-order skew-Gaussian filtering, Gauss-Hermite quadrature.

1

Introduction

> It is not knowledge, but the act of learning, not possession, but the act of getting there, which grants the greatest enjoyment
>
> Carl Friedrich Gauss

This dissertation focuses on the nonlinear filter for a dynamic systems with a state-space model within Bayesian estimation framework. We consider linear system driven by skew-Gaussian process as well as nonlinear system driven by non-Gaussian process and present three types of nonlinear filters to tackle the nonlinearity in system dynamics, measurements model and the non-Gaussianness in the statistcal model of the noise.

1.1 State Estimation in Discrete Time

Estimation theory may trace back to 1795 when Carl Friedrich Gauss, who is accredited to the method of least squares, needed to predict the motions of planets and comets from telescopic measurements (Mendel, 1995). Until 1960, the Kalman filter

(Kalman, 1960), one of very few breakthroughs in estimation theory, marked the beginning of modern estimation theory. As an optimal recursive state estimator, the Kalman filter is computationally efficient and provides insight into the fundamentals of the state estimation. This dissertation is devoted to the state-space approach to filtering in discrete time.

In general, estimation theory can be viewed as the main core of data (or information) processing to the design of computer-implemented filters that process data with uncertainty in an optimal manner. It is at the intersection of the three disciplines, i.e., information processing, signals and systems, and statistics (Li, 2015). The applications of estimation theory have been quite successful in the real world, especially in many physical systems where systems and signals play a role, and its scope of application is still expanding (Van Trees, 2004; Bar-Shalom et al., 2004; Singh et al., 2021).

The state estimation is the (sequential or recursive) state filtering of a (unobserved) dynamic system from (observed) noisy measurements, as shown in Figure 1.1 for ideal situations where observations are received at every time instant of interest. A general state-space model in discrete time, which can be discretized from a continuous-time dynamic model, is of the form (Arasaratnam et al., 2007)

$$x_k = f(x_{k-1}, w_{k-1}, k-1) \tag{1.1}$$

$$z_k = h(x_k, v_k, k) \tag{1.2}$$

A special case of (1.1) and (1.2) is the one with additive noise

$$x_k = f(x_{k-1}, k-1) + w_{k-1} \tag{1.3}$$

$$z_k = h(x_k, k) + v_k \tag{1.4}$$

where $x_k \in \mathbb{R}^{n_x}$ and $z_k \in \mathbb{R}^{n_z}$ are the state and measurement of the system at time k, $\{w_k\}$ is the process noise that drives the dynamic system through the state transition

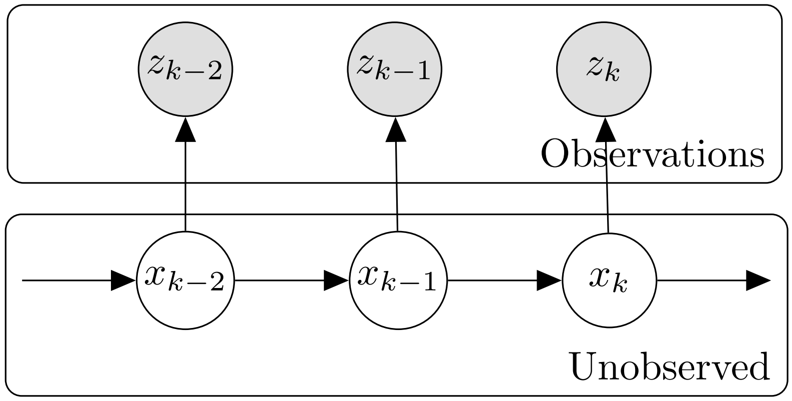

FIGURE 1.1: Ideal state-space model

function f, and $\{v_k\}$ is the measurement noise corrupting the observation of the state through the measurement function h.

The filtering problem is nonlinear if either the system dynamics f or the measurement mode h is nonlinear.

Usually, the noise processes $\{w_k\}$ and $\{v_k\}$ are assumed to be white, mutually independent, and independent of the initial state x_0, which make the state process $\{x_k\}$ a Markov sequence, and so a probabilistic solution may exist.

Depending on the different types of estimation results, i.e., a single value or the whole density of the state, the state estimation consists of point estimation and density estimation, which will be discussed later.

1.2 Optimal Solution under Bayesian Framework

The Bayesian framework is the most commonly used approach to studying the state estimation problem (Ho and Lee, 1964). Following this framework, the optimal estimation solution is closely related to the calculation of the conditional pdf $p(x_k|z^k)$ of x_k given all observations up to time k, i.e., $z^k = \{z_1, \dots, z_k\}$.

Often, the two noise processes $\{w_k\}$ and $\{v_k\}$ are assumed to be independent, and it turns out that the state $\{x_k\}$ is a partially observed Markov process with initial probability density $p(x_0)$ and transition densities $p(x_k|x_{k-1})$, and the observation z_k at time k is conditionally independent of the previous observations z^{k-1} (Li, 2015).

The Bayesian approach for the Markov process x_k can be accomplished as follows:

1. Predict by the Chapman-Kolmogorov equation:

$$p(x_k|z^{k-1}) = \int_{\mathbb{R}^{n_x}} p(x_k|x_{k-1})p(x_{k-1}|z^{k-1})dx_{k-1} \tag{1.5}$$

2. Update by Bayes' rule

$$p(x_k|z^k) = \frac{p(z_k|x_k)p(x_k|z^{k-1})}{\int_{\mathbb{R}^{n_x}} p(z_k|x_k)p(x_k|z^{k-1})dx_k} \tag{1.6}$$

If the Markovian property of x_k and the conditional independence of z_k does not hold, the Bayesian approach is as follows.

$1'$ Predict

$$p(x_k|z^{k-1}) = \int_{\mathbb{R}^{n_x}} p(x_k|x_{k-1}, z^{k-1})p(x_{k-1}|z^{k-1})dx_{k-1} \tag{1.7}$$

$2'$ Update by Bayes' rule

$$p(x_k|z^k) = \frac{p(z_k|x_k, z^{k-1})p(x_k|z^{k-1})}{\int_{\mathbb{R}^{n_x}} p(z_k|x_k, z^{k-1})p(x_k|z^{k-1})dx_k} \tag{1.8}$$

For point estimation, usually the first two moments of $p(x_k|z^k)$ are of interest, that is, the minimum mean-square error (MMSE) estimator of the following

$$\hat{x}_{k|k} = E(x_k|z^k) = \int_{\mathbb{R}^{n_x}} x_k p(x_k|z^k)dx_k$$

$$\text{MSE}(\hat{x}_{k|k}) = P_{k|k} = \int_{\mathbb{R}^{n_x}} (x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})' p(x_k)dx_k$$

The MMSE estimator turns out to minimize not only the mean-square error of the estimator, but also a large class of the Bayesian risks (see Section 2.2).

For density estimation, the posterior density $p(x_k|z^k)$ in (1.6) is sought after. However, to obtain it more resources are required, both technically and computationally, than point estimation, but it provides a complete solution to many state estimation problems (Li and Jilkov, 2010).

1.3 Exact Filtering

Exact filtering delineates a class of filtering problems of which the computation of the desired solution, i.e., the posterior density, is exact. Thus, the exact filtering, more precisely the exact density filtering, simply denotes to those estimation problems whose posterior densities can be obtained analytically.

1.3.1 Importance of Exact Filtering

Exact filter is crucial for many engineering problems. Not only is it theoretically solid, but it also has potential to provide the guidance to approximate nonlinear filtering if no exact filtering solution is available. The most well-known example is the linear systems with *Gaussian assumption*[1] (also known as linear Gaussian estimation). It has a wide application in the real world, and in return, contributes to the development of state estimation theory. The Kalman filter, being the solution to the linear Gaussian system, is optimal not only with respect to point estimation (e.g., in the MMSE sense), but also with is exact respect to density estimation.

The Kalman filter, an exact density filter for linear Gaussian systems, has many merits that bring a lot of insight, intuition, and impetus to the state estimation theory, e.g., it is recursive, finite dimensional, and optimal in almost all statistical senses. As will be given in Section 2.4, many concepts/properties of state estimation theory, e.g., the exactness, the recursitiveness, the finite-dimensionality, etc., deduce

[1] (1.1) and (1.2) (more specifically, (1.3) and (1.4)) are both linear, and the initial state x_0, the process noise $\{w_k\}$, and the measurement noise $\{v_0\}$ are mutually independent and Gaussian distributed.

from the research of the Kalman filtering theory, and such concepts/properties, which are also crucial in practice, enlighten further research, and one of the examples is the results of the recursibility for the linear minimum mean-square error (LMMSE) estimator (Li, 2004).

In nonlinear systems, the exact filtering provides the theoretical foundation of nonlinear filtering approximate methods, for example, the Gaussian filtering is inspired by the Kalman filter. The Gaussian filter, being the density-assumed filter, assumes the distribution of the state and the measurement in (1.3) and (1.4) are jointly Gaussian. Despite being a strong assumption in reality, it still works well in many practical problems. Another important topic of state estimation theory that gets inspired by the exact filtering (specially, the Kalman filter) is the LMMSE estimation, which is still the workhorse of many nonlinear state estimation problems, and achieves a good trade-off between the performance and computation efficiency in applications where the nonlinearity is moderate. A textbook that discusses in detail the LMMSE estimation is (Anderson and Moore, 2005), and this dissertation also gives a brief review of the LMMSE estimation in Section 2.3.

In summary, exact filers have potential to deepen the understanding of the state estimation and may guide the development of new approaches to nonlinear approximate methods. Even though it is arduous, the exact filtering plays a fundamental role in state estimation and is worth further research.

1.3.2 Existing Research of Exact Filtering

Although it is important and deserves more research, an exact filter rarely exists beyond the linear Gaussian world (Daum, 2005). A vivid description of the nonlinear filtering research is given by (Li and Jilkov, 2010):

> In the garden of nonlinear filtering (NLF), there are three trees: continuous-time NLF, discrete-time NLF, and mixed-time NLF. The tree of continuous-

> time NLF is spectacular: Its theoretical flowers blossom with little watering by practical problems, but few of the flowers have turned into application fruits. Most gardeners here care only about the flowers but not the fruits. ... Many of these gardeners are now watering plants in other gardens or lands. In contrast, the tree of discrete-time NLF grows many application fruits but few theoretical flowers. The tree of mixed-time NLF has been largely neglected by most gardeners. As a result, it is smallest and it deserves much better care.

As quoted above, most theoretical results of exact nonlinear filtering are limited only in continuous time, not to mention that existing theoretical results in continuous time, except for the linear Gaussian case, cannot be directly applied nor provide much guidance to implements, which results in a gap between estimation theory and practical applications. Exact filtering in continuous time is beyond the scope of this dissertation, and readers who are interested in the continuous case may refer to (Li and Jilkov, 2010).

Even though exact filtering in discrete time is more appealing for computer-based implementation and more easily understood by researchers, fewer theoretical results exist in discrete time than in continuous time, not to mention that the exact filter often goes with other requirements, e.g., the finite-dimensionality, due to the practical needs, which further makes it more difficult to exist.

For exact fixed finite-dimensional filtering in discrete time, there are very few results, and the two cases that are widely used are the Kalman filter for linear Gaussian models and the discrete-time Wonham filter (Rabiner, 1989) for finite state Markov chains. Further results along these two models can be found in (Elliott and Krishnamurthy, 1999; Krishnamurthy and Evans, 1998). Recent research includes (Rezaie and Li, 2020), who solved the filtering problem on the Gaussian conditional

Markov sequence instead of the Gauss-Markov sequence (Rezaie and Li, 2019).

There are many results on exact filtering, but very few (explicit) exact filters have found of exact filtering except a few noteworthy papers in proving the existence of the exact fixed finite-dimensional filter in discrete time. Sawitzki (1979) discussed the existence of the exact fixed finite-dimensional filter in the exponential family, along with its density propagation. In (Ferrante and Runggaldier, 1990; Ferrante, 1992; GÜnther, 1981), the authors provided a necessary and sufficient condition of the existence of an exact fixed finite-dimensional filter under the following regularities: i) the state $\{x_k\}$ is Markov with an initial density $p(x_0)$ and transition densities $p(x_k|x_{k-1})$, ii) the observation z_k and the state x_k are related by the likelihood function $p(z_k|x_x)$, the posterior density and the likelihood function must be from the exponential family if an exact fixed finite-dimensional filter exists.

Based on the theoretical results above, conjugate families play a fundamental role in exact fixed finite-dimensional filters. Along this direction, some filters have been proposed, e.g., (Ferrante and Giummolé, 1995; Girón and Rojano, 1994; Benavoli et al., 2020), even though such construction is usually ad-hoc and hardly has any application. One of the classes that finds application in the robust estimation is the exact filtering in the elliptical distributions (Girón and Rojano, 1994). It has a similar propagation strategy as the Kalman filter by assuming the state and two noises are jointly elliptical distributed with no correlation involved. However, these three components are generally dependent unless Gaussian distributed (Kollo, 2005).

By allowing the dimension of the exact filter not being fixed, some exact but increasing finite-dimensional filters can be found in some special classes, e.g., the state estimator problem for linear systems in Gaussian noise with the coefficients being functions of a finite-state Markov chain (Elliott et al., 1996). More recent work on the exact filtering with increasing dimension is related to the application to the skew-Gaussian distribution, e.g., (He et al., 2018; Rezaie and Eidsvik, 2014);

however, both suffer a lot from the increasing of dimensionality over time.

In recent years, research of seeking for exact filtering grows in many areas, one example is the extended object tracking. Models which can describe the state dynamics as well as the object extension are explored intensely. Some of the relevant work are dealing with the state dynamic with random matrix whose form being selected deliberately (Koch, 2008; Feldmann et al., 2010; Lan and Li, 2016), so that the joint state and extension, being modeled as the inverse-Wishart Gaussian distribution, can be propagated exactly. However, no such expected result is obtained. Recent work that affords a closed-form prediction of the joint state and extension is (Bartlett et al., 2020) by employing a non-central inverse Wishart distribution to model the state transition density of the target extent.

1.4 Approximate Filtering

Since exact nonlinear filters usually do not exist, numerous approximate solutions to the Bayesian state estimation problems have been proposed over the last few decades. These approaches, depending on either density filtering or point filtering, can be classified into the following groups (Li and Jilkov, 2012, 2004), and some of them will be elaborated in more details later:

1.4.1 Approximate Density Filtering

- **Sequential Monte Carlo methods:** The sequential Monte Carlo method is a set of simulation-based methods to deal with densities without making explicit assumption (Doucet et al., 2001). Unlike the grid based methods whose grid is predetermined or adaptive, these methods employ a random sampling strategy (the importance sampling) to approximate the filtering density. These approaches suffer from the curse of dimensionality and the degeneration phenomenon that all sampling points will collapse to a single point after some time. Careful design of the density representation points as opposed to random particles includes the feedback particle filter (Taghvaei and Mehta, 2021) and the ensemble particle filter (Evensen, 1994).

- **Grid-based methods:** Grid-based methods rely heavily on general-purpose numerical techniques. By approximating the true density using sepcial function over a uniformly distributed grid located at the "area of interest", these methods evolve along with the time-consuming computation of convolution, and the grid design. Their performance is guaranteed by the theory of probability, that is, the true density can be approached by the pointwise limit of a monotonic increasing sequence of non-negative simple functions (Chung, 2001). For more details, see, e.g., the point mass method (Šimandl et al., 2006).

- **Spectral methods:** Spectral methods approximate the density by a linear combination of some basis functions from a functional space. Usually the basis is chosen by some representation theory, e.g., the generalized Edgeworth series expansion (Challa et al., 2000), and Wiener chaos expansion (Lototsky, 2006). Their estimation performance is guaranteed by the approximation theory in an inner product space. However, the required number of terms is usually large in order to achieve the desirable accuracy.

- **Parametric family approximation:** Parametric family methods approximate the filtering density at each time by a member or a subset of a parametric family of distributions, then the density filtering problem is simplified to estimate the finite-dimensional parameter of the family (Kushner, 1967). Their performance may be achieved by minimizing some density measures, e.g., the Kullback-Leibler divergence. Popular methods include variational Bayesian methods (Smidl and Quinn, 2008), Gaussian sum methods (Alspach and Sorenson, 1972), and reduced statistics method (Kulhavỳ, 1996).

- **Optimal interpolation:** Inspired by the deterministic interpolation for function approximation, one can approximate a probability density function by some interpolation methods, e.g., the B-splines (He et al., 2014). These methods may use fewer knots while maintain the same estimation accuracy and do not have degeneracy problems as the sequential Monte Carlo methods do.

1.4.2 Approximate Point Filtering

- **Function approximation:** Approximating a nonlinear function by its linear counterpart is a natural idea to solve the nonlinear state estimation problem. The conventional method is the first order Taylor series expansion, which results in the extended Kalman filter. However, in the situation where some higher terms cannot be negligible, iterated extended Kalman filter (Denham and Pines, 1966) and second-order extended Kalman filters (e.g., (Henriksen, 1982; Jazwinski, 2007)) were given to compensate the performance loss. Note that there always exist cases that this technique fail or even diverge.

- **Sampling-based moment approximation:** Sampling-based methods approximate the quantities of interest (usually, the mean and covariance) directly. Unlike the Monte Carlo method, the moment approximation techniques include: the Gauss-Hermite filter (GHF) based on the Gauss-Hermite quadrature (GHQ) (Ito and Xiong, 2000; Arasaratnam et al., 2007), the unscented filter (UF) using the unscented transformation (UT) (Julier and Uhlmann, 2004), the cubature Kalman filter (CKF) by the cubature rule (Arasaratnam and Haykin, 2009), etc. They are either restricted to Gaussian assumptions or reliant on the LMMSE update at the cost of neglecting potentially effective information about higher moments.

- **Stochastic model approximation:** Stochastic model approximation methods are a promising way to solve nonlinear state estimation problem by considering both the accuracy and the efficiency. Usually an optimal linear stochastic model is employed to accomplish this task. It turns out that the deterministic sampling methods belong to stochastic linear model approximation (Lefebvre et al., 2002).

1.5 Gaussian Filtering

In this section, nonlinear filtering with Gaussian assumption is introduced neglecting the linear case (since this result is well-known). Unless stated otherwise, the terms "Gaussian filtering (or filters)" refers to filtering (or filters) with Gaussian assumption in nonlinear systems in this dissertation.

The Gaussian filter rests on two assumptions: i) the conditional state probability density $p(x_k|z^{k-1})$ at each step k is assumed to be Gaussian; ii) the conditional state and measurement probability density $p(x_k, z_k|z^{k-1})$ at each step k is also jointly Gaussian. Based on these two assumptions, nonlinear estimation turns out to be a successive approximation of the Gaussian densities by moment-matching, and some efficient numerical integration methods are required for recursive filtering.

Based on different integration methods, different Gaussian filters have been proposed. The implementation of the Gaussian-Hermite quadrature (GHQ) by the tensor rule was used early in (McReynolds, 1975), and rediscovered in (Ito and Xiong, 2000). A sparse-grid version of GHQ was proposed in (Jia et al., 2012). In (Arasaratnam and Haykin, 2009), the Gaussian cubature integration was used tde in the spherical-radial coordinate system instead of the Cartesian coordinate system. Higher Gaussian cubature rules were discussed in (Jia et al., 2014). However, performance evaluation of these numerical integration is still inadequate and a discussion by some numerical studies given in (Wu et al., 2006). More details will be discussed in Chapter 5.

1.6 Skew-Gaussian Filtering

Even though Gaussian filters provide a good compromise between estimation accuracy and computation efficiency, the Gaussian assumption is insufficient for many nonlinear filtering problems, especially in situations where information of higher moments (such as skewness and kurtosis) is not negligible.

In this section, the state estimation with skew-Gaussian assumption is introduced in linear/nonlinear systems.

1.6.1 Skew-Gaussian Filtering in Nonlinear Systems

In general, even nonlinear transformation (i.e., (1.1), (1.2) or both) of Gaussian densities often results in asymmetry and distortion of other nice properties of the posterior (Julier, 1998), which always invalidates the Gaussian assumption if the Markov property of $\{x_k\}$ is assumed (Rezaie and Li, 2019). So, there is room for improvement if information about higher moments is incorporated in nonlinear estimation.

In order to model more complex nonlinear problems with some simple tractable form beyond the Gaussian assumption, especially the skewness, a skew-symmetric representation of distributions was proposed in the statistical community, see, e.g., (Genton, 2004). By multiplying a perturbation function (known as the skewing function), the skew-symmetric representation has potential to transfer any symmetric distribution (known as the base pdf) to a skewed one. Moreover, such a representation can approximate any density at any desirable accuracy (Ma and Genton, 2004).

Based on the skew-symmetric representation of distributions, the first-order skew-Gaussian filter is presented with application to a tracking problem. This skew-Gaussian type filter can capture the departure from the symmetry of Gaussian densities for practical problems, and may attain a reasonable compromise between mathematical tractability and shape flexibility of the filtering pdf. More details will be presented in Chapter 4.

1.6.2 Skew-Gaussian Filtering in Linear Systems

Linear systems with non-Gaussian state/output, especially the skewed ones, occur in practice (Huang et al., 2017b; Wu, 1993; Wu and Chang, 1996). However, for problems where skewness exists in linear systems, the Kalman filter does not perform optimally due to the violation of the linear Gaussian assumption. As such, a linear skew-Gaussian system where the state, the process noise, and the measurement noise are skew-Gaussian is proposed in Chapter 3.

The linear skew-Gaussian system can model skewness caused by linear hidden truncation, and its filtering result has an exact recursive finite-dimensional filtering form, i.e., the SGF. Like the linear Gaussian estimation, the linear skew-Gaussian estimation provides a theoretical and applicable approach to the state estimation with skewness.

By incorporating additional parameters to model the skewness, the linear skew-Gaussian estimation not only includes the linear Gaussian estimation as a special case, but also provides one closed-form exact filter to state estimation with higher moments being considered. For example, in nonlinear filtering where the skewness is considered, the filtering results can be applied by linearization methods e.g., the first-order Taylor expansion, the statistical linearization, etc.

1.7 Research Motivation

State estimation becomes difficult beyond the linear Gaussian estimation. However, systems with non-Gaussian state and/or measurements abounds in practice. Many nonlinear filters, especially those based on LMMSE estimation, only propagate the first two moments without paying attention to other important information carried

Table 1.1: Outline of the dissertation

	Gaussian Assumption	Skew-Gaussian Assumption
Linear System	The Kalman filter (density filtering)	The Skew-Gaussian filter (density filtering)
Nonlinear System	Optimized GHQ rule (point filtering)	skew-symmetric representation of distributions (point filtering)

by higher moments. By considering the effect of skewness in the linear/nonlinear system, we proposed three methods, shown in Table 1.1, neglecting the linear case (since this result is well-known).

We first consider the effect of skewness in linear systems, and propose a linear skew-Gaussian system, a generalization of the linear Gaussian system, and derive its corresponding filter. Like the Kalman filter, our SGF, as the filter of linear skew-Gaussian system, is a recursive fixed dimensional exact density filter. The SGF not only has an analytical recursion of a small dimension akin to that of the Kalman filter, but also is efficient, which is comparable to the Kalman filter.

For nonlinear systems with significant skewness, nonlinear filters that only use the first two moments often do not perform well. A skew-symmetric representation of distributions is employed in order to model such complex nonlinear problems involving higher moments. Based on a first-order skew-Gaussian representation, a novel method for nonlinear point estimation is developed. The proposed first-order skew-Gaussian filter (FOSGF) is more general than the traditional Gaussian filters and LMMSE-based nonlinear filters, which propagate only the first two moments. Numerical results illustrate that our FOSGF can outperform conventional nonlinear filtering methods.

As for nonlinear systems with negligible skewness, the Gaussian filtering, is widely used in such situations. Gaussian filters usually obtain good performance efficiently by different quadrature methods. However, the performance of different choices of quadrature points is not predictable, and often degenerates if the integrand is not close to a polynomial. An optimization method to improve the performance of quadrature integration is proposed in this dissertation. It attempts to guarantee the integration accuracy by approximating the integrand as a polynomial, which fits quadrature rules well. The problem of optimizing the quadrature rule through this approximation is formulated and solved as a nonlinear least-squares problem with linear constraints. A new quadrature Gaussian filter is developed and compared with several popular nonlinear filters through simulation of two nonlinear examples.

1.8 Dissertation Outline

This rest of the dissertation, consisting of six chapters, is organized as follows.

Chapter 2 presents the basics of estimation theory, Gaussian/skew-Gaussian densities and its properties used in later chapters.

Chapter 3 proposes a fixed finite-dimensional recursive exact skew-Gaussian filter, which is based on the paper "Recursive Fixed Dimensional Exact Density Filtering for Discrete-Time Linear Skew-Gaussian Systems" submitted to Automatica.

Chapter 4 presents a nonlinear point filter based on skew-symmetric representation of the Gaussian distribution, which is based on the paper "Nonlinear State Estimation Using Skew-Symmetric Representation of Distributions" in 22th International Conference on Information Fusion, 2019.

Chapter 5 proposes the optimized Gauss-Hermite quadrature rule for nonlinear Gaussian filters, which is based on the paper "Optimized Gauss-Hermite Quadrature with Application to Nonlinear Filtering" in 21st International Conference on Information Fusion, 2018.

Chapter 6 draws conclusions and discusses future work.

2

Preliminaries

> Philosophy is the science of
> estimating values
>
> Manly P. Hall

This chapter is devoted to introducing the concepts and results of estimation theory. Gaussian/skew-Gaussian densities, along with their properties and some related theoretical results used in this dissertation, are also summarized.

2.1 Some Concepts in Estimation Theory

The estimation problem deals with the determination of those quantities that cannot be measured directly. In general, there exists a distinction between the estimation of signals (i.e., states) and the estimation of parameters. A *parameter* is a time-invariant quantity that characterises a mode. A *signal*, however, is often treated as a time-variant quantity (Mendel, 1995).

The quantity of interest which is to be estimated is called *estimand*, and an *estimator* $\hat{x}$ of an estimand x is a general rule (more precisely, a function) that passes from every input to the output under the following relationship

$$z^n \triangleq \{z_1, \ldots, z_n\}, \quad \text{with} \quad z_i = h(x_i, v_i, i), \quad v_i \sim (0, R_i)$$

Since z^n is random, as a function of random variables, $\hat{x}$ is also random. An estimator is an *optimal estimator* if it is best in some sense.

The result of the estimator (estimation rule) given a set of data is called the *estimate*, which is a realization of the estimator. The term "estimate" is often interchangeable with "estimator".

2.2 MMSE and Bayesian Estimator

By the *cost function* we denote a function $C(x, \hat{x})$ that assigns a cost to all pairs $(x, \hat{x})$ over the range of interest, and the Bayes risk is the expectation of the cost function $E(C(x, \hat{x}))$. In many cases of interest it is realistic to assume that the cost depends only on the estimation error, which is defined by

$$\tilde{x} \triangleq x - \hat{x}$$

and the Bayes risk is of the form $E(C(\tilde{x}))$. Accordingly, the Bayes estimator is given by

$$\hat{x}(z) \triangleq \arg\min_{\hat{x}(z)} E[C(\tilde{x})]$$

The Bayes estimation provides a unified framework to almost all important estimation methods, among which, the most widely used Bayes risk is of quadratic form, i.e.,

$$C(\tilde{x}) = \tilde{x}'\tilde{x}$$

It turns out that the MMSE estimate is the Bayesian estimation with the cost function of quadratic form, and the estimator is the conditional mean, i.e.,

$$\hat{x}(z) = \arg\min_{\hat{x}|z} E(C_1(\tilde{x}) \mid z) \Longrightarrow \hat{x}(z) = E(x|z)$$

and its corresponding MMSE matrix is defined by

$$\text{MSE}(\hat{x}) = E[(x-\hat{x})(x-\hat{x})'] = E(\tilde{x}\tilde{x}')$$

Actually, the conditional mean is optimal for a larger class of Bayes risk functions. The following two lemmas are the extension of the optimality of the conditional mean.

Lemma 2.1 (Van Trees (2004))**.** *Suppose that the cost function $C(\tilde{x})$ is a symmetric, convex-upward function and that the a posteriori density $p(x|z)$ is symmetric about its conditional mean; that is,*

$$C(\tilde{x}) = C(-\tilde{x})$$
$$C(\lambda x_1 + (1-\lambda)x_2) \leq \lambda C(x_1) + (1-\lambda)C(x_2)$$

for any $\lambda \in (0,1)$. The Bayesian estimator that minimizes any cost function in this class is identical to the conditional mean.

A proof of this lemma can be found in (Van Trees, 2004, p.240).

Lemma 2.2 (Van Trees (2004))**.** *Suppose that the cost function $C(\tilde{x})$ is a symmetric, non-decreasing function and that the a posteriori density $p(x|z)$ is a symmetric (about its conditional mean), unimodal function that satisfies the condition*

$$\lim_{x\to\infty} C(x)p(x|z) = 0$$

The Bayesian estimator that minimizes any cost function in this class is identical to the conditional mean.

A proof can be found in (Van Trees, 2004, p.378).

The significance of these two lemmas should not be undetermined. They ensure that whenever the a posteriori densities satisfy the assumptions given above, the estimates so obtained will be optimal for a large class of cost functions. Clearly, the Gaussian density satisfies the above assumptions.

2.3 LMMSE Estimation

Another main concept that will be used in this dissertation is the LMMSE estimation.

The LMMSE estimation, which propagates only the mean and covariance of the state., plays a key role in the state update of nonlinear point estimation, and it is still the workhorse of the nonlinear state point estimation nowadays.

The LMMSE estimator, which minimizes the mean-squared error (MSE) among all linear estimators (Anderson and Moore, 2005) is given as

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + P_{xz}P_z^{-1}(z_k - \hat{z}_{k|k-1}) \tag{2.1}$$

along with the MSE of $\hat{x}_{k|k}$ given by

$$P_{k|k} = P_{k|k-1} - P_{xz}P_z^{-1}P_{xz}' \tag{2.2}$$

where

$$\begin{aligned}
\hat{x}_{k|k-1} &= E(x_k|z^{k-1}) \\
\hat{z}_{k|k-1} &= E(z_k|z^{k-1}) \\
P_{k|k-1} &= E[(x_k - \hat{x}_{k|k-1})(\cdot)'] \\
P_{xz} &= E[(x_k - \hat{x}_{k|k-1})(z_k - \hat{z}_{k|k-1})' \\
P_z &= E[(z_k - \hat{z}_{k|k-1})(\cdot)']
\end{aligned}$$

where we use $(\cdot)$ to represent the term right before it.

Obtaining an LMMSE update is converted to finding the corresponding moments involved in (2.1) and (2.2), and different approximation methods to evaluate these moments can be used (see, e.g., (Ito and Xiong, 2000; Julier and Uhlmann, 2004; Arasaratnam and Haykin, 2009)).

2.4 Some Concepts in Exact Filtering

A state-space model in discrete time can be viewed as one for a partially observable stochastic process $\{(x_k, z_k)\}$ with unobservable $x_k \in \mathbb{R}^{k_x}$ and observable $z_k \in \mathbb{R}^{k_z}$. A filter is *exact* if it determines exactly the estimator or the distribution of x_n conditioned on available observations z^n (see (1.5) and (1.6)) without any approximation. In general, this conditional density is infinite dimensional. However, there exist filters which admit finite dimensional statistics even though they are rare, and we call such a filter *finite-dimensional* if it only involves finite-dimensional statistics for the posterior density, and a finite-dimensional filter whose dimension is unchanged with the increasing number of observations is called *fixed dimensional* (Li and Jilkov, 2010). A precise definition is given as follows by (GÜnther, 1981),

Definition 2.1. *(GÜnther, 1981) Let* $\{p(x;a) : a \in \mathbb{R}^{n_a}\}$ *be a parameterized set of densities on* $\mathbb{R}^{n_x}$*. A functions* $c : \mathbb{R}^{n_a} \times \mathbb{R}^{n_z} \to \mathbb{R}^{n_a}$ *is a finite-dimensional (*n_a*-dimensional) filter for* $\{(x_k, z_k)\}$ *if for* $k \geq 1$*,*

$$\{p(x_k|z^k) = p(x_k; c(a_{k-1}, z_k))\}$$

given

$$\{p(x_{k-1}|z^{k-1}) = p(x_{k-1}; a_{k-1}) \text{ for some } a_{k-1} \in \mathbb{R}^{n_a}\}$$

If the initial distribution of x_0 is $p(x_0; a_0)$, then the distribution at time k conditional on the observations z^k is $p(x_k; a_k)$, and $\{a_k\}$ is called the *filtering process*, defined by

$$a_k = c(a_{k-1}, z_k)$$

The recursibility of a filter is usually another requirement in practice, which means a recursive form of the filter exists (Li, 2004).

Since most practical applications requires computation and storage in each recursion to be finite and thus the filter to be finite dimensional in that it only involves some finite-dimensional statistic for the filtering density, and if the dimension does not increase over time, the filter is called fixed finite-dimensional. In general, for most estimation problems, the exact recursive filter is infinite dimension and rarely admits a finite dimension (Li and Jilkov, 2010).

2.5 Review of Gaussian Distribution

The Gaussian distribution is the most encountered distribution, and is the fundamental of many other important distributions, e.g., the chi-square distribution, the Rayleigh distribution, etc. The Gaussian distribution, which has a lot of nice properties, plays a central role in linear/nonlinear state estimation.

2.5.1 Mathematically Simplicity

Although it has an "abnormal" pdf and a cdf without closed form, the Gaussian distribution behaves "normally" in many aspects (Li, 2015):

1. It lives in the linear world:

    - Jointly Gaussian vectors after any linear (affine) transformation remain to be jointly Gaussian
    - For Jointly Gaussian vectors, independence is equivalent to (almost) surely no linear correlation.
    - For two jointly Gaussian vectors, the overall optimal estimator of one vector given the other is linear in the other vector;
    - Given the first two moments, there exists one and only one Gaussian distribution.

2. It exists in the real world:

    - The principle in the exponential law of error is accepted and observed by the public.
    - The central limit theorem provides a justification to make the ubiquitous Gaussian assumption for many real-world sources of uncertainty.

2.5.2 Gaussian Assumption in Nonlinear Filtering

In Gaussian filters, the Gaussian assumption can be justified as follows.

From the Viewpoint of Information

The Gaussian distribution plays a special role in the Fisher information matrix (FIM) and Cramer-Rao lower bound (CRLB). It is the worst case in terms of FIM and CRLB, that is, among all distributions of the same first two moments, the Gaussian distribution (at least in the asymptotic case) gives the smallest FIM. More precisely, estimating x using data z with the likelihood function $L(x|z)$ being Gaussian, has the largest CRLB among those with distributions of the same first two moments (Li, 2015).

In fact, a similar statement also holds in terms of Shannon information: The Gaussian distribution is the one that has the maximum entropy (i.e., is most uncertain) among all distributions with zero mean and the same covariance when the differential entropy exists.

In all, assuming the Gaussianity is equivalent to consider the worst case, which is a robust assumption in some sense (Roth et al., 2016).

From the Viewpoint of Projection

Even though most nonlinear estimation problems rarely permit the Gaussian distribution, many nonlinear estimation methods rely on the Gaussian assumption for its simplicity. One possible way to think is the probability-based estimation (Kulhavỳ, 1996).

Instead of using "data matching" methods, e.g., the least-squares method, a "probability matching" method is employed to find a pdf from a given family of densities by minimizing its distance from the true but unknown density. Usually the Kullback-Leibler divergence is adopted. If the given family is constrained to be the Gaussian family, moment matching of the first two moments turns out to be equivalent to minimizing the Kullback-Leibler divergence (Barber, 2012).

2.6 Review of Skew-Gaussian Distribution

In this section some preliminary knowledge about the skew-Gaussian distribution is presented, and a skew-Gaussian process is introduced and analyzed.

2.6.1 Background

The skew-Gaussian distribution was studied in great detail, in the statistical community (see, e.g., (Azzalini, 2013)). Its form, given in Section 2.6.2, is closely related to the skew-symmetric representation of distributions in Section 4.2.1; however, it is special in many aspects among all skewness construction.

The skew-Gaussian distribution has many nice properties that resemble the Gaussian distribution (see Section 2.6.2 for more details), and problems obeying the skew-Gaussian distribution exist in many practical situations. It has a clear physical realization in reality—it models problems where a linear hidden truncation mechanism takes place (Arellano-Valle et al., 2006). One practical example is selective sampling in social and economic studies. A simple scheme which retains the essential ingredients of the skew-Gaussian distribution involves two linear models of the form

$$Z = x'\beta + \epsilon_1$$

$$W = w'\gamma + \epsilon_2$$

where x and w are vectors of explanatory variables being regarded as fixed, β and γ are parameters of appropriate dimensions, and $\epsilon = (\epsilon_1, \epsilon_2)'$ is a two-dimensional Gaussian distribution with correlation coefficient ρ. The first linear model is of interest, but we do not observe Z and W directly. In fact, Z is observed only when $\gamma_1 < W < \gamma_2$. Under such a linear hidden truncation scheme, the distribution of Z is skew-Gaussian.

From the scheme above, if the linear hidden truncation (i.e., $\gamma_1 < W < \gamma_2$) does not exist or $\rho = 0$, the distribution of Z degenerates to Gaussian. As such, the family of skew-Gaussian densities includes the Gaussian as a special case. Since the skew-Gaussian density is grounded on practical demands and its properties are appealing (more details of properties are discussed in Chapter 3), the research of the state estimation based on the skew-Gaussian assumption attracts attention (Genton, 2004; He et al., 2018; Rezaie and Eidsvik, 2014).

2.6.2 Skew-Gaussian Density and Its Properties

In this subsection, the skew-Gaussian distribution is investigated in detail, and its properties, which contribute to the mainstay of our skew-Gaussian filter, are presented in Section 3.3. The exposition of Properties 2.1 and 2.3–2.5 is relied on (Arellano-Valle and Azzalini, 2006) and that of Properties 2.2 and 2.6 is based on (González-Farías et al., 2004).

The following definition is our generalization (in a bilaterally truncation version) of the skew-Gaussian distribution introduced in (Arellano-Valle and Azzalini, 2006), where only the upper limit was considered (i.e., $\gamma_1 = -\infty$).

Definition 2.2. *Consider $n, m \geq 1$, $u \in \mathbb{R}^n$, $\gamma_1, \gamma_2 \in \mathbb{R}^m$ with $\gamma_1 \leq \gamma_2$ componentwisely. Δ is an arbitrary $n \times m$ matrix, and Σ and Γ are positive definite matrices of dimensions $n \times n$ and $m \times m$, respectively. Let $\Phi_n([a, b]; u, \Sigma)$ be the integral of $\mathcal{N}_n(x; u, \Sigma)$ over the domain $a \leq x \leq b$ held componentwisely. We say $x \in \mathbb{R}^n$ is a skew-Gaussian random variable, denoted by*

$$x \sim SG_{n,m}([\gamma_1, \gamma_2], u, \Omega)$$

if its pdf is given by

$$\mathcal{N}_n(x; u, \Sigma) \frac{\Phi_m([\gamma_1, \gamma_2], \Delta'\Sigma^{-1}(x - u), \Gamma - \Delta'\Sigma^{-1}\Delta)}{\Phi_m([\gamma_1, \gamma_2]; \Gamma)} \tag{2.3}$$

provided that

$$\Omega = \begin{bmatrix} \Gamma & \Delta' \\ \Delta & \Sigma \end{bmatrix} \tag{2.4}$$

is positive definite.

Here we write Γ, Δ and Σ in a compact form to emphasize the positive definiteness of Ω, which represents the covariance of some random variable (Property 2.4). Note that u, Σ, and Δ can be deemed location, scale, and skewness parameters, respectively. In general, u and Σ are not the mean and covariance of the skew-Gaussian distribution, although they are highly related to the first three moments.

A contour plot is shown in Figure 2.1 with $\gamma_1 = [0, 0]'$, $\gamma_2 = [5, 5]'$, $u = [0, 0]'$ and a random realization

$$\Omega = \begin{bmatrix} 7.82 & -0.03 & 0.67 & 0.29 \\ -0.03 & 1.79 & -1.87 & -0.78 \\ 0.67 & -1.87 & 16.71 & -3.22 \\ 0.29 & -0.78 & -3.22 & 1.53 \end{bmatrix}$$

Remark 2.1. *The skew-Gaussian density degenerates to its symmetric part* $\mathcal{N}_n(x; u, \Sigma)$ *if* $\Delta = 0$ *or* $[\gamma_1 = (-\infty, \ldots, -\infty)', \gamma_2 = (\infty, \ldots, \infty)']$.

The skew-Gaussian distribution (2.3) has many properties that resemble the Gaussian distribution. In the following discussion, we omit the dimension indices for simplicity.

Suppose

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathrm{SG}([\gamma_1, \gamma_2], u, \Omega) \tag{2.5}$$

and u and Ω in (2.4) are partitioned accordingly as

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad \Delta = \begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_{21}' \\ \Sigma_{21} & \Sigma_2 \end{bmatrix} \tag{2.6}$$

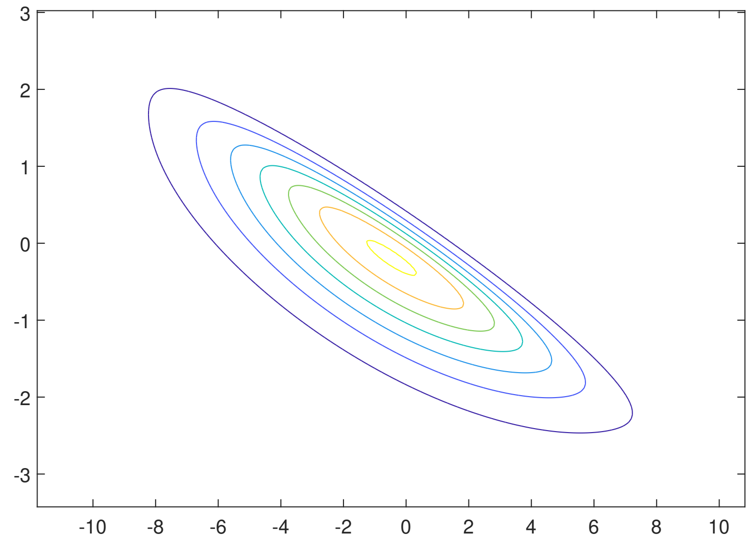

Figure 2.1: Contour of skew-Gaussian in two dimensions

Property 2.1 (Closed under marginalization (Arellano-Valle et al., 2006))**.** *The marginal density of x_1 is skew-Gaussian with density*

$$\mathcal{N}(x_1; u_1, \Sigma_1)\frac{\Phi([\gamma_1, \gamma_2], \Delta_1'\Sigma_1^{-1}(x_1 - u_1), \Gamma - \Delta_1'\Sigma_1^{-1}\Delta_1)}{\Phi([\gamma_1, \gamma_2]; \Gamma)}$$

with u_1 and Σ_1 in (2.6)*, while keeping γ_1, γ_2, and Γ unaltered.*

Remark 2.2. *The marginalization of a skew-Gaussian distribution behaves the same as the Gaussian distribution does, and it is instrumental in the well-definedness of the skew-Gaussian process presented in Section 2.6.3.*

Property 2.2 (Closed under linear transformation (Arellano-Valle et al., 2006))**.** *Suppose we have the following linear transformation (more precisely, affine transformation)*

$$y = Ax + b, \qquad x \sim SG([\gamma_1, \gamma_2], u, \Omega_x) \tag{2.7}$$

where

$$\Omega_x = \begin{bmatrix} \Gamma & \Delta' \\ \Delta & \Sigma \end{bmatrix}$$

Then

$$y \sim SG(y; [\gamma_1, \gamma_2], Au + b, \Omega_y) \tag{2.8}$$

where

$$\Omega_y = \begin{bmatrix} \Gamma & \Delta' A' \\ A\Delta & A\Sigma A' \end{bmatrix}$$

provided that Ω_y is positive definite.

Property 2.3 (Closed under conditioning (Arellano-Valle et al., 2006))**.** *Given* (2.5)*, the density of x_1 conditioned on x_2 is still skew-Gaussian,*

$$(x_1 \mid x_2) \sim SG([\gamma_{1,1|2}, \gamma_{2,1|2}], u_{1|2}, \Omega_{1|2})$$

with

$$\Omega_{1|2} = \begin{bmatrix} \Gamma_{1|2} & \Delta'_{1|2} \\ \Delta_{1|2} & \Sigma_{1|2} \end{bmatrix}$$

where

$$u_{1|2} = u_1 + \Sigma'_{21}\Sigma_2^{-1}(x_2 - u_2) \quad \Sigma_{1|2} = \Sigma_1 - \Sigma'_{21}\Sigma_2^{-1}\Sigma_{21}$$

$$\Delta_{1|2} = \Delta_1 - \Sigma'_{21}\Sigma_2^{-1}\Delta_2 \quad \Gamma_{1|2} = \Gamma - \Delta'_2\Sigma_2^{-1}\Delta_2$$

$$\gamma_{i,1|2} = \gamma_i - \Delta'_2\Sigma_2^{-1}(x_2 - u_2) \quad i = 1, 2$$

A proof of Property 2.3 is given in Appendix B.

Remark 2.3. *The combination of Property 2.2 and Property 2.3 guarantees the closedness of posterior skew-Gaussian density in a linear system, which is the backbone of our skew-Gaussian filter derived in Section 3.3.*

Proposition 2.1 (Moments up to the third order)**.** *Let*

$$\mu_1 = \frac{1}{\Phi([\gamma_1, \gamma_2]; \Gamma)} \int_{\gamma_1 \le t \le \gamma_2} t\mathcal{N}(t; \Gamma)\ \mathrm{d}t \tag{2.9}$$

$$\mu_2 = \frac{1}{\Phi([\gamma_1, \gamma_2]; \Gamma)} \int_{\gamma_1 \le t \le \gamma_2} tt'\mathcal{N}(t; \Gamma)\ \mathrm{d}t \tag{2.10}$$

$$\mu_3 = \frac{1}{\Phi([\gamma_1, \gamma_2]; \Gamma)} \int_{\gamma_1 \le t \le \gamma_2} t vec'(tt')\mathcal{N}(t; \Gamma)\ \mathrm{d}t \tag{2.11}$$

The mean, covariance and skewness (i.e., the third central moments) of the skew-Gaussian distribution (2.3) *are*

$$\bar{m}_1 = \mathrm{E}(x) = u + \Delta\Gamma^{-1}\mu_1 \tag{2.12}$$

$$\bar{m}_2 = \mathrm{cov}(x) = \Sigma - \Delta\Gamma^{-1}(\Gamma + \mu_1\mu_1' - \mu_2)\Gamma^{-1}\Delta' \tag{2.13}$$

$$\begin{aligned} \bar{m}_3 &= \mathrm{E}[(x - \mathrm{E}(x))[(x - \mathrm{E}(x))']^{\otimes 2}] \\ &= \Delta\Gamma^{-1}\{2\mu_1 \,\mathrm{vec}'(\mu_1\mu_1') - [\mu_1 \,\mathrm{vec}'(\mu_2) + \mu_2 \otimes \mu_1' \\ &\quad + \mu_1' \otimes \mu_2] + \mu_3\}(\Delta\Gamma^{-1} \otimes \Delta\Gamma^{-1})' \end{aligned} \tag{2.14}$$

A proof of Proposition 2.1 is given in Appendix A.

Remark 2.4. *(a) Equations* (2.9)–(2.11) *are the first three non-central moments of Gaussian density $\mathcal{N}(x; 0, \Gamma)$ being doubly-truncated with the linear constraint $\gamma_1 \leq x \leq \gamma_2$ holding componentwisely. The evaluation of the multivariate Gaussian integral with the linear domain constraints is well studied in statistics and machine learning. Efficient algorithms of moderate dimensional integral exist (e.g., see (Kan and Robotti, 2017; Wilhelm and Manjunath, 2010)).*

(b) By (2.12) *and* (2.13)*, the MMSE skew-Gaussian estimator follows from the SGF shown in Section 3.3.*

The physical meaning of the skew-Gaussian distribution can be found by the following property via a hidden truncation process.

Property 2.4 (Hidden truncation representation (Arellano-Valle et al., 2006))**.** *Suppose*

$$\begin{bmatrix} x^* \\ x \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ u \end{bmatrix}, \begin{bmatrix} \Gamma & \Delta' \\ \Delta & \Sigma \end{bmatrix} \right) \tag{2.15}$$

Then, the density of x conditioned on $\gamma_1 \leq x^* \leq \gamma_2$ is skew-Gaussian:

$$\big(x \mid (\gamma_1 \leq x^* \leq \gamma_2)\big) \sim SG([\gamma_1, \gamma_2], u, \Omega), \quad \Omega = \begin{bmatrix} \Gamma & \Delta' \\ \Delta & \Sigma \end{bmatrix}$$

where x^* is called the latent random variable. This representation will be used in the definition of the skew-Gaussian process in Section 2.6.3.

Another direction of generating the skew-Gaussian distribution follows the idea of orthogonalization of random variables.

Property 2.5 (Orthogonal representation (Arellano-Valle et al., 2006))**.** *Following* (2.15)*, construct two independent variables*

$$\begin{aligned} v_0 &= x^* \sim \mathcal{N}(0, \Gamma) \\ v_1 &= x - \mathrm{E}(\breve{x} \mid x^*) = x - \Delta\Gamma^{-1}x^* \\ v_1 &\sim \mathcal{N}(u, \Sigma - \Delta\Gamma^{-1}\Delta') \end{aligned}$$

where $\breve{x}$ is the zero-mean part of x. Let $v_0^{(\gamma_1,\gamma_2)}$ be the doubly-truncated v_0 (i.e., $v_0^{(\gamma_1,\gamma_2)} = (v_0 \mid \gamma_1 \le v_0 \le \gamma_2)$). Then

$$(v_1 + \Delta\Gamma^{-1}v_0^{(\gamma_1,\gamma_2)}) \sim SG([\gamma_1, \gamma_2], u, \Omega), \quad \Omega = \begin{bmatrix} \Gamma & \Delta' \\ \Delta & \Sigma \end{bmatrix}$$

Clearly, v_0 and v_1 are the Gram–Schmidt orthogonal decomposition components of $[(x^*)', x']'$ in (2.15), and such a method is useful for simulation. For generating doubly-truncated Gaussian random variables, please refer to (Botev, 2016; Chopin, 2011).

Property 2.6 (Joint distribution of independent skew-Gaussian variables (Arellano–Valle et al., 2006; González-Farías et al., 2004))**.** *Suppose $x_1 \sim SG_{n_1,m_1}([\gamma_1^1, \gamma_2^1], u_1, \Omega_1)$ and $x_2 \sim SG_{n_2,m_2}([\gamma_1^2, \gamma_2^2], u_2, \Omega_2)$ are independent, where*

$$\Omega_1 = \begin{bmatrix} \Gamma_1 & \Delta_1' \\ \Delta_1 & \Sigma_1 \end{bmatrix}, \quad \Omega_2 = \begin{bmatrix} \Gamma_2 & \Delta_2' \\ \Delta_2 & \Sigma_2 \end{bmatrix}$$

Then

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim SG_{n_1+n_2,m_1+m_2}([\gamma_1, \gamma_2], u, \Omega) \tag{2.16}$$

where

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad \gamma_i = \begin{bmatrix} \gamma_i^1 \\ \gamma_i^2 \end{bmatrix}, \quad i = 1, 2$$

$$\Omega = \begin{bmatrix} \Gamma_1 & 0 & \Delta_1' & 0 \\ 0 & \Gamma_2 & 0 & \Delta_2' \\ \Delta_1 & 0 & \Sigma_1 & 0 \\ 0 & \Delta_2 & 0 & \Sigma_2 \end{bmatrix}$$

2.6.3 Skew-Gaussian Process

In this subsection we introduce a new stochastic process, i.e., the skew-Gaussian process.

Definition 2.3. *Let T be an index set. Consider functions $u : T \to \mathbb{R}$, $\Delta : T \to \mathbb{R}$, and $\Sigma : T \times T \to \mathbb{R}$. We say $\{x_t, t \in T\}$ is a* ***skew-Gaussian process*** *w.r.t. the latent random variable $x^* \in \mathbb{R}^m$ (i.e., w.r.t. the parameters $\gamma_1, \gamma_2 \in \mathbb{R}^m$ and a positive definite matrix $\Gamma \in \mathbb{R}^m \times \mathbb{R}^m$) if for every finite set $J = \{t_1, \cdots, t_n\} \subset T$ with any $n > 0$, the vector $x = [x_{t_1}, \cdots, x_{t_n}]'$ is skew-Gaussian distributed:*

$$x \sim SG([\gamma_1, \gamma_2]; u(J), \Omega(J))$$

provided that

$$\Omega(J) = \begin{bmatrix} \Gamma & \Delta(J)' \\ \Delta(J) & \Sigma(J, J) \end{bmatrix}$$

is positive definite.

The skew-Gaussian process is constructed in terms of its finite-dimensional marginals through the Kolmogorov extension theorem (Wong, 1985). The key requirement of the theorem is that the finite-dimensional marginals should be consistent in the following sense: If J_n and J_m are two ordered finite sets from T such that $J_n \subseteq J_m$, then the cdf P_{J_n} must be equal to the cdf P_{J_m} with the appropriate variable set to $+\infty$. For instance,

$$P_{t_1}(x_1) = P_{t_1,t_2}(x_1, +\infty)$$

This is guaranteed by Property 2.1, and its uniqueness of the probability measure also follows from the Kolmogorov extension theorem.

Remark 2.5. *(a) The skew-Gaussian process defined includes the Gaussian process as a special case when $\Delta = 0$ or there are no constraints over the domain of the latent random variable x^*.*

(b) If $\Delta \neq 0$, the states at different times of a skew-Gaussian process are correlated, and the skew-Gaussian process so defined cannot be independent, so it does not include the white stochastic process whose finite marginalization is jointly skew-Gaussian. More specifically, it precludes any white process from being truly skew-Gaussian.

3

Linear Skew-Gaussian Estimation

> The simplest explanation is usually the best one
>
> Occam's razor

3.1 Introduction

As mentioned early in Section 1.2, the filtering problem can be formulated from the Bayesian perspective, where it computes the posterior density of the state at time k given all observed data up to time k (Ho and Lee, 1964). This Bayesian approach offers a unifying methodology for solving the general problems of estimation, and provides a theoretical foundation for filtering methods in practice.

Conceptually, density filtering, which follows the Bayesian approach in state estimation, provides a thorough solution to the filtering problem, but elegant theoretical results are available only for exact density filtering (Li and Jilkov, 2010). Exact density filtering, where its computed posterior density is exact, forms a solid foundation for further development in theory and application. For example, it offers guidance to approximate density filtering, which has been applied to many practical problems and has shown exceptional value when point filtering is inadequate (Ristic et al., 2003). However, obtaining the exact posterior density is in general intractable, if not impossible, since it usually involves infinitely many unknowns or variables and cannot be directly applied in most practical situations.

In reality, finite dimensionality is a prerequisite for implementing density filtering, since practical applications require computation and storage to be finite. Consequently, exact density filtering of a finite dimension makes itself the only possible exact density filtering with practical value. However, exact density filtering problems rarely permit finite-dimensional solutions.

Efficiency of density filtering is another crucial requirement for most engineering applications, especially for real-time applications. In situations where no exact yet efficient density filtering is available, approximation methods are often employed. Due to limited processing resources in practice, exact density filtering of finite dimension may still be insufficient in situations where the dimension is finite but increasing over time. Therefore, fixed dimensional exact filtering, especially with a small dimension, is of extreme importance and interest from both theoretical and practical perspectives.

Despite its rare existence in applications, fixed dimensional exact filtering, which obtains some fixed and finite dimensional sufficient statistics recursively, contributes to density filtering. A well-known example is the state filtering for a linear Gaussian system, where the celebrated Kalman filter provides a complete recursion of the posterior Gaussian density. Another example is the system identification of linear Gaussian system (Elliott and Krishnamurthy, 1999), where a filter-based expectation maximization algorithm was proposed for parameter estimation of the system. However, few results are known beyond the linear Gaussian system and its limited variants, except some procedures made by using conjugate pairs (Vidoni, 1999; Ferrante and Vidoni, 1999) for linear system and some simple finite-dimensional filters for nonlinear systems, e.g., the Beneš filters (Beneš, 1981) and Daum's filters (Daum, 1986).

In this chapter, we study the density filtering problem in linear skew-Gaussian systems, which is a generalization of the linear Gaussian system, and its recursive fixed dimensional exact density filter, i.e., the skew-Gaussian filter (SGF), which is linear (more precisely affine, it, unfortunately, is standard practice in estimation community) in measurements. The SGF not only has an analytical recursion of a small dimension akin to that of the Kalman filter, but also possesses an efficiency comparable to the Kalman filter.

In retrospect, for the conventional linear Gaussian system, the assumed Gaussianity and Markov property furnish a sufficient condition for the existence of a fixed dimensional density filter. The Gaussianity admits that the posterior expectation and covariance are sufficient statistics for the whole density. The Markov property, similar to the recursibility of an estimator (Li, 2004), principally implies that the state distribution at a time contains all the information needed for its later propagation. However, a linear Gaussian model cannot address satisfactorily many practical demands since correlated non-Gaussian noise abounds in inference problems. For example, the radar measurements may present non-Gaussian and correlated behavior due to high sampling frequency (Wu and Chang, 1996) and tracking glint (Wu, 1993), and the distributions of data collected in many engineering areas are non-Gaussian in situations where impulsive interference or outliers arise (Nurminen et al., 2018; Zoubir et al., 2012). The Kalman filter and its variants suffer from performance degradation in the presence of non-Gaussian process/measurement noise.

The ever-growing demand of modeling beyond the linear Gaussian system motivates the exploration of non-Gaussian systems. The literature abounds with approaches to modeling flexible linear non-Gaussian systems, especially in robust filtering, where the non-Gaussianity is often characterized of a skewed and/or heavy-tailed nature (Girón and Rojano, 1994; Wang and Balakrishnan, 2002). However, many of these models are not justified well enough since the design of the noise distributions lacks practical justification (e.g., the assumptions of the Wishart or normal-gamma distribution for the noise covariance were mainly for simplicity and convenience for which no general guarantees have been established (Agamennoni et al., 2012; Huang et al., 2017b)). In addition, exact solutions exist only in very limited cases with restrictive assumptions (e.g., the requirement of the same degree of freedom in the robust t-distribution based Kalman filter), and most filters either require strong approximation (e.g., the mean field variational Bayesian method (Zhang et al., 2018; Huang et al., 2017a)—a strong assumption on the structure of jointly hierarchical distribution)—or resort to powerful approximation methods, such as the sequential Monte Carlo method based on conditional independence in the model (Doucet et al., 2001).

In this chapter, we propose a new linear model for the skew-Gaussian process. This model is broader than the conventional linear Gaussian model and has a potential to handle problems where asymmetry in the densities of state and/or correlated noise is not negligible. The key is the skew-Gaussian process, which aroused interests in the study of parametric classes of probability distributions in the statistical and machine learning communities (Azzalini, 2013; Genton, 2004; Benavoli et al., 2020). The skew-Gaussian distribution can represent a broad class of densities with shape flexibility other than the Gaussian while maintains mathematical tractability. On the one hand, problems obeying a skew-Gaussian distribution exist in many practical situations, e.g., filtering of the state with quantized measurement (Sukhavasi and Hassibi, 2009) or based on some event-based scheduling (He et al., 2018). On the other hand, the skew-Gaussian process, which models problems where a linear hidden truncation mechanism takes place (Arellano-Valle et al., 2006), may also have potential to formulate some type of skewed and correlated noise impacted by some common latent random factors truncation in reality.

An extensive literature exists of linear system models using, e.g., the elliptical distributions or the skew-elliptical distributions. In (Girón and Rojano, 1994), a closed form of a Kalman-like filter was proposed under the assumption of a jointly elliptical distribution of the initial state and two noise processes without correlation. In (Benavoli et al., 2020; Rezaie and Eidsvik, 2014), two linear models based on the skew-Gaussian distribution were proposed, but they suffer from problems of either a restrictive structure on the linear system, which is rare in applications, or high computational complexity due to a rapidly growing dimension. A recent work with application to skew-Gaussian was (He et al., 2018), where the skew-Gaussianity is triggered by an event-based scheduling, but it also suffers from the curse of dimensionality.

Compared with the aforementioned models, our linear skew-Gaussian model admits a fixed dimensional exact filtering algorithm, i.e., the SGF. Including the Kalman filter as a special case, the SGF has an efficient, recursive solution for updating all parameters of the posterior distribution without any approximation. Its implementation has two parts: the Gaussian part and the skewness part, which propagate simultaneously in both prediction and update stages and are linear in measurements, as the Kalman filter does.

This chapter is organized as follows. Section 3.2 formulates the linear skew-Gaussian system. A recursive fixed dimensional filter based on the linear skew-Gaussian model is derived in Section 3.3. Simulation results and analysis are shown in Section 3.4.

3.2 Linear Skew-Gaussian Model

Consider the following discrete-time linear system

$$\begin{aligned} x_k &= F_{k-1}x_{k-1} + G_{k-1}w_{k-1} \\ z_k &= H_k x_k + v_k \end{aligned} \tag{3.1}$$

where $x_k \in \mathbb{R}^{n_x}$ is the state of the system at time k, $\{w_k\}$ is the process noise due to disturbance and modeling error, $z_k \in \mathbb{R}^{n_z}$ is the measurement, and $\{v_k\}$ is the measurement noise.

Instead of using the Gaussian assumption, we propose to use a more general assumption called the **skew-Gaussian assumption** as follows:

1. $\{w_k\}$ and $\{v_k\}$ are skew-Gaussian processes with block diagonal Σ, that is, for every k,

$$w^k = [w_k', \cdots, w_0']' \sim \mathrm{SG}([\gamma_1^w, \gamma_2^w], 0, \Omega^{w,k})$$

$$v^k = [v_k', \cdots, v_1']' \sim \mathrm{SG}([\gamma_1^v, \gamma_2^v], 0, \Omega^{v,k})$$

where

$$\Omega^{w,k} = \begin{bmatrix} \Gamma^w & (\Delta^{w,k})' \\ \Delta^{w,k} & Q^k \end{bmatrix}$$

$$\Omega^{v,k} = \begin{bmatrix} \Gamma^v & (\Delta^{v,k})' \\ \Delta^{v,k} & R^k \end{bmatrix}$$

Here

$$\Delta^{w,k} = [(\Delta_k^w)', \ldots, (\Delta_1^w)']', \qquad \Delta^{v,k} = [(\Delta_k^v)', \ldots, (\Delta_1^v)']'$$

$$Q^k = \mathrm{diag}([Q_k, \cdots, Q_0]), \qquad R^k = \mathrm{diag}([R_k, \cdots, R_1])$$

2. The initial state is skew-Gaussian:

$$x_0 \sim \mathrm{SG}([\gamma_1^x, \gamma_2^x], u_0, \Omega_0^x) \tag{3.2}$$

where

$$\Omega_0^x = \begin{bmatrix} \Gamma^x & (\Delta_0^x)' \\ \Delta_0^x & \Sigma_0 \end{bmatrix}$$

3. x_0, $\{w_k\}$, and $\{v_k\}$ are mutually independent.

The linear model (3.1) together with the skew-Gaussian assumption is the **linear skew-Gaussian assumption**. It reduces to the linear Gaussian assumption if the corresponding $\Delta = 0$ or $[\gamma_1 = (-\infty, \ldots, -\infty)', \gamma_2 = (\infty, \ldots, \infty)']$.

Clearly, this model has potential to formulate more complex situations than the linear Gaussian model does, especially if the skewed state or observation is significant. Moreover, this modeling is exact in situations where the noises at different times are correlated through some common random factors, which are modeled by the latent random variables. For example, in the selective sampling scheme, the samples are only observable when some indicator variable exceeds some threshold (Azzalini, 2013).

3.3 Recursive Finite-Dimensional Filter for Linear Skew-Gaussian System

In this section, we present a recursive finite-dimensional filter based on the linear skew-Gaussian model of Section 3.2.

3.3.1 Exact Skew-Gaussian Filter (SGF)

In this subsection, we connect the idea of the skew-Gaussian assumption in Section 3.2 with the properties of skew-Gaussian distribution in Section 2.6 to present the main results of our chapter. The following two theorems constitute one cycle of filtering for system (3.1).

Theorem 3.1 (Prediction)**.** *The distribution of x_k conditioned on measurements z^{k-1} under the linear skew-Gaussian assumption is given as ($x_{k|k-1} \sim SG$ is a short-hand notation for the pdf $p(x_k \mid z^{k-1})$ being skew-Gaussian)*

$$x_{k|k-1} \sim SG([\hat{\gamma}_{1,k|k-1}, \hat{\gamma}_{2,k|k-1}], \hat{u}_{k|k-1}, \Omega_{k|k-1})$$

where

$$\Omega_{k|k-1} = \begin{bmatrix} \Gamma_{k|k-1} & \Delta'_{k|k-1} \\ \Delta_{k|k-1} & \Sigma_{k|k-1} \end{bmatrix}$$

and the corresponding parameters are updated as:

- *Gaussian part (same as the prediction in the Gaussian case):*

$$\begin{aligned} \hat{u}_{k|k-1} &= F_{k-1}\hat{u}_{k-1|k-1} \\ \Sigma_{k|k-1} &= F_{k-1}\Sigma_{k-1|k-1}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1} \end{aligned}$$

- *Skewness part:*

$$\begin{aligned} \hat{\gamma}_{i,k|k-1} &= \hat{\gamma}_{i,k-1|k-1}, \quad i = 1, 2 \\ \Delta_{k|k-1} &= F_{k-1}\Delta_{k-1|k-1} + [0, G_{k-1}\Delta^w_{k-1}, 0] \\ \Gamma_{k|k-1} &= \Gamma_{k-1|k-1} \end{aligned} \tag{3.3}$$

The 0's in (3.3) have dimensions $n_x \times n_{\gamma^x}$ and $n_x \times n_{\gamma^v}$, respectively, where $n_{\gamma^x} \triangleq \dim(\gamma_1^x)$ and $n_{\gamma^v} \triangleq \dim(\gamma_1^v)$.

Theorem 3.2 (Update)**.** *The distribution of x_k conditioned on measurements z^k under the linear skew-Gaussian assumption is given as*

$$x_{k|k} \sim SG([\hat{\gamma}_{1,k|k}, \hat{\gamma}_{2,k|k}], \hat{u}_{k|k}, \Omega_{k|k})$$

where

$$\Omega_{k|k} = \begin{bmatrix} \Gamma_{k|k} & \Delta'_{k|k} \\ \Delta_{k|k} & \Sigma_{k|k} \end{bmatrix}$$

and the corresponding parameters are updated as:

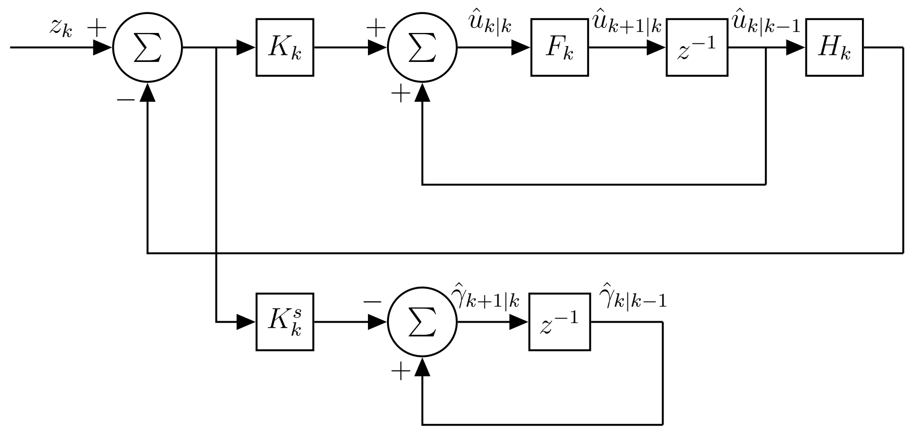

FIGURE 3.1: Flowchart of the propagation of $\hat{u}_{k|k-1}$ and $\hat{\gamma}_{k|k-1}$

- *Gaussian part (same as the update in the Gaussian case):*

$$\begin{aligned}
S_k &= H_k \Sigma_{k|k-1} H_k' + R_k \\
K_k &= \Sigma_{k|k-1} H_k' S_k^{-1} \\
\hat{u}_{k|k} &= \hat{u}_{k|k-1} + K_k(z_k - H_k \hat{u}_{k|k-1}) \\
\Sigma_{k|k} &= \Sigma_{k|k-1} - K_k S_k K_k'
\end{aligned}$$

- *Skewness part:*

$$\begin{aligned}
K_k^s &= (H_k \Delta_{k|k-1} + [0, \Delta_k^v])' S_k^{-1} && (3.4) \\
\hat{\gamma}_{i,k|k} &= \hat{\gamma}_{i,k|k-1} - K_k^s (z_k - H_k \hat{u}_{k|k-1}) \\
\Delta_{k|k} &= (I - K_k H_k) \Delta_{k|k-1} - [0, K_k \Delta_k^v] && (3.5) \\
\Gamma_{k|k} &= \Gamma_{k|k-1} - K_k^s S_k (K_k^s)'
\end{aligned}$$

The dimensions of the 0's in (3.4) and (3.5) are $n_z \times (n_{\gamma^x} + n_{\gamma^w})$ and $n_x \times (n_{\gamma^x} + n_{\gamma^w})$, where $n_{\gamma^w} \triangleq \dim(\gamma_1^w)$. Theorems 3.1 and 3.2 together provide a recursion of the posterior distribution. Noteworthily, the SGF is linear in measurements from the formulas above.

Proofs of Theorems 3.1 and 3.2 are given in Appendix C, and more details are shown in Appendix D.

The algorithm of the SGF is summarized in Algorithm 1. The propagation structures of $\hat{\gamma}_{i,k|k-1}$ and $\Delta_{k|k-1}$ are depicted in Figures 3.1 and 3.2, respectively.

Algorithm 1: The Skew-Gaussian Filter

Initialization: $\hat{u}_{0|0} = u_0^x$, $\Sigma_{0|0} = \Sigma_0^x$, $\Delta_{0|0} = [\Delta_0^x, 0]$,
$\hat{\gamma}_{i,0|0} = [(\gamma_i^x)', (\gamma_i^w)', (\gamma_i^v)']'$, $\Gamma_{0|0} = \text{diag}(\Gamma^x, \Gamma^w, \Gamma^v)$

for $k = 1$ to N

  Input: $\hat{u}_{k-1|k-1}$, $\Sigma_{k-1|k-1}$, $\Delta_{k-1|k-1}$, $\Gamma_{k-1|k-1}$, $\hat{\gamma}_{i,k-1|k-1}$, z_k, Δ_{k-1}^w, Δ_k^v

  Time Update: ;

  $\hat{u}_{k|k-1} = F_{k-1}\hat{u}_{k-1|k-1}$;

  $\Sigma_{k|k-1} = F_{k-1}\Sigma_{k-1|k-1}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1}$;

  $\hat{\gamma}_{i,k|k-1} = \hat{\gamma}_{i,k-1|k-1}$;

  $\Delta_{k|k-1} = F_{k-1}\Delta_{k-1|k-1} + [0, G_{k-1}\Delta_{k-1}^w, 0]$;

  $\Gamma_{k|k-1} = \Gamma_{k-1|k-1}$

  Measurement Update: ;

  $S_k = H_k\Sigma_{k|k-1}H'_k + R_k$;

  $K_k = \Sigma_{k|k-1}H'_kS_k^{-1}$;

  $\hat{u}_{k|k} = \hat{u}_{k|k-1} + K_k(z_k - H_k\hat{u}_{k|k-1})$;

  $\Sigma_{k|k} = \Sigma_{k|k-1} - K_kS_kK'_k$;

  $K_k^s = (H_k\Delta_{k|k-1} + [0, \Delta_k^v])'S_k^{-1}$ $\hat{\gamma}_{i,k|k} = \hat{\gamma}_{i,k|k-1} - K_k^s(z_k - H_k\hat{u}_{k|k-1})$;

  $\Delta_{k|k} = (I - K_kH_k)\Delta_{k|k-1} - [0, K_k\Delta_k^v]$;

  $\Gamma_{k|k} = \Gamma_{k|k-1} - K_k^sS_k(K_k^s)'$;

endfor

Output: $\hat{u}_{k|k}$, $\Sigma_{k|k}$, $\Delta_{k|k}$, $\Gamma_{k|k}$, $\hat{\gamma}_{i,k|k}$

3.3.2 Discussions

We can obtain a batch form of the propagation of the skewness parameter $\Delta_{k|k-1}$. For each time i, define

$$K_i^p = F_iK_i = F_i\Sigma_{i|i-1}H'_iS_i^{-1} \tag{3.6}$$

$$F_i^p = F_i - K_i^pH_i \tag{3.7}$$

$$\Phi_{i,j}^p = \begin{cases} F_{i-1}^p \cdots F_j^p & i > j \\ I & i = j \end{cases} \tag{3.8}$$

$$\Delta_{k|k-1} = \begin{bmatrix}\Delta_{k|k-1}^x, & \Delta_{k|k-1}^w, & \Delta_{k|k-1}^v\end{bmatrix} \tag{3.9}$$

We have

$$\Delta^x_{k|k-1} = \Phi^p_{k,0}\Delta^x_0 \tag{3.10}$$

$$\Delta^w_{k|k-1} = \sum_{i=1}^{k} \Phi^p_{k,i} G_{i-1}\Delta^w_{i-1} \tag{3.11}$$

$$\Delta^v_{k|k-1} = -\sum_{i=1}^{k} \Phi^p_{k,i} K^p_{i-1}\Delta^v_{i-1} \tag{3.12}$$

The superscripts in (3.9) indicate the sources of the skewness that contribute to the state. The Φ^p_i in (3.7) has an identical algebraic expression with the so-called closed-loop state transition in the innovations model of $\hat{x}_{k|k-1}$ in the Kalman filter (Kailath et al., 2000, p.324), and correspondingly, K^p_i in (3.6) is of the same algebraic form as the predicted gain in the Kalman filter. Note that $K^p_0\Delta^v_0 = 0$ since no measurement is available in (3.1) at the initial time (so no v_0 is considered).

The initialization of $\Delta_{0|0}$ needs to be augmented to include the skewness from the process and measurements noise, so the initial density of x_0 is reformulated as:

$$x_0 \sim \mathrm{SG}([\hat{\gamma}_{1,0|0}, \hat{\gamma}_{2,0|0}], \hat{u}_{0|0}, \Omega_{0|0}) \tag{3.13}$$

where

$$\hat{\gamma}_{i,0|0} = \left[(\gamma^x_i)', (\gamma^w_i)', (\gamma^v_i)'\right]'$$

$$\Delta_{0|0} = [\Delta^x_{0|0}, 0, 0]$$

$$\Gamma_{0|0} = \mathrm{diag}(\Gamma^x, \Gamma^w, \Gamma^v)$$

$$\Omega_{0|0} = \begin{bmatrix} \Gamma_{0|0} & \Delta'_{0|0} \\ \Delta_{0|0} & \Sigma^x_0 \end{bmatrix}$$

One can verify that the above reformulation does not alter the density of the initial state x_0 in (3.2); that is, the density formulas of (3.2) and (3.13) are the same.

For point filtering, we can immediately obtain the corresponding MMSE skew-Gaussian filter along with the MSE matrix when applying (2.12) and (2.13) at each time index.

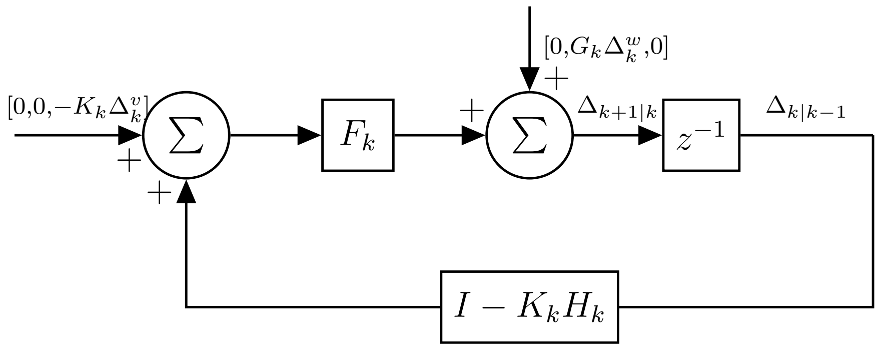

FIGURE 3.2: Flowchart of the propagation of $\Delta_{k|k-1}$

3.3.3 A Special Case—MMSE-SGF

A special case of the SGF is that the only skewness comes from the initial state, which may have applications where the sensor data scheduling applied to the initial state only (He et al., 2018).

By letting Δ_k^w and Δ_k^v equal to zero, we have a lemma for this special case.

Lemma 3.3. *Given system* (3.1) *with the linear skew-Gaussian assumption except that*

1*) $w_k \sim \mathcal{N}(0, Q_k)$, $v_k \sim \mathcal{N}(0, R_k)$, *and* $\{w_k\}$ *and* $\{v_k\}$ *are Gaussian white noises*

one cycle of the SGF of Theorems 3.1 and 3.2 for this special case degenerates as follows:

- *Prediction*
  - *Gaussian part: the same formulas as the general SGF*
  - *Skewness part: the same formulas as the general SGF except*

$$\Delta_{k|k-1} = F_{k-1}\Delta_{k-1|k-1}$$

- *Update:*
  - *Gaussian part: the same formulas as the general SGF*

– *Skewness part: the same formulas as the general SGF except*

$$K_k^s = \Delta'_{k|k-1} H'_k S_k^{-1}$$

$$\begin{aligned}\Delta_{k|k} &= \Delta_{k|k-1} - K_k H_k \Delta_{k|k-1} \\ &= \Sigma_{k|k} \Sigma_{k|k-1}^{-1} \Delta_{k|k-1}\end{aligned}$$

3.3.4 Computation Complexity

In this subsection we evaluate the computational complexity and memory requirements of the SGF.

Let n_x, n_z and n_w be the dimensions of the state x, measurement z, and process noise w. Recall that n_{γ^x}, n_{γ^w} and n_{γ^v} are the dimensions of γ_1^x, γ_1^w and γ_1^v defined in Section 3.2. Then, in each iteration,

- Computational cost: The SGF involves only matrix addition, multiplication and inverse.
  - Gaussian part: the same as the Kalman filter, which has $\mathcal{O}(n_x^3 + n_x^2 n_w + n_x^2 n_z + n_x n_z^2 + n_z^3)$ flops.
  - Skewness part: the main computation lies in the matrix multiplication, and the complexity is $\mathcal{O}((n_x n_z + n_z^2)(n_{\gamma^x} + n_{\gamma^w} + n_{\gamma^v}) + n_z(n_{\gamma^x} + n_{\gamma^w} + n_{\gamma^v})^2)$ flops.
- Memory requirements: The SGF stores only the filtered quantities at each time, and then they can be discarded.
  - Gaussian part: the same as the Kalman filter. The storage is $\mathcal{O}(n_x^2)$.
  - Skewness part: $\mathcal{O}((n_{\gamma^x} + n_{\gamma^w} + n_{\gamma^v})^2)$.

3.3.5 Comparison of Kalman filter and MMSE-SGF

While the Kalman filter can be viewed as an realization of the LMMSE estimation, the MMSE-SGF can not be. Even though all the parameters in SGF are propagated in a linear manner with respect to the measurements, the calculation of its MMSE estimator (more specially, the conditional mean) at each time step no longer follows a linear form, which can be seen from (2.9) and (2.12). In addition, unlike the Kalman filter whose recursibility is accomplished by the linear update directly from its estimator at previous time step, the MMSE-SGF has a recursive linear update from its parameters other than the MMSE estimator itself, thus, there is no contradiction to the theoretical results of recursibility in (Li, 2004).

3.4 Simulation Study

In this section, the proposed theoretical results of the SGF is demonstrated in an example of one-dimensional target tracking, where a target moves at a nearly constant velocity (CV) with state dynamics

$$x_k = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x_{k-1} + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} w_{k-1} \tag{3.14}$$

The state $x_k \triangleq [\mathrm{p}_k, \mathrm{v}_k]'$, the sampling interval $T = 1\mathrm{s}$, and a total motion duration is 30s.

The measurement equation is

$$z_k = \begin{bmatrix} 1, & 0 \end{bmatrix} x_k + v_k \tag{3.15}$$

where only the position is observed at each time.

The initial state x_0 is generated from a skew-Gaussian distribution with

$$u_0 = \begin{bmatrix} 10^3\mathrm{m} \\ 10\mathrm{m/s}^2 \end{bmatrix} \qquad \Sigma_0 = \begin{bmatrix} 100\mathrm{m}^2 & 0 \\ 0 & 25\mathrm{m}^2/\mathrm{s}^4 \end{bmatrix}$$

$$\Gamma^x = \sigma_x^2 = 12510 \qquad \Delta^x = \begin{bmatrix} 10^3 \\ 250 \end{bmatrix} \qquad \begin{bmatrix} \gamma_1^x \\ \gamma_2^x \end{bmatrix} = \begin{bmatrix} -10\sigma_x \\ 0 \end{bmatrix}$$

In this CV motion, both the target and the radar experience linear hidden truncation processes. Thus, w_k and v_k are both skew-Gaussian processes.

The parameters of the process noise $\{w_k\}$ are $u_k^w = 0$, $Q_k = 0.1\text{m}^2/\text{s}^2$, $\Gamma^w = 1.03$, $\Delta_k^w = 0.05$, and

$$\begin{bmatrix} \gamma_1^w \\ \gamma_2^w \end{bmatrix} = \begin{bmatrix} -10\sigma_w \\ 0 \end{bmatrix}$$

The parameters of the measurement noise $\{v_k\}$ are $u_k^v = 0$, $R_k = 10^2\ \text{m}^2$, $\Gamma^v = 2000$, $\Delta_k^v = 57.735$, and

$$\begin{bmatrix} \gamma_1^v \\ \gamma_2^v \end{bmatrix} = \begin{bmatrix} -10\sigma_v \\ 0 \end{bmatrix}$$

3.4.1 Density Filtering

In this simulation, the SGF and the state-of-the-art Kalman filter are compared over 200 Monte Carlo runs. Note that the Kalman filter is a Gaussian density filter, since a Gaussian density is determined by its first two moments. Figures 3.3 and 3.4 depict the contour plots of the posterior densities of one run by the SGF and the Kalman filter with the same level increments at time 5s and 20s, respectively. The strong skewness from the initial state mainly contribute to the skewness in Figure 3.3. As time goes, the impact from the initial state dies out, and the skewness from the process noise and measurement noise play key roles. As shown in Figure 3.4, the contour plot from the SGF at time 25s has a different skewed tendency compared with time 5s, and the plot is squeezed anti-diagonally, i.e., from the top-right to the bottom-left, compared with the symmetric plot from the Kalman filter.

Figures 3.5 and 3.6 plot the Pearson's moment coefficient of skewness of the position and the velocity on one run, defined as

$$\text{skewness} = \frac{\text{E}[x - \text{E}(x)]^3}{[\text{var}(x)]^{\frac{3}{2}}}$$

These theoretical skewness plots result from (2.13) and (2.14) of the SGF and the fact that the Kalman filter always gives zero skewness. The skewness at the early stages has three sources: the initial state, the process noise and the measurement noise. As time goes, the skewness effect due to the initial state dies out.

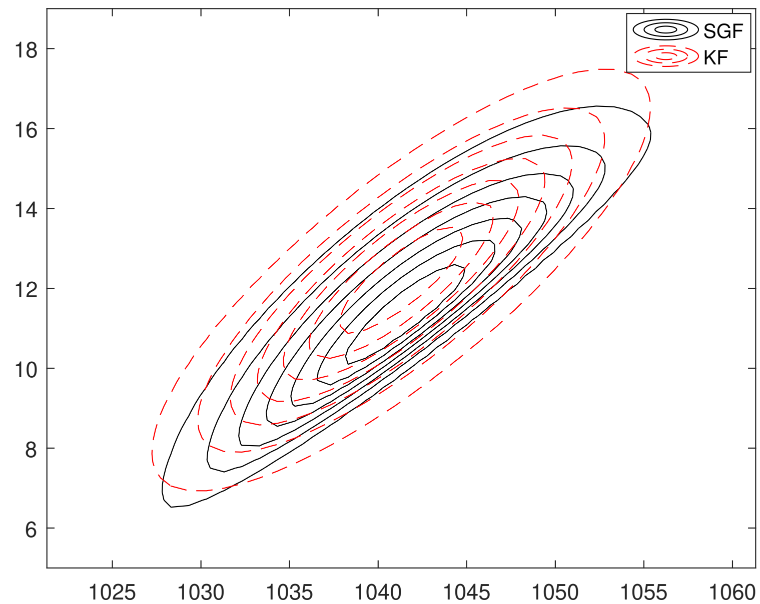

Figure 3.3: Contours at time 5s

3.4.2 MMSE Point Filtering

The MMSE-SGF (by using (2.12) and (2.13)) is also compared with the Kalman filter. Note that the Kalman filter performs sub-optimally because of the non-Gaussianity and the colored noises. As depicted in Figures 3.7 and 3.8, the MMSE-SGF also outperforms the Kalman filter in both position and velocity. The velocity has less difference, perhaps because the skewed measurement noise is directly imposed on the position, but not the velocity.

3.4.3 Computation Cost

Table 3.1 shows the relative computational costs in terms of execution time. The SGF and the Kalman filter use about the same computation resources, whereas the

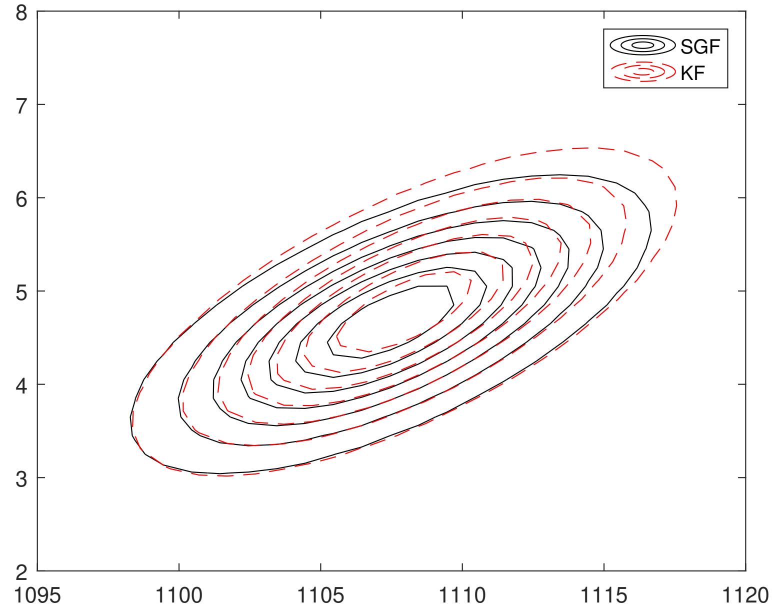

FIGURE 3.4: Contours at time 25s

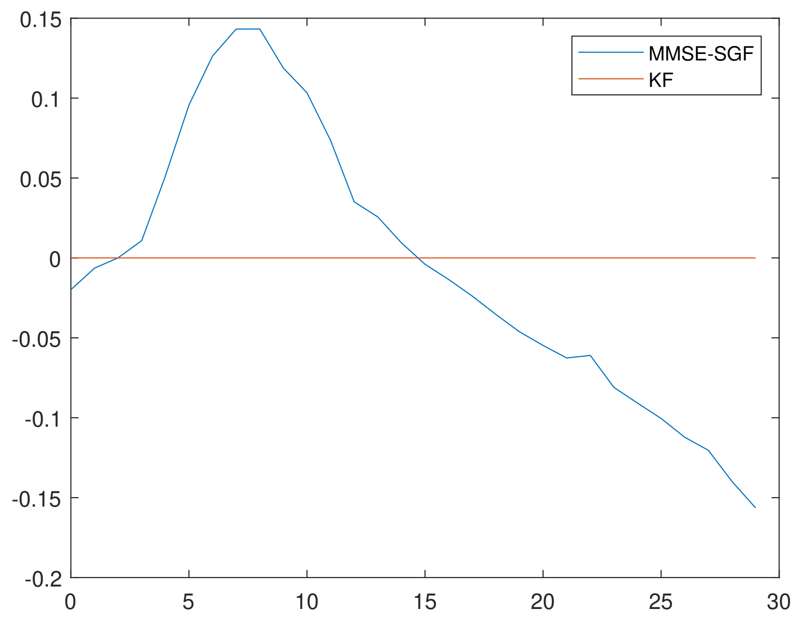

FIGURE 3.5: Skewness of Position

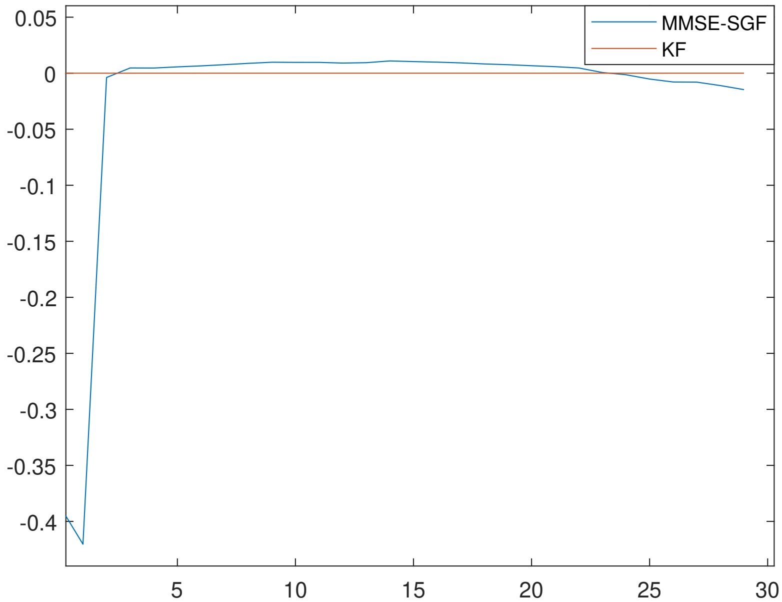

FIGURE 3.6: Skewness of Velocity

Table 3.1: Relative Computation Time

Methods	KF	SGF	MMSE-SGF	Integration
Relative Time	1	1.23	246.84	245.59

MMSE-SGF, as an application of our density SGF in this specific example, demands more computation resource to calculate the state estimate and error covariance based on (2.9) and (2.10), depending on the integration methods employed. It also shows that the main time consumption is due to numerical integration, which takes up to 99.5% in the total computation time of MMSE-SGF.

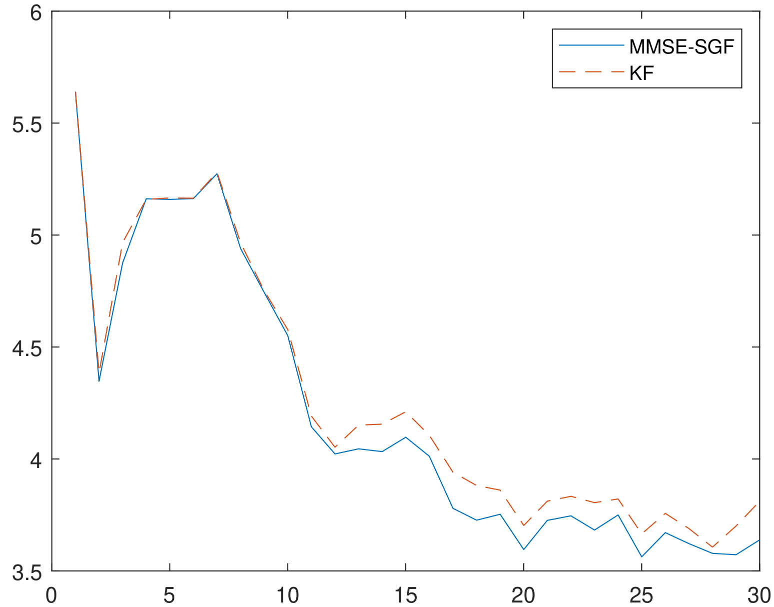

FIGURE 3.7: RMSEs of Position

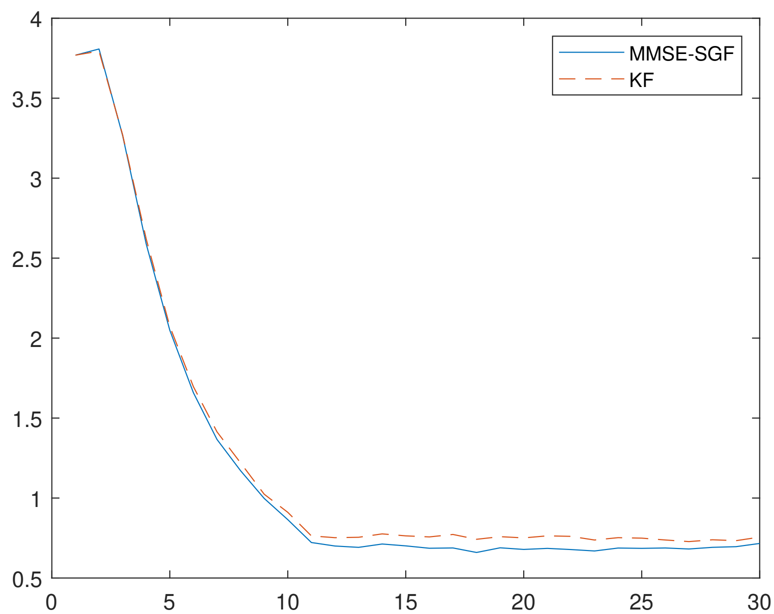

FIGURE 3.8: RMSEs of Velocity

4

Nonlinear Filtering Using Skew-Symmetric Representation of Distributions

Nothing occurs at random, but
everything for a reason and by
necessity.

Leucippus

4.1 Introduction

As is well known, nonlinear estimation in a closed-form has been established only for the linear Gaussian case and several special nonlinear cases (Li and Jilkov, 2010). The general nonlinear estimation poses a challenge since it requires propagating the probability density function (pdf) of the state, which is an infinite-dimensional problem (see, e.g., (Jazwinski, 2007)). Therefore, many practical nonlinear filtering methods in finite dimension have been developed by calling different numerical approximations. In this chapter, nonlinear point estimation, which estimates the quantity of interest without directly obtaining the posterior distribution of the state, is considered due to its simplicity and adequacy for many practical problems.

In general, non-linearity of many practical problems distorts symmetry and other nice properties of the posterior distribution of interest (Julier, 1998). However, most nonlinear point filtering methods, proposed under the framework of LMMSE estimation, exploit only the first two moments (i.e., the mean and the covariance) (Li and Jilkov, 2004). The most prevailing methods, which were shown in Section 1.4.2, are either restricted to Gaussian assumptions or reliant on the LMMSE update at the cost of neglecting potentially effective information about higher moments. So, there is room for improvement if information about higher moments is incorporated.

In order to model more complex nonlinear problems in simple tractable forms, especially the skewness and the kurtosis, a skew-symmetric representation of distributions, which has a product form of a symmetric pdf (known as the base pdf) times a perturbation function (known as the skewing function), is employed in this chapter. Such a representation of distributions, which strictly includes its symmetric base pdf as a special case, has additional parameters to capture the departure from the symmetry for practical problems. Furthermore, such a class may attain a reasonable compromise between mathematical tractability and shape flexibility of the pdf.

A skew-symmetric representation of distributions, in particular with a Gaussian base pdf, is flexible to model data sets abounding in areas such as economics, finance, engineering, and biomedical science. Its application to nonlinear filtering is the topic of this chapter. We develop a nonlinear filtering method based on a first-order skew-Gaussian representation. Contrary to the conventional nonlinear point filters, our proposed nonlinear filter potentially has higher estimation accuracy as a result of the additional information about the third moment involved and a nonlinear state update rule applied accordingly.

The chapter is organized as follows. A skew-symmetric representation of distributions is introduced in Section 4.2, and its application, a first-order skew-Gaussian nonlinear filter, is proposed in Section 4.3. Simulation results and relevant analysis are shown in Section 4.4.

4.2 Skew-Symmetric Representation of Distributions

In this section a skew-symmetric representation of distributions is introduced first. As shown in (Wang et al., 2004), it is noteworthy that any multivariate pdf admits a skew-symmetric representation. Next, a flexible skew-symmetric representation of distributions is presented, in which the skewing part is a polynomial. The main advantage of this representation is that it can approximate a skew-symmetric representation of distributions arbitrarily well if the polynomial in the skewing part has a high enough degree. Last, a first-order skew-Gaussian representation and its properties are discussed in detail, which is the backbone of our skew-Gaussian nonlinear filtering to be presented in Section 4.3.

4.2.1 Skew-Symmetric Representation

Definition 4.1. *Given any pdf* p*:* $\mathbb{R}^n \to \mathbb{R}_+$ *that is symmetric about a location parameter* $u \in \mathbb{R}^n$*, i.e.,* $p(u-x) = p(x-u)$*, and any function* $\pi : \mathbb{R}^n \to [0,1]$ *satisfying* $\pi(-x) = 1 - \pi(x)$*, a function of the following form is a legitimate pdf and is called a skew-symmetric pdf:*

$$2p(x-u)\pi(x-u) \tag{4.1}$$

For different problems, the symmetric base pdf $p(\cdot)$ can be chosen appropriately as the Gaussian distribution, the t-distribution, the elliptical distribution, etc. The so-called skewing function $\pi(\cdot)$ provides a skewed way to reallocate points of the base pdf $p(x)$.

Remark 4.1. *The skew-symmetric form* (4.1) *includes the symmetric pdf* $p(\cdot)$ *as a special case if* $\pi(x) = \frac{1}{2}$*.*

Remark 4.2 (Wang et al. (2004)). *The condition of the skewing function $\pi(x)$ can be easily satisfied if the following construction is obeyed for a skew-symmetric representation*

$$2p(x-u)G(\gamma(x-u)) \tag{4.2}$$

where $p(\cdot)$ is a symmetric pdf about 0, $G : \mathbb{R} \to [0,1]$ is the cumulative distribution function (cdf) of a continuous symmetric random variable whose pdf is symmetric about 0, and $\gamma : \mathbb{R}^n \to \mathbb{R}$ is any odd function.

Remark 4.3 (Wang et al. (2004)). *If the same base pdf $p(\cdot)$ is used in (4.1) and (4.2), the two classes of distributions are equivalent. Without loss of generality, let $u = 0$. Let $\pi(\cdot) \triangleq G[\gamma(\cdot)] : \mathbb{R}^n \to \mathbb{R}$. Then,*

$$\pi(-x) = G[\gamma(-x)] = G[-\gamma(x)] = 1 - G[\gamma(x)] = 1 - \pi(\cdot)$$

So, every pdf in the form of (4.2) must be in the form of (4.1). Conversely, for every $\pi(\cdot)$ in (4.1), there exists a (actually many) cdf $G(\cdot)$ (whose pdf is symmetric about 0) with the inverse function $G^{-1}(\cdot)$ such that

$$\gamma(x) = G^{-1}[\pi(x)] = G^{-1}[1 - \pi(-x)] = -\gamma(-x)$$

that is, $\gamma(\cdot)$ is odd. So,

$$G[\gamma(x)] = G[G^{-1}[\pi(x)]] = \pi(x)$$

that is, every function in the form of (4.1) can be written in the form of (4.2).

Remark 4.4 (Azzalini and Capitanio (2003)). *The legitimacy for (4.1) to be a pdf can be easily validated by the equivalent form (4.2). That is, let $X \sim p$ and $Y \sim G$ be two independent random variables. Then*

$$\begin{aligned}\int_{\mathbb{R}^n} G(\gamma(z-u))p(z-u)dz = \int_{\mathbb{R}^n} G(\gamma(z))p(z)dz &= E[P\{Y \le \gamma(X) \mid X\}] \\ &= P\{Y \le \gamma(X)\} = \frac{1}{2}\end{aligned}$$

The last equality above holds because Y and $\gamma(X)$ are independent random variables whose pdfs are both symmetric about 0. Thus,

$$\int_{\mathbb{R}^n} 2p(x-u)G[\gamma(x-u)]dx = 1$$

So, $f(x) \triangleq 2p(x-u)\pi(x-u) \geq 0$, $\forall x$ and $\int_{\mathbb{R}^n} f(x)dx = 1$, and thus $f(x)$ is a pdf.

Remark 4.5 (Wang et al. (2004)). *Given any pdf g: $\mathbb{R}^n \to \mathbb{R}_+$ and $x_0 \in \mathbb{R}^n$, there exists a unique skew-symmetric representation of $g(x)$:*

$$g(x) = 2p_g(x-x_0)\pi_g(x-x_0)$$

where

$$p_g(x) = \frac{g(x_0+x)+g(x_0-x)}{2}$$

$$\pi_g(x) = \frac{g(x_0+x)}{g(x_0+x)+g(x_0-x)}$$

4.2.2 Flexible Skew-Symmetric Distributions

Let $x \in \mathbb{R}^n$ and $\mathcal{P}_{2m-1}(x)$ denote an odd polynomial of highest degree $2m-1$, meaning that the degrees $\sum_{j=1}^n i_j$ of all terms $x_1^{i_1}x_2^{i_2}\cdots x_n^{i_n}$ in $\mathcal{P}_{2m-1}(x)$ are all odd and their maximum is $2m-1$. A flexible skew-symmetric representation is the restriction of (4.2)), that is,

$$2p(x-u)G(\mathcal{P}_{2m-1}(x-u)) \tag{4.3}$$

where $p(\cdot)$ is a symmetric pdf about 0, and $G : \mathbb{R} \to [0,1]$ is an arbitrary cdf of a continuous symmetric random variable mentioned above. In practice, a popular choice of $G(\cdot)$ is the univariate cdf corresponding to the symmetric pdf $p(\cdot)$.

Like the Stone-Weierstrass theorem, the following proposition proved in (Ma and Genton, 2004) shows that a skew-symmetric representation (4.1) can be approximated arbitrarily well by a flexible skew-symmetric representation (4.3) for $\mathcal{P}_{2m-1}(x)$ of an appropriate degree.

Proposition 4.1 (Ma and Genton (2004)). *Under the condition that both* (4.2) *and* (4.3) *have the same symmetric base pdf* $p(\cdot)$, *and* $p(\cdot)$ *and* $\pi(\cdot)$ *are continuous, the class of flexible skew-symmetric distributions is dense in the class of skew-symmetric distributions under the* L^{∞} *norm.*

The above proposition encourages us to seek for a polynomial approximation in the skewing part. Moreover, it is desirable that such an approximation gives a simple solution to nonlinear point estimation problems.

4.2.3 First-Order Skew-Gaussian Distribution

One typical representative of a flexible skew-symmetric class of distributions is the first-order skew-Gaussian distribution (Azzalini and Capitanio, 2003), denoted by $\mathrm{FOSG}(x; u, \Omega, \alpha)$. Its pdf is

$$2\mathcal{N}(x; u, \Omega)\Phi(\alpha'(x - u)) \tag{4.4}$$

where $\mathcal{N}(x; u, \Omega)$ denotes the Gaussian pdf with mean u and covariance Ω, $\Phi(\cdot)$ is the cdf of the univariate standard Gaussian distribution, and the skewness parameter α is a column vector of the same dimension as x.

The class of first-order skew-Gaussian distributions provides a skew extension of the multivariate Gaussian family by adding an extra parameter α to regulate the skewness. Note that (4.4) with $\alpha = 0$ reduces to Gaussian pdf $\mathcal{N}(x; u, \Omega)$.

Several one-dimensional first-order skew-Gaussian distributions are shown in Figure 4.1, where $u = 0$, $\Omega = 1$, and $\alpha = 0$, -1.5, and 1, respectively.

One major reason for choosing such a form is its mathematical tractability and simplicity when dealing with some fundamental operations, such as linear transformation, marginalization and conditioning, and these properties lay a foundation for the nonlinear filtering method to be proposed in Section 4.3.

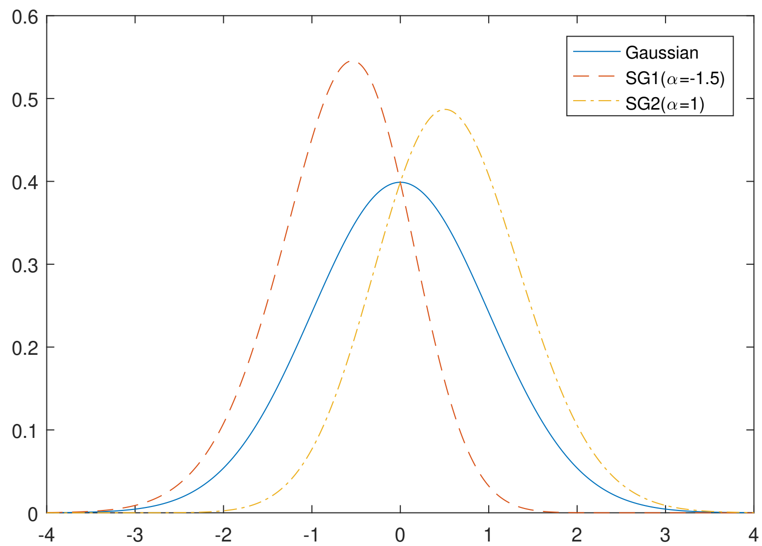

FIGURE 4.1: One-dimensional skew-Gaussian density with different values of the skewness parameter

Define a parameter δ with the following relationship with the skewness parameter α:

$$\delta = \frac{\Omega\alpha}{\sqrt{1+\alpha'\Omega\alpha}}, \qquad \alpha = \frac{\Omega^{-1}\delta}{\sqrt{1-\delta'\Omega^{-1}\delta}} \tag{4.5}$$

where, as one can clearly see, δ and α are uniquely determined by each other.

To study central moments of the density (4.4), we use its cumulant generating function (CGF), given by

$$K(t) = \frac{1}{2}t'\Omega t + u't + \log(2\Phi(\delta' t)) \tag{4.6}$$

The first three central moments, derived from (4.6), are

$$\begin{aligned} E(x) = u + \sqrt{\frac{2}{\pi}}\delta, \qquad \operatorname{cov}(x) = \Omega - \frac{2}{\pi}\delta\delta' \\ E[(x-Ex)((x-Ex)')^{\otimes 2}] = \left(\frac{4}{\pi}-1\right)\sqrt{\frac{2}{\pi}}\delta\,\operatorname{vec}'(\delta\delta') \end{aligned} \tag{4.7}$$

where $\otimes$ denotes the kronecker product, $A^{\otimes k} = \underbrace{A \otimes A \otimes \cdots \otimes A}_{k \text{ times}}$ and the operator vec$(\cdot)$ converts a matrix to a column vector by stacking column by column.

A derivation of (4.7) is presented in Appendix E.

The first-order skew-Gaussian distribution has many properties that resemble the Gaussian distribution (Azzalini and Capitanio, 2003).

Suppose that $x = [x_1', x_2']' \sim \text{FOSG}(x; u, \Omega, \alpha)$ and the vectors are partitioned as

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \qquad \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{bmatrix}$$

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \qquad \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$$

Marginal distribution

The marginal distribution of x_1 is

$$2\mathcal{N}(x_1; u_1, \Omega_{11})\Phi(\bar{\alpha}_1'(x_1 - u_1))$$

where

$$\bar{\alpha}_1 = \frac{\alpha_1 + \Omega_{11}^{-1}\Omega_{12}\alpha_2}{\sqrt{1 + \alpha_2'\Omega_{2|1}\alpha_2}} = \frac{\Omega_{11}^{-1}\delta_1}{\sqrt{1 - \delta_1'\Omega_{11}^{-1}\delta_1}}$$

and

$$\Omega_{2|1} = \Omega_{22} - \Omega_{12}'\Omega_{11}^{-1}\Omega_{12}$$

The corresponding CGF is

$$K_{x_1}(t) = \frac{1}{2}t'\Omega_{11}t + u_1't + \log(2\Phi(\delta_1't)) \tag{4.8}$$

Conditional distribution

The CGF of the conditional distribution of x_1 given x_2 is

$$K_{1|2}^c(t) = \frac{1}{2}t'\Omega_{1|2}t + u_{1|2}'t + \log(\Phi(x_0 + \delta_{1|2}'t)) - \log\Phi(x_0) \tag{4.9}$$

where

$$\begin{aligned}
u_{1|2} &= u_1 + \Omega_{12}\Omega_{22}^{-1}(x_2 - u_2) \\
\Omega_{1|2} &= \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{12}' \\
\delta_{1|2} &= \frac{\Omega_{1|2}\alpha_1}{\sqrt{1 + \alpha_1'\Omega_{1|2}\alpha_1}} \\
\bar{\alpha}_2 &= \frac{\alpha_2 + \Omega_{22}^{-1}\Omega_{21}\alpha_1}{\sqrt{1 + \alpha_1'\Omega_{1|2}\alpha_1}} = \frac{\Omega_{22}^{-1}\delta_2}{\sqrt{1 - \delta_2'\Omega_{22}^{-1}\delta_2}} \\
x_0 &= \bar{\alpha}_2'(x_2 - u_2)
\end{aligned}$$

Comparing (4.8) with (4.9): these two CGFs share a similar form except that an additional constant term is added to (4.9) and the update terms $u_{1|2}$ and $\Omega_{1|2}$ follow the exact update rule, as those in LMMSE update (2.1)–(2.2).

The first three conditional central moments, derived from (4.9), are

$$\begin{aligned}
E(x_1|x_2) &= u_{1|2} + \frac{\mathcal{N}(x_0)}{\Phi(x_0)}\delta_{1|2} \\
\operatorname{cov}(x_1|x_2) &= \Omega_{1|2} - \left[x_0\frac{(x_0)}{\Phi(x_0)} + \big(\frac{(x_0)}{\Phi(x_0)}\big)^2\right]\delta_{1|2}\delta_{1|2}' \\
\bar{M}^3(x_1|x_2) &= E[(x_1 - E(x_1|x_2))(x_1 - E(x_1 \mid x_2))'^{\otimes 2} \mid x_2] \qquad (4.10) \\
&= \left[(x_0^2 - 1)\frac{(x_0)}{\Phi(x_0)} + x_0\big(\frac{\mathcal{N}(x_0)}{\Phi(x_0)}\big)^2\right. \\
&\quad \left. + 2\big(\frac{\mathcal{N}(x_0)}{\Phi(x_0)}\big)^3\right]\delta_{1|2}\operatorname{vec}'(\delta_{1|2}\delta_{1|2}')
\end{aligned}$$

where $\mathcal{N}(x_0)$ and $\Phi(x_0)$ denote the univariate standard Gaussian pdf and cdf, respectively, evaluated at x_0.

A derivation of (4.10) is similar to that of (4.7), which is given in the Appendix E.

Linear transformation

Suppose that $x \sim \text{FOSG}(x; u, \Omega, \alpha)$ and A is a non-singular matrix such that $A'\Omega A$ is of full rank. Then

$$A'x \sim \text{FOSG}(x; Au, A'\Omega A, A^{-1}\alpha)$$

4.3 Nonlinear Filtering Using Skew-Symmetric Representation

As discussed in Section 4.2.1, any pdf can be decomposed as a symmetric base pdf times a skewing function. Furthermore, the proposition in Section 4.2.2 demonstrates that the special skewing function of an odd-polynomial is capable to approximate any pdf. Considering the mathematical tractability and simplicity, a nonlinear filtering method based on a first-order skew-Gaussian distribution is proposed to incorporate the information of the third central moment or the skewness of the state.

4.3.1 Design of First-Order Skew-Gaussian Filtering

Similar in the spirit to the Gaussian filter in the first-order skew-Gaussian filter (FOSGF) is based on the following ideas:

1. Successively approximate the pdfs of $y_{k|k-1}$ (with $y = [x', z']'$) and $x_{k|k}$ at each stage through a first-order skew-Gaussian distribution by moment matching:

$$y_{k|k-1} \sim \text{FOSG}(y_k; u^y_{k|k-1}, \Omega^y_{k|k-1}, \alpha^y_{k|k-1}) \tag{4.11}$$

$$x_{k|k} \sim \text{FOSG}(x_k; u^x_{k|k}, \Omega^x_{k|k}, \alpha^x_{k|k}) \tag{4.12}$$

2. Apply the nonlinear estimation rule (4.10), which is based on Bayes' rule, to the state update:

$$p(x_k \mid z^k) = \frac{p(x_k, z_k \mid z^{k-1})}{\int p(x_k, z_k \mid z^{k-1}) dx_k}$$

Note that if $[(\alpha^y_{k|k-1})', (\alpha^x_{k|k})'] = 0$, the FOSGF reduces to the conventional Gaussian filter (Ito and Xiong, 2000).

Let $\hat{x}_{k-1|k-1}$, $P_{k-1|k-1}$ and $\alpha^x_{k-1|k-1}$ denote the estimates of the state, covariance and skewness parameter at time $k-1$, respectively. Given these quantities, one cycle of the FOSGF is summarized as follows.

1. **Prediction**:

   Under assumption (4.12) at time $k-1$, the parameters $u^x_{k-1|k-1}$ and $\Omega^x_{k-1|k-1}$ are, by using (4.7),

$$u^x_{k-1|k-1} = \hat{x}_{k-1|k-1} - \sqrt{\frac{2}{\pi}}\delta^x_{k-1|k-1}$$

$$\Omega^x_{k-1|k-1} = P_{k-1|k-1} + \frac{2}{\pi}\delta^x_{k-1|k-1}(\delta^x_{k-1|k-1})'$$

   where $\delta^x_{k-1|k-1}$ can be calculated by (4.5) as

$$\delta^x_{k-1|k-1} = \frac{\Omega\alpha^x_{k-1|k-1}}{\sqrt{1+(\alpha^x_{k-1|k-1})'\Omega\alpha^x_{k-1|k-1}}}$$

   (a) Calculate the predicted mean of the state

$$\begin{aligned}\hat{x}_{k|k-1} &= E[x_k \mid z^{k-1}] \\ &\approx \int (f_{k-1}(x) + w_{k-1})\mathrm{FOSG}(x; u^x_{k-1|k-1}\Omega^x_{k-1|k-1}, \alpha^x_{k-1|k-1})\ \mathrm{d}x \\ &= \int f_{k-1}(x)\Phi((\alpha^x_{k-1|k-1})'(x-u^x_{k-1|k-1}))\mathcal{N}(x; u^x_{k-1|k-1}, \Omega^x_{k-1|k-1})\ \mathrm{d}x\end{aligned}$$

   (b) Calculate the predicted covariance of the state

$$\begin{aligned}P_{k|k-1} &= E[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})' \mid z^{k-1}] \\ &\approx \int (x_k - \hat{x}_{k|k-1})(\cdot)'\mathrm{FOSG}(x; u^x_{k-1|k-1}, \Omega^x_{k-1|k-1}, \alpha^x_{k-1|k-1})\ \mathrm{d}x \\ &= \int (f_{k-1}(x) - \hat{x}_{k|k-1})(\cdot)'\Phi((\alpha^x_{k-1|k-1})'(x-u^x_{k-1|k-1})) \\ &\quad \times \mathcal{N}(x; u^x_{k-1|k-1}, \Omega^x_{k-1|k-1})\ \mathrm{d}x + Q_k\end{aligned}$$

(c) Calculate the predicted third central moment of the state

$$
\begin{aligned}
\bar{M}^{3,x}_{k|k-1} &= E[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})'^{\otimes 2} \mid z^{k-1}] \\
&\approx \int (x_k - \hat{x}_{k|k-1})(\cdot)'^{\otimes 2}\Phi((\alpha^x_{k-1|k-1})'(x - u^x_{k-1|k-1})) \\
&\quad \times \mathcal{N}(x; u^x_{k-1|k-1}, \Omega^x_{k-1|k-1})dx
\end{aligned}
$$

With assumption (4.11) at time $k-1$ and the marginalization property, we have

$$
x_{k|k-1} \sim \text{FOSG}(x_k; u^x_{k|k-1}, \Omega^x_{k|k-1}, \alpha^x_{k|k-1})
$$

where $\delta^x_{k|k-1}$ is computed directly by moment matching, which will be discussed in Section 4.3.2, and $\alpha^x_{k|k-1}$ can be computed by (4.5) as

$$
a^x_{k|k-1} = \frac{(\Omega^x_{k|k-1})^{-1}\delta^x_{k|k-1}}{\sqrt{1 - (\delta^x_{k|k-1})'(\Omega^x_{k|k-1})^{-1}\delta^x_{k|k-1}}}
$$

and

$$
\begin{aligned}
u^x_{k|k-1} &= \hat{x}_{k|k-1} - \sqrt{\frac{2}{\pi}}\delta^x_{k|k-1} \\
\Omega^x_{k|k-1} &= P_{k|k-1} + \frac{2}{\pi}\delta^x_{k|k-1}(\delta^x_{k|k-1})'
\end{aligned}
$$

2. **Update**:

(a) Calculate the predicted mean of the measurement

$$
\begin{aligned}
\hat{z}_{k|k-1} &= E[z_k \mid z^{k-1}] \\
&\approx \int (h_k(x) + v_k)\text{FOSG}(x; u^x_{k|k-1}, \Omega^x_{k|k-1}, \alpha^x_{k|k-1})\ \mathrm{d}x \\
&= \int h_k(x)\Phi((\alpha^x_{k|k-1})'(x - u^x_{k|k-1}))\mathcal{N}(x; u^x_{k|k-1}, \Omega^x_{k|k-1})\ \mathrm{d}x
\end{aligned}
$$

(b) Calculate the predicted covariance of the measurement

$$\begin{aligned} P_z &= E[(z_k - \hat{z}_{k|k-1})(z_k - \hat{z}_{k|k-1})' \mid z^{k-1}] \\ &\approx \int (h_k(x) - \hat{z}_{k|k-1})(\cdot)'\Phi((\alpha^x_{k|k-1})'(x - u^x_{k|k-1})) \\ &\quad \times \mathcal{N}(x; u^x_{k|k-1}, \Omega^x_{k|k-1})\ \mathrm{d}x + R_k \end{aligned}$$

(c) Calculate the predicted cross-covariance between the state and the measurement

$$\begin{aligned} P_{xz} &= E[(x_k - \hat{x}_{k|k-1})(z_k - \hat{z}_{k|k-1})' \mid z^{k-1}] \\ &\approx \int [(x - \hat{x}_{k|k-1})][h_k(x) - \hat{z}_{k|k-1}]'\Phi((\alpha^x_{k|k-1})'(x - u^x_{k|k-1})) \\ &\quad \times \mathcal{N}(x; u^x_{k|k-1}, \Omega^x_{k|k-1})\ \mathrm{d}x \end{aligned}$$

(d) Calculate the predicted third central moment of the measurement

$$\begin{aligned} \bar{M}^{3,z}_{k|k-1} &= E[(z_k - \hat{z}_{k|k-1})(z_k - \hat{z}_{k|k-1})'^{\otimes 2} \mid z^{k-1}] \\ &\approx \int (z_k - \hat{z}_{k|k-1})(\cdot)'^{\otimes 2}\Phi((\alpha^x_{k|k-1})'(x - u^x_{k|k-1})) \\ &\quad \times \mathcal{N}(x; u^x_{k|k-1}, \Omega^x_{k|k-1})\ \mathrm{d}x \end{aligned}$$

With assumption (4.11) at time k and the marginalization property, we have

$$z_{k|k-1} \sim \mathrm{FOSG}(z_k; u^z_{k|k-1}, \Omega^z_{k|k-1}, \alpha^z_{k|k-1})$$

where $\delta^z_{k|k-1}$is computed directly by moment matching, which will be discussed in Section 4.3.2, and $\alpha^z_{k|k-1}$ can be computed by (4.5) as

$$a^z_{k|k-1} = \frac{(\Omega^z_{k|k-1})^{-1}\delta^z_{k|k-1}}{\sqrt{1 - (\delta^z_{k|k-1})'(\Omega^z_{k|k-1})^{-1}\delta^z_{k|k-1}}}$$

and

$$u^z_{k|k-1} = \hat{z}_{k|k-1} - \sqrt{\frac{2}{\pi}}\delta^z_{k|k-1}$$

$$\Omega^z_{k|k-1} = P_z + \frac{2}{\pi}\delta^z_{k|k-1}(\delta^z_{k|k-1})'$$

$$\Omega^{xz}_{k|k-1} = P_{xz} + \frac{2}{\pi}\delta^x_{k|k-1}(\delta^z_{k|k-1})'$$

Stacking the above results at time k, the jointly distribution of the state and the measurement (4.11) is

$$\begin{bmatrix} x_{k|k-1} \\ z_{k|k-1} \end{bmatrix} \sim \text{FOSG}\left(\begin{bmatrix} x_k \\ z_k \end{bmatrix}; u^y_{k|k-1}, \Omega^y_{k|k-1}, \alpha^y_{k|k-1} \right)$$

where

$$u^y_{k|k-1} = \begin{bmatrix} u^x_{k|k-1} \\ u^z_{k|k-1} \end{bmatrix}$$

$$\alpha^y_{k|k-1} = \begin{bmatrix} \alpha^x_{k|k-1} \\ \alpha^z_{k|k-1} \end{bmatrix}$$

$$\Omega^y_{k|k-1} = \begin{bmatrix} \Omega^x_{k|k-1} & \Omega^{xz}_{k|k-1} \\ (\Omega^{xz}_{k|k-1})' & \Omega^z_{k|k-1} \end{bmatrix}$$

Update the mean $\hat{x}_{k|k}$ and covariance $P_{k|k}$ by (4.10):

$$\hat{x}_{k|k} = E(x_k \mid z^k) = u^x_{k|k} + \frac{\mathcal{N}(\tilde{z}_{k|k-1})}{\Phi(\tilde{z}_{k|k-1})}\delta_{k|k}$$

$$\begin{aligned} P_{k|k} &= E(x_k - \hat{x}_{k|k} \mid z^k)(\cdot)' \\ &= \Omega^x_{k|k} - \left[\tilde{z}_{k|k-1}\frac{\mathcal{N}(\tilde{z}_{k|k-1})}{\Phi(\tilde{z}_{k|k-1})} + (\frac{\mathcal{N}(\tilde{z}_{k|k-1})}{\Phi(\tilde{z}_{k|k-1})})^2\right]\delta_{k|k}(\delta_{k|k})' \end{aligned}$$

where

$$u^x_{k|k} = u^x_{k|k-1} + \Omega^{xz}_{k|k-1}(\Omega^z_{k|k-1})^{-1}(z_{k|k-1} - u^z_{k|k-1})$$

$$\Omega^x_{k|k} = \Omega^x_{k|k-1} - \Omega^{xz}_{k|k-1}(\Omega^z_{k|k-1})^{-1}(\Omega^{xz}_{k|k-1})'$$

$$\tilde{z}_{k|k-1} = (\bar{\alpha}^z_{k|k-1})'(z_{k|k-1} - u^z_{k|k-1})$$

$$\bar{\alpha}^z_{k|k-1} = \frac{\alpha^z_{k|k-1} + (\Omega^z_{k|k-1})^{-1}(\Omega^{xz}_{k|k-1})'\alpha^x_{k|k-1}}{\sqrt{1 + (\alpha^x_{k|k-1})'\Omega^x_{k|k}\alpha^x_{k|k-1}}}$$

$$\delta_{k|k} = \frac{\Omega^x_{k|k}\alpha^x_{k|k-1}}{\sqrt{1 + (\alpha^x_{k|k-1})'\Omega^x_{k|k}\alpha^x_{k|k-1}}}$$

and the skew parameter $\delta^x_{k|k}$ is updated by equating the third central moment of the unconditional (4.7) and conditional ((4.10)) first-order skew-Gaussian distribution: $\delta^x_{k|k} = \sqrt[3]{c}\delta_{k|k}$, where

$$c = \frac{(\tilde{z}^2_{k|k-1} - 1)\frac{\mathcal{N}(\tilde{z}_{k|k-1})}{\Phi(\tilde{z}_{k|k-1})} + \tilde{z}_{k|k-1}\left(\frac{(\tilde{z}_{k|k-1})}{\Phi(\tilde{z}_{k|k-1})}\right)^2 + 2\left(\frac{(\tilde{z}_{k|k-1})}{\Phi(\tilde{z}_{k|k-1})}\right)^3}{(\frac{4}{\pi} - 1)\sqrt{\frac{2}{\pi}}}$$

Note that the integrals involved in FOSGF are all Gaussian-type integrals, and traditional numerical integration methods (e.g., the Gauss-Hermite quadrature rule) can be employed.

4.3.2 Determination of Skewness Parameter

The determination of the skewness parameter α or δ is crucial in our nonlinear filtering method. From (4.7) or (4.10), one can see that δ is totally determined by the third central moment. However, there exists mismatch in structure between the analytic form and the numerical results. In other words, the following holds in general,

$$E[(x - \bar{x})(\cdot)'^{\otimes 2}] = E[(x - \bar{x})\text{vec}'((x - \bar{x})(\cdot)')] \neq a\delta\,\text{vec}'(\delta\delta')$$

where $a = (\frac{4}{\pi} - 1)\sqrt{\frac{2}{\pi}}$ is the coefficient in (4.7).

Because of the relationship between the parameter δ and the skewness of the assumed density, which is usually defined by its third central moment, we explore some skewness measures to estimate δ by moment matching.

In this section, three unnormalized versions of the multivariate skewness measure are employed, and the corresponding δ exists in each case, which is also summarized in Table 4.1.

1. The marginal third central moment $s_1(x) = E[x_1^3, \cdots, x_n^3]'$, and correspondingly

$$\delta_1 = \sqrt[3]{\frac{s_1}{a}}$$

2. The skewness measure $s_2(x) = E(\|x\|^2 x) = E[(\sum_{i=1}^n x_i^2)x]$ proposed in (Móri et al., 1994), and correspondingly

$$\delta_2 = \mathrm{sgn}(s_2)\frac{s_2}{a^{\frac{1}{3}}(s_2' s_2)^{\frac{1}{3}}}$$

where $\mathrm{sgn}(x)$ is an odd function that extracts the sign of x.

3. The skewness measure $s_3(x) = E[(\mathbf{1}_n x')'(\cdot)x] = E[(\sum_{i,j=1}^n x_i x_j)x]$ presented in (Ito and Xiong, 2000), and correspondingly

$$\delta_3 = \frac{s_3}{a^{\frac{1}{3}}(\mathbf{1}_n' s_3)^{\frac{2}{3}}}$$

where $\mathbf{1}_n$ is the n-dimensional column vector of all 1's.

Evidently, all measures provide n-dimensional skewness characteristics by elaborately selecting some components from the third central moment, of which only the third method includes all the components.

Table 4.1: δ's for different multivariate skewness measure

Skewness measures	δ's
$s_1(x) = E[x_1^3, \cdots, x_n^3]'$	$\delta_1 = \sqrt[3]{\frac{s_1}{a}}$
$s_2(x) = E(\Vert x\Vert^2 x) = E[(\sum_{i=1}^n x_i^2)x]$	$\delta_2 = \mathrm{sgn}(s_2)\frac{s_2}{a^{\frac{1}{3}}(s_2' s_2)^{\frac{1}{3}}}$
$s_3(x) = E[(\mathbf{1}_n x')'(\cdot)x] = E[(\sum_{i,j=1}^n x_i x_j)x]$	$\delta_3 = \frac{s_3}{a^{\frac{1}{3}}(\mathbf{1}_n' s_3)^{\frac{2}{3}}}$

4.4 Simulation Results

In this section, the performance of the FOSGF is demonstrated over a reentry vehicle (RV) problem, and is compared with UF and GHF.

The RV problem is to estimate accurately the position and velocity of a vehicle, whose trajectory is illustrated in Figure 4.2, by a radar with range and bearing measurements. This type of tracking problem is one of the most extensively studied applications considered by many authors (see, e.g., (Julier and Uhlmann, 2004; Daum, 2005; Athans et al., 1968)). The nonlinearity of this problem increases as the vehicle is approaching the earth surface.

The dynamic, after the discretization, is given as

$$\begin{aligned}
x_1(k) &= x_1(k-1) + Tx_3(k-1) \\
x_2(k) &= x_2(k-1) + Tx_4(k-1) \\
x_3(k) &= x_3(k-1) + T[D(k-1)x_3(k-1) \\
&\quad + G(k-1)x_1(k-1) + v_1(k-1)] \\
x_4(k) &= x_4(k-1) + T[D(k-1)x_4(k-1) \\
&\quad + G(k-1)x_2(k-1) + v_2(k-1)] \\
x_5(k) &= x_5(k-1) + Tv_3(k-1)
\end{aligned} \tag{4.13}$$

where $D(k)$ is a drag-related force term, $G(k)$ is a gravity-related force term, and $v(k) = [v_1(k), v_2(k), v_3(k)]'$ is zero-mean process Gaussian noise.

The force terms are given by

$$D(k) = \beta_0 \exp(x_5(k)) \exp(\frac{R_0 - R(k)}{H_0}) V(k)$$

$$G(k) = -\frac{Gm_0}{R^3(k)}$$

where

$$R(k) = \sqrt{x_1^2(k) + x_2^2(k)}$$

$$V(k) = \sqrt{x_3^2(k) + x_4^2(k)}$$

The motion of the vehicle is measured by a radar located at $(x_r, y_r) = (6374.05\,\text{km}, 0\,\text{km})$:

$$r_r(k) = \sqrt{(x_1(k) - x_r)^2 + (x_2(k) - y_r)^2} + w_1(k)$$

$$\theta(k) = \tan^{-1}\left(\frac{x_2(k) - y_r}{x_1(k) - x_r}\right) + w_2(k)$$

where $w(k) = [w_1(k), w_2(k)]'$ consists of two zero-mean independent Gaussian noises with variances of 1 km and 0.017 rad, respectively.

In this two-dimensional scenario, $\beta_0 = -0.59783$, $H_0 = 13.406\,\text{km}$, $Gm_0 = 3.9860 \times 10^5\,\text{km}^3 / \text{s}^2$, and $R_0 = 6374\,\text{km}$. The sampling time T was 0.5 s with 150 total time indices. All the parameter adopted from (Julier and Uhlmann, 2004).

The true initial conditions for the vehicle were

$$x_0^{\text{true}} = [6500.4, 349.14, -1.8093, -6.7967, 0.6932]'$$

$$P_0^{\text{true}} = \text{diag}([10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-6}])$$

The initial conditions assumed by the filter were

$$x_0^{\text{true}} = [6500.4, 349.14, -1.8093, -6.7967, 1]'$$

$$P_0^{\text{true}} = \text{diag}([10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 1])$$

The 3-point Gauss-Hermite quadrature rule was applied in both GHF and FOSGF, so 9 sigma points were used in UF, and 243 quadrature points were used in both GHF and FOSGF.

For simplicity, the skewness parameter δ in the update (4.12) was set to 0, and the third method of determining the skewness parameter was used in the prediction (4.11), which was discussed in Section 4.3.2.

The results of 100 Monte Carlo runs are depicted in Figure 4.3 and Figure 4.4. The FOSGF (solid line) outperforms both UF (dashed line) and GHF (dash-dot line) regarding the position error, and the UF performs the worst since it used the fewest quantization points. However, the FOSGF and GHF, which outperform the UF, have similar performance in velocity.

Table 4.2 shows the computational costs of the three filtering methods in terms of execution time. The FOSGF needs more computational resources, which stems from the calculation of the third moments and the matrix inverse operation for the skewness parameter (α, δ) at each step of filtering.

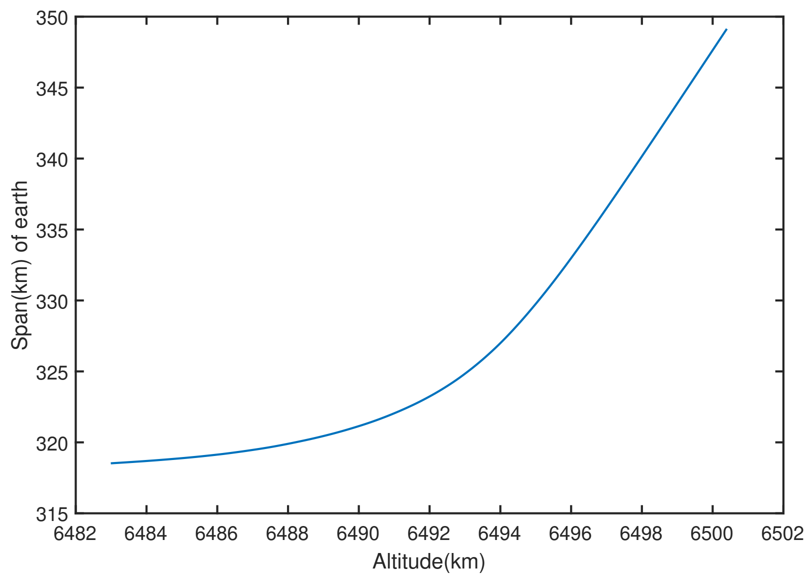

FIGURE 4.2: Trajectory of the two-dimensional RV problem.

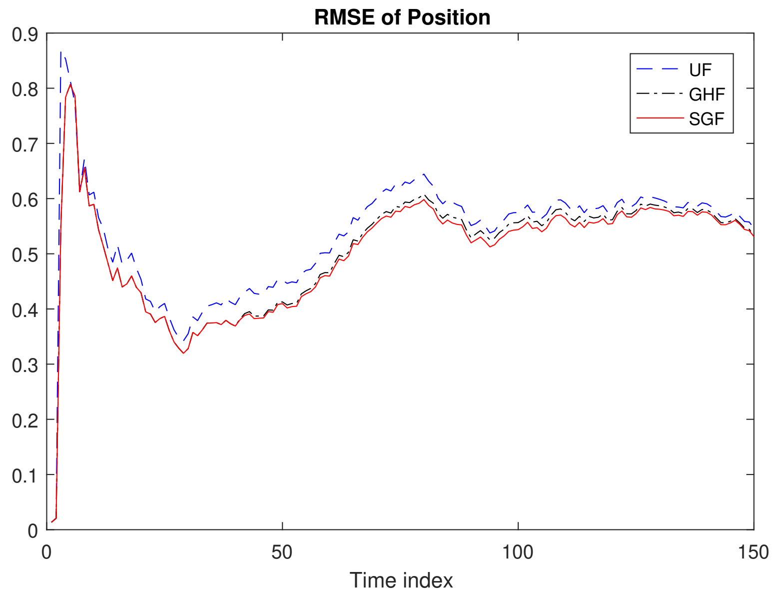

FIGURE 4.3: Position RMSEs (km)

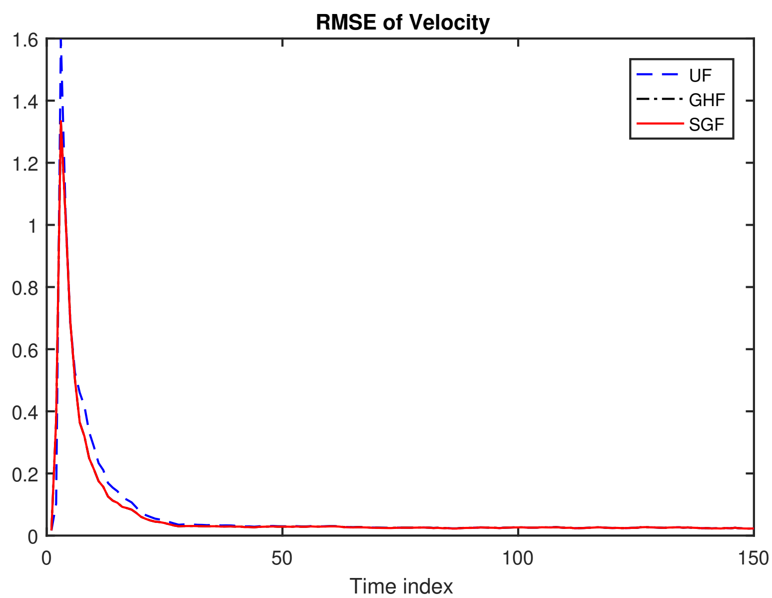

FIGURE 4.4: Velocity RMSEs (km / s)

Table 4.2: Computation Time (s) for One Run

Methods	UF	GHF	FOSGF
Time	0.0138	0.0562	0.7880
Relative Time	1	4.0725	57.1014

5

Optimized Gauss-Hermite Quadrature

5.1 Introduction

Many nonlinear problems involve multi-dimensional integrals that are not analytically tractable. For example, the Gaussian filter requires computing of the integrals (5.9)–(5.15). There are different approaches for numerical integration, e.g., the rectangular rule, the trapezoidal rule, and Simpson's rule (Davis and Rabinowitz, 2007). Among them, the quadrature-based rules, which approximate the integral by a weighted sum of the integrand values over grid points, are widely used, thus reducing the numerical integration to determining the grid points and their associated weights. In particular, the Gauss-Hermite quadrature (GHQ) rule is of special interest because of its close relationship with the Gaussian density.

The GHQ, in which the quadrature points and the weights are determined deterministically, can achieve accurate integral results in cases where the integrand is close to a polynomial. Unfortunately, if the integrand is far from a polynomial of a given degree, a large error may occur in numerical approximation (Kushner, 1967). Therefore, in order to avoid large error, the integrand should be converted to one that fits the theory of GHQ better.

In this chapter, an improvement of the quadrature-based rule is proposed. We convert the original integral to a new integral such that the new integrand is a closely-approximated polynomial, and then employ the regular quadrature rule. In this proposed approach, we give guidance to convert the integral in an optimal sense, which reduces to finding the optimal quadrature points and weights. Furthermore, we systematically formulate the optimal conversion as a nonlinear least-squares problem with linear constraints.

Another main topic of this chapter is the application to nonlinear filtering. Regarding nonlinear filtering, great efforts have been made, and several nonlinear filtering methods have been proposed under the framework of linear minimum mean squared error (LMMSE) estimation. The most popular methods include the Gauss-Hermite filter (GHF) (Ito and Xiong, 2000) using the GHQ rule, the unscented filter (UF) (Julier and Uhlmann, 2004) using the unscented transformation (UT), the cubature Kalman filter (CKF) (Arasaratnam and Haykin, 2009) using the cubature rule, the sparse-grid quadrature filter (SGQF) (Jia et al., 2012) using the sparse-grid quadrature (SGQ) rule, and the divided difference filter of the first and the second order (DD1, DD2) (Nørgaard et al., 2000) using Stirling's interpolation. A summary of point estimation methods was given in Section 1.4.2.

Our second contribution is the development of a nonlinear filtering algorithm by using the new quadrature-based rule. The proposed nonlinear filter potentially has higher estimation accuracy since this rule has higher integration accuracy. Illustrative examples are provided to verify the higher accuracy of our quadrature-based rule over conventional rules and to show the effectiveness of our proposed quadrature-based nonlinear filter.

This chapter is organized as follows. A brief introduction of the conventional multi-dimensional Gauss-Hermite quadrature rules is presented in Section 5.2. The new quadrature-based rule is proposed in Section 5.3, and its application to nonlinear filtering is given in Section 5.4. Numerical simulation results are shown in Section 5.5.

5.2 Existing Quadrature-Based Rules

In this section two typical multi-dimensional quadrature-based rules, i.e., the GHQ based on the tensor product method and the GHQ based on the sparse-grid quadrature method, are briefly reviewed. They both use the same univariate quadrature rule (Stoer and Bulirsch, 2002) to determine quadrature points. However, their main difference is in the extension to form a multi-dimensional grid.

For simplicity, the quadrature method is illustrated w.r.t. the standard Gaussian distribution, and a linear transformation is required for non-standard Gaussian densities.

The GHQ is an approximation for the univariate integral of the following kind

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(x)e^{-\frac{x^2}{2}}\,\mathrm{d}x \approx I_1(g) \triangleq \sum_{i=1}^{m} w_i g(x_i)$$

where the quadrature points $\{(x_i, w_i)\}$ are determined by a quadrature rule, and this approximation $I_1(g)$ becomes exact if $g(x)$ is a polynomial of degree up to $2m-1$.

5.2.1 Tensor Product Method

The tensor product method is a direct extension for a multi-dimensional grid from the univariate GHQ.

Let L_1 denote the accuracy level of the univariate quadrature rule, indicating that $I_1(g)$ is exact for all univariate polynomials $g(x)$ of the form x^k with $0 \le k \le 2L_1 - 1$. Ψ_{L_1} denotes the set of the univariate quadrature points and weights with accuracy level L_1.

The multi-dimensional GHQ rule with accuracy level L_1 at each dimension is

$$
\begin{aligned}
\int_{\mathbb{R}^n} \boldsymbol{g}(\boldsymbol{x})\mathcal{N}(\boldsymbol{x};\mathbf{0},\boldsymbol{I}_n)\,\mathrm{d}\boldsymbol{x} &\approx \sum_{(q_1,w_1)\in\Psi_{L_1}} \cdots \sum_{(q_n,w_n)\in\Psi_{L_1}} w_1\cdots w_n \boldsymbol{g}(q_1,\cdots,q_n) \\
&= I_n(\boldsymbol{g})
\end{aligned}
$$

The multivariate integral approximation becomes exact if the integrand $g(x)$ is a polynomial of the form $x_1^{i_1}x_2^{i_2}\cdots x_n^{i_n}$ with $1 \le i_k \le 2m-1$ for $k = 1, 2, \ldots, n$.

The tensor product rule is an easy and without-thinking-twice way to the extension to multi-dimensional problem, however, the computational costs rise exponentially as the dimension n increases and become prohibitive for high-dimensional problems. This is known as the curse of dimensionality of deterministic numerical integration.

5.2.2 Sparse-Grid Quadrature Method

The sparse-grid quadrature method, as explored in (Heiss and Winschel, 2008), extends the univariate quadrature rule to a multi-dimensional grid by the Smolyak rule (Gerstner and Griebel, 1998).

Denote the accuracy level of n-variate quadrature rule as L_n, meaning that $I_{L_n}(\boldsymbol{g})$ is exact w.r.t. $\mathcal{N}(\boldsymbol{x};\mathbf{0},\boldsymbol{I}_n)$ for all polynomials $\boldsymbol{g}(\boldsymbol{x})$ of the form $x_1^{i_1}x_2^{i_2}\cdots x_n^{i_n}$ with $\sum_{j=1}^n i_j \le 2L_n - 1$. The corresponding sparse-grid quadrature is

$$
\begin{aligned}
\int_{\mathbb{R}^n} \boldsymbol{g}(\boldsymbol{x})\mathcal{N}(\boldsymbol{x};\mathbf{0},\boldsymbol{I}_n)\,\mathrm{d}\boldsymbol{x} &\approx \sum_{q=L-n}^{L-1} \sum_{\Xi\in\mathbf{N}_q^n} \sum_{(q_1,w_1)\in\Psi_{i_1}} \cdots \sum_{(q_n,w_n)\in\Psi_{i_n}} \boldsymbol{g}(q_1,\cdots,q_n) \\
&\qquad \cdot \left[(-1)^{L-q-1}\binom{L-q-1}{n-1}\prod_{i=1}^n w_i\right] \\
&= I_{n,L_n}(\boldsymbol{g})
\end{aligned}
$$

where $\Xi = (i_1,\cdots,i_n)$ and Ψ_{i_j} is the set of the univariate quadrature points and weights with accuracy level i_j $(1 \le j \le n)$. The set of n-tuples of the polynomial degrees is defined by

$$\mathbf{N}_q^n = \begin{cases} \{\Xi : \sum_{j=1}^n i_j = n+q\} & q \geq 0 \\ \emptyset & q < 0 \end{cases}$$

where q is an auxiliary parameter with the range $[L-n, L-1]$.

A detailed algorithm to generate these quadrature points can be found in (Jia et al., 2012).

5.3 Proposed Quadrature-Based Rules

The Stone-Weierstrass theorem states that every continuous function $g(x)$ can be uniformly approximated by a polynomial, which means that the quadrature rule has the potential to achieve arbitrarily high accuracy as long as there are enough quadrature points. However, due to the exponential increase in the number of quadrature points in the design of a multi-dimensional grid and computational power, usually a univariate quadrature rule with a fixed number m of quadrature points is used. Under these conditions, if the integrand $g(x)$ is far from a polynomial of up to $2m-1$ degrees, a large error may occur in numerical integration (Kushner, 1967). A decent approach to circumvent this problem is, instead of the original integral, to consider the following arrangement

$$\int_{\mathbb{R}^n} \boldsymbol{g}(\boldsymbol{x})\mathcal{N}(\boldsymbol{x}; \mathbf{0}, \boldsymbol{I}_n)\ \mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^n} \frac{\boldsymbol{g}(\boldsymbol{x})\mathcal{N}(\boldsymbol{x}; \mathbf{0}, \boldsymbol{I}_n)}{\mathcal{N}(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{P})} \mathcal{N}(x; \bar{\boldsymbol{x}}, \boldsymbol{P})\ \mathrm{d}x \tag{5.1}$$

The idea is to find $(\bar{\boldsymbol{x}}, \boldsymbol{P})$ such that

$$\bar{\boldsymbol{g}}(\boldsymbol{x}) = \frac{\boldsymbol{g}(\boldsymbol{x})\mathcal{N}(\boldsymbol{x}; \mathbf{0}, \boldsymbol{I}_n)}{\mathcal{N}(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{P})}$$

is closest to some polynomial of up to $2m-1$ degrees. The new quadrature-based rule uses the regular quadrature rule to approximate $\int_{R^n} \bar{\boldsymbol{g}}(\boldsymbol{x})\mathcal{N}(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{P})\ \mathrm{d}\boldsymbol{x}$ instead of the original integral $\int_{R^n} \boldsymbol{g}(\boldsymbol{x})\mathcal{N}(\boldsymbol{x}; \mathbf{0}, \boldsymbol{I}_n)\ \mathrm{d}\boldsymbol{x}$.

5.3.1 *One-Dimensional Case*

Let $g(x) : \mathbb{R} \to \mathbb{R}$ and suppose the set of grid points $\{(x_i, y_i)\}_{i=1}^{N}$ with

$$y_i = y(x_i; \bar{x}, \sigma) = \frac{g(x_i)\mathcal{N}(x_i; 0, 1)}{\mathcal{N}(x_i; \bar{x}, \sigma^2)}$$

is given.

The problem of finding $(\overline{x}, \sigma^2)$ in (5.1) with m predetermined quadrature points can be formulated as a nonlinear least-squares (LS) problem with linear constraints

$$\min_{\bar{x},\sigma,\boldsymbol{B}} \sum_{i=1}^{N} [y(x_i; \bar{x}, \sigma) - \mathcal{P}_{2m-1}(x_i, \boldsymbol{B})]^2 \quad \text{s.t.} \quad \sigma > 0 \tag{5.2}$$

where $\mathcal{P}_{2m-1}(x) = \sum_{i=0}^{2m-1} \beta_i x^i$, and the coefficient vector $\boldsymbol{B} = [\beta_0, \beta_{1,} \cdots, \beta_{2m-1}]'$ is unknown.

Let $\boldsymbol{Y} = [y_1, y_2, \cdots, y_N]'$. Then the matrix form of (5.2) is

$$\min_{\hat{x},\sigma,\boldsymbol{B}} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}) \quad \text{s.t.} \quad \sigma > 0 \tag{5.3}$$

where $\boldsymbol{X}$ is an $N \times (2m)$ design matrix whose i-th row is $[1, x_i, x_i^2, \cdots, x_i^{2m-1}]$, $i = 1, 2, \ldots, N$.

Note that the formulation can be simplified if $N \gg 2m$. The solution of problem (5.2) for $\boldsymbol{B}$ as a function of $(\overline{x}, \sigma^2)$ is $\hat{\boldsymbol{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$. Substituting it into (5.3) gives

$$\min_{\hat{x},\sigma} \boldsymbol{Y}'\boldsymbol{P}^{\perp}\boldsymbol{Y} \quad \text{s.t.} \quad \sigma > 0 \tag{5.4}$$

where $\boldsymbol{P}^{\perp} = \boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is a projection matrix and $\boldsymbol{I}_N$ is the N-dimensional identity matrix.

The above is the so-called variable projection method (Golub and Pereyra, 2003), which is typically used for separable nonlinear LS problems, and a routine to solve such a nonlinear LS problem can be used directly (e.g., the trust-region-reflective LS algorithm).

The grid design will be discussed in Section 5.3.3.

5.3.2 Multi-Dimensional Case

The one-dimensional case can be extended to the multi-dimensional case naturally.

Let $\boldsymbol{g}(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ and suppose the set of grid points $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^N$ with

$$\boldsymbol{y}_i = \boldsymbol{y}(\boldsymbol{x}_i; \bar{\boldsymbol{x}}, \boldsymbol{\Sigma}) = \frac{\boldsymbol{g}(\boldsymbol{x}_i)\mathcal{N}(\boldsymbol{x}_i; \boldsymbol{0}, \boldsymbol{I}_n)}{\mathcal{N}(\boldsymbol{x}_i; \bar{\boldsymbol{x}}, \boldsymbol{\Sigma})}$$

is given.

The optimization problem with m predetermined quadrature points can be formulated as

$$\min_{\bar{\boldsymbol{x}}, \boldsymbol{\Sigma}, \boldsymbol{B}} (\boldsymbol{Y}_n - \boldsymbol{X}_n \boldsymbol{B}_n)'(\boldsymbol{Y}_n - \boldsymbol{X}_n \boldsymbol{B}_n) \quad \text{s.t.} \quad \boldsymbol{\Sigma} > \boldsymbol{0} \tag{5.5}$$

where $\boldsymbol{Y}_n = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N]'$, and the structure of the multivariate polynomial basis $\boldsymbol{X}_n$ and the coefficient vector $\boldsymbol{B}_n$ is based on the extension rule mentioned in Section 5.2.1 or Section 5.2.2.

Let the i-th grid points be $\boldsymbol{x}_i = [x_{i,1}, x_{i,2}, \cdots, x_{i,n}]'$.

(1) *Tensor product method* (Section 5.2.1). Each element of the i-th row of $\boldsymbol{X}_n$ is of the form $x_{i,1}^{j_1} x_{i,2}^{j_2} \cdots x_{i,n}^{j_n}, 0 \le j_1, j_2, \cdots, j_n \le 2m-1$, and the coefficient vector $\boldsymbol{B}_n$ is $(2m)^n$ -dimensional. For instance, if $n = 2$ and $m = 2$, the i-th row of $\boldsymbol{X}_n$ is

$$\begin{aligned}(\boldsymbol{X}_n)_i = [&1, x_{i,1}, x_{i,2}, x_{i,1}^2, x_{i,1}x_{i,2}, x_{i,2}^2, x_{i,1}^3, x_{i,1}^2 x_{i,2}, x_{i,1}x_{i,2}^2, x_{i,2}^3, x_{i,1}^3 x_{i,2}, x_{i,1}^2 x_{i,2}^2,\\ &x_{i,1}x_{i,2}^3, x_{i,1}^3 x_{i,2}^2, x_{i,1}^2 x_{i,2}^3, x_{i,1}^3 x_{i,2}^3]\end{aligned}$$

(2) *Sparse-grid quadrature method* (Section 5.2.2). Each element of the i-th row of $\boldsymbol{X}_n$ is of the form $x_{i,1}^{j_1} x_{i,2}^{j_2} \cdots x_{i,n}^{j_n}, 0 \le \sum_{k=1}^n j_k \le 2m-1$, and the coefficient

vector $\boldsymbol{B}_n$ is $\binom{n+2m-1}{n}$-dimensional. For example, if $n = 2$ and $m = 2$, the i-th row of $\boldsymbol{X}_n$ is

$$(\boldsymbol{X}_n)_i = [1, x_{i,1}, x_{i,2}, x_{i,1}x_{i,2}, x_{i,1}^2, x_{i,2}^2, x_{i,1}^2 x_{i,2}, x_{i,1}x_{i,2}^2, x_{i,1}^3, x_{i,2}^3] \tag{5.6}$$

Problem (5.5) can be reduced if $\boldsymbol{X}_n$ has full column rank

$$\min_{\hat{\boldsymbol{x}}, \boldsymbol{\Sigma}} \boldsymbol{Y}_n' \boldsymbol{P}_n^{\perp} \boldsymbol{Y}_n \quad \text{s.t.} \quad \boldsymbol{\Sigma} > \boldsymbol{0} \tag{5.7}$$

where $\boldsymbol{P}_n^{\perp} = \boldsymbol{I} - \boldsymbol{X}_n(\boldsymbol{X}_n'\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n'$ and $\boldsymbol{I}$ is the identity matrix of a compatible dimension. Then (5.7) can be transformed into a nonlinear LS problem with linear constraints as follows.

Let $\boldsymbol{\Sigma} = \sqrt{\boldsymbol{\Sigma}}'\sqrt{\boldsymbol{\Sigma}}$. Here the Cholesky factorization is preferred, and hence $\sqrt{\boldsymbol{\Sigma}}$ is an upper triangular matrix with $\frac{(1+n)n}{2}$ unknown parameters. We define a new vector $\boldsymbol{x}^{\text{new}} = [\bar{\boldsymbol{x}}', \text{vec}(\sqrt{\boldsymbol{\Sigma}})']'$, where the operator $\text{vec}(\cdot)$ converts matrix $\sqrt{\boldsymbol{\Sigma}}$ to a column vector by a column-by-column stacking with all elements below the main diagonal removed. Then (5.7) becomes, for any $k = 1, 2, \ldots, n$,

$$\min_{\boldsymbol{x}^{\text{new}}} [\boldsymbol{Y}_n(\boldsymbol{x}^{\text{new}})]' \boldsymbol{P}_n^{\perp} \boldsymbol{Y}_n(\boldsymbol{x}^{\text{new}}) \quad \text{s.t.} \quad \boldsymbol{1}'_{n+\frac{(1+k)k}{2}} \boldsymbol{x}^{\text{new}} > 0 \tag{5.8}$$

where $\sqrt{\boldsymbol{\Sigma}}$ is $n \times n$ dimensional and k represents its column number, and vector $\boldsymbol{1}_i$ is the $\frac{n^2+3n}{2}$-dimensional column vector with 1 at the i-th position and 0 otherwise. Note that (5.4) is a special case of (5.8) for $\boldsymbol{x}^{\text{new}} = [\bar{x}, \sigma]'$. Existing nonlinear LS methods, such as the trust-region-reflective LS algorithm, can be employed to solve (5.8) directly.

For $\boldsymbol{G}(\boldsymbol{x}) = [\boldsymbol{g}_1(\boldsymbol{x}), \boldsymbol{g}_2(\boldsymbol{x}), \cdots, \boldsymbol{g}_m(\boldsymbol{x})]'$, (5.8) is applied at each component $\boldsymbol{g}_i(\boldsymbol{x})$, $i = 1, 2, \ldots, m$.

5.3.3 Grid Design

Grid design is another crucial component for the new quadrature-based rule. In general, a good grid should have the following properties

(1) There should be sufficiently many yet not too many points;

(2) Sample in regions with a high probability.

As a result, an adaptive grid is preferred. In this chapter a simple grid design is used—propagate the grid with the mean and covariance of the approximated Gaussian density. Given the approximated density $\mathcal{N}(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{P})$, the implementation is to take the Cartesian product of points evenly scattered along the eigenvector direction of the covariance $\boldsymbol{P}$ and centered at the mean $\bar{\boldsymbol{x}}$. The location of the grid points at each principal axis is controlled by two parameters p and κ, where $p = \Pr\{(\boldsymbol{x} - \bar{\boldsymbol{x}})'\boldsymbol{P}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \leq d^2\}$ for some $d > 0$ and κ is the number of grid points allotted to each principal axis. With known d and κ, the locations of each grid point can be easily calculated.

The grid design problem is also extensively studied in point-mass nonlinear filtering, and a comprehensive summary of grid-based design and implementation is provided in (Šimandl et al., 2006).

5.4 Nonlinear Filtering Using Optimized Quadrature Rules

A representative of nonlinear filtering is the Gaussian type filter, e.g., Gaussian filter, whose main task is to approximate the estimated quantities in the form of the Gaussian integral. The optimized quadrature rules in Section 5.3 can be applied in such nonlinear filtering.

For simplicity, consider the nonlinear filtering problem for a discrete-time system (1.3) and (1.4) with additive noises,

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k, k) + \boldsymbol{w}_k \tag{1.3}$$

$$\boldsymbol{z}_k = \boldsymbol{h}(\boldsymbol{x}_k, k) + \boldsymbol{v}_k \tag{1.4}$$

where $\boldsymbol{w}_k \sim (\boldsymbol{0}, \boldsymbol{Q}_k)$ and $\boldsymbol{v}_k \sim (\boldsymbol{0}, \boldsymbol{R}_k)$ are white noises. We assume the initial state $\boldsymbol{x}_0 \sim \mathcal{N}(\hat{\boldsymbol{x}}_0, \boldsymbol{P}_0)$ and $\boldsymbol{x}_0$, $\boldsymbol{w}_k$, and $\boldsymbol{v}_k$ are mutually independent.

5.4.1 Gaussian Type Filtering

The Gaussian filter rests on two assumptions: i) the conditional state probability density $p(\boldsymbol{x}_k|\boldsymbol{z}^{k-1})$ at each step k is assumed to be Gaussian; ii) the conditional state and measurement probability density $p(\boldsymbol{x}_k, \boldsymbol{z}_k|\boldsymbol{z}^{k-1})$ at each step k is also jointly Gaussian.

Let $p(\boldsymbol{x}_k|\boldsymbol{z}^{k-1})$ and $p(\boldsymbol{x}_k|\boldsymbol{z}^{k})$ denote the state densities of the prediction and the update, respectively. A generic Gaussian filter based on the Bayesian framework is given as follows, (Ito and Xiong, 2000),

- Prediction.

$$
\begin{aligned}
p(\boldsymbol{x}_k|\boldsymbol{z}^{k-1}) = \int_{\mathbb{R}^n} & \frac{1}{[(2\pi)^n \det \boldsymbol{Q}_{k-1}]^{1/2}} \exp\big[-\frac{1}{2}(\boldsymbol{x}_k - \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1))'\boldsymbol{Q}_{k-1}^{-1} \\
& \times (\boldsymbol{x}_k - \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1))\big] p(\boldsymbol{x}_{k-1}|\boldsymbol{z}^{k-1}) \, \mathrm{d}\boldsymbol{x}_{k-1} \\
\approx & \, \mathcal{N}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})
\end{aligned}
$$

where, by Fubini's theorem, the mean and covariance are given by

$$
\begin{aligned}
\hat{\boldsymbol{x}}_{k|k-1} &= \int_{\mathbb{R}^n} \boldsymbol{x}_k p(\boldsymbol{x}_k|\boldsymbol{z}^{k-1}) \, \mathrm{d}\boldsymbol{x}_k \\
&= \int_{R^n} \bigg(\int_{R^n} \frac{\boldsymbol{x}_k}{[(2\pi)^n \det \boldsymbol{Q}_{k-1}]^{1/2}} \exp\big[-\frac{1}{2}(\boldsymbol{x}_k - \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1))'\boldsymbol{Q}_{k-1}^{-1} \\
&\quad \times (\boldsymbol{x}_k - \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1))\big] \, \mathrm{d}\boldsymbol{x}_k \bigg) p(\boldsymbol{x}_{k-1}|\boldsymbol{z}^{k-1}) \, \mathrm{d}\boldsymbol{x}_{k-1} \\
&= \int_{\mathbb{R}^n} \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1) p(\boldsymbol{x}_{k-1}|\boldsymbol{z}^{k-1}) \, \mathrm{d}\boldsymbol{x}_{k-1} \\
&\approx \int_{\mathbb{R}^n} \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1) \mathcal{N}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1}) \, \mathrm{d}\boldsymbol{x}_{k-1} \qquad (5.9)
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{P}_{k|k-1} &= \int_{\mathbb{R}^n} (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1})' p(\boldsymbol{x}_k|\boldsymbol{z}^{k-1})\ \mathrm{d}\boldsymbol{x}_k \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \frac{(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1})'}{[(2\pi)^n \det \boldsymbol{Q}_{k-1}]^{1/2}} \exp\big[-\frac{1}{2}(\boldsymbol{x}_k - \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1)' \boldsymbol{Q}_{k-1}^{-1} \right. \\
&\quad \left. \times (\boldsymbol{x}_k - \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1))\big]\ \mathrm{d}\boldsymbol{x}_k \right) p(\boldsymbol{x}_k|\boldsymbol{z}^{k-1})\ \mathrm{d}\boldsymbol{x}_{k-1} \\
&= \int_{\mathbb{R}^n} (\boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1) - \hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1) - \hat{\boldsymbol{x}}_{k|k-1})' \\
&\quad \times p(\boldsymbol{x}_k|\boldsymbol{z}^{k-1})\ \mathrm{d}\boldsymbol{x}_{k-1} + \boldsymbol{Q}_k \\
&\approx \int_{\mathbb{R}^n} (\boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1) - \hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1) - \hat{\boldsymbol{x}}_{k|k-1})' \\
&\quad \times \mathcal{N}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})\ \mathrm{d}\boldsymbol{x}_{k-1} + \boldsymbol{Q}_k
\end{aligned}
\tag{5.10}
$$

- Update

$$
\begin{aligned}
p(\boldsymbol{x}_k|\boldsymbol{z}^k) &= c \exp\big[-\frac{1}{2}(\boldsymbol{z}_k - \boldsymbol{h}(\boldsymbol{x}_k, k))' \boldsymbol{R}_k^{-1} (\boldsymbol{z}_k - \boldsymbol{h}(\boldsymbol{x}_k, k))\big] p(\boldsymbol{x}_k|\boldsymbol{z}^{k-1}) \\
&\approx \mathcal{N}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})
\end{aligned}
$$

with c being the normalization factor. Since

$$
\begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{z}_k \end{bmatrix} \Bigg| \boldsymbol{z}^{k-1} \sim \mathcal{N}\left(\begin{bmatrix} \hat{\boldsymbol{x}}_{k|k-1} \\ \hat{\boldsymbol{z}}_{k|k-1} \end{bmatrix}, \begin{bmatrix} \boldsymbol{P}_{k|k-1} & \boldsymbol{C}_{k|k-1} \\ \boldsymbol{C}'_{k|k-1} & \boldsymbol{S}_{k|k-1} \end{bmatrix} \right)
$$

The update formula of the mean and the covariance are

$$
\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{C}_{k|k-1} \boldsymbol{S}_{k|k-1}^{-1} (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}) \tag{5.11}
$$

$$
\boldsymbol{P}_{k|k} = \boldsymbol{P}_{k|k-1} - \boldsymbol{C}_{k|k-1} \boldsymbol{P}_{k|k-1}^{-1} \boldsymbol{C}'_{k|k-1} \tag{5.12}
$$

where, accordingly,

$$\hat{\boldsymbol{z}}_{k|k-1} = \int_{\mathbb{R}^n} \boldsymbol{h}(\boldsymbol{x}_k, k)\mathcal{N}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1}) \, \mathrm{d}\boldsymbol{x}_k \tag{5.13}$$

$$\boldsymbol{C}_{k|k-1} = \int_{\mathbb{R}^n} (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1})(\boldsymbol{h}(\boldsymbol{x}_k, k) - \hat{\boldsymbol{z}}_{k|k-1})'\mathcal{N}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1}) \, \mathrm{d}\boldsymbol{x}_k \tag{5.14}$$

$$\boldsymbol{S}_{k|k-1} = \int_{\mathbb{R}^n} (\boldsymbol{h}(\boldsymbol{x}_k, k) - \hat{\boldsymbol{z}}_{k|k-1})(\boldsymbol{h}(\boldsymbol{x}_k, k) - \hat{\boldsymbol{z}}_{k|k-1})'\mathcal{N}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1}) \, \mathrm{d}\boldsymbol{x}_k + \boldsymbol{R}_k \tag{5.15}$$

Note that the LMMSE estimation is applied at the update stage (5.11)–(5.12), and the moments involved can be approximated by employing numerical integration, e.g., the quadrature-based rules.

5.4.2 Proposed Optimized Quadrature-Based Filter

The traditional quadrature-based filters, e.g., the Gauss-Hermite filter (GHF) (see, e.g., (Ito and Xiong, 2000; Arasaratnam and Haykin, 2009)) and sparse-grid quadrature filter (SGQF) (see, e.g., (Jia et al., 2012)), employ the multi-dimensional quadrature rules, discussed in Section 5.2.1 and Section 5.2.2, to approximate the quantities in (5.11)–(5.12) and (5.13)–(5.15) at each time. The optimized quadrature-based filter uses the optimized quadrature rules, discussed in Section 5.3.1 and Section 5.3.2, for nonlinear filtering.

Denote the dynamic function and measurement function in (1.3) and (1.4) as

$$\boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_{n_x}(\boldsymbol{x})]'$$
$$\boldsymbol{h}(\boldsymbol{x}) = [h_1(\boldsymbol{x}), h_2(\boldsymbol{x}), \cdots, h_{n_z}(\boldsymbol{x})]'$$

where the time index k is omitted for simplicity.

Let $\hat{\boldsymbol{x}}_{k-1|k-1}$ and $\boldsymbol{P}_{k-1|k-1}$ denote the estimate of the state and the corresponding covariance at time $k-1$, respectively.

Let $[a^i]$ and $[a^{i,j}]$ denote a vector and a matrix whose component at i and (i,j) is a^i and $a^{i,j}$, respectively. One cycle of the new quadrature-based nonlinear filter is summarized as follows,

(1) Prediction.

Given the set of sample points $\{\boldsymbol{x}_p\}_{p=1}^N$ w.r.t. $\mathcal{N}(\boldsymbol{x}; \hat{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{P}_{k-1|k-1})$, $\boldsymbol{X}_n$ in (5.6) is known.

(a) Calculate the predicted mean in (5.9):

for $i = 1$ to n_x

Determine each element y_p of $\boldsymbol{Y}_n$ in (5.8) with

$$y_p = \frac{f_i(\boldsymbol{x}_p)\mathcal{N}(\boldsymbol{x}_p; \hat{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{P}_{k-1|k-1})}{\mathcal{N}(\boldsymbol{x}_p; \overline{\boldsymbol{x}}, \boldsymbol{P})}$$

Apply the optimization method in (5.8) with $\boldsymbol{Y}_n$ and $\boldsymbol{X}_n$ to find the optimal $(\bar{\boldsymbol{x}}, \boldsymbol{P})$;

Use the regular quadrature rule w.r.t. $\mathcal{N}(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{P})$ to find $\hat{x}^i_{k|k-1}$.

end

$$\hat{\boldsymbol{x}}_{k|k-1} = [\hat{x}^i_{k|k-1}]$$

(b) Calculate the predicted covariance in (5.12):

for $i = 1$ to n_x and $j = 1$ to n_x

Determine each element y_p of $\boldsymbol{Y}_n$ in (5.8)) with

$$y_p = \frac{f_i(\boldsymbol{x}_p)f_j(\boldsymbol{x}_p)\mathcal{N}(\boldsymbol{x}_p; \hat{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{P}_{k-1|k-1})}{\mathcal{N}(\boldsymbol{x}_p; \bar{\boldsymbol{x}}, \boldsymbol{P})}$$

Apply the optimization method in (5.8) with $\boldsymbol{Y}_n$ and $\boldsymbol{X}_n$ to find the optimal $(\bar{\boldsymbol{x}}, \boldsymbol{P})$;

Use the regular quadrature rule w.r.t. $\mathcal{N}(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{P})$ to find $P^{i,j}_{k|k-1}$.

end

$$\boldsymbol{P}_{k|k-1} = [P^{i,j}_{k|k-1}] - \hat{\boldsymbol{x}}_{k|k-1}\hat{\boldsymbol{x}}'_{k|k-1} + \boldsymbol{Q}_k$$

(2) Update.

Update the new set of sample points $\{\boldsymbol{x}_p\}_{p=1}^{N}$ w.r.t. $\mathcal{N}(\boldsymbol{x}; \hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})$, and then the new $\boldsymbol{X}_n$ in (5.8) is known.

(a) Calculate the predicted measurement in (5.13)

for $i = 1$ to n_z

Determine each element y_p of $\boldsymbol{Y}_n$ in (5.8) with

$$y_p = \frac{h_i(\boldsymbol{x}_p)\mathcal{N}(\boldsymbol{x}_p; \hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})}{\mathcal{N}(\boldsymbol{x}_p; \bar{\boldsymbol{x}}, \boldsymbol{P})}$$

Apply the optimization method in (5.8) with $\boldsymbol{Y}_n$ and $\boldsymbol{X}_n$ to find the optimal $(\bar{\boldsymbol{x}}, \boldsymbol{P})$;

Use the regular quadrature rule w.r.t. $\mathcal{N}(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{P})$ to find $\hat{\boldsymbol{z}}^{i}_{k|k-1}$.

end

$$\hat{\boldsymbol{z}}_{k|k-1} = [\hat{\boldsymbol{z}}^{i}_{k|k-1}]$$

(b) Calculate the covariance of the predicted measurement in (5.15):

for $i = 1$ to n_z and $j = 1$ to n_z

Determine each element y_p of $\boldsymbol{Y}_n$ in (5.8) with

$$y_p = \frac{h_i(\boldsymbol{x}_p)h_j(\boldsymbol{x}_p)\mathcal{N}(\boldsymbol{x}_p; \hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})}{\mathcal{N}(\boldsymbol{x}_p; \bar{\boldsymbol{x}}, \boldsymbol{P})}$$

Apply the optimization method in (5.8) with $\boldsymbol{Y}_n$ and $\boldsymbol{X}_n$ to find the optimal $\bar{\boldsymbol{x}}, \boldsymbol{P}$);

Use the regular quadrature rule w.r.t. $\mathcal{N}(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{P})$ to find $C_{\boldsymbol{z}}^{i,j}$.

end

$$\boldsymbol{C}_{\boldsymbol{z}} = [C_{\boldsymbol{z}}^{i,j}] - \hat{\boldsymbol{z}}_{k|k-1}\hat{\boldsymbol{z}}'_{k|k-1} + \boldsymbol{R}_k$$

(c) Calculate the covariance of the predicted measurement in (5.14):

for $i = 1$ to n_x and $j = 1$ to n_z

Determine each element y_p of $\boldsymbol{Y}_n$ in (5.8) with

$$y_p = \frac{x_p^i h_j(\boldsymbol{x}_p)\mathcal{N}(\boldsymbol{x}_p; \hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})}{\mathcal{N}(\boldsymbol{x}_p; \bar{\boldsymbol{x}}, \boldsymbol{P})}$$

where x_p^i is the i-th component of $\boldsymbol{x}_p$.

Apply the optimization method in (5.8) with $\boldsymbol{Y}_n$ and $\boldsymbol{X}_n$ to find the optimal $(\bar{\boldsymbol{x}}, \boldsymbol{P})$;

Use the regular quadrature rule w.r.t. $\mathcal{N}(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{P})$ to find $C_{\boldsymbol{xz}}^{i,j}$.

end

$$\boldsymbol{C}_{\boldsymbol{xz}} = [C_{\boldsymbol{xz}}^{i,j}] - \hat{\boldsymbol{x}}_{k|k-1}\hat{\boldsymbol{z}}'_{k|k-1}$$

(d) Update the mean and the covariance by (5.11) and (5.12).

The Gaussian filtering algorithm can be extended easily to the Gaussian-sum filtering framework, so we omit the details.

5.5 Simulation

In this section we demonstrate the feasibility of our proposed filter using four test examples. We compare in performance our filter with other well-known filters—EKF, UF, GHF, and SGQF.

5.5.1 Static One-Dimensional Case

We consider the following integral

$$\int_{-\infty}^{\infty} \cos^2(x)\mathcal{N}(x;\bar{x},\sigma^2)\ \mathrm{d}x = \frac{1}{2}\,\mathrm{Re}(e^{2i\bar{x}-2\sigma^2}) + \frac{1}{2}$$

with $\bar{x} = 0$, $\sigma = 1$, and 100 grid points evenly scattered over the 3σ interval.

Table 5.1: Univariate integral $L_1 = 3$

Methods	Result	Absolute Error	Error Percentage
True value	0.567668		
GHQ	0.675259	0.107592	18.95%
Proposed	0.607035	0.033936	5.98%

Table 5.2: Univariate integral $L_1 = 4$

Methods	Result	Absolute Error	Error Percentage
True value	0.567668		
GHQ	0.537401	0.030265	5.33%
Proposed	0.569969	0.002301	0.41%

The numerical results of using different approximations with accuracy levels $L_1 =$ 3 and 4 are shown in Table 5.1 and Table 5.2, respectively. Both quadrature methods have higher accuracy as L_1 increases. Our quadrature-based method reduces the error of GHF from 18.95% to 5.98% and from 5.33% to 0.41%.

5.5.2 Static Multi-Dimensional Case

Next, we consider a more complex integral with two random variables,

$$\iint_{\mathbb{R}^2} \cos^2(x_1)\cos^2(x_2)\mathcal{N}(\boldsymbol{x};\bar{\boldsymbol{x}},\Sigma)\ \mathrm{d}\boldsymbol{x} = \frac{1}{4}(ab+a+b+1)$$

where

$$a = \mathrm{Re}(e^{i[2,0]\bar{\boldsymbol{x}}-\frac{1}{2}[2,0]\Sigma[2,0]'})$$

$$b = \mathrm{Re}(e^{i[0,2]\bar{\boldsymbol{x}}-\frac{1}{2}[0,2]\Sigma[0,2]'})$$

Here we choose $\bar{\boldsymbol{x}} = [x_1, x_2]' = [1, -2]'$, $\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}$, $p = 0.9$ and $\kappa = 15$.

Table 5.3: Multivariate integral $L_2 = 3$

Methods	Result	Absolute Error	Error Percentage
True value	0.233095		
GHQ	0.170293	0.062802	26.94%
Proposed GHQ	0.207429	0.025666	11.01%
SGQ	-0.051951	0.285047	122.29%
Proposed SGQ	0.081892	0.151203	64.86%

Table 5.4: Multivariate integral $L_2 = 4$

Methods	Result	Absolute Error	Error Percentage
True value	0.233095		
GHQ	0.242499	0.009403	4.03%
Proposed GHQ	0.236071	0.002975	1.27%
SGQ	0.497539	0.264443	113.44%
Proposed SGQ	0.227325	0.005770	2.47%

Table 5.3 and Table 5.4 show the results with $L_2 = 3$ and 4, respectively. Our methods have better performance compared with the conventional quadrature methods. The relative error is reduced from 26.94% to 11.01% w.r.t. GHQ and from 122.29% to 64.86% w.r.t. SGQ for $L_2 = 3$. Note that the error of the SGQ is due to possible non-positiveness of the weights. Similarly our quadrature-based method lowers the error from 4.03% to 1.27% w.r.t. GHQ and from 113.44% to 2.47% w.r.t. SGQ for $L_2 = 4$.

5.5.3 Dynamic One-Dimensional Case

A target (see Figure 5.1) is assumed to be moving along the x-axis, and the system model is

$$x_{k+1} = x_k + 1 + w_k$$

The platform, starting from $[0, 20]$, is moving horizontally at $y_{p,k} = 20\,\mathrm{m}$ with velocity $\dot{x}_{p,k} = 4\,\mathrm{m}\,/\,\mathrm{s}$. The measurement is the distance between the target and the platform,

$$z_k = \sqrt{(x_k - x_{p,k})^2 + y_{p,k}^2} + v_k$$

Here the total time is 15 s, the initial target state is $x_0 \sim \mathcal{N}(30, 50)$, $w_k \sim \mathcal{N}(0, 0.01)$, $v_k \sim \mathcal{N}(0, 0.25)$, $p = 0.9$, and $\kappa = 100$.

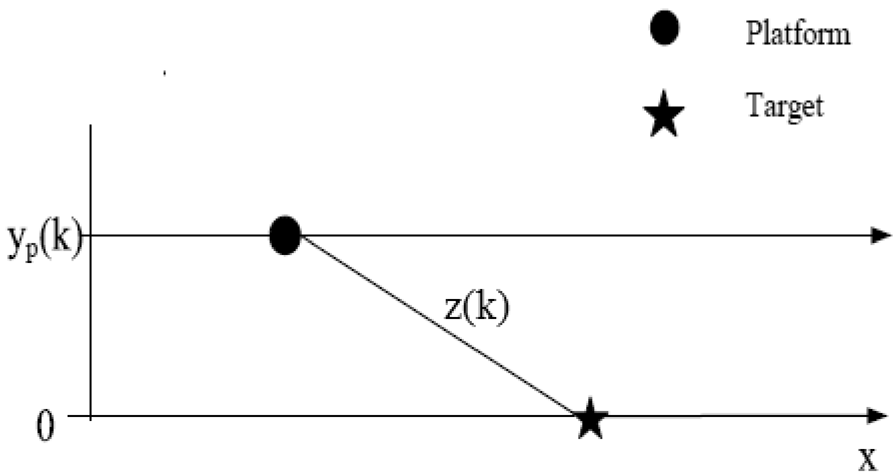

FIGURE 5.1: Platform and Target

Table 5.5: Relative Computation Time for One Run of Each Method (1D)

Methods	$L_2 = 2$	$L_2 = 3$
EKF	1	1
UF	1.47	1.47
GHQ	1.54	1.60
Proposed GHQ	688.92	774.27

The root-mean-square errors (RMSEs) from 100 Monte Carlo runs of the EKF, UF, GHF, and our quadrature-based nonlinear filter based on GHQ (NQF-GH) are shown in Figure 5.2 and Figure 5.3. It can be seen that in both cases our NQF-GH (line with squares) outperforms the other three filters in terms of estimation accuracy, and this enhancement is more significant at accuracy level $L_1 = 2$.

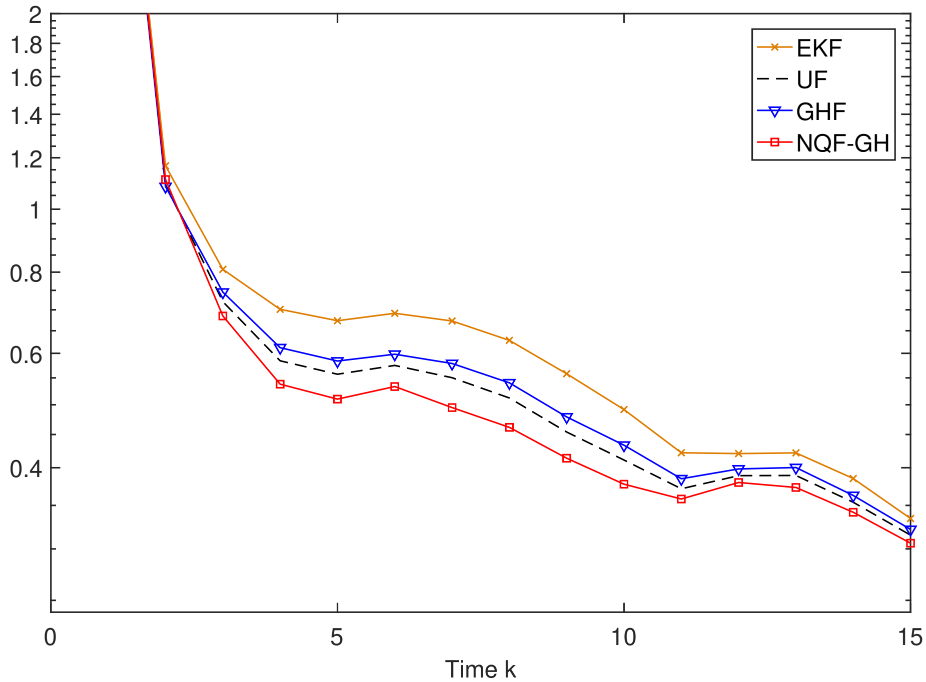

FIGURE 5.2: RMSEs (accuracy level $L_1 = 2$)

The computational loads of the four filtering methods are listed in Table 5.5 in terms of execution time relative to the EKF. The NQF-GH needs much more computation than the others since it involves optimization.

5.5.4 Dynamic Multi-Dimensional Case

The scenario we used to test the performance of the filtering is the same as the one in (Bellaire et al., 1995). It is a pendulum without external forces, and the equation of motion is $\ddot{\theta}(t) = -\frac{g}{L}\sin\theta(t)$ according to Newton's second law. Here θ is the angle subtended by a rod with length L, and g is the gravitational constant. The state is $\boldsymbol{x}_k = [\theta_k, \dot{\theta}_k]'$, and the model of motion and observation is

$$\boldsymbol{x}_{k+1} = \begin{bmatrix} x_{k,1} + \tau_s x_{k,2} - \frac{g\tau_s}{2L}\tau_s \sin x_{k,1} \\ x_{k,2} - \frac{g\tau_s}{2L}\sin x_{k,1} - \frac{g\tau_s}{2L}\sin(x_{k,1} + \tau_s x_{k,2}) \end{bmatrix} + \begin{bmatrix} \frac{\tau_s^2}{2} \\ \tau_s \end{bmatrix} w_k$$

$$z_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{x}_k + v_k$$

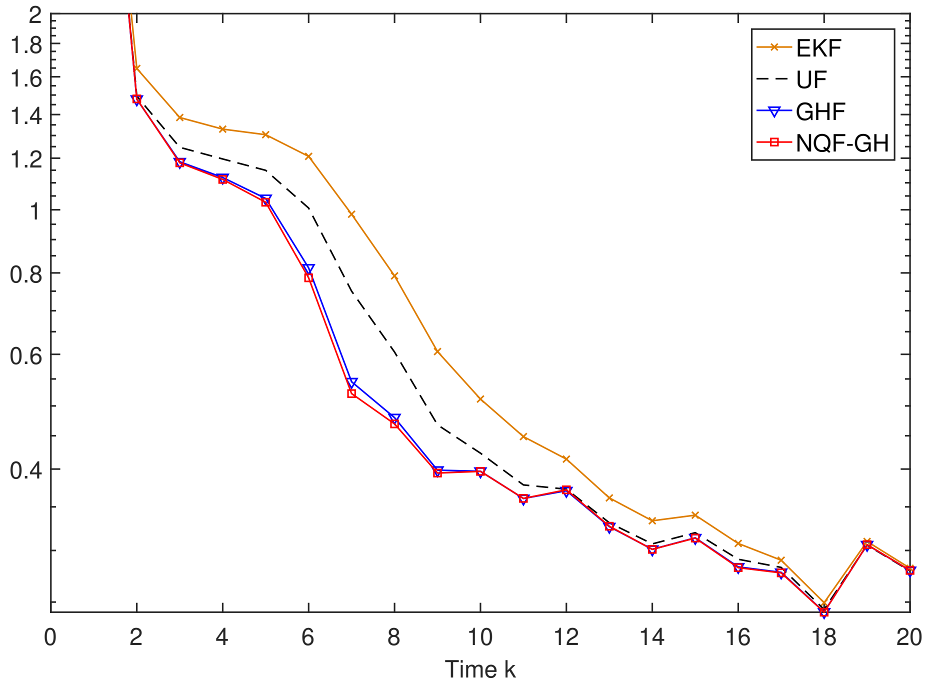

FIGURE 5.3: RMSEs (accuracy level $L_1 = 3$)

It is assumed that $\boldsymbol{x}_0 \sim \mathcal{N}\big([1.47, 0]', \begin{bmatrix} \frac{5}{3} & 0 \\ 0 & 0.01 \end{bmatrix}\big)$ and $\frac{g}{L} = 55$. Then the period of the pendulum is $T = 4\sqrt{\frac{L}{2G}} \int_0^{\theta_0} \frac{1}{\sqrt{\cos\theta - \cos\theta_0}}\,\mathrm{d}\theta \approx 0.9782$, and the sampling period is $\tau_s = \frac{1}{30}T$. $w_k \sim \mathcal{N}(0, 0.01)$ and $v_k \sim \mathcal{N}(0, 0.01)$, $p = 0.9$, and $\kappa = 15$. One swing is considered in this scenario.

In the scenario, six filters are compared: the EKF, UF, GHF, SGQF, our quadrature-based nonlinear filter based on GHQ (NQF-GH), and our quadrature-based nonlinear filter based on SGQ (NQF-SGQ).

This example permits the one-step predicted mean and covariance to be calculated analytically if no process noise exists.

$$\hat{\boldsymbol{x}}_{1|0}^{\text{true}} = \begin{bmatrix} 1.419 \\ -1.551 \end{bmatrix}, \qquad \boldsymbol{P}_{1|0}^{\text{true}} = \begin{bmatrix} 1.654 & -0.122 \\ -0.122 & 4.262 \end{bmatrix}$$

The mean and covariance predicted by EKF and UK are

$$\hat{\boldsymbol{x}}_{1|0}^{\mathrm{EKF}} = \begin{bmatrix} 1.354 \\ -3.569 \end{bmatrix}, \qquad \boldsymbol{P}_{1|0}^{\mathrm{EKF}} = \begin{bmatrix} 1.629 & 0.104 \\ 0.104 & 0.0169 \end{bmatrix}$$

$$\hat{\boldsymbol{x}}_{1|0}^{\mathrm{UF}} = \begin{bmatrix} 1.416 \\ -1.645 \end{bmatrix}, \qquad \boldsymbol{P}_{1|0}^{\mathrm{UF}} = \begin{bmatrix} 1.661 & 0.0312 \\ 0.031 & 4.276 \end{bmatrix}$$

Our quadrature-based nonlinear filters are listed below compared with GHF and SGQF for accuracy level $L_2 = 3$,

1) GHF vs. NQF-GH

$$\hat{\boldsymbol{x}}_{1|0}^{\mathrm{GHF}} = \begin{bmatrix} 1.416 \\ -1.645 \end{bmatrix}, \qquad \boldsymbol{P}_{1|0}^{\mathrm{GHF}} = \begin{bmatrix} 1.661 & 0.031 \\ 0.031 & 7.438 \end{bmatrix}$$

$$\hat{\boldsymbol{x}}_{1|0}^{\mathrm{NQF\text{-}GH}} = \begin{bmatrix} 1.419 \\ -1.553 \end{bmatrix}, \qquad \boldsymbol{P}_{1|0}^{\mathrm{NQF\text{-}GH}} = \begin{bmatrix} 1.651 & -0.179 \\ -0.179 & 4.399 \end{bmatrix}$$

2) SGQF vs. NQF-SGQ

$$\hat{\boldsymbol{x}}_{1|0}^{\mathrm{SGQF}} = \begin{bmatrix} 1.416 \\ -1.645 \end{bmatrix}, \qquad \boldsymbol{P}_{1|0}^{\mathrm{SGQF}} = \begin{bmatrix} 1.661 & 0.031 \\ 0.031 & 7.439 \end{bmatrix}$$

$$\hat{\boldsymbol{x}}_{1|0}^{\mathrm{NQF\text{-}SGQ}} = \begin{bmatrix} 1.419 \\ -1.554 \end{bmatrix}, \qquad \boldsymbol{P}_{1|0}^{\mathrm{NQF\text{-}SGQ}} = \begin{bmatrix} 1.651 & -0.169 \\ -0.169 & 4.405 \end{bmatrix}$$

Our quadrature-based nonlinear filter has higher estimation accuracy in mean and covariance at the one-step prediction, especially of the $(1,2)$-th element, $(2,1)$-th element, and $(2,2)$-th element of the covariance.

The results of 100 Monte Carlo runs are depicted in Figure 5.4 and Figure 5.5. For $L_2 = 3$, 5 sigma points were used in the UF, and 9 and 13 quadrature points were used in the GHF and SGQF, respectively. There are three groups in $L_2 = 3$. The NQF-GH (dashed line with circles) and the NQF-SGQ (solid line with squares), which both belong to the group at the lowest position, perform better than the second group, which includes the GHF, SGQF, and UF. They are all much better than the

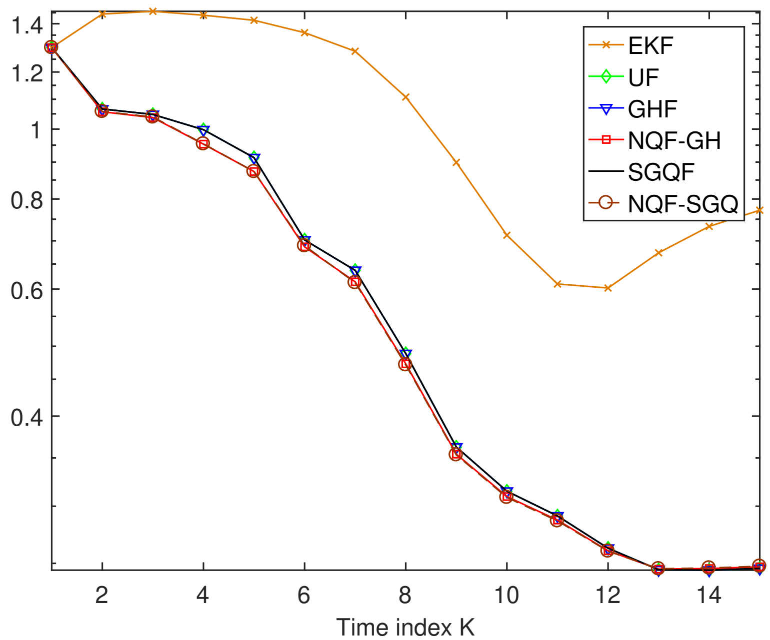

FIGURE 5.4: RMSEs (accuracy level $L_2 = 3$)

EKF. For $L_2 = 4$, 16 and 29 quadrature points were used in the GHF and SGQF, respectively. The GHF (solid line with triangles) and SGQF (solid line) outperform the UF (solid line of diamonds) at both the initial stage and the final stage, but they are still worse than our NQF-GH and NQF-SGQ. The improvements are more significant when the accuracy level L_2 is smaller.

Table 5.6 shows the computational costs of the six filtering methods in terms of execution time relative to the EKF. The EKF, UF, GHF, and SGQF require

Table 5.6: Relative Computation Time for One Run of Each Method (2D)

Methods	$L_2 = 3$	$L_2 = 4$
EKF	1	1
UF	1.02	1.04
GHQ	1.13	1.18
Proposed GHQ	233.03	239.62
SGQ	1.11	1.14
Proposed SGQ	231.32	239.03

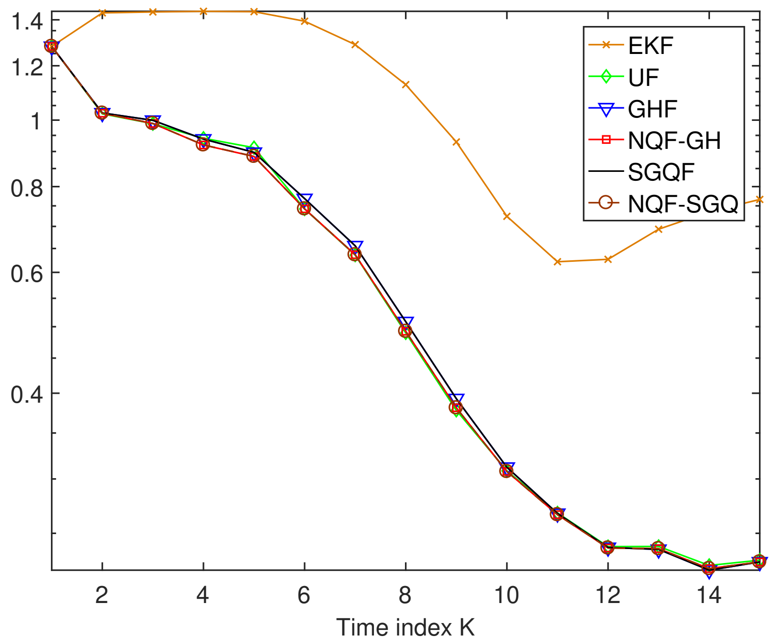

FIGURE 5.5: RMSEs (accuracy level $L_2 = 4$)

roughly the same execution time due to the low dimensionality of this scenario. Compared with the GHF and SGQF, the NQF-GH and NQF-SGQ need much more computational time since they both require optimization at each step of filtering.

6

Conclusions and Future Work

> The first rule of discovery is to have brains and good luck. The second rule of discovery is to sit tight and wait till you get a good idea.
>
> George Pólya

In this dissertation, we considered state estimation with a *symmetric* Gaussian assumption or an *asymmetric* skew-Gaussian assumption under linear/nonlinear systems, and developed three methods: the skew-Gaussian filter in linear systems, the first-order skew-Gaussian filter based on skew-symmetric representation of distributions, and the optimized Gaussian-Hermite quadrature rule. These three results, presented in order of importance from our perspectives, have demonstrated their power in theory and application, and their performance in state estimation has been analyzed. We highlight the main results obtained in Chapters 3–5.

For linear systems with the skew-Gaussian assumption, the linear skew-Gaussian system and its filtering result have been proposed in Chapter 3. The linear skew-Gaussian estimation provides a complete solution in an analytical form to the filtering problem where linear hidden truncation is involved in linear systems, and its filtering result (i.e., the SGF), derived from the Bayesian perspective, has an exact recursive finite-dimensional form in discrete time, which can be readily implemented by digital devices without any approximation. Besides the practical value mentioned above, the linear skew-Gaussian estimation also provides theoretical value for state estimation theory. First, it expands the theory of linear Gaussian estimation by considering the effect of skewness from the initial state and the two noises, and it degenerates to the well-known linear Gaussian estimation if no skewness exists. Second, it is one of few explicit exact density filters now available in discrete time, which is efficient and is comparable to the Kalman filter. Third, by propagating the symmetric part (i.e, the Gaussian part, exactly same as in the Kalman filter) and the skewness part (with the calculation of the cumulative skewness from the state and the two noises) simultaneously but separately, it provides a theoretical framework for nonlinear estimation where information of the third moment benefits filtering, and may inspire more research on nonlinear point estimation with the third moments involved, compared with the traditional LMMSE framework.

As for nonlinear systems where information of higher moments of the state, especially the skewness, is not negligible, the skew-symmetric representation of distributions has been adopted to approximately solve such estimation problems. The existence of such a representation for every pdf and its simple form (i.e., the product of a symmetric base pdf and a skewing function), make it possible to handle complex modeling problems well. Moreover, the flexible skew-symmetric form guarantees to approximate every pdf with arbitrary accuracy, which provides theoretical support for its application in nonlinear state estimation. Following this direction, the first-order skew-Gaussian representation is employed in nonlinear estimation problems. Even though the construction of the first-order skew-Gaussian distribution follows from the skew-symmetric representation of distributions, the first-order skew-

Gaussian distribution has a close relationship with the skew-Gaussian density studied in Chapter 3. More specifically, the first-order skew-Gaussian density is a special case of the skew-Gaussian density with one-dimensional linear hidden truncation of form $x^* \geq 0$ in (2.15). The corresponding FOSGF, based on the first-order skew-Gaussian distribution, has been proposed in Chapter 4. The FOSGF, in which the first three moments involved in nonlinear state estimation, can achieve better estimation accuracy if higher moments carry useful information for state estimation. Simulation results have demonstrated the effectiveness of the proposed FOSGF, and its computation consumption increased by the extra computation of the third moment compared with the LMMSE estimation. Further work may include exploring more effective and efficient ways to determine the skewness parameter and exploiting higher-order skew-Gaussian nonlinear filtering methods. Moreover, nonlinear filtering based on other symmetric base pdfs, such as student's t-distributions and elliptical distributions, is also noteworthy to research.

For nonlinear state estimation problems where the Gaussian filter is employed, the optimized Gaussian-Hermite quadrature rule can be applied to achieve results of Gaussian filters with better accuracy. It is worth noting that the optimized Gaussian-Hermite quadrature rule was proposed to improve the approximation accuracy of Gaussian integrals by using quadrature methods, so it has a wide applications besides the application to Gaussian filtering. By converting the original Gaussian type integral into a form that would result in higher approximation accuracy, the optimized GHQ involves an optimization problem of finding the optimal quadrature points and weights. We formulated this problem as a nonlinear least-squares problem with linearly inequality constraints, so any existing standard routine, such as the trust-region-reflective least-squares algorithm, can be directly used. With application to nonlinear filtering problems, four simulation experiments have been presented comparing our quadrature rule with other prevailing ones. The performance is better than the others, and the improvement is significant especially for cases of low degree. The main computation consumption results from the adaptive sampling and optimization with constraint at each step. Further work may includes exploring efficient grid design and reformulating the problem to nonlinear least-squares problems with applicable constraints such that efficient methods, e.g., the Levenberg-Marquadrt algorithm, can be used.

Even though, by employing linearization methods, the SGF in linear systems can be directly applied to nonlinear estimation problems with skewness being considered, it still has differences with the FOSGF with respect to application domain, computation load and performance. First, being a special form of the skew-Gaussian density (2.15), the first-order skew-Gaussian density (4.4) has fewer parameters to be estimated, so the FOSGF requires less computation than the SGF employed in linearized nonlinear systems. Second, the application of the SGF in nonlinear filtering problems requires the information of skewness from the initial state, the process noise and measurement noise, and if such information is not available or hard to obtain, the SGF may fail to work appropriately. Third, the SGF is suitable in situations where the skewness comes from the two noises, in other words, if the skewness of state estimation mainly comes from the nonlinear system transformation, e.g., systems with high nonlinearity in Gaussian noises, the FOSGF may be a better choice. Fourth, for well-studied nonlinear systems corrupted by skewed noises, the SGF adopted in linearized nonlinear systems, is much more capable of capture the skewness of the state density sought after. Beside the statements made above, further research is needed to give more detail on the performance of these two methods based on different nonlinear scenarios.

As for the optimized GHQ rule in Chapter 5, its structure is thoroughly different from that of the SGF and the FOSGF. Not limit to the Gaussian filtering, the optimized GHQ rule is designed to improve all Gaussian type integrals, so it has a promising application in other areas, such as computer science and statistics where the calculation of Gaussian integrals are needed. By converting the optimized GHQ problem into a nonlinear least squares optimization, many efficient existing algorithm can be adopted directly to obtain the optimized GHQ, which helps the optimized GHQ rule to be developed into a standard integration method in software. One connection between the optimized GHQ and the other two methods above is that it can be applied to moment matching to help reduce estimation error.

Appendix A

Proof of Proposition 2.1

Before presenting our proof, we refer to the following lemma, which plays an indispensable role in our proof of Proposition 2.1.

Lemma A.1 (He et al. (2018)). *Consider $u_x \in R^n$, $x \in R^n$, $y \in R^m$ and symmetric positive definite matrices $\Sigma_x \in R^{n\times n}$, $\Sigma_y \in R^{m\times m}$. Then*

$$\mathcal{N}(x; u_x, \Sigma_x)\mathcal{N}(y; Ax + b, \Sigma_y) = \mathcal{N}(x; u_x^*, \Sigma_x^*)\mathcal{N}(y; u_y^*, \Sigma_y^*)$$

where

$$\begin{aligned} u_x^* &= u_x + \Sigma_x A'(A\Sigma_x A' + \Sigma_y)^{-1}(y - Au_x - b) \\ \Sigma_x^* &= (\Sigma_x^{-1} + A'\Sigma_y^{-1}A)^{-1} \\ u_y^* &= Au_x + b, \quad \Sigma_y^* = A\Sigma_x A' + \Sigma_y \end{aligned}$$

For brevity, in the sequel we write $\lambda = \Phi([\gamma_1, \gamma_2]; \Gamma)$, and $(\cdot)$ (or $[\cdot]$) to represent the term with parentheses (or brackets) right before it.

Proof of Proposition 2.1.

$$\begin{aligned}
\bar{m}_1 &= \int_{\mathbb{R}^n} x\mathrm{SG}(x;[\gamma_1,\gamma_2],u,\Omega)\ \mathrm{d}x \\
&= \frac{1}{\lambda}\int_{\mathbb{R}^n}\int_{\gamma_1}^{\gamma_2} x\mathcal{N}(x;u,\Sigma)\mathcal{N}(y;\Delta'\Sigma^{-1}(x-u),\Gamma-\Delta'\Sigma^{-1}\Delta)\ \mathrm{d}y\ \mathrm{d}x \\
&= \frac{1}{\lambda}\int_{\gamma_1}^{\gamma_2}\int_{\mathbb{R}^n} x\mathcal{N}(x;u+\Delta\Gamma^{-1}y,\Sigma-\Delta\Gamma^{-1}\Delta')\cdot\mathcal{N}(y;\Gamma)\ \mathrm{d}y\ \mathrm{d}x \\
&= \frac{1}{\lambda}\int_{\gamma_1}^{\gamma_2}(u+\Delta\Gamma^{-1}y)\mathcal{N}(y;\Gamma)\ \mathrm{d}y
\end{aligned}$$

The third equality employs Lemma A.1 as well as the commutative law of multiple integral. By removing the parentheses we get (2.12) from the fourth equality.

$$\begin{aligned}
\bar{m}_2 &= \int_{\mathbb{R}^n}[x-\mathrm{E}(x)][\cdot]'\mathrm{SG}(x;[\gamma_1,\gamma_2],u,\Omega)\ \mathrm{d}x \\
&= \frac{1}{\lambda}\int_{\mathbb{R}^n}\int_{\gamma_1}^{\gamma_2}[x-\mathrm{E}(x)][\cdot]'\mathcal{N}(y;\Gamma)\mathcal{N}(x;u+\Delta\Gamma^{-1}y,\Sigma-\Delta\Gamma^{-1}\Delta')\ \mathrm{d}y\ \mathrm{d}x \\
&= \frac{1}{\lambda}\int_{\gamma_1}^{\gamma_2}\int_{\mathbb{R}^n}\big[(x-u-\Delta\Gamma^{-1}y)+\Delta\Gamma^{-1}(y-\mu_1)\big][\cdot]' \\
&\quad\cdot\mathcal{N}(x;u+\Delta\Gamma^{-1}y,\Sigma-\Delta\Gamma^{-1}\Delta')\mathcal{N}(y;\Gamma)\ \mathrm{d}x\ \mathrm{d}y \\
&= \Sigma-\Delta\Gamma^{-1}\Delta'+\frac{1}{\lambda}\int_{\gamma_1}^{\gamma_2}[\Delta\Gamma^{-1}(y-\mu_1)][\cdot]'\mathcal{N}(y;\Gamma)\ \mathrm{d}y
\end{aligned}$$

The last equality is obtained by integrating w.r.t. x first and observing that the integral of the cross-term is zero. The final result (2.13) follows from expanding the terms in the parenthesis.

Proof of the skewness (2.14) is more tedious.

$$\begin{aligned}
\bar{m}_3 &= \int_{\mathbb{R}^n} [x - \mathrm{E}(x)][\cdot]' \otimes (\cdot)' \mathrm{SG}(x; [\gamma_1, \gamma_2], u, \Omega) \, \mathrm{d}x \\
&= \frac{1}{\lambda} \int_{\mathbb{R}^n} \int_{\gamma_1}^{\gamma_2} [x - \mathrm{E}(x)][\cdot]' \otimes [\cdot]' \mathcal{N}(y; \Gamma) \mathcal{N}(x; u + \Delta\Gamma^{-1} y, \Sigma - \Delta\Gamma^{-1}\Delta') \, \mathrm{d}y \, \mathrm{d}x \\
&= \frac{1}{\lambda} \int_{\mathbb{R}^n} \int_{\gamma_1}^{\gamma_2} \{(x - u - \Delta\Gamma^{-1} y) + [u + \Delta\Gamma^{-1} y - \mathrm{E}(x)]\}\{\cdot\}' \otimes \{\cdot\}' \\
&\quad \cdot \mathcal{N}(x; u + \Delta\Gamma^{-1} y, \Sigma - \Delta\Gamma^{-1}\Delta') \mathcal{N}(y; \Gamma) \, \mathrm{d}y \, \mathrm{d}x
\end{aligned}$$

The above double integral can be simplified as

$$\begin{aligned}
&\frac{1}{\lambda} \int_{\gamma_1}^{\gamma_2} (\Delta\Gamma^{-1} y - \Delta\Gamma^{-1}\mu_1) \, \mathrm{vec}'[(\cdot)(\cdot)'] \mathcal{N}(y; \Gamma) \, \mathrm{d}y \\
&= \frac{1}{\lambda} \Delta\Gamma^{-1} \int_{\gamma_1}^{\gamma_2} (y - \mu_1) \, \mathrm{vec}'[(y - \mu_1)(y - \mu_1)'](\Delta\Gamma^{-1} \otimes \Delta\Gamma^{-1})' \mathcal{N}(y; \Gamma) \, \mathrm{d}y
\end{aligned}$$

The equality is from the following relation

$$\mathrm{vec}(ABC) = (C' \otimes A) \, \mathrm{vec}(B)$$

for any matrices A, B and C of compatible dimensions (Kollo, 2005, Proposition 1.3.14, p.89). Further, after removing the parentheses, the result (2.14) follows immediately.

An alternative proof can be observed by using Property 2.5 and applying the properties of moments of sum of independent random variables. We also verified (2.12)–(2.14) by computer simulations with many examples. □

Another way of proving Proposition 2.1 is given below, and it can be easily extended to derive the higher moments.

Proof. By Property 2.5, x can be decomposed into two independent variables, i.e.,

$$x = v_1 + \Delta\Gamma^{-1} v_0^{(\gamma_1, \gamma_2)}$$

where

$$v_0 \sim \mathcal{N}(0, \Gamma) \qquad v_1 \sim \mathcal{N}(u, \Sigma - \Delta\Gamma^{-1}\Delta')$$

Then

$$\mathrm{E}(x) = \mathrm{E}(v_1 + \Delta\Gamma^{-1}v_0^{(\gamma_1,\gamma_2)}) = u + \Delta\Gamma^{-1}\mathrm{E}(v_0^{(\gamma_1,\gamma_2)}) = u + \Delta\Gamma^{-1}\mu_1$$

where μ_1 is the mean of $v_0^{(\gamma_1,\gamma_2)}$ shown in (2.9).

$$\begin{aligned}
\mathrm{cov}(x) &= \mathrm{cov}(v_1 + \Delta\Gamma^{-1}v_0^{\gamma_1,\gamma_2}) \\
&= \mathrm{cov}(v_1) + \mathrm{cov}(\Delta\Gamma^{-1}v_0^{\gamma_1,\gamma_2}) \\
&= \Sigma - \Delta\Gamma^{-1}\Delta' + \Delta\Gamma^{-1}\,\mathrm{cov}(v_0^{(\gamma_1,\gamma_2)})\Gamma^{-1}\Delta' \\
&= \Sigma - \Delta\Gamma^{-1}(\Gamma + \mu_1\mu_1' - \mu_2)\Gamma^{-1}\Delta'
\end{aligned}$$

where μ_2 is the the second (non-central) moment of $v_0^{(\gamma_1,\gamma_2)}$ shown in (2.10).

$$\begin{aligned}
\mathrm{E}[(x - \mathrm{E}(x))[(x - \mathrm{E}(x))']^{\otimes 2}] &= \mathrm{E}[(\check{v}_1 + \Delta\Gamma^{-1}\check{v}_0^{(\gamma_1,\gamma_2)})[(\cdot)']^{\otimes 2}] \\
&= \mathrm{E}(\Delta\Gamma^{-1}\check{y}[(\Delta\Gamma^{-1}\check{y})']^{\otimes 2})
\end{aligned}$$

where $\check{(\cdot)}$ is the zero-mean part and $\check{y} = \check{v}_0^{(\gamma_1,\gamma_2)}$. Note that the last equation above is of the exact form of the last integral on the right column of page 10. Since

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD) \text{ for conforming matrices}$$

we have

$$\mathrm{E}(\Delta\Gamma^{-1}\check{y}[(\Delta\Gamma^{-1}\check{y})']^{\otimes 2}) = \Delta\Gamma^{-1}\mathrm{E}[\check{y}(\check{y}')^{\otimes 2}](\Delta\Gamma^{-1} \otimes \Delta\Gamma^{-1})'$$

Expanding the terms in expectation, and use

$$(A \otimes B)' = A' \otimes B' \qquad a' \otimes b = ba' = b \otimes a' \text{ if } a, b \text{ are vectors}$$

we have

$$
\begin{aligned}
\mathrm{E}[\check{y}(\check{y}')^{\otimes 2}] &= \mathrm{E}[(y-\mu_1)(y'\otimes y' - \mu_1'\otimes y' - y'\otimes \mu_1' + \mu_1'\otimes \mu_1')] \\
&= \mathrm{E}(yy'\otimes y' - y\mu_1'\otimes y' - yy'\otimes \mu_1' + y\mu_1'\otimes \mu_1' - \mu_1 y'\otimes y' \\
&\quad + \mu_1\mu_1'\otimes y' + \mu_1 y'\otimes \mu_1' - \mu_1\mu_1'\otimes \mu_1') \\
&= \mathrm{E}(y\mu_1'\otimes \mu_1' + \mu_1\mu_1'\otimes y' - \mu_1\mu_1'\otimes \mu_1') - \mathrm{E}(y\mu_1'\otimes y' + yy'\otimes \mu_1' + \mu_1 y'\otimes y') \\
&\quad + \mathrm{E}(yy'\otimes y') \\
&= 2\mu_1 \operatorname{vec}'(\mu_1\mu_1') - [\mu_1 \operatorname{vec}'(\mu_2) + \mu_2\otimes \mu_1' + \mu_1'\otimes \mu_2] + \mu_3
\end{aligned}
$$

where μ_3 is the the third (non-central) moment of $v_0^{(\gamma_1,\gamma_2)}$ shown in (2.11). $\square$

Appendix B

Proofs of Property 2.3

By Properties 2.1 and 2.4, the jointly skew-Gaussian density $p_{X_1,X_2}(x_1,x_2)$ and its marginal density $p_{X_2}(x_2)$ are of the forms

$$p_{\upsilon_1,\upsilon_2|\gamma_1\le\upsilon^*\le\gamma_2}(x_1,x_2), \qquad p_{\upsilon_2|\gamma_1\le\upsilon^*\le\gamma_2}(x_2)$$

where

$$\begin{bmatrix}\upsilon^*\\ \upsilon_1\\ \upsilon_2\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\ u_1\\ u_2\end{bmatrix}, \begin{bmatrix}\Gamma & \Delta_1' & \Delta_2'\\ \Delta_1 & \Sigma_1 & \Sigma_{21}'\\ \Delta_2 & \Sigma_{21} & \Sigma_2\end{bmatrix}\right) \tag{B.1}$$

on the condition $\gamma_1 \le \upsilon^* \le \gamma_2$. By Bayes' rule, the density of x_1 conditioned on x_2 is

$$p_{\upsilon_1|\upsilon_2,\gamma_1\le\upsilon^*\le\gamma_2}(x_1 \mid x_2) = \frac{p_{\upsilon_1,\upsilon_2|\gamma_1\le\upsilon^*\le\gamma_2}(x_1,x_2)}{p_{\upsilon_2|\gamma_1\le\upsilon^*\le\gamma_2}(x_2)}$$

From (B.1),

$$\left(\begin{bmatrix}\upsilon^*\\ \upsilon_1\end{bmatrix} \middle| \ \upsilon_2\right) \sim \mathcal{N}\left(\begin{bmatrix}\Delta_2'\Sigma_2^{-1}(\upsilon_2-u_2)\\ u_{1|2}\end{bmatrix}, \begin{bmatrix}\Gamma_{1|2} & \Delta_{1|2}'\\ \Delta_{1|2} & \Sigma_{1|2}\end{bmatrix}\right)$$

with $u_{1|2}$, $\Sigma_{1|2}$, $\Delta_{1|2}$ and $\Gamma_{1|2}$ given in Property 2.3. Apply Property 2.4 on the following transformed form

$$\left(\begin{bmatrix} \upsilon^{**} \\ \upsilon_1 \end{bmatrix} \middle| \; \upsilon_2 \right) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ u_{1|2} \end{bmatrix}, \begin{bmatrix} \Gamma_{1|2} & \Delta'_{1|2} \\ \Delta_{1|2} & \Sigma_{1|2} \end{bmatrix} \right) \tag{B.2}$$

with $\gamma_1 - \Delta_2'\Sigma_2^{-1}(\upsilon_2 - u_2) \leq \upsilon^{**} \leq \gamma_2 - \Delta_2'\Sigma_2^{-1}(\upsilon_2 - u_2)$. Property 2.3 is easily obtained by letting $\upsilon_1 = x_1$, $\upsilon_2 = x_2$ and applying Property 2.4 to (B.2).

Appendix C

Proofs of Theorems 3.1 and 3.2

In the sequel we present a proof of Theorem 3.2 first, and then a proof of Theorem 3.1.

Proof of Theorem 3.2. The proof is by induction. It is easy to check the base case that (3.13) is true.

For time $k \geq 1$, given the system (3.1),

$$x_k = \Phi_{k,0} x_0 + \sum_{i=1}^{k} \Phi_{k,i} G_{i-1} w_{i-1} \triangleq \Phi_{k,0} x_0 + w_{k,0}$$

$$\begin{aligned} z^k &= \begin{bmatrix} H_k \Phi_{k,0} \\ \vdots \\ H_1 \Phi_{1,0} \end{bmatrix} x_0 + \begin{bmatrix} H_k w_{k,0} + v_k \\ \vdots \\ H_1 w_{1,0} + v_1 \end{bmatrix} \\ &\triangleq \mathcal{O}^k x_0 + \zeta^k \end{aligned}$$

where $\Phi_{i,j}$ is the state transition matrix from x_j to x_i:

$$\Phi_{i,j} = \begin{cases} F_{i-1} \cdots F_j & i > j \\ I & i = j \end{cases}$$

and $\mathcal{O}^k$ is termed the observability map.

By Properties 2.2 and 2.6,

$$\begin{bmatrix} x_k \\ z^k \end{bmatrix} \sim \mathrm{SG}\left(\begin{bmatrix} \gamma_1^x, \gamma_2^x \\ \gamma_1^w, \gamma_2^w \\ \gamma_1^v, \gamma_2^v \end{bmatrix}, \begin{bmatrix} \Phi_{k,0} u_0 \\ \mathcal{O}^k u_0 \end{bmatrix}, \Omega^* \right)$$

where

$$\Omega^* = \left[\begin{array}{c|cc} \Gamma & \Delta_{x_k}' & \Delta_{z^k}' \\ \hline \Delta_{x_k} & \Sigma_{x_k} & \Sigma_{x_k z^k} \\ \Delta_{z^k} & (\Sigma_{x_k z^k})' & \Sigma_{z^k} \end{array} \right] \tag{C.1}$$

and

$$\Gamma = \mathrm{diag}(\Gamma^x, \Gamma^w, \Gamma^v)$$

$$\begin{bmatrix} \Delta_{x_k} \\ \Delta_{z^k} \end{bmatrix} = \begin{bmatrix} \Phi_{k,0}\Delta_0^x & \sum_{i=1}^k \Phi_{k,i} G_{i-1} \Delta_{i-1}^w & 0 \\ H_k \Phi_{k,0} \Delta_0^x & H_k \sum_{i=1}^k \Phi_{k,i} G_{i-1} \Delta_{i-1}^w & \Delta_k^v \\ \vdots & \vdots & \vdots \\ H_1 \Phi_{1,0} \Delta_0^x & H_1 \Phi_{1,1} G_0 \Delta_0^w & \Delta_1^v \end{bmatrix}$$

$$\triangleq \begin{bmatrix} \Phi_{k,0}\Delta_0^x & \sum_{i=1}^k \Phi_{k,i} G_{i-1} \Delta_{i-1}^w & 0 \\ \Delta_{z^k}^x & \Delta_{z^k}^w & \Delta_{z^k}^v \end{bmatrix}$$

From Property 2.4, Ω^* in (C.1) is the genuine covariance of

$$\left[(x_0^*)', (w^*)', (v^*)', (x_k)', (z^k)' \right]'$$

Note that the lower right block of (C.1) has the identical covariance structure with the covariance between the state x_k and z^k in the Kalman filter. Therefore, the same triangular factorization of covariance can be applied algebraically (see, e.g., (Kailath et al., 2000, p.323-325)). Thus,

$$\begin{bmatrix} \Sigma_{x_k} & \Sigma_{x_k z^k} \\ (\Sigma_{x_k z^k})' & \Sigma_{z^k} \end{bmatrix} = \begin{bmatrix} \Phi_{k,0} \Sigma_0 \Phi_{k,0}' + \Sigma_{w_{k,0}} & \Sigma_{x_k e^k} (U^k)' \\ U^k (\Sigma_{x_k e^k})' & U^k S^k (U^k)' \end{bmatrix} \tag{C.2}$$

where

$$
\begin{aligned}
\Sigma_{w_{k,0}} &= \sum_{i=1}^{k} \Phi_{k,i} G_{i-1} Q_{i-1} G'_{i-1} \Phi'_{k,i} \\
\Sigma_{x_k e^k} &= \left[\Phi_{k,k} K_k S_k, \cdots, \Phi_{k,1} K_1 S_1\right] \\
&= \left[\Phi_{k,k} K_k S_k, \Sigma_{x_k e^{k-1}}\right] \qquad \text{(C.3)} \\
S^k &= \mathrm{diag}(S_k, \cdots, S_1) \\
U^k &= \begin{bmatrix} I & H_k \Phi_{k,k} K^p_{k-1} & \cdots & H_k \Phi_{k,2} K^p_1 \\ & I & \cdots & H_{k-1} \Phi_{k-1,2} K^p_1 \\ & & \ddots & \vdots \\ & & & I \end{bmatrix}
\end{aligned}
$$

Also the inverse of U^k exists and is

$$
(U^k)^{-1} = \begin{bmatrix} I & -H_k \Phi^p_{k,k} K^p_{k-1} & \cdots & -H_k \Phi^p_{k,2} K^p_1 \\ & I & \cdots & -H_{k-1} \Phi^p_{k-1,2} K^p_1 \\ & & \ddots & \vdots \\ & & & I \end{bmatrix}
$$

where $\Phi^p_{m,n}$ is defined by (3.8). Further, $\Phi^p_{m,n}$ and $\Phi_{m,n}$ are related as follows: For $m > n$,

$$
\begin{aligned}
\Phi^p_{m,n} &= \Phi_{m,n} - \sum_{i=1}^{m-n} \Phi_{m,n+i} K^p_{n+i-1} H_{n+i-1} \Phi^p_{n+i-1,n} \\
&= \Phi_{m,n} - \sum_{i=1}^{m-n} \Phi^p_{m,n+i} K^p_{n+i-1} H_{n+i-1} \Phi_{n+i-1,n}
\end{aligned} \qquad \text{(C.4)}
$$

and

$$
\Phi^p_{n,n} = \Phi_{n,n} = I, \quad \forall n \in \mathbb{N}
$$

The factorization in (C.2) allows us to write the following global expression which relates $\hat{u}_{i|i-1}$ to z^i,

$$
(U^k)^{-1}(z^k - \mathcal{O}^k u_0) = [z_i - H_i \hat{u}_{i|i-1}]_{i=1}^k \qquad \text{(C.5)}
$$

where $[\cdot]_{i=1}^{k}$ stands for a k-vector of which the ith element has the form in the bracket. Thus, the Gaussian part follows from Property 2.3 that

$$
\begin{aligned}
\hat{u}_{k|k} &= \Phi_{k,0}u_0 + \Sigma_{x_k e^k}(U^k)'(U^k S^k (U^k)')^{-1}(z^k - \mathcal{O}^k u_0) \\
&= \Phi_{k,0}u_0 + \left[K_k, \quad F_{k-1}\Sigma_{x_{k-1}e^{k-1}}(S^{k-1})^{-1}\right](U^k)^{-1}(z^k - \mathcal{O}^k u_0) \\
&= \Phi_{k,0}u_0 + F_{k-1}\Sigma_{x_{k-1}e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}(z^{k-1} - \mathcal{O}^{k-1}u_0) + K_k(z_k - H_k\hat{u}_{k|k-1}) \\
&= \hat{u}_{k|k-1} + K_k(z_k - H_k\hat{u}_{k|k-1}) \\
\Sigma_{k|k} &= \Sigma_{x_k} - \Sigma_{x_k z^k}\Sigma_{z^k}\Sigma'_{x_k z^k} \\
&= F_{k-1}\Sigma_{x_{k-1}}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1} - \left[K_k S_k \quad F_{k-1}\Sigma_{x_{k-1}e^{k-1}}S^{k-1}\right](S^k)^{-1}[\cdot]' \\
&= F_{k-1}\Sigma_{x_{k-1}}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1} - K_k S_k K'_k - F_{k-1}\Sigma_{x_{k-1}e^{k-1}}(S^{k-1})^{-1}(\Sigma_{x_{k-1}e^{k-1}})'F'_{k-1} \\
&= F_{k-1}\Sigma_{k-1|k-1}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1} - K_k S_k K'_k \\
&= \Sigma_{k|k-1} - K_k S_k K'_k
\end{aligned}
$$

For the skewness part, first observe that

$$(U^k)^{-1}\Delta_{z^k} = (U^k)^{-1}\left[\Delta^x_{z^k}, \Delta^w_{z^k}, \Delta^v_{z^k}\right] \tag{C.6}$$

where the terms in (C.6) follow from (C.4) as

$$(U^k)^{-1}\Delta^x_{z^k} = [H_i\Phi^p_{i,0}\Delta^x_0]_{i=1}^{k} \tag{C.7}$$

$$(U^k)^{-1}\Delta^w_{z^k} = [H_i\sum_{j=1}^{i}\Phi^p_{i,j}G_{j-1}\Delta^w_{j-1}]_{i=1}^{k} \tag{C.8}$$

$$(U^k)^{-1}\Delta^v_{z^k} = [\Delta^v_i - H_i\sum_{j=1}^{i}\Phi^p_{i,j}K^p_{j-1}\Delta^v_{j-1}]_{i=1}^{k} \tag{C.9}$$

Here (C.7) and (C.8) are obtained by applying the second equality in (C.4), and (C.9) follows directly from matrix multiplication. Also, $K^p_0 = 0$ and $\Delta^v_0 = 0$ are used to deal with the initial terms.

Let

$$\Delta_{k|k-1} \triangleq \left[\Delta^x_{k|k-1}, \Delta^w_{k|k-1}, \Delta^v_{k|k-1}\right]$$

From (C.4), (C.7)–(C.9), and

$$\begin{aligned}\Sigma_{x_k e^{k-1}} = F_{k-1}\Sigma_{x_{k-1}e^{k-1}} &= \left[\Phi_{k,k-1}K_{k-1}S_{k-1}, \cdots, \Phi_{k,1}K_1S_1\right] \\ &= \left[\Phi_{k,k}K^p_{k-1}S_{k-1}, \cdots, \Phi_{k,2}K^p_1S_1\right]\end{aligned}$$

it admits that, by applying Property 2.3 to (C.2),

$$\begin{aligned}\Delta^x_{k|k-1} &= \Phi_{k,0}\Delta^x_0 - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta^x_{z^{k-1}} \\ &= \Phi_{k,0}\Delta^x_0 - \sum_{i=2}^{k}\Phi_{k,i}K^p_{i-1}H_{i-1}\Phi^p_{i-1,0}\Delta^x_0 \\ &= \Phi^p_{k,0}\Delta^x_0 \\ \Delta^w_{k|k-1} &= \sum_{i=1}^{k}\Phi_{k,i}G_{i-1}\Delta^w_{i-1} - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta^w_{z^{k-1}} \\ &= \sum_{i=1}^{k}\Phi_{k,i}G_{i-1}\Delta^w_{i-1} - \sum_{i=2}^{k}\Phi_{k,i}K^p_{i-1}H_{i-1}\sum_{j=1}^{i-1}\Phi^p_{i-1,j}G_{j-1}\Delta^w_{j-1} \\ &= \sum_{i=1}^{k}\Phi^p_{k,i}G_{i-1}\Delta^w_{i-1} \\ \Delta^v_{k|k-1} &= 0 - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta^v_{z^{k-1}} \\ &= -\sum_{i=2}^{k}\Phi_{k,i}K^p_{i-1}\Delta^v_{i-1} + \sum_{i=2}^{k}\Phi_{k,i}K^p_{i-1}H_{i-1}\sum_{j=1}^{i-1}\Phi^p_{i-1,j}K^p_{j-1}\Delta^v_{j-1} \\ &= -\sum_{i=1}^{k}\Phi^p_{k,i}K^p_{i-1}\Delta^v_{i-1}\end{aligned} \tag{C.10}$$

where the final results of $\Delta^x_{k|k-1}$, $\Delta^w_{k|k-1}$ and $\Delta^v_{k|k-1}$ are obtained by employing the first equality in (C.4) and use the fact that $K^p_0 = 0$, $\Delta^w_0 = 0$ and $\Delta^v_0 = 0$.

Let

$$\Delta_{k|k} \triangleq \left[\Delta^x_{k|k}, \Delta^w_{k|k}, \Delta^v_{k|k}\right] \tag{C.11}$$

where the terms in (C.11) follows from (C.3), (C.7)–(C.9) and (C.10) as

$$
\begin{aligned}
\Delta^x_{k|k} &= \Phi_{k,0}\Delta^x_0 - \Sigma_{x_k e^k}(S^k)^{-1}(U^k)^{-1}\Delta^x_{z^k} \\
&= \Delta^x_{k|k-1} - K_k H_k \Phi^p_{k,0}\Delta^x_0 \\
&= (I - K_k H_k)\Delta^x_{k|k-1} \\
\Delta^w_{k|k} &= \sum_{i=1}^{k} \Phi_{k,i} G_{i-1}\Delta^w_{i-1} - \Sigma_{x_k e^k}(S^k)^{-1}(U^k)^{-1}\Delta^w_{z^k} \\
&= \Delta^w_{k|k-1} - K_k H_k \sum_{i=1}^{k} \Phi^p_{k,i} G_{i-1}\Delta^w_{i-1} \\
&= (I - K_k H_k)\Delta^w_{k|k-1} \\
\Delta^v_{k|k} &= 0 - \Sigma_{x_k e^k}(S^k)^{-1}(U^k)^{-1}\Delta^v_{z^k} \\
&= \Delta^v_{k|k-1} - K_k(\Delta^v_k - H_k \sum_{i=1}^{k} \Phi^p_{k,i} K^p_{i-1}\Delta^v_{i-1}) \\
&= (I - K_k H_k)\Delta^v_{k|k-1} - K_k \Delta^v_k
\end{aligned}
$$

By (C.5) and (C.7)–(C.9), it turns out that, for $i = 1, 2$,

$$
\begin{aligned}
\hat{\gamma}_{i,k|k} &= \gamma_i - \Delta'_{z^k}(U^k S^k (U^k)')^{-1}(z^k - \mathcal{O}^k u_0) \\
&= \gamma_i - ((U^{k-1})^{-1}\Delta_{z^{k-1}})'(S^{k-1})^{-1}(U^{k-1})^{-1}(z^{k-1} - \mathcal{O}^{k-1}u_0) \\
&\quad - \begin{bmatrix} (H_k \Delta^x_{k|k-1})' \\ (H_k \Delta^w_{k|k-1})' \\ (\Delta^v_k + H_k \Delta^v_{k|k-1})' \end{bmatrix} S_k^{-1}(z_k - H_k \hat{u}_{k|k-1}) \\
&= \hat{\gamma}_{i,k|k-1} - K^s_k(z_k - H_k \hat{u}_{k|k-1})
\end{aligned}
$$

where $\gamma_i = \left[(\gamma_i^x)', (\gamma_i^w)', (\gamma_i^v)'\right]'$ and

$$\begin{aligned}
\Gamma_{k|k} &= \Gamma - \Delta'_{z^k}(U^k S^k (U^k)')^{-1}\Delta_{z^k} \\
&= \Gamma - ((U^k)^{-1}\Delta_{z^k})'(S^k)^{-1}(U^k)^{-1}\Delta_{z^k} \\
&= \Gamma - ((U^{k-1})^{-1}\Delta_{z^{k-1}})'(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta_{z^{k-1}} \\
&\quad - \begin{bmatrix} (H_k\Delta^x_{k|k-1})' \\ (H_k\Delta^w_{k|k-1})' \\ (\Delta^v_k + H_k\Delta^v_{k|k-1})' \end{bmatrix} S_k^{-1} \begin{bmatrix} (H_k\Delta^x_{k|k-1})' \\ (H_k\Delta^w_{k|k-1})' \\ (\Delta^v_k + H_k\Delta^v_{k|k-1})' \end{bmatrix}' \\
&= \Gamma_{k|k-1} - K_k^s S_k (K_k^s)'
\end{aligned}$$

□

Proof of Theorem 3.1. Following the same procedure as shown in the proof of Theorem 3.2, by $\Sigma_{x_k e^{k-1}} = F_{k-1}\Sigma_{x_{k-1}e^{k-1}}$, we have for the Gaussian part,

$$\begin{aligned}
\hat{u}_{k|k-1} &= \Phi_{k,0}u_0 + \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}(z^{k-1} - \mathcal{O}^{k-1}u_0) \\
&= F_{k-1}\hat{u}_{k-1|k-1} \\
\Sigma_{k|k-1} &= \Sigma_{x_k} - \Sigma_{x_k z^{k-1}}\Sigma_{z^{k-1}}\Sigma'_{x_k z^{k-1}} \\
&= F_{k-1}\Sigma_{x_{k-1}}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1} - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}S^{k-1}(\Sigma_{x_k e^{k-1}}(S^{k-1})^{-1})' \\
&= F_{k-1}\Sigma_{k-1|k-1}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1}
\end{aligned}$$

and for the skewness part

$$
\begin{aligned}
\Delta^x_{k|k-1} &= \Phi_{k,0}\Delta^x_0 - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta^x_{z^{k-1}} \\
&= F_{k-1}\Delta^x_{k-1|k-1} \\
\Delta^w_{k|k-1} &= \sum_{i=1}^{k}\Phi_{k,i}G_{i-1}\Delta^w_{i-1} - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta^w_{z^k} \\
&= F_{k-1}\sum_{i=1}^{k-1}\Phi_{k-1,i}G_{i-1}\Delta^w_{i-1} + G_{k-1}\Delta^w_{k-1} - F_{k-1}\Sigma_{x_{k-1}e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta^w_{z^k} \\
&= F_{k-1}\Delta^w_{k-1|k-1} + G_{k-1}\Delta^w_{k-1} \\
\Delta^v_{k|k-1} &= 0 - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta^v_{z^{k-1}} \\
&= 0 - F_{k-1}\Sigma_{x_{k-1}e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta^v_{z^{k-1}} \\
&= F_{k-1}\Delta^v_{k-1|k-1} \\
\Gamma_{k|k-1} &= \Gamma - ((U^{k-1})^{-1}\Delta_{z^{k-1}})'(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta_{z^{k-1}} \\
&= \Gamma_{k-1|k-1}
\end{aligned}
$$

and, for $i = 1, 2$,

$$
\begin{aligned}
\hat{\gamma}_{i,k|k-1} &= \gamma_i - ((U^{k-1})^{-1}\Delta_{z^{k-1}})'(S^{k-1})^{-1}(U^{k-1})^{-1}(z^{k-1} - \mathcal{O}^{k-1}u_0) \\
&= \hat{\gamma}_{i,k-1|k-1}
\end{aligned}
$$

$\square$

Appendix D

More Details of Theorems 3.1 and 3.2

Proof of Theorems 3.1 and 3.2. The proof is by induction. It is easy to check that the base case (3.13) is true.

For time $k \geq 1$, given the system (3.1),

$$\begin{aligned} x_k &= \Phi_{k,0} x_0 + \sum_{i=1}^{k} \Phi_{k,i} G_{i-1} w_{i-1} \triangleq \Phi_{k,0} x_0 + w_{k,0} \\ z^k &= \begin{bmatrix} H_k \Phi_{k,0} \\ \vdots \\ H_1 \Phi_{1,0} \end{bmatrix} x_0 + \begin{bmatrix} H_k w_{k,0} + v_k \\ \vdots \\ H_1 G_0 w_0 + v_1 \end{bmatrix} \\ &\triangleq \mathcal{O}^k x_0 + \zeta^k \end{aligned}$$

where $\Phi_{i,j}$ is the state transition matrix from x_j to x_i:

$$\Phi_{i,j} = \begin{cases} F_{i-1} \cdots F_j & i > j \\ I & i = j \end{cases}$$

and $\mathcal{O}^k$ is termed the observability map. Then,

$$\begin{bmatrix} x_k \\ z^k \end{bmatrix} = \mathbf{F}_k \begin{bmatrix} x_0 \\ w^{k-1} \\ v^k \end{bmatrix} \tag{D.1}$$

where

$$w^{k-1} = \left[(w_{k-1})', \cdots, (w_0)'\right]'$$

$$v^k = \left[(v_k)', \cdots, (v_1)'\right]'$$

and

$$\mathbf{F}_k = \left[\begin{array}{c|ccccc|cccc} \Phi_{k,0} & \Phi_{k,k}G_{k-1} & \Phi_{k,k-1}G_{k-2} & \cdots & \Phi_{k,1}G_0 & 0 & 0 & \cdots & 0 \\ H_k\Phi_{k,0} & H_k\Phi_{k,k}G_{k-1} & H_k\Phi_{k,k-1}G_{k-2} & \cdots & H_k\Phi_{k,1}G_0 & I & 0 & \cdots & 0 \\ H_{k-1}\Phi_{k-1,0} & 0 & H_{k-1}\Phi_{k-1,k-1}G_{k-2} & \cdots & H_{k-1}\Phi_{k-1,1}G_0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ H_1\Phi_{1,0} & 0 & 0 & \cdots & H_1\Phi_{1,1}G_0 & 0 & 0 & \cdots & I \end{array}\right]$$

From Property 2.6,

$$\begin{bmatrix} x_0 \\ w^{k-1} \\ v^k \end{bmatrix} \sim \mathrm{SG}(\begin{bmatrix} \gamma_1^x & & \gamma_2^x \\ \gamma_1^w & , & \gamma_2^w \\ \gamma_1^v & & \gamma_2^v \end{bmatrix}, \begin{bmatrix} u_0 \\ 0 \\ 0 \end{bmatrix}, \Omega)$$

where

$$\Omega = \begin{bmatrix} \Gamma^x & & & (\Delta_0^x)' & & \\ & \Gamma^w & & & (\Delta^{w,k-1})' & \\ & & \Gamma^v & & & (\Delta^{v,k})' \\ \Delta_0^x & & & \Sigma_0^x & & \\ & \Delta^{w,k-1} & & & Q^{k-1} & \\ & & \Delta^{v,k} & & & R^k \end{bmatrix}$$

and

$$\Delta^{w,k-1} = \left[(\Delta_{k-1}^w)', (\Delta_{k-2}^w)', \cdots, (\Delta_0^w)'\right]'$$

$$\Delta^{v,k} = \left[(\Delta_k^v)', (\Delta_{k-1}^v)', \cdots, (\Delta_1^v)'\right]'$$

$$Q^{k-1} = \mathrm{diag}([Q_{k-1}, \cdots, Q_0])$$

$$R^k = \mathrm{diag}([R_k, \cdots, R_1])$$

Since (D.1) is linear, the following is from Property 2.2

$$\begin{bmatrix} x_k \\ z^k \end{bmatrix} \sim \mathrm{SG}(\begin{bmatrix} \gamma_1^x & & \gamma_2^x \\ \gamma_1^w & , & \gamma_2^w \\ \gamma_1^v & & \gamma_2^v \end{bmatrix}, \begin{bmatrix} \Phi_{k,0}u_0 \\ \mathcal{O}^k u_0 \end{bmatrix}, \Omega^*) \tag{D.2}$$

where

$$\Omega^* = \left[\begin{array}{c|cc} \Gamma & \Delta'_{x_k} & \Delta'_{z^k} \\ \hline \Delta_{x_k} & \Sigma_{x_k} & \Sigma_{x_k z^k} \\ \Delta_{z^k} & (\Sigma_{x_k z^k})' & \Sigma_{z^k} \end{array}\right]$$

and

$$\begin{aligned}
\Gamma &= \operatorname{diag}(\Gamma^x, \Gamma^w, \Gamma^v) \\
\begin{bmatrix} \Delta_{x_k} \\ \Delta_{z^k} \end{bmatrix} &= \begin{bmatrix} \Phi_{k,0}\Delta_0^x & \sum_{i=1}^k \Phi_{k,i} G_{i-1} \Delta_{i-1}^w & 0 \\ H_k \Phi_{k,0}\Delta_0^x & H_k \sum_{i=1}^k \Phi_{k,i} G_{i-1} \Delta_{i-1}^w & \Delta_k^v \\ \vdots & \vdots & \vdots \\ H_1 \Phi_{1,0}\Delta_0^x & H_1 \Phi_{1,1} G_0 \Delta_0^w & \Delta_1^v \end{bmatrix} \\
&\triangleq \begin{bmatrix} \Phi_{k,0}\Delta_0^x & \sum_{i=1}^k \Phi_{k,i} G_{i-1} \Delta_{i-1}^w & 0 \\ \Delta_{z^k}^x & \Delta_{z^k}^w & \Delta_{z^k}^v \end{bmatrix}
\end{aligned}$$

Note that the lower right block of Ω^* has the identical covariance structure with the covariance between the state x_k and z^k in the Kalman filter. Therefore, the same triangular factorization of covariance can be applied algebraically (see, e.g., (Kailath et al., 2000, p.323-325)). Thus,

$$\begin{bmatrix} \Sigma_{x_k} & \Sigma_{x_k z^k} \\ (\Sigma_{x_k z^k})' & \Sigma_{z^k} \end{bmatrix} = \begin{bmatrix} \Phi_{k,0}\Sigma_0\Phi'_{k,0} + \Sigma_{w_{k,0}} & \Sigma_{x_k e^k}(U^k)' \\ U^k(\Sigma_{x_k e^k})' & U^k S^k (U^k)' \end{bmatrix}$$

where

$$\begin{aligned}
\Sigma_{w_{k,0}} &= \sum_{i=1}^k \Phi_{k,i} G_{i-1} Q_{i-1} G'_{i-1} \Phi'_{k,i} \\
\Sigma_{x_k e^k} &= \left[\Phi_{k,k} K_k S_k, \cdots, \Phi_{k,1} K_1 S_1\right] \\
&= \left[\Phi_{k,k} K_k S_k, \Sigma_{x_k e^{k-1}}\right] \\
S^k &= \operatorname{diag}(S_k, \cdots, S_1) \\
U^k &= \begin{bmatrix} I & H_k \Phi_{k,k} K_{k-1}^p & H_k \Phi_{k,k-1} K_{k-2}^p & \cdots & H_k \Phi_{k,2} K_1^p \\ & I & H_{k-1}\Phi_{k-1,k-1} K_{k-2}^p & \cdots & H_{k-1}\Phi_{k-1,2} K_1^p \\ & & & \ddots & \vdots \\ & & & & I \end{bmatrix}
\end{aligned}$$

The inverse of U^k exists and is

$$(U^k)^{-1} = \begin{bmatrix} I & -H_k\Phi^p_{k,k}K^p_{k-1} & -H_k\Phi^p_{k,k-1}K^p_{k-2} & \cdots & -H_k\Phi^p_{k,2}K^p_1 \\ & I & -H_{k-1}\Phi^p_{k-1,k-1}K^p_{k-2} & \cdots & -H_{k-1}\Phi^p_{k-1,2}K^p_1 \\ & & & \ddots & \vdots \\ & & & & I \end{bmatrix}$$

Later we will use $(U^k)^{-1}$ and the following relation to derive our main results.

Lemma D.1. *For $m > n$, we have*

$$\begin{aligned} \Phi^p_{m,n} &= \Phi_{m,n} - \sum_{i=1}^{m-n} \Phi_{m,n+i}K^p_{n+i-1}H_{n+i-1}\Phi^p_{n+i-1,n} \\ &= \Phi_{m,n} - \sum_{i=1}^{m-n} \Phi^p_{m,n+i}K^p_{n+i-1}H_{n+i-1}\Phi_{n+i-1,n} \end{aligned}$$

and

$$\Phi^p_{n,n} = \Phi_{n,n} = I, \quad \forall n \in \mathbb{N}$$

Proof of Lemma D.1. For the first equality, by expanding the summation in increasing order w.r.t. i,

$$\begin{aligned} \text{RHS} &= \Phi_{m,n} - \Phi_{m,n+1}K^p_nH_n\Phi^p_{n,n} - \cdots \\ &= \Phi_{m,n+1}(F_n - K^p_nH_n) - \cdots \\ &= \Phi_{m,n+1}\Phi^p_{n+1,n} - \Phi_{m,n+2}K^p_{n+1}H_{n+1}\Phi^p_{n+1,n} - \cdots \\ &= \Phi_{m,n+2}(F_{n+1} - K^p_{n+1}H_{n+1})\Phi^p_{n+1,n} - \cdots \\ &= \Phi_{m,n+2}\Phi^p_{n+2,n} - \Phi_{m,n+3}K^p_{n+2}H_{n+2}\Phi^p_{n+2,n} - \cdots \\ &\ \ \vdots \\ &= \Phi_{m,m-1}\Phi^p_{m-1,n} - \Phi_{m,m}K^p_{m-1}H_{m-1}\Phi^p_{m-1,n} \\ &= \Phi_{m,m}\Phi^p_{m,n} \\ &= \text{LHS} \end{aligned}$$

The second equality follows from $F_k^p = F_k - K_k^p H_k$ and $\Phi_{k,k}^p = I$ for $\forall k \in \mathbf{N}$, the third equality is from $\Phi_{n+1,n}^p = F_n^p$. By continuing in the same manner, we get the result.

For the second equality, analogous procedure applied by expanding the summation in decreasing order w.r.t. i.

$$\begin{aligned}
\text{RHS} &= \Phi_{m,n} - \Phi_{m,m}^p K_{m-1}^p H_{m-1} \Phi_{m-1,n} - \cdots \\
&= (F_{m-1} - K_{m-1}^p H_{m-1}) \Phi_{m-1,n} - \cdots \\
&= \Phi_{m,m-1}^p \Phi_{m-1,n} - \Phi_{m,m-1}^p K_{m-2}^p H_{m-2} \Phi_{m-2,n} - \cdots \\
&= \Phi_{m,m-1}^p (F_{m-2} - K_{m-2}^p H_{m-2}) \Phi_{m-2,n} - \cdots \\
&= \Phi_{m,m-2}^p \Phi_{m-2,n} - \Phi_{m,m-2}^p K_{m-3}^p H_{m-3} \Phi_{m-3,n} - \cdots \\
&\ \ \vdots \\
&= \Phi_{m,n+1}^p \Phi_{n+1,n} - \Phi_{m,n+1}^p K_n^p H_n \Phi_{n,n} \\
&= \Phi_{m,n}^p \Phi_{m,m} \\
&= \text{LHS}
\end{aligned}$$

The second relation is from the definition of $\Phi_{n,n}^p$ and $\Phi_{n,n}$. □

The factorization in (C.2) allows us to write the following global expression which relates $\hat{u}_{i|i-1}$ to z^i,

$$(U^k)^{-1}(z^k - \mathcal{O}^k u_0) = [z_i - H_i \hat{u}_{i|i-1}]_{i=1}^k \tag{C.5}$$

where $[\cdot]_{i=1}^k$ stands for a k-vector of which the ith element has the form in the bracket.

Proof of (C.5). The proof is by induction.

For $n = 1$, we have

$$z_1 - H_1 F_0 u_0 = z_1 - H_1 \hat{u}_{1|0}$$

holds since $\hat{u}_{1|0} = F_0 u_0$.

Suppose that for $n \leq k-1$, the relation, i.e.,

$$(U^{k-1})^{-1}(z^{k-1} - \mathcal{O}^{k-1}u_0) = [z_i - H_i\hat{u}_{i|i-1}]_{i=1}^{k-1}$$

holds. Then for $n = k$,

$$\text{LHS} = (z_k - H_k\Phi_{k,0}u_0) - H_k\left[\Phi^p_{k,k}K^p_{k-1}, \Phi^p_{k,k-1}K^p_{k-2}, \cdots, \Phi^p_{k,2}K^p_1\right](z^{k-1} - \mathcal{O}^{k-1}u_0)$$

Note that

$$\left[\Phi^p_{k,k}K^p_{k-1}, \Phi^p_{k,k-1}K^p_{k-2}, \cdots, \Phi^p_{k,2}K^p_1\right] = \left[\Phi_{k,k}K^p_{k-1}, \cdots, \Phi_{k,2}K^p_1\right](U^{k-1})^{-1} \qquad \text{(D.3)}$$

by applying the first equality in Lemma D.1 to the RHS of the above. Then,

$$\text{LHS} = z_k - H_k\left(\Phi_{k,0}u_0 + \left[\Phi_{k,k}K^p_{k-1}, \cdots, \Phi_{k,2}K^p_1\right][z_i - H_i\hat{u}_{i|i-1}]_{i=1}^{k-1}\right)$$

by substituting (D.3) above and using the assumption that it holds for $n \leq k-1$. Also by the recursion between $\hat{u}_{i|i-1}$ and $\hat{u}_{i-1|i-2}$,

$$\begin{aligned}
\hat{u}_{k|k-1} &= F_{k-1}\hat{u}_{k-1|k-2} + K^p_{k-1}(z_{k-1} - H_{k-1}\hat{u}_{k-1|k-2}) \\
&= F_{k-1}F_{k-2}\hat{u}_{k-2|k-3} + F_{k-1}K^p_{k-2}(z_{k-2} - H_{k-2}\hat{u}_{k-2|k-3}) \\
&\quad + K^p_{k-1}(z_{k-1} - H_{k-1}\hat{u}_{k-1|k-2}) \\
&\quad \vdots \\
&= \Phi_{k,0}u_0 + \left[\Phi_{k,k}K^p_{k-1}, \cdots, \Phi_{k,2}K^p_1\right][z_i - H_i\hat{u}_{i|i-1}]_{i=1}^{k-1} \qquad \text{(D.4)}
\end{aligned}$$

Thus, substituting (D.4),

$$\text{LHS} = z_k - H_k\hat{u}_{k|k-1} = \text{RHS}$$

$\square$

- The Gaussian part

  The derivation of $\hat{u}_{k|k}$ follows from, in Property 2.3,

$$u_{1|2} = u_1 + \Sigma_{12}\Sigma_2^{-1}(x_2 - u_2)$$

applying Property 2.3 to (D.2) and note that the triangular factorization in (C.2),

$$
\begin{aligned}
\hat{u}_{k|k} &= \Phi_{k,0}u_0 + \Sigma_{x_k z^k}(\Sigma_{z^k})^{-1}(z^k - \mathcal{O}^k u_0) \\
&= \Phi_{k,0}u_0 + \Sigma_{x_k e^k}(U^k)'(U^k S^k (U^k)')^{-1}(z^k - \mathcal{O}^k u_0) \\
&= \Phi_{k,0}u_0 + \Sigma_{x_k e^k}(S^k)^{-1}(U^k)^{-1}(z^k - \mathcal{O}^k u_0) \\
&= \Phi_{k,0}u_0 + \left[\Phi_{k,k}K_k, \cdots, \Phi_{k,1}K_1\right][z_i - H_i \hat{u}_{i|i-1}]_{i=1}^{k} \\
&= \Phi_{k,0}u_0 + \left[\Phi_{k,k}K_{k-1}^p, \cdots, \Phi_{k,2}K_1^p\right][z_i - H_i \hat{u}_{i|i-1}]_{i=1}^{k-1} + K_k(z_k - H_k \hat{u}_{k|k-1}) \\
&= \hat{u}_{k|k-1} + K_k(z_k - H_k \hat{u}_{k|k-1})
\end{aligned}
$$

The second and third equalities hold because of the triangular factorization in (C.2) and (C.3). The second last equality holds because of (D.4).

The derivation of $\Sigma_{k|k}$ also follows from, in Property 2.3,

$$\Sigma_{1|2} = \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{21}$$

as

$$
\begin{aligned}
\Sigma_{k|k} &= \Sigma_{x_k} - \Sigma_{x_k z^k}\Sigma_{z^k}\Sigma'_{x_k z^k} \\
&= F_{k-1}\Sigma_{x_{k-1}}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1} - \Sigma_{x_k e^k}(U^k)'(U^k S^k (U^k)')^{-1}U^k(\Sigma_{x_k e^k})' \\
&= F_{k-1}\Sigma_{x_{k-1}}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1} - \Sigma_{x_k e^k}(S^k)^{-1}(\Sigma_{x_k e^k})' \\
&= F_{k-1}\Sigma_{x_{k-1}}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1} - \left[K_k S_k, F_{k-1}\Sigma_{x_{k-1}e^{k-1}}S^{k-1}\right](S^k)^{-1}[\cdot]' \\
&= F_{k-1}(\Sigma_{x_{k-1}} - \Sigma_{x_{k-1}e^{k-1}}(S^{k-1})^{-1}\Sigma'_{x_{k-1}e^{k-1}})F'_{k-1} \\
&\quad + G_{k-1}Q_{k-1}G'_{k-1} - K_k S_k K'_k \\
&= F_{k-1}\Sigma_{k-1|k-1}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1} - K_k S_k K'_k \\
&= \Sigma_{k|k-1} - K_k S_k K'_k
\end{aligned}
$$

The fourth equality holds due to (D.4), and the second last equality holds by observing the structure of $\Sigma_{k-1|k-1}$.

- The skewness part

  First we prove that

  $$(U^k)^{-1}\Delta_{z^k} = (U^k)^{-1}\left[\Delta^x_{z^k}, \Delta^w_{z^k}, \Delta^v_{z^k}\right]$$
  $$= \left[H_i\Phi^p_{i,0}\Delta^x_0, \quad H_i\textstyle\sum_{j=1}^i \Phi^p_{i,j}G_{j-1}\Delta^w_{j-1}, \quad \Delta^v_i - H_i\textstyle\sum_{j=1}^i \Phi^p_{i,j}K^p_{j-1}\Delta^v_{j-1}\right]_{i=1}^k$$

  Here is a derivation for:

  - First term

    $$\begin{aligned}(U^k)^{-1}\Delta^x_{z^k} &= (U^k)^{-1}[H_i\Phi_{i,0}\Delta^x_0]_{i=1}^k \\ &= \begin{bmatrix} I & -H_k\Phi^p_{k,k}K^p_{k-1} & \cdots & -H_k\Phi^p_{k,2}K^p_1 \\ & I & \cdots & -H_{k-1}\Phi^p_{k-1,2}K^p_1 \\ & & \ddots & \vdots \\ & & & I \end{bmatrix}\begin{bmatrix} H_k\Phi_{k,0}\Delta^x_0 \\ H_{k-1}\Phi_{k-1,0}\Delta^x_0 \\ \vdots \\ H_1\Phi_{1,0}\Delta^x_0 \end{bmatrix} \\ &= [H_i\Phi_{i,0}\Delta^x_0 - H_i\sum_{j=2}^i \Phi^p_{i,j}K^p_{j-1}H_{j-1}\Phi_{j-1,0}\Delta^x_0]_{i=1}^k \\ &= [H_i\Phi_{i,0}\Delta^x_0 - H_i\sum_{j=2}^i \Phi^p_{i,j}K^p_{j-1}H_{j-1}\Phi_{j-1,0}\Delta^x_0 - H_i\Phi^p_{i,1}\underbrace{K^p_0}_{0}H_0\Phi_{0,0}\Delta^x_0]_{i=1}^k \\ &= [H_i\Phi_{i,0}\Delta^x_0 - H_i\sum_{j=1}^i \Phi^p_{i,j}K^p_{j-1}H_{j-1}\Phi_{j-1,0}\Delta^x_0]_{i=1}^k \\ &= [H_i\Phi^p_{i,0}\Delta^x_0]_{i=1}^k\end{aligned}$$

    In each row × column operation in the second line, the second equality in (C.4) (Lemma A.1) is applied, and note that $K^p_0 = 0$ (since no prediction happens from x_{-1} to x_0), so we can add one more term in each row of the fourth line.

– Second term

$$
\begin{aligned}
(U^k)^{-1}\Delta^w_{z^k} &= (U^k)^{-1}[H_i \sum_{j=1}^{i} \Phi_{i,j} G_{j-1}\Delta^w_{j-1}]_{i=1}^k \\
&= (U^k)^{-1} \begin{bmatrix} H_k\Phi_{k,k} & H_k\Phi_{k,k-1} & \cdots & H_k\Phi_{k,1} \\ 0 & H_{k-1}\Phi_{k-1,k-1} & \cdots & H_{k-1}\Phi_{k-1,1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & H_1\Phi_{1,1} \end{bmatrix} \begin{bmatrix} G_{k-1}\Delta^w_{k-1} \\ G_{k-2}\Delta^w_{k-2} \\ \vdots \\ G_0\Delta^w_0 \end{bmatrix} \\
&= \begin{bmatrix} H_k\Phi^p_{k,k} & H_k\Phi^p_{k,k-1} & \cdots & H_k\Phi^p_{k,1} \\ 0 & H_{k-1}\Phi^p_{k-1,k-1} & \cdots & H_{k-1}\Phi^p_{k-1,1} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & H_1\Phi^p_{1,1} \end{bmatrix} \begin{bmatrix} G_{k-1}\Delta^w_{k-1} \\ G_{k-2}\Delta^w_{k-2} \\ \vdots \\ G_0\Delta^w_0 \end{bmatrix} \\
&= [H_i \sum_{j=1}^{i} \Phi^p_{i,j} G_{j-1}\Delta^w_{j-1}]_{i=1}^k
\end{aligned}
$$

The second equality of (C.4) is applied again to get the matrix at the third line.

– Third term

$$
\begin{aligned}
(U^k)^{-1}\Delta^v_{z^k} &= \begin{bmatrix} I & -H_k\Phi^p_{k,k}K^p_{k-1} & \cdots & -H_k\Phi^p_{k,2}K^p_1 \\ & I & \cdots & -H_{k-1}\Phi^p_{k-1,2}K^p_1 \\ & & \ddots & \vdots \\ & & & I \end{bmatrix} \begin{bmatrix} \Delta^v_k \\ \Delta^v_{k-1} \\ \vdots \\ \Delta^v_1 \end{bmatrix} \\
&= [\Delta^v_i - H_i \sum_{j=2}^{i} \Phi^p_{i,j} K^p_{j-1} \Delta^v_{j-1}]_{i=1}^k \\
&= [\Delta^v_i - H_i \sum_{j=2}^{i} \Phi^p_{i,j} K^p_{j-1} \Delta^v_{j-1} - H_i \Phi^p_{i,1} \underbrace{K^p_0 \Delta^v_0}_{0}]_{i=1}^k \\
&= [\Delta^v_i - H_i \sum_{j=1}^{i} \Phi^p_{i,j} K^p_{j-1} \Delta^v_{j-1}]_{i=1}^k
\end{aligned}
$$

$K_0^p \Delta_0^v = 0$ because both K_0^p and Δ_0^v are equal to 0, since neither prediction nor measure is available at the initial time.

Proof of the batch form (3.10), (3.11), *and* (3.12). Now we prove the batch form $\Delta_{k|k-1}$.

By Property 2.3,

$$\begin{aligned}\Delta_{k|k-1} &= \Phi_{k,0}\Delta_{x_k} - \Sigma_{x_k z^{k-1}}(\Sigma_{z^{k-1}})^{-1}\Delta_{z^{k-1}} \\ &= \Phi_{k,0}\Delta_{x_k} - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta_{z^{k-1}}\end{aligned}$$

For brevity,

$$\Delta_{k|k-1} \triangleq \left[\Delta^x_{k|k-1}, \Delta^w_{k|k-1}, \Delta^v_{k|k-1}\right]$$

Note that

$$\Sigma_{x_k e^{k-1}} = \left[\Phi_{k,k} K^p_{k-1} S_{k-1}, \cdots, \Phi_{k,2} K^p_1 S_1\right]$$

By (C.4), (C.7) and $K_0^p = 0$,

$$\begin{aligned}\Delta^x_{k|k-1} &= \Phi_{k,0}\Delta_0^x - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta^x_{z^{k-1}} \\ &= \Phi_{k,0}\Delta_0^x - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}[H_i\Phi^p_{i,0}\Delta_0^x]_{i=1}^{k-1} \\ &= \Phi_{k,0}\Delta_0^x - \sum_{i=2}^{k}\Phi_{k,i}K^p_{i-1}H_{i-1}\Phi^p_{i-1,0}\Delta_0^x \\ &= \Phi_{k,0}\Delta_0^x - \sum_{i=2}^{k}\Phi_{k,i}K^p_{i-1}H_{i-1}\Phi^p_{i-1,0}\Delta_0^x - \Phi_{k,1}\underbrace{K_0^p}_{0}H_0\Phi^p_{0,0}\Delta_0^x \\ &= \Phi_{k,0}\Delta_0^x - \sum_{i=1}^{k}\Phi_{k,i}K^p_{i-1}H_{i-1}\Phi^p_{i-1,0}\Delta_0^x \\ &= \Phi^p_{k,0}\Delta_0^x\end{aligned}$$

By (C.4) and (C.8), and $\Phi_{k,k} = \Phi^p_{k,k} = I$,

$$
\begin{aligned}
\Delta^w_{k|k-1} &= \sum_{i=1}^{k} \Phi_{k,i} G_{i-1} \Delta^w_{i-1} - \Sigma_{x_k e^{k-1}} (S^{k-1})^{-1} (U^{k-1})^{-1} \Delta^w_{z^{k-1}} \\
&= \sum_{i=1}^{k} \Phi_{k,i} G_{i-1} \Delta^w_{i-1} - \sum_{i=2}^{k} \Phi_{k,i} K^p_{i-1} H_{i-1} \sum_{j=1}^{i-1} \Phi^p_{i-1,j} G_{j-1} \Delta^w_{j-1} \\
&= \sum_{i=1}^{k} \Phi_{k,i} G_{i-1} \Delta^w_{i-1} - \begin{bmatrix} \Phi_{k,k} K^p_{k-1} H_{k-1} & \Phi_{k,k-1} K^p_{k-2} H_{k-2} & \cdots & \Phi_{k,2} K^p_1 H_1 \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} \Phi^p_{k-1,k-1} & \Phi^p_{k-1,k-2} & \cdots & \Phi^p_{k-1,1} \\ & \Phi^p_{k-2,k-2} & \cdots & \Phi^p_{k-2,1} \\ & & \ddots & \vdots \\ & & & \Phi^p_{1,1} \end{bmatrix} \begin{bmatrix} G_{k-2} \Delta^w_{k-2} \\ G_{k-3} \Delta^w_{k-3} \\ \vdots \\ G_0 \Delta^w_0 \end{bmatrix} \\
&= \Phi_{k,k} G_{k-1} \Delta^w_{k-1} + \sum_{i=0}^{k-2} \Phi_{k,i+1} G_i \Delta^w_i \\
&\quad - \begin{bmatrix} \Phi_{k,k} K^p_{k-1} H_{k-1} \Phi^p_{k-1,k-1} & \cdots & \sum_{i=1}^{k-1} \Phi_{k,k-i+1} K^p_{k-i} H_{k-i} \Phi^p_{k-i,1} \end{bmatrix} \begin{bmatrix} G_{k-2} \Delta^w_{k-2} \\ \vdots \\ G_0 \Delta^w_0 \end{bmatrix} \\
&= \Phi_{k,k} G_{k-1} \Delta^w_{k-1} + \sum_{i=0}^{k-2} (\Phi_{k,i+1} - \sum_{j=1}^{k-i-1} \Phi_{k,i+j+1} K^p_{i+j} H_{i+j} \Phi^p_{i+j,i+1}) G_i \Delta^w_i \\
&= \Phi^p_{k,k} G_{k-1} \Delta^w_{k-1} + \sum_{i=1}^{k-1} \Phi^p_{k,i} G_{i-1} \Delta^w_{i-1} \\
&= \sum_{i=1}^{k} \Phi^p_{k,i} G_{i-1} \Delta^w_{i-1}
\end{aligned}
$$

The sixth equality holds by using (C.4) again in the parentheses.

By (C.4), (C.9) and $\Phi_{k,k} = \Phi^p_{k,k} = I$,

$$\Delta^v_{k|k-1} = 0 - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}(U^{k-1})^{-1}\Delta^v_{z^{k-1}}$$

$$= -\begin{bmatrix}\Phi_{k,k}K^p_{k-1} & \cdots & \Phi_{k,2}K^p_1\end{bmatrix}[\Delta^v_i - H_i \sum_{j=1}^{i} \Phi^p_{i,j}K^p_{j-1}\Delta^v_{j-1}]^{k-1}_{i=1}$$

$$= -\sum_{i=2}^{k} \Phi_{k,i}K^p_{i-1}\Delta^v_{i-1} + \sum_{i=2}^{k} \Phi_{k,i}K^p_{i-1}H_{i-1}\sum_{j=1}^{i-1}\Phi^p_{i-1,j}K^p_{j-1}\Delta^v_{j-1}$$

$$= -\Phi_{k,k}K^p_{k-1}\Delta^v_{k-1} - \sum_{i=1}^{k-1}(\Phi_{k,i} - \sum_{j=1}^{k-i}\Phi_{k,i+j}K^p_{i+j-1}H_{i+j-1}\Phi^p_{i+j-1,i})K^p_{i-1}\Delta^v_{i-1}$$

$$= -\sum_{i=1}^{k}\Phi^p_{k,i}K^p_{i-1}\Delta^v_{i-1}$$

The fourth equality is obtained in an exact way as for the third equality in $\Delta^w_{k|k-1}$.

Based on the above we can get the exact form of

$$\Delta_{k|k} \triangleq \left[\Delta^x_{k|k}, \Delta^w_{k|k}, \Delta^v_{k|k}\right]$$

where, by (C.3) and the derivation of $\Delta^x_{k|k-1}$,

$$\Delta^x_{k|k} = \Phi_{k,0}\Delta^x_0 - \Sigma_{x_k e^k}(S^k)^{-1}(U^k)^{-1}\Delta^x_{z^k}$$

$$= \Phi_{k,0}\Delta^x_0 - \begin{bmatrix}\Phi_{k,k}K_k & \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}\end{bmatrix}[H_i\Phi^p_{i,0}\Delta^x_0]^k_{i=1}$$

$$= \Phi_{k,0}\Delta^x_0 - \Sigma_{x_k e^{k-1}}(S^{k-1})^{-1}[H_i\Phi^p_{i,0}\Delta^x_0]^{k-1}_{i=1} - K_kH_k\Phi^p_{k,0}\Delta^x_0$$

$$= \Delta^x_{k|k-1} - K_kH_k\Phi^p_{k,0}\Delta^x_0$$

$$= (I - K_kH_k)\Delta^x_{k|k-1}$$

and by (C.3) and the derivation of $\Delta^w_{k|k-1}$,

$$
\begin{aligned}
\Delta^w_{k|k} &= \sum_{i=1}^{k} \Phi_{k,i} G_{i-1} \Delta^w_{i-1} - \Sigma_{x_k e^k} (S^k)^{-1} (U^k)^{-1} \Delta^w_{z^k} \\
&= \sum_{i=1}^{k} \Phi_{k,i} G_{i-1} \Delta^w_{i-1} - \begin{bmatrix} \Phi_{k,k} K_k & \Sigma_{x_k e^{k-1}} (S^{k-1})^{-1} \end{bmatrix} [H_i \sum_{j=1}^{i} \Phi^p_{i,j} G_{j-1} \Delta^w_{j-1}]_{i=1}^{k} \\
&= \Delta^w_{k|k-1} - \Sigma_{x_k e^{k-1}} (S^{k-1})^{-1} [H_i \sum_{j=1}^{i} \Phi^p_{i,j} G_{j-1} \Delta^w_{j-1}]_{i=1}^{k-1} - \Phi_{k,k} K_k H_k \sum_{i=1}^{k} \Phi^p_{k,i} G_{i-1} \Delta^w_{i-1} \\
&= \Delta^w_{k|k-1} - \Phi_{k,k} K_k H_k \sum_{i=1}^{k} \Phi^p_{k,i} G_{i-1} \Delta^w_{i-1} \\
&= \Delta^w_{k|k-1} - K_k H_k \Delta^w_{k|k-1} \\
&= (I - K_k H_k) \Delta^w_{k|k-1}
\end{aligned}
$$

The second last equality holds because we derived $\Delta^w_{k|k-1} = \sum_{i=1}^{k} \Phi^p_{k,i} G_{i-1} \Delta^w_{i-1}$ above.

Similarly,

$$
\begin{aligned}
\Delta^v_{k|k} &= 0 - \Sigma_{x_k e^k} (S^k)^{-1} (U^k)^{-1} \Delta^v_{z^k} \\
&= - \begin{bmatrix} \Phi_{k,k} K_k & \Sigma_{x_k e^{k-1}} (S^{k-1})^{-1} \end{bmatrix} [\Delta^v_i - H_i \sum_{j=1}^{i} \Phi^p_{i,j} K^p_{j-1} \Delta^v_{j-1}]_{i=1}^{k} \\
&= \Delta^v_{k|k-1} - K_k (\Delta^v_k - H_k \sum_{i=1}^{k} \Phi^p_{k,i} K^p_{i-1} \Delta^v_{i-1}) \\
&= (I - K_k H_k) \Delta^v_{k|k-1} - K_k \Delta^v_k
\end{aligned}
$$

The second last equality holds because we derived $\Delta^v_{k|k-1} = \sum_{i=1}^{k} \Phi^p_{k,i} K^p_{i-1} \Delta^v_{i-1}$ above. $\square$

By (C.5) and (C.7)–(C.9), it turns out that, for $i = 1, 2$,

$$\begin{aligned}
\hat{\gamma}_{i,k|k} &= \gamma_i - \Delta'_{z^k} \Sigma_{z^k}^{-1} (z^k - \mathcal{O}^k u_0) \\
&= \gamma_i - \Delta'_{z^k} (U^k S^k (U^k)')^{-1} (z^k - \mathcal{O}^k u_0) \\
&= \gamma_i - ((U^{k-1})^{-1} \Delta_{z^{k-1}})' (S^{k-1})^{-1} (U^{k-1})^{-1} \\
&\quad \cdot (z^{k-1} - \mathcal{O}^{k-1} u_0) - \begin{bmatrix} (H_k \Delta^x_{k|k-1})' \\ (H_k \Delta^w_{k|k-1})' \\ (\Delta^v_k + H_k \Delta^v_{k|k-1})' \end{bmatrix} \\
&\quad \cdot S_k^{-1} (z_k - H_k \hat{u}_{k|k-1}) \\
&= \hat{\gamma}_{i,k|k-1} - K^s_k (z_k - H_k \hat{u}_{k|k-1})
\end{aligned}$$

where $\gamma_i = \left[(\gamma_i^x)', (\gamma_i^w)', (\gamma_i^v)' \right]'$ and the third equality holds by substituting the batch form of $(U^{k-1})^{-1} \Delta_{z^{k-1}}$ and $(U^{k-1})^{-1} (z^{k-1} - \mathcal{O}^{k-1} u_0)$ in (C.6) and (C.6), respectively.

$$\begin{aligned}
\Gamma_{k|k} &= \Gamma - \Delta'_{z^k} \Sigma_{z^k}^{-1} \Delta_{z^k} \\
&= \Gamma - \Delta'_{z^k} (U^k S^k (U^k)')^{-1} \Delta_{z^k} \\
&= \Gamma - ((U^k)^{-1} \Delta_{z^k})' (S^k)^{-1} (U^k)^{-1} \Delta_{z^k} \\
&= \Gamma - ((U^{k-1})^{-1} \Delta_{z^{k-1}})' (S^{k-1})^{-1} (U^{k-1})^{-1} \Delta_{z^{k-1}} \\
&\quad - \begin{bmatrix} (H_k \Delta^x_{k|k-1})' \\ (H_k \Delta^w_{k|k-1})' \\ (\Delta^v_k + H_k \Delta^v_{k|k-1})' \end{bmatrix} S_k^{-1} [\cdot]' \\
&= \Gamma_{k|k-1} - K^s_k S_k (K^s_k)'
\end{aligned}$$

□

Appendix E

The derivation of Equation (4.7)

To derive the unconditional cumulant generation function (4.6) of the first-order skew-Gaussian pdf (4.4), the following lemma is used.

Lemma E.1. *If $X \sim \mathcal{N}_n(0, \Sigma_n)$, then, for any $m \times n$ matrix G and $m \times 1$ vector a, $m \leq n$,*

$$E\{\Phi_m(a + GX)\} = \Phi_m\Big\{\big[(I_m + G\Sigma_n G')^{\frac{1}{2}}\big]^{-1} a\Big\}$$

where $\mathcal{N}_n(0, \Sigma_n)$ denotes the n-dimensional Gaussian pdf with mean 0 and covariance Σ_n, Φ_m denotes the cdf of the m-dimensional Gaussian $\mathcal{N}_n(0, I_m)$, and $(\cdot)^{\frac{1}{2}}$ denotes the positive definite square root of a positive definite matrix.

Proof. Let $Y \sim \mathcal{N}_m(0, I_m)$, and X and Y be independent. Since

$$\Phi_m(a + GX) = P\{Y \leq a + GX \mid X\}$$

we have

$$
\begin{aligned}
E\{\Phi_m(a+GX)\} &= E[P\{Y \le a+GX \mid X\}] \\
&= P\{Y \le a+GX\} \\
&= \Phi_m\Big\{\big[(I_m + G\Sigma_n G')^{\frac{1}{2}}\big]^{-1} a\Big\}
\end{aligned}
$$

where the second equality follows from the total expectation theorem and the last equality holds because $Y - GX \sim \mathcal{N}_m(0, I_m + G\Sigma_n G')$.

All inequalities here are component-wise. □

Using Lemma E.1, the cumulant generating function of (4.4) is easily derived from its definition, as

$$
\begin{aligned}
K(t) &= \log \int_{\mathbb{R}^n} e^{t'x} 2N(x; u, \Omega)\Phi(\alpha'(x-u))dx \\
&= \log \frac{2}{\sqrt{|2\pi\Omega|}} \int_{\mathbb{R}^n} \exp(-\frac{1}{2}(x-\Omega t-u)'\Omega^{-1}(x-\Omega t-u) + \frac{1}{2}t'\Omega t + u't) \\
&\quad \times \Phi(\alpha'(x-u))dx \\
&= \log 2\exp(\frac{1}{2}t'\Omega t + u't) \int_{\mathbb{R}^n} \Phi(\alpha'(x-u))\mathcal{N}(x; \Omega t - u, \Omega)dx \\
&= \log 2\exp(\frac{1}{2}t'\Omega t + u't)\Phi(\frac{\alpha'\Omega}{\sqrt{1+\alpha'\Omega\alpha}}t) \\
&= \frac{1}{2}t'\Omega_{11}t + u_1't + \log(2\Phi(\delta_1't))
\end{aligned}
$$

where the second equality follows from completing the square, and the fourth equality holds because of Lemma E.1.

Let x and y be two column vectors. Define the k-th derivative of y w.r.t. x as

1. $\frac{dy}{dx} = \frac{d}{dx}y' = y' \otimes \frac{d}{dx}$

2. $\frac{d^k y}{dx^k} = \frac{d}{dx}\left(\frac{d^{k-1}y}{dx^{k-1}}\right) = \frac{d}{dx}y' \otimes \frac{d}{dx'} \otimes \cdots \otimes \frac{d}{dx'}$ (k times)

The next theorem provides a straightforward connection between central moments and cumulants of low order.

Theorem E.2 (Kollo (2005))**.** *Let x be a random p-dimensional vector, $c_i(x)$ be the i-th cumulant, $m_i(x)$ and $\bar{m}_i(x)$ be the i-th moment and central moment, respectively. Then,*

1. $c_1(x) = m_1(x) = E(x)$;
2. $c_2(x) = \bar{m}_2(x) = \mathrm{cov}(x)$;
3. $c_3(x) = \bar{m}_3(x) = E[(x - E(x))(x - E(x))'^{\otimes 2}]$.

Proposition E.1. *The first three central moments of the pdf* (4.4) *are given as*

$$m_1(x) = \left.\frac{dK(t)}{dt}\right|_{t=0} = \Omega t + u + \left.\frac{\mathcal{N}(\delta' t)}{\Phi(\delta' t)}\right|_{t=0} = u + \sqrt{\frac{2}{\pi}}\delta$$

$$\bar{m}_2(x) = \left.\frac{d}{dt}K(t) \otimes \frac{d}{dt'}\right|_{t=0} = \Omega - \left.\left[\frac{\mathcal{N}^{(1)}(\delta' t)}{\Phi(\delta' t)} + \left(\frac{\mathcal{N}(\delta' t)}{\Phi(\delta' t)}\right)^2\right]\delta\delta'\right|_{t=0}$$

$$= \Omega - \frac{2}{\pi}\delta\delta'$$

$$\bar{m}_3(x) = \left.\frac{d}{dt}K(t) \otimes \frac{d}{dt'} \otimes \frac{d}{dt'}\right|_{t=0}$$

$$= \left.\left[2\left(\frac{\mathcal{N}(\delta' t)}{\Phi(\delta' t)}\right)^3 - 3\frac{\mathcal{N}(\delta' t)\mathcal{N}^{(1)}(\delta' t)}{(\Phi(\delta' t))^2} + \frac{\mathcal{N}^{(2)}(\delta' t)}{\Phi(\delta' t)}\right]\delta vec'(\delta\delta')\right|_{t=0}$$

$$= \left(\frac{4}{\pi} - 1\right)\sqrt{\frac{2}{\pi}}\delta vec'(\delta\delta')$$

where $\mathcal{N}^{(i)}(\delta' t)$ denotes the i-th derivative of the univariate Gaussian pdf evaluated at $\delta' t$.

We omit our derivation of Proposition E.1 here because a proof of its generalization is given in Appendix A.

Remark E.1. *The central moments of the conditional first-order skew-Gaussian pdf can be derived similarly as in Proposition E.1 by simply replacing* $\delta' t$ *with* $x_0 + \delta'_{1|2} t$.

Bibliography

Agamennoni, G., Nieto, J. I., and Nebot, E. M. (2012), "Approximate inference in state-space models with heavy-tailed noise," *IEEE Transactions on Signal Processing*, 60, 5024–5037.

Alspach, D. and Sorenson, H. (1972), "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Transactions on Automatic Control*, 17, 439–448.

Anderson, B. D. and Moore, J. B. (2005), *Optimal Filtering*, Dover Publications.

Arasaratnam, I. and Haykin, S. (2009), "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, 54, 1254–1269.

Arasaratnam, I., Haykin, S., and Elliott, R. J. (2007), "Discrete-time nonlinear filtering algorithms using Gauss–Hermite quadrature," *Proceedings of the IEEE*, 95, 953–977.

Arellano-Valle, R. B. and Azzalini, A. (2006), "On the unification of families of skew-normal distributions," *Scandinavian Journal of Statistics*, 33, 561–574.

Arellano-Valle, R. B., Branco, M. D., and Genton, M. G. (2006), "A unified view on skewed distributions arising from selections," *Canadian Journal of Statistics*, 34, 581–601.

Athans, M., Wishner, R., and Bertolini, A. (1968), "Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements," *IEEE Transactions on Automatic Control*, 13, 504–514.

Azzalini, A. (2013), *The Skew-Normal and Related Families*, vol. 3, Cambridge University Press.

Azzalini, A. and Capitanio, A. (2003), "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 367–389.

Bar-Shalom, Y., Li, X. R., and Kirubarajan, T. (2004), *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*, John Wiley & Sons.

Barber, D. (2012), *Bayesian Reasoning and Machine Learning*, Cambridge University Press.

Bartlett, N. J., Renton, C., and Wills, A. G. (2020), "A closed-form prediction update for extended target tracking using random matrices," *IEEE Transactions on Signal Processing*, 68, 2404–2418.

Bellaire, R. L., Kamen, E. W., and Zabin, S. M. (1995), "New nonlinear iterated filter with applications to target tracking," in *Signal and Data Processing of Small Targets 1995*, vol. 2561, pp. 240–251.

Benavoli, A., Azzimonti, D., and Piga, D. (2020), "Skew Gaussian processes for classification," *Machine Learning*, 109, 1877–1902.

Beneš, V. E. (1981), "Exact finite-dimensional filters for certain diffusions with nonlinear drift," *Stochastics: An International Journal of Probability and Stochastic Processes*, 5, 65–92.

Botev, Z. I. (2016), "The normal law under linear restrictions: simulation and estimation via minimax tilting," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 125–148.

Challa, S., Bar-Shalom, Y., and Krishnamurthy, V. (2000), "Nonlinear filtering via generalized Edgeworth series and Gauss-Hermite quadrature," *IEEE Transactions on Signal Processing*, 48, 1816–1820.

Chopin, N. (2011), "Fast simulation of truncated Gaussian distributions," *Statistics and Computing*, 21, 275–288.

Chung, K.-L. (2001), *A Course in Probability Theory*, Academic press.

Daum, F. (1986), "Exact finite-dimensional nonlinear filters," *IEEE Transactions on Automatic Control*, 31, 616–622.

Daum, F. (2005), "Nonlinear filters: beyond the Kalman filter," *IEEE Aerospace and Electronic Systems Magazine*, 20, 57–69.

Davis, P. J. and Rabinowitz, P. (2007), *Methods of Numerical Integration*, Courier Corporation.

Denham, W. F. and Pines, S. (1966), "Sequential estimation when measurement function nonlinearity is comparable to measurement error." *AIAA journal*, 4, 1071–1076.

Doucet, A., De Freitas, N., Gordon, N. J., et al. (2001), *Sequential Monte Carlo Methods in Practice*, Springer.

Elliott, R. J. and Krishnamurthy, V. (1999), "New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models," *IEEE Transactions on Automatic Control*, 44, 938–951.

Elliott, R. J., Dufour, F., and Sworder, D. D. (1996), "Exact hybrid filters in discrete time," *IEEE Transactions on Automatic Control*, 41, 1807–1810.

Evensen, G. (1994), "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of Geophysical Research: Oceans*, 99, 10143–10162.

Feldmann, M., Fränken, D., and Koch, W. (2010), "Tracking of extended objects and group targets using random matrices," *IEEE Transactions on Signal Processing*, 59, 1409–1420.

Ferrante, M. (1992), "On the existence of finite-dimensional filters in discrete time," *Stochastics: An International Journal of Probability and Stochastic Processes*, 40, 169–179.

Ferrante, M. and Giummolé, F. (1995), "Finite dimensional filters for a discrete-time nonlinear system with generalized gaussian white noise," *Stochastics and Stochastic Reports*, 53, 195–211.

Ferrante, M. and Runggaldier, W. J. (1990), "On necessary conditions for the existence of finite-dimensional filters in discrete time," *Systems & control letters*, 14, 63–69.

Ferrante, M. and Vidoni, P. (1999), "A Gaussian-generalized inverse Gaussian finite-dimensional filter," *Stochastic processes and their applications*, 84, 165–176.

Genton, M. G. (2004), *Skew-Elliptical Distributions and Their Applications: A Journey beyond Normality*, CRC Press.

Gerstner, T. and Griebel, M. (1998), "Numerical integration using sparse grids," *Numerical Algorithms*, 18, 209–232.

Girón, F. J. and Rojano, J. C. (1994), "Bayesian Kalman filtering with elliptically contoured errors," *Biometrika*, 81, 390–395.

Golub, G. and Pereyra, V. (2003), "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Problems*, 19.

González-Farías, G., Domínguez-Molina, A., and Gupta, A. K. (2004), "Additive properties of skew normal random vectors," *Journal of Statistical Planning and Inference*, 126, 521–534.

GÜnther, S. (1981), "Finite dimensional filter systems in discrete time," *Stochastics: An International Journal of Probability and Stochastic Processes*, 5, 107–114.

He, L., Chen, J., and Qi, Y. (2018), "Event-based state estimation: optimal algorithm with generalized closed skew normal distribution," *IEEE Transactions on Automatic Control*, 64, 321–328.

He, X., Sithiravel, R., Tharmarasa, R., Balaji, B., and Kirubarajan, T. (2014), "A spline filter for multidimensional nonlinear state estimation," *Signal Processing*, 102, 282–295.

Heiss, F. and Winschel, V. (2008), "Likelihood approximation by numerical integration on sparse grids," *Journal of Econometrics*, 144, 62–80.

Henriksen, R. (1982), "The truncated second-order nonlinear filter revisited," *IEEE Transactions on Automatic Control*, 27, 247–251.

Ho, Y. and Lee, R. (1964), "A Bayesian approach to problems in stochastic estimation and control," *IEEE Transactions on Automatic Control*, 9, 333–339.

Huang, Y., Zhang, Y., Wu, Z., Li, N., and Chambers, J. (2017a), "A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices," *IEEE Transactions on Automatic Control*, 63, 594–601.

Huang, Y., Zhang, Y., Shi, P., Wu, Z., Qian, J., and Chambers, J. A. (2017b), "Robust Kalman filters based on Gaussian scale mixture distributions with application to target tracking," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49, 2082–2096.

Ito, K. and Xiong, K. (2000), "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, 45, 910–927.

Jazwinski, A. H. (2007), *Stochastic Processes and Filtering Theory*, Courier Corporation.

Jia, B., Xin, M., and Cheng, Y. (2012), "Sparse-grid quadrature nonlinear filtering," *Automatica*, 48, 327–341.

Jia, B., Xin, M., and Cheng, Y. (2014), "Relations between sparse-grid quadrature rule and spherical-radial cubature rule in nonlinear Gaussian estimation," *IEEE Transactions on Automatic Control*, 60, 199–204.

Julier, S. J. (1998), "Skewed approach to filtering," in *Signal and Data Processing of Small Targets 1998*, vol. 3373, pp. 271–282.

Julier, S. J. and Uhlmann, J. K. (2004), "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, 92, 401–422.

Kailath, T., Sayed, A. H., and Hassibi, B. (2000), *Linear Estimation*, Prentice Hall.

Kalman, R. E. (1960), "A new approach to linear filtering and prediction problems," *ASME J. Basic Engineering*, Series 82D.

Kan, R. and Robotti, C. (2017), "On moments of folded and truncated multivariate normal distributions," *Journal of Computational and Graphical Statistics*, 26, 930–934.

Koch, J. W. (2008), "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Transactions on Aerospace and Electronic Systems*, 44, 1042–1059.

Kollo, T. (2005), *Advanced Multivariate Statistics with Matrices*, Springer.

Krishnamurthy, V. and Evans, J. (1998), "Finite-dimensional filters for passive tracking of markov jump linear systems," *Automatica*, 34, 765–770.

Kulhavỳ, R. (1996), *Recursive Nonlinear Estimation: A Geometric Approach*, Springer.

Kushner, H. (1967), "Approximations to optimal nonlinear filters," *IEEE Transactions on Automatic Control*, 12, 546–556.

Lan, J. and Li, X. R. (2016), "Tracking of extended object or target group using random matrix: New model and approach," *IEEE Transactions on Aerospace and Electronic Systems*, 52, 2973–2989.

Lefebvre, T., Bruyninckx, H., and De Schuller, J. (2002), "Comment on "a new method for the nonlinear transformation of means and covariances in filters and estimators"," *IEEE Transactions on Automatic Control*, 47, 1406–1409.

Li, X. R. (2004), "Recursibility and optimal linear estimation and filtering," in *IEEE Conference on Decision and Control*, vol. 2, pp. 1761–1766.

Li, X. R. (2015), *Lecture Notes: Fundamentals of Applied Estimation and Filtering*, University of New Orleans.

Li, X. R. and Jilkov, V. P. (2004), "A survey of maneuvering target tracking: approximation techniques for nonlinear filtering," in *Signal and Data Processing of Small Targets 2004*, vol. 5428, pp. 537–550.

Li, X. R. and Jilkov, V. P. (2010), "A survey of maneuvering target tracking-part VIa: Density-based exact nonlinear filtering," in *Signal and Data Processing of Small Targets 2010*, vol. 7698.

Li, X. R. and Jilkov, V. P. (2012), "A survey of maneuvering target tracking-part VIc: approximate nonlinear density filtering in discrete time," in *Signal and Data Processing of Small Targets 2012*, vol. 8393.

Lototsky, S. V. (2006), "Wiener chaos and nonlinear filtering," *Applied Mathematics and Optimization*, 54, 265–291.

Ma, Y. and Genton, M. G. (2004), "Flexible class of skew-symmetric distributions," *Scandinavian Journal of Statistics*, 31, 459–468.

McReynolds, S. (1975), "Multidimensional Hermite-Gaussian quadrature formulae and their application to nonlinear estimation," *Proc. 6th Symp. Nonlinear Estimation.*

Mendel, J. M. (1995), *Lessons in Estimation Theory for Signal Processing, Communications, and Control*, Pearson Education.

Móri, T. F., Rohatgi, V. K., and Székely, G. (1994), "On multivariate skewness and kurtosis," *Theory of Probability and Its Applications*, 38, 547–551.

Nørgaard, M., Poulsen, N. K., and Ravn, O. (2000), "New developments in state estimation for nonlinear systems," *Automatica*, 36, 1627–1638.

Nurminen, H., Ardeshiri, T., Piché, R., and Gustafsson, F. (2018), "Skew-t filter and smoother with improved covariance matrix approximation," *IEEE Transactions on Signal Processing*, 66, 5618–5633.

Rabiner, L. R. (1989), "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 77, 257–286.

Rezaie, J. and Eidsvik, J. (2014), "Kalman filter variants in the closed skew normal setting," *Computational Statistics & Data Analysis*, 75, 1–14.

Rezaie, R. and Li, X. R. (2019), "Gaussian conditionally Markov sequences: Dynamic models and representations of reciprocal and other classes," *IEEE Transactions on Signal Processing*, 68, 155–169.

Rezaie, R. and Li, X. R. (2020), "Destination-directed trajectory modeling, filtering, and prediction using conditionally markov sequences," *IEEE Transactions on Aerospace and Electronic Systems*, 57, 820–833.

Ristic, B., Arulampalam, S., and Gordon, N. (2003), *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House.

Roth, M., Hendeby, G., and Gustafsson, F. (2016), "Nonlinear Kalman filters explained: A tutorial on moment computations and sigma point methods," *Journal of Advances in Information Fusion*, 11, 47–70.

Sawitzki, G. (1979), "Exact filtering in exponential families: discrete time," in *Stochastic Control Theory and Stochastic Differential Systems*, pp. 554–558, Springer.

Šimandl, M., Královec, J., and Söderström, T. (2006), "Advanced point-mass method for nonlinear state estimation," *Automatica*, 42, 1133–1145.

Singh, K. K., Kumar, S., Dixit, P., and Bajpai, M. K. (2021), "Kalman filter based short term prediction model for COVID-19 spread," *Applied Intelligence*, 51, 2714–2726.

Smidl, V. and Quinn, A. (2008), "Variational Bayesian filtering," *IEEE Transactions on Signal Processing*, 56, 5020–5030.

Stoer, J. and Bulirsch, R. (2002), *Introduction to Numerical Analysis*, Springer.

Sukhavasi, R. T. and Hassibi, B. (2009), "The Kalman like particle filter: optimal estimation with quantized innovations/measurements," in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pp. 4446–4451.

Taghvaei, A. and Mehta, P. G. (2021), "Optimal Transportation Methods in Nonlinear Filtering," *IEEE Control Systems Magazine*, 41, 34–49.

Van Trees, H. L. (2004), *Detection, Estimation, and Modulation Theory, part I: Detection, Estimation, and Linear Modulation Theory*, John Wiley & Sons.

Vidoni, P. (1999), "Exponential family state space models based on a conjugate latent process," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 213–221.

Wang, F. and Balakrishnan, V. (2002), "Robust Kalman filters for linear time-varying systems with stochastic parametric uncertainties," *IEEE Transactions on Signal Processing*, 50, 803–813.

Wang, J., Boyer, J., and Genton, M. G. (2004), "A skew-symmetric representation of multivariate distributions," *Statistica Sinica*, pp. 1259–1270.

Wilhelm, S. and Manjunath, B. G. (2010), "tmvtnorm: package for the truncated multivariate normal distribution," *Sigma*, 2, 25.

Wong, E. (1985), *Stochastic Processes in Engineering Systems*, Springer-Verlag.

Wu, W.-R. (1993), "Target racking with glint noise," *IEEE Transactions on Aerospace and Electronic Systems*, 29, 174–185.

Wu, W.-R. and Chang, D.-C. (1996), "Maneuvering target tracking with colored noise," *IEEE Transactions on Aerospace and Electronic Systems*, 32, 1311–1320.

Wu, Y., Hu, D., Wu, M., and Hu, X. (2006), "A numerical-integration perspective on Gaussian filters," *IEEE Transactions on Signal Processing*, 54, 2910–2921.

Zhang, C., Bütepage, J., Kjellström, H., and Mandt, S. (2018), "Advances in variational inference," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41, 2008–2026.

Zoubir, A. M., Koivunen, V., Chakhchoukh, Y., and Muma, M. (2012), "Robust estimation in signal processing: tutorial-style treatment of fundamental concepts," *IEEE Signal Processing Magazine*, 29, 61–80.

Vita

Haozhan Meng received the B.S. and M.S. degrees from Xi'an Jiaotong University in Electrical Engineering and University of New Orleans in Mathematics in 2010 and 2019, respectively. His research interests include nonlinear state estimation, sequential analysis, stochastic modeling, and their applications to target tracking.