Advanced Kalman Filtering Approaches to Bayesian State Estimation

Michael Roth
Cover illustration: The blue lines illustrate consecutive state vectors of a 40-dimensional Lorenz model that is often used a benchmark problem for ensemble Kalman filtering.
To Kora, Clara and Constantin
Abstract

Bayesian state estimation is a flexible framework to address relevant problems at the heart of existing and upcoming technologies. Application examples are obstacle tracking for driverless cars and indoor navigation using smartphone sensor data. Unfortunately, the mathematical solutions of the underlying theory cannot be translated to computer code in general. Therefore, this thesis discusses algorithms and approximations that are related to the Kalman filter (KF).

Four scientific articles and an introduction with the relevant background on Bayesian state estimation theory and algorithms are included. Two articles discuss nonlinear Kalman filters, which employ the KF measurement update in nonlinear models. The numerous variants are presented in a common framework and the employed moment approximations are analyzed. Furthermore, their application to target tracking problems is discussed. A third article analyzes the ensemble Kalman filter (EnKF), a Monte Carlo implementation of the KF that has been developed for high-dimensional geoscientific filtering problems. The EnKF is presented in a simple KF framework, including its challenges, important extensions, and relations to other filters. Whereas the aforementioned articles contribute to the understanding of existing algorithms, a fourth article devises novel filters and smoothers to address heavy-tailed noise. The development is based on Student's t distribution and provides simple recursions in the spirit of the KF. The introduction and articles are accompanied by extensive simulation experiments.
Populärvetenskaplig sammanfattning


Kalmanfiltret används ofta tillsammans med ett antagande om gaussiska brus-signaler. Dock förekommer ofta i praktiken brusvärden sporadiskt, dessa passar dåligt in på det gaussiska antagandet. Därför undersöks metoder baserat på student t-fördelat brus som bättre beskriver stora avvikelser från det normala. Ett av resultaten i avhandlingen är ett filter baserat på detta brusantagandet. Den nya algoritmen är olinjär och det visas att det ger bra resultat i ett mer utmanande scenario när filtret jämförs med andra vanligt förekommande metoder.

Acknowledgments

I begin writing these lines in January 2016 on one of those cloudy empty-corridor Saturdays at uni. These days mark the end of my time as PhD student at the automatic control group in Linköping. It has been an exciting journey during which I have had the privilege to investigate relevant research problems, take PhD courses by distinguished lecturers, and visit parts of the world for conferences and workshops. Of course I also got to know the hard work associated with PhD life.

The following people have helped me in completing my studies.

First, I would like to thank my supervisors Fredrik Gustafsson and Gustaf Hendeby. Fredrik’s productivity is unique and I have learned a lot from our collaborations in research, teaching, and master’s supervision. Gustaf Hendeby joined in as co-supervisor in 2013. Especially during the last year, our regular discussions and his comments have helped to finalize this thesis.

Furthermore, I would like to acknowledge the head of the automatic control group, Svante Gunnarsson. If Svante is the head, then Ninna Stensgård is the heart of our group. She and the former coordinator Åsa Karmelind have helped with many administrative aspects. I would also like to thank Mikael Norrlöf for establishing the contact to Fredrik that led to my employment.

The outcomes of this thesis have been developed with several co-authors at Linköping University: Tohid Ardeshiri, Carsten Fritsche, Emre Özkan, Umut Orguner, as well as Fredrik and Gustaf. The cooperation and the many discussions are gratefully acknowledged. Furthermore, Jonas Linder and Erik Hedberg have given valuable comments on the introductory chapters of this thesis. Thank you.

I have been lucky enough to share my time as PhD student with colleagues who soon became friends. Therefore, I would like to thank Tohid, Niklas Wahlström, George Mathai, Ylva Jung, André Carvalho Bittencourt, Manon Kok, Bram Dil, Erik, Jonas, Johan Dahlin, Saikat Saha, Biqiang Mu, Christian Lyzell, Antonios Pitarokoilis, Henrik Ohlsson, Christian Lundquist, Sina Khoshfetrat Pakzad, and the other colleagues of the control group. We shared offices, apartments, conference trips, Kårallen lunches, PhD courses, and many memorable conversations. Especially Tohid (with family) and Niklas (with Nicky) are acknowledged for being so welcoming during the first months in Sweden.

Beyond the control group, I would like to thank the other research fellows in the MC Impulse program; Oliver Heirich at DLR, Oberpfaffenhofen; Per Sundbom and Laila Hübbert at the Department of Medical and Health Sciences, Linköping University; Marcus Baum and Shishan Yang at the Data Fusion Group, University of Göttingen; and Michael Meyer zu Hörste at DLR, Braunschweig.

I also want to thank our family friends in Linköping, especially Dörte, Jens, Magdalena, Erik, Silvana, and Matthias, who have supported us in many ways. Moreover, I want to acknowledge all friends who came to visit us in Sweden, especially Lucas, Ansgar, and my university friends from Berlin. After returning to Germany, the “cucina mobile”, Peggy, Jens, Marianne, Konstantin, Kathleen, and Constantin have helped me feel at home again quickly. Thank you!

My academic studies would not have been possible without the support of my parents Ingrid and Christian and my brothers Thomas and Andreas. Especially
Thomas has helped in so many ways. I hope that we will soon see each other more often. Also, I want to thank the Neupert family Gabi, Franz, Helga, and Gregor for all the support. Thank you Gabi and Ingrid for taking care of the kids on most afternoons.

Finally, I want to thank my wife and Kids. Liebe Kora, ich bin dankbar und stolz, dass du meine Frau und die Mutter unserer Kinder bist. Danke für das Abenteuer Schweden, was mit meinem Abschluss ein Ende findet. Danke für die Unterstützung als du von der Traufe (Referendariat) in den Regen (Beruf und Kids in Braunschweig; Partner in Schweden) geschickt wurdest. Danke für all die schönen Momente früher, jetzt, und in der Zukunft! Liebe Clara, lieber Consti, zwei tollere Kinder gibt es nicht noch einmal auf der Welt. Ich bin so froh, dass es euch gibt und freue mich, dass ich bald wieder mehr Zeit für euch habe!

Linköping, Sweden
Michael Roth

Financial support for the work in this thesis has been provided by the MC Impulse project, a European commission FP7 Marie Curie initial training network; the project Cooperative Localization (CoopLoc) funded by Swedish Foundation for Strategic Research (SSF); the Swedish strategic research center Security Link; and the project Scalable Kalman Filters granted by the Swedish Research Council (VR).
Contents

I Background

1 Introduction 3
  1.1 Motivation ................................................. 3
  1.2 Outline and Contributions ................................. 9

2 Bayesian State Estimation 15
  2.1 Stochastic State-Space Models ............................. 16
  2.2 Bayesian State Estimation via Basic Density Manipulations ... 17
  2.3 Elegant Sequential Solutions ............................... 19
  2.4 Instructive Scalar Examples ............................... 20
    2.4.1 “The particle filter example” ........................... 20
    2.4.2 A Student’s t random walk observed in noise ............ 21
    2.4.3 Saturated measurements ............................... 24

3 State Estimation Algorithms 27
  3.1 Particle Filters and Smoothers ............................. 27
  3.2 The Kalman Filter and the Rauch-Tung-Striebel Smoother ........ 30
  3.3 Nonlinear Kalman Filters .................................. 31
  3.4 Student’s t Filters and Smoothers .......................... 33
  3.5 Ensemble Kalman Filters .................................. 36

4 Concluding Remarks 41

A On Probability Density Functions 45
  A.1 Relevant Results and Definitions .......................... 45
  A.2 Important Probability Density Functions .................... 46

B Exact Bayesian State Estimation for Scalar Problems 49
  B.1 The Algorithmic Treatment of Scalar Densities ............... 49
  B.2 Point Mass Filters and Smoothers ......................... 50
  B.3 A Likelihood-free Filter .................................. 51
<p>| Contents |
|----------|---|
| 3.4 A scalar example | 128 |
| 4 A Student’s $t$ Filter | 130 |
| 4.1 A simple filter based on intermediate approximations | 130 |
| 4.2 Approximation by a joint $t$ density | 132 |
| 4.3 Marginal approximation via matrix parameter adjustment | 135 |
| 5 Algorithmic Properties and Extensions | 136 |
| 5.1 An inherited optimality property | 138 |
| 5.2 Square root implementation | 138 |
| 5.3 Application to nonlinear models | 139 |
| 6 A Student’s $t$ Smoother | 140 |
| 7 Simulation Examples | 141 |
| 7.1 A scalar example | 141 |
| 7.2 Drone tracking | 142 |
| 8 Concluding Remarks | 147 |
| A On Quadratic Forms of Partitioned Vectors | 147 |
| B The Kalman filter and RTS smoother | 148 |
| B.1 The Kalman filter equations | 148 |
| B.2 A compact derivation of the RTS smoother | 148 |
| Bibliography | 150 |
| C The Ensemble Kalman Filter: A Signal Processing Perspective | 153 |
| 1 Introduction | 156 |
| 2 Filtering and EnKF literature | 157 |
| 3 A Signal Processing Introduction to the Ensemble Kalman Filter | 159 |
| 3.1 A brief Kalman filter review | 159 |
| 3.2 The ensemble idea | 160 |
| 3.3 The EnKF time update | 160 |
| 3.4 The EnKF measurement update | 161 |
| 3.5 The EnKF gain | 162 |
| 4 Some Properties and Challenges of the EnKF | 162 |
| 4.1 Asymptotic convergence results | 162 |
| 4.2 Computational complexity | 163 |
| 4.3 Sampling and coupling effects for finite ensemble size | 163 |
| 5 Important Extensions to the EnKF | 164 |
| 5.1 Sequential updates | 164 |
| 5.2 Model knowledge in the EnKF and square-root filters | 165 |
| 5.3 Ensemble inflation | 166 |
| 5.4 Localization | 166 |
| 5.5 The EnKF gain and least squares | 169 |
| 6 Relations to Other Algorithms | 169 |
| 7 Instructive Simulation Examples | 170 |
| 7.1 A scalar linear Gaussian model | 170 |
| 7.2 The particle filter benchmark | 171 |
| 7.3 Batch smoothing using the EnKF | 173 |
| 7.4 The 40-dimensional Lorenz model | 178 |
| 8 Concluding Remarks | 181 |</p>
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bibliography</td>
<td>183</td>
</tr>
<tr>
<td>D</td>
<td>EKF/UKF Maneuvering Target Tracking</td>
<td>189</td>
</tr>
<tr>
<td>1</td>
<td>Introduction</td>
<td>191</td>
</tr>
<tr>
<td>2</td>
<td>Coordinated Turn Models</td>
<td>192</td>
</tr>
<tr>
<td>2.1</td>
<td>From particle motion to discrete-time CT models</td>
<td>192</td>
</tr>
<tr>
<td>2.2</td>
<td>Coordinated turn with polar velocity</td>
<td>194</td>
</tr>
<tr>
<td>2.3</td>
<td>Coordinated turn with Cartesian velocity</td>
<td>195</td>
</tr>
<tr>
<td>2.4</td>
<td>Process noise assumptions</td>
<td>195</td>
</tr>
<tr>
<td>3</td>
<td>EKF and UKF for Nonlinear Filtering</td>
<td>196</td>
</tr>
<tr>
<td>3.1</td>
<td>Bayesian filtering theory</td>
<td>196</td>
</tr>
<tr>
<td>3.2</td>
<td>Kalman filter adaptations for nonlinear filtering</td>
<td>196</td>
</tr>
<tr>
<td>3.3</td>
<td>The role of the state coordinate system</td>
<td>197</td>
</tr>
<tr>
<td>4</td>
<td>Experiment and Discussion</td>
<td>198</td>
</tr>
<tr>
<td>4.1</td>
<td>Experiment setup</td>
<td>198</td>
</tr>
<tr>
<td>4.2</td>
<td>Simulation results and discussion</td>
<td>199</td>
</tr>
<tr>
<td>4.3</td>
<td>Summary</td>
<td>201</td>
</tr>
<tr>
<td>5</td>
<td>Concluding Remarks</td>
<td>204</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>205</td>
</tr>
</tbody>
</table>
Part I

Background
This chapter gives the motivation for addressing Bayesian state estimation as research topic, lists the contributions, and gives an outline of the thesis.

1.1 Motivation

Bayesian state estimation is an exciting research topic in many aspects.

First, it addresses the important question of how to extract relevant information from sensor data in the presence of uncertainty. This task is at the very heart of existing and future technologies that influence our everyday lives. For example, autonomous cars must continuously estimate the position of all objects (cars, bicyclists, pedestrians, etc.) in their vicinity in order to drive safely. The car itself and many of the objects are moving, and there is information collected by sensors on the car (cameras, radar, lidar, etc.). The resulting challenge of estimating positions and velocities that change over time from a stream of measurements is best treated as a Bayesian state estimation problem. A similar underlying structure can be found in as diverse challenges as weather prediction, climate analysis, navigation, trading price estimation in finance, and speech processing via tracking of formant frequencies. Bayesian state estimation provides a powerful and flexible framework to approach all of them.

Second, the number of sensors in our environment is ever increasing. Accelerometers, gyroscopes, barometers, and magnetometers can be found in most phones today, whereas such a range of sensors could only be seen in aircraft a few decades ago. Using Bayesian state estimation, the sensor information can be processed to yield useful information, for example, for indoor navigation. Similar statements can be made also for the vast amount of unconventional “sensor” data as collected by surveys, apps, and websites.

Third, Bayesian state estimation is based on a mathematically rigorous for-
Introduction

Figure 1.1: An air traffic control scenario with several aircraft. The airport tower is equipped with a radar.

mulation in terms of stochastic state-space models and conditional probability distributions. This formulation even allows for conceptual solutions. A research challenge arises because the elegant mathematical results cannot be translated to computer code in all but a few special cases. Hence, Bayesian state estimation is often about sacrificing some of the mathematical rigor and finding viable approximations that yield implementable algorithms. One example are the Kalman filters discussed in this thesis.

Finally, Bayesian state estimation is an established topic researched since the 1950s and there are available tools such as Kalman or particle filters. However, new challenges arise with upcoming technologies. For example, the miniaturized sensors in our phones exhibit different errors than the high-end accelerometers and gyroscopes used before. Also the estimation challenges become more difficult, for example, with the many types of objects and their behavior that must be addressed in object tracking for autonomous cars. Research is required to adapt the existing algorithms to these new challenges.

The above concepts and challenges are further specified and illustrated using a number of examples.

**Air traffic control**

A classical application area of Bayesian state estimation is air traffic control [13], and here used to explain the notion of Bayesian filtering. Figure 1.1 illustrates the scenario of aircraft in the vicinity of an airport. The tower staff must know the positions and velocities of all aircraft in order to coordinate the traffic safely. The requirement to estimate the positions online makes this a filtering problem.

Because the aircraft are usually well separated in space, we consider one aircraft at a time. The state vector \( x \) would comprise the position and velocity in
three dimensions. From knowledge about the common aircraft types and their behavior, a model of how $x$ changes over time can be devised. Typically this model appears in the form of a difference equation that describes how $x_{k+1}$, the state at time instance $k + 1$, depends on $x_k$. Of course, the precise thrust and steering actions of the pilot are unknown to the airport staff. This uncertainty is included via a random input $v_k$, called process noise, in the state difference equation. Unfortunately, $x_k$ cannot be measured directly. However, the airport radar measures the range, the bearing and elevation angles, and the range rate of the aircraft at discrete time instances $k$. These observations make up a measurement vector $y_k$. Its relation to $x_k$ is described by a measurement equation. Measurement errors are modeled using a measurement noise vector $e_k$. In the Bayesian framework, the noise $v_k$ and $e_k$ and the initial state $x_0$ (when the aircraft is first registered by the tower) are modeled as random variables with a known probability distribution. A common and often well suited choice is the Gaussian distribution.

Now, exact Bayesian filtering amounts to sequentially updating the conditional probability density function $p(x_k | y_{1:k})$, which describes all the knowledge about $x_k$ given $y_{1:k} = \{y_1, ..., y_k\}$ in a probabilistic manner. Less mathematical and more common in air traffic control is that $x_k$ is described with an estimate $\hat{x}_{k|k}$ and a covariance matrix $P_{k|k}$. The latter describes the uncertainty in the estimate. Typically, Kalman filters or algorithms based on Kalman filters are used to compute $\hat{x}_{k|k}$ and $P_{k|k}$ in operational air traffic control systems.

Already this example includes some of the mathematical challenges that impede exact Bayesian filtering and call for approximations. Although the state evolution can be described well by a linear difference equation, the radar measurements $y_k$ are nonlinearly related to $x_k$. As a consequence, the conditional distribution of $x_k$ cannot be Gaussian, even if the noise is Gaussian. Non-Gaussian densities are more difficult to represent in a computer, compared to a Gaussian distribution which is fully described by its mean value and covariance matrix. However, the nonlinearities in aircraft tracking problems are mild and Kalman filters can be applied using linearization or sampling based solutions.

A conceptual similarity between tracking aircraft and object tracking for autonomous cars can be seen. The latter challenge is, however, more difficult because the sensor platform moves. Furthermore, the variability in object behavior is larger and the sensors on the car (cameras, radar, lidar) entail more diverse measurement errors, for example, in different weather conditions.

Papers A and D discuss Kalman filters for nonlinear models and their application for aircraft tracking in great detail.

### A drone tracking scenario

The next example is related to, but more challenging than the filtering in air traffic control. Figure 1.2 illustrates the scenario. We consider a confined area that is observed by a number of cameras, for example, a yard of some industry or government building. Fences and walls may keep out intruders on foot. However, a drone or unmanned aerial vehicle (UAV) is more difficult to prevent from entering. Bayesian state estimation can be used to track intruding drones, for
A drone tracking scenario. A yard is surveyed with four cameras. A highly maneuvering drone enters. Trees occlude the camera views and their moving leaves complicate the image processing.

In comparison to the air traffic control problem, new challenges can be seen. Whereas the aircraft are usually cooperative in air traffic control, an intruding drone acts so as to conceal its actions. Furthermore, a drone is much more agile than a larger aircraft due to its size and actuation. The cameras record images that can be used to detect a drone using image processing and provide angle or position measurements to a tracking filter. The image processing, however, is a difficult task in itself and further complicated by moving trees in the background. Hence, large measurement errors or outliers are likely to occur.

Both the increased maneuverability and the erroneous measurements can be modeled by employing heavy-tailed distributions for the process and measurement noise. The "heavy tails" refer to an increased probability of generating large values. Figure 1.3 illustrates this concept by comparing a Gaussian with a Student's $t$ distribution that exhibits heavy tails. The probability densities appear very similar and so do most of the shown 50 random samples. However, the $t$ distribution generates a few values that are far from 0. In the tracking example, this behavior can be utilized to model sudden accelerations or turns of the drone as well as outliers in the angle or position measurements.

Paper B addresses the problem of filtering in heavy-tailed noise and describes the development of state estimation algorithms based on Student's $t$ distribution.

**Numerical weather prediction**

The third example is a state estimation problem that most readers will relate to: the weather forecast. The associated scientific discipline is known as numerical weather prediction [95] and concerned with large physical models, measurements by satellites or weather stations, and a lot of uncertainty. Although all ingredients to formulate a Bayesian state estimation are present, numerical weather prediction comes with its own challenges. Before delving into them, we note that
1.1 Motivation

Figure 1.3: The probability density functions and 50 samples each of a Gaussian distribution $\mathcal{N}(0,1)$ and a heavy-tailed Student’s $t$ distribution $\text{St}(0,0.8,3)$.

Figure 1.4: Temperature forecast for Europe on January 24, 2017.

we now investigate a prediction instead of a filtering problem. The objective is slightly different as we are interested in $x_{k+l}$ for $l \geq 1$ given the measurements $y_{1:k}$. That is, we try to find the prediction density $p(x_{k+l} | y_{1:k})$. Prediction and filtering, however, go hand in hand and one is needed to solve the other.

Figure 1.4 contains the temperature forecast for Europe on January 24, 2017, as provided by a German news website\(^1\). The temperature is illustrated as one of the atmospheric quantities that make up our weather. Others include air pressure, wind speed, or humidity. A useful weather forecast must provide these quantities for many geographic locations. As a consequence, the state dimension $n$ can be in the order of millions. This impedes the use of Kalman filters that require the processing and storage of $n \times n$ covariance matrices. Furthermore, the models that describe the state evolution are formulated in terms of partial diffe-

\(^1\)http://wetter.tagesschau.de/europawelt/, accessed on Jan 24, 2017.
rential equations, due to the underlying physics. In order to apply discrete-time state estimation as introduced in the air traffic control example, these PDEs must be discretized in space and time. The result is typically a nonlinear black-box simulator that propagates $x_k$ to $x_{k+1}$. Due to the chaotic behavior of weather models, such state transitions are only valid for short sampling times. Moreover, nonlinear measurement relations appear for satellite observations.

An algorithm to address the above difficulties is the ensemble Kalman filter (EnKF), a Monte Carlo sampling implementation of the Kalman filter. The algorithm propagates and adjusts ensembles of $N \ll n$ state realizations instead of $\hat{x}_k|k$ and $P_k|k$ to achieve a reduction in computing complexity. Interestingly, EnKF are seldom discussed outside of the geoscientific literature. Therefore, Paper C discusses the algorithm in a signal processing context, including its challenges and potential ideas that can be beneficial also for other state estimation algorithms.

The search for MH370

The last example is about Bayesian state estimation to obtain a complete picture of the state distribution of $x_k$ given $y_{1:k}$. In that respect it goes beyond the task of finding a “best estimate and its uncertainty” that is common to the previous examples.

On March 8, 2014, Malaysia Airlines Flight 370 (MH370) went missing over the Indian Ocean. A precise crash location could not be determined from the available radar and communication data. Consequently, immediate search and rescue missions were confronted with a too large search area to be successful. Later, all the data were gathered and processed by the Defense Science and Technology (DST) Group, an Australian government agency, to devise the probability density function of the crash location in Figure 1.5. The DST analysis is documented in [31] and uses the Bayesian state estimation framework. In particular, aircraft motion and measurement models were devised and combined with the available data using a particle filter. The algorithm simulates trajectory hypotheses and maintains only those that comply with the data. Unlikely trajectories are discarded systematically. The probability density function in Figure 1.5 is computed from the hypotheses of the final position. Also illustrated is the proposed $10^3$ square kilometer large search area. In comparison, the search area for the successful recovery of Air France flight 447 was a disk of 17000 square kilometers. Although MH370 could not be recovered yet and the search was suspended in January 2017, this example illustrates the usefulness of Bayesian state estimation and the picture it can provide about uncertain states.

The interpretation of Bayesian state estimation as the search for conditional probability density functions is useful for understanding the potential and limitations of different state estimation algorithms. Therefore, we highlight this perspective throughout Part I of this thesis.
1.2 Outline and Contributions

The thesis consists of two parts, with introductory material in Part I and four scientific papers in Part II.

Part I: Background

The first part introduces the theoretical framework for the material in Part II. Chapter 2 defines the Bayesian state estimation problem and Chapter 3 introduces algorithms for filtering and smoothing. The presentation is compact and the focus is simple examples rather than mathematical derivations. Concluding remarks are given in Chapter 4. Furthermore, two appendices give a brief overview of probability density functions and numerical methods for Bayesian state estimation for the scalar case.

A contribution is the illustration of exact prediction, filtering, and smoothing densities and their approximations as provided by different algorithms for a number of instructive scalar examples. Such a visual presentation is seldom seen in the available literature and can facilitate a better understanding of the underlying mathematical concepts.

Parts of the material in Part I have been published in

Part II: Publications

The second part consists of four articles that discuss different algorithms for Bayesian state estimation in greater detail. Being the lead author of the included papers, the author of this thesis has contributed the majority of text, derivations, and simulations. Of course, all of the papers evolved from a collaborative effort and discussions about the technical content, mathematical derivations, and the presentation of the material. Notable contributions by the co-authors include the discussion of exact filtering in $t$ noise based on derivations by T. Ardeshiri, described in Paper B, and the simulations of Paper A which were performed in close collaboration with G. Hendeby.

Paper A

Paper A is an edited version of


The paper is a tutorial on approximate Kalman filtering for nonlinear state-space models. In its main focus are different sigma point Kalman filters and the methods to compute mean values and covariance matrices (moments) of nonlinearly transformed random vectors that they utilize. The in-depth presentation covers the basics of Bayesian filtering; Kalman filter approaches including extended KF (EKF), unscented KF (UKF), divided-difference KF, and numerical integration or cubature KF (CKF); structural properties of moment computation problems; and sigma point methods based on density approximation, function approximation, and numerical integration, as well as their relations. An extensive simulation section complements the theoretical discussion.

The contributions include

- the presentation of nonlinear Kalman filters in a simple common framework which highlights the computation of mean values and covariance matrices (the first two moments) as the central challenge. This unifying perspective facilitates a quick and critical assessment of the ever increasing number of KF variants in the literature.

- the introduction of the moment computation problem and a detailed analysis of its theoretical properties, including structural reformulations and “tricks of the trade” that are exploited in sigma point KF. Results for the Gaussian distribution are extended to the wider class of elliptically contoured distributions.

- a discussion of analytical solutions to the moment computation problem based on Taylor series. A compact formulation of degree-$d$ Taylor polynomials as high-dimensional matrix-vector products is introduced, and reveals moment expressions similar to the linear case.
1.2 Outline and Contributions

- a survey of sigma point methods to approximately solve the moment computation problem using $N$ weighted samples, the sigma points. Discussed are Monte Carlo integration, the unscented transformation (UT), numerical integration rules, and interpolation approaches. Similarities, advantages, and challenges of the respective methods are highlighted.

- an extensive simulation section in which the methods are compared.

The tutorial is based on the licentiate thesis


Parts of the material were furthermore presented in

M. Roth and F. Gustafsson. An efficient implementation of the second order extended Kalman filter. In 14th International Conference on Information Fusion (FUSION), Chicago, USA, July 2011.

**Paper B**

Paper B is an edited version of


Motivated by upcoming challenges in tracking problems with highly maneuvering targets and outlier-corrupted measurements, this paper investigates the use of Student’s $t$ distribution to devise simple and robust filtering and smoothing algorithms.

The contributions include

- a discussion of Student’s $t$ distribution and important differences to the Gaussian distribution, for example, the role of the scale/covariance matrix.

- a discussion of elliptically contoured distributions and results that they share with the Gaussian, for example, the conditional mean and scale matrix that relate to the Kalman filter measurement update.

- a discussion of an exact filtering step with Student’s $t$ noise, the lack of a closed form solution, and the re-appearing Kalman filter equations.

- simple filtering and smoothing algorithms that have a clear relation to the Kalman filter but with a measurement-dependent update of $P_{jk}$, derived from joint $t$ density assumptions.

- a simulation section which demonstrates the potential of the devised algorithms.
Parts of the material were published in

M. Roth, E. Özkan, and F. Gustafsson. A Student’s t filter for heavy
tailed process and measurement noise. In 38th International Confe-
rence on Acoustics, Speech, and Signal Processing (ICASSP), Vancou-
ver, Canada, May 2013,

the author’s licentiate thesis

M. Roth. Kalman Filters for Nonlinear Systems and Heavy-Tailed
Noise. Licentiate thesis, Linköping University, Linköping, Sweden,
Sept. 2013,

and a technical report on Student’s t distribution

M. Roth. On the multivariate t distribution. Technical Report LiTH-
ISY-R-3059, Automatic Control, Linköping University, Linköping, Swe-
den, Apr. 2013.

Paper C

Paper C is an edited version of

M. Roth, G. Hendeby, C. Fritsche, and F. Gustafsson. The ensemble

It investigates the ensemble Kalman filter (EnKF), which has been developed to
approach geoscientific estimation problems with state dimensions $n$ in the order
of millions. In contrast to the Kalman filter that is impeded by the requirement
to store and process $n \times n$ covariance matrices, the EnKF achieves a reduced com-
putational complexity by propagating an ensemble of $N < n$ random states. Inter-
estingly, the EnKF has received little attention in the Bayesian filtering literature
compared to its impact in differen
t geoscientific disciplines. Nevertheless, there
are good reasons to cast a closer look at the EnKF because of its scalability, its
applicability in nonlinear and non-Gaussian problems, and the EnKF localiza-
tion techniques to systematically update only parts of the state vector with each
measurement update.

The contributions include

• a survey of the extensive EnKF literature that highlights the relevant EnKF
publications.

• a compact derivation of the EnKF as a Monte Carlo sampling implementa-
tion of the Kalman filter, without the often encountered geoscientific termin-
ology of the EnKF literature.

• a discussion of algorithmic properties and challenges for small ensembles
($N < n$) that follow from the random sampling and the coupling in the en-
semble.
1.2 Outline and Contributions

• a discussion of EnKF extensions to reduce adverse effects via square root implementation and localization techniques.

• a simulation section that illustrates the theoretical properties on geoscientific and signal processing benchmark problems.

Parts of the material were published in


Paper D

Paper D is an edited version of

M. Roth, G. Hendeby, and F. Gustafsson. EKF/UKF maneuvering target tracking using coordinated turn models with polar/Cartesian velocity. In 17th International Conference on Information Fusion (FUSION), Salamanca, Spain, July 2014.

The paper is more application oriented than the others. It discusses the role of state coordinate frames on the performance of different Kalman filter variants in an aircraft tracking problem.

The contributions include

• a discussion of different stochastic discrete-time coordinated turn models derived from continuous-time particle motion.

• a simulation study in which extended Kalman filters (first and second order EKF) and unscented Kalman filters (UKF) are compared on a maneuvering target tracking benchmark problem.

• a detailed analysis of filter performance with respect to the chosen model and the noise parameters.

• the conclusion that models with polar velocity appear favorable in comparison to those with Cartesian velocity.

Related publications

The following publication on particle filters and smoothers is related to this thesis, but not discussed further.

Bayesian State Estimation

This chapter provides a compact introduction to Bayesian state estimation. An unconventional but instructive view is provided by showing the relation to basic operations on probability density function. Furthermore, the elegant yet abstract mathematical results are clarified by plotting the posterior densities for scalar examples. Such a visual presentation of exact Bayesian state estimation densities is seldom found in the available literature.

We assume that the information about the state $x_k$ given a sequence of measurements $y_{1:L} = \{y_1, \ldots, y_L\}$ can be described by a conditional probability density function $p(x_k | y_{1:L})$ that contains all the information about $x_k$ in a probabilistic manner. The computation of the posterior density $p(x_k | y_{1:L})$ for $k = 0, \ldots, L$ is the task of Bayesian state estimation.

We consider only one state $x_k$ at a time, which is called marginal state estimation. In contrast, the objective of joint state estimation is to find the density $p(x_{0:k} | y_{1:L})$ of a sequence of states $x_{0:k}$.

Depending on the temporal relations of $k$ and the measurement horizon $L$, three different estimation problems can be distinguished. The case $L < k$ is called prediction, $L = k$ filtering, and $L > k$ smoothing or retrodiction.

In general it is not feasible to compute $p(x_k | y_{1:L})$ analytically because the posterior cannot be described by a finite number of parameters. This impedes its exact representation in a digital computer. Furthermore, the appearing integral equations are difficult to solve and do not yield closed form expressions except for a few special cases. However, efficient numerical approximations exist for scalar problems. Such solutions are described in Appendix B and utilized to illustrate the posterior densities for the examples in this chapter.

The literature on Bayesian state estimation is rich and dates back to [74]. A classic yet informative monograph is [87]. Timely presentations with modern

---

1 Posterior as opposed to prior; after including the observations $y_{1:L}$. 

Bayesian state estimation algorithms are provided in [64] or [161], for example. Also the present author’s licentiate thesis [138] and the papers of Part II contain introductions.

## 2.1 Stochastic State-Space Models

Bayesian state estimation is based on stochastic state-space models. These provide a flexible and powerful framework for describing real world phenomena. In a general form, a state transition and a measurement equation

\[
x_{k+1} = f(x_k, v_k), \quad (2.1a)
\]

\[
y_k = h(x_k, e_k), \quad (2.1b)
\]

describe the evolution of the \(n\)-dimensional state \(x_k \in \mathcal{X}\) to \(x_{k+1} \in \mathcal{X}\) and its relation to the \(m\)-dimensional measurement \(y_k \in \mathcal{Y}\), respectively. The inputs \(v_k \in \mathcal{V}\) and \(e_k \in \mathcal{E}\) are called process and measurement noise, respectively. The general form of (2.1) can be further specified in many cases. Common are nonlinear models with additive noise \(v_k\) and \(e_k\) and, of course, the linear state-space model

\[
x_{k+1} = Fx_k + Gv_k, \quad (2.2a)
\]

\[
y_k = Hx_k + e_k, \quad (2.2b)
\]

which is the basis of the Kalman filter.

In order to make (2.1) a stochastic state-space model, a probabilistic description of the noise terms \(v_k\), \(e_k\), and the initial state \(x_0\) must be provided for all relevant \(k\). The simplest and most common assumption is that \(v_k\) and \(e_k\) are white noise sequences that are mutually independent and independent of \(x_0\). Then, their joint probability density function can be factored as

\[
p(x_0, v_0: k-1, e_1: k) = p(x_0) \prod_{l=1}^{k} p(v_{l-1}) p(e_l). \quad (2.3)
\]

Some cases of practical relevance do not allow such a factorization. For example, \(v_k\) and \(e_k\) are no longer independent if a feedback controller is employed to influence \(x_{k+1}\). Often, only the moments

\[
E(x_0) = \hat{x}_0, \quad E(v_k) = 0, \quad E(e_k) = 0,
\]

\[
cov(x_0) = P_0, \quad cov(v_k) = Q, \quad cov(e_k) = R \quad (2.4)
\]

are specified. In the Gaussian case, these parameters provide a full description of \(p(x_0, v_0: k-1, e_1: k)\). Because of the randomness in \(v_k\), \(e_k\), and \(x_0\), also \(x_k\) and \(y_k\) are random variables. This justifies the search for a posterior density as the main task in Bayesian state estimation.

In general, all above parameters and functions can be time-varying. An extra time index on, say, \(f\), \(F\), or \(Q\), is omitted for simplicity.

In the independent white noise case of (2.3), (2.1) is a hidden Markov model for which \(x_{k+1}\) is independent of \(x_{0:k-1}\) given \(x_k\), among a number of other conditional independence properties. Such models are often described by a transition
2.2 Bayesian State Estimation via Basic Density Manipulations

The here presented view reduces Bayesian state estimation to the fundamental operations of transformation, marginalization, and conditioning of probability density functions (Appendix A). Such considerations date back to [74] and are helpful to devise algorithms based on intermediate approximations. Examples include the nonlinear Kalman filters of Paper A or the Student's $t$ filter of Paper B, which assume intermediate Gaussian and Student's $t$ densities to obtain convenient recursions. Furthermore, the shown procedure also accommodates models that lack expressions for the transition density and likelihood function of (2.5).

**Prediction**

Prediction is the simplest problem with regard to the involved operations. However, it is the most difficult problem in terms of achievable performance.

We begin with one-step-ahead prediction, where the objective is to compute $p(x_{k+1} | y_{1:k})$ given a filtering posterior $p(x_k | y_{1:k})$. For independent white noise,

$$p(x_k, v_k | y_{1:k}) = p(x_k | y_{1:k}) p(v_k)$$

(2.6)

is the joint density of the random inputs $x_k$ and $v_k$ that create $x_{k+1}$ in the state difference equation (2.1a). Other densities $p(x_k, v_k | y_{1:k})$ are conceivable for known probabilistic relations between $x_k$ and $v_k$. Now, $x_{k+1}$ is a transformed random variable and its density can be computed under certain requirements on the state transition function. For example, if there exists an auxiliary variable $\xi_k \in \Xi$ with a one-to-one mapping $\psi : \mathcal{X} \times \mathcal{V} \rightarrow \mathcal{X} \times \Xi$, then the joint density $p(x_{k+1}, \xi_k | y_{1:k})$ can be obtained from a transformation theorem [67] of Appendix A.

The sought after prediction density is then obtained by marginalization

$$p(x_{k+1} | y_{1:k}) = \int p(x_{k+1}, \xi_k | y_{1:k}) d\xi_k.$$  

(2.7)

Hence, prediction requires the operations of transformation and marginalization. Prediction several time steps ahead can be approached in the same way. If there is correlation between, say $v_k$ and $v_{k-1}$, more variables need to be kept in the transformation to not neglect any probabilistic dependencies. In fact, the order of marginalization and transformation can be exchanged in any way if carried out exactly. Of course, this general procedure can often be simplified. For...
example, in linear Gaussian models (2.2) with (2.4), prediction can be achieved by transforming the mean and covariance matrices according to the rules for the Gaussian distribution.

**Filtering**

The filtering problem amounts to including new observations $y_k$. We assume that a prediction density $p(x_k | y_{1:k-1})$ is available. For independent measurement noise, we must transform

$$p(x_k, e_k | y_{1:k-1}) = p(x_k | y_{1:k-1}) p(e_k).$$

(2.8)

Other $p(x_k, e_k | y_{1:k-1})$ are conceivable for dependent $x_k$ and $e_k$. From a transformation with the measurement function (2.1b) follows the joint prediction density $p(x_k, y_k | y_{1:k-1})$. Again, the introduction of auxiliary variables and marginalization might be required.

The filtering result is obtained by conditioning on $y_k$ via

$$p(x_k | y_{1:k}) = \frac{p(x_k, y_k | y_{1:k-1})}{p(y_k | y_{1:k-1})},$$

(2.9)

with a normalization constant

$$p(y_k | y_{1:k-1}) = \int p(x_k, y_k | y_{1:k-1}) dx_k.$$  

(2.10)

Some distributions provide closed-form expressions for conditional densities, for example, the Gaussian or Student’s $t$ distribution. Hence, the linear Gaussian case (2.2) with (2.4) facilitates exact filtering via the conditioning results of the Gaussian distribution. Furthermore, density approximations can be employed to utilize conditioning results in approximate filtering algorithms, as shown in Papers A and B.

Again, attention is required if there are probabilistic dependencies among the noise variables. For example, if $v_k$ and $e_k$ are correlated in the feedback control case, then also $e_k$ must be included in (2.6) and the transformation and marginalization operations before conditioning.

**Smoothing**

Smoothing problems do not require any operations beyond the transformation, marginalization, and conditioning of the filtering problem. However, the challenge becomes significantly more difficult because more variables must be involved, even for the white noise case. Also, the sequential nature of the prediction and filtering problems is less apparent in the smoothing problem.

Without the requirements of online processing, smoothing is simplest viewed as a batch problem. From the joint density (2.3) follows the joint density of
p(x_{0:L}, y_{1:L}) via transformation, perhaps with the help of marginalization of auxiliary variables. Then, conditioning can be applied to obtain the joint smoothing density p(x_{0:L} | y_{1:L}). Finally,

\[ p(x_k | y_{1:L}) = \int p(x_{0:L} | y_{1:L}) \, dx_{1:k-1} \, dx_{k+1:L} \]  

(2.11)

yields the marginal smoothing density.

Again, the order of operations is irrelevant if carried out exactly. Of course, it is desirable to obtain sequential solutions that can be implemented as a forward and a backward iteration that adjusts some parameters. For linear Gaussian models the Rauch-Tung-Striebel smoother provides such a solution.

### 2.3 Elegant Sequential Solutions

Although insightful and easily adjustable to probabilistic dependencies in the noise, the density viewpoint of the previous section has some drawbacks. For example, the transformation of random variables can be complicated to put down in functional form. In contrast, elegant recursive solutions for prediction and filtering can be devised for Markov models with (2.5). Good references for their derivation are [64, 161].

From the Markov property of the model follow some conditional independence properties

\[
\begin{align*}
p(x_{k+1} | x_k, y_{1:k}) &= p(x_{k+1} | x_k), \quad (2.12a) \\
p(y_k | x_k, y_{1:k-1}) &= p(y_k | x_k), \quad (2.12b) \\
p(x_k | x_{k+1}, y_{1:L}) &= p(x_k | x_{k+1}, y_{1:k}), \quad L > k, \quad (2.12c)
\end{align*}
\]

that are utilized to derive the following posterior densities.

#### Prediction and filtering

An early reference for the following iteration is [87]. Using the transition density and likelihood of (2.5), the one-step-ahead prediction and filtering densities are given by

\[
\begin{align*}
p(x_{k+1} | y_{1:k}) &= \int p(x_{k+1} | x_k) p(x_k | y_{1:k}) \, dx_k, \quad (2.13a) \\
p(x_k | y_{1:k}) &= \frac{p(y_k | x_k) p(x_k | y_{1:k-1})}{p(y_k | y_{1:k-1})}, \quad (2.13b)
\end{align*}
\]

where

\[
p(y_k | y_{1:k-1}) = \int p(y_k | x_k) p(x_k | y_{1:k-1}) \, dx_k \tag{2.13c}
\]

is a normalization constant.

For the linear model (2.2) with Gaussian noise, the above densities are Gaussian. Their parameters can be obtained by the Kalman filter [161].
20 2 Bayesian State Estimation

Smoothing

A backward recursion for smoothing based on the results (2.13) was suggested in [97]. For \( L > k \), the prediction and filtering results are processed in

\[
p(x_k | y_{1:L}) = \int p(x_{k+1} | x_k) p(x_{k+1} | y_{1:L}) \, dx_{k+1} / p(x_{k+1} | y_{1:k}). \tag{2.14}
\]

For the linear model (2.2) with Gaussian noise, the smoothing density is Gaussian and its parameters are given by the Rauch-Tung-Striebel smoother [161].

2.4 Instructive Scalar Examples

This section illustrates Bayesian state estimation via numerically computed posterior densities. The utilized point mass algorithms and a novel filter for intractable likelihood functions are described in Appendix B.

2.4.1 “The particle filter example”

The first example is a nonlinear growth model that is often used as benchmark problem for particle filters [11, 60, 97]. The state evolution and measurement are governed by

\[
x_{k+1} = \frac{x_k}{2} + 25 \frac{x_k}{1 + x_k^2} + 8 \cos(1.2(k + 1)) + v_k, \tag{2.15a}
\]
\[
y_k = \frac{x_k^2}{20} + e_k, \tag{2.15b}
\]

with independent Gaussian noise \( v_k \sim \mathcal{N}(0,10) \), \( e_k \sim \mathcal{N}(0,1) \), and the initial state \( x_0 \sim \mathcal{N}(0,1) \). The state transition function includes a deterministic input \( 8 \cos(1.2(k + 1)) \) that can be interpreted as a time varying function \( f \) in (2.1a) or as a non-zero mean value of the process noise \( v_k \).

The one-step-ahead prediction, filtering, and smoothing posterior densities for the time instance \( k = 10 \) of one realization are given in Figure 2.1. Also illustrated is an inverse mapping \( \pm \sqrt{20y_k} \) of the measurement function under the assumption that \( e_k = 0 \).

Some interesting observations can be made. The prediction and filtering densities are bimodal and one of the filtering peaks is close to the mapping of \( y_k \). The appearance of several modes can be related to the \( x_k^2 \) in the measurement equation and the ambiguity it entails. With respect to potential point estimates, the mean value would entail a larger error than the maximum a posteriori (MAP) estimate. In the smoothing density only the mode closer to the true state persists. Figure 2.2 shows similar results for a number of consecutive time instances. For \( k = 0 \), the initial filtering density is shown. The first measurement arrives at \( k = 1 \). Unlike the illustration of Figure 2.1, the densities appear unimodal at many times. Also, the nonlinearity in the state transition does not appear to introduce extra modes. That is, a unimodal filtering density results in a unimodal
2.4 Instructive Scalar Examples

Figure 2.1: Prediction (blue), filtering (black), and smoothing (orange) densities for one time instance of the “particle filter example”. The black vertical line marks the true state and the red lines a mapping of the measurement back into the state space.

prediction for the next time step. The prediction densities appear always broad because of the additive noise with large variance. The smoothing is performed with a backward pass that is initialized with the filtering density at \( k = L = 15 \). Although the smoothing often yields narrower densities in comparison to the filtering results, there are instances without notable difference \((k = 5, 10, 11)\). The true states are always well covered by the filtering and smoothing densities.

2.4.2 A Student’s \( t \) random walk observed in noise

The second example is a random walk that is observed in noise. The state transition and measurement equation are

\[
\begin{align*}
    x_{k+1} &= x_k + v_k, \quad (2.16a) \\
    y_k &= x_k + e_k. \quad (2.16b)
\end{align*}
\]

The noise and initial state are Student’s \( t \) distributed with \( v_k \sim \text{St}(0, 1, 3), e_k \sim \text{St}(0, 1, 3), \) and \( x_0 \sim \text{St}(0, 1, 3) \). Because of the low degrees of freedom 3, the distributions are heavy-tailed and produce large values occasionally. Despite the simplest possible model, compact analytical expressions for the posterior densities cannot be derived. A discussion of an exact filtering step that highlights the challenges is provided in Paper B.

Figure 2.3 shows the one-step-ahead prediction, filtering, and smoothing results for several consecutive time instances of a single realization. Again, the first measurement arrives at \( k = 1 \) and the smoothing is carried out for \( L = 15 \). The densities are unimodal most of the times. Caused by a measurement outlier, however, two modes appear at \( k = 9 \). The bimodality can be traced to the multiplication of \( p(x_k | y_{1:k-1}) \) with the likelihood \( p(y_k | x_k) \) in (2.13b) and has been described as “moment of indecision” [2]. Another outlier at \( k = 12 \) yields a heavily skewed filtering posterior. The smoothing result appear unimodal again and close to the prediction for both \( k = 9 \) and \( k = 12 \).

The fact that the densities in Figure 2.3 are unimodal at most times motivates
Figure 2.2: Prediction (blue), filtering (black), and smoothing (orange) densities for several consecutive time instances of the “particle filter example”. The green dots mark the true states.
Figure 2.3: Prediction (blue), filtering (black), and smoothing (orange) densities for several consecutive time instances of the Student’s t random walk. The green and red dots mark the true states and measurements, respectively.
for the development of unimodal filters in the spirit of the Kalman filter, which is the subject of Paper B.

### 2.4.3 Saturated measurements

The last example considers the model

\[
\begin{align*}
    x_{k+1} &= 0.7x_k + v_k, \\
    y_k &= \text{sat}(x_k + e_k),
\end{align*}
\]  

(2.17a) \hspace{1cm} (2.17b)

with \(v_k \sim \mathcal{N}(0, 1), e_k \sim \mathcal{N}(0, 0.5)\), and \(x_0 \sim \mathcal{N}(0, 0.1)\). The sat function truncates its input beyond \(-1.5\) and \(1.5\). Hence, the measurement model reflects the realistic case of a sensor with a limited range, for example, a saturating accelerometer or microphone.

Filtering using point mass approaches is difficult for (2.17) because the likelihood \(p(y_k | x_k)\) is difficult to obtain for the many-to-one saturation function. However, the filter of Appendix B.3 can be applied easily. The resulting densities for one realization are shown in Figure 2.4. The measurements are saturated at several times. The corresponding filtering and prediction densities exhibit significant probability mass beyond the measurement range.
Figure 2.4: Prediction (blue), filtering (black), and smoothing (orange) densities for several consecutive time instances of the saturated measurements example. The green and red dots mark the true states and measurements, respectively.
This chapter presents algorithms for approximately solving Bayesian filtering and smoothing problems.

### 3.1 Particle Filters and Smoothers

Particle filters and smoothers (PF and PS) are not the central subject of this thesis. The reason to present them first is their asymptotically exact approximation of the Bayesian state estimation densities of Section 2.3 as the number of particles tends to infinity.

Early Monte Carlo filters date back to [70] but the first PF [60], with the essential resampling step, was not presented until the early 1990s. Overview articles on PF [11, 63], PF and PS [34, 57], as well as recent text books [64, 136, 161] provide detailed discussions of particle algorithms. Therefore, only the basic PF and PS are listed here.

The PF approximation to the filtering density (2.13) is based on importance sampling. That is, $N$ particles $x_k^{(i)}$ and weights $w_k^{(i)}$ are propagated such that expected values with respect to the posterior densities appear as finite sums,

$$
\int g(x_k)p(x_k | y_{1:k}) dx_k \approx \sum_{i=1}^{N} w_k^{(i)} g(x_k^{(i)}),
$$

(3.1)

for some function $g(x_k)$ with finite expected value. Hence, the particle approximation can be seen as Dirac mixture density

$$
p(x_k | y_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(x_k - x_k^{(i)}).
$$

(3.2)

The particles $x_k^{(i)}$ are drawn from proposal distributions that can depend on previous particles as well as the entire measurement history $y_{1:k}$. A particularly simple
choice [60] samples from \( p(x_{k+1} | x^{(i)}_k) \), which amounts to a time update
\[
x^{(i)}_{k+1} = f(x^{(i)}_k, v^{(i)}_k), \quad i = 1, \ldots, N.
\]
Here, \( v^{(i)}_k \) is an independent process noise realization.

The measurement update processes \( y_k \) to compute the weights. For the above proposal, the weight update is given by
\[
w^{(i)}_k \propto w^{(i)}_{k-1} p(y_k | x^{(i)}_k), \quad i = 1, \ldots, N.
\]
Subsequent normalization ensures that \( \sum_{i=1}^N w^{(i)}_k = 1 \).

Recursive execution of (3.3) and (3.4) does not yield a functioning filter since the measurements have no effect on the particle locations. Consequently, particles explore regions where the true posterior has no probability mass. This yields a degeneracy problem in which all weights except one become 0. As a remedy, a resampling step was introduced in [60]. Particles are randomly duplicated or discarded according to their weights, to yield a new set of equally weighted particles \( w^{(i)}_k = 1/N \) for the next time update.

Particle smoothing can be achieved in different ways [34, 57, 161]. The algorithm of [34] maintains the particles of a previously run PF and computes the smoothing weights
\[
w^{(i)}_{k|L} = \sum_{j=1}^N w^{(j)}_{k+1} \frac{w^{(i)}_k p(x^{(j)}_{k+1} | x^{(i)}_k)}{\sum_{l=1}^N w^{(l)}_k p(x^{(j)}_{k+1} | x^{(l)}_k)},
\]
in a backward pass.

We already mentioned the asymptotically exact approximation of the Bayesian state estimation densities as \( N \to \infty \) [33]. Furthermore, basic PF and PS are simple to implement and applicable to arbitrary nonlinear and non-Gaussian state-space models. However, PF and PS have some limitations too. Specifically, the basic variants work only well for low-dimensional problems. The reason lies in the difficulty to generate meaningful samples in high-dimensional spaces, and the resulting breakdown of importance sampling. Furthermore, the inherent random sampling can be problematic for too small \( N \) when the noise exhibits rare events such as outliers. PF and PS always entail some Monte Carlo variations which should be assessed thoroughly for the problem at hand. In comparison to Kalman filters, the computational load is large. The complexity increases with \( N \) in the PF and \( N^2 \) in the presented PS.

However, there is vivid development on PF and PS. For example, marginalized PF [34, 63, 73, 156] and PS [108] exploit linear sub-structures of the model to combine KF approaches with particle methods. This facilitates their application in higher-dimensional problems.

The performance of the PF and PS (3.3)–(3.5) with \( N = 5000 \) is shown for the “particle filter example” of Section 2.4.1. Figure 3.1 illustrates the results. The continuous densities are obtained by kernel density estimation (Appendix B.1) with the weights and particles. Except for some variations due to the random
Figure 3.1: Particle filter and smoother prediction (blue), filtering (black), and smoothing (orange) densities for several consecutive time instances of the “particle filter example”. The green dots mark the true states.
sampling and a smoother appearance due to the kernel density estimation, the PF and PS resemble the exact Bayesian state estimation densities in Figure 2.2 very well.

3.2 The Kalman Filter and the Rauch-Tung-Striebel Smoother

The Kalman filter (KF) is known as the workhorse of estimation [13] and, indeed, forms the basis of all discussed algorithms in Part II of this thesis. The seminal paper [94] appeared just in time to solve an important navigation problem in the Apollo program [62], which largely contributed to the success of the KF and led to increased research efforts for state-space models in general. Interestingly, the application in the Apollo program already required an approximation of the KF theory via a linearization approach that became known as the extended KF.

The literature on the KF is rich. Classic references include [3, 93]. An extensive discussion is also found in the author’s licentiate thesis [146]. Therefore, only a brief coverage is provided here.

The KF can be derived as an optimal filter for linear systems (2.2) with known noise statistics (2.4). The optimality refers to the KF being the best unbiased linear filter in the minimum variance sense [3]. From a Bayesian perspective, the KF solves the filtering recursion (2.13) in linear systems with Gaussian noise [161]. The one-step-ahead prediction and filtering densities are Gaussian in this case, and the KF provides their mean vectors and covariance matrices.

The algorithm can be split into a time and a measurement update, respectively, which adjust state estimates and their covariance matrices. The time update provides a predicted estimate and its covariance

\[
\dot{x}_{k+1|k} = F\dot{x}_{k|k},
\]

\[
P_{k+1|k} = FP_{k|k}F^T + Q.
\]

In preparation for the measurement update, a prediction of the output \(\hat{y}_{k|k-1}\) and its covariance \(S_k\), as well as the cross-covariance \(M_k\) between the state and output are computed. Furthermore, it is convenient to introduce the Kalman gain \(K_k\). The expressions are

\[
\hat{y}_{k|k-1} = H\dot{x}_{k|k-1},
\]

\[
M_k = P_{k|k-1}H^T,
\]

\[
S_k = HP_{k|k-1}H^T + R,
\]

\[
K_k = M_kS_k^{-1}.
\]

The actual measurement update combines the prediction results with the measurement \(y_k\) to obtain the filtered estimate and covariance

\[
\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - \hat{y}_{k|k-1}),
\]

\[
P_{k|k} = P_{k|k-1} - K_kS_kK_k^T.
\]
A smoothing extension to the KF is given by the Rauch-Tung-Striebel (RTS) algorithm [135]. Using the smoother gain

\[ G_k = P_{k|k} F^T P_{k+1|k}^{-1}, \]

the KF results are processed in a backward iteration to yield smoothed estimates and covariance matrices

\[
\hat{x}_{k|L} = \hat{x}_{k|k} + G_k (\hat{x}_{k+1|L} - \hat{x}_{k+1|k}),
\]

\[
P_{k|L} = P_{k|k} + G_k (P_{k+1|L} - P_{k+1|k}) G_k^T.
\]

The KF and RTS smoother are powerful tools that have proven their success in many engineering applications. However, there are limits to their performance. We show this by revisiting the Student’s \( t \) random walk of Section 2.4.2. The true posterior densities in Figure 2.3 serves as reference solution. We run the KF and RTS smoother with the correct variance parameters of the \( t \) noise distributions, which yields the optimal linear filter and smoother. The estimates and variances of (3.6), (3.8), and (3.10) are computed and illustrated as Gaussian densities in Figure 3.2.

A first thing to note is that the Gaussian bell for the initial filtering density at \( k = 0 \) (Figure 3.2) appears wider than the exact initial filtering density (Figure 2.3). This is because the KF employs the true variance of \( p(x_0) \), which is large due to the heavy tails of the \( t \) distribution. For well-behaved measurements, the KF covers the true state well. However, at \( k = 9 \) and \( k = 12 \) outliers appear. Whereas the true filtering density becomes broader (Figure 2.3), the KF shifts its probability mass towards the measurement and no longer covers the true state well. In contrast to the exact smoothing results (Figure 2.3), this behavior is not corrected by the RTS backward pass (Figure 3.2).

### 3.3 Nonlinear Kalman Filters

Nonlinear KF apply the measurement update of the KF (3.8) for nonlinear models (2.1). Many different variants exists, ranging from the linearization based EKF [3], interpolation approaches [86, 128], variants based on numerical integration rules [8, 86], to “unscented” KF [90, 92]. In contrast to the analytical EKF, most nonlinear KF employ some form of sampling which led to the term sigma point KF [173] in the literature. Also nonlinear variants of the Rauch-Tung-Striebel smoother have been developed, but these are not discussed in this thesis. A comprehensive overview is given in [161]. Nonlinear KF are the main subject of Paper A which discusses the different variants in great detail.

The employed measurement update (3.8) can be motivated from an intermediate Gaussian approximation

\[
p(x_k, y_k | y_{1:k-1}) \approx \mathcal{N}
\left(
\begin{bmatrix}
  x_k \\
  y_k \\
  \hat{x}_{k|k-1} \\
  \hat{y}_{k|k-1}
\end{bmatrix}
\bigg|
\begin{bmatrix}
  P_{k|k-1} \\
  M_k \\
  S_k
\end{bmatrix}
\right).
\]

(3.11)
Figure 3.2: Kalman filter prediction (blue), filtering (black), and RTS smoothing (orange) densities for several consecutive time instances of the Student’s random walk. The green and red dots mark the true states and measurements, respectively.
The parameters of (3.11) are computed as mean values and covariance matrices of nonlinearly transformed (Gaussian) random variables. The employed methods to compute these moments distinguish the respective filters. Sigma point KF, for instance, employ sampling based schemes that can be motivated from function approximation or interpolation [86, 128], numerical integration [8, 86], or density approximation perspectives [90, 92].

Common to all nonlinear KF is the representation of predicted and filtered states by a mean value and a covariance, often interpreted as the parameters of a Gaussian distribution. This simple representation is convenient, but also has some limitations. For example, multimodal densities are difficult to represent by a single Gaussian component.

To give a glimpse of different nonlinear KF variants in action, we revisit the “particle filter example” of Section 2.4.1. The exact prediction and filtering densities of Figure 2.2 are compared to the Gaussian approximations of two nonlinear KF.

First, we investigate the interpolation based divided-difference KF [86, 128] which can be seen as a variant of the EKF, but with numerically computed derivatives in the linearization steps. The obtained prediction and filtering densities are shown in Figure 3.3. Already after processing the first measurement at \( k = 1 \), the filtering result completely fails to represent the exact (bimodal) density (Figure 2.2). Similar failures occur at several time steps and lead to large estimation errors. Apparently, the inherent linear approximation cannot represent the nonlinear state transition and measurement equations (2.15) adequately.

Second, we investigate a nonlinear KF that uses Monte Carlo integration with \( N = 1000 \) samples to compute the moments of (3.11). The increased accuracy in the moment computations comes at the price of an increased computational load. The prediction and filtering results are shown in Figure 3.4 and appear very broad in comparison to the divided-difference KF (Figure 3.3). The Monte Carlo KF provides larger variance parameters. Still, the results do not reflect the exact densities well (Figure 2.2). Apparently, more accurate moment computation schemes do not necessarily lead to a working KF. More advanced methods are required to adequately represent multimodal densities in severely nonlinear and non-Gaussian filtering problems.

It should be noted that nonlinear KF are successfully applied in many state estimation problems. However, it is important to be aware of their limitations.

### 3.4 Student’s t Filters and Smoothers

Many real world phenomena cannot be modeled by a single Gaussian distribution, for example, measurement outliers or jumps in the process noise. Hence, challenges can be expected for the application of the KF and the RTS smoother in models with heavy-tailed noise. A potential alternative to the Gaussian is Student’s t distribution, which can exhibit heavy tails for low degrees of freedom.

Paper B therefore describes Student’s t distribution and the challenges that appear in linear models (2.2) with t noise. Exact filtering does not yield closed form
Figure 3.3: Divided-difference Kalman filter prediction (blue) and filtering (black) densities for several consecutive time instances of the “particle filter example”. The green dots mark the true states.
Figure 3.4: Monte Carlo Kalman filter prediction (blue) and filtering (black) densities for several consecutive time instances of the “particle filter example”. The green dots mark the true states.
recursions as in the Gaussian case. Nevertheless, simple filtering and smoothing algorithms can be devised. The main result of the paper is the $t$ filter that appears similar to the KF except for a matrix update

$$P_{k|k} = \frac{\eta_k + (y_k - \hat{y}_k)^T S_k^{-1} (y_k - \hat{y}_k)}{\eta_k + m} (P_{k|k-1} - K_k S_k K_k^T)$$

(3.12)

that nonlinearly depends on the measurement $y_k$. Furthermore, extra scaling factors are applied to the matrix parameters within the time and measurement update. This is required to, for example, replace Gaussian noise densities by $t$ densities in an optimal way. A smoother derivation based on Student’s $t$ distribution arrives at the Rauch-Tung-Striebel algorithm, but with the parameters of a previously run $t$ filter.

We here show the potential of the $t$ filter and smoother on the Student’s $t$ random walk example of Section 2.4.2. The exact state estimation densities are illustrated in Figure 2.3. We also recall that the KF and RTS smoother exhibit problems with the outliers in this example (Figure 3.2). The results for the $t$ filter and smoother are shown in Figure 3.5. For $k = 9$ and $k = 12$, the filtering density is large to account for the discrepancy between the predictions and the measurement outliers. In contrast to the KF, the resulting filtering posterior covers the true state well and reflect the exact density (Figure 2.3) better. The smoothing results appear close to the prediction results for $k = 9$ and $k = 12$. Despite the simple algorithms, an improvement is obtained by $t$ filtering and smoothing.

### 3.5 Ensemble Kalman Filters

The ensemble Kalman filter (EnKF) is a sampling based implementation of the KF that has been developed for extremely high-dimensional state estimation problems. The first EnKF [41] was suggested in the early 1990s, during a time period that also saw the first particle filter [60] and sigma point KF [90]. In contrast to the latter two, the EnKF has not gained as much interest among state estimation researchers. Perhaps one reason for this is that most of its development is documented in geoscientific rather than signal processing journals, often with many application specific details. However, its applicability in high-dimensional problems that cannot be solved by “traditional” KF and the fact that the EnKF can be used in nonlinear and non-Gaussian state-space models, motivates a closer investigation.

Paper C provides a detailed review that derives the EnKF from the KF, shows its relations to other nonlinear filters, its challenges, and a number of algorithmic extensions. Hence, only the nonlinear EnKF variant used in the upcoming examples is shown here.

Similar to the PF, the EnKF propagates an ensemble of $N$ state realizations. For a nonlinear model (2.1), the EnKF time update simulates each member of the filtering ensemble $x_{k|k}^{(i)}$ with an independent noise realization $v_k^{(i)}$,

$$x_{k+1|k}^{(i)} = f(x_{k|k}^{(i)}, v_k^{(i)}), \quad i = 1, \ldots, N.$$  

(3.13)
Figure 3.5: Student’s t filter and smoother prediction (blue), filtering (black), and RTS smoothing (orange) densities for several consecutive time instances of the Student’s random walk. The green and red dots mark the true states and measurements, respectively.
The time update requires the creation of an output ensemble which includes independent measurement noise realizations $e^{(i)}_k$,

$$y^{(i)}_{k|k-1} = h(x^{(i)}_{k|k-1}, e^{(i)}_k), \quad i = 1, \ldots, N. \quad (3.14)$$

The samples $x^{(i)}_{k|k-1}$ and $y^{(i)}_{k|k-1}$ are used to compute the EnKF gain $\bar{K}_k$. Finally, the measurement update is given by

$$x^{(i)}_{k|k} = x^{(i)}_{k|k-1} + \bar{K}_k(y_k - y^{(i)}_{k|k-1}), \quad i = 1, \ldots, N. \quad (3.15)$$

For linear models (2.2), the above scheme converges to the KF as $N \to \infty$. That is, the sample statistics of the ensemble then correspond to the KF mean and covariance. Often, however, the EnKF is applied with $N \ll n$ ensemble members, where $n$ is the state dimension. Then, a tendency to underestimate the uncertainty can be seen and further measures, such as localization or ensemble inflation, are required. In contrast to basic PF, the EnKF adjusts its “particles” in the time and the measurement update.

The simple structure of the EnKF facilitates its application to nonlinear models, as shown above. However, nonlinear EnKF lack a theoretical foundation in the Bayesian filtering solution and do not converge to (2.13) as $N \to \infty$ in general. We illustrate this by running an EnKF with $N = 1000$ on the “particle filter example” of Section 2.4.1. The exact filtering and prediction results are given in Figure 2.2. Kernel density estimation (Appendix B.1) is used to obtain continuous densities from the prediction and filtering ensembles. The EnKF results are shown in Figure 3.6 and fail to represent the exact densities at many time steps.

A more positive outcome is achieved for the saturated measurements example of Section 2.4.3. The EnKF densities in Figure 3.7 represent the exact solution in Figure 2.4 very well.
Figure 3.6: Ensemble Kalman filter prediction (blue) and filtering (black) densities for several consecutive time instances of the “particle filter example”. The green dots mark the true states.
Figure 3.7: Ensemble Kalman filter prediction (blue) and filtering (black) densities for several consecutive time instances of the saturated measurements example. The green and red dots mark the true states and measurements, respectively.
The examples in the introduction highlight the relevance of Bayesian state estimation for many existing and upcoming technologies such as air traffic control, autonomous driving, drone tracking, and the search for missing aircraft. The mathematical formulation of prediction, filtering, and smoothing in terms of stochastic state-space models and conditional probability distributions in Chapter 2 allows for the needed flexibility and even provides elegant recursive solutions. However, approximations are required to translate the mathematical results to the algorithms of Chapter 3 and the papers in Part II. This thesis contributes to the understanding of existing approximate filters and the design of new algorithms for the relevant challenge of heavy-tailed noise.

Research on existing algorithms is important to understand, critically assess, and improve the ever-increasing number of filters and smoothers in the literature. Two of the included papers are therefore tutorial or review articles. First, Paper A provides a detailed analysis of existing nonlinear Kalman variants and investigates them in a common framework that might not appear obvious from the individual publications. This can facilitate much quicker access for researchers who are novel to the field, but also provide new insights for the more experienced readers. The detailed analysis of the encountered moment computation problem and the survey of sigma point methods reveals similarities and limitations. Furthermore, it can serve as starting point for novel developments. Second, Paper C explains the ensemble Kalman filter as an algorithm that has been developed by geoscientific rather than signal processing researchers. Its analysis from a Kalman filter perspective reveals a simple yet flexible algorithm. This might not be obvious from the extensive geoscientific literature that ever so often discusses application specific details. Moreover, the paper includes a discussion of limitations and extensions that are often essential for the EnKF. In addition to the above two articles, Paper D addresses a specific tracking problem using nonlinear Kal-
man filters. Here, the performance analysis with respect to the noise parameters contributes to the understanding of different filter and model combinations. Finally, the illustrative examples in Chapters 2 and 3 of Part I have been included to make Bayesian state estimation more accessible. In summary, understanding can be seen as a main theme throughout the thesis.

In Paper B, Student's $t$ distribution is investigated because it can reflect real world phenomena such as measurement outliers and jumps in the process noise more realistically than the Gaussian distribution. Results for the $t$ distribution are then exploited to arrive at simple algorithms in the spirit of the Kalman filter. Therefore, simplicity can be highlighted as second main theme of this thesis.

The research of this thesis hopefully enriches the reader's toolbox with a wide range of algorithms. Still, some open questions remain for the practitioner and answers can only be provided under consideration of the specific problem in many cases. For example, it is difficult to say whether a certain Kalman filter can solve a nonlinear problem. Its success depends on how nonlinear the involved functions are and how well a Gaussian distribution approximates the exact posterior density. The lack of a reference solution, unlike in the examples of Chapters 2 and 3, complicates the situation further. However, having a range of simple and well understood algorithms certainly helps to pick the right filter.

The algorithms of this thesis should not only be seen as out-of-the-box solutions to arbitrary state estimation problems. It is often useful to have a closer look at the state-space models at hand to see if the estimation problem can be divided into smaller sub-problems. For example, the height tracking in air traffic control is often separated from estimation of a horizontal position. Each sub-problem can then be approached with a dedicated filter. A remaining challenge is to devise methods that combine the filter results for all sub-problems in the best way.

Many directions for future work can be devised from the material in this thesis. For instance, the examples in Chapter 2 showed that significant improvements can be obtained via smoothing. It is therefore interesting to investigate smoothing variants of the EnKF in Paper C, that facilitate improved state estimation in high-dimensional models. Also, further research on the Student's $t$ filter and smoother of Paper B is required. The provided exact filtering step for $t$ noise can be a starting point for novel algorithms. Furthermore, a combination of $t$ filters and smoothers for nonlinear models with the sigma point methods of Paper A is an interesting route that has gained some attention in the recent literature. Even when the offline (fixed-interval) smoothing of this thesis is not affordable, a small lag is conceivable in many applications. Therefore, research on fixed-lag smoothers that process $\hat{y}_k$ to estimate $x_{k-l}$, for a small lag $l > 0$, can be beneficial.
Appendix
On Probability Density Functions

This appendix lists some results from probability theory that are helpful to understand Chapter 2. More thorough accounts can be found in many textbooks, for example, [17] or [67].

A.1 Relevant Results and Definitions

A random variable \( x \) has a probability distribution instead of admitting a single value. Many relevant distributions can be represented by a probability density function \( p(x) \). The probability of \( x \) being in a region \( \mathcal{R} \) is given by

\[
\Pr\{x \in \mathcal{R}\} = \int_{\mathcal{R}} p(x) \, dx. \tag{A.1}
\]

Further requirements on \( p(x) \) are

\[
p(x) \geq 0, \quad x \in \mathcal{X}; \quad p(x) = 0, \quad x \notin \mathcal{X}; \quad \int_{\mathcal{X}} p(x) \, dx = 1, \tag{A.2}
\]

where the set \( \mathcal{X} \) contains all values that \( x \) can admit.

An extension of the above are joint densities \( p(x, y) \) of two random variables \( x \) and \( y \). If \( p(x, y) = p(x)p(y) \) then \( x \) and \( y \) are independent.

The marginal density \( p(x) \) can be obtained from \( p(x, y) \) via marginalization

\[
p(x) = \int p(x, y) \, dy. \tag{A.3}
\]

Conditioning is the operation for including observations. The conditional density of \( x \) given a known \( y \) is

\[
p(x | y) = \frac{p(x, y)}{p(y)} = \frac{p(y | x)p(x)}{p(y)}. \tag{A.4}
\]
The last expression is known as Bayes’ theorem, with the likelihood \( p(y|x) \).

If \( x \) with density \( p_x(x) \) undergoes a transformation to form \( z = g(x) \), then the density of \( z \) can be obtained via a transformation theorem [67]. For a one-to-one mapping \( g : \mathcal{X} \rightarrow \mathcal{Z} \) with inverse mapping \( h : \mathcal{Z} \rightarrow \mathcal{X} \),

\[
p(z) = p_x(h(z))|\det(Dh(z))|, \quad \text{for } z \in \mathcal{Z}, \tag{A.5}
\]

and \( p(z) = 0 \) otherwise. Here \( |\det(Dh(z))| \) is the absolute value of the Jacobian determinant and has its origin in the underlying change of variables. A proof and the extension for many-to-one mappings \( g \) can be found in [67].

Mixture densities provide a framework to construct complex probability density functions from simpler components. This is useful to create multimodal densities that exhibit several maxima, for example. From \( N \) valid density functions \( p_i(x) \) and \( N \) weights \( w_i \geq 0 \) with \( \sum_i w_i = 1 \), a mixture density can be constructed as follows

\[
p(x) = \sum_{i=1}^{N} w_i p_i(x). \tag{A.6}
\]

The notion of a density can be too complex for many purposes. Often, a single point estimate \( \hat{x} \) is required, for example, for processing in a control algorithm. The most common choice of a point estimate \( \hat{x} \) is the mean value

\[
E(x) = \int_{\mathcal{X}} x p(x) dx, \tag{A.7}
\]

also known as the first moment of \( x \). Such mean estimates \( \hat{x} \) minimize the variance \( \text{var}(x - \hat{x}) \) [3]. Another common choice for \( \hat{x} \) is \( \text{argmax}_x p(x) \), in the context of Bayesian estimation often termed maximum a posteriori estimate\(^1\). A third potential \( \hat{x} \) is the median, for which

\[
\int_{-\infty}^{\hat{x}} p(x) dx = \int_{\hat{x}}^{\infty} p(x) dx = \frac{1}{2}. \tag{A.8}
\]

Unlike the mean and maximum \( \hat{x} \), the above median only applies to scalar \( x \). However, there are some generalization to the vector case. The median \( \hat{x} \) minimizes \( E(|x - \hat{x}|) \) [3].

The uncertainty of \( x \) is often specified by a covariance matrix

\[
\text{cov}(x) = \int_{\mathcal{X}} (x - E(x))(x - E(x))^T p(x) dx. \tag{A.9}
\]

Zero entries in the covariance matrix correspond to components of \( x \) that are uncorrelated.

### A.2 Important Probability Density Functions

This section lists a number of relevant probability distributions with their density functions.

\(^1\)For a posterior density \( p(x) \).
A.2 Important Probability Density Functions

The Gaussian Distribution

An \(n\)-dimensional random variable \(x\) with Gaussian distribution \(\mathcal{N}(\hat{x}, P)\) is fully characterized by its mean \(\hat{x}\) and covariance matrix \(P\). For a positive definite matrix \(P\), the Gaussian probability density function is given by

\[
\mathcal{N}(x; \hat{x}, P) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(P)}} \exp\left(-\frac{1}{2} (x - \hat{x})^T P^{-1} (x - \hat{x})\right).
\]  
(A.10)

Gaussian random variables that are uncorrelated are also independent. The marginal and conditional distribution of a partitioned Gaussian random vector are again Gaussian.

In the state estimation context, the Gaussian distribution is closely related to the Kalman filter. For example, the approximate nonlinear Kalman filters discussed in Papers A and D can be derived from intermediate Gaussian assumptions.

Student’s \(t\) distribution

A more detailed account of the \(t\) distribution is given in Paper B and the technical report [139] by the author of this thesis.

Student’s \(t\) distribution is a relative of the Gaussian distribution that can exhibit heavy tails for certain parameter choices. An \(n\)-dimensional \(t\) random variable \(x\) with distribution \(\mathcal{S}(\hat{x}, P, \nu)\) is fully characterized by a mean vector \(\hat{x}\), a positive (semi-)definite scale matrix \(P\), and the scalar degrees of freedom \(\nu > 0\). Lower degrees of freedom \(\nu\) result in heavier tails, that is, a higher probability of generating values far from \(\hat{x}\). For positive definite \(P\), the \(t\) probability density function is given by

\[
\mathcal{S}(x; \hat{x}, P, \nu) = \frac{\Gamma\left(\frac{\nu + n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu \pi}} \left(1 + \frac{1}{\nu} (x - \hat{x})^T P^{-1} (x - \hat{x})\right)^{-\frac{\nu + n}{2}}.
\]  
(A.11)

Alternatively, the \(t\) density can be written as an infinite Gaussian mixture with a latent Gamma-distributed variable \(\lambda\).

\[
\mathcal{S}(x; \hat{x}, P, \nu) = \int \mathcal{N}(x; \hat{x}, \frac{1}{\lambda} P) \mathcal{G}(\lambda; \frac{\nu}{2}, \frac{\nu}{2}) d\lambda,
\]  
(A.12)

where the Gamma density in shape/rate parametrization is given by

\[
\mathcal{G}(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta), \quad \lambda > 0.
\]  
(A.13)

Although the \(t\) distribution shares the convenient results for marginalization and linear transformations of the Gaussian distribution, there are some differences. The covariance is given by

\[
\text{cov}(x) = \frac{\nu}{\nu - 2} P
\]  
(A.14)
and finite only for $\nu > 2$. In general, $\nu$ determines the existence of moments. For example, $\nu = 1$ yields the Cauchy distribution that does not have a mean value. Uncorrelated $t$ variables are not independent in general. The conditional distribution of partitioned $t$ vectors is again $t$, but with increased degrees of freedom and a scale matrix that depends nonlinearly on the observed value. These results are employed to devise simple approximate filtering and smoothing algorithms in Paper B.

**The Dirac distribution**

The Dirac density $\delta(x)$ is a mathematical construct to describe deterministic values and relationships in a probabilistic context. Because $\delta(x)$ is a generalized function, it cannot be evaluated as such. Instead, the Dirac density is defined via integrals

$$\int f(x)\delta(x - \hat{x})dx = f(\hat{x}).$$

(A.15)

Dirac densities appear as limit cases of other probability density functions. For example, Gaussian densities tend to Dirac densities for $P \to 0$.

In the state estimation context, the samples and weights of a particle filter can be seen as a mixture of Dirac densities. A similar interpretation can be given to the sigma points and weights in unscented Kalman filters (Paper A) and the ensemble in the ensemble Kalman filter (Paper C).
Exact Bayesian State Estimation for Scalar Problems

This appendix presents methods to numerically compute the Bayesian state estimation densities of Chapter 2. A discussion of the representation of probability density functions in a computer is followed by a brief presentation of grid-based point mass filters and an algorithm for filtering in the case of intractable likelihood functions. In principle, the considerations are not limited to scalar problems. However, the computational load associated with grid-based approaches impedes the use in high-dimensional problems.

B.1 The Algorithmic Treatment of Scalar Densities

The following section denotes random samples with superscripts in parentheses (as in $x^{(i)}$) and grid values with bare superscripts (as in $x^i$).

Uniformly sampled densities

Assume a random $x \in \mathbb{R}$ with continuous density $p(x)$ and its probability mass within a known interval. That is, $p(x) = 0$ outside of that interval. A reasonable representation of $p(x)$ in a digital computer can be achieved by storing $p(x^i)$ for grid points $x^i$, $i = 1, \ldots, M$. For sufficiently dense uniformly spaced grids with increments $\Delta$ we can associate a point mass $p(x^i) \Delta$ to each $x^i$. Expected values simplify to finite sums $E(f(x)) = \sum_{i=1}^{M} f(x^i)p(x^i)\Delta$. An important aspect is that algorithms often yield densities that do not integrate to 1. Then, normalization can be applied to ensure $\sum_{i=1}^{M} p(x^i)\Delta = 1$. Also non-uniform grids are conceivable. However, the computational gain seldom justifies the extra book-keeping efforts in grid-based methods for the scalar case. A simpler route is to employ particle methods which automatically provide an adaptive and non-uniform grid.
Random samples from sampled densities

Given uniformly spaced $x^i$ and $p(x^i)$, realizations of $x$ can be generated via the inversion method [17, 32] which uses the cumulative distribution function and uniform random numbers. Below is a simple algorithm that furthermore adds noise to obtain values in between grid points. There is a close relation to the resampling algorithms in particle filters [11].

Algorithm for sampling from uniformly sampled densities

1: Sample $u^{(i)} \sim U(0, 1)$.
2: Pick the largest $x^i$ such that $\sum_{j=1}^i p(x^j) \Delta < u^{(i)}$.
3: Add uniform noise $\sim U(-\frac{\Delta}{2}, \frac{\Delta}{2})$ to form $x^{(i)}$.

Kernel density estimates from samples

Kernel density estimation [17] can be used to approximate $p(x^j)$ from (weighted) random samples. For $x^{(i)}$ and $w^{(i)}$ with $\sum_{i=1}^N w^{(i)} = 1$, the density at $x^j$ is given by

$$
\hat{p}(x^j) = \sum_{i=1}^N w^{(i)} N(x^j; x^{(i)}, \sigma^2).
$$

Here, $\sigma^2$ is the variance of the Gaussian kernels. The concept can be applied to vector $x$ with the only change that a corresponding multivariate Gaussian density is used. Also other kernels are conceivable, but the Gaussian is a simple and smooth choice. Put simply, a kernel density estimate can be seen as a scaled and smoothed histogram.

An example of a kernel density estimate is shown in Figure B.1 for a Gaussian mixture $p(x)$. The grid comprises $M = 1000$ points $x^j$ between 0 and 10, the kernel variance is $\sigma^2 = 0.015$, and two different data sets with $N = 200$ and $N = 1000$ are used. A single scaled kernel is also included. The $N = 1000$ estimate closely resembles $p(x)$, whereas $N = 200$ yields many local modes. Such variations from $p(x)$ are common for few samples and small kernel variances.

B.2 Point Mass Filters and Smoothers

Point mass approaches [64, 158] evaluate the Bayesian filtering and smoothing solutions of Section 2.3 for a finite number of grid points. Hence, the model must be formulated in terms of a transition density and likelihood function (2.5).

The first point mass filter was presented in [20]. Later, advanced versions were discussed in [97, 100]. The first related smoother was presented in [97] along the theoretical smoothing density (2.14). Instead of mere point masses, however, piecewise linear density approximations were employed. Both [20, 100] discuss how to best choose the grid values at each time step, and how to rotate rectangular grids in higher dimensions. For the one-dimensional case, however,
B.3 A Likelihood-free Filter

The point mass approach of the previous section requires evaluation of the transition density and likelihood (2.5), which can be difficult for models (2.1). The particle filters of Section 3.1 alleviate this requirement by processing the likelihood only. Here, we present a likelihood-free filtering algorithm that combines random sampling with grid based approaches. The algorithm is used to compute the filtering density in the saturation example of Section 2.4.

The filter relies on the sampling and kernel density estimation techniques of Appendix B.1. The user determines a grid by choosing \( M \) points \( x^i \), the number of

\[
p(x_{k+1}^i | y_{1:k}) \propto \sum_{j=1}^{M} p(x_{k+1}^j | x_k^i) p(x_k^i | y_{1:k}). \tag{B.2a}
\]

\[
p(x_k^i | y_{1:k}) \propto p(x_k^i | y_{1:k-1}) p(y_k | x_k^i). \tag{B.2b}
\]

Normalization ensures that the above approximate sampled densities integrate to 1. For grid increments \( \Delta \), the point mass \( p(x_k^i | y_{1:k-1}) \Delta \) associated with \( x^i \) justifies the name of the algorithm. The computations require \( M^2 \) evaluations of the transition density, which dominates the cost.

A backward iteration to obtain the smoothing density at \( x^i \) is given by

\[
p(x_k^i | y_{1:L}) \propto p(x_k^i | y_{1:k}) \sum_{j=1}^{M} p(x_{k+1}^j | x_k^i) p(x_{k+1}^j | y_{1:L}) \frac{p(x_{k+1}^j | y_{1:k})}{p(x_k^i | y_{1:k})}. \tag{B.3}
\]

Again, \( M^2 \) evaluations of the transition density are required.

### Figure B.1: Gaussian mixture density and two kernel density estimates. The scaled component illustrates the kernel width.

we can skip such considerations and focus on straightforward approximation of (2.13) and (2.14) for a fixed uniform grid.

The only, yet crucial, user choice is a uniformly spaced grid with \( M \) values \( x^i \) that cover the relevant region of the state space. Then, forward iterations for the sampled one-step-ahead and filtering densities at \( x^i \) are given by

\[
p(x_{k+1}^i | y_{1:k}) \propto \sum_{j=1}^{M} p(x_{k+1}^j | x_k^i) p(x_k^i | y_{1:k}). \tag{B.2a}
\]

\[
p(x_k^i | y_{1:k}) \propto p(x_k^i | y_{1:k-1}) p(y_k | x_k^i). \tag{B.2b}
\]
random samples $N$, and variances $\sigma_x^2$ and $\sigma_y^2$ for two-dimensional kernel density estimation in $x_k$ and $y_k$.

The time update requires generating $N$ samples $x_k^{(i)}$ and $v_k^{(i)}$ from $p(x_k | y_{1:k})$ and $p(v_k)$, respectively. Because $p(x_k | y_{1:k})$ is only available for the grid points $x'$, the sampling method of Appendix B.1 must be used. The samples are then transformed and used to compute a kernel density estimate of $p(x_{k+1} | y_{1:k})$ for the grid points,

$$x_{k+1}^{(i)} = f\left(x_k^{(i)}, v_k^{(i)}\right), \quad i = 1, \ldots, N, \tag{B.4a}$$

$$p(x_{k+1}^j | y_{1:k}) \propto \sum_{i=1}^N N(x_{k+1}^j; x_{k+1}^{(i)}, \sigma_x^2) N(y_k; y_k^i, \sigma_y^2), \quad j = 1, \ldots, M. \tag{B.4b}$$

For the measurement update, the samples $x_{k|k-1}^{(i)}$ are supplemented with $N$ measurement noise realizations $e_k^{(i)}$. Then, output samples are generated and processed via kernel density estimation,

$$y_k^{(i)} = h(x_{k|k-1}^{(i)}, e_k^{(i)}), \quad i = 1, \ldots, N, \tag{B.5a}$$

$$p(x_k^j | y_{1:k}) \propto \sum_{i=1}^N N(x_k^j; x_{k|k-1}^{(i)}, \sigma_x^2) N(y_k; y_k^i, \sigma_y^2), \quad j = 1, \ldots, M. \tag{B.5b}$$

In fact, (B.5b) is a two-dimensional kernel density estimate of $p(x_k^j, y_k | y_{1:k-1})$. However, according to (A.4) we are only interested in this joint density for a specific $y_k$ to compute the conditional density. The desired filtering result is obtained after normalization. The computations can be sped up by computing only those kernels that actually contribute to $p(x_k^j | y_{1:k})$, that is, by only processing $y_k^{(i)}$ close to $y_k$. 
Bibliography


Part II

Publications
Papers

The articles associated with this thesis have been removed for copyright reasons. For more details about these see:

http://urn.kb.se/resolve?urn:nbn:se:liu:diva-134867
PhD Dissertations
Division of Automatic Control
Linköping University


