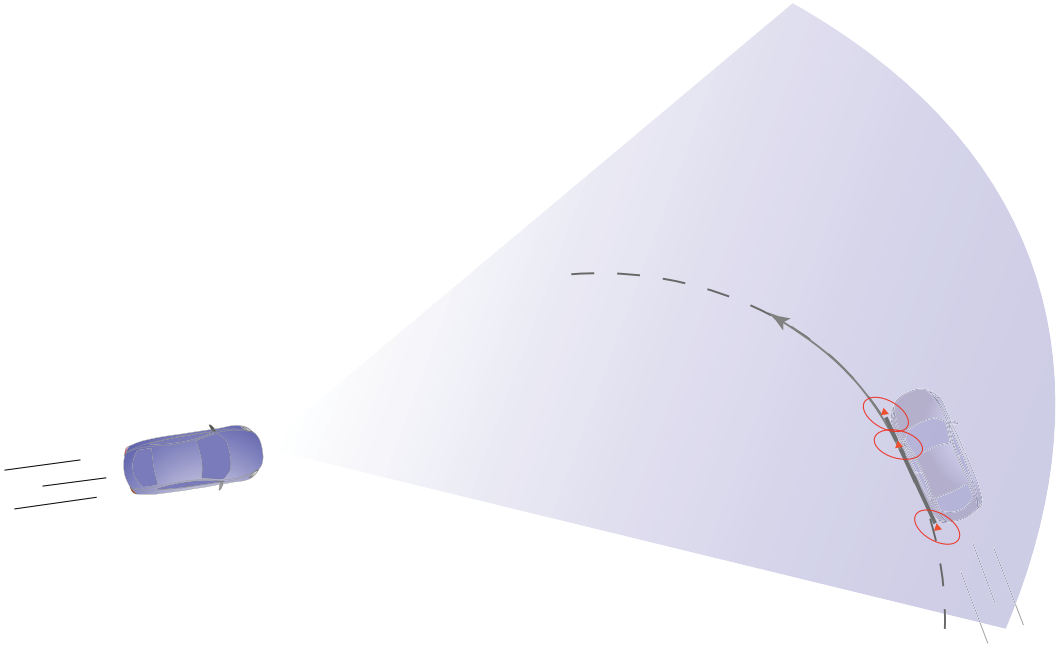
CHALMERS



Tracking and radar sensor modelling for automotive safety systems

Lars Danielsson

Department of Signals and Systems
Signal Processing Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2010

Thesis for the degree of Doctor of Philosophy

Tracking and radar sensor modelling for automotive safety systems

by

Lars Danielsson

CHALMERS

Department of Signals and Systems
Chalmers University of Technology
Göteborg, Sweden 2010

Göteborg 2010

Tracking and radar sensor modelling for automotive safety systems
Lars Danielsson

This thesis has been prepared using LaTeX.

Doktorsavhandlingar vid Chalmers Tekniska Högskola
ISBN 978-91-7385-383-5
Series No 3064
ISSN 0346-718X

Department of Signals and Systems
Signal Processing Group
Chalmers University of Technology
SE-412 96 Göteborg, Sweden

Phone: +46 (0)31 59 2809
Fax: +46 (0)31 59 6670
E-mail: ldanie25@volvocars.com

Printed by Chalmers Reproservice
Göteborg, Sweden, March 2010

To my fiancée Marja

Tracking and radar sensor modelling for automotive safety systems

Lars Danielsson
Department of Signals and Systems
Chalmers University of Technology

Abstract

This thesis studies the problem of tracking in the setting of an automotive safety system. In particular, it considers the problem of estimating the surrounding traffic situation using observations from radar sensors. In this context, we develop two accurate radar sensor models and vehicle tracking algorithms, when multiple measurement can be obtained from each object. The first model describes the radar return from a vehicle as originating from a known set of point features, whereas the second approach jointly estimates the position of the reflecting point features and the position of the extended object. Both models incorporate novel approaches for describing the limited resolution of the sensor and the resulting tracking frameworks effectively exploit the information in all the detections from the vehicle.

Additionally, we investigate the use of radar measurements in a probability hypothesis density (PHD) framework for constructing maps over the stationary objects around a vehicle. By proposing new data clustering and merging methods we manage to exploit the inherent structure in the map. The efficiency of the PHD framework both in the measurement update and in the representation of the map is thereby improved considerably.

Besides models for accurately describing the measurements, we also propose a new vehicle motion model that describes the driver as an optimal controller with preferences (described by a parameter vector) which are tracked over time. The proposed framework enables more accurate predictions and offers a formal treatment of the model uncertainties.

Finally, we present a modular sensor data fusion functional architecture, tailored for development of automotive safety systems. The ambition with this paper is to illustrate how the other findings in the thesis can be implemented in practice.

Keywords: extended target tracking, radar sensor models, motion models, automotive safety systems, sensor data fusion

Acknowledgment

I would like to thank my supervisor Professor Mats Viberg at Chalmers for his coaching, support and letting me join his research group. A group consisting of the most inspiring bunch of researchers one can find. A special thanks to my co-supervisor, colleague and friend Lennart (Letnart) Svensson, you are a truly remarkable supervisor and you have made this journey most interesting.

My gratitude goes also out to Volvo Car Corporation and my managers over the year for allowing me to pursue my Ph.D. degree as a part of my work assignment; Michael Edvardsson and Dr. Tomas Andersson for believing in me from the start and Mats Lestander for enabling me to focus on my studies for the last years. Looking forward to putting this theory into practise in a future active safety system.

During my studies I have had the privilege of collaborating with some of the most talented people. Fredrik Sandblom, Joakim Sörstedt and Lennart Svensson ensured that I got a flying start while Malin Lundgren, Daniel Svensson and Christan Lundquist made for an interesting finale. This thesis would not have the same quality without your contributions and it would definitely not have been this much fun.

I would like to acknowledge the help I have received with proofreading the manuscript by Lennart Svensson, Daniel Svensson[1], Malin Lundgren and Mattias Brännström. You have not only contributed to improving the thesis, you have also ensured that at least someone has read it (for this I apologise).

Last, but definitely not least, I would like to express my most heartfelt gratitude to my fiancée Marja. You have taken a big load these last months, now it is my turn to make it up to you.

Göteborg, March, 2010
Lars Danielsson

[1]An especial thanks goes to Daniel Svensson for proofreading the 160 most important characters in the thesis, Thanks!.

List of publications

This thesis is based on the following publications:

Paper I
L. Danielsson, F. Sandblom, L. Svensson, and J. Sörstedt., A radar sensor model for improved tracking of vehicles using possibly unresolved detections. Submitted to *IEEE Transactions on Aerospace and Electronic Systems.*

Paper II
D. Svensson, M. Ulmke, and L. Danielsson., Multi-target sensor resolution model for arbitrary target numbers, In *Proceedings of SPIE Signal and Data Processing of Small Targets 2010*, Orlando, April, 2010.

Paper III
L. Danielsson, M. Lundgren, and L. Svensson., Adaptive radar sensor model for tracking structured extended objects, Submitted to *IEEE Transactions on Aerospace and Electronic Systems.*

Paper IV
C. Lundquist, L. Danielsson, and F. Gustavsson., Road Mapping using Radar Measurements in a Probability Hypothesis Density Filter, Submitted to *IEEE Transaction on Signal Processing.*

Paper V
J. Sörstedt, L. Svensson, F. Sandblom, and L. Danielsson., A new vehicle motion model for improved predictions and situation assessment. Submitted to *IEEE Transactions on Intelligent Transportation Systems.*

Paper VI
F. Bengtsson and L. Danielsson., A design architecture for sensor data fusion systems with application to automotive safety. In *Proceedings of Intelligent Transport Systems World Congress, 2008*, New York, November 2008.

Other publications by the author, omitted in the thesis:

- F. Bengtsson and L. Danielsson. Designing a real time sensor data fusion system with application to automotive safety. Technical Report, Chalmers University of Technology, Mars 2008. ISSN: 1403-266X.
- J. Gunnarsson, L. Svensson, L. Danielsson, and F. Bengtsson. Tracking vehicles using radar detections. In *Proceedings of IEEE Intelligent Vehicles Symposium, 2007*, Istanbul, June 2007.
- J. Gunnarsson, L. Svensson, F. Bengtsson, and L. Danielsson. Joint driver intention classification and tracking of vehicles. In *Nonlinear Statistical Signal Processing Workshop 2006*, Cambridge, UK, September 2006.
- L. Danielsson, H. Lind, and S. Jonasson. INSAFES HCI principles for integrated ADAS applications. In *Lecture Notes in Computer Science, Universal Access in Human-Computer Interaction. Ambient Interaction*, volume 4555, pages 339 - 348. Springer, 2007.
- L. Danielsson, H. Lind, E. Bekiaris, M. Gemou, A. Amditis, M. Miglietta, P. Stålberg. HMI Principles for Lateral Safe Applications. In *Lecture Notes in Computer Science, Universal Access in Human-Computer Interaction. Ambient Interaction*, volume 4555, pages 330 - 338. Springer, 2007.
- A. Amditis, N. Floudas; U. Kaiser-Dieckhoff, T. Hackbarth; B. Van Den Broek, M. Miglietta, L. Danielsson, M. Gemou, E. Bekiaris. Integrated vehicle's lateral safety: the LATERAL SAFE experience. In *IET Intelligent Transport Systems*, volume 2(1), pages 15 - 26, March, 2008
- A. Polychonopoulos, M. Gemou, L. Danielsson, A. Amditis, E. Bekiaris. Design of a friendly HMI combining multiple sources of warnings, for the Lateral Safe applications. In Proceedings *24th Annual European Conference on Human Decision Making and Manual Control EAM2005*, Athens, Greece, October, 2005

Contents

Part I
Introduction

Chapter 1

Introduction

Considerable effort is put into developing *preventive safety systems*[1] that assist the driver in safely guiding the vehicle to the desired destination. By monitoring and analysing the traffic situation, these systems aim to support the driver in preventing accidents from occurring in the first place. The system can be seen as being composed of three layers, each responsible for their own separate part. *A perception layer* that uses sensors to perceive the current traffic scenario, *a decision layer* that uses this information to detect hazards and decide on the appropriate action, and, *an action layer* that executes the decided action.

Some of these systems base their decisions on observations of the position of surrounding vehicles estimated using external object sensors mounted on the vehicle, such as radar and vision sensors. The sensor observations are often influenced by noise and clutter, and hence need to be refined before being transmitted to the decision layer. A tracking algorithm uses a series of these observations and tries to accurately estimate, e.g., the current positions and future trajectories of adjacent vehicles. Together with the sensors. the tracking algorithm forms a tracking system, which provides a refined perception to the decision layer. For certain decisions, even more accurate and reliable information is needed. Typically this is achieved by combining or

[1] Also known as active safety systems. These two terms are used interchangeably throughout this thesis.

fusing information from multiple sensors. In this case, the tracking system is also called a sensor data fusion system.

This thesis considers tracking systems (or parts of tracking systems), designed to support these applications with accurate and reliable information about the surrounding traffic situation. Although tracking and fusion systems have been used extensively in military applications, such as surveillance systems and airborne radar target tracking, this thesis highlights aspects that are of particular interest in the application to automotive safety systems. For example, in airborne radar target tracking it is commonly assumed that the origin of the observations made by the radar can be accurately modelled as a point source. This is often a valid assumption when tracking aircrafts at a distance of tens of kilometres. However, in automotive scenarios, where the distance to the objects of interest is much shorter, the radar is typically capable of detecting multiple features (reflection centres) on the same object.

Clearly, receiving multiple detections from a object, e.g. a vehicle, offers a possibility to extract more detailed information about the object. To address this, we propose more accurate radar sensor models and tracking algorithms when multiple detections can be obtained from each object. Paper I and III focus on tracking moving objects and aim at improving the tracking of vehicles in a traffic environment. Paper II concentrates on resolution models for general objects (or group of objects), made up of multiple reflectors which may or may not be resolved. The goal in Paper IV is to use radar measurements and the probability hypothesis density (PHD) framework to construct a map over the stationary objects around a vehicle.

Apart from modelling the radar observations and using them to estimate the traffic environment, this thesis studies a model for accurately describing the motion of vehicles. In Paper V we propose a new motion model that describes the driver as an optimal controller with preferences (described by a parameter vector) which are tracked over time. In the last paper, Paper VI, we propose a modular sensor data fusion functional architecture, tailored for development of automotive active safety systems. The ambition with this paper is to illustrate how the other findings can be implemented in practice.

This work was conducted within the IVSS[2] sponsored project, sensor fusion for safety (SEFS). The aim of the project is to develop methods of combining information from different sensors in order to get a better view of the environment around the vehicle. By providing this information, it is possible to construct more effective preventive safety systems.

1.1 Thesis outline and contributions

The thesis is divided into two parts. The first part gives an introduction and background to the subject at hand, and convey some basic theory to those not familiar with tracking theory and radar sensor modelling. The second part contains the author's contributions in the form of publications. An overview of the introductory chapters are given below together with a discussion on the scientific contributions in the publications.

Part I: Introduction

Chapter 1 *Introduction*
This general introduction to the thesis includes discussions of contributions and future work.

Chapter 2 *Preventive safety systems*
This chapter gives an introduction to preventive safety systems. The motivation for bringing these systems to the market is discussed and put into context alongside other measures of reducing the number of dead and injured in traffic-related accidents. The basic components that compose a preventive safety system are presented together with the different aspects required to be considered in the design. This chapter focuses on giving the reader the perspective needed to understand the reasoning behind the following chapters.

[2]Intelligent vehicle safety system project is a Swedish research project sponsored by the Swedish road authority and Vinnova. The aim of the project was to strengthen Sweden's competence in the field of IVSS and preventive safety systems.

Chapter 3 *Tracking Theory*
This chapter presents the different components of a tracking system along with the underlaying theory. The aim is to give the reader the appropriate background to appreciate the publications and to give an overview of the different methods that are available. This will range from filtering techniques and the models therein, to data association methods and track handling schemes.

Chapter 4 *Radar sensor modelling*
To better position the contributions in this thesis related to radar sensor modelling, this introductory chapter on radar sensors is included. The discussions are primarily focused on the signal processing and signal modelling perspectives of radar. That is, discussions about radar system performance analysis and components are omitted to benefit the understanding of the characteristics of the radar signal and the radar measurements.

Part II: Publications

Paper I *A radar sensor model for improved tracking of vehicles using possibly unresolved detections*
This paper is concerned with the problem of tracking vehicles using radar detections. In particular, we deal with problems where multiple detections are received from each vehicle. Previous research shows that automotive radar sensors mainly receive detections from a discrete set of strong reflection centres. However, due to the limited resolution of the sensor, some reflection centres may be detected individually while others are clustered and, hence, only render a single joint detection. In this paper we propose a radar sensor model capable of describing multiple detections from each vehicle and the effects of a limited sensor resolution. We also suggest appropriate approximations to reduce the inherent complexity of the data association problem and to make the model suitable in a tracking framework. Using these approximations we derive a tractable vehicle tracking framework that effectively exploits the information in all the detections from the vehicle.

Paper II *Multi-target sensor resolution model for arbitrary target numbers*

In this paper we focus on modelling the resolution capability of a radar sensor. In many radar tracking problems the observed objects are so closely spaced that they cannot always be resolved by the sensor(s). Typical examples are those discussed in Paper I. Ignoring the limited sensor resolution in a tracking system may lead to degraded tracking performance, in particular unwanted track-losses. In this paper we extend the resolution model, given for two partially unresolved objects in [33], to the case of arbitrary object numbers. We also derive the effects of the resolution model to the multi-target likelihood function and the possible data associations. Further, it is shown how the model can be integrated into the Joint Probabilistic Data Association Filter (JPDAF).

Paper III *Adaptive radar sensor model for tracking structured extended objects*

In this paper we propose a general tracking framework where we jointly estimate the position of an extended object and adapt the description of the object-generated radar measurements. The extended object is modelled as a set of structured radar reflection centres, where we assume that the structure is known but both the number of reflection centres and their positions on the structure are unknown. Additionally, we incorporate the radar resolution model proposed in Paper II to consider the effects of limited sensor resolution. As a result, we have a tracking framework capable of adapting the description of the extended object as the range and aspect angle to the object change.

Paper IV *Road Mapping using Radar Measurements in a Probability Hypothesis Density Filter*

In this paper we investigate the use of radar measurements in the probability hypothesis density (PHD) framework to construct a map over surrounding stationary objects. The main advantage with using a PHD filter is that it avoids the detection, the data association and the track handling problems in conventional multiple-target tracking. It also gives a parsimonious

representation of the map in contrast to grid based methods. Two original contributions further releases the complexity of the algorithm: first, a new data clustering algorithm is suggested to sort the components of the PHD in a few different clusters, which considerably improves the description of appearing objects used in the prediction step of the PHD filter. Second, a merging step is proposed to simplify the map representation in the PHD filter.

Paper V *A new vehicle motion model for improved predictions and situation assessment*

Most of us can agree that the motion of a vehicle under normal conditions is controlled by the driver. Even so, many vehicle motion models ignore this fact or, alternatively, model the drivers influence as a zero-mean white Gaussian noise process. This paper proposes an alternative motion model framework where the expected action of the driver is included as a control input. This enables us to consider and model different driver intentions, such as the desire to drive safely and comfortably, and how it influences the motion of the vehicle. The preferences of the driver are described by a cost function and the expected control input is derived by seeing the driver as an optimal controller. We show that this description gives an opportunity to more accurately predict the future trajectory, as well as to get an understanding of the current traffic situation.

This framework was originally presented in [51] and further developed in [21]. Here, it is extended by considering the interaction with other vehicles and, more importantly, with the ability to estimate the uncertainty in the expected driver control input. The latter is a very useful property when applying the framework in, e.g., tracking algorithms.

Paper VI *A design architecture for sensor data fusion systems with application to automotive safety*

When designing a sensor data fusion system that should work in real-time using online sensor data, several practical issues need to be considered, in addition to those discussed in classical text books. This paper highlights some of these aspects and considers, in particular, issues important in an

automotive context. Here, we elaborate on the different design choices and solutions to practical problems that arise when fusing data from different asynchronous sources. Based on these experiences, a modular fusion framework, suitable for a preventive safety system, is proposed and motivated. In the proposed framework it is relatively easy to exchange the sensor configuration and software components.

1.2 Future work

Elaborations on future extensions in the discussed areas are given below.

Sensor resolution modelling

The radar resolution model proposed in Paper II and used in Paper III has thus far been applied in traditional tracking frameworks using classical data association algorithms. However, the problem considered in Paper III can also be formulated using the *finite set statistics* FISST framework and *random finite sets* RFS developed by Mahler [38]. To the authors' knowledge, no framework exist for incorporating limited resolution considerations into FISST based methods. It would be interesting to adjust the FISST algorithms to handle these problems.

Structured object tracking

In Paper III, we demonstrate the possibility for jointly tracking an extended object and estimating the parameters of the sensor model. This is accomplished by imposing a rather rigid line structure to describe both the object and the position of the radar reflecting features. Although this structure proved to work well in the evaluated scenario, it would probably be limiting in more general cases.

The results from the evaluation merit an exploration of other structures for describing an extended object using the same filtering framework. This

could, for example, be a more general graph representation where the radar reflection features are only loosely interacting. The structure chould then adapt more freely to changes in object response, as the radar views it, e.g., from different angles. This type of model would probably fit well also for tracking a group of objects.

Mapping

From the PHD filter proposed in Paper IV, we obtain a general map describing where the radar measurements indicate that there are stationary objects. Although a radar sensor gives numerous detections on stationary objects, it will not give us the complete picture of the infrastructure surrounding the vehicle. However, the information obtained is fairly up to date and reliable in terms of relative positioning.

On the other hand, the map databases typically used for navigation are becoming more and more advanced and are enhanced with information about, e.g., curvature of the road and the geometry of an intersection. This information can be very detailed but it may also suffer from relative positioning errors and that the map data base is out-dated.

Hence, there is a potential in fusing the information in the radar based map with that in the map database to obtain a detailed locally updated map with high relative positioning accuracy.

Chapter 2

Preventive safety systems

Every year, far too many are involved in traffic-related accidents resulting in injury or death. In the EU alone, there are around 1.3 million injured and 43,000 fatalities every year [18]. Over the years, effective countermeasures have been taken reduce these numbers, such as building better and safer roads or developing vehicles crashworthiness. Although these countermeasures have had a considerable effect on lowering the number of fatal accidents, there is still a long way to go.

In this chapter preventive safety systems are explained and put it in context alongside other measures to lower the number of dead and severely injured in traffic-related accidents. An historical background on in-vehicle safety systems is given and the motivational forces that drive these systems to the market are highlighted. Plus discussion around the three basic layers, proposed in [42], that together compose a preventive safety system, i.e., *perception*, *decision* and *action layer*. This serves as an introduction and motivation for the next chapter where the different aspects of the perception layer will be discussed in more detail.

2.1 Background of in-vehicle safety systems

Historically, there been two dominating strategies for reducing the number of traffic-related fatalities, either by regulatory changes and developing the

road infrastructure to reduce the probability of an accident, e.g., introducing speed limits and constructing signal-controlled intersections, or by developing in-vehicle *(passive) protective systems* that mitigate the injuries when the accident becomes a reality, e.g., crumple zones and seat belts. The first strategy aims at designing a road transport system that is safe to travel on, whereas the latter is a matter of designing safe vehicles. Both are very important factors in reducing the number of severe injuries.

2.1.1 Protective safety systems

Newton's First Law of Motion states in part that, an object in motion will stay in motion - unless influenced by an external unbalanced force. This has been the focus of automotive research for more then a half century. The aim has been and remains to disperse the kinetic energy in an accident to protect the vehicle's occupants, i.e., protective safety system. Systems that we today see as standard components of a car have emerged over the year, e.g., bumpers, seat belts, crumple zones and air-bags. Research shown also that the introduction of these systems has played a vital role in reducing the severity of injuries [44].

2.1.2 Preventive safety systems

In resent years a new type of safety systems have been introduced. In contrast to the protective safety systems, these in-vehicle systems are designed to help the driver to prevent, avoid or mitigate an impending collision or accident. One of the first preventive safety systems introduced as early as 1978 was the *Anti-lock Braking System* (ABS). By monitoring the rotational speed of the wheels during braking, the system actively optimises the brake pressure to give maximum contact between the tyres and the road surface while not locking the wheels. This enables the driver to steer the vehicle even under emergency braking manoeuvres and shortens the stopping distance significantly in comparison with if the wheels where allowed to skid.

In 1995, *Electronic Stability Control* (ESC) systems where introduced on the market. ESC supports the driver in maintaining the desired path when

experiencing loss of optimal contact with the road surface. In an evasive manoeuvre or when negotiating a curve on a slippery road, ESC detects if and when the vehicle starts to slip and applies different braking force on each individual wheel to stabilise the car[1]. Although the system was first introduced as an option on high end cars, in 2006 ESC was offered as standard in 40% of the passenger vehicle models in EU [17]. In the U.S., legislation mandates that 100% of all vehicles weighing under 10,000 lbs (approximately 4.5 tons) should be equipped with ESC by 2012 [40].

Following the success of ABS and ESC, more and more advanced preventive safety systems are being introduced on the market. Through sensing not only the motion of vehicle hosting the system but also perceiving the traffic situation around the vehicle, they can detect and classify dangerous situations. Depending on the severity and timing, the systems make decisions whether to:

- inform the driver of the possible danger
- warn the driver of an impending accident
- actively assist or intervene in order to ultimately avoid the accident or mitigate its consequences.

Some of the systems are listed below. A more comprehensive list of current as well as upcoming systems can be found in [9], where there also is a discussion of their effectiveness and possible side effects.

Forward collision warning systems monitor the traffic situation in front of the vehicle, utilising external object sensors such as radar, lidar and/or vision systems. The system alerts the driver when it is determined that there is a high probability of a rear-end collision with the closest vehicle in path.

Collision avoidance by braking is an application that, if detecting that a collision is likely and that the driver does not react, automatically

[1]In order for a driver to have the same ability, there needs to be four brake pedals in the car, one for each wheel.

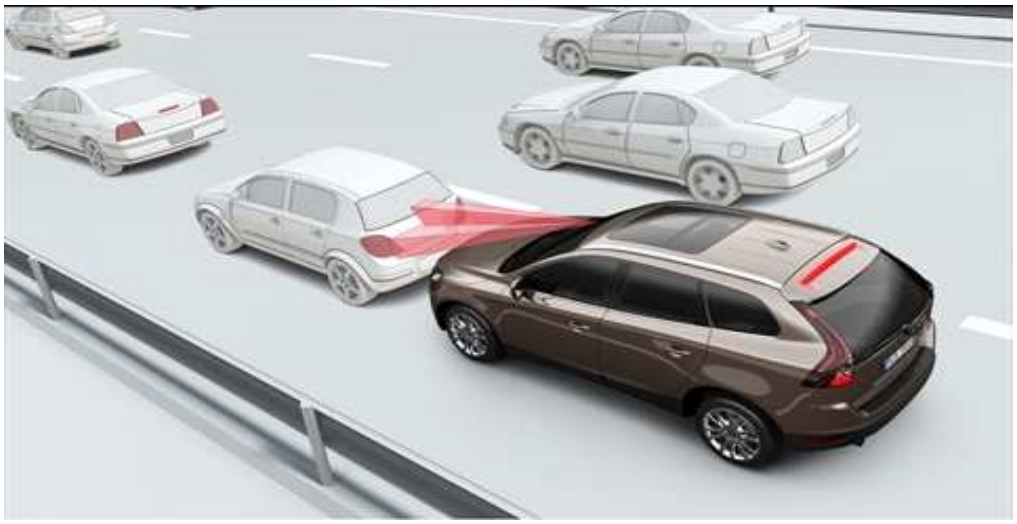


Figure 2.1: Collision avoidance system launched in 2008 on Volvo XC60. City safety automatically avoids low speed collisions of up to 15 kph and mitigates collisions for speeds up to 30 kph.

applies the brake to avoid the collision. An example of such a system is the Volvo City Safety system, which automatically avoids collision for a host speed of up to 15 kph, see Fig. 2.1.

Collision mitigation by braking functions in a similar manner as the collision avoidance by braking systems, by actively reducing the collision speed of the vehicle either by, amplifying the driver-initiated brake pressure or autonomously applying the brakes when the system determines that a collision is unavoidable. The difference being that the main aim is not to avoid the collision but rather to reduce its severity.

Lane departure warning typically uses a camera or laser that measures the vehicle position between the lane markings. If it detects that the vehicle begins to wander outside the lane without the use of indicators an alert is issued to the driver. The application targets unintentional lane changes and road departures.

Lane change aid supports the driver during lane change manoeuvres. By monitoring the traffic in the left and right adjacent lanes, the system

informs the driver of approaching cars with a speed and distance which would make it unsafe to change lanes.

The main focus of this thesis is on supplying this type of preventive safety systems with the information that they need in order to make them as efficient and reliable as possible.

2.1.3 Integrated safety systems

In the wake of the preventive safety system, synergies with the traditional protective safety systems have been identified. The information perceived by the preventive system can be used to earlier prepare the protective system of an impending collision. By utilising this information, the seat can be put in a better position for a collision, the seat belt can be pre-tightened and the air bag system can be pre-fired to enable a smoother deployment.

2.2 Motivation and safety benefit

Traffic safety research have shown that traditional protective safety systems, though proven to be very effective [9, 44], are reaching a ceiling when it comes to further significantly decreasing the number of serious injuries and fatalities. However, research also show that the preventive safety system ESC is estimated to have lowered the fatal single-vehicle crashes involving cars by 30-50% and by as much as 50-70% for SUVs [17]. This indicates that with the introduction of new preventive safety system, there is a possibility of changing the trend.

This is a step in the right direction, but as mentioned in the introduction, there is still a long way to go. When analysing the cause of accidents, research concludes that a majority of accidents involve human inattention [54, 32], which is not considered by these early systems. An analysis of American rear-end collisions [55], concludes that in over 78% of the cases, the driver did not take any corrective action before the collision. This is the main motivation for introducing the more advanced preventive safety systems discussed in Section 2.1.2. The general aim of these systems, is to aid drivers when driving.

The potential for these more advanced systems is clear. For example, research on the preliminary safety benefits of a forward collision warning system shows that it has the potential to prevent 51% of all police reported rear-end crashes [20] in the U.S. As yet however, there is no conclusive evidence that these new preventive safety systems have a real impact on real-world safety. The systems are relatively new on the market and the penetration is still quite low. There have been naturalistic studies [41, 14], or *field operational tests* (FOT), where more people get the opportunity to drive cars with and without these systems. During which a large quantity of data is collected and analysed to determine if these system have a positive impact on driving. Although, rear-end collisions are a major accident scenario, they occur relatively seldom. Therefore, the FOTs were only able to make vague statements such as,

> *"Forward Collision Warning may be capable of assisting drivers in avoiding crashes within their own lane by improving awareness of a developing conflict and prompting its recognition."* - [41]

Although research shows that there is a great potential for these advanced preventive safety systems, the need for continued research in this area is evident.

2.3 Basic system components

When analysing the examples of preventive safety system listed above, one can identify three separate basic components or layers that they all have in common. The systems need to:

1. perceive the traffic environment and the host vehicle motion
2. detect dangerous situations and make decisions on whether to inform, warn or intervene
3. convey this decision to the driver or to the vehicle.

In [42] these three layers are denoted, *perception*, *decision* and *action*, respectively. To make an effective preventive safety system, all these layers need to be carefully designed. To be able to detect a dangerous situation, the perception layer needs to supply the decision layer with sufficiently accurate and reliable information. Similarly, it does not matter how good you are at detecting dangerous situations if you have no means of conveying the information to the driver or making efficient interventions to avoid or mitigate the collision. When designing a preventive safety system all these different principles need to come together, e.g., tracking and estimation theory in the perception layer, detection and threat assessment methods in the decision layer and human behavioural studies as well as ergonomics in the action layer. Improved knowledge in all of these layers will result in the most efficient system.

In the sections to follow there will be a more detailed discussion of the different aspects in each layer.

2.3.1 Perception layer

In order to make the right decision the systems require robust and reliable information about the current traffic situation. This is retrieved by equipping the vehicle with sensors, such as:

- gyros and accelerometers to measure the host vehicle motion
- radar and lidar sensors to measure the traffic scenario around the vehicle
- vision system to measure the road and classify objects, e.g., cars, trucks and pedestrians.

Given the information from these sensors, the perception tries to estimate both the ego vehicle motion as well as the position and motion of the surrounding vehicles with associated uncertainties. This is typically done with a tracking system or a fusion system (in the event that data from several sensors is combined or fused). Both these systems are frameworks that use

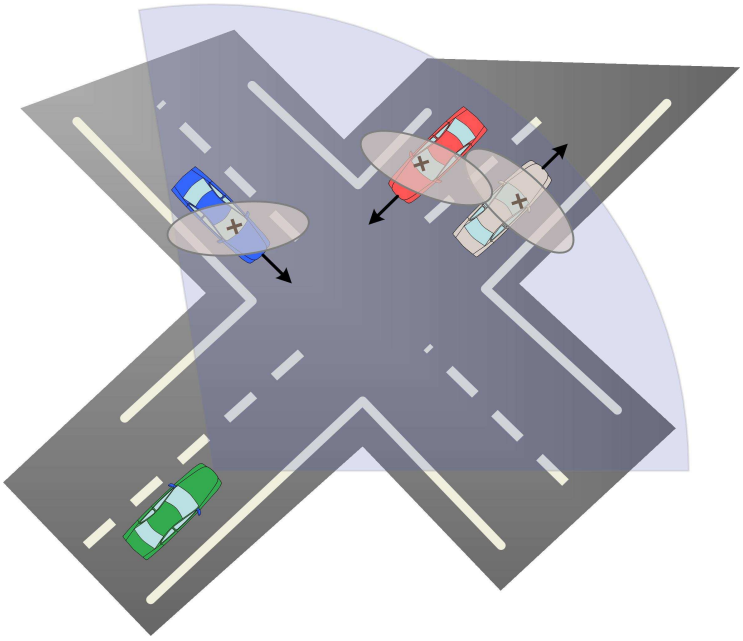


Figure 2.2: Example of a perception system mounted on a car observing the traffic situation at an intersection. The estimated positions of the vehicles are shown together with their associated uncertainty regions.

past and present measurements to more accurately estimate vehicle position and future trajectories. The frameworks apply statistical models of objects' motion and measurement response to incorporate a-priori knowledge about, e.g., typical motion of vehicles and sensor characteristics. These models are often called *motion model* and *measurement model*, respectively. Contributions to the modelling of vehicle motion is presented in Paper V, whereas contributions to modelling of a radar sensor can be found in Paper I - IV. In Paper VI, a real-time implementation of a fusion system for automotive preventive safety system is presented.

2.3.2 Decision layer

Given the estimates of vehicle positions in the vicinity of the host vehicle including uncertainty measures, it is up the decision layer, i.e., preventive safety applications, to detect dangerous situations and make a decision whether to

inform, warn or intervene. This can typically be done in two stages, threat assessment and feature. The former is responsible for detecting hazards and calculate a corresponding threat level, e.g., how difficult it is to avoid a collision. The latter then makes a decision on an appropriate action and when this should be taken, e.g., warn the driver of an impending rear-end collision when there is still enough time for the driver to react or apply the brakes to mitigate a collision when the accident is unavoidable.

Depending on the intrusiveness of the decision, the requirement for false alarm rates could be very different. For example, if a warning application gives a false warning it is irritating for the driver (which ultimately leads to the driver switching off the system), whereas with an application that autonomously applies the brakes the consequences are much more serious. It all boils down to the information provided by the perception. In order to make intrusive/effective applications the reliability of the information supplied by the perception needs to be very good.

Even though this part of a preventive safety system is not covered in the thesis, it is important to know its needs in order to design effective perception algorithms.

2.3.3 Action layer

When the decision layer has decided on an appropriate action it needs to be executed. Depending upon if the chosen course of action is to inform or warn the driver or if it is to intervene, the action layer has to consider different aspects. For an intervention, it has to activate and manage the appropriate actuators that are to perform the intervention. In the case of a warning on the other hand, an intuitive message needs to be conveyed to the driver that persuades him/her to take appropriate action. Figure 2.3 shows an example of such a warning from a forward collision warning application, where red light is projected onto the windscreen in conjunction with an audible alert. The row of red lights is reminiscent of the brake lights of the preceding vehicle indicating that the driver should apply the brakes.

If there are several preventive safety applications integrated in the same

Figure 2.3: A warning of an impending rear-end collision issued as a row of red lights projected onto the windscreen.

vehicle, either requesting the attention of the driver or the actuators, these requests need to be managed. Research shows that integration of several information and warning applications in a vehicle have the could reduce the effectiveness of each individual application and increase the workload for the driver, [34]. It is also possible that applications may request conflictive actuator responses. In these cases there is a clear need for mechanisms to manage warnings and actuator requests. Such mechanisms for integrating multiple systems in the same vehicle are presented in [13].

2.4 Summary

To conclude the discussion, three types of in-vehicle safety systems have been applied to reduce the number of dead and injured on the roads.

protective safety system designed to passively protect the driver and passenger in the case of an accident. This could be crumple zones that are constructed to absorb energy at the time of collision or seat belts that hold the driver firmly.

preventive safety system is an active system that monitors the driver, host vehicle and/or surrounding traffic in order to apply active measures to avoid or mitigate an accident.

integrated safety system is a combination of the two above, where the protective safety system is enhanced using information about an impending accident from the preventive safety systems.

The remainder of this thesis, will mainly focus on the different aspects of preventive safety systems, although some of these aspects also apply for integrated safety systems.

Chapter 3

Bayesian tracking theory

A tracking system, consisting of one or several sensors and tracking algorithms, is an essential part of a preventive safety system. It supplies the decision algorithms with the perception of the surrounding traffic environment based on the observations made by the sensors. External object sensor systems, such as radar, lidar and vision, report (noisy) measurements originating from a variety of sources. The measurements could originate from objects of interest, referred to as targets[1], but also clutter or spurious detections due to, e.g.,, thermal noise in the sensor. The aim of the tracking algorithms is to detect and confirm the true targets and collect the set of observations that originate from the same target into tracks. From these tracks, the quantities of interest, such as target position and motion, future predicted path and object characteristics, are estimated.

This chapter tries to cover the basic Bayesian tracking theory needed to solve this problem. Good text books that cover this topic or at least important parts of it are, e.g. [12, 45, 7]. In this thesis, we focus the discussion on the state space approach to model the quantities of interest and focus on the discrete-time representation of the problem. That is, all observations are collected at discrete-times t_k, where $k \in \mathbb{N}$ denotes the discrete time index.

[1]The expression originates from military applications of tracking theory and also be is extensively used in the literature. The term target will used in this thesis, even though preventive safety systems are more concerned with avoiding an object rather then hitting a target.

The time between two consecutive observations is specified by the possibly time dependent sampling interval, $T_s(k) \triangleq t_k - t_{k-1}$. At time k, the complete information required to describe the system is summarised in the *discrete-time state vector* $\mathbf{x}_k \in \mathbb{R}^n_x$. In tracking applications, $\mathbf{x}_k$ typically includes kinematic information about the targets of interest and can be divided into sub-vectors, representing the state for each target,

$$\mathbf{x}_k = [(\mathbf{x}_k^1)^T, (\mathbf{x}_k^2)^T, \ldots, (\mathbf{x}_k^{n_t(k)})^T]^T, \tag{3.1}$$

where $n_t(k)$ is the number of targets at time k.

Noisy observations related to $\mathbf{x}_k$ are supplied to the tracking algorithm (at each time instance) in the form of a measurement vector, $\mathbf{y}_k \in \mathbb{R}^{n_y}$. The complete observed information available on the system up to and including time k is collected in,

$$\mathbf{Y}_{1:k} = [(\mathbf{y}_1)^T, (\mathbf{y}_2)^T, \ldots, (\mathbf{y}_k)^T]^T. \tag{3.2}$$

The aim of a Bayesian tracking algorithm is to calculate the joint probability density of all target-positions given all available information, i.e., to find the *posterior probability density function (pdf)*,

$$p(\mathbf{x}_k \big| \mathbf{Y}_{1:k}), \tag{3.3}$$

also called the posterior density in short. All available statistical information about the system is contained in this probability density and from this, as we will see later, it is possible to calculate an optimal estimate of $\mathbf{x}_k$, including measures on the estimation accuracy.

Preventive safety systems, as with many other systems, rely on obtaining an estimate of the quantity of interest, e.g., current traffic situation, as soon as new measurements are received. A convenient approach to render this solution, is a recursive tracking algorithm. In this formulation, it is not necessary to store previous data nor to reprocess old data as new data become available[2]. Figure 3.1 illustrates this approach for a complete multi-target

[2]Assuming that there are no measurements arriving out of sequence, *i.e* measurements

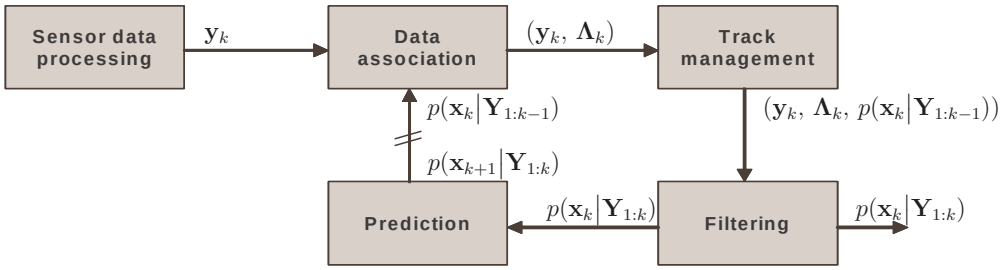


Figure 3.1: Sequential multi-target tracking process. Here, $\mathbf{y}_k$ represents the measurements on the system, $\mathbf{\Lambda}_k$ contains the measurement to track associations, $p(\mathbf{x}_k|\mathbf{Y}_{1:k-1})$ and $p(\mathbf{x}_{k+1}|\mathbf{Y}_{1:k})$ are respectively the prediction densities at the beginning and at the end of a filter iteration and $p(\mathbf{x}_k|\mathbf{Y}_{1:k})$ is the desired posterior density.

tracking system [12] and the basic blocks are briefly described below. A more detailed description of the blocks are given in the coming sections.

Data association: as new measurements, $\mathbf{y}_k$, arrive they are associated with the existing tracks, using their predicted densities, $p(\mathbf{x}_k^i|\mathbf{Y}_{1:k-1})$, i.e., the densities of each target, i, at time k based on measurements up to and including time $k-1$. Using this distribution, this block is able to form the most likely connections between the new measurements and existing tracks. These associations are specified in the data association matrix, $\mathbf{\Lambda}_k$, such that if $\mathbf{\Lambda}_k(i,j) = 1$ if measurement j is associated with track i.

Track management: in each iteration in the tracking system there is an uncertainty regarding how many tracks that are present. The track management is responsible for ensuring that tracks gets initiated and validated if data indicates that a new track is present. It also ensures that tracks that are no longer observed or exhibit unlikely behaviour, get deleted.

Filtering: the old and newly initiated tracks are updated with their associated measurements to calculate the desired posterior density, $p(\mathbf{x}_k|\mathbf{Y}_{1:k})$.

that were made prior to the time of the current state estimate.

Prediction: to align the known tracks in time with the expected new measurements arriving at time $k+1$, the tracks are predicted and $p(\mathbf{x}_{k+1}|\mathbf{Y}_{1:k})$ is calculated. This concludes the track update cycle and we are back at the beginning, except that the information in $\mathbf{y}_k$ is now incorporated in the tracks.

This chapter continues by examining the recursive Bayesian filtering problem in Section 3.1. Here, the conceptual formulation of the problem is presented together with exact and approximative solutions. In Section 3.2, two statistical models for solving the recursive tracking problem are introduced, i.e., the motion model and the measurement model. The former describes how the state vector evolves over time and is mainly used in the prediction module in Fig. 3.1. The latter relates the state to the noisy measurements and is used both in the data association and the filtering blocks. Different aspects of the observation-to-track problem are discussed in Section 3.3 and the chapter is concluded with different schemes to handle initiation and deletion of tracks in Section 3.4.

3.1 Recursive filtering problem

In this section, the recursive Bayesian filtering problem is presented. Here it is assumed that there are no association uncertainties and we disregard all track handling considerations. We start by introducing the general solution to the problem in Section 3.1.1 and continue with showing how it can be solved analytically under linear and Gaussian assumptions, resulting in the well known Kalman filter presented in Section 3.1.2. However, for many practical problems, the assumptions used to derive the Kalman filter are too restrictive. To alleviate some of these restrictions, approximative solutions to the general tracking problem are presented in Sections 3.1.3 - 3.1.5.

3.1.1 General recursive filtering problem

The problem at hand is to recursively calculate the posterior probability density function, $p(\mathbf{x}_k|\mathbf{Y}_{1:k})$. In the Bayesian framework, the quantities of

interest are seen as unknown stochastic variables. In this setting it is possible, and necessary, to describe *a-priori* information about the random parameter in the form of a prior density

$$p(\mathbf{x}_0|\mathbf{y}_0) \triangleq p(\mathbf{x}_0), \tag{3.4}$$

i.e., the initial information about the parameter before any observations are done. A discussion of designing a suitable prior can be found in 3.4.1.

Furthermore, to recursively estimate the target positions we require at least two probabilistic models. One describing the evolution of the state vector, called the motion model, and one relating the noisy measurements to the state, called measurement or sensor model. We assume that the former can be written as

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}), \tag{3.5}$$

where $\mathbf{f}_{k-1}(\cdot)$ is a possibly nonlinear and time dependent function and $\mathbf{v}_{k-1}$ is a stochastic noise process explaining unforseen deviations and perturbations from the expected behaviour. This relation is used to specify the conditional pdf $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, i.e., the probability density of $\mathbf{x}_k$ given that the previous position $\mathbf{x}_{k-1}$ is known.

The measurement model is similarly formed as

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{w}_k), \tag{3.6}$$

where $\mathbf{h}_k(\cdot)$ is a possibly nonlinear and time-dependent function describing how the state is related to the measurements, and $\mathbf{w}_k$ is a stochastic noise process to account for measurement noise and sensor modelling uncertainty. As for the motion model, the measurement model has a probabilistic interpretation. In this sense, (3.6) is regarded as the likelihood function of $\mathbf{x}_k$ given the measurement $\mathbf{y}_k$, denoted $p(\mathbf{y}_k|\mathbf{x}_k)$[3] or $\mathcal{L}(\mathbf{x}_k|\mathbf{y}_k)$. Often, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and $p(\mathbf{y}_k|\mathbf{x}_k)$ are used interchangeably with (3.5) and (3.6), respectively. Further-

[3]Note that the likelihood $p(\mathbf{y}_k|\mathbf{x}_k)$ should be seen as a function of $\mathbf{x}_k$ and is not a probability density function. Hence, the motivation for the alternative notation which is clearer in this way.

more, both noise process, $\mathbf{v}_k$ and $\mathbf{w}_k$, are assumed to be white and mutually independent.

Assuming that the density $p(\mathbf{x}_{k-1}|\mathbf{Y}_{1:k-1})$ from the previous recursion is known, the posterior density can be obtained in two steps, prediction and measurement update. The prediction step involves using (3.5) and the *Chapman-Kolmogorov equation* (or law of total probability) to calculate,

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{Y}_{1:k-1}) &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{Y}_{1:k-1}) p(\mathbf{x}_{k-1}|\mathbf{Y}_{1:k-1}) = \\ &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{Y}_{1:k-1}) d\mathbf{x}_{k-1}. \end{aligned} \tag{3.7}$$

Here we use the fact that $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{Y}_{1:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$ as (3.5) stipulates that all information needed to calculate $\mathbf{x}_k$ is summarised in the previous state $\mathbf{x}_{k-1}$. In statistics, it is stated that (3.5) is a *Markov process* of order one.

In the succeeding update step, the predicted density in (3.7) is updated using the information in the new measurement, $\mathbf{y}_k$, to form the desired posterior distribution. According to Bayes rule,

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{Y}_{1:k}) &= p(\mathbf{x}_k|\mathbf{y}_k, \mathbf{Y}_{1:k-1}) = \\ &= \frac{p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{Y}_{1:k-1}) p(\mathbf{x}_k|\mathbf{Y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{Y}_{1:k-1})} = \\ &= \frac{p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{Y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{Y}_{1:k-1})}, \end{aligned} \tag{3.8}$$

where the normalising constant in the denominator can be obtained by marginalising known distributions as

$$p(\mathbf{y}_k|\mathbf{Y}_{1:k-1}) = \int p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{Y}_{1:k-1}) d\mathbf{x}_k. \tag{3.9}$$

Starting from (3.4), it is possible to recursively calculate $p(\mathbf{x}_k|\mathbf{Y}_{1:k})$ for all k by repeatedly using (3.5) to predict the prior to the time of the measurement according to (3.7) and updating with new information, in the form of a measurement, using (3.6) according to (3.8).

In the Bayesian sense it is possible to calculate the optimal estimate of $\mathbf{x}_k$, given the posterior distribution $p(\mathbf{x}_k|\mathbf{Y}_{1:k})$, under any optimality condition. It is possible to calculate the *minimum mean square error* (MMSE) estimate by taking the conditional expectation on $\mathbf{x}_k$ as,

$$\hat{\mathbf{x}}_{k|k}^{\text{MMSE}} = \mathbb{E}\{\mathbf{x}_k|\mathbf{Y}_{1:k}\} = \int \mathbf{x}_k p(\mathbf{x}_k|\mathbf{Y}_{1:k}) d\mathbf{x}_k, \tag{3.10}$$

or the *maximum a posteriori* (MAP) estimate as the maximum of $p(\mathbf{x}_k|\mathbf{Y}_{1:k})$,

$$\hat{\mathbf{x}}_{k|k}^{\text{MAP}} \triangleq arg \max_{\mathbf{x}_k} p(\mathbf{x}_k|\mathbf{Y}_{1:k}). \tag{3.11}$$

The optimal recursive solution to the Bayesian filtering problem is supplied by (3.7) and (3.8). However, in practice these expressions are seldom analytically solvable. Even if they were solvable for general distributions, it could imply the need to keep track of infinite many moments in order to describe $p(\mathbf{x}_k|\mathbf{Y}_{1:k})$. However, under special conditions/assumptions, it is possible to derive an optimal solution that is also tractable. If both the process model and the measurement model are linear with additive Gaussian noise the solution results in the well known Kalman filter, discussed further in Section 3.1.2. In Sections 3.1.3 - 3.1.5, different approximation methods are presented when the linear Gaussian assumptions do not hold.

3.1.2 Kalman Filter

In this section, an important case where (3.7) and (3.8) can be solved analytically is presented. This is accomplished by imposing the following assumptions on the system,

Assumption 1 (Linear and Gaussian models)

1. *the prior pdf,* $p(\mathbf{x}_0|\mathbf{y}_0)$*, is a Gaussian distribution*

2. *the process noise,* $\mathbf{v}_{k-1}$*, and the measurement noise,* $\mathbf{w}_k$*, processes are assumed to be independent for all* k *and Gaussian*

3. *both the process model and measurement model are assumed to depend*

linearly on $\mathbf{x}_k$, $\mathbf{v}_k$ *and* $\mathbf{w}_k$.

Using these assumptions the system model can be expressed as a linear system of equations,

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{v}_{k-1}, \tag{3.12}$$

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{w}_k, \tag{3.13}$$

where $\mathbf{v}_{k-1} \sim \mathcal{N}(0, \mathbf{Q}_{k-1})$ and $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_k)$[4]. Recursively solving (3.7) and (3.8) analytically under these assumptions results in the well known Kalman filter [28]. Before presenting the solution we discuss some interesting properties of Gaussian linear models which, as we will see, make them convenient to work with in this context. Proofs of these theorems can be found in standard statistical texts, such as [29].

Theorem 3.1 (Gaussian linear transformation)
For any linear transformation, $\mathbf{T}$*, of a Gaussian random vector,* $\mathbf{x}$*, such that*

$$\mathbf{y} = \mathbf{T}\mathbf{x}, \tag{3.14}$$

the resulting random vector is also Gaussian with

$$\mathbb{E}\{\mathbf{y}\} = \mathbf{T}\mathbb{E}\{\mathbf{x}\}, \tag{3.15}$$

$$Cov\{\mathbf{y}\} = \mathbf{T}Cov\{\mathbf{x}\}\mathbf{T}^T. \tag{3.16}$$

□

Theorem 3.2 (Sum of Gaussian random vectors)
If $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{P}_\mathbf{x})$ *and* $\mathbf{y} \sim \mathcal{N}(\bar{\mathbf{y}}, \mathbf{P}_\mathbf{y})$ *are two independent Gaussian vectors,* $\mathbf{z} = \mathbf{x} + \mathbf{y}$*, is also Gaussian with,*

$$\bar{\mathbf{z}} = \bar{\mathbf{x}} + \bar{\mathbf{y}}, \tag{3.17}$$

$$\mathbf{P}_\mathbf{z} = \mathbf{P}_\mathbf{x} + \mathbf{P}_\mathbf{y}. \tag{3.18}$$

[4]The Kalman filter can easily be expanded to handle non-zero mean noise vectors, although not covered here.

These two theorems give us the nice property that, under the linear and Gaussian assumptions, the transformations in (3.12) and (3.13) will always produce Gaussian random vectors with statistics that are easily calculated.

Theorem 3.3 (Conditional pdf of multivariate Gaussian)
If two random vectors, $\mathbf{x}$ *and* $\mathbf{y}$*, are jointly Gaussian with mean vector* $[\mathbb{E}\{\mathbf{x}\}^T\ \mathbb{E}\{\mathbf{y}\}^T]^T$ *and covariance matrix*

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{xx} & \mathbf{P}_{xy} \\ \mathbf{P}_{yx} & \mathbf{P}_{yy} \end{bmatrix}. \tag{3.19}$$

then the conditional pdf $p(\mathbf{x}|\mathbf{y})$ *is also Gaussian with*

$$\mathbb{E}\{\mathbf{x}|\mathbf{y}\} = \mathbb{E}\{\mathbf{x}\} + \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}(\mathbf{y} - \mathbb{E}\{\mathbf{y}\}) \tag{3.20}$$

$$\mathbf{P}_{x|y} = \mathbf{P}_{xx} - \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}\mathbf{P}_{yx} \tag{3.21}$$

Note that Theorem 3.3 also stipulates that the MMSE estimate (3.10) of $\mathbf{x}$, given $\mathbf{y}$, for the Gaussian linear case can be written in the form,

$$\hat{\mathbf{x}}(\mathbf{y}) = \mathbb{E}\{\mathbf{x}|\mathbf{y}\} = \mathbb{E}\{\mathbf{x}\} + \mathbf{K}(\mathbf{y} - \mathbb{E}\{\mathbf{y}\}), \tag{3.22}$$

where the expected value of $\mathbf{x}$ is updated by the difference between the observed $\mathbf{y}$ and its expectation scaled with some gain, $\mathbf{K} = \mathbf{P}_{xy}\mathbf{P}_{yy}^{-1}$.

Theorem 3.4 (Additive conditional expectation)
The conditional expectation of the Gaussian random vector $\mathbf{x}$ *given two jointly Gaussian and independent random vectors* $\mathbf{y}_1$ *and* $\mathbf{y}_2$ *can be formulated as,*

$$\mathbb{E}\{\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2\} = \mathbb{E}\{\mathbf{x}|\mathbf{y}_1\} + [\mathbb{E}\{\mathbf{x}|\mathbf{y}_2\} - \mathbb{E}\{\mathbf{x}\}]\,. \tag{3.23}$$

This property of the conditional expectation follows from Theorem 3.3 and is convenient when finding a sequential formulation of the MMSE estimate of $\mathbf{x}$ based on two independent data sets. Theorem 3.4 could be interpreted as; if we have an MMSE estimate of $\mathbf{x}$ based on the data set $\mathbf{y}_1$ and an MMSE estimate of $\mathbf{x}$ based on the orthogonal (uncorrelated) data $\mathbf{y}_2$ the MMSE estimate based on both of these data sets can be found by simply adding them together.

Using the above properties of a linear Gaussian system and examining (3.7) and (3.8), it is clear that the following Gaussian densities are of interest,

$$p(\mathbf{x}_{k-1}\big|\mathbf{Y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_{k-1};\ \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}), \tag{3.24}$$

$$p(\mathbf{x}_{k}\big|\mathbf{Y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_{k};\ \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}), \tag{3.25}$$

$$p(\mathbf{x}_{k}\big|\mathbf{Y}_{1:k}) = \mathcal{N}(\mathbf{x}_{k};\ \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}), \tag{3.26}$$

and consequentially only the mean and the covariance of these distributions need to be calculated. Assuming that (3.24) is available for the previous recursion and using (3.12), the prediction step (3.7) resolves to,

$$\hat{\mathbf{x}}_{k|k-1} = \mathbb{E}\{\mathbf{x}_k\big|\mathbf{Y}_{1:k-1}\} = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1} \tag{3.27}$$

$$\mathbf{P}_{k|k-1} = \mathrm{Cov}\{\mathbf{x}_k\big|\mathbf{Y}_{1:k-1}\} = \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}. \tag{3.28}$$

Before disclosing the measurement update step, we introduce the innovation in the new measurement,

$$\tilde{\mathbf{y}}_k \triangleq \mathbf{y}_k - \mathbb{E}\{\mathbf{y}_k|\mathbf{Y}_{1:k-1}\}, \tag{3.29}$$

which can be viewed as the new information in $\mathbf{y}_k$ that is uncorrelated with $\mathbf{Y}_{1:k-1}$ (and hence $\hat{\mathbf{x}}_{k|k-1}$). Using Theorem 3.1 and 3.2 it is easy to show that $\tilde{\mathbf{y}}_k$ is Gaussian with zero mean and conditional covariance

$$\mathbf{S}_k = \mathrm{Cov}\{\tilde{\mathbf{y}}_k\big|\mathbf{Y}_{1:k-1}\} = \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k. \tag{3.30}$$

Using the innovation, it is possible to find a convenient decomposition of the updated estimate,

$$\mathbb{E}\{\mathbf{x}_k \big| \mathbf{Y}_{1:k}\} = \mathbb{E}\{\mathbf{x}_k \big| \tilde{\mathbf{y}}_k, \mathbf{Y}_{1:k-1}\} \tag{3.31}$$

and using Theorem 3.3

$$\begin{aligned}\mathbb{E}\{\mathbf{x}_k \big| \tilde{\mathbf{y}}_k, \mathbf{Y}_{1:k-1}\} &= \mathbb{E}\{\mathbf{x}_k \big| \mathbf{Y}_{1:k-1}\} + \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{yy}}^{-1}(\tilde{\mathbf{y}}_k - \mathbb{E}\{\tilde{\mathbf{y}}_k\}) = \\ &= \mathbb{E}\{\mathbf{x}_k \big| \mathbf{Y}_{1:k-1}\} + \mathbf{P}_{\mathbf{xy}}\mathbf{S}_k^{-1}(\mathbf{y}_k - \mathbb{E}\{\mathbf{y}_k | \mathbf{Y}_{1:k-1}\}) \end{aligned} \tag{3.32}$$

Evaluating (3.32) using (3.12) and (3.13) results in the Kalman filter update,

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}) \tag{3.33}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k(\mathbf{S}_k)^{-1}\mathbf{K}_k^T, \tag{3.34}$$

where

$$\mathbf{K}_k = \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{yy}}^{-1} = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1} \tag{3.35}$$

is known as the Kalman gain matrix.

Clearly, the state estimate $\hat{\mathbf{x}}_{k|k}$, calculated using the Kalman filter is the optimal state estimate in the MMSE sense, under the linear Gaussian assumptions. It is also possible to show that the Kalman filter is the best linear estimator, or *linear minimum mean square error* (LMMSE) estimator, even if the process and measurement noise is non-Gaussian [29].

Example 3.1 (Scalar Kalman filter)
To demonstrate the different steps in the Kalman filter we use a simple example of tracking a scalar quantity, x_k, which evolves according to this very simple model,

$$x_k = 1.2x_{k-1} + v_{k-1}, \tag{3.36}$$

where $v_{k-1} \sim \mathcal{N}(0, 0.1)$. We also make noisy observations, y_k which relate to

the state as,

$$y_k = x_k + w_k, \tag{3.37}$$

where $w_k \sim \mathcal{N}(0, 0.5)$. *From the previous iteration of the filter we have the density,*

$$x_{k-1}|\mathbf{Y}_{1:k-1} \sim \mathcal{N}(1, .3). \tag{3.38}$$

Using the models described by (3.36) - (3.38) *we have all the information we need to construct a Kalman filter. Suppose we at time* k *receive a measurement,* $y_k = 1.5$, *the first iteration of the Kalman filter becomes,*

$$\mathbf{x}_k \big| \mathbf{Y}_{1:k-1} \sim \mathcal{N}(1.2,\, 0.53) \tag{3.39}$$

$$\mathbf{x}_{k|k} \big| \mathbf{Y}_{1:k-1} \sim \mathcal{N}(1.4,\, 0.26) \tag{3.40}$$

The different steps in the Kalman filter to arrive at this posterior density are illustrated in Fig. 3.2. Note that the predicted density is much flatter than the prior as uncertainty is introduced in the prediction. However, in the resulting posterior density the uncertainty decreased is due to the measurement update.

□

3.1.3 The extended Kalman filter

For many practical filtering applications the assumptions made to derive the classic Kalman filter are too restrictive. Typically, either the motion or the measurement models (or both) are not accurately described as linear models. In addition, if more accurate models exist that depart from the linear and Gaussian assumption, there is often no analytical solution to (3.7) and (3.8). In these cases, suboptimal methods for approximatively solving the Bayesian filtering equations should be used. The extended Kalman filter (EKF) is one such suboptimal method, to expand the Kalman filter framework to also handle nonlinear models. The general idea is to linearise the nonlinear models

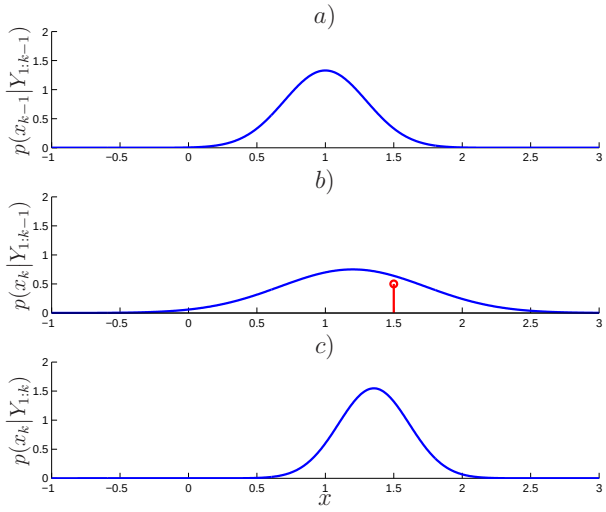


Figure 3.2: a) The prior density $p(x_{k-1}|Y_{1:k-1})$. b) Predicted density $p(x_{k-1}|Y_{1:k-1})$ in blue and the measurement y_k in red. c) The posterior density $p(x_{k-1}|Y_{1:k-1})$ after the measurement update.

by first order Taylor expansion. Using the first order Taylor approximation of the nonlinear models and assuming independent white Gaussian noise ensures that the linear Gaussian assumptions hold for the linearised system.

Assume that we have models in the general form as in (3.5) and (3.6), where $\mathbf{v}_{k-1}$ is zero-mean process noise with covariance matrix $\mathbf{Q}_{k-1}$ and $\mathbf{w}_k$ is zero-mean measurement noise with covariance matrix $\mathbf{R}_k$. By linearising the models using first order Taylor expansion around their conditional expectations we can approximate the nonlinear models as,

$$\begin{aligned}\mathbf{x}_k &\approx \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{v}_{k-1})\big|_{\mathbf{x}_{k-1}=\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{v}_{k-1}=\mathbf{0}} + \\ &\quad + \widehat{\mathbf{F}}_{\mathbf{x}_{k-1}}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}) + \widehat{\mathbf{F}}_{\mathbf{v}_{k-1}}(\mathbf{v}_{k-1} - \mathbf{0}), \qquad (3.41)\\ \mathbf{y}_k &\approx \mathbf{h}_k(\mathbf{x}_k, \mathbf{w}_k)\big|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}, \mathbf{w}_k=\mathbf{0}} + \\ &\quad + \widehat{\mathbf{H}}_{\mathbf{x}_k}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) + \widehat{\mathbf{H}}_{\mathbf{w}_k}(\mathbf{w}_k - \mathbf{0}), \qquad (3.42)\end{aligned}$$

where

$$\widehat{\mathbf{F}}_{\mathbf{x}_{k-1}} = [\nabla_{\mathbf{x}_{k-1}} \mathbf{f}_{k-1}^T(\mathbf{x}_{k-1}, \mathbf{v}_{k-1})]^T \big|_{\mathbf{x}_{k-1}=\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{v}_{k-1}=\mathbf{0}} \tag{3.43}$$

$$\widehat{\mathbf{F}}_{\mathbf{v}_{k-1}} = [\nabla_{\mathbf{v}_{k-1}} \mathbf{f}_{k-1}^T(\mathbf{x}_{k-1}, \mathbf{v}_{k-1})]^T \big|_{\mathbf{x}_{k-1}=\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{v}_{k-1}=\mathbf{0}} \tag{3.44}$$

$$\widehat{\mathbf{H}}_{\mathbf{x}_k} = [\nabla_{\mathbf{x}_k} \mathbf{h}_k^T(\mathbf{x}_k, \mathbf{w}_k)]^T \big|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}, \mathbf{w}_k=\mathbf{0}} \tag{3.45}$$

$$\widehat{\mathbf{H}}_{\mathbf{w}_k} = [\nabla_{\mathbf{w}_k} \mathbf{h}_k^T(\mathbf{x}_k, \mathbf{w}_k)]^T \big|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}, \mathbf{w}_k=\mathbf{0}}. \tag{3.46}$$

Using these linear approximative models and applying them in the Kalman filter framework results in the following recursive estimator:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}_{k-1}(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{0}) \tag{3.47}$$

$$\mathbf{P}_{k|k-1} = \widehat{\mathbf{F}}_{\mathbf{x}_{k-1}} \mathbf{P}_{k-1|k-1} \widehat{\mathbf{F}}_{\mathbf{x}_{k-1}}^T + \widehat{\mathbf{F}}_{\mathbf{v}_{k-1}} \mathbf{Q}_{k-1} \widehat{\mathbf{F}}_{\mathbf{v}_{k-1}}^T \tag{3.48}$$

$$\hat{\mathbf{x}}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \widehat{\mathbf{H}}_{\mathbf{x}_k} \hat{\mathbf{x}}_{k|k-1}) \tag{3.49}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T, \tag{3.50}$$

where

$$\mathbf{S}_k = \widehat{\mathbf{H}}_{\mathbf{x}_k} \mathbf{P}_{k|k-1} \widehat{\mathbf{H}}_{\mathbf{x}_k}^T + \widehat{\mathbf{H}}_{\mathbf{w}_k} \mathbf{R}_k \widehat{\mathbf{H}}_{\mathbf{w}_k}^T \tag{3.51}$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \widehat{\mathbf{H}}_{\mathbf{x}_k}^T \mathbf{S}_k^{-1} \tag{3.52}$$

In order for the first order EKF to be applicable, the estimation error, $\tilde{\mathbf{x}}_k = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k}$, must be kept sufficiently small. This is something that is difficult to guarantee, as the estimation error can build up over time. If it does not hold, the estimate will suffer in accuracy and it could lead to instability issues of the filter. There are also cases when the EKF is an inconsistent estimator, i.e., the estimator error is not zero-mean with a covariance that matches that calculated by the filter [46]. One should also mention that it is often tedious to compute (3.43) - (3.46).

There have been attempts to improve the performance of the EKF, e.g., by including a second-order term in the Taylor expansion [4], or by iteratively finding the linearisation point in the measurement update step [7]. The reasoning behind the latter improvement is that you have more information

about $\mathbf{x}_k$ after the measurement update step and are therefore able to choose a better linearisation point resulting in a more applicable linearisation.

One should bear in mind that if (3.5) and (3.6) are mildly nonlinear, and/or the noise is relatively small (compared to the nonlinearity), the EKF has proven in practise to produce sufficiently accurate estimates. It is after all probably the most widely used estimator for nonlinear problems.

Example 3.2 (Linearisation of a nonlinear motion model)
This example will try to demonstrate the effects that linearisation has on the estimated prediction density, $p(\mathbf{x}_{k|k-1}|\mathbf{Y}_{1:k-1})$, if the motion model is nonlinear. We are interested in describing the prediction density of the discrete-time state vector,

$$\mathbf{x}_k = \begin{bmatrix} \xi_x(k) \\ \xi_y(k) \\ \psi(k) \\ v(k) \\ \dot{\psi}(k) \\ a(k) \end{bmatrix}, \tag{3.53}$$

where (ξ_x, ξ_y) is the position of the vehicle in global coordinates, ψ is the heading of the vehicle, v and a are the velocity and acceleration in that direction, respectively, and $\dot{\psi}$ is the yaw rate of the vehicle. To describe how $\mathbf{x}_k$ evolves as a function of time, we use the simplified bicycle model as in [10, 22]. In discrete time this motion model has the following approximative

form:

$$\xi_x(k) = \xi_x(k-1) + v(k-1)T_s\cos(\psi(k-1)) \cdot \alpha_1(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}, T_s) + \sin(\psi(k-1)) \cdot \alpha_2(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}, T_s), \tag{3.54}$$

$$\xi_y(k) = \xi_y(k-1) + v(k-1)T_s\sin(\psi(k-1)) \cdot \alpha_1(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}, T_s) + \cos(\psi(k-1)) \cdot \alpha_2(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}, T_s), \tag{3.55}$$

$$\psi(k) = \psi(k-1) + \dot{\psi}(k-1)T_s + v_{\ddot{\psi}}\frac{T_s^2}{2} \tag{3.56}$$

$$v(k) = v(k-1) + a(k-1)T_s + v_{\dot{a}}\frac{T_s^2}{2} \tag{3.57}$$

$$\dot{\psi}_k = \dot{\psi}(k-1) + v_{\ddot{\psi}}T_s \tag{3.58}$$

$$a(k) = a(k-1) + v_{\dot{a}}T_s. \tag{3.59}$$

where $\alpha_i(\cdot)$ *are nonlinear functions of the previous state* $\mathbf{x}_{k-1}$*, the noise process,* $\mathbf{v}_{k-1}$*, and the sample time* T_s*. Furthermore,* $\mathbf{v}_{k-1} = \left[v_{\ddot{\psi}},\, v_{\dot{a}}\right]^T$ *is a zero-mean and Gaussian noise process with covariance matrix*

$$\mathbf{Q}_{k-1} = \begin{bmatrix} q_{\ddot{\psi}} & 0 \\ 0 & q_{\dot{a}} \end{bmatrix}. \tag{3.60}$$

To appreciate this example, is it not important to fully understand the proposed motion model. However, what is important to understand is that the position of the vehicle at time k*,* $(\xi_x(k), \xi_y(k))$*, is a nonlinear function of* $\mathbf{x}_{k-1}$ *and* $\mathbf{v}_{k-1}$*. For example, uncertainty in the heading of the vehicle,* ψ*, will propagate through the trigonometric functions to the position (see* (3.54) *and* (3.55)*).*

Assume that we at time k *have a prior distribution*

$$p(\mathbf{x}_{k-1}\big|\mathbf{Y}_{1:k-1}) \approx \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}), \tag{3.61}$$

where

$$\hat{\mathbf{x}}_{k-1|k-1} = [0,\ 0,\ \pi/6,\ 25,\ \pi/2,\ 4]^T, \tag{3.62}$$

$$\mathbf{P}_{k-1|k-1} = diag([1, 1, \pi/15, 1, \pi/6, 6]). \tag{3.63}$$

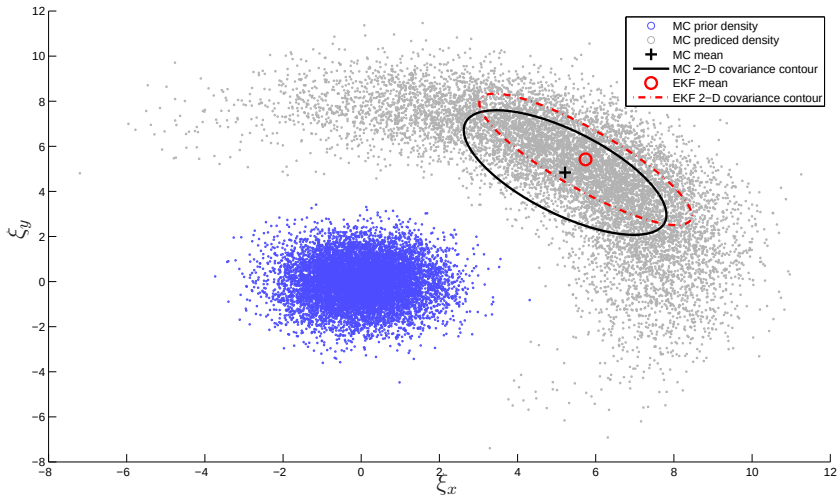


Figure 3.3: Comparison of the true mean and covariance to that estimated by EKF of a Gaussian distribution propagated through a nonlinear function, (3.54)-(3.59). In the figure is only the (ξ_x, ξ_y) components of the state vector displayed. In the simulations we use $T_s = 0.3s$ and $\mathbf{Q}_{k-1} = \text{diag}([0.1, 1])$.

Suppose the tedious work of calculating the Jacobians (3.43) *and* (3.44) *for the discrete-time motion model* (3.54) - (3.59) *is already done, we can estimate the mean and covariance of the predicted density* $p(\mathbf{x}_k | \mathbf{Y}_{1:k-1})$ *according to* (3.47) - (3.48). *In Figure 3.3 this estimated mean and covariance is compared to the "true" mean and covariance estimated through Monte Carlo simulations. From the simulations it is clear that the* EKF *estimate has a significant bias in the mean, and the covariance is under estimated for this highly nonlinear example.*

□

3.1.4 Unscented Kalman Filter

Relatively recently, another extension of the Kalman filter framework has been proposed [24], called the *unscented Kalman filter (*UKF*)*. Instead of linearising the nonlinear functions $\mathbf{f}_{k-1}(\cdot)$ and $\mathbf{h}_k(\cdot)$, this approach uses the *unscented transform* to approximate moments of $p(\mathbf{x}_k | \mathbf{Y}_{1:k-1})$ and $p(\mathbf{y}_k | \mathbf{Y}_{1:k-1})$.

The estimate of $\hat{\mathbf{x}}_{k|k}$ is then formed as an LMMSE estimator using the estimated moments from the unscented transform [37].

We will start by explaining the unscented transform and continue with showing how it is used in an LMMSE estimator.

Unscented transform

The basic idea behind the unscented transform (UT) is that it is easier to approximate a probability distribution than it is to approximate an arbitrary nonlinear function or transformation. The principle is; for any given probability distribution, $p(\mathbf{x})$, a set of points, called *sigma points*, are deterministically selected such that their mean and covariance equals that of $\mathbf{x} \in \mathbb{R}^{n_\mathbf{x}}$. By transforming these sigma points through a nonlinear mapping, $g(\cdot)$, statistics about $p(\mathbf{z}) = p(g(\mathbf{x}))$ can be estimated from these transformed points.

We denote a set of sigma points with their associated weights as $\mathcal{S} = \{\mathcal{X}^{(i)}, W^{(i)}, i = 0 \dots p\}$, where $W^{(i)}$ is such that

$$\sum_{i=0}^{p} W^{(i)} = 1. \tag{3.64}$$

Note that $W^{(i)}$ can take both positive or negative values, but to ensure an unbiased estimate they all have to sum to one. The sigma points should be chosen in such a way that they capture at least[5] the mean and the covariance of $\mathbf{x}$, i.e., these two conditions should hold,

$$\bar{\mathbf{x}} = \mathbb{E}\{\mathbf{x}\} = \sum_{i=0}^{p} W^{(i)} \mathcal{X}^{(i)}, \tag{3.65}$$

$$\mathbf{P}_{\mathbf{xx}} = \mathrm{Cov}\{\mathbf{x}\} = \sum_{i=0}^{p} W^{(i)} (\mathcal{X}^{(i)} - \bar{\mathbf{x}})(\mathcal{X}^{(i)} - \bar{\mathbf{x}})^T. \tag{3.66}$$

[5]There are methods of choosing sigma points that claim to also capture higher moments under certain conditions [27].

By transforming the sigma points through the nonlinear mapping, $\mathbf{z} = g(\mathbf{x})$,

$$\mathcal{Z}^{(i)} = g(\mathcal{X}^{(i)}). \tag{3.67}$$

The mean and the covariance of $\mathbf{z}$ can be estimated according to

$$\bar{\mathbf{z}} \approx \sum_{i=0}^{p} W^{(i)} \mathcal{Z}^{(i)}, \tag{3.68}$$

$$\mathbf{P}_{\mathbf{zz}} \approx \sum_{i=0}^{p} W^{(i)} (\mathcal{Z}^{(i)} - \bar{\mathbf{z}})(\mathcal{Z}^{(i)} - \bar{\mathbf{z}})^T. \tag{3.69}$$

If the sigma points are selected such that (3.65) and (3.66) hold, it is claimed in [25] that the estimates in (3.68) is accurate to the second order and (3.69) to the first order of the Taylor expansion of the true mean and covariance.

There are different ways of choosing the set of sigma points, depending on which statistics of the transformed distribution that should be captured. In [27] there is a discussion about these different methods as well as derivations of their properties. Here the method called symmetric set is listed. The idea is to select one point in the mean and $2n_x$ points symmetrically and equally spaced (in a statistical sense) on a contour of $\mathbf{P}_{\mathbf{x}}$, inversely proportional to their associated weight. The method is as follows

$$\begin{aligned}
&\mathcal{X}^{(0)} = \bar{\mathbf{x}}, && W^{(0)} = \tfrac{\kappa}{n_{\mathbf{x}}+\kappa}, && i = 0 \\
&\mathcal{X}^{(i)} = \bar{\mathbf{x}} + \left(\sqrt{(n_{\mathbf{x}}+\kappa)\mathbf{P}_{\mathbf{x}}}\right)_i, && W^{(i)} = \tfrac{1}{2(n_{\mathbf{x}}+\kappa)}, && i = 1, \ldots, n_{\mathbf{x}} \\
&\mathcal{X}^{(i)} = \bar{\mathbf{x}} - \left(\sqrt{(n_{\mathbf{x}}+\kappa)\mathbf{P}_{\mathbf{x}}}\right)_{(i-n_{\mathbf{x}})}, && W^{(i)} = \tfrac{1}{2(n_{\mathbf{x}}+\kappa)}, && i = (n_{\mathbf{x}}+1), \ldots, 2n_{\mathbf{x}}
\end{aligned} \tag{3.70}$$

where $n_{\mathbf{x}}$ is the dimension of the state vector and $\left(\sqrt{(n_{\mathbf{x}}+\kappa)\mathbf{P}_{\mathbf{x}}}\right)_i$ is the i^{th} column of any matrix $\mathbf{L}$ such that

$$\left(\sqrt{(n_{\mathbf{x}}+\kappa)\mathbf{P}_{\mathbf{x}}}\right) = \mathbf{L}\mathbf{L}^T. \tag{3.71}$$

The design parameter κ can be viewed as a scaling constant that determines

at which covariance contour the sigma points are positioned, as

$$|\mathcal{X}^{(i)} - \bar{\mathbf{x}}| \propto \sqrt{n_{\mathbf{x}} + \kappa}. \tag{3.72}$$

Setting $\kappa > 0$ tend to position the points further from the mean and choosing a $\kappa < 0$ will do the opposite. This could be used to abate the scaling effects caused when the dimension of the state vector, $n_{\mathbf{x}}$, is large. If the sigma points are too far apart the risk is that too much importance is given to aspects of $g(\cdot)$ that are far from the mean. This results in a covariance estimate that is perhaps too general, and does not accurately describe the covariance locally, especially if $g(\cdot)$ has discontinuities. Setting $\kappa = 3 - n_{\mathbf{x}}$ will result in a dimensional independent scaling, which is the optimal setting if $\mathbf{x}$ is a Gaussian [27].

An alternative view on κ, is that it regulates how much weight is given to the sigma point assigned to the mean, i.e.,

$$W^{(0)} = \frac{\kappa}{(n_{\mathbf{x}} + \kappa)}. \tag{3.73}$$

Similar reasoning as above holds for this interpretation as well. If large weight (large κ) is put on the mean, the other sigma points compensate by moving further away and vice versa. Note, however, that setting $\kappa < 0$ will result in a negative weighting of $\mathcal{X}^{(0)}$. Although this is acceptable, as the sigma set does not represent a distribution, caution is needed as this could lead to a non-positive definite covariance estimate, see (3.69). An amendment to this problem is to use the scaled unscented transform, originally proposed in [26]. In this unscented transform the sigma point selection method is re-parameterised and a covariance compensation term is added to ensure that the covariance estimates are positive definite.

Compared to the method used in the EKF, where $g(\cdot)$ is first linearised using first order Taylor expansion and then $\bar{\mathbf{z}}$ and $\mathbf{P}_{\mathbf{zz}}$ are estimated, the UT propagates the sigma points through the nonlinear function and then makes the estimates. As the standard EKF truncates the Taylor series at the first order (assumes that all higher terms are small or zero), the estimated

mean and covariance is only accurate to the first order. In contrast, the UT manages to capture the mean accurately to the second order and the covariance to the first. Moreover, as there is no truncation, some parts of the higher order terms are still present [53]. The latter means that if you have information about the higher order moments of $\bar{\mathbf{z}}$ and $\mathbf{P}_{\mathbf{zz}}$ this can be used when selecting the κ parameter. It can be shown, by examining the Taylor expansion of (3.68) and (3.69), that the κ parameters determine how much of the higher order terms are included [25].

Filter equations

Again, we return to the general nonlinear filtering problem defined by (3.7) and (3.8). Here we will see how the UT can be used to approximate these distributions' mean and covariance, with the aim of finding an LMMSE estimator of $\mathbf{x}_k$.

From text books, such as [29], we know that the LMMSE estimator can be written on the form,

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{yy}}^{-1}(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}) \tag{3.74}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{yy}}^{-1}\mathbf{P}_{\mathbf{xy}}^{T} \tag{3.75}$$

where

$$\mathbf{P}_{\mathbf{xy}} = \mathbb{E}\{(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})^T\} \tag{3.76}$$

$$\mathbf{P}_{\mathbf{yy}} = \mathbb{E}\{(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})^T\}. \tag{3.77}$$

The relation above holds if and only if, $\hat{\mathbf{x}}_{k|k-1}$, is uncorrelated with the new information in $\tilde{\mathbf{y}}_k = (\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})$, where $\hat{\mathbf{x}}_{k|k-1}$ and $\hat{\mathbf{y}}_{k|k-1}$ are LMMSE estimates of $\mathbf{x}_k$ and $\mathbf{y}_k$ given $\mathbf{Y}_{1:k-1}$, respectively. To put things in perspective, it should be noted that the estimator in (3.74) - (3.75) in the linear and

Gaussian case, turn into the Kalman filter as

$$\mathbf{P}_{\mathbf{xy}} = \mathbf{P}_{k|k-1}\mathbf{H}_k^T \tag{3.78}$$

$$\mathbf{P}_{\mathbf{yy}} = \mathbf{S}_k \tag{3.79}$$

and the Kalman gain matrix can consequentially be identified as

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1} = \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{yy}}^{-1} \tag{3.80}$$

Here, we will demonstrate how the unscented transform can be used to make nonlinear estimates of the mean and covariance needed to calculate (3.74) and (3.75) and that the necessary LMMSE conditions hold. As usual, this is done in two steps, prediction and measurement update.

The first step is to estimate the prediction density assuming that $\hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$ are known. Here, we introduce an augmented state vector,

$$\mathbf{x}_k^a = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{v}_k \\ \mathbf{w}_{k+1} \end{bmatrix}, \tag{3.81}$$

of dimension $n_{\mathbf{x}}^a = n_{\mathbf{x}} + n_{\mathbf{v}} + n_{\mathbf{w}}$ and with corresponding covariance matrix

$$\mathbf{P}_{k|k}^a = \begin{bmatrix} \mathbf{P}_{k|k} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{k+1} \end{bmatrix}. \tag{3.82}$$

Using this notation we can represent $p(\mathbf{x}_{k-1}^a|\mathbf{Y}_{1:k-1})$ as the sigma points set

$$\mathcal{X}_{k-1|k-1}^{a,(0)} = \hat{\mathbf{x}}_{k-1|k-1}^a, \qquad W^{(0)} = \frac{\kappa}{(n_{\mathbf{x}}^a + \kappa)} \tag{3.83}$$

$$\mathcal{X}_{k-1|k-1}^{a,(i)} = \hat{\mathbf{x}}_{k-1|k-1}^a + \left(\sqrt{\gamma \mathbf{P}_{k-1|k-1}^a}\right)_i, \qquad i = 1, \ldots, n_{\mathbf{x}}^a \tag{3.84}$$

$$\mathcal{X}_{k-1|k-1}^{a,(i)} = \hat{\mathbf{x}}_{k-1|k-1}^a - \left(\sqrt{\gamma \mathbf{P}_{k-1|k-1}^a}\right)_{i-n_{\mathbf{x}}^a}, \; i = (n_{\mathbf{x}}^a + 1), \ldots, 2n_{\mathbf{x}}^a \tag{3.85}$$

$$W^{(i)} = \frac{1}{2(n_{\mathbf{x}}^a + \kappa)}, \qquad i = 1, \ldots, 2n_{\mathbf{x}}^a, \tag{3.86}$$

where $\gamma = n_{\mathbf{x}_k^a} + \kappa$ and

$$\hat{\mathbf{x}}_{k-1|k-1}^a = \begin{bmatrix} \hat{\mathbf{x}}_{k-1|k-1} \\ \mathbb{E}\{\mathbf{v}_{k-1}\} \\ \mathbb{E}\{\mathbf{w}_k\} \end{bmatrix}. \tag{3.87}$$

By forming the sigma point set with the augmented state vector we are able to propagate and estimate the possibly nonlinear influence that the process and measurement noise has on the state vector and measurement vector, respectively[6]. Note also that the sigma point set $\mathcal{X}_{k-1|k-1}^a$ can be divided partitioned as

$$\mathcal{X}_{k-1|k-1}^a = \begin{bmatrix} \mathcal{X}_{k-1|k-1}^x \\ \mathcal{X}_{k-1}^v \\ \mathcal{X}_k^w \end{bmatrix}. \tag{3.88}$$

The estimates $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ can be formed by propagating $\mathcal{X}_{k-1|k-1}^a$ through the augmented process model

$$\mathcal{X}_{k|k-1}^{x,(i)} = \mathbf{f}_{k-1}(\mathcal{X}_{k-1|k-1}^{x,(i)}, \mathcal{X}_{k-1|k-1}^{v,(i)}), \tag{3.89}$$

and approximate

$$\hat{\mathbf{x}}_{k|k-1} \approx \sum_{i=0}^{2n_{\mathbf{x}}^a} W^{(i)} \mathcal{X}_{k|k-1}^{x,(i)}, \tag{3.90}$$

$$\mathbf{P}_{k|k-1} \approx \sum_{i=0}^{2n_{\mathbf{x}}^a} W^{(i)} (\mathcal{X}_{k|k-1}^{x,(i)} - \hat{\mathbf{x}}_{k|k-1})(\mathcal{X}_{k|k-1}^{x,(i)} - \hat{\mathbf{x}}_{k|k-1})^T. \tag{3.91}$$

[6]If the process noise and/or measurement noise is additive, there is no need to augment the state vector. The noise influence can be exactly accounted for by adding the noise covariances to the estimated covariances of the non augmented sigma points.

The updated sigma point set is now

$$\mathcal{X}^a_{k|k-1} = \begin{bmatrix} \mathcal{X}^x_{k|k-1} \\ \mathcal{X}^v_{k-1} \\ \mathcal{X}^w_k \end{bmatrix}. \tag{3.92}$$

which can be used to estimate $\hat{\mathbf{y}}_{k|k-1}$ and the corresponding covariance $\mathbf{P_{yy}}$ but also the cross-covariance $\mathbf{P_{xy}}$. Again we apply the unscented transform,

$$\mathcal{Y}^{(i)}_{k|k-1} = \mathbf{h}_k(\mathcal{X}^{x,(i)}_{k|k-1}, \mathcal{X}^{w,(i)}_k) \tag{3.93}$$

$$\hat{\mathbf{y}}_{k|k-1} \approx \sum_{i=0}^{2n^a_{\mathbf{x}}} W^{(i)} \mathcal{Y}^{(i)}_{k|k-1} \tag{3.94}$$

$$\mathbf{P_{yy}} \approx \sum_{i=0}^{2n^a_{\mathbf{x}}} W^{(i)} (\mathcal{Y}^{(i)}_{k|k-1} - \hat{\mathbf{y}}_{k|k-1})(\mathcal{Y}^{(i)}_{k|k-1} - \hat{\mathbf{y}}_{k|k-1})^T \tag{3.95}$$

$$\mathbf{P_{xy}} \approx \sum_{i=0}^{2n^a_{\mathbf{x}}} W^{(i)} (\mathcal{X}^{x,(i)}_{k|k-1} - \hat{\mathbf{x}}_{k|k-1})(\mathcal{Y}^{(i)}_{k|k-1} - \hat{\mathbf{y}}_{k|k-1})^T \tag{3.96}$$

As $\hat{\mathbf{y}}_{k|k-1}$ is calculated using all available information up to and including time $k-1$, the innovation, $(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})$, and the predicted state estimate, $\hat{\mathbf{x}}_{k|k-1}$, are uncorrelated. Hence, we can use (3.74) and (3.75) to calculate

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{P_{xy}}\mathbf{P}^{-1}_{\mathbf{yy}}(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}) \tag{3.97}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{P_{xy}}\mathbf{P}^{-1}_{\mathbf{yy}}\mathbf{P}^T_{\mathbf{xy}} \tag{3.98}$$

which is the desired estimate and error covariance. To summarise the Unscented Kalman filter, it:

1. estimates $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ using the unscented transform, where also the process noise is propagated through $\mathbf{f}_{k-1}(\cdot)$

2. similarly calculates estimates of $\hat{\mathbf{y}}_{k|k-1}$, $\mathbf{P_{xy}}$ and $\mathbf{P_{yy}}$ using the updated sigma point set (3.92)

3. finally updates $\hat{\mathbf{x}}_{k|k}$ and $\mathbf{P}_{k|k}$ on the form of an LMMSE estimator.

Example 3.3 (Unscented transform of a nonlinear motion model) *Returning to the problem defined in Example 3.2, but instead of using linearisation of* (3.54) - (3.59) *we here apply the unscented transform to estimate* $p(\mathbf{x}_k|\mathbf{Y}_{1:k-1})$. *Defining the augmented mean and the covariance of the augmented state vector as,*

$$\hat{\mathbf{x}}^a_{k-1|k-1} = \begin{bmatrix} \hat{\mathbf{x}}^a_{k-1|k-1} \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{P}^a_{k-1|k-1} = \begin{bmatrix} \mathbf{P}_{k-1|k-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{k-1} \end{bmatrix}. \tag{3.99}$$

We select sigma points according to (3.83) - (3.86), *where*

$$\kappa = 3 - n_{\mathbf{x}^a_k} = -5 \tag{3.100}$$

in accordance with previous discussion on optimal choice for Gaussian distributions. Finally, we use (3.89) *to propagate the sigma points through the motion model and* (3.90) - (3.91) *to estimate the mean and covariance of the prediction density. As in Example 3.2, the result is compared to Monte Carlo estimates of the true mean and covariance shown in Figure 3.4, together with the propagated sigma points.*

It is clear from the figure that the unscented transform is capable of making an unbiased mean estimate as well as accurately capturing the covariance.

□

3.1.5 Particle Filters

The methods described previously all use mean and covariance to describe the posterior density, $p(\mathbf{x}_k|\mathbf{Y}_{1:k})$, either exactly as in the Kalman filer or as an approximation as in the EKF and UKF. In many cases the posterior density is adequately described by these two moments. The particle filter algorithms, however, offer another type of description which enables a general and straightforward parametrisation for a large a family of problems. These algorithms rely on *Monte Carlo approximations* and *importance sampling* to describe $p(\mathbf{x}_k|\mathbf{Y}_{1:k})$. An excellent book covering this subject is [45], and

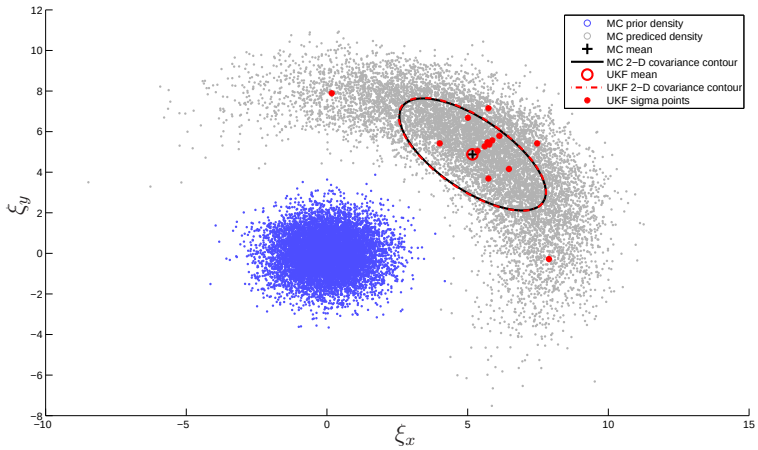


Figure 3.4: Comparison of the true mean and covariance to that estimated by UKF of a Gaussian distribution propagated through a non-linear function, (3.54) - (3.59). In the figure is only the (ξ_x, ξ_y) components of the state vector displayed. In the simulations we use $T_s = 0.3s$ and $\mathbf{Q}_{k-1} = \text{diag}([0.1, 1])$.

application of particle filters can be found in [3, 16, 15].

The principle of Monte Carlo approximation builds upon the notion that, besides the function itself also statistics of a distribution, i.e. moments, can be numerically approximated using samples of the distribution, also called *particles*.

Example 3.4 (Monte Carlo approximation)
Suppose we have $M \gg 1$ independent samples drawn from a pdf, $p(\mathbf{x})$ as,

$$\mathbf{x}^{(m)} \sim p(\mathbf{x}), \quad m = 1 \dots M, \tag{3.101}$$

we can approximate $p(\mathbf{x})$ and its general expectation function as,

$$p(\mathbf{x}) \approx \frac{1}{M} \sum_{m=1}^{M} \delta(\mathbf{x} - \mathbf{x}^{(m)}) \tag{3.102}$$

$$\mathbb{E}\{g(\mathbf{x})\} \approx \frac{1}{M} \sum_{m=1}^{M} \int g(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}^{(m)}) d\mathbf{x} = \frac{1}{M} \sum_{m=1}^{M} g(\mathbf{x}^{(m)}) \tag{3.103}$$

The accuracy of this approximation is dependent on the complexity of $p(\mathbf{x})$

and how many samples, M, that are used to describe it.

□

At first glance, Monte Carlo approximation may seem very similar to the aforementioned unscented transform. In this setting, however, the samples are drawn stochastically and not chosen deterministically, and typically the number of particles are far more numerous than the number of sigma points. The particles are generally able to capture higher moments of the distribution, whereas the sigma points are only designed to accurately describe the first two moments. In situations where the pdf is not nearly Gaussian (not accurately depicted by just mean and covariance) this will prove to be a very limiting restriction.

The method described in Example 3.4 relies upon the fact that it is possible to generate samples from the density, $p(\mathbf{x})$. Generally, this is not computationally straightforward except for some specific functions[7]. To circumvent this difficulty importance sampling can be used; that is instead of sampling directly from $p(\mathbf{x})$, generate samples from a similar distribution called the *importance density* $q(\mathbf{x})$. Using these samples, the approximations (3.102) and (3.103) can be expressed as,

$$p(\mathbf{x}) = \frac{p(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x}) \approx \frac{1}{M} \sum_{m=1}^{M} \frac{p(\mathbf{x}^{(m)})}{q(\mathbf{x}^{(m)})} \delta(\mathbf{x} - \mathbf{x}^{(m)}) = \sum_{m=1}^{M} w^{(m)} \delta(\mathbf{x} - \mathbf{x}^{(m)}) \tag{3.104}$$

$$\mathbb{E}\{g(\mathbf{x})\} \approx \sum_{m=1}^{M} \int g(\mathbf{x}) w^{(m)} \delta(\mathbf{x} - \mathbf{x}^{(m)}) d\mathbf{x} = \sum_{m=1}^{M} w^{(m)} g(\mathbf{x}^{(m)}), \tag{3.105}$$

where $w^{(m)} = \frac{p(\mathbf{x}^{(m)})}{M q(\mathbf{x}^{(m)})}$ is a (sample dependent) importance weight to compensate that we are not sampling from the desired distribution. Depending on our knowledge of $p(\mathbf{x})$, is it sometimes only possible to calculate $\tilde{w}^{(m)} \propto w^{(m)}$.

[7]For instance, in the case when $\mathbf{x}$ has a Gaussian or uniform distribution for example, one can in MATLAB easily use the functions `randn()` and `rand()`, respectively, to generate samples from these distributions.

However, it is easily amended by normalisation,

$$w^{(m)} = \frac{\tilde{w}^{(m)}}{\sum_{m=1}^{M} \tilde{w}^{(m)}} \tag{3.106}$$

In this case, the only requirement on $p(\mathbf{x})$ is that it should be possible to evaluate the function pointwise up to a normalising constant, a much looser assumption than if it should be practical to generate samples from it. The requirement of the importance density is that it should be easy to generate samples from and that it has the same support as $p(\mathbf{x})$, that is

$$p(\mathbf{x}) > 0 \Rightarrow q(\mathbf{x}) > 0 \quad \forall \mathbf{x} \in \mathbb{R}^{n_{\mathbf{x}}}. \tag{3.107}$$

The latter condition is to ensure that we are able to generate samples from the complete range of $\mathbf{x}$ and that the weights $p(\mathbf{x})/q(\mathbf{x})$ have an upper bound for all $\mathbf{x}$.

Particle filter algorithm

Here we will demonstrate how Monte Carlo approximation and importance sampling can be used sequentially to calculate the posterior distribution, $p(\mathbf{x}_k|\mathbf{Y}_{1:k})$, again using the general nonlinear models (3.5) and (3.6). Before deriving the filter we introduce $\mathbf{X}_{1:k-1} = \left[(\mathbf{x}_1)^T, (\mathbf{x}_2)^T, \ldots, (\mathbf{x}_k)^T\right]$, which contains a whole trajectory of state vectors. Suppose that we at time $k-1$ have a set of particles with associated normalised weights, $\{\mathbf{X}_{1:k-1}^{(m)}, w_{k-1}^{(m)}\}_{m=1}^{M}$. This particle set approximates $p(\mathbf{X}_{1:k-1}|\mathbf{Y}_{1:k-1})$ such that

$$p(\mathbf{X}_{1:k-1}|\mathbf{Y}_{1:k-1}) \approx \sum_{m=1}^{M} w_{k-1}^{(m)} \delta(\mathbf{X}_{1:k-1} - \mathbf{X}_{1:k-1}^{(m)}), \tag{3.108}$$

$$w_{k-1}^{(m)} \propto \frac{p(\mathbf{X}_{1:k-1}^{(m)}|\mathbf{Y}_{1:k-1})}{q(\mathbf{X}_{1:k-1}^{(m)}|\mathbf{Y}_{1:k-1})}, \qquad \sum_{m=1}^{M} w_{k-1}^{(m)} = 1. \tag{3.109}$$

Furthermore, assume that we can factorise the importance functions as,

$$q(\mathbf{X}_{1:k}|\mathbf{Y}_{1:k}) \triangleq q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_k) q(\mathbf{X}_{1:k-1}|\mathbf{Y}_{1:k-1}). \tag{3.110}$$

Using importance function on this form enables simple generation of new particles, $\mathbf{X}_{1:k}^{(m)} \sim q(\mathbf{X}_{1:k}|\mathbf{Y}_{1:k})$, by adjusting each of the old particles, $\mathbf{X}_{1:k-1}^{(m)} \sim q(\mathbf{X}_{k-1}|\mathbf{Y}_{1:k-1})$, by appending $\mathbf{x}_k^{(m)} \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^{(m)}, \mathbf{y}_k)$. Here, we assume that the $\mathbf{x}_k$ is a first order Markov process and hence only dependent on its previous state. To conclude the filter update we just need to update the importance weights. Remembering the general measurement update equation (3.8) we can express the posterior distribution as

$$\begin{aligned} p(\mathbf{X}_{1:k}|\mathbf{Y}_{1:k}) &= \frac{p(\mathbf{y}_k|\mathbf{X}_{1:k}, \mathbf{Y}_{1:k-1})p(\mathbf{X}_{1:k}|\mathbf{Y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{Y}_{1:k-1})} \\ &= \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{X}_{1:k-1}, \mathbf{Y}_{1:k-1})p(\mathbf{X}_{1:k-1}|\mathbf{Y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{Y}_{1:k-1})} \\ &\propto p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{X}_{1:k-1}|\mathbf{Y}_{1:k-1}) \end{aligned} \tag{3.111}$$

An expression for updating the unnormalised importance weights can be found by substituting (3.110) and (3.111) in (3.109),

$$\begin{aligned} \tilde{w}_k^{(m)} &= \frac{p(\mathbf{y}_k|\mathbf{x}_k^{(m)})p(\mathbf{x}_k^{(m)}|\mathbf{x}_{k-1}^{(m)})p(\mathbf{X}_{1:k-1}^{(m)}|\mathbf{Y}_{1:k-1})}{q(\mathbf{x}_k^{(m)}|\mathbf{x}_{k-1}^{(m)}, \mathbf{y}_k)q(\mathbf{X}_{1:k-1}^{(m)}|\mathbf{Y}_{1:k-1})} \\ &= \frac{p(\mathbf{y}_k|\mathbf{x}_k^{(m)})p(\mathbf{x}_k^{(m)}|\mathbf{x}_{k-1}^{(m)})}{q(\mathbf{x}_k^{(m)}|\mathbf{x}_{k-1}^{(m)}, \mathbf{y}_k)} w_{k-1}^{(m)}, \end{aligned} \tag{3.112}$$

and by normalisation we have

$$w_k^{(m)} = \frac{\tilde{w}_k^{(m)}}{\sum_{m=1}^{M} \tilde{w}_k^{(m)}}. \tag{3.113}$$

From the updated particle set $\{\mathbf{x}_k^{(m)}, w_k^{(m)}\}_{m=1}^{M}$, it is possible to form an approximative MMSE estimate of the quantities of interest as,

$$\hat{\mathbf{x}}_k^{MMSE} = \mathbb{E}\{\mathbf{x}_k|\mathbf{Y}_{1:k}\} \approx \sum_{m=1}^{M} w_k^{(m)} \mathbf{x}_k^{(m)}. \tag{3.114}$$

The accuracy of (3.114) depends on how many particles are used in the filter as well as how effectively these particles are utilised. The latter could be

somewhat controlled by choosing an appropriate importance function. In [16] it is proven that choosing $q(\mathbf{x}_k^{(m)}|\mathbf{x}_{k-1}^{(m)}, \mathbf{y}_k) = p(\mathbf{x}_k^{(m)}|\mathbf{x}_{k-1}^{(m)}, \mathbf{y}_k)$ minimises the variance in the weights, $w_k^{(m)}$. In the general nonlinear filtering problem, it could be difficult to generate samples from this pdf. A good approximation is, however, to estimate the distribution using either an EKF or an UKF filter. Another simple, but suboptimal, choice is to use the motion model, $p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(m)})$, and therefore ignore the new information in the measurement when generating new particles.

In [16] it is also shown that the variance of the importance weights can only grow for each iteration. In a few iterations the mass of the normalised weights will be concentrated in a handful of particles, and only these particles contribute in describing the posterior density. As a result, the Monte Carlo approximation will degenerate. One answer to the problem would be to resample the particles if the importance weights get concentrated in just a few particles. By mapping the degenerated particle set $\{\mathbf{x}_k^{(m)}, w_k^{(m)}\}_{m=1}^M$ into a new particle set with uniform weights, $\{\mathbf{x}_k^{*(m)}, 1/M\}_{m=1}^M$, such that

$$\Pr\{\mathbf{x}_k^{*(m)} = \mathbf{x}_k^{(m)}\} = w_k^{(m)}, \tag{3.115}$$

the new particle set will be likely to contain more good particles that previously had large weights than particles with low weights. This is sometimes appropriately called the survival of the fittest.

3.2 Models

In the model-based formulation of the sequential tracking problem that is discussed in this thesis, there is a need for two probabilistic models, one describing the uncertainty in target motion and the other describing the uncertainty in measurement origin. These models are called motion model (or dynamic model) and sensor (or measurement model) and are here assumed to be in the form of (3.5) and (3.6), respectively. It is clear that if information exists about the probable motion of a target and sensor characteristics,a tracking system should benefit from including this information.

Consequently, will the performance of the tracking system also be dependent on the accuracy of both these models.

This section discusses the design of these two models. In Section 3.2.1 different aspects of constructing target motion models are considered, whereas Section 3.2.2 focuses on modelling sensor measurement origin and the uncertainties therein.

3.2.1 Motion models

The motion model in a discrete-time tracking framework aims at describing the evolution of the state vector in-between the measurement updates. That is, assuming the previous position of target i is known, $\mathbf{x}_{k-1}^i$, the motion model predicts the target's position at the time of the next expected observation, $\mathbf{x}_k^i$, including uncertainty measures. In preventive safety applications, motion models are also used to predict the future traffic scenario. Precise vehicle motion models are of high importance, as these systems need to give prior warning or intervene before the dangerous situation occur. In Figure 3.5 5 is an example, where a vehicle motion model can be used to determine if a car is likely to brake to take account of the slower vehicle in front or if it is more likely to overtake. By comparing the motion indicated by the two models with the actual motion of the vehicle, a decision can be made on which of these hypotheses is more likely.

From the discussion in Section 3.1 discrete-time motion models are considered on the from,

$$\mathbf{x}_k^i = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}^i, \mathbf{v}_{k-1}),$$

where $\mathbf{v}_{k-1}$ is a noise process used to describe motion uncertainties. There are three inherent uncertainties that need to be considered, namely,

- the uncertainty in possible deviations/changes from the path stipulated by $\mathbf{x}_{k-1}$ between $k-1$ and k
- the uncertainty in parameters used to describe $\mathbf{f}_{k-1}(\cdot)$

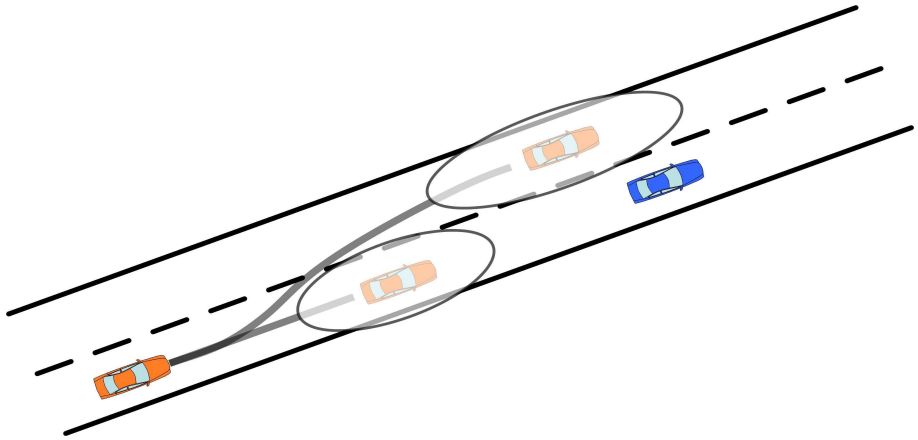


Figure 3.5: Multi hypothesis motion model for prediction of the future trajectory of the ego vehicle.

- the uncertainty that originates from approximations or simplifications when designing $\mathbf{f}_{k-1}(\cdot)$

To find this model, there are two choices. Either to firstly formulate the model in a continuous-time differential equation model, the solution of which is then discretised, or describing the discrete-time motion model directly. Often it is more convenient to start by formulating the continuous-time model, as motion is a continuous-time phenomenon. However, depending upon the complexity of the model, it could be difficult to find an analytical solution to the discretisation. In such cases it is easier to construct the motion model directly in discrete time. The main difference between these approaches is the interpretation of the resulting noise process. The examples used in this section are directly formulated in discrete time.

An extensive overview of different target motion models is given in [35]. Helpful books that also cover the subject are [12] and [7]. These references mainly focus on kinematic models, extensively used to describe the motion of, e.g. aircrafts. More accurate motion models, which specialise in describing the motion of cars, have emerged since the introduction of preventive safety systems, relying on accurate description of the vehicle dynamics, e.g., ESC discussed in Chapter 2. An example of one such model is the *single track model*, used in, e.g. [2]. Even though the primary use of this model is for

control applications, a simplified version of the model is successfully applied in a tracking framework in [22] and [10]. Examples of commonly used motion models are given next.

Constant velocity model

A frequently used and simple model to describe vehicle motion is the constant velocity model (CV). This model is based upon the assumption that the accelerations posed on the vehicle between two samples are constant and drawn from a zero-mean white noise process. Hence, this model is sometimes called discrete white acceleration model [7]. Using this assumption the motion of the vehicle can be modelled as,

$$\mathbf{x}_k = \mathbf{F}^{\text{cv}}\mathbf{x}_{k-1} + \mathbf{G}^{\text{cv}}\mathbf{v}_{k-1} \tag{3.116}$$

where

$$\mathbf{F}^{\text{cv}} = \begin{bmatrix} 1 & 0 & T_s & 0 \\ 0 & 1 & 0 & T_s \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{G}^{\text{cv}} = \begin{bmatrix} T_s^2/2 & 0 \\ 0 & T_s^2/2 \\ T_s & 0 \\ 0 & T_s \end{bmatrix} \tag{3.117}$$

where the state is parameterised in the inertial coordinates $(\xi_x(k), \xi_y(k))$ as $\mathbf{x}_k = [\xi_x(k), \xi_y(k), \dot{\xi}_x(k), \dot{\xi}_y(k)]^T$. The covariance matrix of the white acceleration noise process is,

$$\text{Cov}\{\mathbf{v}_{k-1}\} = \begin{bmatrix} \sigma_{\ddot{x}}^2 & 0 \\ 0 & \sigma_{\ddot{y}}^2 \end{bmatrix}. \tag{3.118}$$

The covariance matrix of the noise term in (3.116) then becomes,

$$\text{Cov}\{\mathbf{G}^{\text{cv}}\mathbf{v}_{k-1}\} = \begin{bmatrix} \sigma_{\ddot{x}}^2 T_s^4/4 & 0 & \sigma_{\ddot{x}}^2 T_s^3/2 & 0 \\ 0 & \sigma_{\ddot{y}}^2 T_s^4/4 & 0 & \sigma_{\ddot{y}}^2 T_s^3/2 \\ \sigma_{\ddot{x}}^2 T_s^3/2 & 0 & \sigma_{\ddot{x}}^2 T_s^2 & 0 \\ 0 & \sigma_{\ddot{y}}^2 T_s^3/2 & 0 & \sigma_{\ddot{y}}^2 T_s^2 \end{bmatrix} \tag{3.119}$$

Note that the unit of $\sigma_{\ddot{x}}$ and $\sigma_{\ddot{y}}$ is m/s^2, i.e., accelerations as would be expected.

Constant acceleration model

A slightly more complicated motion model is the constant acceleration model (CA). In this model, the acceleration *increments* are assumed to be a zero-mean white process. This results in a model containing position, speed and acceleration in each dimension as,

$$\mathbf{x}_k = \mathbf{F}^{\text{ca}}\mathbf{x}_{k-1} + \mathbf{G}^{\text{ca}}\mathbf{v}_{k-1} \tag{3.120}$$

where

$$\mathbf{F}^{\text{ca}} = \begin{bmatrix} 1 & 0 & T_s & 0 & T_s^2/2 & 0 \\ 0 & 1 & 0 & T_s & 0 & T_s^2/2 \\ 0 & 0 & 1 & 0 & T_s & 0 \\ 0 & 0 & 0 & 1 & 0 & T_s \end{bmatrix}, \quad \mathbf{G}^{\text{ca}} = \begin{bmatrix} T_s^2/2 & 0 \\ 0 & T_s^2/2 \\ T_s & 0 \\ 0 & T_s \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.121}$$

and the state vector is extended to also include the accelerations, according to $\mathbf{x}_k = [\xi_x(k),\, \xi_y(k),\, \dot{\xi}_x(k),\, \dot{\xi}_y(k),\, \ddot{\xi}_x(k),\, \ddot{\xi}_y(k)]^T$. The noise term in the CA model becomes,

$$\text{Cov}\{\mathbf{G}^{\text{ca}}\mathbf{v}_{k-1}\} = \begin{bmatrix} \sigma_{\ddot{x}}^2 T_s^4/4 & 0 & \sigma_{\ddot{x}}^2 T_s^3/2 & 0 & \sigma_{\ddot{x}}^2 T_s^2/2 & 0 \\ 0 & \sigma_{\ddot{y}}^2 T_s^4/4 & 0 & \sigma_{\ddot{y}}^2 T_s^3/2 & 0 & \sigma_{\ddot{y}}^2 T_s^2/2 \\ \sigma_{\ddot{x}}^2 T_s^3/2 & 0 & \sigma_{\ddot{x}}^2 T_s^2 & 0 & \sigma_{\ddot{x}}^2 T_s & 0 \\ 0 & \sigma_{\ddot{y}}^2 T_s^3/2 & 0 & \sigma_{\ddot{y}}^2 T_s^2 & 0 & \sigma_{\ddot{y}}^2 T_s \\ \sigma_{\ddot{x}}^2 T_s^2/2 & 0 & \sigma_{\ddot{x}}^2 T_s & 0 & \sigma_{\ddot{x}}^2 & 0 \\ 0 & \sigma_{\ddot{y}}^2 T_s^2/2 & 0 & \sigma_{\ddot{y}}^2 T_s & 0 & \sigma_{\ddot{y}}^2 \end{bmatrix} \tag{3.122}$$

The unit of the standard deviation of noise process is m/s^2. A clear advantage with both these models is that they are linear in both the state and the noise

process. This property make them suitable motion models to use in a Kalman filter.

3.2.2 Sensor models

The sensor model tries to capture the probabilistic relation between the sensor measurements and the state vector. To recap (3.6), we assume that it is possible to describes this relation using

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{w}_k),$$

where $\mathbf{h}(\cdot)$ is a possibly nonlinear function and $\mathbf{w}_k$ is a noise process that describe the uncertainty in the origin of the measurement. The interpretation of $\mathbf{h}(\cdot)$ can typically be divided into two parts,

1. a deterministic mapping between the state vector to the measurement vector, $\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{0})$, describing what in $\mathbf{x}_k$ is measured and how it is measured.

2. a probabilistic description of the measurement noise including uncertainties that originate from simplification and assumptions made in the mapping described above that influence the uncertainty in the origin of the measurement.

To illustrate this we use a simple example where we are tracking a vehicle in Cartesian coordinates on which we make radar observations in polar coordinates.

Example 3.5 (Radar sensor model for a point target)
Suppose we have the state vector parameterised in Cartesian coordinates as,

$$\mathbf{x}_k = \begin{bmatrix} \eta_x \\ \eta_y \\ \dot{\eta}_x \\ \dot{\eta}_y \end{bmatrix}. \tag{3.123}$$

In the system we have a radar sensor delivering measurements in polar coordinates, e.g., range and azimuth (angle). A typical radar sensor model is formed as

$$\mathbf{y}_k = \begin{bmatrix} r \\ \phi \end{bmatrix} = \begin{bmatrix} \sqrt{\eta_x^2 + \eta_y^2} \\ \arctan\left(\frac{\eta_y}{\eta_x}\right) \end{bmatrix} + \begin{bmatrix} w_r \\ w_\phi \end{bmatrix} \tag{3.124}$$

where w_r and w_ϕ is measurement noise in range and azimuth, respectively. In (3.124) *it is clear that*

$$\mathbf{y}_k = \begin{bmatrix} \sqrt{\eta_x^2 + \eta_y^2} \\ \arctan\left(\frac{\eta_y}{\eta_x}\right) \end{bmatrix} \tag{3.125}$$

describes the deterministic mapping between the different coordinate systems and the w_r and w_ϕ accounts for the uncertainty in measurement origin.

□

Note that, the whole state vector is not observed by the sensor in Example 3.5. The only information that the tracking system can derive about the velocity of a target, $(\dot{\eta}_x, \dot{\eta}_y)$, is coupled through the target motion model. Only by using both these models the tracking framework is able to draw conclusion about the target velocity.

The purpose of the sensor model in a multi-target tracking system is twofold. It is used in the measurement update in the filter to determine not only how the information in the measurement is related to the state but also how much the tracking algorithm should trust the new measurement over the information already collected in the previous state estimate. Recalling the Kalman filter update equation (3.33),

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}), \\ \mathbf{K}_k &= \mathbf{P}_{\mathbf{xy}}\mathbf{P}_{\mathbf{yy}}^{-1}, \end{aligned}$$

where $\mathbf{P}_{\mathbf{yy}} = \mathbb{E}\{(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})^T\}$ is the covariance of the new information in the measurement describing the uncertainty therein and $\mathbf{P}_{\mathbf{xy}} = \mathbb{E}\{(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1})^T\}$ describes the mapping (correlation) between

the innovation and the state. Both of these statistics are highly dependent on the sensor model, as $\hat{\mathbf{y}}_{k|k-1} = \mathbb{E}\{\mathbf{h}_k(\mathbf{x}_k, \mathbf{w}_k)|\mathbf{Y}_{1:k-1}\}$. If the measurement is informative, i.e., $\mathbf{P_{yy}}$ is small in relation to $\mathbf{P_{xy}}$, the filter will trust the measurement more and vice versa if the measurement is less informative. Similarly, if the measurement noise is underestimated in the model compared to the real measurement noise, the estimates from the filter will become noisy. On the other hand, if the noise is overestimated, the information in the measurement will not be efficiently exploited. It is clear that no filter is optimal as long as the models used in the filter is inaccurate.

The other purpose of the sensor model is to find the likely measurements-to-track associations in the data association. Using the measurement model it is possible to determine how likely it is that a certain track gives rise to a certain measurement. These aspects of the sensor model are covered more in detail Section 3.3. For further discussion on sensor models, a useful survey in [36] and text books that cover the subject are [12] and [5].

3.3 Data association

Up until now we have disregarded any uncertainty in measurement-to-track assignments, *i.e* it has been known to us which measurement should update which track. In practical multi-target tracking systems, this is however seldom known. Methods for finding likely measurements-to-track associations need to be in place to handle this uncertainty, known as the data association problem.

Example 3.6 (Data association problem)
Suppose we have mounted a radar sensor on a vehicle observing the traffic in front of the car as in Fig. 3.6. At time k, we are tracking $n_t(k) = 2$ targets but we receive $n_m(k) = 11$ measurements. The data association problem is to find which of these $n_m(k)$ (if any) likely belong to which of the current $n_t(k)$ tracks.

To easily remove highly unlikely measurements-to-track associations, a gate is often constructed around each track. Measurements falling outside

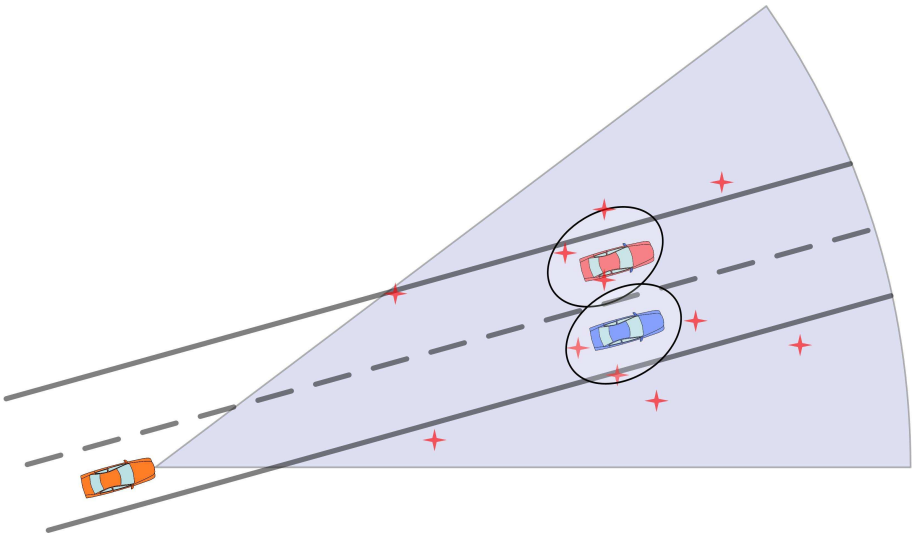


Figure 3.6: A radar mounted on a vehicle observes two targets in front. The measurements from the radar are depicted as red stars. It is clear that the radar sees both objects, but there are also many false alarms or clutter measurements. It is also uncertain which measurement truly originates from the two cars. To remove unlikely data associations, a so-called gate is constructed around each object, drawn as an ellipse. All measurements falling outside this gate will not be considered in the data association.

these gates are not considered as candidates for association. The gates in this example are depicted as circles in Fig. 3.6.

□

The discussion in this section is limited to systems that can be separated as

$$\mathbf{x}_k^i = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}^i, \mathbf{v}_{k-1}^i) \tag{3.126}$$

$$\mathbf{y}_k^i = \mathbf{h}_k(\mathbf{x}_k^i, \mathbf{w}_k^i) \tag{3.127}$$

where $\mathbf{x}_k^i$ is the state vector of the i^{th} track and $\mathbf{y}_k^i$ is a measurement originating from that track. It is further assumed that at most one measurement originates from each track. Even though we show in Paper I-III, that this is not always a valid assumption.

There are several methods available for solving the data association problem. A good overview is given in [12] and [5]. In [12] there is also an extension to more advanced methods not covered in this thesis, such as *multi-hypothesis tracking* (MHT). Here, we will focus the discussion to two common types of data association methods, nearest neighbour association, in Section 3.3.2, and all-neighbour association, in Section 3.3.3. The conceptual difference between the two is that the nearest neighbour algorithms find one likely measurement for each track, whereas the all-neighbour methods assign multiple measurements to each track, but weigh their contribution depending on how likely it is that each originate from the track. However, we start by examining a convenient method of reducing the computational load be removing unlikely measurements-to-track associations. This method is called gating and is explained in Section 3.3.1.

3.3.1 Gating

As shown in Example 3.6, gates are constructed around each track, or rather the predicted measurement of each track, $\hat{\mathbf{y}}^i_{k|k-1}$, to limit the number of possible measurement-to-track associations that are examined by the data association algorithm. Consider the residual between a measurement, $\mathbf{y}^j_k$, and the expected measurement from the i^{th} track, $\hat{\mathbf{y}}^i_{k|k-1}$,

$$\tilde{\mathbf{y}}^{i,j}_{k|k-1} = \mathbf{y}^j_k - \hat{\mathbf{y}}^i_{k|k-1}. \tag{3.128}$$

In the Kalman filter we know that this residual (or innovation) is Gaussian distributed as,

$$\tilde{\mathbf{y}}^{i,j}_{k|k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}^i_k), \tag{3.129}$$

and in the EKF and UKF we believe it is approximately so. Using this measure we are able to calculate the statistical distance, $d^2_{i,j}$, between the measurement, $\mathbf{y}^j_k$, and the expected measurement, $\hat{\mathbf{y}}^i_{k|k-1}$, as the norm of the residual,

$$d^2_{i,j} = (\tilde{\mathbf{y}}^{i,j}_{k|k-1})^T (\mathbf{S}^i_k)^{-1} (\tilde{\mathbf{y}}^{i,j}_{k|k-1}), \tag{3.130}$$

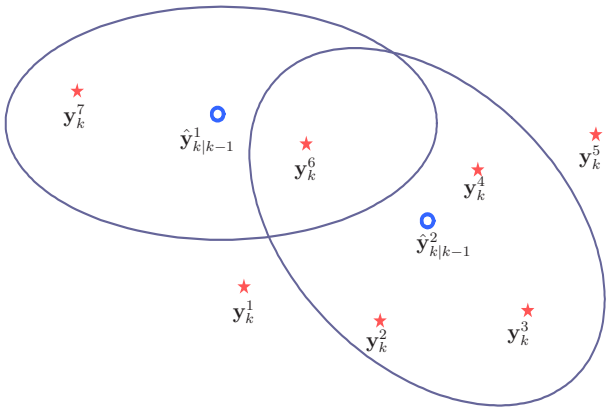


Figure 3.7: Two elliptic gates centred around the predicted measurement of each track.

where $d_{i,j}^2$, if (3.129) holds, is a χ_M^2 random variable for the correct measurement-to-track pairing, M being the dimension of the measurement vector. It is possible to express an elliptic gate using this measure as,

$$(\tilde{\mathbf{y}}_{k|k-1}^{i,j})^T(\mathbf{S}_k^i)^{-1}(\tilde{\mathbf{y}}_{k|k-1}^{i,j}) \leq G \tag{3.131}$$

where G is the gate size. All measurements that satisfy the gate criterion $d_{i,j}^2 \leq G$ is said to pass the gate and will be considered in the data association algorithm. The assumption that $d_{i,j}^2 \sim \chi_M^2$ can be used to determine a suitable gate size, according to the probability that the true measurement from the i^{th} track will fall with in the gate of size G, for details see, e.g., [12]. In Fig. 3.7, the gating of two tracks is demonstrated. Both predicted measurements, $\hat{\mathbf{y}}_{k|k-1}^1$ and $\hat{\mathbf{y}}_{k|k-1}^2$, are shown together with their elliptic gates. From the gating process will measurements, $\mathbf{y}_k^1$ and $\mathbf{y}_k^5$, will be removed by the data association algorithm from further processing. Note also that measurement $\mathbf{y}_k^6$ falls within the gate of both tracks.

As the accuracy of the predicted measurement of a track is both dependent on the motion model to calculate $\hat{\mathbf{x}}_{k|k-1}^i$, and the sensor model to retrieve, $\hat{\mathbf{y}}_{k|k-1}^i$, the precision of both these models will influence the gating performance as well as the opportunity to create correct data associations.

3.3.2 Nearest neighbour association

The concept of nearest neighbour association is to find the one measurement that is in some sense nearest to the predicted measurement of each track. The simplest method is called *nearest neighbour* (NN), and is a greedy method in which each track picks the measurement with the smallest statistical distance $d^2_{i,j}$. Clearly, this could result in two tracks choosing the same measurement to update their estimates, and can lead to the tracks converging to describe the same target. NN can be seen as locally minimising the association distance for each individual track separately.

Another natural approach would be to find the global minimum distance considering all tracks simultaneously, under the restriction that a measurement can only be associated with one track. This approach is referred to in the literature as *global nearest neighbour* (GNN) or *single hypothesis tracking* [12], as it tries to find the single most probable measurement-to-track assignment. To ensure that poor quality tracks (large prediction covariance) do not steal measurements from high quality tracks, it is common to modify (3.130) to the general statistical distance,

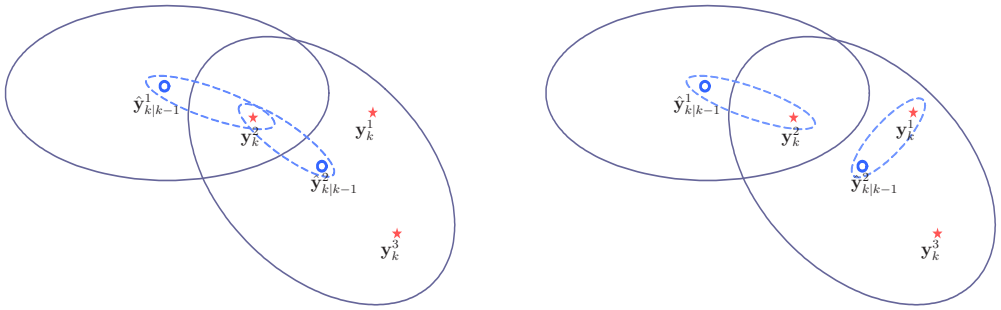
$$\breve{d}^2_{i,j} = d^2_{i,j} + \ln(|\mathbf{S}^i_k|), \tag{3.132}$$

where $|\cdot|$ is the determinant. Additionally, some arbitrary large distance can be given to those measurements that fall outside the gate of a track. The aim of GNN is then to find the unique measurement-to-track assignments $(1, \lambda_1), \ldots, (n_t(k), \lambda_{n_t(k)})$ that solves,

$$\min_{\{\lambda_i\}} \sum_{i=1}^{n_t(k)} \breve{d}^2_{i,\lambda_i}, \tag{3.133}$$

where $\lambda_i \in (0, \ldots, n_m(k))$ indicates which measurement track i has been assigned. The value $\lambda_i = 0$ stipulates that track i is not assigned a measurement.

The main difficulty with GNN is to find an efficient way to solve the assignment problem (3.133). In [12], there is a discussion of different alternatives,



(a) Data association using NN

(b) Data association using GNN

Figure 3.8: Comparison between NN and GNN data association methods.

where the auction algorithm [11] is suggested as a suitable candidate. Solving (3.133) using one of these algorithms results in the data association matrix $\mathbf{\Lambda}_k(i, \lambda_i) = 1$, indicating which measurement, λ_i, is assigned to track i.

Figure 3.8 shows a comparison between the NN and GNN, where one of the tracks only has one measurement in its gate. The comparison clearly shows the suboptimal behaviour of the NN algorithms as it assigns measurement, $\mathbf{y}^2_k$, to both tracks.

3.3.3 All-neighbour association

In contrast to the data association methods discussed in Section 3.3.2, the all-neighbour association methods, as the name implies, uses all the measurements that fall inside the gate of a track. The idea is to weigh each measurement, j, in accordance with the probability that it originated from track i, denoted p_{ij}. There are two methods that use this approach for assigning multiple measurements to tracks. The first one was *probabilistic data association* (PDA) proposed in [8], which was later extended to *joint probabilistic data association* (JPDA) in [19]. The only difference between the two is the way the association probabilities are calculated, p_{ij}. Similarly to NN, PDA considers each track separately, whereas JPDA formes global hypotheses to calculate the probabilities. We start by explaining PDA and later expand to consider JPDA.

To calculate the association probabilities, p_{ij}, these algorithms forms data

association hypothesis. Suppose $n_m^i(k)$ measurements fall within the gate of track i at time k, then there are $n_m^i(k)+1$ hypotheses that can be formed. One suggesting that none of the measurements originate from track i,i.e., $\lambda_i = 0$, and that measurement j is the valid measurement from track i, denoted $\lambda_i = j$. Using these hypotheses we can calculate the association probabilities,

$$p_{ij} = \Pr\{\lambda_i = j \big| \mathbf{Y}_{1:k-1}\}, \tag{3.134}$$

where $j = 0 \dots n_m^i(k)$ and p_{i0} is the probability that non of the measurements are valid. The association probabilities for all tracks can be summarised in the association matrix as, $\mathbf{\Lambda}_k(i,j) = p_{ij}$. Suggestions on how to calculate these probabilities can be found in [12] and [8]. Using these probabilities it is possible to calculate an equivalent innovation to use in the measurement update,

$$\tilde{\mathbf{y}}_k^i = \sum_{j=1}^{n_m^i(k)} p_{ij} \tilde{\mathbf{y}}_k^{i,j}, \tag{3.135}$$

where $\tilde{\mathbf{y}}_k^{i,j}$ is the innovation of measurement j to track i. To compensate that clutter is also weighted into the state update, the resulting error covariance needs to be adjusted,

$$\mathbf{P}_{k|k} = p_{i0}\mathbf{P}_{k|k-1} + (1-p_{i0})\mathbf{P}_{k|k}^* + d\mathbf{P}_{k|k} \tag{3.136}$$

where the state covariance updated with the correct measurement is,

$$\mathbf{P}_{k|k}^* = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T \tag{3.137}$$

and the spread of the innovation terms contribute with,

$$d\mathbf{P}_{k|k} \triangleq \mathbf{K}_k \left[\sum_{j=1}^{n_m^i(k)} p_{ij} (\tilde{\mathbf{y}}_k^{i,j})(\tilde{\mathbf{y}}_k^{i,j})^T - (\tilde{\mathbf{y}}_k^i)(\tilde{\mathbf{y}}_k^i)^T \right] \mathbf{K}_k^T \tag{3.138}$$

This concludes the formulation of the PDA algorithm. Additional information about the PDA algorithm can be found in [31], where also practical examples of PDA implementations are provided.

As previously mentioned, the only difference between PDA and JPDA, is the way the hypotheses are constructed. PDA suffer from the same deficiency as NN, in that it is possible for tracks to share measurements. In multi-target environments will this lead to degeneration of the tracking performance. JPDA mitigates this deficiency by calculating the association probabilities, p_{ij}, jointly across all tracks. In these calculations assignment hypotheses are formed where there are no assignment conflict

$$\lambda^n = \{\lambda_1^n, \lambda_2^n, \ldots, \lambda_{n_t(k)}^n : \text{if } \lambda_i^n = \lambda_j^n \text{ then } i = j \text{ or } \lambda_i^n = \lambda_j^n = 0\} \tag{3.139}$$

The joint association probability can the be calculated as,

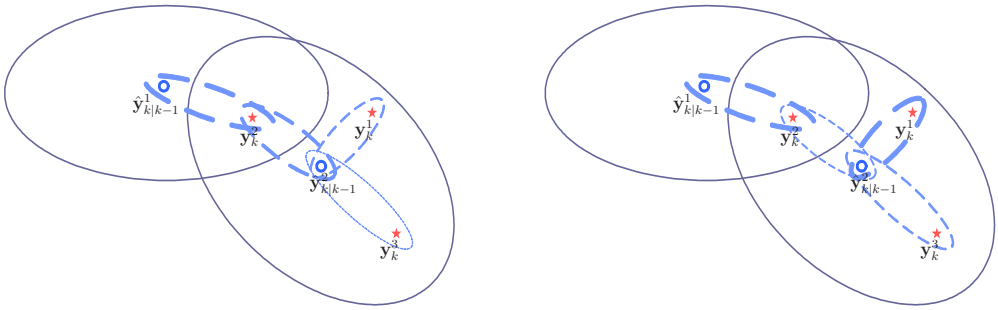
$$p_{ij} = \sum_{\{\lambda^n = j\} \in \lambda^n} \Pr\{\lambda^n \big| \mathbf{Y}_{1:k-1}\}. \tag{3.140}$$

This will ensure that a measurement that is likely to originate from one track will have low weight (probability) for all the other tracks. More on how to calculate these global probabilities can be found in [19].

In Fig. 3.9, a similar comparison is made between the PDA and JPDA as previously made between NN and GNN. As these methods computes a weighted innovation from all measurements inside the gate, how much weight is assigned a certain measurement is indicated by the thickness of the connection. The conclusion being that less cross influence between sharing measurements is achieved using the JPDA.

3.4 Track management

In order to apply the methods discussed in the previous sections we have one major aspect left to consider, the aspect of track management. Up until now we have assumed that all tracks have already been validated and that



(a) Data association using PDA

(b) Data association using JPDA

Figure 3.9: Comparison between PDA and JPDA data association methods.

we have no uncertainty regarding how many true targets we are observing. Track management methods make sure that only tracks of sufficient quality gets reported to the receiving decision algorithms. The life of a track can be divided into three stages,

track initialisation: a tentative track gets started, for those measurements that are likely to originate from new targets, by choosing a suitable prior, $p(\mathbf{x}_0 \big| \mathbf{y}_0)$, to initialise the tracking filter. This enables us to associate further measurements to the tentative track using standard data association methods.

track validation: if observations indicate that the tentative track resembles a true target rather then just random noise or clutter, its status gets upgraded to a validated track, i.e., a track that can be trusted by the receiving decision algorithms as describing a real target.

track deletion: if the track is no longer visible by any sensor or the tracking algorithm is too uncertain by the target state, the track gets removed and will no longer be considered.

In this section we discuss different methods of track initialisation, validation and track deletion.

3.4.1 Track initialisation

When initiating a track in a Bayesian tracking filter it is essential to design a prior density, $p(\mathbf{x}_0|\mathbf{y}_0)$, that incorporates the information about the state, $\mathbf{x}_0$, before any observations are made. Depending on the problem the construction of a proper prior could sometimes be difficult. In the example of tracking vehicles, prior knowledge is available in that vehicles tend to travel on the road and with a velocity in the same direction as the road. This type of information could, for example, be included into the prior density. It is also possible to have different priors to represent different hypotheses in the track initialisation. A radar sensor is only able to measure radial velocity, as such it is hard to distinguish between a stationary vehicle and a vehicle travelling perpendicular to the sensor. By designing a prior for both hypotheses it is possible to let future measurements determine which hypothesis is the correct one.

It is, however, also possible to initiate the filter with a non-informative prior, indicating that no information about the state exists before making any observations. In a Kalman filter framework this would be to chose a prior as,

$$\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_0), \tag{3.141}$$

where $\mathbf{P}_0$ is chosen to be much larger in relation to the measurement uncertainty. In this case both the mean and covariance estimates, after the first iteration of the Kalman filter, will be largely based on the information in the associated measurement and its uncertainty.

Further discussion on suitable priors in a tracking framework are presented in [7] and a more general discussion of the choice of prior density can be found in [47].

3.4.2 Track validation

Before tentative tracks get reported further, the track management need to make sure that it describes a real target and is not a result of spurious

clutter. A frequently used but *ad hoc* method is the M/N principle. Here, a track is validated if M measurements out of N possible are associated with the tentative track. This principle for track validation is frequently used together with nearest neighbour data association methods.

More formal methods for track validation in all-neighbour association algorithms can be found in [6] and [39]. The latter derives the PDA algorithm without the initial assumption of track existence. The resulting *integrated probabilistic data association* (IPDA) is thus capable of both expressing probability of target existence and data association. A track is validated using this framework when the probability of existence exceeds a certain threshold.

A general measure for track validation which support standard detection test, such as the *sequential probability ratio test* (SPRT), is the formulation of track score first proposed in [48]. The probability ratio used in SPRT to accommodate track validation can be found by calculating,

$$s_k^i = \frac{\Pr\{\{\}\,\mathcal{H}_T^i \big| \mathbf{Y}_{1:k}\}}{\Pr\{\{\}\,\mathcal{H}_{FA}^i \big| \mathbf{Y}_{1:k}\}} = \frac{p(\mathbf{Y}_{1:k}\big|\mathcal{H}_T^i)\Pr\{\{\}\,\mathcal{H}_T^i\}}{p(\mathbf{Y}_{1:k}\big|\mathcal{H}_{FA}^i)\Pr\{\{\}\,\mathcal{H}_{FA}^i\}}, \tag{3.142}$$

where $\mathcal{H}_T^i$ is the hypothesis that track i describes a true target and $\mathcal{H}_{FA}^i$ is the hypothesis that it is a false alarm (or clutter). Assuming independent measurement noise a sequential formulation of the track score can be found as

$$s_k^i = \frac{p(\mathbf{y}_k\big|\mathcal{H}_T^i)}{p(\mathbf{y}_k\big|\mathcal{H}_{FA}^i)} \cdot \frac{p(\mathbf{Y}_{1:k-1}\big|\mathcal{H}_T^i)\Pr\{\{\}\,\mathcal{H}_T^i\}}{p(\mathbf{Y}_{1:k-1}\big|\mathcal{H}_{FA}^i)\Pr\{\{\}\,\mathcal{H}_{FA}^i\}} = \frac{p(\mathbf{y}_k\big|\mathcal{H}_T^i)}{p(\mathbf{y}_k\big|\mathcal{H}_{FA}^i)} s_{k-1}^i. \tag{3.143}$$

Similarly as in the IPDA algorithm, a tentative track is validated if the track score exceeds a certain threshold. A detailed derivation of the hypotheses probabilities can be found in [12].

3.4.3 Track deletion

The deletion of tracks can also be performed by the use of the track score defined in (3.142). In this case if the score falls below a certain threshold. There are also other more ad hoc methods, such as deleting the track if no

observations have been associated with the track in N_D consecutive scans.

Chapter 4

Radar sensor modelling

There exist a vast amount of literature describing all parts of a radar system, from the design of the important oscillator circuits to advanced signal processing specialities. Probably the most classical reference work is the introductional book by Skolnik [49]. Other good reference books include [50], directed towards non-specialists, and [43], having a signal processing point of view. In this thesis, the discussion is primarily focused on the signal processing and signal modelling perspectives of the radar sensor. That is, discussions about radar system performance analysis and components are omitted to benefit the understanding of the characteristics of the radar signal and the radar measurements. This will also serve as an introduction to the contribution in the appended papers dealing with radar sensor modelling.

4.1 Brief history of radar and its applications

The word radar originates from the acronym, RADAR, for "radio detection and ranging," but is now considered an ordinary English noun. The word has two meanings today [1]. The first meaning is a method of detecting distant objects and determining their position, speed, material composition, or other characteristics, by causing radio waves to be reflected from them and analysing the reflected waves. The second meaning is the equipment used in such detecting.

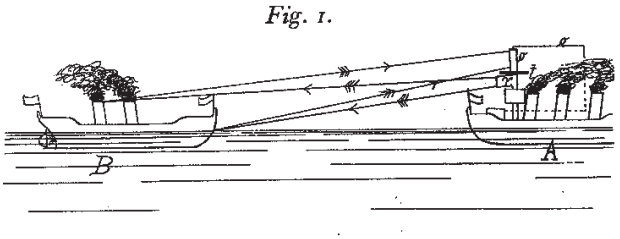


Figure 4.1: The first radar patent describing a system for detecting ships in fog.

The first patent on radar technology was granted in 1904 to the German engineer Hülsmeier [23]. As depicted in Fig. 4.1, the intended application was a collision warning system for ships where radio waves were used to detect approaching ships in dense fog. The development was accelerated and spread in the 1930s, very much driven by military necessities, which still today is the dominant user and developer of radar technology. Some military applications of radar sensors include surveillance, navigation, and weapon-guidance for ground, sea and air vehicles.

Today the use of radar systems is far from limited to military applications. Most of us have probably been subjected the police traffic radar, used to enforce speed limits, and have watched a weather forecast showing images from a meteorological radar depicting the movement of rain clouds. As Hülsmeier originally intended, radar sensors are frequently used in collision avoidance systems for ships and, as indicated by this thesis, they are beginning to have the same role for automotive safety systems.

4.2 Radar system description

This thesis focuses on what is called a *detection* radar which is used in, e.g., automotive safety systems. The basic aim of a detection radar is to detect the presence of objects and to estimate the range and angle to these objects as well as their range rate (radial velocity). This is accomplished by analysing the return of a directed and modulated radio signal (electromagnetic wave)

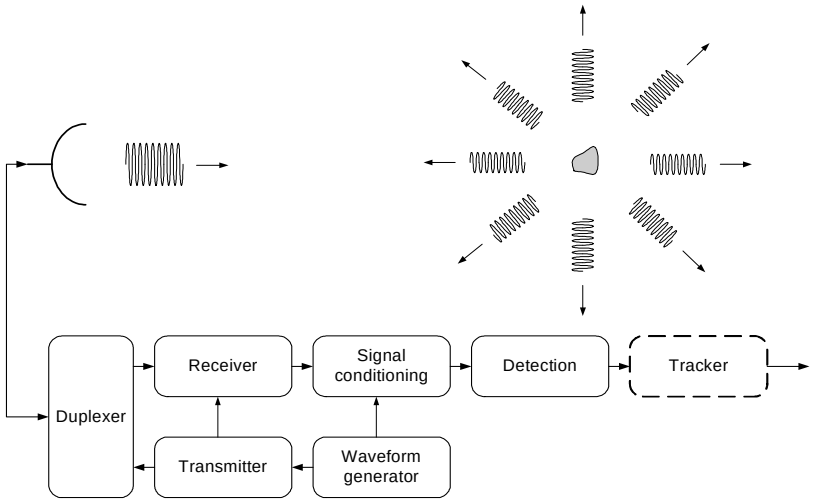


Figure 4.2: One general description of a monostatic detection radar where the emphasis is put on the signal processing parts of the radar system.

sent from the radar antenna and scattered back both from the objects of interest as well as from other objects, known as clutter. In Figure 4.2, one example of a generic description of a monostatic[1] detection radar system is shown, where the emphasis is put on the signal processing parts of the system. A walk-through of the different steps is given below.

We start with the design of the transmitted signal. The shape of the radar signal is first constructed at baseband by the waveform generator. The choice of waveform, e.g. simple pulse, linear frequency modulated pulse or continuous wave signal, will very much influence the resolution and accuracy in both the estimates of the range to objects and their range rates. The baseband signal is modulated by the carrier frequency and amplified in the transmitter stage before it is sent to the antenna through the duplexer.

The antenna radiates the modulated waveform as a directed electromagnetic wave, where the directivity is determined by the antenna beam pattern (proportional to the size of the antenna). The narrower the beam pattern, the more concentrated the electromagnetic wave is. Objects illuminated by

[1]Monostatic means that both the transmit and receive antenna are either the same or at least collocated.

the radio wave scatters the energy in different directions and a small portion is reflected back to the radar. The back-scattered signal is received by the antenna and then amplified and down-converted to baseband in the receiver stage. The down-conversion to a baseband signal is typically performed in a couple of steps, each synchronised with the carrier frequency which modulates the transmitted signal (coherent radar).

As only a small part of the transmitted energy is reflected back and intercepted by the radar antenna, the received signal has a very low amplitude and is thus sensitive to corruption by thermal noise in the receiver and atmospheric noise intercepted by the antenna. A low signal-to-noise (SNR) will badly affect the performance of the detector. Hence, to improve the SNR before the detection stage some signal conditioning is usually performed. Assuming additive white Gaussian noise, maximised SNR is achieved by filtering the received baseband signal with a conjugated and time-reversed version of the transmitted radar waveform, a so called *matched filter* [30].

Once the SNR is maximised, it is up to the detection stage to decide whether there are reflections from objects present in the received signal or not. This decision is typically made using statistical decision theory considering two hypotheses, is there a reflected radar signal present in the received signal during a certain time interval or is there only noise. In most cases, the decision is made by comparing the amplitude at the output of the matched filter against a threshold, which could either be set a priori or determined adaptively from radar data. If the amplitude is above the threshold, it is assumed that there is an object present, and vice versa. If an object reflection is detected, the signal can be further analysed to more accurately estimate the range and angle to the object as well as its range rate.

4.3 Basic radar measurement principles

As aforementioned, the aim of a detection radar is to detect objects and to estimate their positions (in terms of range and angle[2]) as well as their range

[2]Automotive applications are typically only interested in the bearing to objects and not the elevation angle. Hence, this thesis only considers the position of objects in range

rates. In this section, we briefly discuss the basic properties of the estimates (measurements) from a detection radar on the form given in Fig. 4.2. The estimates are discussed in terms of measurement principles and in terms of resolution and accuracy.

The accuracy is to what precision the radar can correctly determine the position of a point object, and is presented as the theoretical minimum variance of the estimation error determined by the Cramer-Rao lower bound (CRLB). The resolution, on the other hand, describes the ability of the radar to separate two closely spaced point objects in a certain measurement dimension and report them as two objects and not as one. An inclusive figure of merit of the resolution capability of a radar is however difficult to attain as it is highly dependent on the relative strength of the return signals (from the two closely spaced objects). As a result, we resort to give an intuitive description of which radar parameters that influence the resolution in each dimension and to give a rule-of-thumb expression for the resolution if the reflections from the objects are of equal strength.

4.3.1 Range measurement

Suppose a radar signal sent at time $t = 0$ is reflected from an object and detected at the receiver output at time $t = \tau$. The time it took for the signal to propagate to the object and back, τ, is directly proportional to the distance to the object according to this simple relation

$$r = \frac{c\tau}{2}, \tag{4.1}$$

where r is the range to the object and c is the speed of light. Hence, by measuring the time it takes for a transmitted signal to propagate to an object and back gives us a measure of the distance to that object.

and bearing.

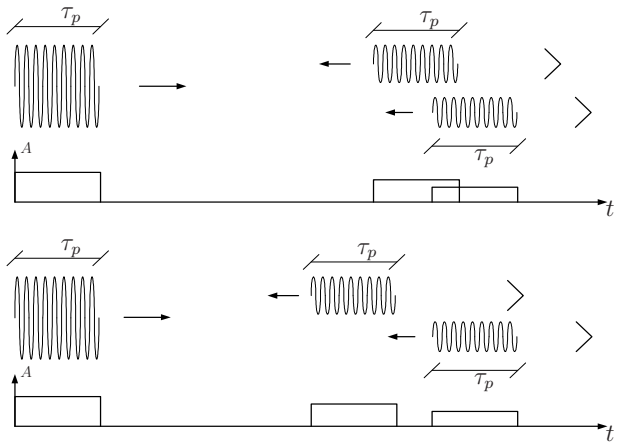


Figure 4.3: Illustration of the range resolution of a pulsed radar. In the top scenario the echo from both objects are grouped in the received signal whereas in the bottom scenario they are well separated.

Range resolution

The capability of a radar to separate two closely spaced targets in range is most intuitively explained considering a pulse radar. In its most simple form, a pulse radar transmits a pulse modulated signal at a single carrier frequency and of a certain duration τ_p (in practice the time between the half-power points of the pulse). In Fig. 4.3, the reflected signal from two point reflectors are shown for two scenarios. In the top scenario, the two point reflectors are not sufficiently separated for the two echoes to be separated in the received signal, whereas in the bottom scenario they are. For the trailing edge of the echo from the first object to arrive at the receiver before the leading edge of the second echo, the objects must be separated by more than

$$\Delta r = \frac{c\tau_p}{2}, \tag{4.2}$$

where Δr is called the *range resolution* of the radar. From (4.2) we deduce that for a pulse radar the range resolution is directly proportional to the pulse duration.

The expression in (4.2), although intuitive, is misleading for other radar waveforms than the simple pulse, such as linear frequency modulated pulses

(pulse compression) or continuous wave signals. Fundamentally, it is not the duration of the pulse that limits the range resolution of the radar but rather the bandwidth of the transmitted signal. A more general expression for (4.2), which is frequently used, is

$$\Delta r = \frac{c}{2B}, \tag{4.3}$$

where B is the bandwidth of the baseband waveform, in practice taken as the 3dB bandwidth. Note that a reasonable approximation of the 3dB bandwidth of a simple pulse is $B \approx 1/\tau_p$.

Range accuracy

Intuitively, the ability to determine the range to an object is dependent on two things: the "sharpness" of the pulse shape and SNR. In other words, the accuracy is dependent on how informative the waveform is and how clear the response from the target is in relation to the noise in the system.

In [49], the CRLB of the range error variance for a point target is derived. Assuming that a band-limited signal with high SNR is processed using a matched filter and that the noise is additive and described by a Gaussian density with zero-mean, the standard deviation in range error is bounded by

$$\sigma_r \geq \frac{c}{2}\frac{1/\beta}{\sqrt{2E/N_0}}, \tag{4.4}$$

where $2E/N_0$ is the SNR after the matched filter and β is the effective bandwidth of the transmitted waveform. The effective bandwidth is given by,

$$\beta^2 = \frac{\int_{-\infty}^{\infty}(2\pi f)^2|S(f)|^2 df}{\int_{-\infty}^{\infty}|S(f)|^2 df} = \frac{\int_{-\infty}^{\infty}(2\pi f)^2|S(f)|^2 df}{E}, \tag{4.5}$$

where $S(f)$ is the Fourier transform of the transmitted baseband waveform. The effective bandwidth of a signal is a measure of its spread of energy in the frequency domain, analogous to variance in probability theory.

As intuition predicted, the variance of the range error is inversely propor-

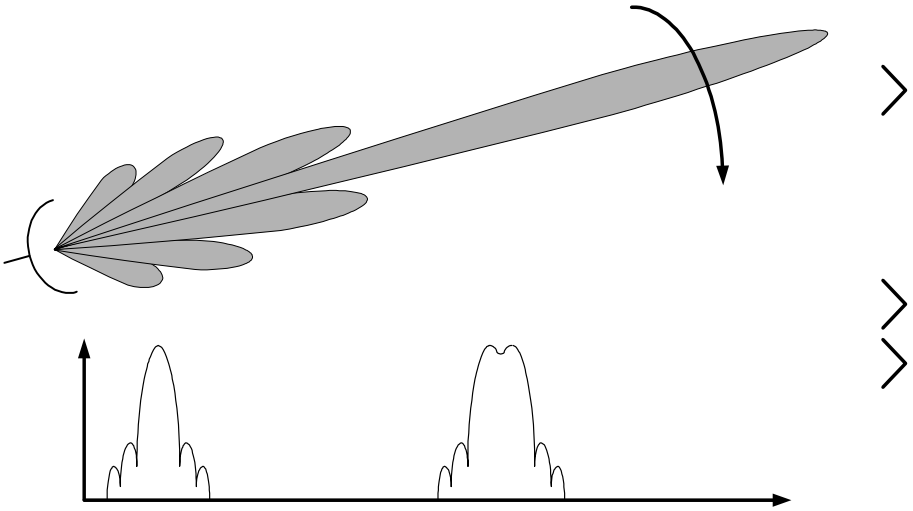


Figure 4.4: Illustration of the angle resolution of a scanning radar.

tional to how informative the waveform is (bandwidth) and the clearness of the response (SNR).

4.3.2 Angular measurement

The angle to a radar reflecting object, denoted φ, can be deduced by concentrating the radar energy in a narrow antenna beam, by estimating the direction of arrival of the reflected wave, or by combining the two. The first principle, shown in Fig. 4.4, is used by scanning radars, which sweeps the area of interest (in a known pattern) with a concentrated beam. Hence, an object is only detected if the antenna is pointed in the direction of the object. Using an antenna array or a monopulse antenna (antennas with multiple receive channels), the direction of arrival of the reflected radar signal can be estimated by sampling the instantaneous response from different parts of the antenna.

Angular resolution

Similar to the difficulty of separating two close objects in range, the radar can not resolve closely spaced objects in angle either. The concept is illustrated in Fig. 4.4 for a scanning radar, but similar reasoning holds also for other types

of angle estimators. In the figure, there are three stationary point objects that are located at the same range from the sensor and that are equally good radar reflectors. The antenna radiates energy according to the antenna beam pattern and when the beam is swept from top to bottom, the return power is registered for each direction as shown in the bottom plot. The return from each object is at its maximum when the beam is pointed directly at it but a response is given at other angles as well. As a result, the return power from each object is spread in angle according to the antenna beam pattern and, whereas the top object is clearly distinguishable, the returns from the bottom pair are merged together.

From the considered example it is clear that the capability of the radar to resolve closely spaced objects in angle is dependent on the width of the antenna pattern. A practical definition of angle resolution would hence be

$$\Delta\varphi = \theta_{3\text{dB}}, \tag{4.6}$$

where $\theta_{3\text{dB}}$ is the beam-width between the half-power points in the antenna pattern. However, the fact that two objects are separated more than $\Delta\varphi$ is not a guarantee for that they are actually resolved. For example, if the return power from one of the objects is much greater than from the other, the return from the weaker object will drown.

Angular accuracy

The angular accuracy is, similar as for range, dependent on the SNR and the sharpness of the antenna beam pattern. The CRLB of the angular estimation error is derived in Chapt. 10 of [49] and is expressed as,

$$\sigma_{\varphi/\lambda} \geq \frac{1/\gamma}{\sqrt{2E/N_0}}, \tag{4.7}$$

where $\sigma_{\phi/\lambda}$ is the standard deviation for the angular error normalised by the wave length, λ, of the carrier and γ is the effective aperture width (analogous

to the effective bandwidth). The effective aperture width is given by

$$\gamma^2 = \frac{\int_{-\infty}^{\infty}(2\pi x)^2|A(x)|^2dx}{\int_{-\infty}^{\infty}|A(x)|^2dx}, \tag{4.8}$$

where $A(x)$ is the aperture illumination function describing the amplitude distribution across the aperture (x direction).

For a rectangular receiving aperture of width D, the effective aperture width is found as [49]

$$\gamma = \frac{\pi}{\sqrt{3}}D,$$

and the half-power beam with is given as $\theta_{3\text{dB}} \approx 0.88\lambda/D$. The minimum theoretical angular error can thus be expressed as,

$$\sigma_\varphi \geq \frac{\sqrt{3}}{\pi}\frac{\lambda/D}{\sqrt{2E/N_0}} \approx \frac{0.628\theta_{3\text{dB}}}{\sqrt{2E/N_0}}. \tag{4.9}$$

Hence, to get a finer angular resolution, one can increase the transmitted power or make a more concentrated beam by increasing the antenna aperture (size) D or shortening wave length λ (by increasing the carrier frequency).

4.3.3 Range rate measurement

All objects having a relative radial velocity (range rate), denoted v_r, to the radar sensor will change the frequency of the reflected electromagnetic wave. This phenomenon is known as the *Doppler effect*.

Suppose an object is at a distance R from a radar emitting an electromagnetic wave with wave length λ. The 2-way path between the radar and object is $2R/\lambda$ number of wave lengths, each corresponding to a phase rotation of 2π. In total, the signal has had a phase rotation

$$\delta\phi = \frac{4\pi R}{\lambda}, \tag{4.10}$$

during its transit to and from the object. As a result, at a given time, the

phase of the received signal, ϕ_r, will lag that of the transmitted signal, ϕ_t, according to,

$$\phi_r(t) = \phi_t(t) - \delta\phi. \tag{4.11}$$

If there is a relative motion between the sensor and the object, R will change as a function of time and, consequently, also will $\delta\phi$. A change in phase with respect to time corresponds to a frequency. If we denote the distance travelled by the wave as $2\tilde{R}(t)$ when there is a relative motion, this frequency is given by

$$\omega_\delta = \frac{d\delta\phi}{dt} = \frac{4\pi}{\lambda}\frac{d\tilde{R}(t)}{dt} \approx \frac{4\pi v_r}{\lambda}, \tag{4.12}$$

where the approximation is accurate as long as $v_r \ll c$.

Assuming that the radial velocity is constant during the time for it takes the electromagnetic wave to travel to the object and back, the received angular frequency, given by the time derivative of (4.11), is

$$\frac{d\phi_r(t)}{dt} = \omega_c - \frac{4\pi v_r}{\lambda} = \omega_c + 2\pi f_d, \tag{4.13}$$

where ω_c is the carrier angular frequency and f_d is called the doppler frequency given by,

$$f_d = -\frac{2v_r}{\lambda}. \tag{4.14}$$

The frequency of the return signal can thus be written as,

$$f_r = f_c + f_d. \tag{4.15}$$

From (4.14) it is possible to deduce that the received frequency will increase if $v_r < 0$ (closing velocity) and decrease if $v_r > 0$ (opening velocity). Hence, by measuring the difference in frequency between the transmitted and received signal, the radar can estimate the relative radial velocity of the object.

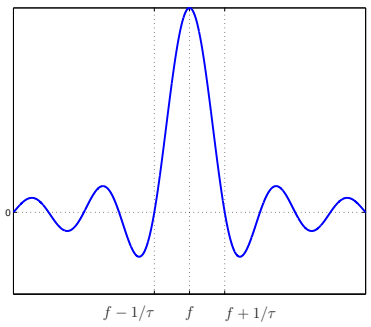


Figure 4.5: The frequency content of a τ seconds long sinusoidal with frequency f.

Range rate resolution

As explained in the previous section the range rate of an object is manifested as a doppler shift in the frequency of the return signal. How well a radar can distinguish two closely spaced targets in range rate can hence be regarded as a frequency estimation problem.

Suppose that we have a signal $s(t)$ with a duration of τ seconds, which contains a single frequency, f. The total frequency content of this signal is described by a sinc-shaped function centred at f and where the first null is at $f \pm 1/\tau$ as shown in Fig. 4.5. Analogous to the angular resolution discussed in Section 4.3.2, if two signals are too close in frequency relative to the width of the main lobe of the sinc-function, the two individual signals will be hard to distinguish. As for the angular resolution, the frequency resolution is often defined as the half-power (-3dB) width of the single frequency response of a signal of length τ, given by

$$\Delta f = f_{3\text{dB}} = \frac{1}{\tau}. \tag{4.16}$$

The range rate resolution is thus,

$$\Delta \dot{r} = \frac{\lambda}{2}\Delta f = \frac{\lambda}{2\tau}. \tag{4.17}$$

From (4.17) we can deduce that in order to have a good range rate res-

olution we need to observe the return signal from the object under a long time. This is the opposite to the requirement for a good range resolution using the simple pulse radar. There, the pulse should be as short as possible to get good range resolution.

Range rate accuracy

The CRLB for the range rate error is derived in similar manner as for range and angle, resulting in the expression,

$$\sigma_{\dot{r}} \geq \frac{\lambda}{2} \frac{1/\alpha}{\sqrt{2E/N_0}}. \tag{4.18}$$

The parameter α is defined as the effective time duration of the transmitted signal $s(t)$, given by

$$\alpha^2 = \frac{\int_{-\infty}^{\infty} (2\pi t)^2 s^2(t) dt}{\int_{-\infty}^{\infty} s^2(t) dt} = \frac{\int_{-\infty}^{\infty} (2\pi t)^2 s^2(t) dt}{E}, \tag{4.19}$$

which can be viewed as a normalised second order moment of the transmitted signal, $s^2(t)$, in this case taken around $t = 0$.

As with range and angular error, the error in range rate can be made arbitrary small by increasing the SNR. Although in practice there is an upper limit on how large SNR that can be obtained. If the SNR is fixed, however, the accuracy in range and range rate are coupled. In this case the range error is inversely proportional to the effective bandwidth occupied by the signal, β, whereas the range rate error is inversely proportional to the effective time duration of the signal, α. It is a well-known mathematical fact that a narrow time signal has a wide spectrum and that a wide time signal has a narrow spectrum and that both the time signal and the frequency spectrum cannot be made arbitrary small simultaneously. This leads to the so-called uncertainty relation

$$\beta\alpha \geq \pi \tag{4.20}$$

which is proven in Chap. 10 of [49]. From (4.20) we can express to which accuracy range and range rate can be measured simultaneously as,

$$\min\{\sigma_r\sigma_{\dot{r}}\} = \frac{\lambda c}{4}\frac{1}{\beta\alpha\sqrt{2E/N_0}} \leq \frac{\lambda c}{4}\frac{1}{\pi\sqrt{2E/N_0}}. \tag{4.21}$$

That is, range and range rate can be simultaneously estimated to as small theoretical error as desired by designing the radar to give sufficient SNR or, for fixed SNR, choose a waveform that gives a large value of $\beta\alpha$. A waveform which has a large $\beta\alpha$ has both a long duration and a wide bandwidth. Typically, practical constraints, such as power limitations and the time to scan the area of interest limit the theoretical accuracy.

4.4 Signal model

This section introduces the basic signal processing models used in a detection radar. A more thorough account is given in the books [43] and [49].

4.4.1 Transmit signal

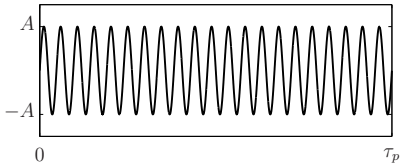
The transmitted signal from a radar is described by

$$s_t(t) = a(t)cos(\omega_c t + \theta(t)), \tag{4.22}$$

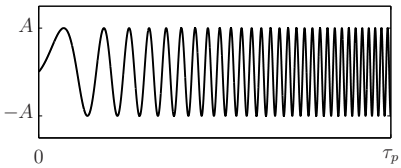
where ω_c is the carrier at radio frequency, $a(t)$ is the envelop function describing the amplitude modulation of the signal, and $\theta(t)$ accounts for phase and frequency modulations of the carrier. Using the model in (4.22), a simple pulsed waveform can be described by setting $\theta(t)$ to zero and $a(t)$ as a periodic rectangular function of width τ_p. Other examples of possible waveforms are depicted in Fig. 4.6.

To facilitate later analysis it is convenient to represent the real signal in (4.22) by its complex equivalent

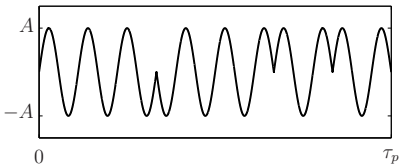
$$s_t(t) = a(t)e^{j(\omega_c t+\theta(t))}. \tag{4.23}$$



(a) Simple pulse



(b) Linear frequency modulated pulse



(c) Binary phase-coded pulse

Figure 4.6: Three different examples of modulated waveforms.

Further, the baseband part of (4.23) is called the *complex envelop* of the waveform, defined as

$$u(t) = a(t)e^{j\theta(t)}, \tag{4.24}$$

and describes the amplitude and phase or frequency modulation applied to the carrier.

The design of the waveform directly influences many of the basic radar performance metrics. For example, the amplitude of the waveform directly affects the signal-to-noise ratio and consequently the accuracy in all measurement dimensions. As we saw in the previous section, the spread of the signal spectrum (bandwidth) determines the resolution and accuracy of the range estimates and the time duration of the transmitted signal influences the resolution and accuracy in range rate.

In a matched filter receiver, as the one depicted in Fig. 4.2, the resolution properties of a waveform is described by its *ambiguity function* defined by the integral

$$\chi(t, \omega_{\mathrm{d}}) = \frac{1}{E} \int_{-\infty}^{\infty} u(s)u^*(s-t)e^{j\omega_{\mathrm{d}}s} ds. \tag{4.25}$$

The ambiguity function describes the output of the matched filter receiver

(correlation receiver) in terms of how well the filter is matched to the received signal in time delay, t, and doppler shift, ω_d. Consequently, the amplitude maximum is found as $|\chi(0,0)| \geq |\chi(t,\omega_\mathrm{d})|$, i.e., when the received signal and the matched filter are perfectly matched. Thus, an object whose response is perfectly matched to the receiver will give rise to an output proportional to $\chi(t,0)$ (for simplicity we assume that the range delay of the returned pulse corresponds to time 0). The return from another object with similar SNR, but with slightly different time delay, τ, and doppler shift, ω_d, will for the same filter give rise to an output proportional to $\chi(t-\tau,\omega_\mathrm{d})$. If $\chi(t,0)$ and $\chi(t-\tau,\omega_\mathrm{d})$ are not significantly different, the return from the two objects are difficult to resolve.

From a resolution point of view, it is most favourable that $\chi(t,\omega_\mathrm{d})$ has a peak at $(0,0)$ and is zero everywhere else. However, it is possible to show that, irrespectively of the waveform, $|\chi(0,0)| = 1$, and that the volume under $|\chi(t,\omega_\mathrm{d})|^2$ (energy) is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\chi(t,\omega_\mathrm{d})|^2 dt\, d\omega_\mathrm{d} = 1. \tag{4.26}$$

From these two properties of the ambiguity function, it is possible to deduce that even if we make the peak at $(0,0)$ narrow, the limit in the amplitude at this point, and the constant volume of $|\chi(t,\omega_\mathrm{d})|^2$ means that the mass of $|\chi(t,\omega_\mathrm{d})|$ must appear at other places, further away from the central peak. As a result, if the waveform offers good resolution in the centre (where t and ω_d are zero), ambiguities will appear off-centre in the form of large values.

4.4.2 Radar equation

The performance of the radar, in terms of the probability of detecting an object and in terms measurement precision, is highly dependent on the SNR of the received signal after the matched filter. The expected SNR in an echo from an object is expressed by the radar equation, in terms of system and object parameters. There are many flavours of this equation depending on which aspects of the system that are of interest. A more detailed discussion

is found in, e.g., [49]. In this thesis we consider the equation on the following form

$$SNR(r,\varphi) = \frac{E_t G_t(\varphi) A_r(\varphi) RCS}{(4\pi)^2 r^4 k T_s C_B L}, \tag{4.27}$$

where r is the range and φ is the azimuth angle to the object reflecting the signal. The parameters in (4.27) are defined as

E_t : energy in the transmitted signal waveform $[Ws]$
$G_t(\varphi)$: antenna gain of the transmit antenna in the φ direction [ ]
$A_r(\varphi)$: effective receive aperture for an incident wave at φ $[m^2]$
RCS : radar cross section (RCS) of the reflecting object $[m^2]$
kT_sC_B : noise energy $[Ws]$
L : general loss term [ ].

The transmitted energy, E_t, is determined by the amplitude and length of the complex envelop $u(t)$. The parameters G_t and A_r depend on the directivity (and size) of the transmit and receive antennas, respectively. The noise primarily enters the signal as thermal noise in the receiver, which is described by the factor kT_sC_B, where k is Boltzmanns' constant and T_s is the system temperature. The term C_B accounts for the fact that the received signal passes through a set of filters. In the ideal case, C_B is one, which occurs when the received signal is filtered by a matched filter.

All of the parameters mentioned above are relatively easy to determine based on the design of the radar system. What is more difficult to know a priori, however, is the RCS of the reflecting object. The RCS is a measure of how well the object reflects the radar energy, and for a nontrivial object, this is sensitive to changes in distance and aspect ratio of the object relative to the sensor. This, among other things, is discussed in the following sections.

4.4.3 Object return model

Assuming that the reflection process is linear and frequency independent over the bandwidth of the signal, the complex representation of the reflected

signal at the receiver (before filtering) from an ideal point object at range r and at an azimuth angle φ can be written as

$$\begin{aligned} s_r(t) &= A(r,\varphi) g_{\text{rcs}} a(t-\tau) e^{j[\omega_c(t-\tau)+\theta(t-\tau)+\omega_{\text{d}}t+\phi]} \\ &= A(r,\varphi) g_{\text{rcs}} e^{j\phi} s_t(t-\tau) e^{j\omega_{\text{d}}t}, \end{aligned} \tag{4.28}$$

where $g_{\text{rcs}} = \sqrt{RCS}$, and

$$A(r,\varphi) = \sqrt{\frac{G_t(\varphi) A_r(\varphi)}{(2\pi)^2 r^4 L}} \tag{4.29}$$

is the attenuation. The time delay, τ, is the time it takes for the signal to travel from the radar to the object and back, and the frequency shift, ω_{d}, is due to the doppler effect described in Section 4.3.3. Further, the reflected signal is subjected to an unknown phase shift, ϕ. The expression in (4.28) is thus an attenuated, time delayed, as well as frequency and phase shifted version of the transmitted signal. The attenuation is described by $A(r,\varphi)$ and is due to general losses in the system, energy dilution (energy spreading) during propagation, and the gain of the transmit and receive antenna in the direction of the object.

The objects under consideration thus far have been regarded as ideal point sources, where all objects are assumed to reflect the radar signal in a single point, and to have the same (constant) radar cross section. A more complex object can be modelled as a set of ideal point objects, where the i^{th} point object is positioned at a range r_i and angle φ_i. The return from this type of object is thus an aggregation of the returns from the individual point objects,

$$\begin{aligned} s_r(t) &= \sum_i A(r_i,\varphi_i) g_{\text{rcs},i} e^{j\phi_i} a(t-\tau_i) e^{j[\omega_c(t-\tau_i)+\theta(t-\tau_i)+\omega_{\text{d},i}t]} \\ &= \sum_i A(r_i,\varphi_i) g_{\text{rcs},i} e^{j\phi_i} s_t(t-\tau_i) e^{j\omega_{\text{d},i}t}. \end{aligned} \tag{4.30}$$

From (4.30) we deduce that the return from a more complex object gives a more spread response, in time (range), angle and frequency (doppler) than

the ideal point object. The spreading of the signal affects the measurement accuracies negatively compared to the ideal point source cases discussed in Section 4.3. This effect is small if the spread is small in relation to the resolution in the different dimensions (point source assumption), but if the spreading is even and comparable to the resolution, the effect is quite noticeable (extended object).

The time delay (range) differences between the individual reflectors in relation to the carrier frequency and the induced individual phase shifts, determine if the returns from the point objects are added coherently (constructively) or incoherently. As a result, the RCS of a complex object fluctuates with just small changes in range. This is easily seen for a simple pulse, i.e., where $\theta(t)$ is zero, and assuming that the spread in time delay, angle and doppler is small compared to the resolution cell. In this case, the received signal can be written as,

$$\begin{aligned} s_r(t) &= \sum_i A(r_i, \varphi_i) g_{\mathrm{RCS},i} e^{j(-\omega_c \tau_i + \phi_i)} a(t - \tau_i) e^{j(\omega_c t + \omega_{\mathrm{d},i} t)} \\ &\approx \left(\sum_i g_{\mathrm{RCS},i} e^{j(-4\pi r_i/\lambda_\mathrm{c} + \phi_i)} \right) A(r, \varphi) a(t - \tau) e^{j(\omega_c t + \omega_\mathrm{d} t)}, \end{aligned} \tag{4.31}$$

where r, φ and ω_d is the mean of the position of the point reflectors in respective dimension. The RCS of the complex object is an aggregation of the contribution from each reflector. By defining the complex sum,

$$g_\mathrm{RCS} = \sum_i g_{\mathrm{RCS},i} e^{j(-4\pi r_i/\lambda_\mathrm{c} + \phi_i)}, \tag{4.32}$$

the RCS of the complex object is defined as $|g_\mathrm{RCS}|^2$. Assume that the number of point reflectors is large, that the RCS of the individual point reflectors, $g_{\mathrm{rcs},i}$, are identically distributed and that the phase shifts $(-4\pi r_i/\lambda_\mathrm{c} + \phi_i)$ are independent and uniformly distributed. The central limit theorem then gives us

$$\mathrm{Re}\{g_\mathrm{RCS}\}, \mathrm{Im}\{g_\mathrm{RCS}\} \sim \mathcal{N}(0, \sigma^2_\mathrm{RCS}). \tag{4.33}$$

From (4.33) it follows that the RCS of a complex object can be modelled as

$$|g_{\text{RCS}}|^2 \sim \text{Exp}\left(\frac{1}{2\sigma^2_{\text{RCS}}}\right), \tag{4.34}$$

where $\text{E}\{|g_{\text{RCS}}|^2\} = 2\sigma^2_{\text{RCS}}$. This is the Swerling I (and Swerling II) model [52] of fluctuating target RCS. Other assumptions about the properties of the components in (4.32) will lead to slightly different models, see e.g. [43]. However, it is clear that a small change in the range to the object (fraction of a wavelength) will cause the terms in (4.32) to add up completely differently and the resulting RCS is thus an independent realisation of (4.33). Additionally, if the aspect angle changes, some of the point reflectors in (4.32) are no longer visible, while other points on the object appear. This leads to a change in the expected object RCS, $\text{E}\{|g_{\text{RCS}}|^2\}$.

4.4.4 Receiver and matched filtering

The aim of the receiver is to demodulate the received signal to baseband and to extract the relevant information from the signal in order to facilitate the detection of objects and to estimate their properties. The demodulation is typically done with an *In-phase* and *Quadrature* mixer. Using an I/Q demodulator, the general received signal in (4.30) is transformed to a baseband signal on the form,

$$\begin{aligned} s_b(t) &= \sum_i A(r_i, \phi_i) g_{\text{rcs}} u(t - \tau_i) e^{j\omega_{\text{d},i} t} + n(t) \\ &= I(t) + jQ(t), \end{aligned} \tag{4.35}$$

where $u(t)$ is the complex envelop of the waveform, $n(t)$ describes the thermal noise in the receiver, and

$$I(t) = \text{Re}\{s_b(t)\} \tag{4.36}$$

$$Q(t) = \text{Im}\{s_b(t)\}, \tag{4.37}$$

are the real signal representations of the complex received signal. The noise is modelled as an additive white complex Gaussian noise process with variance σ_n^2 in each component (due to bandwidth limitations the noise is in practice non-white).

Given that the return signal only consists of one ideal point object (there is only one term in the sum in (4.35)), it is proven in [30] that the SNR of the signal is maximised using a matched filter (correlation receiver), with impulse response

$$h(t) = u^*(\tau_{\text{int}} - t)e^{-j\omega_{\text{d}}(\tau_{\text{int}}-t)}, \tag{4.38}$$

where ω_{d} is the doppler shift caused by the range rate of the object, and τ_{int} is an integration time used to make the filter causal. The integration time is typically chosen such that $h(t) = 0$ for $t \notin [0,\ \tau_{\text{int}}]$, i.e., for a simple pulse system $\tau_{\text{int}} = \tau_{\text{p}}$. To simplify the following discussions, the amplitude of $u(t)$ is set to unity and the amplitude of the transmitted signal is instead incorporated into $A(r, \phi)$.

Applying the matched filter in (4.38) to the general signal in (4.35), and denoting $A_i = A(r_i, \varphi_i)g_{\text{rcs}}$, and $\omega_{\text{d},i} = \omega_{\text{d}} + \Delta\omega_i$, the output of the matched filter can be written as,

$$\begin{aligned}
s_b * h(t) &= \int_{-\infty}^{\infty} \sum_i A_i u(s-\tau_i) e^{j(\omega_{\text{d}}+\Delta\omega_i)s} u^*(s+\tau_{\text{int}}-t) e^{-j\omega_{\text{d}}(s+\tau_{\text{int}}-t)} ds \\
&= \sum_i A_i \int_{-\infty}^{\infty} u(s-\tau_i) u^*(s+\tau_{\text{int}}-t) e^{j(\omega_{\text{d}}+\Delta\omega_i)s - j\omega_{\text{d}}(s+\tau_{\text{int}}-t)} ds \\
&= e^{j\omega_{\text{d}}(t-\tau_{\text{int}})} \sum_i A_i \int_{-\infty}^{\infty} u(s-\tau_i) u^*(s+\tau_{\text{int}}-t) e^{j\Delta\omega_i s} ds \\
&= e^{j\omega_{\text{d}}(t-\tau_{\text{int}})} \sum_i A_i \chi(t-\tau_i-\tau_{\text{int}}, \Delta\omega_i),
\end{aligned} \tag{4.39}$$

where $\chi(\cdot,\cdot)$ is the *ambiguity function* of the transmitted waveform as defined in (4.25). As previously discussed, the resolution capability of the sensor is in large part determined by the choice of waveform design. From (4.39) we deduce that the signal-dependent output of the matched filter from a complex

object is the superposition of scaled and shifted ambiguity functions.

In (4.39), the filter response of the signal part of the received signal is derived. To describe the complete output of the receiver, we also need to consider the effect of the matched filtering on the noise process $n(t)$. Clearly, low-pass filtering (which is what the matched filter performes) of a random process will lead to time correlation. This correlation is described by the autocorrelation function of the filtered noise process, defined as

$$\begin{aligned} R(\tau) &= \text{Cov}\left\{n * h(t), n * h(t+\tau)\right\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(\omega)|^2 S_n(\omega) e^{j\omega\tau} d\omega \\ &= \frac{2\sigma_n^2}{2\pi}\int_{-\infty}^{\infty} |H(\omega)|^2 e^{j\omega\tau} d\omega, \end{aligned} \tag{4.40}$$

where $H(\omega)$ is the frequency response of the matched filter and $S_n(\omega) = 2\sigma_n^2$ is the power spectral density of $n(t)$. As the matched filter has unit energy, the variance of the noise process after the matched filter is,

$$R(0) = 2\sigma_n^2. \tag{4.41}$$

Consequently, the variance of the noise is unaffected by the matched filter, but its time-correlation is increased.

4.4.5 Detection

Suppose we want to detect if there is an object at a certain distance, corresponding to a time delay τ, and with a certain range rate, corresponding to a doppler shift ω_{d}. This problem can be formulated as a binary detection problem, where the two hypotheses are:

$$\mathcal{H}_1 \ : s_b(t) = A g_{\text{RCS}} u(t-\tau) e^{j\omega_{\text{d}} t} + n(t), \qquad \tau < t \leq \tau + \tau_{\text{int}} \tag{4.42}$$

$$\mathcal{H}_0 \ : s_b(t) = n(t), \qquad \tau < t \leq \tau + \tau_{\text{int}}, \tag{4.43}$$

where A is the deterministic part of the amplitude of the return signal. As the phase of the signal is noninformative, a sufficient statistic for detection

can be formed as [30]

$$\Lambda = |s_b * h(t)|^2\big|_{t=\tau+\tau_{\text{int}}} = I_m^2(t) + Q_m^2(t)\big|_{t=\tau+\tau_{\text{int}}}, \tag{4.44}$$

where $h(t)$ is the impulse response of the filter matched to the signal we want to detect, and I_m and Q_m are the output of the matched filter in the I and Q channels, respectively. In the case of a signal present ($\mathcal{H}_1$), the expression in (4.44) becomes

$$\Lambda = 2A^2|g_{\text{RCS}}|^2 + 2\sigma_n^2, \tag{4.45}$$

whereas in the case of only noise ($\mathcal{H}_0$), we have

$$\Lambda = 2\sigma_n^2. \tag{4.46}$$

Hence, if the energy in the return signal, $2A^2|g_{\text{RCS}}|^2$, is large in relation to the power spectral density of the noise, $2\sigma_n^2$, i.e. high SNR, these two hypotheses are well separated.

The optimal decision that maximises the probability of detection (making a correct detection) for a given probability of false alarm (detecting a false target), is found as

$$\Lambda \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{4.47}$$

where γ is a threshold.

In the discussed problem, it is assumed that the time delay and doppler shift of the reflected signal is known. This is, however, not very realistic in a real radar situation. As a result, we need to sample the matched filter at all possible time-delays and doppler shifts. Furthermore, since it is unlikely that we will obtain a sample at the precise time delay and doppler shift of the signal, we suffer from additional performance losses.

Bibliography

[1] *The American Heritage Science Dictionary*. Houghton Mifflin Company, 2005.

[2] J. Ackermann and T. Bunte. Yaw disturbance attenuation by robust decoupling of car steering. *Control Engineering Practice*, 5(8):1131–1136, August 1997.

[3] M.S. Arulampalam, N. Gordon, M. Orton, and B. Ristic. A variable structure multiple model particle filter for gmti tracking. *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat.No.02EX5997)*, vol.2:927 – 34, 2002.

[4] M. Athans, R. Wishner, and A. Bertolini. Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements. *Automatic Control, IEEE Transactions on*, October 1968.

[5] Y. Bar-Shalom and W. Dale Blair. *Multitarget-Multisensor Tracking Volume III: Applications and Advances*. Artech House, Norwood, MA, 2000.

[6] Y. Bar-Shalom, K.C. Chang, and H.A.P. Blom. Automatic track formation in clutter with a recursive algorithm. *Decision and Control, 1989., Proceedings of the 28th IEEE Conference on*, pages 1402–1408 vol.2, 13-15 Dec 1989.

[7] Y. Bar-Shalom, X. Rong Li, and T. Kirubarajan. *Estimation with applicarions to tracking and navigation*. Wiley, New York, NY, 2001.

[8] Y. Bar-Shalom and E. Tse. Tracking in a cluttered environment with probabilistic data association. *Automatica*, 11, September 1975.

[9] M. Bayly, B. Fildes, M. Regan, and K. Young. Review of crash effectiveness of intelligent transport systems. Trace project deliverable D4.1.1.

[10] F. Bengtsson and L. Danielsson. Designing a real time sensor data fusion system with application to automotive safety. Technical Report R007/2008, Signal and Systems, Chalmers University of Technology, Göteborg, Sweden, April 2008.

[11] D.P. Bertsekas. The auction algorithm for assignment and other network flow problems: a tutorial. *Interfaces*, 20(4):133 – 49, 1990.

[12] S. Blackman and R. Popoli. *Design and analysis of modern tracking systems*. Artech House, Norwood, MA, 1999.

[13] Lars Danielsson, Henrik Lind, and Stig Jonasson. Insafes hci principles for integrated adas applications. In Constantine Stephanidis, editor, *HCI (6)*, volume 4555 of *Lecture Notes in Computer Science*, pages 339–348. Springer, 2007.

[14] T. A. Dingus, S.G. Klauer, V.L. Neale, A. Petersen, S.E. Lee, J. Sudweeks, M.A. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z.R. Doerzaph, J. Jermeland, and R.R. Knipling. The 100-car naturalistic driving study, phase ii results of the 100-car field experiment. Technical Report DOT HS 810 593, Virginia Tech Transportation Institute, Washington, DC, April 2006.

[15] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer Verlag, New York, 2001.

[16] A. Doucet, S. Godsill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statist. Comput.*, 10(3):197–208, 2000.

[17] S. Ferguson. The effectiveness of electronic stability control in reducing real-world crashes: A literature review. *Traffic Injury Prevention*, 8(4):329 – 338, December 2007. ISSN: 1538-9588.

[18] Directorate-General for Energy and Transport with Eurostat. *Statistical pocket book 2007*, volume Energy and transport in figures 2007, Part 3 : Transport. European Commission, Luxenbourg, 2007.

[19] T. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal of Oceanic Engineering*, 8, July 1983.

[20] NHTSA Benefits Working Group. *Preliminary Assessment of Crash Avoidance Systems Benefits*. Washington, DC, October 1996.

[21] J. Gunnarsson, L. Svensson, F. Bengtsson, and L. Danielsson. Joint driver intention classification and tracking of vehicles. In *Nonlinear Statistical Signal Processing Workshop 2006*, Cambridge, UK, September 2006.

[22] J. Gunnarsson, L. Svensson, L. Danielsson, and F. Bengtsson. Tracking vehicles using radar detections. In *Intelligent Vehicles Symposium, 2007. Proceedings IEEE*, Istanbul, June 2007.

[23] C.H.R. Hülsmeier. Verfahren, um entfernte metallische gegenstände mittels elektrischer wellen einem beobachter zu melden. *Eurpean Patent Office, Patentschrift DE165546*, 30, 1904.

[24] S. Julier and J. Uhlmann. A new extension of the kalman filter to nonlinear systems. In *The Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls, Multi Sensor Fusion, Tracking and Resource Management II*, 1997.

[25] S. J. Julier and J. K. Uhlmann. A general method for approximating nonlinear transformations of probability distributions. Can be downloaded from http://www.cs.uwc.ac.za/ kschef/.

[26] S.J. Julier. The scaled unscented transformation. *American Control Conference, 2002. Proceedings of the 2002*, 6:4555–4559 vol.6, 2002.

[27] S.J. Julier and J.K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3), 2004.

[28] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D. Journal Basic Eng.*, 82:24–45, 1960.

[29] S. M. Kay. *Fundamentals of Statistical Signal Processing - Estimation Theory*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[30] S.M. Kay. *Fundamentals of statistical signal processing - detection theory*. Prentice-Hall, 1998.

[31] T. Kirubarajan and Y. Bar-Shalom. Probabilistic data association techniques for target tracking in clutter. *Proceedings of the IEEE*, 92(3):536–557, Mar 2004.

[32] S.G. Klauer, T. A. Dingus, V. L. Neale, , J. D. Sudweeks, and D. J. Ramsey. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. Technical Report DOT HS 810 594, Virginia Tech Transportation Institute, Washington, DC, April 2006.

[33] W. Koch and G. Van Keuk. Multiple hypothesis track maintenance with possibly unresolved measurements. *Aerospace and Electronic Systems, IEEE Transactions on*, 33(3):883–892, July 1997.

[34] T. C. Lansdown, N. Brook-Carter, and T. Kersloot. Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. *Ergonomics*, 47(1):91–104, 2004.

[35] X. Rong Li and V. P. Jilkov. Survey of maneuvering target tracking. Part I. Dynamic models. *Aerospace and Electronic Systems, IEEE Transactions on*, 39(4):1333–1364, 2003.

[36] X. Rong Li and V.P. Jilkov. Survey of maneuvering target tracking. Part III: Measurement models. In *Proceedings of the 2001 SPIE Conference on Signal and Data Processing of Small Targets*, volume 4473, San Diego, CA, 2001.

[37] X. Rong Li and V.P. Jilkov. Survey of maneuvering target tracking. Part VI. Dynamic models. *Signal and Data Processing of Small Targets 2004, Proceedings of SPIE*, 5428:537–550, 2004.

[38] R.P.S. Mahler. *Statistical multisource-multitarget information fusion.* Artech House, Inc. Norwood, MA, USA, 2007.

[39] D. Musicki, R. Evans, and S. Stankovic. Integrated probabilistic data association. *Automatic Control, IEEE Transactions on*, 39(6):1237–1241, Jun 1994.

[40] National Highway Traffic Safety Administration (NHTSA). Federal motor vehicle safety standards; electronic stability control systems; controls and displays. Federal Rule, FMVSS: 126, Parts: 49 CFR Parts 571 and 585.

[41] National Highway Traffic Safety Administration (NHTSA). Automotive collision avoidance system field operational test report: Methodology and results. Technical Report DOT HS 809 900, University of Michigan Transportation Research Institute (UMTRI) and General Motors Research and Development Center, Washington, DC, August 2005.

[42] A. Polychronopoulos, A. Amditis, U. Scheunert, and T. Tatschke. Revisiting JDL model for automotive safety applications: the PF2 functional model. *2006 9th International Conference on Information Fusion*, pages 1–7, July 2006.

[43] M.A. Richards. *Fundamentals of radar signal processing.* McGraw-Hill, 2005.

[44] M. Richter, H.C. Pape, D. Otte, and C. Krettek. Improvements in passive car safety led to decreased injury severity - a comparison between the 1970s and 1990s. *Injury*, 36(4):484–488, April 2005. ISSN: 0020-1383.

[45] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman filter: particle filters for tracking applications.* Artech House, Norwood, MA, 2004.

[46] B. Ristic, A. Farina, D. Benvenuti, and M.S. Arulampalam. Performance bounds and comparison of nonlinear filters for tracking a ballistic object on re-entry. *Radar, Sonar and Navigation, IEE Proceedings* -, April 2003.

[47] C. P. Robert. *The Bayesian Chioce, Second Edition.* Springer, New York, NY, 2001. ISBN 0-387-95231-4.

[48] R.W. Sittler. An optimal data association problem in surveillance theory. *IEEE Transactions on Military Electronics*, MIL-8(2):125 – 139, 1964.

[49] M.I. Skolnik, editor. *Introduction to radar systems, third edition.* McGraw-Hill Companies, 2001.

[50] G. W. Stimson, editor. *Introduction to Airborne Rada, second edition.* SciTech Publishing, 1998.

[51] L. Svensson and J. Gunnarsson. A new motion model for tracking of vehicles. In *14th IFAC Symposium on System Identification*, Newcastle, Australia, March 2006.

[52] P. Swerling. Probability of detection for fluctuating targets. *Information Theory, IRE Transactions on*, 6(2):269–308, 1960.

[53] R. van der Merwe. Sigma-point kalman filters for probabilistic inference in dynamic state-space models. PhD Dissertations, OGI School of Science and Engineering at Oregon Health and Science University, April 2004.

[54] J. S. Wang, R. R. Knipling, and M. J. Goodman. The role of driver inattention in crashes: new statistics from the 1995 crashworthiness data system. In *40th Annual Proceedings: Association for the Advancement of Automotive Medicine*, pages 377–392, Vancouver, British Columbia, Canada, June 1996.

[55] C. Wiacek and W. Najm. Driver/vehicle characteristics in rear-end pre-crash scenarios based on the general estimates system (ges). Technical Report 1999-01-0817, SAE technial papar series, Warrendale, PA, 1999.