Imperial College London
Department of Electrical & Electronic Engineering

Tracking and Estimation Algorithms for Bearings Only Measurements

Attila Can Özelçi

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering of Imperial College London, December 2012

Declaration of Originality

This thesis is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering of Imperial College London. All the work contained within this thesis is either my own or appropriately referenced.

Acknowledgments

My sincere gratitude goes to Prof. Richard Vinter, whose help, encouragement, kindness and generosity have been paramount to the the completion of this thesis. It is under his supervision and guidance that I have not only acquired the knowledge I have today, but, more importantly, have learnt to learn. The fruitful discussions with Dr. J.M.C. Clark and Dr. Aris Kountouriotis and their helpful insights are also gratefully acknowledged.

I would also like to thank my parents who have always put my education above all of their priorities. Last but definitely not least, these past years would not have been the same without the support, love and friendship of Anastassia, and of my friends, notably Christos, Mario, Philip, Daniel, Thulasi, Andrea and Michele.

Abstract

The Bearings-only tracking problem is to estimate the state of a moving object from noisy observations of its direction relative to a sensor. The Kalman filter, which provides least squares estimates for linear Gaussian filtering problems is not directly applicable because of the highly nonlinear measurement function of the state, representing the bearings measurements and so other types of filters must be considered. The shifted Rayleigh filter (SRF) is a highly effective moment-matching bearings-only tracking algorithm which has been shown, in 2D, to achieve the accuracy of computationally demanding particle filters in situations where the well-known extended Kalman filter and unscented Kalman filter often fail.

This thesis has two principal aims. The first is to develop accurate and computationally efficient algorithms for bearings-only tracking in 3D space. We propose algorithms based on the SRF, that allow tracking, in the presence of clutter, of both nonmaneuvering and maneuvering targets. Their performances are assessed, in relation to competing methods, in highly challenging tracking scenarios, where they are shown to match the accuracy of high-order sophisticated particle filters, at a fraction of the computational cost.

The second is to design accurate and consistent algorithms for bearings-only simultaneous localization and mapping (SLAM). The difficulty of this problem, originating from the uncertainty in the position and orientation of the sensor, and the absence of range information of observed landmarks, motivates the use of advanced bearings-only tracking algorithms. We propose the quadrature-SRF SLAM algorithm, which is a moment-matching filter based on the SRF, that numerically evaluates the exact mean and covariance of the posterior. Simulations illustrate the accuracy and consistency of its estimates in a situation where a widely used moment-matching algorithm fails to produce consistent estimates. We also propose a Rao-Blackwellized SRF implementation of a particle filter, which, however, does not exhibit favorable consistency properties.

Contents

Declaration of Originality **2**

Acknowledgments **3**

Abstract **4**

Contents **5**

List of Figures **8**

List of Tables **11**

List of Abbreviations **12**

Chapter 1. Introduction **14**

Chapter 2. Estimation Theory and Algorithms **20**

2.1 The Frequentist approach 21

2.1.1 The Maximum Likelihood estimator 21

2.1.2 The Least Squares estimator 22

2.2 The Bayesian Approach 23

2.2.1 The Kalman Filter 25

2.3 Nonlinear Filtering and Approximate Bayesian Estimators 30

2.3.1 Moment Matching Filters 31

2.3.2 Sequential Monte Carlo Methods 39

2.4 Target Tracking Models 45

2.4.1 The Discretized Continuous White Noise Acceleration Model 46

2.4.2 The Discretized Continuous Wiener Process Acceleration Model . . 47
2.4.3 The 3D Constant-Turn Model 48
2.4.4 The Interacting Multiple Model Algorithm 51
2.4.5 The Bearings-only Measurement Model 55
2.5 Bearings-only Tracking Algorithms 56
2.5.1 The Extended Kalman Filter in Cartesian Coordinates 56
2.5.2 The Extended Kalman Filter in Modified Spherical Coordinates . . 57
2.5.3 The Range-parameterized Extended Kalman Filter 61

Chapter 3. Bearings only tracking in 3D **64**
3.1 The Shifted Rayleigh Filter 65
3.1.1 Derivation of Shifted Rayleigh Filter Equations 68
3.1.2 Extension to Data Association 86
3.1.3 Extension to Multiple-Model Maneuvering Target Tracking 94
3.2 The Rauch-Tung-Striebel Shifted Rayleigh Smoother 101

Chapter 4. 3D Bearings-Only Tracking Applications **105**
4.1 Conduct of Simulations 105
4.1.1 Filter Initialization 105
4.1.2 Comparison Metrics 107
4.2 The Posterior Cramér-Rao Lower Bound 108
4.3 Single Sensor Tracking of a Non-Maneuvering Target in Clutter 110
4.4 The Shifted Rayleigh Particle Filter 117
4.5 Multiple Sensor Tracking of a Maneuvering Target in Clutter 121

Chapter 5. The Shifted Rayleigh Filter in Bearings-only SLAM **138**
5.1 Formulation of the Bearings-Only SLAM problem 139
5.2 Bearings-only EKF SLAM 141

5.3 The quadrature-SRF SLAM algorithm 144
5.3.1 Derivation of the Algorithm 144
5.3.2 Landmark Initialization 154
5.4 The Rao-Blackwellized SRPF-SLAM Algorithm 155
5.4.1 Derivation of the Algorithm 156
5.5 Performance of the quadrature-SRF SLAM and RBSRPF SLAM Algorithms 158
5.5.1 Implementation of the IDP-EKF SLAM Algorithm 159
5.5.2 Implementation of the quadrature-SRF SLAM Algorithm 160
5.5.3 Implementation of the RBSRPF SLAM Algorithm 163
5.5.4 Simulation Results 164

Chapter 6. Conclusion **169**

6.1 Achievements 169
6.2 Future Work 172

Bibliography **174**

List of Figures

2.1 Inadequacy of linear approximations of nonlinear transformations 35

2.2 Markov chain for 3 state system. 52

2.3 Azimuth and elevation angles θ and α in the 3D measurement frame 56

3.1 Comparison of the azimuth bearing density with the 'wrapped' Normal density for $\mathbf{y} = (0, 1)^{\mathrm{T}}$ and $\sigma = 0.2, 0.4, \ldots, 1$ 80

3.2 Effect of 'standard' measurement noise on the displacement vector along constant latitudinal and longitudinal lines of the unit sphere 83

3.3 Comparison of marginal azimuth (a) and elevation (b) angle densities with the 'wrapped' normal density for $\alpha = 0$ and $\sigma = 0.1$, 0.2, 0.3 85

3.4 Comparison of marginal azimuth (a) and elevation (b) angle densities with the normal density for $\alpha = \frac{\pi}{4}$ and $\sigma = 0.1$, 0.2, 0.3 85

3.5 Comparison of marginal azimuth (a) and elevation (b) angle densities with the normal density for $\alpha = \frac{\pi}{3}$ and $\sigma = 0.1$, 0.2, 0.3 86

3.6 Comparison of marginal azimuth (a) and elevation (b) angle densities with the normal density for $\alpha = \frac{\pi}{2}$ and $\sigma = 0.1$, 0.2, 0.3 86

3.7 Correct modelling of the bearings density of clutter 93

(a) Uniform spacial distribution of clutter projected onto the unit sphere 93

(b) Non-uniform distribution of clutter in the measurement space 93

3.8 Incorrect modelling of the bearings density of clutter 93

(a) Non-uniform spacial distribution of clutter projected onto the unit sphere 93

(b) Uniform distribution of clutter in the measurement space 93

4.1 Target/sensor platform geometry (projected on the xy plane) 110

4.2 Sample azimuth/elevation measurements for $p_c = 0.5$ 112

4.3 Azimuth/elevation bearing rates (no clutter) 113

4.4 Target range RMS error ($p_c = 0.5$, 200 MC runs) 114

4.5 Percentage of track divergence versus probability of clutter 115

4.6 Range TARMSE of the target over 10 repetitions of 200 MC runs, with $p_c = 0.5$. 119

4.7 Percentage of track divergence over 10 repetitions of 200 MC runs, with $p_c = 0.5$. 120

4.8 Target/sensor geometry (projected on the xy plane) 123

4.9 Target range from sensor 1 124

4.10 Target speed 124

4.11 Target acceleration 125

4.12 Sample realization of azimuth and elevation bearing measurements at sensor 1 for $p_c = 0.5$. 126

4.13 Azimuth and elevation bearing rates in the absence of clutter at sensor 1. . 127

4.14 Sample realization of azimuth and elevation bearing measurements at sensor 2 for $p_c = 0.5$. 127

4.15 Azimuth and elevation bearing rates in the absence of clutter at sensor 2. . 128

4.16 Range and velocity RMSEs of the target state estimates for $p_c = 0.5$ using the *first* set of models, over 200 MC runs. 131

4.17 Mode probability estimates for $p_c = 0.5$ using the *first* set of models, over 200 MC runs. 132

4.18 Percentage of track divergence against probability of clutter, using the *first* set of models. 134

4.19 Range and velocity RMSEs ($p_c = 0.5$, *second* set of models) 135

4.20 Mode probability estimates for $p_c = 0.5$ using the *second* set of models, over 200 MC runs. 136

4.21 Percentage of track divergence against probability of clutter, using the *second* set of models. 137

5.1 Geometry of bearings-only SLAM on the plane 140

5.2 Sample run of the quadrature-SRF SLAM algorithm. 164

5.3 Vehicle position and heading errors of the quadrature-SRF SLAM, IDP-EKF SLAM and RBSRPF SLAM algorithms. 166

5.4 Average NEES of vehicle state estimates over 50 MC runs using the quadrature-SRF SLAM, IDP-EKF SLAM and RBSRPF SLAM algorithms. 167

List of Tables

4.1 TARMS range errors (m) . . . 115

4.2 Computation times (relative) . . . 116

4.3 TARMS range errors (m) . . . 117

4.4 PF algorithm using posterior approximation as proposal distribution . . . 118

4.5 Initializing parameters of target state . . . 130

4.6 TARMS range errors (m) - *first* set of models . . . 133

4.7 TARMS range errors (m) - *second* set of models . . . 135

4.8 Computation times (relative) . . . 137

5.1 One iteration of the RBSRPF SLAM Algorithm . . . 157

5.2 TARMS vehicle position and heading errors and computational cost . . . 168

List of Abbreviations

EKF	Extended Kalman Filter
EKPF	Extended Kalman Particle Filter
EnKF	Ensemble Kalman Filter
FLIR	Forward Looking InfraRed
GBP	Generalized Pseudo-Bayesian
IDP	Inverse Depth Parametrization
IEKF	Iterated Extended Kalman Filter
IMM	Interacting Multiple Model
LLSE	Linear Least Squares Estimator
LSE	Least Squares Estimator
MAP	Maximum A Posteriori
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimator
MMSE	Minimum Mean Square Error
MPEKF	Modified Polar Coordinates Extended Kalman Filter
MSCEKF	Modified Spherical Coordinates Extended Kalman Filter
NEES	Normalized Estimation Error Squared
PDA	Probabilistic Data Association
PDF	Probability Density Function
PCRLB	Posterior Cramer-Rao Lower Bound
PF	Particle Filter
RBSRPF	Rao-Blackwellized Shifted Rayleigh Particle Filter
RMSE	Root Mean Square Error
RPEKF	Range-Parameterized Extended Kalman Filter

RPEKS	Range-Parameterized Extended Kalman Smoother
RTS	Rauch-Tung-Striebel
RV	Random Variable
SIR	Sampling Importance Resampling
SLAM	Simultaneous Localization And Mapping
SRF	Shifted Rayleigh Filter
SRF3C	Shifted Rayleigh Filter for measurements in 3D with Clutter
SRMF3C	Shifted Rayleigh Mixture Filter for measurements in 3D with Clutter
SRPF	Shifted Rayleigh Particle Filter
SRS3C	Shifted Rayleigh Smoother for measurements in 3D with Clutter
TARMSE	Time-Averaged Root Mean Square Error
UKF	Unscented Kalman Filter
UPF	Unscented Particle Filter

Chapter 1

Introduction

Tracking and estimation are closely related and overlapping topics in mathematics and engineering. Estimation theory provides the mathematical framework for inferring the parameters or states of possibly stochastic and time-varying systems, given uncertain observations of their outputs. It encompasses statistical as well as probabilistic analysis in the formulation of algorithms that provide probabilistic estimates of quantities of interest. Depending on the application, a choice is often required in terms of the estimation paradigm to be used. Two main approaches are available, namely the frequentist and the Bayesian frameworks. Whereas the former presupposes the estimated variables to be deterministic, Bayesian methods assume these same variables to be random and governed by a probability density function (pdf). Least squares and maximum-likelihood methods, which fall into the category of frequentist techniques and are known for their optimality properties, are often preferred when system parameters need to be inferred offline from large sets of observations. On the other hand, Bayesian methods, which are sequential in nature, provide a convenient and efficient framework for online dynamic estimation problems. Bayesian estimation is characterized by a prediction and correction cycle, whereby the latest estimate of the probability density function (pdf) of the state is propagated in time and updated in light of new observations.

Tracking is usually formalized as a dynamic state estimation problem, and as such, can be treated as an application area of estimation theory. It is concerned with the inference of the kinematics of moving objects (or targets), based on various possible types of noisy observations. It is of high practical importance in both military and civilian applications. These include the tracking of military aircraft, missiles, surface vessels, submarines and land vehicles, as well as civilian air traffic control. Observations can be acquired, among others, by infrared sensors, radars, phased array antennas, sonars,

cameras or a combination of these. While all these sensors exhibit different advantages and weaknesses, they are most importantly characterized by the quantities they measure, which are typically range, 'range-rate' or bearings, sampled at a certain frequency.

Although frequentist methods have been used to some extend in target tracking, owing to its convenient sequential structure, it is the Bayesian framework which has attracted the most attention from researchers and practitioners. The now well-known Kalman filter, introduced in 1960 by Rudolf E. Kalman, represents an efficient and optimal Bayesian estimation algorithm for the class of linear systems with Gaussian noise. Its extensive use in the U.S. space program cemented its popularity and lead to its widespread utilization. More importantly, it provided the basis for more sophisticated methods which were required to address nonlinear estimation problems.

The Kalman filter provides, at every time step, the linear least squares estimate of the parameters of the true posterior pdf of the state. If, however, either the system dynamics or the measurement function are not linear-Gaussian, the resulting non-Gaussian posterior can no longer be adequately characterized by the linear Kalman filter equations. Suboptimal methods are therefore resorted to, which approximate the true posterior by various means.

The extended Kalman filter, which makes linear approximations of nonlinear system dynamics and measurement functions, is a suboptimal adaptation of the Kalman Filter to nonlinear estimation problems. However, it involves two sources of approximation: the linearization of system and/or measurement functions, and the matching of the true non-Gaussian posterior by a Gaussian pdf with mean and covariance evaluated using the linear Kalman filter equations.

Various other suboptimal nonlinear filters have been developed to help better address the shortcomings of the extended Kalman filter. One example is the unscented Kalman filter, which is based on the deterministic approximation of non-Gaussian densities rather than the linearization of nonlinear system and/or measurement functions. Although generally more accurate and stable than the extended Kalman filter, it also

makes use of the 'moment-matching' approximation, whereby the true posterior is modelled by a Gaussian. Particle filters, on the other hand, are a class of Monte Carlo algorithms which aim to overcome the effects of nonlinearities by a random sample based representation of the true non-Gaussian posterior. This leads to superior accuracy, in comparison to moment-matching methods, but at an exponentially higher computational cost.

In the context of target tracking, bearings-only tracking represents an important subproblem in which the target kinematics need to be inferred using only measurements of the bearings of the target relative to the sensor. The highly nonlinear measurement function characterizing bearings-only tracking makes it a challenging problem. This is compounded, in single sensor scenarios, by the lack of range information in the measurements. Nevertheless, under certain conditions of observability, unambiguous target kinematics can be reconstructed with bearings-only measurements from a single sensor. Bearings-only tracking is of high importance, for instance in applications requiring the use of passive sensors (such as infrared sensors, cameras, or also radars and phased array sensors operated in passive mode). Because they do not emit energy, such sensors allow the tracking of targets without revealing one's presence. They are also central to certain specialized tracking systems, such as those of heat-seeking missiles utilizing forward looking infrared sensors (FLIRs).

The recently introduced shifted Rayleigh filter (SRF) is a highly accurate moment-matching algorithm designed specifically to address the nonlinearity of the bearings-only measurement function. Given a Gaussian prior, the SRF evaluates the exact first and second order moments of the true posterior, and approximates this pdf by a Gaussian with matched moments. This is achieved through a reformulation of the measurement equation, which results in a measurement pdf practically equivalent to that produced by the standard model. As a consequence, the SRF introduces no approximation errors, besides moment-matching. Extensive simulations in 2D have established its accuracy, which is superior to that of the extended and unscented Kalman filters, and on par with that of particle filters, at a fraction of the computational cost.

An important focus of this thesis is bearings-only target tracking in 3D. It is a

topic which has, so far, attracted much less attention in the literature than the simpler bearings-only tracking problem in 2D. We address a variety of 3D tracking problems, where the nonlinearity of the measurement model is compounded by data association uncertainty and/or switching system dynamics, which represent important challenges to tracker accuracy and robustness. A number of algorithms are proposed, to deal with these nonlinearities and their performances compared to those of competing methods.

In this thesis, we consider three important bearings-only tracking problems. The first one is the bearings-only tracking of a nonmaneuvering target in 3D, with measurement origin uncertainty (in the form of clutter). We propose an extension to the SRF, labelled SRF3C, which allows for the presence of clutter. Central to this algorithm, is the derivation of the pdf of the predicted bearings measurement. Simulation results are provided, in a challenging single sensor tracking scenario (involving high clutter probability, poor target initialization, high rate of change of measured bearings and low observability), where the new algorithm produces accurate and reliable estimates, similar to those obtained by a sophisticated particle filter, while several commonly used moment-matching filters exhibit catastrophic failure.

Using the same tracking scenario, we then incorporate our algorithm within the well-known Rauch-Tung-Striebel smoothing framework to provide an efficient and highly accurate fixed-interval smoother, called the RTS-SRS3C. We also investigate the benefits of using our filtering algorithm to generate proposal distributions within a particle filter. The resulting shifted Rayleigh particle filter (SRPF), is shown to achieve improved accuracy, at a given computational cost, compared to similar particle filters utilizing, instead of the transitional prior, proposal densities that incorporate the latest observation.

We then propose the shifted Rayleigh mixture filter for 3D measurements with clutter (SRMF3C), for bearings-only tracking of multiple-model maneuvering targets, in 3D, in the presence of clutter. This algorithm, which uses the same structure as the 'shifted Rayleigh mixture filter', is adaptated to bearings-only tracking in 3D, through the use of the previously calculated pdf of the predicted bearings measurement. Simulation results of a multiple sensor tracking scenario reveal its superiority to the traditionally

used interacting-multiple-model filters. Results from a multiple model SRPF are also reported, which show accuracy levels similar to that of our algorithm, but at a much higher computational cost.

The final problem that we investigate in this thesis is that of bearings-only simultaneous localization and mapping (SLAM). This is an estimation problem whereby an agent constucts a map of an environment in which it is randomly navigating, using bearings-only measurements of features in the environment, while at the same time inferring its own location and heading within this map. Bearings-only SLAM is central, among others, to computer vision, autonomous indoor and outdoor navigation and land, airborne and submarine exploration. The uncertainty associated with the location and orientation of the sensor, relative to which all observations are made, is one of the principal challenges of SLAM. Bearings-only SLAM, moreover, involves the additional difficulty of estimating the (unobservable) range of landmarks upon initialization. Various algorithms that address the challenges of bearings-only SLAM have been proposed, based on the extended Kalman filter and on particle methods. It is documented, however, that although satisfactory in practical implementation, they can lead to inconsistent estimates.

In view of the high accuracy of the SRF and its robustness to poor target initialization in standard bearings-only target tracking scenarios, we propose two bearings-only SLAM algorithms based on the SRF: the quadrature-SRF SLAM and the RBSRPF SLAM algorithms. The first one is a moment-matching filter, in which, given a Gaussian approximation of the prior, the exact first and second order moments of the posterior are numerically evaluated. The second algorithm is a Rao-Blackwellization particle filter, where the estimation problem is broken down into two sub-problems: the estimation the vehicle heading (carried out with a particle filter) and the conditional estimation of the relative positions of the vehicle and the landmarks (obtained using the SRF equations). Both algorithms are tested in a simulated environment, and their performance is compared with that of the well-known extended Kalman filter using inverse depth parametrization. The quadrature-SRF SLAM algorithm is shown to produce not only accurate but, more importantly, consistent estimates of the vehicle position. The other algorithms, however, report optimistic vehicle estimates which do not pass consistency tests.

The thesis is organized as follows. This chapter provides the general context and motivation of the problems addressed in the sequel. Chapter 2 represents a background of some important estimation and tracking algorithms that are presented in the literature, many of which are used in the later chapters. The main theoretical work related to the development of the 3D bearings-only tracking algorithms is presented in chapter 3, while chapter 4 regroups the associated simulation results and related discussions. Chapter 5 contains a review of the bearings-only SLAM problem, along with the derivation of the two novel SLAM algorithms, their simulation results and discussions. Finally, concluding remarks and suggestions for future work are given in chapter 6.

Chapter 2

Estimation Theory and Algorithms

Estimation is the process of inferring the values of quantities of interest through inaccurate and indirect observations. These quantities can either be time-invariant or be the state of a dynamic system described by a stochastic equation. Dealing with non-time varying variables, one speaks of 'parameter estimation', whereas dynamic systems require 'state estimation' methods. According to the nature of the problem at hand, and also its formulation, two main estimation paradigms are available: the frequentist and Bayesian approaches. Although typically employed for parameter estimation and state estimation respectively, their areas of application contain significant overlap. The general premise is as follows. Let $\mathbf{x}_k \in \mathbb{R}^{n_x}$ denote a vector of time-varying parameters and $\mathbf{z}_k \in \mathbb{R}^{n_z}$ the observation (also called measurement), both at at time t_k ($k \in \mathbb{N}$ denotes the time index), related through the possibly nonlinear and time-varying measurement function h_k

$$\mathbf{z}_k \;=\; h(k, \mathbf{x}_k) + \mathbf{w}_k, \tag{2.0.1}$$

with $\mathbf{w}_k$ representing random disturbances of known statistics. Using the notation $\mathbf{Z}_k = \{\mathbf{z}_i\}_{i=1}^{k}$ to denote the collection of all observations up to time t_k, we define an estimator

$$\hat{\mathbf{x}}_k \;\triangleq\; \hat{\mathbf{x}}(k, \mathbf{Z}_k), \tag{2.0.2}$$

as a function of the observations that estimates the value of $\mathbf{x}_k$ in some sense. The following provides a brief overview of the two major approaches to solving the estimation problem.

2.1 The Frequentist approach

The frequentist (or simply non-Bayesian) approach is used extensively in parameter estimation problems. In contrast to Bayesian methods, it makes no assumptions about prior information on the variables to be estimated. Instead, it provides a framework to evaluate the parameters of the model under consideration purely so as to 'explain' or 'fit' the available observations in some, possibly optimal, sense. Although commonly employed in parameter estimation, this method can also be applied to dynamic state estimation [8]. Two of the most widely encountered frequentist estimators, namely the maximum-likelihood and the least squares estimators, are outlined.

2.1.1 The Maximum Likelihood estimator

The maximum likelihood estimator (MLE) is a point estimator that evaluates the parameters which maximize the likelihood function

$$\Lambda_{\mathbf{Z}_k}(\mathbf{x}_k) \triangleq p(\mathbf{Z}_k \mid \mathbf{x}_k) \tag{2.1.1}$$

defined as the pdf of the observations conditioned on the parameters. The MLE, in other words, yields the parameter values that are the most likely to have generated the observations collected up to the present time. Assuming that the likelihood function $\Lambda_{\mathbf{Z}_k}$ is computable, the MLE of $\mathbf{x}_k$ is given by

$$\hat{\mathbf{x}}_k^{\mathrm{ML}} \;=\; \arg\max_{\mathbf{x}_k} \Lambda_{\mathbf{Z}_k}(\mathbf{x}_k) \;=\; \arg\max_{\mathbf{x}_k} p\left(\mathbf{Z}_k \mid \mathbf{x}_k\right), \tag{2.1.2}$$

and, assuming further that $\Lambda_{\mathbf{Z}_k}$ is differentiable and has only one local maximum, the MLE is a solution of

$$\frac{d\Lambda_{\mathbf{Z}_k}(\mathbf{x}_k)}{d\mathbf{x}_k} \;=\; \frac{dp\left(\mathbf{Z}_k \mid \mathbf{x}_k\right)}{d\mathbf{x}_k} \;=\; 0. \tag{2.1.3}$$

Provided the observations are independent (i.e. the measurement noise sequence is independent), the likelihood function $\Lambda_{\mathbf{Z}_k}$ can be expressed in a more manageable multiplicative form as

$$\Lambda_{\mathbf{Z}_k}(\mathbf{x}_k) \;=\; \prod_{i=1}^{k} p\left(\mathbf{z}_i \mid \mathbf{x}_i\right). \tag{2.1.4}$$

Analytical solutions to (2.1.3) often are not possible and optimization tools are used to evaluate the MLE. The problem is then replaced by the equivalent, but computationally more convenient, task of minimizing the negative log-likelihood function. The MLE has a number of desirable asymptotic properties, namely: consistency, efficiency and normality [36]. Despite its formulation as a 'batch' estimator (where the arrival of a new observation requires the entire data to be reprocessed), the MLE has sequential variants based on a 'sliding-window' of observations. These implementations, while foregoing the asymptotic properties, allow real-time implementation.

2.1.2 The Least Squares estimator

The least squares estimator (LSE) is another widely used 'batch' method in parameter estimation. In it simplest form, it aims to find the parameters of a (possibly nonlinear) function that minimize the sum of squared errors or deviations from the actual observations. A straightforward refinement is to incorporate a weighting which indicates the level of confidence placed on the individual observations, reflected by their variance. The most commonly used weighting matrix is the inverse of the observation noise covariance matrix $\mathbf{R}_i$. The LSE is thus expressed as

$$\hat{\mathbf{x}}_k^{LS} = \arg\min_{\mathbf{x}_k} \left\{ \sum_{i=1}^{k} \left[\mathbf{z}_i - h(i, \mathbf{x}_k)\right]^{\mathrm{T}} \mathbf{R}_i^{-1} \left[\mathbf{z}_i - h(i, \mathbf{x}_k)\right] \right\}. \tag{2.1.5}$$

Assuming that the observation noise terms $\mathbf{w}_i$ are independent and identically distributed (iid) Gaussian random variables with zero mean and covariance $\mathbf{R}_i$, the LSE coincides with the MLE [8]. Although the least squares method allows for minimization of the errors over nonlinear functions, this task is often too challenging. Consequently, the linear least squares estimator (LLSE), minimizing the sum of squared errors over the more restrictive class of linear functions, is frequently resorted to. We note that the LSE has a convenient sequential equivalent known as the recursive least squares (RLS) estimator.

The computational complexity of non-Bayesian estimators, linked to the function minimization and batch processing requirements, often restricts their use to off-line computation on acquired data sets. Meanwhile, their sequential variants, although convenient, do not exhibit the attractive asymptotic properties of their 'batch' counterparts.

Bayesian methods, being intrinsically sequential, help to address this shortcoming.

2.2 The Bayesian Approach

The starting assumption of the Bayesian approach is the treatment of the parameters to be estimated as random variables. As a result, this requires the availability of 'a priori' knowledge about them, which is characterized by a 'prior' pdf. From there, one can define the notion of a 'posterior' pdf $p(\mathbf{x}_k \mid \mathbf{Z}_k)$, that incorporates the latest observation into the prior to reflect all the available information in $\mathbf{Z}_k$. Then, the posterior can be propagated in time to provide the prior (or prediction) ahead of the following observation, concluding the recursion. Thus, assuming that an initializing density $p(\mathbf{x}_0 \mid \mathbf{z}_0)$ is available, the process can be perpetuated as a cycle of prediction and update stages. This framework defines the concept of the recursive Bayesian filter, whereby a prior belief is refined in light of new information, without the need for storing and reprocessing past data.

Suppose we have an expression for the posterior pdf $p(\mathbf{x}_{k-1} \mid \mathbf{Z}_{k-1})$ at time $k-1$. Given a (possibly time-variant) state transition function f_{k-1} mapping the previous state $\mathbf{x}_{k-1}$, a deterministic control input $\mathbf{u}^s_{k-1}$ and a process noise term $\mathbf{v}_{k-1}$ into the state $\mathbf{x}_k$ at time k

$$\mathbf{x}_k \;=\; f_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}^s_{k-1}, \mathbf{v}_{k-1}), \tag{2.2.1}$$

and a (possibly time-variant) measurement function h_k relating the current state $\mathbf{x}_k$ and a measurement noise term $\mathbf{w}_k$ into a measurement $\mathbf{z}_k$ at time k

$$\mathbf{z}_k \;=\; h_k(\mathbf{x}_k, \mathbf{w}_k), \tag{2.2.2}$$

the prediction density of the state at time k is obtained, using the Chapman-Kolmogorov equation and noticing that (2.2.1) represents a first-order Markov process, as

$$p(\mathbf{x}_k \mid \mathbf{Z}_{k-1}) \;=\; \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1} \mid \mathbf{Z}_{k-1})\, d\mathbf{x}_{k-1}, \tag{2.2.3}$$

where the density $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$ is evaluated using the system equation (2.2.1) and the known statistics of $\mathbf{v}_{k-1}$.

The update stage at time k is then carried out upon arrival of the observation $\mathbf{z}_k$ using Bayes' rule:

$$\begin{aligned} p(\mathbf{x}_k \mid \mathbf{Z}_k) &= p(\mathbf{x}_k \mid \mathbf{z}_k, \mathbf{Z}_{k-1}) \\ &= \frac{p(\mathbf{z}_k \mid \mathbf{x}_k, \mathbf{Z}_{k-1})\, p(\mathbf{x}_k \mid \mathbf{Z}_{k-1})}{p(\mathbf{z}_k \mid \mathbf{Z}_{k-1})} \\ &= \frac{p(\mathbf{z}_k \mid \mathbf{x}_k)\, p(\mathbf{x}_k \mid \mathbf{Z}_{k-1})}{p(\mathbf{z}_k \mid \mathbf{Z}_{k-1})} \end{aligned} \tag{2.2.4}$$

where the normalizing constant can be expressed as:

$$p(\mathbf{z}_k \mid \mathbf{Z}_{k-1}) \;=\; \int p(\mathbf{z}_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{Z}_{k-1})\, d\mathbf{x}_k. \tag{2.2.5}$$

The estimated pdf of $\hat{\mathbf{x}}_k$, evaluated sequentially through (2.2.3) and (2.2.4) reflects all the information contained in the measurement set $\mathbf{Z}_k$. From there, one can compute, at each time step, an optimal estimate of $\mathbf{x}_k$ with respect to any criterion, along with its degree of accuracy. Two of the most common choices are:

- The minimum mean square error (MMSE) estimate defined as

$$\hat{\mathbf{x}}_k^{\mathrm{MMSE}} \;=\; \arg\min_{\hat{\mathbf{x}}_k} E\left[(\hat{\mathbf{x}}_k - \mathbf{x}_k)^2 \mid \mathbf{Z}_k\right], \tag{2.2.6}$$

  of which the solution can be obtained by setting the gradient of the mean of the squared norm of the error to zero:

$$\nabla_{\hat{\mathbf{x}}_k} E\left[(\hat{\mathbf{x}}_k - \mathbf{x}_k)^{\mathrm{T}} (\hat{\mathbf{x}}_k - \mathbf{x}_k) \mid \mathbf{Z}_k\right] \;=\; 2\left(\hat{\mathbf{x}}_k - E\left[\mathbf{x}_k \mid \mathbf{z}_k\right]\right) \;=\; 0, \tag{2.2.7}$$

  leading to:

$$\hat{\mathbf{x}}_k^{\mathrm{MMSE}} \;=\; E\left[\mathbf{x}_k \mid \mathbf{Z}_k\right] \;\triangleq\; \int_{-\infty}^{\infty} \mathbf{x}_k\, p(\mathbf{x}_k \mid \mathbf{Z}_k)\, d\mathbf{x}_k, \tag{2.2.8}$$

  which is the conditional mean of $\mathbf{x}_k$.

- The maximum a posteriori (MAP) estimate, which is the argument $\mathbf{x}_k$ that maximizes the posterior $p(\mathbf{x}_k \mid \mathbf{Z}_k)$:

$$\hat{\mathbf{x}}_k^{\mathrm{MAP}} \;\triangleq\; \arg\max_{\hat{\mathbf{x}}_k} p(\mathbf{x}_k \mid \mathbf{Z}_k). \tag{2.2.9}$$

Note that for the case of a Gaussian posterior, the mean of the distribution coincides with its mode. Thus the MMSE and MAP estimates are identical.

The recursive Bayesian solution defined by (2.2.3) and (2.2.4) provides a conceptual framework for the filtering problem. Although optimal estimates can be obtained from this scheme, it is only of practical value for the class of linear system and observation functions with Gaussian noise. Generally, outside of these restrictive conditions, propagating the non-Gaussian posterior becomes an intractable nonlinear filtering problem. Whereas in the linear-Gaussian case the state mean and covariance estimates fully specify the posterior, non-Gaussian pdfs, introduced by nonlinear transformations of densities, require an infinite number of parameters that need to be computed and stored. Since most practical applications involve nonlinear dynamics and/or observation functions, suboptimal or approximate solutions need to be resorted to. However, despite its limited applicability, the linear-Gaussian version of the Bayesian filter, known as the Kalman filter, provides the basis for more sophisticated methods.

2.2.1 The Kalman Filter

The Kalman filter ([51]) is the algorithmic formalization of the conceptual linear-Gaussian Bayesian filter. It makes the assumption of linear system and observation equations f_{k-1} and h_k and Gaussian process and measurement noises $\mathbf{v}_{k-1}$ and $\mathbf{w}_k$. Relations (2.2.1) and (2.2.2) can then be expressed as:

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}^s_{k-1} + \mathbf{v}_{k-1} \tag{2.2.10}$$

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{w}_k \tag{2.2.11}$$

where $\mathbf{F}_{k-1}$ is an $n_x \times n_x$ state transition matrix, $\mathbf{H}_k$ is an $n_z \times n_x$ observation matrix and $\mathbf{G}_{k-1}$ is an $n_x \times n_u$ control-input matrix. The process and measurement noise terms $\mathbf{v}_{k-1}$ and $\mathbf{w}_k$ are mutually independent zero-mean white Gaussian with known covariance matrices $\mathbf{Q}_{k-1}$ and $\mathbf{R}_k$. The initial state $\mathbf{x}_0$ is a Gaussian random variable, independent of the two noise sequences, and with specified mean and covariance matrix. The Kalman filter recursion is then defined by the following equations:

Prediction:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1} + \mathbf{G}_{k-1}\mathbf{u}^s_{k-1} \tag{2.2.12}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}^{\mathrm{T}}_{k-1} + \mathbf{Q}_{k-1} \tag{2.2.13}$$

$$\mathbf{S}_k = \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}^{\mathrm{T}}_k + \mathbf{R}_k \tag{2.2.14}$$

Update:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}^{\mathrm{T}}_k\mathbf{S}^{-1}_k \tag{2.2.15}$$

$$\nu_k = \mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} \tag{2.2.16}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\nu_k \tag{2.2.17}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{S}_k\mathbf{K}^{\mathrm{T}}_k \tag{2.2.18}$$

where ν_k is known as the innovation (or observation residual) and $\mathbf{S}_k$ denotes the innovation covariance matrix. The matrix $\mathbf{K}_k$ is the Kalman gain, which specifies the weight given to the latest observation. A "large" value of the Kalman gain will be indicative of an inaccurate state prediction and an accurate observation, while a "small" gain will reflect the contrary. Note that $\mathbf{P}_{k|k-1}$, $\mathbf{S}_k$, $\mathbf{K}_k$ and $\mathbf{P}_{k|k}$ are independent of the observation $\mathbf{z}_k$ and thus can be computed offline.

Derivation of The Kalman Filter Equations

Before proving (2.2.17) and (2.2.18), we introduce two theorems which provide the fundamentals of linear least squares estimation.

Theorem 2.2.1 (**Solution to the Linear Last Squares Estimation Problem).** *Let* $\mathbf{x}$ *and* $\mathbf{z}$ *be joint random vectors of dimensions* n *and* m*, respectively, with means* $\mathbf{m_x}$ *and* $\mathbf{m_z}$ *and covariance*

$$E\left[\begin{pmatrix} \mathbf{x}-\mathbf{m_x} \\ \mathbf{z}-\mathbf{m_z} \end{pmatrix}\begin{pmatrix} \mathbf{x}^{\mathrm{T}}-\mathbf{m_x}^{\mathrm{T}} & \mathbf{z}^{\mathrm{T}}-\mathbf{m_z}^{\mathrm{T}} \end{pmatrix}\right] = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}, \tag{2.2.19}$$

where $\mathbf{P}_{22} > 0$.

The LLSE of $\mathbf{x}$ *given* $\mathbf{z}$ *is*

$$\hat{\mathbf{x}} \;=\; \mathbf{m_x} + \mathbf{P}_{12}\mathbf{P}_{22}^{-1}(\mathbf{z} - \mathbf{m_z}) \tag{2.2.20}$$

and the covariance of the error is

$$cov\{\mathbf{x} - \hat{\mathbf{x}}\} \;=\; \mathbf{P}_{11} - \mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21} \tag{2.2.21}$$

Proof. $\mathbf{z}$ and $\hat{\mathbf{x}}$ being linearly related, we express $\hat{\mathbf{x}}$ as $\hat{\mathbf{x}} \;=\; \mathbf{A}^*\mathbf{z} + \mathbf{a}^*$. The aim is then to find $\mathbf{A}^*$ and $\mathbf{a}^*$ that minimize

$$\phi(\mathbf{A}, \mathbf{a}) \;=\; E\left[||\mathbf{x} - \mathbf{A}\mathbf{z} - \mathbf{a}||^2\right]. \tag{2.2.22}$$

By fixing $\mathbf{A}$ and setting the gradient of the error to 0:

$$\nabla_{\mathbf{a}}\phi\left(\mathbf{A}, \mathbf{a}\right) \;=\; E\left[-2\left(\mathbf{x} - \mathbf{A}\mathbf{z} - \mathbf{a}\right)\right] \;=\; 0, \tag{2.2.23}$$

we obtain

$$\mathbf{a}^* \;=\; \mathbf{m_x} - \mathbf{A}^*\mathbf{m_z}. \tag{2.2.24}$$

$\mathbf{A}^*$ must now minimize $\psi(\mathbf{A}) = \phi(\mathbf{A}, \mathbf{a}(\mathbf{A}))$, which can be written as:

$$\begin{aligned}\psi(\mathbf{A}) \;&= E\left[||\mathbf{x} - \mathbf{A}\mathbf{z} - \mathbf{m_x} + \mathbf{A}\mathbf{m_z}||^2\right] \\ &= E\left[\left|\left|\mathbf{x}' - \mathbf{A}\mathbf{z}'\right|\right|^2\right],\end{aligned} \tag{2.2.25}$$

where $\mathbf{x}'$ and $\mathbf{z}'$ are the centered random variables $\mathbf{x} - \mathbf{m_x}$ and $\mathbf{z} - \mathbf{m_z}$. Adding and subtracting $\mathbf{A}^*$ we obtain:

$$\begin{aligned}\psi(\mathbf{A}) \;&= \psi\left(\mathbf{A} - \mathbf{A}^* + \mathbf{A}^*\right) \\ &= E\left[\left|\left|\mathbf{x}' - \mathbf{A}^*\mathbf{z}' - (\mathbf{A} - \mathbf{A}^*)\mathbf{z}'\right|\right|^2\right] \\ &= \text{trace } E\left[\left(\mathbf{x}' - \mathbf{A}^*\mathbf{z}' - (\mathbf{A} - \mathbf{A}^*)\mathbf{z}'\right)\left(\mathbf{x}' - \mathbf{A}^*\mathbf{z}' - (\mathbf{A} - \mathbf{A}^*)\mathbf{z}'\right)^{\mathrm{T}}\right] \\ &= \text{trace } E\left[\left(\mathbf{x}' - \mathbf{A}^*\mathbf{z}'\right)\left(\mathbf{x}' - \mathbf{A}^*\mathbf{z}'\right)^{\mathrm{T}}\right] - \text{trace } E\left[\left(\mathbf{x}' - \mathbf{A}^*\mathbf{z}'\right)\left((\mathbf{A} - \mathbf{A}^*)\mathbf{z}'\right)^{\mathrm{T}}\right] \\ &\quad - \text{trace } E\left[\left((\mathbf{A} - \mathbf{A}^*)\mathbf{z}'\right)\left(\mathbf{x}' - \mathbf{A}^*\mathbf{z}'\right)^{\mathrm{T}}\right] + \text{trace } E\left[\left((\mathbf{A} - \mathbf{A}^*)\mathbf{z}'\right)\left((\mathbf{A} - \mathbf{A}^*)\mathbf{z}'\right)^{\mathrm{T}}\right]\end{aligned}$$

(2.2.26)

Observing that

$$E\left[\left(\mathbf{x}' - \mathbf{A}^*\mathbf{z}'\right)\mathbf{z}'^{\mathrm{T}}\right] \;=\; E\left[\mathbf{x}'\mathbf{z}'^{\mathrm{T}}\right] - \mathbf{A}^* E\left[\mathbf{z}'\mathbf{z}'^{\mathrm{T}}\right] \;=\; \mathbf{P}_{12} - \mathbf{A}^*\mathbf{P}_{22}$$

can be set to 0 by choosing $\mathbf{A}^* = \mathbf{P}_{12}\mathbf{P}_{22}^{-1}$, we then get (for that choice of $\mathbf{A}^*$):

$$\psi(\mathbf{A}) \;=\; E\left[\left|\left|\mathbf{x}' - \mathbf{A}^*\mathbf{z}'\right|\right|^2\right] + E\left[\left|\left|(\mathbf{A} - \mathbf{A}^*)\mathbf{z}'\right|\right|^2\right] \tag{2.2.27}$$

The first term does not depend on $\mathbf{A}$, while the second term, which is non-negative can be set to zero by the very choice $\mathbf{A} = \mathbf{A}^*$. Thus the minimizing value of $\mathbf{A}$ is $\mathbf{P}_{12}\mathbf{P}_{22}^{-1}$ and we have:

$$\hat{\mathbf{x}} \;=\; \mathbf{m}_{\mathbf{x}} + \mathbf{P}_{12}\mathbf{P}_{22}^{-1}(\mathbf{z} - \mathbf{m}_{\mathbf{z}}) \tag{2.2.28}$$

The covariance of the LLSE error is given by:

$$\begin{aligned}\mathrm{cov}\{\mathbf{x}\} \;&= E\left[\left((\mathbf{x} - \mathbf{m}_{\mathbf{x}}) - \mathbf{P}_{12}\mathbf{P}_{22}^{-1}(\mathbf{z} - \mathbf{m}_{\mathbf{z}})\right)\left((\mathbf{x} - \mathbf{m}_{\mathbf{x}}) - \mathbf{P}_{12}\mathbf{P}_{22}^{-1}(\mathbf{z} - \mathbf{m}_{\mathbf{z}})\right)^{\mathrm{T}}\right] \\ &= \mathbf{P}_{11} - 2\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21} + \mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21} \\ &= \mathbf{P}_{11} - \mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21}. \end{aligned} \tag{2.2.29}$$

■

Theorem 2.2.2 (**LSE for Jointly Normal Random Vectors (Jazwinski)).**
Let $\mathbf{x}$ *and* $\mathbf{z}$ *be jointly normally distributed, as in (2.2.19). Then the conditional density of* $\mathbf{x}$ *given* $\mathbf{z}$ *is normal with mean*

$$\mathbf{m}_{\mathbf{x}|\mathbf{z}} \;=\; \mathbf{m}_{\mathbf{x}} + \mathbf{P}_{12}\mathbf{P}_{22}^{-1}\left(\mathbf{z} - \mathbf{m}_{\mathbf{z}}\right) \tag{2.2.30}$$

and covariance matrix

$$cov\{\mathbf{x} \mid \mathbf{z}\} \;=\; \mathbf{P}_{11} - \mathbf{P}_{21}\mathbf{P}_{22}^{-1}\mathbf{P}_{21}. \tag{2.2.31}$$

Proof. See [45] ■

Using the recursive Bayesian framework described by (2.2.3) and (2.2.4), the conditional

density of $\mathbf{x}_k$ given $\mathbf{z}_k$ obtained by the Kalman filter is arrived at through the recursion:

$$p(\mathbf{x}_{k-1} \mid \mathbf{Z}_{k-1}) = \mathcal{N}(\mathbf{x}_{k-1};\, \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}) \tag{2.2.32}$$

$$p(\mathbf{x}_k \mid \mathbf{Z}_{k-1}) = \mathcal{N}(\mathbf{x}_k;\, \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \tag{2.2.33}$$

$$p(\mathbf{x}_k \mid \mathbf{Z}_k) = \mathcal{N}(\mathbf{x}_k;\, \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}), \tag{2.2.34}$$

where the parameters of the normal densities have yet to be determined.

Fixing $k \geq 1$, we obtain the predicted state and observation by taking conditional expectations with respect to $\mathbf{Z}_{k-1}$ across equations (2.2.10) and (2.2.11):

$$\hat{\mathbf{x}}_{k|k-1} = E\left[\mathbf{x}_k \mid \mathbf{Z}_{k-1}\right] = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1} + \mathbf{G}_{k-1}\mathbf{u}^s_{k-1} \tag{2.2.35}$$

$$\hat{\mathbf{z}}_{k|k-1} = E\left[\mathbf{z}_k \mid \mathbf{Z}_{k-1}\right] = \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} \tag{2.2.36}$$

The state prediction error and observation prediction error (or innovation) can then be expressed as:

$$\tilde{\mathbf{x}}_{k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k-1}\tilde{\mathbf{x}}_{k-1|k-1} + \mathbf{v}_{k-1} \tag{2.2.37}$$

$$\tilde{\mathbf{z}}_{k|k-1} = \mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1} = \mathbf{H}_k\tilde{\mathbf{x}}_{k|k-1} + \mathbf{w}_k = \nu_k \tag{2.2.38}$$

The covariances of the state and observation predictions are in turn given by:

$$\begin{aligned}
\mathbf{P}_{k|k-1} &= E\left[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{x}}^{\mathrm{T}}_{k|k-1} \mid \mathbf{Z}_{k-1}\right] \\
&= \mathbf{F}_{k-1}E\left[\tilde{\mathbf{x}}_{k-1|k-1}\tilde{\mathbf{x}}^{\mathrm{T}}_{k-1|k-1} \mid \mathbf{Z}_{k-1}\right]\mathbf{F}^{\mathrm{T}}_{k-1} + E\left[\mathbf{v}_{k-1}\mathbf{v}^{\mathrm{T}}_{k-1}\right] \\
&= \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}^{\mathrm{T}}_{k-1} + \mathbf{Q}^{\mathrm{T}}_{k-1}
\end{aligned} \tag{2.2.39}$$

$$\begin{aligned}
\mathbf{S}_k &= E\left[\tilde{\mathbf{z}}_{k|k-1}\tilde{\mathbf{z}}^{\mathrm{T}}_{k|k-1}|\mathbf{Z}_{k-1}\right] \\
&= \mathbf{H}_kE\left[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{x}}^{\mathrm{T}}_{k|k-1}|\mathbf{Z}_{k-1}\right]\mathbf{H}^{\mathrm{T}}_k + E\left[\mathbf{w}_k\mathbf{w}^{\mathrm{T}}_k\right] \\
&= \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}^{\mathrm{T}}_k + \mathbf{R}_k
\end{aligned} \tag{2.2.40}$$

and the joint covariance of the state and observation is:

$$E\left[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{z}}^{\mathrm{T}}_{k|k-1} \mid \mathbf{Z}_{k-1}\right] = E\left[\tilde{\mathbf{x}}_{k|k-1}\left(\mathbf{H}_k\tilde{\mathbf{x}}_{k|k-1} + \mathbf{w}_k\right)^{\mathrm{T}} \mid \mathbf{Z}_{k-1}\right]$$

$$= \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} \tag{2.2.41}$$

Treating $\mathbf{x}_k$ and $\mathbf{z}_k$ as variables with joint normal density $p(\mathbf{x}_k, \mathbf{y}_k \mid \mathbf{Z}_{k-1})$, we can apply Thm. 2.2.2 to obtain the conditional mean and covariance of $\mathbf{x}_k$ given $\mathbf{z}_k$:

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k} &= E\left[\mathbf{x}_k \mid \mathbf{Z}_k\right] \\
&= E\left[\mathbf{x}_k \mid \mathbf{Z}_{k-1}\right] + \mathrm{cov}\{\mathbf{x}_k, \mathbf{z}_k \mid \mathbf{Z}_{k-1}\}\left(\mathrm{cov}\{\mathbf{z}_k \mid \mathbf{Z}_{k-1}\}\right)^{-1}\left(\mathbf{z}_k - E\left[\mathbf{z}_k \mid \mathbf{Z}_{k-1}\right]\right) \\
&= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\nu_k, \qquad (2.2.42)
\end{aligned}$$

where

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\mathbf{S}_k^{-1} \tag{2.2.43}$$

and

$$\begin{aligned}
\mathbf{P}_{k|k} &= \mathrm{cov}\{\mathbf{x}_k \mid \mathbf{Z}_{k-1}\} - \mathrm{cov}\{\mathbf{x}_k, \mathbf{z}_k \mid \mathbf{Z}_{k-1}\}\mathrm{cov}\{\mathbf{z}_k|\mathbf{Z}_{k-1}\}^{-1}\mathrm{cov}\{\mathbf{z}_k, \mathbf{x}_k \mid \mathbf{Z}_{k-1}\} \\
&= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1} \\
&= \left[\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right]\mathbf{P}_{k|k-1} \qquad (2.2.44) \\
&= \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^{\mathrm{T}} \qquad (2.2.45)
\end{aligned}$$

Under linear/Gaussian assumptions, the Kalman filter is the MMSE state estimator. If these conditions do not hold, however, the actual posterior densities become non-Gaussian and the Kalman filter only reports their mean and covariance. In this case, the Kalman filter is the linear MMSE estimator.

2.3 Nonlinear Filtering and Approximate Bayesian Estimators

As mentioned before, the linear/Gaussian assumption is rarely justified in practical applications. Most problems involve nonlinear dynamic and/or measurement equations, making the conceptual optimal nonlinear filter inapplicable. For this reason, many suboptimal approaches have been developed to address the nonlinear filtering problem, with varying levels of accuracy, robustness and computational requirements. These nonlinear filters can be classified within two categories according to their general approximating

approach, namely density-approximating and moment-matching. Filters in the first category aim directly at approximating the non-Gaussian pdf of interest by various techniques, including the use of Gaussian mixtures and sequential Monte Carlo methods. Examples include multiple model filters such as the generalized pseudo-Bayesian (GPB) filter, the Interacting Multiple Model (IMM) filter and the range-parameterized extended Kalman filter (RPEKF) and also sequential Monte Carlo methods like the particle filter (PF). Moment matching filters, meanwhile, aim to characterize the posterior pdf by a normal distribution with matched first two moments, which are approximated by various means. For example, the extended Kalman filter (EKF) relies on the linearization of the dynamic and/or measurement equation, while the unscented Kalman filter (UKF) is based on an approximation of the non-Gaussian pdf by a deterministic sampling method.

Filters of the density-approximating category offer the potential for greater accuracy but with higher computational demands, justifying in many cases the use of moment matching alternatives. In the following, we provide an overview of the EKF and the UKF as examples of generic moment-matching filters, and of the sampling-importance-resampling (SIR) particle filter as an example of a Monte Carlo density-approximating method. Various other types of moment-matching and density-approximating filters will later be introduced within the context of target tracking.

2.3.1 Moment Matching Filters

The Extended Kalman Filter

The extended Kalman filter (EKF) [45] is a natural extension of the Kalman filter to the problem of nonlinear filtering. Nonlinear dynamic and/or measurement equations are linearized to provide an approximate representation of the system, at a given operating point, that fits into the standard Kalman filtering framework. Consider the following system and measurement equations with additive noise, which are a special case of (2.2.1) and (2.2.2):

$$\mathbf{x}_k = f_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}^s) + \mathbf{v}_{k-1} \tag{2.3.1}$$

$$\mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{u}_k^m) + \mathbf{w}_k. \tag{2.3.2}$$

The nonlinear functions f_{k-1} and h_k are expressed using a first order Taylor expansion around the latest state estimate $\hat{\mathbf{x}}_{k-1|k-1}$ and state prediction $\hat{\mathbf{x}}_{k|k-1}$, respectively:

$$\mathbf{x}_k = f_{k-1}\left(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_{k-1}^s\right) + \hat{\mathbf{F}}_{k-1}\left[\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}\right] + \text{H.O.T} + \mathbf{v}_{k-1} \tag{2.3.3}$$

$$\mathbf{z}_k = h_k\left(\hat{\mathbf{x}}_{k|k-1}, \mathbf{u}_k^m\right) + \hat{\mathbf{H}}_k\left[\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}\right] + \text{H.O.T} + \mathbf{w}_k, \tag{2.3.4}$$

where

$$\hat{\mathbf{F}}_{k-1} \triangleq \left[\nabla_{\mathbf{x}_{k-1}} f_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}^s)\right]\Big|_{\mathbf{x}_{k-1}=\hat{\mathbf{x}}_{k-1|k-1}} \tag{2.3.5}$$

$$\hat{\mathbf{H}}_k \triangleq \left[\nabla_{\mathbf{x}_k} h_k^{\mathrm{T}}(\mathbf{x}_k, \mathbf{u}_k^m)\right]\Big|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}} \tag{2.3.6}$$

are respectively the Jacobians of the vector-valued functions f_{k-1} and h_k evaluated at the latest state estimate $\hat{\mathbf{x}}_{k-1|k-1}$ and state prediction $\hat{\mathbf{x}}_{k|k-1}$.

Conditioning on the measurements $\mathbf{Z}_{k-1}$ and taking expectations, we obtain (after dropping the higher order terms) the predicted state and measurement:

$$\hat{\mathbf{x}}_{k|k-1} = f_{k-1}\left(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_{k-1}^s\right) \tag{2.3.7}$$

$$\hat{\mathbf{z}}_{k|k-1} = h_k\left(\hat{\mathbf{x}}_{k|k-1}, \mathbf{u}_k^m\right). \tag{2.3.8}$$

The state prediction error $\tilde{\mathbf{x}}_{k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}$ and measurement prediction error $\tilde{\mathbf{z}}_{k|k-1} = \mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}$ are then given by:

$$\begin{aligned}
\tilde{\mathbf{x}}_{k|k-1} &= \hat{\mathbf{F}}_{k-1}\left[\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}\right] + \mathbf{v}_{k-1} = \hat{\mathbf{F}}_{k-1}\tilde{\mathbf{x}}_{k-1|k-1} + \mathbf{v}_{k-1} \\
\tilde{\mathbf{z}}_{k|k-1} &= \hat{\mathbf{H}}_k\left[\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}\right] + \mathbf{w}_k = \hat{\mathbf{H}}_k\tilde{\mathbf{z}}_{k|k-1} + \mathbf{w}_k,
\end{aligned}$$

from which we obtain the predicted state and innovation covariance matrices

$$\begin{aligned}
\mathbf{P}_{k|k-1} &= E[\tilde{\mathbf{x}}_{k|k-1}\tilde{\mathbf{x}}_{k|k-1}^{\mathrm{T}} \mid \mathbf{Z}_{k-1}] \\
&= E[(\hat{\mathbf{F}}_{k-1}\tilde{\mathbf{x}}_{k-1|k-1} + \mathbf{v}_{k-1})(\hat{\mathbf{F}}_{k-1}\tilde{\mathbf{x}}_{k-1|k-1} + \mathbf{v}_{k-1})^{\mathrm{T}} \mid \mathbf{Z}_{k-1}] \\
&= \hat{\mathbf{F}}_{k-1}\mathbf{P}_{k-1|k-1}\hat{\mathbf{F}}_{k-1}^{\mathrm{T}} + \mathbf{Q}_{k-1}
\end{aligned} \tag{2.3.9}$$

and

$$\begin{aligned}\mathbf{S}_k &= E[\tilde{\mathbf{z}}_{k|k-1}\tilde{\mathbf{z}}_{k|k-1}^{\mathrm{T}}|\mathbf{Z}_{k-1}] \\ &= E[(\hat{\mathbf{H}}_k\tilde{\mathbf{z}}_{k|k-1}+\mathbf{w}_k)(\hat{\mathbf{H}}_k\tilde{\mathbf{z}}_{k|k-1}+\mathbf{w}_k)^{\mathrm{T}}|\mathbf{Z}_{k-1}] \\ &= \hat{\mathbf{H}}_k\mathbf{P}_{k|k-1}\hat{\mathbf{H}}_k^{\mathrm{T}}+\mathbf{R}_k, \end{aligned} \tag{2.3.10}$$

where we have used the fact that the state and measurement prediction errors are uncorrelated with the system and measurement noises, respectively.

The standard Kalman filter equations can now be applied using the expressions for the state and measurement predictions and by replacing the matrices $\mathbf{F}_{k-1}$ and $\mathbf{H}_k$ by the jacobians $\hat{\mathbf{F}}_{k-1}$ and $\hat{\mathbf{H}}_k$. Equations (2.2.12)-(2.2.18) then become:

Prediction:

$$\hat{\mathbf{x}}_{k|k-1} = f_{k-1}\left(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_{k-1}^s\right) \tag{2.3.11}$$

$$\mathbf{P}_{k|k-1} = \hat{\mathbf{F}}_{k-1}\mathbf{P}_{k-1|k-1}\hat{\mathbf{F}}_{k-1}^{\mathrm{T}}+\mathbf{Q}_{k-1} \tag{2.3.12}$$

$$\mathbf{S}_k = \hat{\mathbf{H}}_k\mathbf{P}_{k|k-1}\hat{\mathbf{H}}_k^{\mathrm{T}}+\mathbf{R}_k \tag{2.3.13}$$

Update:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\hat{\mathbf{H}}_k^{\mathrm{T}}\mathbf{S}_k^{-1} \tag{2.3.14}$$

$$\nu_k = \mathbf{z}_k - h_k\left(\hat{\mathbf{x}}_{k|k-1}, \mathbf{u}_k^m\right) \tag{2.3.15}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1}+\mathbf{K}_k\nu_k \tag{2.3.16}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1}-\mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^{\mathrm{T}} \tag{2.3.17}$$

We note that unlike the Kalman filter, the EKF requires online computation of the state covariance matrix $\mathbf{P}_{k|k}$. This results from the fact that both Jacobians $\hat{\mathbf{F}}_{k-1}$ and $\hat{\mathbf{H}}_k$ now depend on the state estimate.

Due to its remarkable simplicity and versatility, the EKF has seen extensive use in

many areas of nonlinear filtering. Nevertheless, the basic assumptions made in its derivation need to be well understood and validated for each particular application.

As stated earlier, the EKF requires the linear representations of the dynamic and/or measurement equations to be reasonably close to their nonlinear behaviour in the vicinity of the point of linearization. Severe nonlinearities either in the state dynamics or measurement equations will fail to be captured adequately by local linearization, leading to unacceptable approximation errors and a poor representation of the non-Gaussian posterior. This is illustrated in Figure 2.1 where the pdfs of two normally distributed random variables are transformed using the nonlinear function $f(\cdot)$ and also by its linear approximation $\mathbf{F}(\cdot)$, about the point $E[\mathbf{x}]$. The true pdf (shown in green in the left hand side box) of the transformed low variance density $p'(x)$ is very close to a Gaussian pdf (shown in dashed red lines in the same box) obtained through the linear approximation of the mapping. However, the transformation of the high variance density $p(x)$ results in a skewed pdf (shown in blue in the right hand side box), of which a Gaussian approximation (shown in dashed red lines) is evidently inadequate. In general, linear approximations of nonlinear transformations are satisfactory only as long as the variance of the pdf is 'low' around regions of high nonlinearity of the dynamic or measurement equations. Moreover, it must also be remembered that the EKF forces a Gaussian approximation on the *posterior* $p(\mathbf{x}_k|\mathbf{Z}_k)$. If the true posterior exhibits a significant departure from Gaussianity (such as multiple modes), a Gaussian density will be an inadequate representation of the state estimate.

Another problem specific to linearization schemes is the fact that the Jacobians should really be computed at the true previous and current states. As this is impossible, the estimated and predicted states are used instead, resulting in further errors. Moreover, it must be noted that the bias introduced by nonlinear transformations, together with the mismatch between the real covariances and those computed through linearization can also undermine the effectiveness of the filter [47, 63]. More precisely, the EKF will have a tendency to misrepresent covariances, typically underestimating them, leading to 'overconfident' estimates whose computed variance is below their mean-square error. This

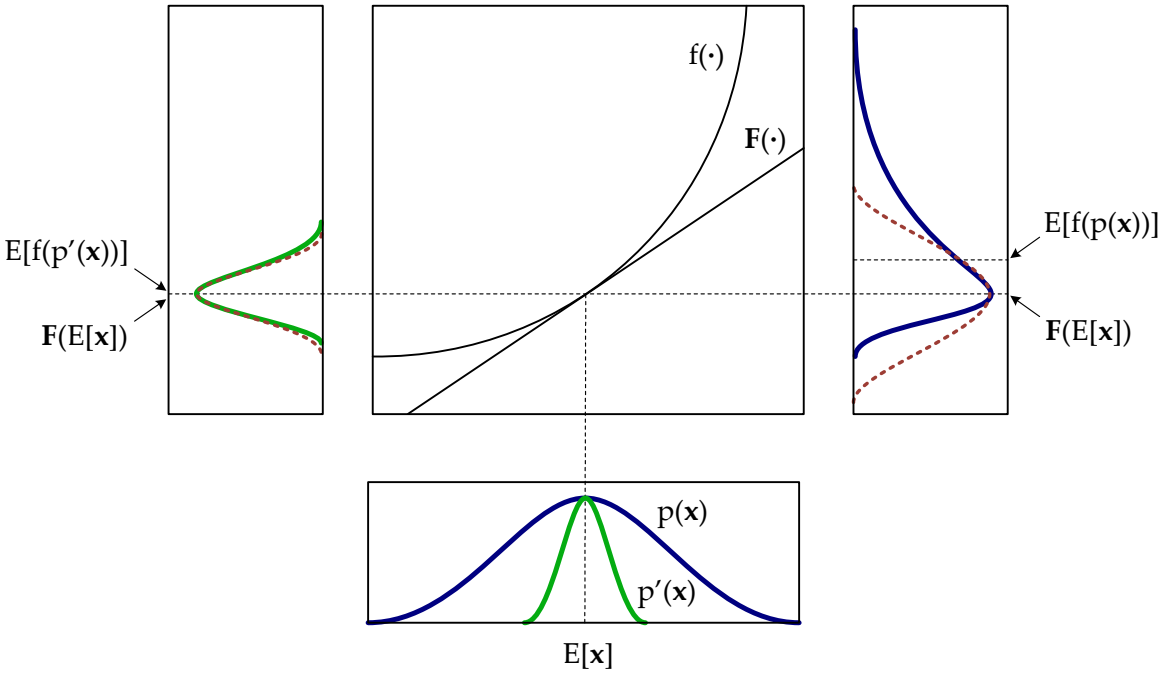


Figure 2.1: Two RVs with Gaussian distributions $p(\mathbf{x})$ and $p'(\mathbf{x})$ are transformed through the nonlinear function $f(\cdot)$ and its linear approximation $\mathbf{F}(\cdot)$ about the mean $E[\mathbf{x}]$. The left hand plot shows the adequacy of the Gaussian approximation while the right hand plot shows the high discrepancy between the actual transformed pdf and its Gaussian approximation.

combination of bias and misestimated covariance, known as *inconsistency*, can eventually lead to filter divergence.

Several refinements have been proposed to address the shortcomings of the EKF. The most natural one is the use of higher order terms in the Taylor series expansion of the dynamic and measurement functions [4, 8]. The 'second-order' EKF, which employs Hessians alongside Jacobians to propagate the state and covariance estimates, produces a quadratic rather than linear approximation of the nonlinearities. Although this strategy allows the filter to better cope with nonlinearities, it comes at a higher computational cost and more importantly still does not address the problems of Gaussian approximation and of the evaluation of the Taylor series about estimated rather true states.

The iterated EKF (iEKF) [24, 45], meanwhile, is a method that iteratively aims to improve the approximation of the measurement function by re-evaluating the Jacobian for a given measurement at the *updated* state, rather than the predicted one, until convergence (usually 2 or 3 iterations are sufficient).

Finally, another widely used 'ad-hoc' modification is noise-inflation, whereby unmodeled errors are accounted for by the addition of positive semi-definite matrices to the predicted covariances. Although helpful in certain scenarios, this method remains a problem specific heuristic adjustment and as such, the intensity of the so-called 'stabilizing noise' needs to be chosen by trial and error. Also, this tuning strategy will not eliminate inconsistency if it is caused by bias in the estimates.

Despite its well-documented shortcomings, the established adequacy of the EKF in many 'benign' problems, together with its low computation demands, continue to justify its use where more sophisticated methods are not needed.

The Unscented Kalman Filter

The unscented Kalman filter [46, 47, 49, 82] is a moment matching filter that relies on the unscented transform proposed by Julier and Uhlmann. It may be interpreted as a linear least squares estimator in which discrete approximations are employed of the distributions

generating the covariances involved. This scheme is based on the intuition that approximating an arbitrary distribution is easier than approximating an arbitrary nonlinear transformation. Therefore, in contrast with the analytical linearization of the functions f_{k-1} and h_k carried out in the EKF, the unscented Transform makes use of statistical linearization for the approximation of densities resulting from nonlinear transformations [36].

Consider the n-dimensional random variable $\mathbf{x}_{k-1}$ with mean $\hat{\mathbf{x}}_{k-1}$ and covariance $\mathbf{P}_{k-1}$. The UKF estimates the posterior density $p(\mathbf{x}_k|\mathbf{Z}_k)$ by a Gaussian distribution using a number of deterministically chosen samples. This is achieved by first generating a set of N weighted so-called *sigma points* $\boldsymbol{\chi}^i_{k-1}, i = 0, \ldots, N-1$ whose sample mean, covariance and all higher odd-ordered central moments exactly match those of the density $p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1})$:

$$\begin{aligned}
\boldsymbol{\chi}^0_{k-1} &= \hat{\mathbf{x}}_{k-1} & W^i &= \frac{\kappa}{(n+\kappa)} & i &= 0 \\
\boldsymbol{\chi}^i_{k-1} &= \hat{\mathbf{x}}_{k-1} + \left(\sqrt{(n+\kappa)\mathbf{P}_{k-1}}\right)_i & W^i &= \frac{\kappa}{2(n+\kappa)} & i &= 1, \ldots, n \\
\boldsymbol{\chi}^i_{k-1} &= \hat{\mathbf{x}}_{k-1} - \left(\sqrt{(n+\kappa)\mathbf{P}_{k-1}}\right)_{i-n} & W^i &= \frac{\kappa}{2(n+\kappa)} & i &= n+1, \ldots, 2n
\end{aligned} \tag{2.3.18}$$

where $\left(\sqrt{(n+\kappa)\mathbf{P}_{k-1}}\right)_i$ denotes the ith row of the matrix square root $\mathbf{L}$ of $(n+\kappa)\mathbf{P}_{k-1}$, such that $(n+\kappa)\mathbf{P}_{k-1} = \mathbf{L}^{\mathrm{T}}\mathbf{L}$, W^i is the weight of the ith point and $\kappa \in \mathbb{R}$ is a design parameter (such that $n + \kappa \neq 0$) whose value can be adjusted to incorporate possible knowledge of the distribution's higher order moments. The choice of matrix square root does not affect the mean and covariance and thus numerically efficient and stable methods such as the Cholesky decomposition can be used.

The sigma points are then propagated through the nonlinear model to produce a set of transformed sigma points

$$\boldsymbol{\chi}^i_{k|k-1} = f_{k-1}\left[\boldsymbol{\chi}^i_{k-1}, \mathbf{u}^s_{k-1}\right] \qquad i = 0, \ldots, 2n \tag{2.3.19}$$

whose sample mean and sample sample covariance

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=0}^{2n} W^i \boldsymbol{\chi}^i_{k|k-1} \tag{2.3.20}$$

$$\mathbf{P}_{k|k-1} = \sum_{i=0}^{2n} W^i(\boldsymbol{\chi}^i_{k|k-1} - \hat{\mathbf{x}}^i_{k|k-1})(\boldsymbol{\chi}^i_{k|k-1} - \hat{\mathbf{x}}^i_{k|k-1})^{\mathrm{T}} + \mathbf{Q}_{k-1} \tag{2.3.21}$$

provide the statistics of the Gaussian approximation of the prediction $p(\mathbf{x}_k|\mathbf{Z}_{k-1})$. It is shown in [46] that these are accurate up to the second order of nonlinearity (and up to the third for Gaussian priors). The transformed sigma points $\boldsymbol{\chi}^i_{k|k-1}$ are then run through the nonlinear observation model h_k, resulting in the set of points

$$\boldsymbol{\mathcal{Z}}^i_{k|k-1} = h_k(\boldsymbol{\chi}^i_{k|k-1}, \mathbf{u}^m_k) \qquad i = 0, \ldots, 2n \tag{2.3.22}$$

whose sample mean and covariance

$$\hat{\mathbf{z}}_{k|k-1} = \sum_{i=0}^{2n} W^i \boldsymbol{\mathcal{Z}}^i_{k|k-1} \tag{2.3.23}$$

$$\mathbf{S}_k = \sum_{i=0}^{2n} W^i(\boldsymbol{\mathcal{Z}}^i_{k|k-1} - \hat{\mathbf{z}}^i_{k|k-1})(\boldsymbol{\mathcal{Z}}^i_{k|k-1} - \hat{\mathbf{z}}^i_{k|k-1})^{\mathrm{T}} + \mathbf{R}_{k-1} \tag{2.3.24}$$

represent the predicted measurement and the innovation covariance, respectively. The update equations of the UKF therefore can be written as

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k} + \mathbf{K}_k(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}) \tag{2.3.25}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^{\mathrm{T}} \tag{2.3.26}$$

where the Kalman gain, expressed in terms of the predicted state/measurement cross-covariance is

$$\mathbf{K}_k = \mathbf{P}^{\mathbf{xz}}_{k|k-1}\mathbf{S}_k^{-1} \tag{2.3.27}$$

$$\mathbf{P}^{xz}_{k|k-1} = \sum_{i=0}^{2n} W^i(\boldsymbol{\chi}^i_{k|k-1} - \hat{\mathbf{x}}^i_{k|k-1})(\boldsymbol{\mathcal{Z}}^i_{k|k-1} - \hat{\mathbf{z}}^i_{k|k-1})^{\mathrm{T}}. \tag{2.3.28}$$

We note that the UKF can be applied to the more general case of non-additive noise by augmenting the state vector with the process noise as described in [46].

The unscented transform does not involve truncation of the series expansions of the functions f_{k-1} and h_k at any particular order and it can be shown ([50]) that this

results in partial incorporation of the higher order terms. Consequently, the higher order moments of the approximation can be "fine tuned" to partially exploit prior knowledge of those of the distribution. This can be done through the choice of the parameter κ, which for Gaussian priors can be set according to the heuristic $n + \kappa = 3$ [47]. In contrast, linearization truncates the series expansion of the nonlinear functions at the first order, resulting in a predicted conditional mean accurate only up to the first order, while the covariance is approximated accurately up to the second order. This is a result of the fact that the correct calculation of the mean of a transformed random variable up to the mth order requires knowledge of both the derivatives of the transformation and the moments of the original distribution up to the mth order [49].

The UKF has the advantage of not requiring the computation of Jacobians or Hessians or indeed any approximation of the nonlinear functions f_{k-1} and h_k. Moreover, the deterministic sampling scheme on which it is based can capture possible discontinuities in the nonlinear transformations if they significantly affect the distribution [46]. All these factors contribute to the established superior accuracy of the UKF over the EKF for nonlinear estimation problems, while imposing comparable computational demands.

More sophisticated variants of the UKF have been proposed, lately, since the initial work by Julier and Uhlmann. Van der Merwe and Wan presented an alternative scheme for the choice of the sigma points and their weights in [82] , involving more tuning possibilities. The same authors have also introduced the square-root unscented Kalman filter [81] as a numerically more stable version of the UKF. Also, Date and Ponomareva ([22, 61]) as well as Tenne and Singh ([72]) have also proposed methods to match the higher order moments of the transformed pdf.

2.3.2 Sequential Monte Carlo Methods

Sequential Monte Carlo (SMC) methods are a class of simulation based estimation algorithms. They rely on the discrete representation of pdfs by weighted sets of samples, or 'particles', from which estimates of interest can be computed. As the number of particles is increased, this representation approaches the functional description of the actual pdf. A major benefit of these techniques is that they enable the characterization of arbitrary

non-Gaussian pdfs arising in nonlinear estimation problems. Although introduced in the 1950s, it wasn't before the 1990s that they attracted research interest [27, 38, 63], coinciding with the availability of sufficiently fast computers. Recent advances have helped increase the efficiency of SMCs and allowed their use in all varieties of estimation problems.

The aim is the estimation of the joint posterior $p(\mathbf{X}_k|\mathbf{Z}_k)$, along with $p(\mathbf{x}_k|\mathbf{Z}_k)$ and any moments of interest of the form

$$I(f) \;=\; \int f(\mathbf{X}_k)p(\mathbf{X}_k|\mathbf{Z}_k)\; d\mathbf{X}_k. \tag{2.3.29}$$

Assuming that $N \gg 1$ samples can be drawn independently from the distribution $p(\mathbf{X}_k|\mathbf{Z}_k)$, a discrete approximation of the full posterior is given by

$$\hat{p}(\mathbf{X}_k|\mathbf{Z}_k) \;=\; \frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{X}_k - \mathbf{X}_k^i), \tag{2.3.30}$$

where δ is the Dirac delta function, and the Monte Carlo (MC) estimate of the integral $I(f)$ becomes the sample mean:

$$\hat{I}(f) \;=\; \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{X}_k^i). \tag{2.3.31}$$

It can be shown that in the limit $N \to \infty$ the estimate $\hat{I}(f)$ is unbiased and will almost surely converge to $I(f)$ (in accordance with the strong law of large numbers). Unfortunately, however, it is not possible to draw samples directly from the posterior itself. Importance sampling [63], as will be described, provides an answer to this problem.

The Particle Filter

Assume that we can draw samples from a density $\pi(\mathbf{X}_k|\mathbf{Z}_k)$, referred to as *importance* or *proposal* density, which is similar to $p(\mathbf{X}_k|\mathbf{Z}_k)$ in the sense that $p(\mathbf{X}_k|\mathbf{Z}_k) > 0 \Rightarrow \pi(\mathbf{X}_k|\mathbf{Z}_k) > 0$ (indicating that the two functions have the same support). It is then possible to estimate the integral $I(f)$ by using an appropriate weighting of the samples

drawn from $\pi(\mathbf{X}_k|\mathbf{Z}_k)$. Indeed the integral can be rewritten in the form

$$I(f) \;=\; \int f(\mathbf{X}_k)p(\mathbf{X}_k|\mathbf{Z}_k)\; d\mathbf{X}_k \;=\; \int f(\mathbf{X}_k)\frac{p(\mathbf{X}_k|\mathbf{Z}_k)}{\pi(\mathbf{X}_k|\mathbf{Z}_k)}\pi(\mathbf{X}_k|\mathbf{Z}_k)\; d\mathbf{X}_k, \tag{2.3.32}$$

which, assuming that $p(\mathbf{X}_k|\mathbf{Z}_k)/\pi(\mathbf{X}_k|\mathbf{Z}_k)$ is upper bounded, can be computed using MC integration by drawing independent samples $\{\mathbf{X}_k^i\}_{i=1}^N$ from the density $\pi(\mathbf{X}_k|\mathbf{Z}_k)$ and evaluating the weighted sum:

$$\hat{I}(f) \;=\; \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{X}_k^i)w_k^i, \tag{2.3.33}$$

where the weights are given by

$$w_k^i \;\propto\; \frac{p(\mathbf{X}_k^i|\mathbf{Z}_k)}{\pi(\mathbf{X}_k^i|\mathbf{Z}_k)}. \tag{2.3.34}$$

The posterior can therefore be represented by the discrete approximation

$$p(\mathbf{X}_k|\mathbf{Z}_k) \;\approx\; \sum_{i=1}^{N} w_k^i\, \delta(\mathbf{X}_k - \mathbf{X}_k^i). \tag{2.3.35}$$

A sequential implementation of this framework can be obtained if the importance density is chosen such that it can be factored in the form:

$$\pi(\mathbf{X}_k|\mathbf{Z}_k) \;\triangleq\; \pi(\mathbf{x}_k|\mathbf{X}_{k-1}, \mathbf{Z}_k)\pi(\mathbf{X}_{k-1}|\mathbf{Z}_{k-1}), \tag{2.3.36}$$

which results ([63]) in the following weight update equation (up to a normalizing constant):

$$w_k^i \;\propto\; w_{k-1}^i \frac{p(\mathbf{z}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{\pi(\mathbf{x}_k^i|\mathbf{X}_{k-1}^i, \mathbf{Z}_k)}. \tag{2.3.37}$$

Also, the need to store the entire history of samples along with the measurements can be avoided with the assumption that $\pi(\mathbf{x}_k|\mathbf{X}_{k-1}, \mathbf{Z}_k) = \pi(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k)$. In this case, the importance density at time k will depend only on $\mathbf{x}_{k-1}$ and $\mathbf{z}_k$. The unnormalized weights will be given by:

$$w_k^i \;\propto\; w_{k-1}^i \frac{p(\mathbf{z}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{\pi(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{z}_k)}, \tag{2.3.38}$$

and the filtered posterior density $p(\mathbf{x}_k|\mathbf{Z}_k)$ will be approximated by:

$$p(\mathbf{x}_k|\mathbf{Z}_k) \approx \sum_{i=1}^{N} w_k^i \, \delta(\mathbf{x}_k - \mathbf{x}_k^i) \tag{2.3.39}$$

This is basis of the sequential importance sampling (SIS) algorithm, which recursively propagates a set of particles and computes their associates weights in order to produce a discrete approximation of the posterior. Statistics of interest such as the mean and covariance can then be estimated by the equations

$$\hat{\mathbf{x}}_{k|k} = \sum_{i=1}^{N} w_k^i \mathbf{x}_k^i \tag{2.3.40}$$

$$\mathbf{P}_{k|k} = \sum_{i=1}^{N} w_k^i \left[\hat{\mathbf{x}}_{k|k} - \mathbf{x}_k^i\right] \left[\hat{\mathbf{x}}_{k|k} - \mathbf{x}_k^i\right]^{\mathrm{T}}. \tag{2.3.41}$$

The major drawback of the SIS framework is the *degeneracy* phenomenon, whereby the variance of the weights inevitably grows over time, eventually leading to all but one particle having near zero normalized weight. This results in a loss of accuracy and a waste of computational resources due to the simulation of particles with negligible weights. An effective method to manage this problem is resampling of the particles when excessive degeneracy is encountered [38]. Degeneracy can be suitably quantified by the *effective sample size* N_{eff}, which can be estimated as:

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N} \left(w_k^i\right)^2}. \tag{2.3.42}$$

Each time N_{eff} falls below a specified threshold, the particle set is resampled so that particles with non-negligible weight are replicated while those with negligible weight are discarded. The obtained N samples are then assigned equal weights of $1/N$. There are several methods to generate these samples, with different levels of simplicity and MC variance. We give here an outline of some of the most commonly employed techniques (detailed reviews are available in [41] and [26]):

Multinomial resampling

Used in the original sampling importance resampling (SIR) filter of [38], also presented in [54], multinomial resampling consists in drawing (with replacement) N independent

samples $\{\mathbf{x}_k^{i\,*}\}_{i=1}^N$ from the approximate discrete distribution (2.3.39) such that $P\{\mathbf{x}_k^{i\,*} = \mathbf{x}_k^j\} = w_k^j$. This results in a new set of i.i.d samples with uniform weights equal to $1/N$. Algorithmically, this scheme starts with the generation of N ordered uniform random numbers $u_n \sim \mathcal{U}(0,1)$ where $n = 1, \ldots, N$. The new samples $\mathbf{x}_k^{i\,*}$ are then chosen following the multinomial distribution so that:

$$\begin{aligned} \mathbf{x}_k^{i\,*} &= \mathbf{x}(F^{-1}(u_n)) \\ &= \mathbf{x}_k^j, \quad \text{with } j \text{ is s.t } u_n \in \left[\sum_{s=1}^{j-1} w_k^s, \sum_{s=1}^{j} w_k^s\right), \end{aligned} \tag{2.3.43}$$

where F^{-1} is the inverse cumulative weight distribution function.

Stratified resampling

In [32], Fearnhead presents a technique based on a well known result from survey sampling theory. It states that the variance of an estimator of a function of a random variable θ based on a sample of N points can be reduced using stratification. By representing the density of θ as a mixture of densities and drawing from each one a number N_i of samples proportional to the weight of the mixture component yields an estimator with variance lower than can be achieved by simple random sampling from the original density [21]. In practice, the stratified resampling scheme can be implemented by generating N ordered random numbers

$$u_k = \frac{(k-1) + \tilde{u}_k}{N}, \quad \text{with } \tilde{u}_k \sim \mathcal{U}(0,1) \tag{2.3.44}$$

and then selecting the samples $\mathbf{x}_k^{i\,*}$ by way of the multinomial distribution.

Systematic resampling

Introduced in [54] (and also used in [12] under the name of 'stratified resampling'), systematic resampling is a deterministic variant of the stratified resampling scheme. Instead of sampling randomly from each subinterval of width $1/N$, a uniform random variable is generated to represent the draw from the first subinterval; subsequent draws are then inferred deterministically at constant intervals of $1/N$. Thus, the algorithm reduces to the generation of N ordered numbers

$$u_k = \frac{(k-1) + \tilde{u}}{N}, \quad \text{with } \tilde{u} \sim \mathcal{U}(0,1), \tag{2.3.45}$$

and the selection of the samples $\mathbf{x}_k^{i\,*}$ is done (as before) through the multinomial distribution. Note that, the interdependence which is introduced between the samples makes the theoretical study of this method difficult (notably in terms of MC variance). Nevertheless, its simplicity and well documented empirical performance make it the preferred choice in most applications.

Despite its apparent benefit, the resampling step introduces other difficulties [63], most notably that of *sample impoverishment*. The continual application of resampling in the filtering process can lead to the sample set consisting only of copies of the same few particles replicated many times. This loss of particle diversity is most severe in cases where process noise is low or absent altogether. Possible ways to counteract this problem are the use of Markov chain Monte Carlo (MCMC) methods [11] and regularization [58].

The final, and crucial, point to consider in particle filter design is the choice of importance density π. To minimize the variance of importance weights, this density needs to be as close to the posterior as possible. The optimal choice is $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{z}_k)_{opt} = p(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{z}_k)$, but it is in practice not possible to sample from it [63]. The most straightforward suboptimal option then is the transitional prior (or prediction density)

$$\pi(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{z}_k) = p(\mathbf{x}_k|\mathbf{x}_{k-1}^i). \tag{2.3.46}$$

This choice has the benefit of simplicity, as it results (through substitution of (2.3.46) in (2.3.37)) in the weights being given by:

$$w_k^i \;\propto\; w_{k-1}^i p(\mathbf{z}_k|\mathbf{x}_k^i). \tag{2.3.47}$$

However, as it does not incorporate the latest measurement $\mathbf{z}_k$, the prior can be a crude approximation of the posterior. Particularly, applications involving low measurement noise variance relative to that of the system noise, can result in the majority of particles being drawn from areas of the state space where the measurement likelihood is low. The effect of this small overlap between prior and likelihood can then only be counteracted by the 'brute-force' use of even larger numbers of particles.

Better approximations of the optimal importance density can be obtained through

(suboptimal) analytical estimates that incorporate the latest measurement $\mathbf{z}_k$. In particular, as described in [80], a bank of EKFs or UKFs can be used to generate, for each particle, an importance density of the form

$$\pi(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{z}_k) = \mathcal{N}(\mathbf{x}_k^i;\, \hat{\mathbf{x}}_{k|k}^i, \mathbf{P}_{k|k}^i), \tag{2.3.48}$$

where $\hat{\mathbf{x}}_{k|k}^i$ and $\mathbf{P}_{k|k}^i$ are the mean and covariance of particle $\mathbf{x}_k^i$, conditioned on $\mathbf{z}_k$, as estimated by the EKF or the UKF. The importance weights, in this case, are given by (2.3.38) with the substitution of the denominator by (2.3.48).

The Ensemble Kalman Filter

The ensemble Kalman filter (EnKF), used primarily in large-scale problems involving up to millions of state variables and observation components, is a Monte Carlo based suboptimal filter that makes use of sample statistics. Although we present here only a brief overview of the algorithm, the reader is referred to [31,42] for more detailed analyses. The EnKF is mostly used in weather forecasting and ocean sciences, where the complexity of the underlying systems preclude the use of filters such as the EKF and UKF, which require storage and computation of the error covariance matrix. Within the EnKF, the system state vector $\mathbf{x}_k$ is represented by an ensemble $\mathbf{X}_k \triangleq (\mathbf{x}_k^1, \ldots, \mathbf{x}_k^q)$ of q vectors, each denoting a possible hypothesis of the state vector, which are propagated forward by the system equations to represent state 'forecasts'. These are then individually updated in an 'analysis' step, with perturbed versions of the measurements, using the standard Kalman filter update equations. The computation of the Kalman gain (which is the same for all ensemble members) as well as the error covariance matrix are carried out using the sample statistics of the ensemble members (in a manner very similar to the UKF where these quantities are evaluated as the sample statistics of sigma points).

2.4 Target Tracking Models

Target tracking is an application area of estimation theory, where the state of a moving object (comprising usually of its position, velocity, acceleration and/or other kinematic attributes of interest) is estimated from noisy observations. These can consist (among others) of a combination of range, range-rate and bearings, sampled at discrete time in-

stants from one or more sensors. The application of Bayesian estimation methods requires knowledge of a mathematical model describing, with sufficient accuracy, the time evolution of the states. Various such models exist and their choice is dictated by the characteristics of the tracking scenario under consideration. Moving targets are nearly always treated as point objects by such dynamic models, and while sacrificing valuable shape and orientation information, this simplification considerably reduces the complexity of the problem.

Target dynamics are typically defined in continuous time to reflect their actual physical behaviour, but discretized for easy reconciliation with the measurement process which is in discrete time. Discretized continuous-time dynamic models are intuitively more consistent with the actual target kinematics than models defined directly in discrete-time [66]. We provide here an outline of some of the most widely used dynamic models in the tracking literature.

2.4.1 The Discretized Continuous White Noise Acceleration Model

The continuous white noise acceleration model (also known as 'nearly constant velocity' model) describes the motion of an object whose acceleration is represented by zero-mean white noise. The acceleration, along a generic coordinate ζ, of an object following such a model is described in continuous time by [8]:

$$\ddot{\zeta}(t) \;=\; \tilde{v}(t), \tag{2.4.1}$$

where $\tilde{v}(t)$ is zero-mean white noise with power spectral density $\tilde{q}$. The corresponding two-dimensional state vector is then

$$\mathbf{x} \;=\; [\zeta \;\; \dot{\zeta}]^{\mathrm{T}}, \tag{2.4.2}$$

and the continuous-time state equation is

$$\dot{\mathbf{x}}(t) \;=\; \mathbf{A}\mathbf{x}(t) + \mathbf{D}\tilde{v}(t), \tag{2.4.3}$$

where

$$\mathbf{A} \;=\; \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{D} \;=\; \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{2.4.4}$$

The corresponding discrete-time state equation (with sampling period T) is

$$\mathbf{x}_k \;=\; \mathbf{F}\mathbf{x}_{k-1} + \mathbf{v}_k, \tag{2.4.5}$$

where the transition matrix is

$$\mathbf{F} \;=\; e^{\mathbf{A}T} \;=\; \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \tag{2.4.6}$$

and the discrete-time noise term is expressed as

$$v_k \;=\; \int_0^T e^{\mathbf{A}(T-t)}\mathbf{D}\tilde{v}(kT+\tau)\,d\tau, \tag{2.4.7}$$

with covariance matrix

$$\mathbf{Q} \;=\; E[\mathbf{v}_k\,\mathbf{v}_k^{\mathrm{T}}] \;=\; \begin{bmatrix} \frac{1}{3}T^3 & \frac{1}{2}T^2 \\ \frac{1}{2}T^2 & T \end{bmatrix} \tilde{q}. \tag{2.4.8}$$

As its name suggests, the nearly constant velocity model is appropriate when instantaneous changes in velocity (accelerations) are 'small' relative to the actual velocity. If they are not, as will be explained next, a third-order model that includes an acceleration component can be used.

2.4.2 The Discretized Continuous Wiener Process Acceleration Model

The continuous Wiener process acceleration model (also known as 'nearly constant acceleration' and 'white noise jerk' model) is a third-order model that describes the motion of an object undergoing slight variations in acceleration (known as jerk), that are modeled as zero-mean white noise. The jerk, along a generic coordinate ζ, of an object following this model is described in continuous time by [8]:

$$\dddot{\zeta}(t) \;=\; \tilde{v}(t). \tag{2.4.9}$$

It can be seen that the acceleration, which is the integral of the jerk, is then a Wiener process, explaining the name of the model. The three-dimensional state vector is now

$$\mathbf{x} \;=\; [\zeta \;\; \dot{\zeta} \;\; \ddot{\zeta}]^{\mathrm{T}}, \tag{2.4.10}$$

and the continuous-time state equation has the same form as (2.4.3), with

$$\mathbf{A} \;=\; \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{D} \;=\; \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{2.4.11}$$

The discretized state equation is of the form (2.4.5), where the transition matrix is now

$$\mathbf{F} \;=\; e^{\mathbf{A}T} \;=\; \begin{bmatrix} 1 & T & \frac{1}{2}T^2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.4.12}$$

and the covariance matrix of the noise term $v(k)$ is

$$\mathbf{Q} \;=\; E[\mathbf{v}_k\,\mathbf{v}_k^{\mathrm{T}}] \;=\; \begin{bmatrix} \frac{1}{20}T^5 & \frac{1}{8}T^4 & \frac{1}{6}T^3 \\ \frac{1}{8}T^4 & \frac{1}{3}T^3 & \frac{1}{2}T^2 \\ \frac{1}{6}T^3 & \frac{1}{2}T^2 & T \end{bmatrix} \tilde{q} \tag{2.4.13}$$

The intensity $\tilde{q}$ of the process noise should be 'small' relative to the actual acceleration that the target undergoes for this model to be an accurate representation of actual kinematics. We note that the nearly constant acceleration model can describe 'linear' motion as well as 'maneuvering' motion where the target course changes. Nevertheless, it is inappropriate in the modeling of maneuvers that obey known kinematic laws. These can be described with more accuracy using a variety of special models (see [66]).

2.4.3 The 3D Constant-Turn Model

Most 2D and 3D target maneuvers are turn motions models. A natural way to derive these is from target kinematics rather than random processes. We focus here on an important class of turn models known as constant turn (sometimes also referred to as 'coordinated turn'), which is characterized by constant target speed and turn rate.

The general motion of an object rotating around a fixed center $\mathbf{p}_0$ is described in terms of its linear and angular velocity vectors $\mathbf{v}$ and $\boldsymbol{\Omega}$ by:

$$\mathbf{v} \;=\; \boldsymbol{\Omega} \times (\mathbf{p} - \mathbf{p}_0). \tag{2.4.14}$$

Differentiating (2.4.14), with the constant angular velocity assumption $\dot{\boldsymbol{\Omega}} = 0$ [66] yields the basic kinematic relation for 3D fixed-center constant angular-velocity motion

$$\mathbf{a} \;=\; \boldsymbol{\Omega} \times \mathbf{v}, \tag{2.4.15}$$

from which the angular velocity vector can be expressed (noting that $\boldsymbol{\Omega} \perp \mathbf{v}$) as

$$\boldsymbol{\Omega} \;=\; \frac{\mathbf{v} \times \mathbf{a}}{\|\mathbf{v}\|^2}. \tag{2.4.16}$$

This relation, describing the kinematics of a constant turn model, implies that $\boldsymbol{\Omega} \perp \mathbf{a}$ and that the motion takes place on a 'maneuver plane' orthogonal to the direction of $\boldsymbol{\Omega}$, which is defined by the velocity and acceleration vectors. Differentiating (2.4.15) and using the fact that $\boldsymbol{\Omega} \perp \mathbf{v}$, an alternative formulation can be arrived at, which involves the turn rate ω [66]:

$$\begin{aligned} \dot{\mathbf{a}} \;&= \boldsymbol{\Omega} \times \mathbf{a} \\ &= \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{v}) \\ &= (\boldsymbol{\Omega} \cdot \mathbf{v})\, \boldsymbol{\Omega} - (\boldsymbol{\Omega} \cdot \boldsymbol{\Omega})\, \mathbf{v} \\ &= -\omega^2 \mathbf{v} \end{aligned} \tag{2.4.17}$$

where

$$\omega \;\triangleq\; \|\boldsymbol{\Omega}\| \;=\; \frac{\|\mathbf{v} \times \mathbf{a}\|}{\|\mathbf{v}\|^2} \tag{2.4.18}$$

As a result, the constant turn (CT) maneuver model can be represented by a second-order Markov process

$$\dot{\mathbf{a}} \;=\; -\omega^2 \mathbf{v} + \tilde{\mathbf{w}}, \tag{2.4.19}$$

where $\tilde{\mathbf{w}}$ is white noise with power spectral density $\sigma_\omega^2 \mathbf{I}_{n\times n}$, with n denoting the dimensionality of the Euclidian space (the intensity σ_ω of the noise can be different along x, y and z directions). Provided $\tilde{\mathbf{w}}$ is small, (2.4.19) can be interpreted as a 'nearly' planar motion model. The state-space representation in 3D cartesian coordinates, with the 9-dimensional state vector $\mathbf{x} = [x, \dot{x}, \ddot{x}, y, \dot{y}, \ddot{y}, z, \dot{z}, \ddot{z}]^{\mathrm{T}}$, is then

$$\dot{\mathbf{x}}(t) = \mathrm{diag}[\mathbf{A}(\omega), \mathbf{A}(\omega), \mathbf{A}(\omega)]\mathbf{x}(t) + \mathrm{diag}[\mathbf{D}, \mathbf{D}, \mathbf{D}]\tilde{\mathbf{w}}(t), \tag{2.4.20}$$

where

$$\mathbf{A}(\omega) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -\omega^2 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{2.4.21}$$

Discretization results in the state equation [66]

$$\mathbf{x}_k = \mathrm{diag}[\mathbf{F}(\omega), \mathbf{F}(\omega), \mathbf{F}(\omega)]\mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \tag{2.4.22}$$

where

$$\mathbf{F}(\omega) = \begin{bmatrix} 1 & \frac{\sin\omega T}{\omega} & \frac{1-\cos\omega T}{\omega^2} \\ 0 & \cos\omega T & \frac{\sin\omega T}{\omega} \\ 0 & -\omega\sin\omega T & \cos\omega T \end{bmatrix} \tag{2.4.23}$$

and the covariance matrix of the discrete-time noise is:

$$\begin{aligned} \mathrm{cov}(\mathbf{w}_k) &= E[\mathbf{n}_k\, \mathbf{n}_k^{\mathrm{T}}] \\ &= \mathrm{diag}[\sigma_\omega^2\mathbf{Q}(\omega), \sigma_\omega^2\mathbf{Q}(\omega), \sigma_\omega^2\mathbf{Q}(\omega)], \end{aligned} \tag{2.4.24}$$

with

$$\mathbf{Q}(\omega) = \begin{bmatrix} \frac{6\omega T - 8\sin\omega T + \sin 2\omega T}{4\omega^5} & \frac{2\sin^4(\omega T/2)}{\omega^4} & \frac{-2\omega T + 4\sin\omega T - sin2\omega T}{4\omega^3} \\ \frac{2sin^4(\omega T/2)}{\omega^4} & \frac{2\omega T - \sin 2\omega T}{4\omega^3} & \frac{\sin^2\omega T}{2\omega^2} \\ \frac{-2\omega T + 4\sin\omega T - \sin 2\omega T}{4\omega^3} & \frac{\sin^2\omega T}{2\omega^2} & \frac{2\omega T + \sin 2\omega T}{4\omega} \end{bmatrix} \sigma_\omega^2. \tag{2.4.25}$$

We can see that it is only through the common turn rate ω that the motions along the x, y and z directions are coupled. Knowledge of ω can either be assumed available, or it can be treated as an unknown state variable to be estimated. Note that the special case where

the maneuver plane is horizontal reduces the model to the so-called 2D 'coordinated turn' model [8, 9].

It is important to note that the 3D constant-turn model is one among many examples of 3D maneuver modes, of which more can be found in [66]. In most tracking applications, knowledge of the exact maneuver dynamics are unknown to the observer. Therefore, more general models such as the 'nearly constant acceleration' model (which encompasses a wide range of possible maneuvers) are often resorted to. Such model mismatches increase tracker robustness to unanticipated dynamics, but come at the cost of lower accuracy, in comparison to model-matched trackers.

2.4.4 The Interacting Multiple Model Algorithm

The interacting multiple-model (IMM) algorithm ([10]) is a computationally efficient suboptimal framework employed in the context of multiple model tracking. Multiple model formulations of tracking problems are widely used to model targets that execute maneuvers. These typically refer to any departure from 'nearly constant velocity' straight-line motion, which is the defining model for the well studied problem of 'target motion analysis' (TMA). In this thesis, we assume that systems operate under one of a finite number of dynamic models (such as those previously discussed). As a result, we treat the maneuvering target tracking problem as one of hybrid estimation in which the state vector $\mathbf{x}_k$ is augmented with a 'mode' parameter r_k, taking values in the discrete set $S = \{1, \ldots, M\}$, and indicating the dynamic model which is in effect during the time interval $(t_{k-1}, t_k]$. The mode sequence $\{r_k\}$ is assumed to follow a (hidden) first-order Markov process, with the unknown transition probabilities set as design parameters.

We define a switching system of the form

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, r_k) + \mathbf{u}_k^s + \mathbf{v}(r_k) \tag{2.4.26}$$

$$\mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{u}_k^m + \mathbf{w}(r_k), \tag{2.4.27}$$

with f describing the mode dependent system dynamics, along with the corresponding mode dependent noise vectors $\mathbf{v}$ and $\mathbf{w}$. The transition probabilities of the system switch-

ing from the ith mode to the jth mode are given as

$$P\{r_k = j | r_{k-1} = i\} \;=\; \pi_{ij}, \tag{2.4.28}$$

and illustrated in figure 2.2

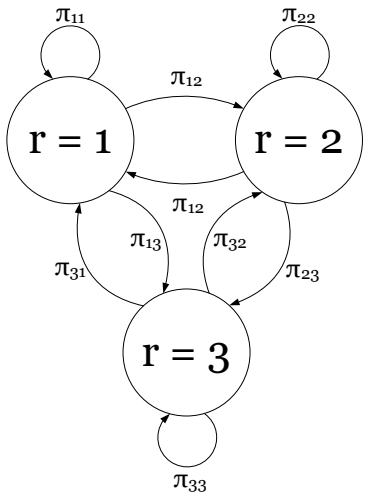


Figure 2.2: Markov chain for 3 state system.

A Bayesian framework can then be used to approximate the posterior probabilities at time k of the M possible modes having been in effect during $(t_{k-1}, t_k]$, and their associated state estimates, conditioned on the measurements up to time k. The posterior state density is given by

$$p(\mathbf{x}_k|\mathbf{Z}_k) \;=\; \sum_{l=1}^{M^k} p(\mathbf{x}_k|r^l_{1:k}, \mathbf{Z}_k) P\{r^l_{1:k}|\mathbf{Z}_k\}, \tag{2.4.29}$$

where $r^l_{1:k}$ denotes the l'th possible mode history up to time k (out of M^k possible sequences). The probabilities $P\{r^l_{1:k}|\mathbf{Z}_k\}$, henceforth denoted μ^l_k, can be expressed as

$$\begin{aligned}
\mu^l_k \;&= P\{r^l_{1:k}|\mathbf{Z}_k\} \\
&= \frac{1}{c}\, p(\mathbf{z}_k|r^l_{1:k}, \mathbf{Z}_{k-1}) P\{r^l_{1:k}|\mathbf{Z}_{k-1}\} \\
&= \frac{1}{c}\, p(\mathbf{z}_k|r^l_{1:k}, \mathbf{Z}_{k-1}) P\{r^j_k, r^s_{1:k-1}|\mathbf{Z}_{k-1}\} \\
&= \frac{1}{c}\, p(\mathbf{z}_k|r^l_{1:k}, \mathbf{Z}_{k-1}) P\{r^j_k|r^s_{1:k-1}, \mathbf{Z}_{k-1}\} P\{r^s_{1:k-1}|\mathbf{Z}_{k-1}\} \\
&= \frac{1}{c}\, p(\mathbf{z}_k|r^l_{1:k}, \mathbf{Z}_{k-1}) P\{r^j_k|r^s_{1:k-1}\} \mu^s_{k-1},
\end{aligned}$$

where c is a normalizing constant. Using the Markov property of the mode switching sequence, we arrive at

$$\begin{aligned}\mu_k^l &= \frac{1}{c}\, p(\mathbf{z}_k|r_{1:k}^l, \mathbf{Z}_{k-1}) P\{r_k^j|r_{k-1}^i\}\mu_{k-1}^s \\ &= \frac{1}{c}\, p(\mathbf{z}_k|r_{1:k}^l, \mathbf{Z}_{k-1})\pi_{ij}\mu_{k-1}^s, \end{aligned} \tag{2.4.30}$$

where i denotes the last mode value of the sequence s.

We can notice from (2.4.29) that the size of the Gaussian mixture representation of the posterior density grows exponentially with time. This results from the need to store *all* past histories of the state (even though the random parameters follow Markov processes) and estimating each one of them with a separate filter. Consequently, strategies need to be employed that limit the number of terms in the Gaussian mixture.

The IMM algorithm, proposed in [10], provides an efficient structure for approximating the exponentially growing mixture representation of the posterior pdf, using a bank of M model-matched filters, as

$$p(\mathbf{x}_k|\mathbf{Z}_k) \;=\; \sum_{j=1}^{M} p(\mathbf{x}_k|r_k^j, \mathbf{z}_k, \mathbf{Z}_{k-1})\mu_k^j, \tag{2.4.31}$$

where μ_k^j represents the posterior probability of the jth mode being in effect during the period $(t_{k-1}, t_k]$. The main feature of the IMM algorithm is the so-called 'mixing' of the model-matched posterior estimates at time $k-1$. This consists in defining, at time k and for $i, j = 1, \ldots, M$, the input to the jth model-matched filter to be the sum of the M model-matched estimates at time $k-1$, weighted by the probabilities of the ith mode having been in effect at time $k-1$, given that the jth mode is in effect at time k. This step avoids the quadratic growth in the number of mixture components, while achieving the accuracy of the more complex GPB2 algorithm ([8]).

One iteration of the IMM filter, of which a full derivation can be found in [8], can be summarized as:

1. Calculation of the mixing probabilities $\mu^{ij}_{k-1|k-1}$ for $i, j = 1, \ldots, M$:

$$\begin{aligned}\mu^{ij}_{k-1|k-1} &= P\{r^i_{k-1}|r^j_k, \mathbf{Z}_{k-1}\} \\ &= \frac{1}{\bar{c}^j} P\{r^j_k|r^i_{k-1}, \mathbf{Z}_{k-1}\} P\{r^i_{k-1}|\mathbf{Z}_{k-1}\} \\ &= \frac{1}{\bar{c}^j} \pi_{ij} \mu^i_{k-1}, \end{aligned} \tag{2.4.32}$$

   where $\bar{c}^j = \sum_{i=1}^M \pi_{ij}\mu^i_{k-1},\ j = 1, \ldots, M$, is a normalizing constant.

2. Mixing of the posterior model-matched estimates at time $k-1$, for $j = 1, \ldots, M$:

$$\bar{x}^j_k = \sum_{i=1}^M \hat{\mathbf{x}}^i_{k-1|k-1} \mu^{ij}_{k-1} \tag{2.4.33}$$

$$\bar{P}^j_k = \sum_{i=1}^M \mu^{ij}_{k-1} \left[\mathbf{P}^i_{k-1|k-1} + \left(\hat{\mathbf{x}}^i_{k-1|k-1} - \bar{\mathbf{x}}^j_{k-1|k-1}\right)\left(\hat{\mathbf{x}}^i_{k-1|k-1} - \bar{\mathbf{x}}^j_{k-1|k-1}\right)^{\mathrm{T}} \right] \tag{2.4.34}$$

3. Model-matched filtering and measurement likelihood calculation for $j = 1, \ldots, M$: given $\{\bar{x}^j_k, \bar{P}^j_k\}$, model-matched filters (such as banks of EKFs, UKFs or other filters) are used to compute estimates $\{\hat{\mathbf{x}}^j_{k|k}, \mathbf{P}^j_{k|k}\}$ and associated measurement likelihoods $L^j_k = p(\mathbf{z}_k|r^j_k, \mathbf{Z}_{k-1})$.

4. Evaluation of posterior mode probabilities μ^j_k, at time k, for $j = 1, \ldots, M$:

$$\begin{aligned}\mu^j_k &= P\{r^j_k|\mathbf{Z}_k\} \\ &= \frac{1}{c} p(\mathbf{z}_k|r^j_k, \mathbf{Z}_{k-1}) P\{r^j_k|\mathbf{Z}_{k-1}\} \\ &= \frac{1}{c} L^j_k \sum_{i=1}^M P\{r^j_k|r^i_{k-1}, \mathbf{Z}_{k-1}\} P\{r^i_{k-1}|\mathbf{Z}_{k-1}\} \\ &= \frac{1}{c} L^j_k \sum_{i=1}^M \pi_{ij} \mu^i_{k-1}, \end{aligned} \tag{2.4.35}$$

   where c is a normalizing constant.

5. Computation of posterior moment-matched density (for output):

$$\hat{\mathbf{x}}_{k|k} = \sum_{j=1}^M \hat{\mathbf{x}}^j_{k|k}\, \mu^j_k \tag{2.4.36}$$

$$\mathbf{P}_{k|k} = \sum_{j=1}^{M} \mu_k^j \left[\mathbf{P}_{k|k}^j + \left(\hat{\mathbf{x}}_{k|k}^j - \hat{\mathbf{x}}_{k|k} \right) \left(\hat{\mathbf{x}}_{k|k}^j - \hat{\mathbf{x}}_{k|k} \right)^{\mathrm{T}} \right] \tag{2.4.37}$$

2.4.5 The Bearings-only Measurement Model

Bearings-only tracking is the problem of estimating the state of a moving target from noisy measurements of its direction, relative to the sensor. The standard form of the discrete-time measurement equation in 3D Euclidian space, as typically represented in the literature ([65]) takes the following form:

$$\mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{w}_k, \tag{2.4.38}$$

where

$$h(\mathbf{x}_k) = \begin{bmatrix} \theta_k \\ \alpha_k \end{bmatrix} = \begin{bmatrix} \arctan\left(x_k / y_k \right) \\ \arctan\left(z_k / \sqrt{x_k^2 + y_k^2} \right) \end{bmatrix}, \tag{2.4.39}$$

and $\mathbf{w}_k$ is zero-mean white Gaussian noise with diagonal covariance matrix. $\theta_k \in [-\pi, \pi]$ and $\alpha_k \in [-\pi/2, \pi/2]$ can be recognized as the azimuth and elevation angles of the line of sight from the sensor to the target. As illustrated in figure 2.3, the cartesian displacement vector $\mathbf{d}$ of a target relative to the sensor can then be expressed, using spherical coordinates, as

$$\mathbf{d} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = r \begin{bmatrix} \sin\theta \cos\alpha \\ \cos\theta \cos\alpha \\ \sin\alpha \end{bmatrix}, \tag{2.4.40}$$

where r is the relative range of the target (also known as 'slant range').

In addition to the nonlinearity of the measurement model, an inherent difficulty of single-sensor bearings-only tracking is the absence of range information in individual measurements. This observability problem which leads to ambiguity in target localization can be resolved by maneuver of the sensor platform. In general, a 'rule of thumb' for ensuring observability is for the sensor platform trajectory to have "one more nonzero derivative than the target trajectory" [8]. For example, tracking of a stationary target

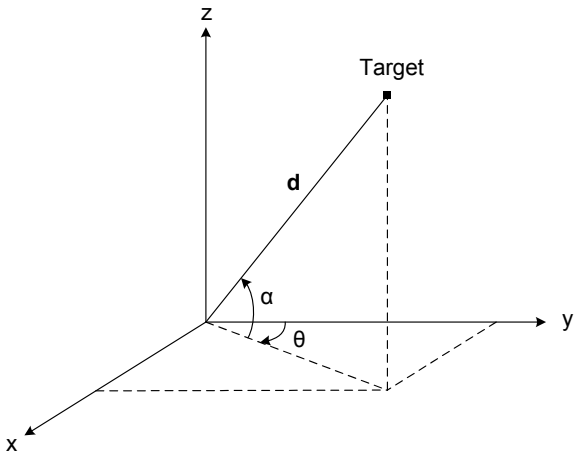


Figure 2.3: Azimuth and elevation angles θ and α in the 3D measurement frame

can only be achieved with a sensor platform moving at nonzero velocity, while a target moving at constant velocity requires the sensor platform to have nonzero acceleration. Although appropriate for most practical scenarios, this is only a *necessary* condition of observability. [44], [34] and [40] present in depth analyses of observability requirements for 3D bearings-only measurements and provide *sufficient* conditions on sensor platform trajectories to ensure observability.

2.5 Bearings-only Tracking Algorithms

Many adaptations and modifications of the previously outlined approximate Bayesian estimators have been applied to the bearings-only tracking problem. We present here some of the most notable algorithms that have appeared in the literature.

2.5.1 The Extended Kalman Filter in Cartesian Coordinates

The EKF presented in section 2.3.1 can be readily applied to bearings-only tracking. Defining the state vector with the position and velocity components of the target in Cartesian coordinates, *relative to the sensor*, as

$$\mathbf{x}_k = \begin{bmatrix} x_k & \dot{x}_k & y_k & \dot{y}_k & z_k & \dot{z}_k \end{bmatrix}^{\mathrm{T}}, \tag{2.5.1}$$

the Jacobian $\hat{\mathbf{H}}_k$ is given by

$$\begin{aligned}\hat{\mathbf{H}}_k &\triangleq [\nabla_{\mathbf{x}_k} h_k(\mathbf{x}_k, \mathbf{u}_k^m)]\Big|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}} \\ &= \left[\begin{array}{cccccc} \frac{y_k}{\rho_k^2} & 0 & -\frac{x_k}{\rho_k^2} & 0 & 0 & 0 \\ -\frac{x_k z_k}{\rho_k r_k^2} & 0 & -\frac{y_k z_k}{\rho_k r_k^2} & 0 & \frac{\rho_k}{r_k^2} & 0 \end{array}\right]\Bigg|_{\mathbf{x}_k=\hat{\mathbf{x}}_{k|k-1}},\end{aligned} \tag{2.5.2}$$

where

$$\rho_k = \sqrt{x_k^2 + y_k^2} \quad \text{and} \quad r_k = \sqrt{x_k^2 + y_k^2 + z_k^2} \tag{2.5.3}$$

are the relative ground range and (slant) range of the target. The recursion provided by equations (2.3.11) to (2.3.17) can then be implemented upon initialization of the state vector and covariance matrix.

The performance and limitations of this filter are discussed in [9] and [1]. The stability of the EKF is known to depend highly on the accuracy of its initialization. High uncertainty about the initial target range estimate can cause excessive linearization errors, resulting in the collapse of the covariance matrix and divergence of the filter [1]. The use of Cartesian coordinates is reported in [2] to be another major cause of covariance matrix ill-conditioning.

2.5.2 The Extended Kalman Filter in Modified Spherical Coordinates

The use of Cartesian coordinates is known often to lead to poor performance in EKFs. This is due, in part, to the coupling of the observable and unobservable components of the state vector estimates which was found to precipitate ill-conditioning of the covariance matrix, and causing filter instability [2]. The extended Kalman filter in modified polar coordinates (MPEKF) was proposed in [2] to address this problem in 2D.

Similarly, the extended Kalman filter in modified spherical coordinates (MSCEKF) is based on a formulation of the 3D bearings-only tracking problem in a coordinate system that decouples the observable and unobservable states. It was first presented in [71] and appeared subsequently in [3, 52] with slight modifications. Instead of the usual Cartesian

position and velocity states, the MSCEKF uses a *relative* state vector $\mathbf{y}_k$ defined as

$$\mathbf{y}_k = \begin{bmatrix} \alpha_k & \dot{\alpha}_k & \theta_k & \omega_k & 1/r_k & \dot{r}_k/r_k \end{bmatrix}^{\mathrm{T}}, \tag{2.5.4}$$

in terms of the elevation angle α_k, the elevation angle-rate $\dot{\alpha}_k$, the azimuth angle θ_k, the inverse target range $1/r_k$, the inverse time-to-go $\dot{r}_k/r_k$ (where $\dot{r}_k$ is the range rate) and

$$\omega_k = \dot{\theta}_k \cos\alpha_k, \tag{2.5.5}$$

with $\dot{\theta}_k$ representing the azimuth angle-rate. For easy reconciliation with [71] and [52] (and given the involved nature of the derivations), we follow the convention used in those papers, where the azimuth angle θ is measured anti-clockwise from the x axis and the elevation angle α is measured downwards from the xy plane. Conversion to this convention from ours (as illustrated in figure (2.3)) is easily carried out by adding $\pi/2$ to the negative azimuth angle and inverting the sign of the elevation angle.

The state vector in MSC is related to its Cartesian version through the transformation

$$\mathbf{y}_k = g(\mathbf{x}_k), \tag{2.5.6}$$

where (dropping the time index k for notational clarity):

$$y_1 = \alpha = \tan^{-1}\left(-z/\sqrt{x^2+y^2}\right) \tag{2.5.7}$$

$$y_2 = \dot{\alpha} = \frac{-\dot{z}(x^2+y^2)+z(\dot{x}x+\dot{y}y)}{(x^2+y^2+z^2)\sqrt{x^2+y^2}} \tag{2.5.8}$$

$$y_3 = \theta = \tan^{-1}(y/x) \tag{2.5.9}$$

$$y_4 = \omega = \frac{\dot{y}x-\dot{x}y}{x^2+y^2}\cos\left(\tan^{-1}\left(-z/\sqrt{x^2+y^2}\right)\right) \tag{2.5.10}$$

$$y_5 = \frac{1}{r} = \left(\sqrt{x^2+y^2+z^2}\right)^{-1} \tag{2.5.11}$$

$$y_6 = \frac{\dot{r}}{r} = \frac{\dot{x}x+\dot{y}y+\dot{z}z}{x^2+y^2+z^2}. \tag{2.5.12}$$

The (now nonlinear) state equation is given in continuous-time by

$$\dot{\mathbf{y}} = f(\mathbf{y}, \mathbf{u}^A) + \tilde{\mathbf{v}}, \tag{2.5.13}$$

and more explicitly

$$\dot{y}_1 = \frac{d}{dt}\alpha = \dot{\alpha} + \tilde{v}_1 \tag{2.5.14}$$

$$\dot{y}_2 = \frac{d}{dt}\dot{\alpha} = -2\left(\frac{\dot{r}}{r}\right)\dot{\alpha} - \omega^2 \tan\alpha + \frac{u_3^A}{r} + \tilde{v}_2 \tag{2.5.15}$$

$$\dot{y}_3 = \frac{d}{dt}\theta = \frac{\omega}{\cos\alpha} + \tilde{v}_3 \tag{2.5.16}$$

$$\dot{y}_4 = \frac{d}{dt}\omega = \left(-2\left(\frac{\dot{r}}{r}\right) + \dot{\alpha}\tan\alpha\right)\omega - \frac{u_2^A}{r} + \tilde{v}_4 \tag{2.5.17}$$

$$\dot{y}_5 = \frac{d}{dt}\frac{1}{r} = -\frac{1}{r}\frac{\dot{r}}{r} + \tilde{v}_5 \tag{2.5.18}$$

$$\dot{y}_6 = \frac{d}{dt}\frac{\dot{r}}{r} = \dot{\alpha}^2 + \omega^2 - \left(\frac{\dot{r}}{r}\right)^2 - \frac{u_1^A}{r} + \tilde{v}_6, \tag{2.5.19}$$

where $\mathbf{u}^A$ is the acceleration vector of the sensor platform in 'antenna' coordinates defined as

$$\mathbf{u}^A = \mathbf{C}_I^A \mathbf{u}^I = \begin{bmatrix} \cos\alpha\cos\theta & \cos\alpha\sin\theta & -\sin\alpha \\ -\sin\theta & \cos\theta & 0 \\ \sin\alpha\cos\theta & \sin\alpha\sin\theta & \cos\alpha \end{bmatrix} \mathbf{u}^I, \tag{2.5.20}$$

and $\mathbf{u}^I$ is the acceleration vector of the sensor platform in 'inertial' coordinates. The noise term $\tilde{\mathbf{v}}$ regroups the unknown accelerations of the target .

The MSCEKF follows the same recursive formulation as the EKF, given by equations (2.3.11) to (2.3.12). The Jacobian $\hat{\mathbf{F}}_{k-1}$ is obtained by discretization of the continuous-time state equations (with the assumption that the change in relative geometry between the target and sensor platform is negligible during the sampling period T)

and linearization (see [71] for full derivation). It is expressed as:

$$\hat{\mathbf{F}} = \begin{bmatrix} 1 & T & 0 & 0 & 0 & 0 \\ -\frac{\omega^2}{\cos^2\alpha}T & 1-2\frac{\dot{r}}{r}T & 0 & -2T\omega\tan\alpha & u_3^A T & -2\dot{\alpha}T \\ \frac{\omega\sin\alpha}{\cos^2\alpha}T & 0 & 1 & \frac{1}{\cos\alpha}T & 0 & 0 \\ \frac{\omega\dot{\alpha}}{\cos^2\alpha}T & T\omega\tan\alpha & 0 & 1+\left(\dot{\alpha}\tan\alpha-2\frac{\dot{r}}{r}\right)T & -u_2^A T & -2\omega T \\ 0 & 0 & 0 & 0 & 1-\frac{\dot{r}}{r}T & -\frac{1}{r}T \\ 0 & 2\dot{\alpha}T & 0 & 2\omega T & -u_1^A T & 1-2\frac{\dot{r}}{r}T \end{bmatrix}. \tag{2.5.21}$$

The covariance in MSC of the discrete-time equivalent of the process noise term $\tilde{\mathbf{v}}$ is approximated as

$$\mathbf{Q}^{MSC} = \mathbf{G}_{\mathbf{x}}\mathbf{Q}\mathbf{G}_{\mathbf{x}}^{\mathrm{T}}, \tag{2.5.22}$$

where

$$\mathbf{G}_{\mathbf{x}} = \nabla_{\mathbf{x}} g(\mathbf{x})\big|_{\mathbf{x}=\hat{\mathbf{x}}_{k-1|k-1}} \tag{2.5.23}$$

is the Jacobian of the Cartesian to MSC transformation evaluated at the current state estimate (in Cartesian coordinates). The measurement equation in MSC is given by

$$\mathbf{z}_k = \begin{bmatrix} \theta_k \\ \alpha_k \end{bmatrix} = \mathbf{H}(\mathbf{y}_k) + \mathbf{w}_k, \tag{2.5.24}$$

where $\mathbf{w}_k$ is white Gaussian noise and the measurement matrix

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.5.25}$$

selects the target bearings from the state vector. We note that the formulation of the tracking problem in MSC effectively translates linear state equations into nonlinear ones while turning the nonlinear measurement equations into linear ones.

The main advantages of the MSCEKF, over the EKF in Cartesian coordinates, are increased stability and more accurate range estimation with moderate maneuvering of the

sensor platform. Target maneuvers, however, are a source of difficulty and special care needs to be taken to deal with simultaneous target and sensor platform maneuvers [3]. Recently, a new formulation of this filter was proposed in [56], where the continuous system dynamics are integrated numerically to provide a more accurate representation of the predicted state mean and covariance. Although less accurate than the Cartesian EKF for low levels of measurement noise, the proposed 'continuous-discrete' EKF in MSC and its variant in log spherical coordinates were reported to outperform it for low measurement accuracy.

2.5.3 The Range-parameterized Extended Kalman Filter

The range-parameterized extended Kalman Filter (RPEKF) [60] is a modification of the standard EKF, aiming to overcome the effects of large initial range uncertainty. The performance of a tracking filter is determined to a large extent by the *coefficient of variation* C_r of the range estimate, defined as $\hat{\sigma}_r/\hat{r}$ (where $\hat{r}$ is the estimated range and $\hat{\sigma}_r$ is its standard deviation). A low C_r contributes to filter stability by ensuring that a change in measured bearings results in only a small change in the range estimate. Therefore, initialization should ideally be carried out with as low a value of C_r as possible. Given a large initial range uncertainty, the range interval of interest (r_{min}, r_{max}) can be subdivided into N_f smaller segments in which separate and independent EKFs are initialized. Imposing the same value of C_r for each segment will ensure comparable performance for all filters.

Assuming that the range errors are uniformly distributed within each segment, the required subdivision can be achieved by setting the boundaries to follow a geometric progression with common ratio ρ such that

$$r_{max} \;=\; r_{min}\,\rho^{N_f}. \tag{2.5.26}$$

The common coefficient of variation will then be given by

$$C_r \;=\; \frac{\sigma_{r_i}}{r_i} \;=\; \frac{2(\rho-1)}{\sqrt{12}(\rho+1)}, \tag{2.5.27}$$

where r_i and σ_{r_i} are the mean of the subinterval i and its standard deviation. Then, the EKF corresponding to the i'th subinterval can be initialized with a Gaussian prior on the range, with mean and standard deviation

$$\hat{r}_i = r_{min} \frac{(\rho^{i-1} + \rho^i)}{2} \tag{2.5.28}$$

$$\hat{\sigma}_{r_i} = r_{min} \frac{(\rho^i - \rho^{i-1})}{\sqrt{12}}. \tag{2.5.29}$$

Although each EKF operates independently using its own estimates, the obtained Gaussian mixture is combined after each time step to a single state estimate. This is done by moment-matching, according to the weight w_k^i associated with each EKF. Denoting the initial range subinterval hypotheses R_i (with $i = 1, \ldots, N_f$), these weights are computed, using Bayes' rule, as

$$\begin{aligned} w_k^i &= p(R_i|\mathbf{Z}_k) \\ &= \frac{p(\mathbf{z}_k|R_i, \mathbf{Z}_{k-1}) p(R_i|\mathbf{Z}_{k-1})}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})} \\ &= \frac{p(\mathbf{z}_k|R_i, \mathbf{Z}_{k-1}) w_{k-1}^i}{\sum_{j=1}^{N_f} p(\mathbf{z}_k|R_j, \mathbf{Z}_{k-1}) w_{k-1}^j}, \end{aligned} \tag{2.5.30}$$

where $p(\mathbf{z}_k|R_i, \mathbf{Z}_{k-1})$ is the measurement likelihood conditioned on the hypothesis that the target originated from the i'th range subinterval. It can be computed as:

$$p(\mathbf{z}_k|R_i, \mathbf{Z}_{k-1}) \sim \mathcal{N}\left(\mathbf{z}_k; \hat{\mathbf{z}}_{k|k-1}^i, \mathbf{S}_k^i\right), \tag{2.5.31}$$

with $\hat{\mathbf{z}}_{k|k-1}^i$ and $\mathbf{S}_k^i$ denoting the predicted measurement and its associated covariance matrix, conditioned on the i'th initial range hypothesis. The denominator of (2.5.30) acts as a normalizing constant ensuring that all the weights add up to unity. To reflect the (near) absence of initial range knowledge, each EKF will typically be initialized with an equal weighting of $1/N_f$.

Reduction of the weighted Gaussian mixture to a single Gaussian is a straightfor-

ward operation. The mean and covariance of the resulting density are given by

$$\hat{\mathbf{x}}_{k|k} = \sum_{i=1}^{N_f} w_k^i \, \hat{\mathbf{x}}_{k|k}^i \tag{2.5.32}$$

$$\mathbf{P}_{k|k} = \sum_{i=1}^{N_f} w_k^i \left[\mathbf{P}_{k|k}^i + (\mathbf{x}_{k|k} - \mathbf{x}_{k|k}^i)(\mathbf{x}_{k|k} - \mathbf{x}_{k|k}^i)^{\mathrm{T}} \right]. \tag{2.5.33}$$

The increased accuracy of the RPEKF over the standard EKF is reported in [60] for scenarios involving very vague initial range knowledge. However, this gain comes at the expense of a decrease in computational efficiency by a factor of N_f. Yet, in many situations, most filter weights will become negligible after maneuver of the sensor platform, indicating that the range ambiguity has been resolved. The corresponding filters can then be 'pruned' with little effect on overall accuracy, helping to reduce computational overhead.

Chapter 3

Bearings only tracking in 3D

The tracking literature contains an abundance of algorithms for bearings-only tracking in 2 dimensions. These are of relevance in many applications where the problem is intrinsically 2-dimensional, such as surface vessel or ground vehicle tracking. Alternatively, 2-dimensional tracking might be a preferable choice (for reasons of computational simplicity) in situations where vertical motion does not need to be estimated jointly with planar motion. This is the case in civilian air traffic control where aircraft maneuvers are mostly restricted to the horizontal plane, and vertical out of plane maneuvers rarely happening simultaneously. In such 'benign' scenarios, decoupled filters for the independent estimation of the xy and z components of the target dynamics are appropriate. Another use of 2D filters can arise from the lack of elevation information, typically absent in passive-sonar measurements [53].

There exist situations, however, that are characterized by strongly coupled dynamics between all 3 coordinates. For these problems, adequate tracking performance can only be achieved by the use of fully coupled 3D estimation algorithms that concurrently evaluate the target motion in all coordinates. Applications requiring such 3D estimation schemes include the tracking of agile aircraft and missiles which exhibit a high degree of maneuverability. Most existing 2D filters (including the EKF, UKF and PF, in cartesian coordinates) can be trivially extended to 3D, while others (such as the MPEKF) can be generalized at the cost of higher complexity (MSCEKF).

We propose in this chapter a novel extension to an existing bearings-only tracking algorithm known as the shifted Rayleigh filter (SRF), appearing in [18–20], enabling its implementation in complex 3D tracking situations. Although readily applicable to both 2D and 3D bearings measurements in its original form, an extension to cope with the

challenges of data association and multiple target dynamic models in 3-dimensional space is yet to be derived. The central element in this extension is the probability density of the predicted bearings, given past measurements, which can be used as a likelihood function, allowing for the proper weighting of multiple measurements in a cluttered environment or the discrimination between different possible target dynamic modes. Owing to the particular formulation of the SRF measurement model, the probability density of the measurement has a non-trivial form, and has so far only been presented for measurements in 2D space ([19]).

The organization of this chapter is as follows. An outline of the (previously published) SRF is provided, followed by the mathematical derivation of the filter, with particular focus on 3-dimensional tracking. The probability density function for 3D measurements, which is the principal contribution of the chapter, is then derived and used to justify the particular measurement model inherent to the SRF. Making use of this measurement density, extensions of the SRF algorithm to handle data association and multiple target dynamic modes are then presented. Finally, the problem of optimal smoothing is touched upon and a shifted Rayleigh adaptation of the well known Rauch-Tung-Striebel smoother is presented.

3.1 The Shifted Rayleigh Filter

We recall that the shifted Rayleigh filter [19, 20] is a moment-matching bearings-only tracking filter. Unlike other moment-matching methods, it is derived specifically to take into account the particular nonlinearities of the bearings measurement equation. Whereas competing moment-matching estimators overcome this difficulty by means of linearization (either analytical, as with the EKF, or statistical as with the UKF), the SRF has the distinctive feature of computing the exact mean and covariance of the non-Gaussian posterior density, given a Gaussian prior. This is made possible through a reformulation of the measurement equation in terms of an 'augmented' measurement of the target displacement vector. At the end of each iteration, the filtered density is approximated by a Gaussian whose first two moments are those of the exact non-Gaussian posterior. We stress that this moment-matching step is the only source of approximation in the SRF

recursion.

It is shown that the modification of the measurement equation, which enables computation of the exact moments, results, in 2D space, in a bearing density which is practically indistinguishable from the Gaussian density of the conventional bearings measurement model. In 3D space, on the other hand, the resulting density of azimuth and elevation angle measurements is shown to coincide with a measurement model that describes the uncertainties better than the standard model does.

Consider an n-dimensional state vector $\mathbf{x}_k$ composed of the kinematic attributes of interest of the target and possibly of the sensor platform (all in Cartesian coordinates), in d-dimensional Euclidian space. Let $\mathbf{d}_k$ be the displacement d-vector of the target, relative to the position of the sensor platform. Finally, let $\mathbf{b}_k$ denote the 'direction cosines' bearing d-vector in the direction of the noisy angular measurement of the target. These quantities are expressed through the equations

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{u}_{k-1}^s + \mathbf{v}_k \tag{3.1.1}$$

$$\mathbf{d}_k = \mathbf{H}\mathbf{x}_k + \mathbf{u}_k^m \tag{3.1.2}$$

$$\mathbf{b}_k = \Pi(\mathbf{d}_k + \mathbf{w}_k), \tag{3.1.3}$$

where $\mathbf{F}_k$ is the possibly time-dependent $n \times n$ system matrix, while $\mathbf{H}$ is the $d \times n$ time-invariant measurement matrix that picks out the cartesian position coordinates of the target. The deterministic control input n-vector $\mathbf{u}_{k-1}^s$ is included for generality. The deterministic measurement input d-vector $\mathbf{u}_k^m$ can be included to represent the position of the sensor platform. If instead, its motion is integrated into the state vector $\mathbf{x}_k$, the measurement matrix $\mathbf{H}$ needs to be appropriately modified to extract the target position relative to that of the sensor platform. The system noise n-vector $\mathbf{v}_{k-1}$ and measurement noise d-vector $\mathbf{w}_k$ are independent zero-mean Gaussian with covariances $\mathbf{Q}_{k-1}^s$ and $\mathbf{Q}_k^m$. Finally, Π denotes the projection of d-dimensional cartesian space onto the d-dimensional unit sphere. Its argument $\mathbf{d}_k + \mathbf{w}_k$, a noise-corrupted version of the displacement, represents an 'augmented' measurement $\mathbf{y}_k$. The bearings unit vector is thus given (in consistence

with the convention of section 2.4.5) by

$$\mathbf{b}_k \;=\; \frac{\mathbf{y}_k}{||\mathbf{y}_k||} = \begin{cases} \big[\sin(\theta_k)\;\cos(\theta_k)\big]^{\mathrm{T}} & \text{for } d=2 \\ \big[\sin(\theta_k)\cos(\alpha_k)\;\cos(\theta_k)\cos(\alpha_k)\;\sin(\alpha_k)\big]^{\mathrm{T}} & \text{for } d=3. \end{cases} \tag{3.1.4}$$

The main characteristic of this setup is the addition of noise to the displacement *before* the nonlinearity introduced by Π. This is at odds with the conventional bearings measurement model where zero-mean Gaussian noise is added to the bearings of the displacement (i.e. *after* the nonlinearity). As will become clear, however, the unconventional model is used in order to simplify the calculations involved in the construction of the tracking algorithm.

Using the formulation of (3.1.1) to (3.1.3) and letting $\hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$ be the mean and covariance of the Gaussian posterior of the state vector at time $k-1$, one cycle of the SRF takes the form of the following prediction and update equations:

Prediction Step:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} + \mathbf{u}^s_{k-1} \tag{3.1.5}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k\mathbf{P}_{k-1|k-1}\mathbf{F}_k^{\mathrm{T}} + \mathbf{Q}^s_{k-1} \tag{3.1.6}$$

$$\mathbf{V}_k = \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} + \mathbf{Q}^m_k \tag{3.1.7}$$

Correction Step:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\mathbf{V}_k^{-1} \tag{3.1.8}$$

$$z_k = (\mathbf{b}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{b}_k)^{-\frac{1}{2}}\mathbf{b}_k^{\mathrm{T}}\mathbf{V}_k^{-1}(\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} + \mathbf{u}^m_k) \tag{3.1.9}$$

$$\gamma_k = (\mathbf{b}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{b}_k)^{-\frac{1}{2}}\rho_d(z_t) \tag{3.1.10}$$

$$\delta_k = (\mathbf{b}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{b}_k)^{-1}(d + z_k\,\rho(z_k) - \rho_d^2(z_k)) \tag{3.1.11}$$

$$\hat{\mathbf{x}}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\mathbf{u}^m_k + \gamma_k\mathbf{K}_k\mathbf{b}_k \tag{3.1.12}$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1} + \delta_k\mathbf{K}_k\mathbf{b}_k\mathbf{b}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}. \tag{3.1.13}$$

$\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ can be recognized as the predicted mean and covariance of the state, while $\mathbf{V}_k$ is the predicted covariance of the 'augmented' measurement $\mathbf{y}_k$ (i.e. the noisy displacement), all at time k given measurements up to time $k-1$. $\mathbf{K}_k$ is the Kalman gain,

and γ_k and δ_k are the conditional mean and variance of $||\mathbf{y}_k||$. Finally, the scalar $\rho_d(z_k)$ is the mean of a 'shifted Rayleigh' variable, evaluated as

$$\rho_d(z) \;=\; \frac{\int_0^\infty s^d e^{-\frac{1}{2}(s-z)^2}\, ds}{\int_0^\infty s^{d-1} e^{-\frac{1}{2}(s-z)^2}\, ds} \tag{3.1.14}$$

3.1.1 Derivation of Shifted Rayleigh Filter Equations

Using the formulation presented in (3.1.1) to (3.1.3) and the augmented measurement $\mathbf{y}_k$, previously defined as

$$\mathbf{y}_k \;=\; \mathbf{H}\mathbf{x}_k + \mathbf{u}_k^m + \mathbf{w}_k, \tag{3.1.15}$$

the MMSE estimate of the state vector $\mathbf{x}_k$, assuming that access to the augmented measurement $\mathbf{y}_k$ is available and that $(\mathbf{x}_k, \mathbf{y}_k)$ are jointly Gaussian, is given by the standard Kalman filter equations as

$$\hat{\mathbf{x}}_k \;=\; (\mathbf{I} - \mathbf{K}_k\mathbf{H})\hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\mathbf{u}_k^m + \mathbf{K}_k\mathbf{y}_k + \xi_k, \tag{3.1.16}$$

where the term

$$\xi_k \;\sim\; \mathcal{N}(\mathbf{0}, (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1}) \tag{3.1.17}$$

represents the error of the state estimate.

The predicted augmented measurement at time k, conditioned on measurements up to time $k-1$ is

$$\mathbf{y}_k \;\sim\; \mathcal{N}(\hat{\mathbf{y}}_{k|k-1}, \mathbf{V}_k), \tag{3.1.18}$$

where

$$\hat{\mathbf{y}}_{k|k-1} \;=\; \mathbf{H}\hat{\mathbf{x}}_{k|k-1} + \mathbf{u}_k^m \quad \text{and} \quad \mathbf{V}_k \;=\; \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^{\mathrm{T}} + \mathbf{Q}_k^m. \tag{3.1.19}$$

Now, let r_k be the range of the predicted augmented measurement $\mathbf{y}_k$ such that

$$\mathbf{y}_k \;=\; r_k\mathbf{b}_k, \tag{3.1.20}$$

and define $\bar{r}_k$ and $\bar{\mathbf{b}}_k$ as the range and bearing vector of the predicted transformed augmented measurement $\bar{\mathbf{y}}_k = \mathbf{V}_k^{-\frac{1}{2}}\mathbf{y}_k$, which has a $\mathcal{N}(\mathbf{V}_k^{-\frac{1}{2}}\hat{\mathbf{y}}_{k|k-1}, \mathbf{I}_{d\times d})$ distribution, so that

$$\bar{r}_k \;=\; \left|\left|\mathbf{V}_k^{-\frac{1}{2}}\mathbf{y}_k\right|\right| \quad \text{and} \quad \bar{\mathbf{b}}_k \;=\; \frac{\mathbf{V}_k^{-\frac{1}{2}}\mathbf{y}_k}{\left|\left|\mathbf{V}_k^{-\frac{1}{2}}\mathbf{y}_k\right|\right|}. \tag{3.1.21}$$

The augmented measurement can then be expressed as

$$\mathbf{y}_k \;=\; \bar{r}_k\mathbf{V}_k^{\frac{1}{2}}\bar{\mathbf{b}}_k, \tag{3.1.22}$$

and substituted back into (3.1.16) to yield

$$\mathbf{x}_k \;=\; (\mathbf{I}-\mathbf{K}_k\mathbf{H})\hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\mathbf{u}_k^m + \bar{r}_k\mathbf{K}_k\mathbf{V}_k^{\frac{1}{2}}\bar{\mathbf{b}}_k + \xi_k. \tag{3.1.23}$$

The above expression assumes direct access to the augmented measurement $\mathbf{y}_k$. This is, however, not the case as only its bearings $\mathbf{b}_k$ are measured. As a result, we need to evaluated the moments of $\mathbf{x}_k$ conditioned on $\mathbf{b}_k$. Since (3.1.23) is defined in terms of the bearings vector $\bar{\mathbf{b}}_k$ of the transformed augmented measurement, we first show that there is a one-to-one (bijective) relationship between $\bar{\mathbf{b}}_k$ and $\mathbf{b}_k$, so that conditioning on one is equivalent to conditioning on the other. This can be shown as follows. Using (3.1.20) together with the definition of $\bar{\mathbf{b}}_k$, we get

$$\begin{aligned}
\bar{\mathbf{b}}_k \;&=\; \frac{r_k\mathbf{V}_k^{-\frac{1}{2}}\mathbf{b}_k}{\left|\left|r_k\mathbf{V}_k^{-\frac{1}{2}}\mathbf{b}_k\right|\right|} \\
&=\; \left[(\mathbf{V}_k^{-\frac{1}{2}}\mathbf{b}_k)^{\mathrm{T}}(\mathbf{V}_k^{-\frac{1}{2}}\mathbf{b}_k)\right]^{-\frac{1}{2}}\mathbf{V}_k^{-\frac{1}{2}}\mathbf{b}_k \\
&=\; (\mathbf{b}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{b}_k)^{-\frac{1}{2}}\mathbf{V}_k^{-\frac{1}{2}}\mathbf{b}_k,
\end{aligned} \tag{3.1.24}$$

while, on the other hand, starting from the definition of $\mathbf{b}_k$ we have

$$\begin{aligned}
\mathbf{b}_k \;&=\; \frac{\mathbf{y}_k}{||\mathbf{y}_k||} \\
&=\; (\bar{r}_k^2\,\bar{\mathbf{b}}_k^{\mathrm{T}}\,\mathbf{V}_k^{\frac{1}{2}}\,\mathbf{V}_k^{\frac{1}{2}}\,\bar{\mathbf{b}}_k)^{-\frac{1}{2}}\,\bar{r}_k\,\mathbf{V}_k^{\frac{1}{2}}\,\bar{\mathbf{b}}_k \\
&=\; (\bar{\mathbf{b}}_k^{\mathrm{T}}\mathbf{V}_k\bar{\mathbf{b}}_k)^{-\frac{1}{2}}\mathbf{V}_k^{\frac{1}{2}}\bar{\mathbf{b}}_k.
\end{aligned} \tag{3.1.25}$$

Comparison of (3.1.24) and (3.1.25) confirms the equivalence of conditioning on $\bar{\mathbf{b}}_k$ and $\mathbf{b}_k$. Therefore, the mean of $\mathbf{x}_k$ conditioned on $\mathbf{b}_k$, can be expressed (using (3.1.23) and assuming independence between ξ_k and $\mathbf{b}_k$) as

$$\begin{aligned} E[\mathbf{x}_k|\mathbf{b}_k] &= (\mathbf{I}-\mathbf{K}_k\mathbf{H})\hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\mathbf{u}_k^m + E[\bar{r}_k|\bar{\mathbf{b}}_k]\mathbf{K}_k\mathbf{V}_k^{\frac{1}{2}}\bar{\mathbf{b}}_k \\ &= (\mathbf{I}-\mathbf{K}_k\mathbf{H})\hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\mathbf{u}_k^m + (\mathbf{b}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{b}_k)^{-\frac{1}{2}}E[\bar{r}_k|\bar{\mathbf{b}}_k]\mathbf{K}_k\mathbf{b}_k, \end{aligned} \tag{3.1.26}$$

where the relation (3.1.24) was used to get from the first to the second line. Similarly, defining the error term

$$\begin{aligned} \tilde{\mathbf{x}}_k &= \mathbf{x}_k - E[\mathbf{x}_k|\mathbf{b}_k] \\ &= \xi_k + (\bar{r}_k - E\big[\bar{r}_k|\bar{\mathbf{b}}_k\big])\mathbf{K}_k\mathbf{V}_k^{\frac{1}{2}}\bar{\mathbf{b}}_k, \end{aligned} \tag{3.1.27}$$

the conditional covariance, given the bearing vector, can be written as

$$\begin{aligned} \mathrm{cov}\big[\mathbf{x}_k|\mathbf{b}_k\big] &= E\big[\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^{\mathrm{T}} \mid \mathbf{b}_k\big] \\ &= \mathrm{cov}\big[\xi_k\big] + E\big[(\bar{r}_k - E\big[\bar{r}_k|\bar{\mathbf{b}}_k\big])(\bar{r}_k - E\big[\bar{r}_k|\bar{\mathbf{b}}_k\big])^{\mathrm{T}} \mid \mathbf{b}_k\big]\mathbf{K}_k\mathbf{V}_k^{\frac{1}{2}}\bar{\mathbf{b}}_k\bar{\mathbf{b}}_k^{\mathrm{T}}\mathbf{V}_k^{\frac{1}{2}}\mathbf{K}_k^{\mathrm{T}} \\ &= (\mathbf{I}-\mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1} + \mathrm{var}\big[\bar{r}_k|\bar{\mathbf{b}}_k\big]\mathbf{K}_k\mathbf{V}_k^{\frac{1}{2}}\bar{\mathbf{b}}_k\bar{\mathbf{b}}_k^{\mathrm{T}}\mathbf{V}_k^{\frac{1}{2}}\mathbf{K}_k^{\mathrm{T}} \\ &= (\mathbf{I}-\mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1} + (\mathbf{b}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{b}_k)^{-1}\mathrm{var}\big[\bar{r}_k|\bar{\mathbf{b}}_k\big]\mathbf{K}_k\mathbf{b}_k\mathbf{b}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}. \end{aligned} \tag{3.1.28}$$

Equations (3.1.26) and (3.1.28) require computation of the terms $E[\bar{r}_k|\bar{\mathbf{b}}_k]$ and $\mathrm{var}\big[\bar{r}_k|\bar{\mathbf{b}}_k\big]$, representing the mean and variance of the range of the transformed augmented measurement, conditioned on the measured bearings. To this end, we first derive the joint probability density function of the range and bearings vector of the transformed augmented measurement $\bar{\mathbf{y}}_k$.

The probability density function of the transformed augmented measurement (3.1.18) can be expressed as

$$p_{\bar{\mathbf{y}}_k}(\bar{\mathbf{y}}_k) = \frac{1}{(2\pi)^{\frac{d}{2}}}e^{-\frac{1}{2}(\bar{\mathbf{y}}_k-\mathbf{V}_k^{-\frac{1}{2}}\hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}}(\bar{\mathbf{y}}_k-\mathbf{V}_k^{-\frac{1}{2}}\hat{\mathbf{y}}_{k|k-1})}. \tag{3.1.29}$$

Conversion from polar to Cartesian ($d = 2$) and spherical to Cartesian ($d = 3$) coordinates can be obtained through the transformations g_p and g_s, respectively, such that:

$$\begin{aligned}\bar{\mathbf{y}}_k &= \begin{cases} g_p(\bar{r}_k, \bar{\theta}_k) & \text{for } d = 2 \\ g_s(\bar{r}_k, \bar{\theta}_k, \bar{\alpha}_k) & \text{for } d = 3 \end{cases} \\ &= \begin{cases} \left[\bar{r}_k \sin(\bar{\theta}_k) \;\; \bar{r}_k \cos(\bar{\theta}_k)\right]^{\mathrm{T}} & \text{for } d = 2 \\ \left[\bar{r}_k \sin(\bar{\theta}_k)\cos(\bar{\alpha}_k) \;\; \bar{r}_k \cos(\bar{\theta}_k)\cos(\bar{\alpha}_k) \;\; \bar{r}_k \sin(\bar{\alpha}_k)\right]^{\mathrm{T}} & \text{for } d = 3, \end{cases} \end{aligned} \tag{3.1.30}$$

where $\bar{\theta}_k$ and $\bar{\alpha}_k$ are the azimuth and elevation angles of the bearings vector of the transformed augmented measurement $\bar{\mathbf{y}}_k$. It follows [1] that the distribution of $\bar{\mathbf{y}}_k$ in either polar or spherical coordinates is:

$$p_{\bar{r}_k, \bar{\mathbf{b}}_k}(\bar{r}_k, \bar{\mathbf{b}}_k) = \begin{cases} \left|\nabla g_p(\bar{r}_k, \bar{\theta}_k)\right| \; p_{\bar{\mathbf{y}}_k}\left(g_p(\bar{r}_k, \bar{\theta}_k)\right) & \text{for } d = 2 \\ \left|\nabla g_s(\bar{r}_k, \bar{\theta}_k, \bar{\alpha}_k)\right| \; p_{\bar{\mathbf{y}}_k}\left(g_s(\bar{r}_k, \bar{\theta}_k, \bar{\alpha}_k)\right) & \text{for } d = 3 \end{cases} \tag{3.1.31}$$

where $\nabla g_p(\bar{r}_k, \bar{\theta}_k)$ and $\nabla g_s(\bar{r}_k, \bar{\theta}_k, \bar{\alpha}_k)$ are the determinants of the Jacobians of the transformations from polar to cartesian ($d = 2$) and spherical to cartesian ($d = 3$) coordinates, evaluated as

$$\begin{aligned}\nabla g_p(\bar{r}_k, \bar{\theta}_k) &= \begin{vmatrix} \frac{\partial \bar{y}_k^1}{\partial \bar{r}_k} & \frac{\partial \bar{y}_k^1}{\partial \bar{\theta}_k} \\ \frac{\partial \bar{y}_k^2}{\partial \bar{r}_k} & \frac{\partial \bar{y}_k^2}{\partial \bar{\theta}_k} \end{vmatrix} \\ &= \begin{vmatrix} \sin(\bar{\theta}_k) & \bar{r}_k \cos(\bar{\theta}_k) \\ \cos(\bar{\theta}_k) & -\bar{r}_k \sin(\bar{\theta}_k) \end{vmatrix} \\ &= \bar{r}_k \end{aligned} \tag{3.1.32}$$

and

$$\nabla g_s(\bar{r}_k, \bar{\theta}_k, \bar{\alpha}_k) = \begin{vmatrix} \frac{\partial \bar{y}_k^1}{\partial \bar{r}_k} & \frac{\partial \bar{y}_k^1}{\partial \bar{\theta}_k} & \frac{\partial \bar{y}_k^1}{\partial \bar{\alpha}_k} \\ \frac{\partial \bar{y}_k^2}{\partial \bar{r}_k} & \frac{\partial \bar{y}_k^2}{\partial \bar{\theta}_k} & \frac{\partial \bar{y}_k^2}{\partial \bar{\alpha}_k} \\ \frac{\partial \bar{y}_k^3}{\partial \bar{r}_k} & \frac{\partial \bar{y}_k^3}{\partial \bar{\theta}_k} & \frac{\partial \bar{y}_k^3}{\partial \bar{\alpha}_k} \end{vmatrix}$$

[1] the (reasonable) assumption is made here that in practice the ratio of the mean of the range of $\bar{\mathbf{y}}_k$ to its standard deviation is low enough for issues of ambiguity in the many-to-one mapping not to arise.

$$
\begin{aligned}
&= \begin{vmatrix} \sin(\bar{\theta}_k)\cos(\bar{\alpha}_k) & \bar{r}_k\cos(\bar{\theta}_k)\cos(\bar{\alpha}_k) & -\bar{r}_k\sin(\bar{\theta}_k)\sin(\bar{\alpha}_k) \\ \cos(\bar{\theta}_k)\cos(\bar{\alpha}_k) & -\bar{r}_k\sin(\bar{\theta}_k)\cos(\bar{\alpha}_k) & -\bar{r}_k\cos(\bar{\theta}_k)\sin(\bar{\alpha}_k) \\ \sin(\bar{\alpha}_k) & 0 & \bar{r}_k\cos(\bar{\alpha}_k) \end{vmatrix} \\
&= -\bar{r}_k^2\cos(\bar{\alpha}_k). \qquad (3.1.33)
\end{aligned}
$$

The densities $p_{\bar{\mathbf{y}}_k}\left(g_p(\bar{r}_k,\bar{\theta}_k)\right)$ and $p_{\bar{\mathbf{y}}_k}\left(g_s(\bar{r}_k,\bar{\theta}_k,\bar{\alpha}_k)\right)$ in (3.1.31) can be expressed under unified notation, in terms of $\bar{r}_k$ and $\bar{\mathbf{b}}_k$ as:

$$
\begin{aligned}
p_{\bar{\mathbf{y}}_k}\left(\bar{r}_k\bar{\mathbf{b}}_k\right) &= \frac{1}{(2\pi)^{\frac{d}{2}}}\, e^{-\frac{1}{2}(\bar{r}_k\bar{\mathbf{b}}_k-\mathbf{V}_k^{-\frac{1}{2}}\hat{\mathbf{y}}_{k|k-1})^{\mathrm{T}}(\bar{r}_k\bar{\mathbf{b}}_k-\mathbf{V}_k^{-\frac{1}{2}}\hat{\mathbf{y}}_{k|k-1})} \\
&= \frac{1}{(2\pi)^{\frac{d}{2}}}\, e^{-\frac{1}{2}(\bar{r}_k^2-2\bar{r}_k\bar{\mathbf{b}}_k^{\mathrm{T}}\mathbf{V}_k^{-\frac{1}{2}}\hat{\mathbf{y}}_{k|k-1}+\hat{\mathbf{y}}_{k|k-1}^{\mathrm{T}}\mathbf{V}_k^{-1}\hat{\mathbf{y}}_{k|k-1}^{\mathrm{T}})} \qquad (3.1.34)
\end{aligned}
$$

and setting $z_k = \bar{\mathbf{b}}_k^{\mathrm{T}}\mathbf{V}_k^{-\frac{1}{2}}\hat{\mathbf{y}}_{k|k-1}$, we get:

$$
\begin{aligned}
p_{\bar{\mathbf{y}}_k}\left(\bar{r}_k\bar{\mathbf{b}}_k\right) &= \frac{1}{(2\pi)^{\frac{d}{2}}}\, e^{-\frac{1}{2}(\bar{r}_k^2-2\bar{r}_k z_k+\hat{\mathbf{y}}_{k|k-1}^{\mathrm{T}}\mathbf{V}_k^{-1}\hat{\mathbf{y}}_{k|k-1})} \\
&= \frac{1}{(2\pi)^{\frac{d}{2}}}\, e^{-\frac{1}{2}(\bar{r}_k-z_k)^2} e^{-\frac{1}{2}(\hat{\mathbf{y}}_{k|k-1}^{\mathrm{T}}\mathbf{V}_k^{-1}\hat{\mathbf{y}}_{k|k-1}-z_k^2)}. \qquad (3.1.35)
\end{aligned}
$$

It follows that the joint probability density of the range and bearings vector of the transformed augmented measurement $\bar{\mathbf{y}}_k$ is given as:

$$
p_{\bar{r}_k,\bar{\mathbf{b}}_k}(\bar{r}_k,\bar{\mathbf{b}}_k) = \begin{cases} \frac{\bar{r}_k}{2\pi}\, e^{-\frac{1}{2}(\bar{r}_k-z_k)^2} e^{-\frac{1}{2}(\hat{\mathbf{y}}_{k|k-1}^{\mathrm{T}}\mathbf{V}_k^{-1}\hat{\mathbf{y}}_{k|k-1}-z_k^2)} & \text{for } d=2 \\ \frac{\bar{r}_k^2\cos(\alpha_k)}{(2\pi)^{\frac{3}{2}}}\, e^{-\frac{1}{2}(\bar{r}_k-z_k)^2} e^{-\frac{1}{2}(\hat{\mathbf{y}}_{k|k-1}^{\mathrm{T}}\mathbf{V}_k^{-1}\hat{\mathbf{y}}_{k|k-1}-z_k^2)} & \text{for } d=3. \end{cases} \qquad (3.1.36)
$$

For both cases, the conditional density of $\bar{r}_k$ given $\bar{\mathbf{b}}_k$ can be expressed, using the rule of conditional probability, as:

$$
\begin{aligned}
p_{\bar{r}_k|\bar{\mathbf{b}}_k}(\bar{r}_k|\bar{\mathbf{b}}_k) &= \frac{p_{\bar{r}_k,\bar{\mathbf{b}}_k}(\bar{r}_k,\bar{\mathbf{b}}_k)}{p_{\bar{\mathbf{b}}_k}(\bar{\mathbf{b}}_k)} \\
&= \frac{\bar{r}_k^{\,d-1}\, e^{-\frac{1}{2}(\bar{r}_k-z_k)^2}}{\int_0^\infty \bar{r}_k^{\,d-1} e^{-\frac{1}{2}(\bar{r}_k-z_k)^2}\, d\bar{r}_k} \\
&= \frac{\bar{r}_k^{\,d-1}\, e^{-\frac{1}{2}(\bar{r}_k-z_k)^2}}{K(z_k)}. \qquad (3.1.37)
\end{aligned}
$$

Evaluation of the first moment of this density yields:

For $d = 2$

$$E\left[\bar{r}_k|\bar{\mathbf{b}}_k\right] \;=\; \rho_2(z_k) \;=\frac{\int_0^\infty \bar{r}_k^{\,2}\, e^{-\frac{1}{2}(\bar{r}_k - z_k)^2} d\bar{r}_k}{\int_0^\infty \bar{r}_k\, e^{-\frac{1}{2}(\bar{r}_k - z_k)^2}\, d\bar{r}_k}, \tag{3.1.38}$$

where carrying out the substitution $u = \bar{r}_k - z_k$, the numerator can be expressed as

$$\begin{aligned}\int_0^\infty \bar{r}_k^{\,2}\, e^{-\frac{1}{2}(\bar{r}_k - z_k)^2} d\bar{r}_k \;&=\; \int_{-z_k}^\infty (u + z_k)^2\, e^{-\frac{1}{2}u^2}\, du \\ &= \int_{-z_k}^\infty u^2 e^{-\frac{1}{2}u^2}\, du + 2z_k \int_{-z_k}^\infty u\, e^{-\frac{1}{2}u^2}\, du + z_k^2 \int_{-z_k}^\infty e^{-\frac{1}{2}u^2}\, du,\end{aligned} \tag{3.1.39}$$

and the denominator as

$$\begin{aligned}\int_0^\infty \bar{r}_k\, e^{-\frac{1}{2}(\bar{r}_k - z_k)^2} d\bar{r}_k \;&=\; \int_{-z_k}^\infty (u + z_k)\, e^{-\frac{1}{2}u^2}\, du \\ &= \int_{-z_k}^\infty u e^{-\frac{1}{2}u^2}\, du + z_k \int_{-z_k}^\infty e^{-\frac{1}{2}u^2}\, du.\end{aligned} \tag{3.1.40}$$

For $d = 3$

$$E\left[\bar{r}_k|\bar{\mathbf{b}}_k\right] \;=\; \rho_3(z_k) \;=\frac{\int_0^\infty \bar{r}_k^{\,3}\, e^{-\frac{1}{2}(\bar{r}_k - z_k)^2} d\bar{r}_k}{\int_0^\infty \bar{r}_k^{\,2} e^{-\frac{1}{2}(\bar{r}_k - z_k)^2}\, d\bar{r}_k}, \tag{3.1.41}$$

where using the same change of variable, the numerator can be expressed as

$$\begin{aligned}\int_0^\infty \bar{r}_k^{\,3}\, e^{-\frac{1}{2}(\bar{r}_k - z_k)^2} d\bar{r}_k \;&=\; \int_{-z_k}^\infty (u + z_k)^3\, e^{-\frac{1}{2}u^2}\, du \\ &= \int_{-z_k}^\infty u^3 e^{-\frac{1}{2}u^2}\, du + 3z_k \int_{-z_k}^\infty u^2 e^{-\frac{1}{2}u^2}\, du + \\ &\quad 3z_k^2 \int_{-z_k}^\infty u e^{-\frac{1}{2}u^2}\, du + z_k^3 \int_{-z_k}^\infty e^{-\frac{1}{2}u^2}\, du\end{aligned} \tag{3.1.42}$$

and the denominator is identical to (3.1.39).

To facilitate computation of these moments, we define the following integral:

$$I_m(z) \;=\; \int_{-z}^\infty r^m e^{-\frac{r^2}{2}}\, dr$$

$$
\begin{aligned}
&= \int_{-z}^{\infty} r^{m-1} r\, e^{-\frac{r^2}{2}}\, dr \\
&= -r^{m-1} e^{-\frac{r^2}{2}}\Big|_{-z}^{\infty} + (m-1)\int_{-z}^{\infty} r^{m-2} e^{-\frac{r^2}{2}}\, dr \\
&= (-z)^{m-1} e^{-\frac{z^2}{2}} + (m-1) I_{m-2}(z) \\
&= (-z)^{m-1} I_1(z) + (m-1) I_{m-2}(z), \qquad (3.1.43)
\end{aligned}
$$

and give explicit formulae for $I_0(z)$ and $I_1(z)$:

$$
I_0(z) = \int_{-z}^{\infty} e^{-\frac{r^2}{2}}\, dr = \sqrt{2\pi} F_{\mathcal{N}}(z) \qquad (3.1.44)
$$

$$
I_1(z) = \int_{-z}^{\infty} r e^{-\frac{r^2}{2}}\, dr = e^{-\frac{z^2}{2}} \qquad (3.1.45)
$$

where $F_{\mathcal{N}}(z)$ is the cumulative distribution function of a $\mathcal{N}(0,1)$ variable, which can be evaluated (in terms of the complementary error function $erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2}\, dt$, found in most programming packages) as

$$
\begin{aligned}
F_{\mathcal{N}}(z) &= \frac{1}{\sqrt{2\pi}} \int_{-z}^{\infty} e^{-\frac{s^2}{2}}\, ds \\
&= \frac{1}{\sqrt{\pi}} \int_{-\frac{z}{\sqrt{2}}}^{\infty} e^{-u^2}\, du \\
&= \frac{1}{2}\, erfc\left(-\frac{z}{\sqrt{2}}\right). \qquad (3.1.46)
\end{aligned}
$$

The moments (3.1.38) and (3.1.41) can now be expressed, in terms of the integrals presented above, as

$$
\begin{aligned}
\rho_2(z_k) &= \frac{I_2(z_k) + 2 z_k I_1(z_k) + z_k^2 I_0(z_k)}{I_1(z_k) + z_k I_0(z_k)} \\
&= \frac{z_k I_1(z_k) + (z_k^2 + 1) I_0(z_k)}{I_1(z_k) + z_k I_0(z_k)} \\
&= \frac{z_k\, e^{-\frac{z^2}{2}} + \sqrt{2\pi}(z_k^2 + 1) F_{\mathcal{N}}(z_k)}{e^{-\frac{z_k}{2}} + \sqrt{2\pi} z_k F_{\mathcal{N}}(z_k)} \qquad (3.1.47) \\
& \qquad (3.1.48)
\end{aligned}
$$

and

$$\begin{aligned}\rho_3(z_k) &= \frac{I_3(z_k)+3z_kI_2(z_k)+3z_k^2I_1(z_k)+z_k^3I_0(z_k)}{I_2(z_k)+2z_kI_1(z_k)+z_k^2I_0(z_k)}\\ &= \frac{z_k^2I_1(z_k)+2I_1(z_k)-3z_k^2I_1(z_k)+3z_kI_0(z_k)+3z_k^2I_1(z_k)+z_k^3I_0(z_k)}{z_kI_1(z_k)+(z_k^2+1)I_0(z_k)}\\ &= \frac{z_k\big[z_kI_1(z_k)+(z_k^2+1)I_0(z_k)\big]+2\big[I_1(z_k)+z_kI_0(z_k)\big]}{z_kI_1(z_k)+(z_k^2+1)I_0(z_k)}\\ &= z_k+\frac{2}{\rho_2(z_k)} \end{aligned} \tag{3.1.49}$$

Having derived expressions for $E\big[\bar{r}_k|\bar{\mathbf{b}}_k\big]$ (for $d=2$ and $d=3$), we turn our attention to evaluating $\mathrm{var}\big[\bar{r}_k|\bar{\mathbf{b}}_k\big]$. The n-th moment of $p_{\bar{r}_k|\bar{\mathbf{b}}_k}(\bar{r}_k|\bar{\mathbf{b}}_k)$ as given in (3.1.37) is:

$$\begin{aligned}E\big[\bar{r}_k^n|\bar{\mathbf{b}}_k\big] &= \frac{1}{K(z_k)}\int_0^\infty \bar{r}_k^{\,n+d-1}\,e^{-\frac{1}{2}(\bar{r}_k-z_k)^2}\,d\bar{r}_k\\ &= \frac{1}{K(z_k)}\left(\int_0^\infty \bar{r}_k^{\,n+d-2}(\bar{r}_k-z_k)\,e^{-\frac{1}{2}(\bar{r}_k-z_k)^2}\,d\bar{r}_k+\int_0^\infty z_k\bar{r}_k^{\,n+d-2}\,e^{-\frac{1}{2}(\bar{r}_k-z_k)^2}\,d\bar{r}_k\right),\end{aligned}$$

where the first expression inside the brackets can be integrated by parts by setting

$$\begin{aligned} u &= \bar{r}_k^{r+d-2} & dv &= (\bar{r}_k-z_k)\,e^{-\frac{1}{2}(\bar{r}_k-z_k)}\,d\bar{r}_k\\ du &= (n+d-2)\bar{r}_k^{n+d-3}\,d\bar{r}_k & v &= -e^{-\frac{1}{2}(\bar{r}_k-z_k)},\end{aligned}$$

which results in:

$$\begin{aligned}E\big[\bar{r}_k^n|\bar{\mathbf{b}}_k\big] &= \frac{1}{K(z_k)}\left(-\bar{r}_k^{r+d-2}\,e^{-\frac{1}{2}(\bar{r}_k-z_k)}\Big|_0^\infty+\int_0^\infty(n+d-2)\bar{r}_k^{n+d-3}\,e^{-\frac{1}{2}(\bar{r}_k-z_k)}\,d\bar{r}_k\right)+\\ &\quad\frac{1}{K(z_k)}\int_0^\infty z_k\bar{r}_k^{\,n+d-2}\,e^{-\frac{1}{2}(\bar{r}_k-z_k)^2}\,d\bar{r}_k\\ &= \frac{(n+d-2)}{K(z_k)}\int_0^\infty \bar{r}_k^{n+d-3}\,e^{-\frac{1}{2}(\bar{r}_k-z_k)}\,d\bar{r}_k+\frac{z_k}{K(z_k)}\int_0^\infty \bar{r}_k^{\,n+d-2}\,e^{-\frac{1}{2}(\bar{r}_k-z_k)^2}\,d\bar{r}_k\\ &= (n+d-2)E\big[\bar{r}_k^{n-2}|\bar{\mathbf{b}}_k\big]+z_kE\big[\bar{r}_k^{n-1}|\bar{\mathbf{b}}_k\big].\end{aligned} \tag{3.1.50}$$

This recursive formula allows computation of the higher order moments of the desired conditional density. In particular, the variance can be expressed in terms of the second

order moment as:

$$\begin{aligned}\mathrm{var}\big[\bar{r}_k|\bar{\mathbf{b}}_k\big] &= E\big[\bar{r}_k^2|\bar{\mathbf{b}}_k\big] - \big(E\big[\bar{r}_k^2|\bar{\mathbf{b}}_k\big]\big)^2 \\ &= d + z_k\rho_d(z_k) - (\rho_d(z_k))^2. \end{aligned} \tag{3.1.51}$$

We can now return to (3.1.26) and (3.1.28) and provide expressions for the *exact* mean and covariance of the posterior density $p(\mathbf{x}_k|\mathbf{b}_k)$:

$$\begin{aligned}E[\mathbf{x}_k|\mathbf{b}_k] &= (\mathbf{I} - \mathbf{K}_k\mathbf{H})\hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\mathbf{u}_k^m + (\mathbf{b}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{b}_k)^{-\frac{1}{2}}E[\bar{r}_k|\bar{\mathbf{b}}_k]\mathbf{K}_k\mathbf{b}_k \\ &= (\mathbf{I} - \mathbf{K}_k\mathbf{H})\hat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\mathbf{u}_k^m + \gamma_k\mathbf{K}_k\mathbf{b}_k \end{aligned} \tag{3.1.52}$$

and

$$\begin{aligned}\mathrm{cov}\big[\mathbf{x}_k|\mathbf{b}_k\big] &= (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1} + (\mathbf{b}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{b}_k)^{-1}\mathrm{var}\big[\bar{r}_k|\bar{\mathbf{b}}_k\big]\mathbf{K}_k\mathbf{b}_k\mathbf{b}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}} \\ &= (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1} + \delta_k\mathbf{K}_k\mathbf{b}_k\mathbf{b}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}}, \end{aligned} \tag{3.1.53}$$

where

$$\gamma_k = E\big[||\mathbf{y}_k|| \mid \mathbf{b}_k\big] = (\mathbf{b}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{b}_k)^{-\frac{1}{2}}E[\bar{r}_k|\bar{\mathbf{b}}_k] \tag{3.1.54}$$

and

$$\delta_k = \mathrm{var}\big[||\mathbf{y}_k|| \mid \mathbf{b}_k\big] = (\mathbf{b}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{b}_k)^{-1}\mathrm{var}\big[\bar{r}_k|\bar{\mathbf{b}}_k\big] \tag{3.1.55}$$

are the mean and variance of the range estimates of $\mathbf{y}_k$.

It can be seen that (3.1.52) and (3.1.53) closely resemble the standard Kalman filter update equations, with the addition of extra terms accounting for the inaccessibility of the augmented measurement. More specifically, $\gamma_k\mathbf{K}_k\mathbf{b}_k$ represents the correction to the state estimate, resulting from the unobservability of $\mathbf{y}_k$ which is instead approximated by

$$E\big[\mathbf{y}_k|\mathbf{b}_k\big] = E\big[||\mathbf{y}_k||\,\mathbf{b}_k|\mathbf{b}_k\big] = E\big[||\mathbf{y}_k||\,|\mathbf{b}_k\big]\mathbf{b}_k = \gamma_k\mathbf{b}_k.$$

Similarly, the addition of the non-negative term $\delta_k \mathbf{K}_k \mathbf{b}_k \mathbf{b}_k^{\mathrm{T}} \mathbf{K}_k^{\mathrm{T}}$ to the covariance estimate is none other than the covariance of $\mathbf{K}_k \mathbf{y}_k$ given $\mathbf{b}_k$, expressed as:

$$\mathrm{cov}\big[\mathbf{K}_k \mathbf{y}_k | \mathbf{b}_k\big] \;=\; \mathrm{cov}\big[||\mathbf{y}_k||\, \mathbf{K}_k \mathbf{b}_k | \mathbf{b}_k\big] \;=\; \mathrm{var}\big[||\mathbf{y}_k|| \,|\mathbf{b}_k\big] \mathbf{K}_k \mathbf{b}_k \mathbf{b}_k^{\mathrm{T}} \mathbf{K}_k^{\mathrm{T}} \;=\; \delta_k \mathbf{K}_k \mathbf{b}_k \mathbf{b}_k^{\mathrm{T}} \mathbf{K}_k^{\mathrm{T}}.$$

Measurement Noise in the Augmented Measurement

The additive augmented measurement noise term $\mathbf{w}_k$ is related to the noise $\tilde{\mathbf{w}}_k$ of the conventional angular measurement model through the scaling

$$\mathbf{w}_k = \|\mathbf{d}_k\| \tilde{\mathbf{w}}_k, \tag{3.1.56}$$

where $\|\mathbf{d}\|_k$ is the relative range of the target from the sensor platform at time k and $\tilde{\mathbf{w}}_k \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{d\times d})$ is independent of $\mathbf{d}_k$. From the law of conditional probability, it follows that $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q}_k^m)$, where

$$\mathbf{Q}_k^m \;=\; E\left[\|\mathbf{d}_k\|^2\right] \sigma^2 \mathbf{I}_{d\times d}. \tag{3.1.57}$$

The exact evaluation of the covariance matrix of the augmented measurement noise at time k requires access to the augmented measurement itself. As this is not available, the covariance is instead approximated by replacing $\|\mathbf{d}_k\|$ by the *predicted* relative range, conditioned on measurements up to time $k-1$. Using (without loss of generality, for the 2-dimensional case), the notation $\mathbf{d}_k = [d_{k1} \;\; d_{k2}]^{\mathrm{T}} = \mathbf{H}\mathbf{x}_k + \mathbf{u}_k^m$ we have:

$$\begin{aligned}
\mathbf{Q}_k^m \;&= \sigma^2 E\left[\|\mathbf{d}_k\|^2 \,|\mathbf{b}_{1:k-1}\right] \mathbf{I}_{d\times d} \\
&= \sigma^2 E\left[(d_{k1}^2 + d_{k2}^2) \,|\mathbf{b}_{1:k-1}\right] \mathbf{I}_{d\times d} \\
&= \sigma^2 \Big(\mathrm{var}\left[d_{k1}|\mathbf{b}_{1:k-1}\right] + \left(E\left[d_{k1}|\mathbf{b}_{1:k-1}\right]\right)^2 + \mathrm{var}\left[d_{k2}|\mathbf{b}_{1:k-1}\right] + \left(E\left[d_{k2}|\mathbf{b}_{1:k-1}\right]\right)^2\Big) \mathbf{I}_{d\times d} \\
&= \sigma^2 \left(\|\mathbf{H}\hat{\mathbf{x}}_{k|k-1} + \mathbf{u}_k^m\|^2 + \mathrm{trace}\left(\mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^{\mathrm{T}}\right)\right) \mathbf{I}_{d\times d}.
\end{aligned} \tag{3.1.58}$$

Depending on the application, an additional term $\mathbf{Q}_k^{tr}$, representing a translational noise covariance matrix, can be added to the expression for $\mathbf{Q}_k^m$. This can account for unknown perturbations in the position of the sensor platform well as random fluctuations in the direction of arrival of signals from an extended target.

This concludes the derivation of the SRF equations. The following section concerns the derivation of the probability density functions, in both 2D and 3D, of the bearings of the augmented measurement. Although a 2D derivation appears in [19], we propose here a slightly different approach which we then generalize to the 3D case. The similarity of the bearings density for measurements in 2D space to that associated with the 'conventional' measurement model is discussed. For the case of measurements in 3D space, an alternative measurement noise model is presented, which provides a better characterization of the uncertainty in azimuth and elevation angles and, moreover, closely matches the bearings density of the SRF measurement model.

The 2D Bearing density

Consider a point in 2 dimensional Euclidian space, described by the normally distributed random displacement vector $\mathbf{y}_k = (x_k,\ y_k)^{\mathrm{T}}$, with mean $\mathbf{d}_k = (d_{k1},\ d_{k2})^{\mathrm{T}}$ and covariance matrix $\mathbf{V}_k$. Suppressing the time index k for notational clarity, the density of $\mathbf{y}$ can be written as

$$p_{\mathbf{y}}(\mathbf{y}) \ = \ \frac{1}{2\pi\left(\det\mathbf{V}\right)^{1/2}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{d})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{y}-\mathbf{d})}. \tag{3.1.59}$$

Using (3.1.31) and (3.1.32), the corresponding density in polar coordinates (r, η) can be expressed as

$$\begin{aligned} p_{r,\eta}(r,\eta) \ &= \left|\nabla g_p(r,\eta)\right| \ p_{\mathbf{y}}\left(g_p(r,\eta)\right) \\ &= \frac{r}{2\pi\left(\det\mathbf{V}\right)^{1/2}} e^{-\frac{1}{2}(g_p(r,\eta)-\mathbf{d})^{\mathrm{T}}\mathbf{V}^{-1}(g_p(r,\eta)-\mathbf{d})} \end{aligned} \tag{3.1.60}$$

and the argument of the exponential can be written in compact quadratic form, such that

$$p_{r,\eta}(r,\eta) \ = \frac{r}{2\pi\left(\det\mathbf{V}\right)^{1/2}} e^{-\frac{1}{2}\left(ar^2+br+c\right)}, \tag{3.1.61}$$

where the coefficients are:

$$a \ = a_{11}\sin^2(\eta) + a_{22}\cos^2(\eta) + (a_{12}+a_{21})\cos(\eta)\sin(\eta) \tag{3.1.62}$$

$$b \ = \ -2a_{11}d_1\sin(\eta) - 2a_{22}d_2\cos(\eta) - (a_{12}+a_{21})(d_1\cos(\eta) + d_2\sin(\eta)) \tag{3.1.63}$$

$$c \ = a_{11}(d_1)^2 + a_{22}(d_2)^2 + (a_{12}+a_{21})d_1d_2, \tag{3.1.64}$$

with a_{ij} denoting the elements of the precision matrix $\mathbf{V}^{-1}$. Completing the square inside the exponential, we arrive at:

$$p_{r,\eta}(r,\eta) \;=\; \frac{r}{2\pi\,(\det\mathbf{V})^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}}\, e^{-\frac{a}{2}\left(r+\frac{b}{2a}\right)^2}.$$

The density of the azimuth angle η is then obtained, by marginalizing out the range r:

$$\begin{aligned} p_\eta(\eta) \;&=\; \int_0^\infty p_{r,\eta}(r,\eta)\,dr \\ &= \frac{1}{2\pi(\det\mathbf{V})^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}} \int_0^\infty r\, e^{-\frac{a}{2}\left(r+\frac{b}{2a}\right)^2} dr, \end{aligned} \tag{3.1.65}$$

which reduces, after the change of variable $u=\sqrt{a}(r+\frac{b}{2a})$, to:

$$\begin{aligned} p_\eta(\eta) \;&= \frac{1}{2\pi(\det\mathbf{V})^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}} \int_{b/(2\sqrt{a})}^\infty \left(\frac{u}{a}-\frac{b}{2a^{3/2}}\right) e^{-\frac{u^2}{2}}\,du \\ &= \frac{1}{2\pi(\det\mathbf{V})^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}} \left[\frac{1}{a} I_1\left(-\frac{b}{2\sqrt{a}}\right) - \frac{b}{2a^{3/2}} I_0\left(-\frac{b}{2\sqrt{a}}\right)\right] \\ &= \frac{1}{2\pi(\det\mathbf{V})^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}} \left[\frac{1}{a}\, e^{-\frac{b^2}{8a}} - \frac{\sqrt{2\pi}b}{2a^{3/2}} F_{\mathcal{N}}\left(-\frac{b}{2\sqrt{a}}\right)\right], \end{aligned} \tag{3.1.66}$$

where the final expression is obtained using the previously defined functions for I_1 and I_0 in (3.1.45) and (3.1.44).

Consider now the 'conventional' angular measurement model in 2D space (in which noise is added to the bearing of the true displacement):

$$\begin{aligned} z_k \;&=\; \tan^{-1}\left(d_{k1}/d_{k2}\right) + w_k \\ &= \theta_k + w_k, \end{aligned} \tag{3.1.67}$$

where w_k is a scalar white noise process with distribution $\mathcal{N}\left(0,\sigma^2\right)$ 'wrapped' onto the interval $[-\pi,\pi]$. We compare the bearing density $p_\eta(\eta)$ with the 'wrapped' Normal density ([43]) resulting from the standard measurement model (3.1.67), expressed as

$$p(z_k) \;=\; \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left(z_k-\theta-2\pi i\right)^2\right). \tag{3.1.68}$$

Figure 3.1 provides a comparison of the two densities, for values of σ up to 1 rad, when $\hat{\mathbf{y}}$ is normalized to $(0,1)^{\mathrm{T}}$ and the covariance of the augmented measurement noise is taken to be $\mathbf{Q}^m = \|\mathbf{d}\|^2\sigma^2\mathbf{I}_{d\times d}$. The difference between the two densities, which never exceeds 0.04 for $\sigma \leq 1$, is clearly negligible. In other words, the bearing density of the augmented measurement is indistinguishable from that of the conventional angular measurement. This establishes the equivalence of modelling a noisy bearing measurement as the measurement of a 'noisy displacement' vector (the augmented measurement).

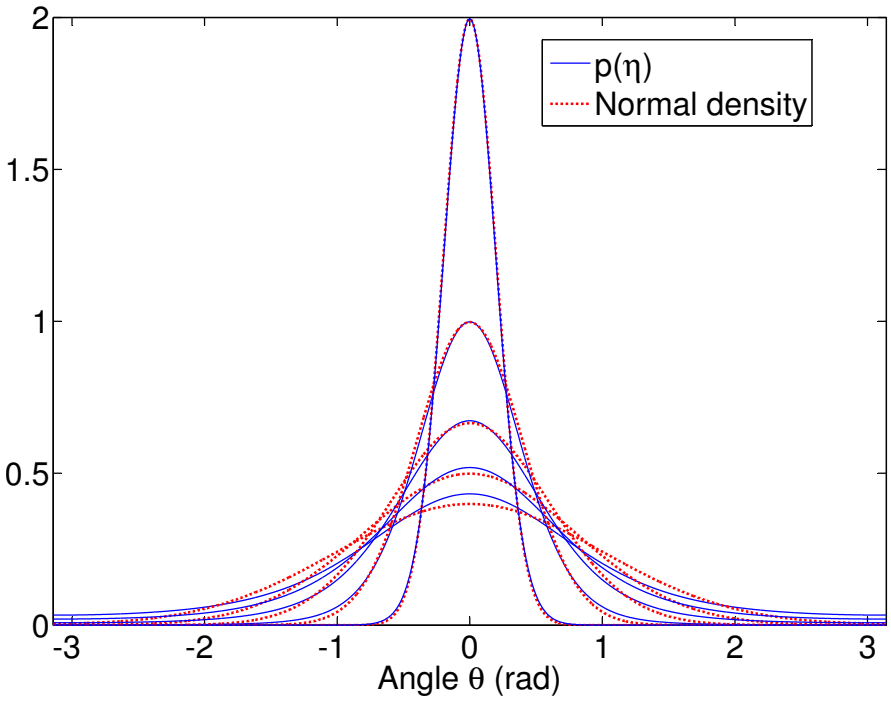


Figure 3.1: Comparison of the azimuth bearing density with the 'wrapped' Normal density for $\mathbf{y} = (0,1)^{\mathrm{T}}$ and $\sigma = 0.2, 0.4, \ldots, 1$

The 3D Bearings density

Take a point in 3 dimensional Euclidian space, described by the normally distributed random displacement vector $\mathbf{y}_k = (x_k,\ y_k,\ z_k)^{\mathrm{T}}$, with mean $\mathbf{d}_k = (d_{k1},\ d_{k2},\ d_{k3})^{\mathrm{T}}$ and covariance matrix $\mathbf{V}_k$. Suppressing the time index k for notational clarity, the density of $\mathbf{y}$ can be written as

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{1}{(2\pi)^{3/2}(\det\mathbf{V})^{1/2}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{d})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{y}-\mathbf{d})}. \tag{3.1.69}$$

Using (3.1.31) and (3.1.33), the corresponding density in spherical coordinates (r, η, ζ) can be expressed as

$$\begin{aligned} p_{r,\eta,\zeta}(r,\eta,\zeta) &= |\nabla g_s(r,\eta,\zeta)|\; p_{\mathbf{y}}\left(g_s(r,\eta,\zeta)\right) \\ &= \frac{r^2\cos(\zeta)}{(2\pi)^{3/2}\left(\det\mathbf{V}\right)^{1/2}}\, e^{-\frac{1}{2}(g_s(r,\eta,\zeta)-\mathbf{d})^{\mathrm{T}}\mathbf{V}^{-1}(g_s(r,\eta,\zeta)-\mathbf{d})} \end{aligned} \tag{3.1.70}$$

and similarly to the 2D case, the argument of the exponential can be written in compact quadratic form, such that

$$p_{r,\eta,\zeta}(r,\eta,\zeta) = \frac{r^2\cos(\zeta)}{(2\pi)^{3/2}\left(\det\mathbf{V}\right)^{1/2}}\, e^{-\frac{1}{2}\left(ar^2+br+c\right)}, \tag{3.1.71}$$

with coefficients:

$$\begin{aligned} a =\; & a_{11}\sin^2(\eta)\cos^2(\zeta) + a_{22}\cos^2(\eta)\cos^2(\zeta) + a_{33}\sin^2(\zeta) + \\ & (a_{12}+a_{21})\cos(\eta)\sin(\eta)\cos^2(\zeta) + \\ & (a_{23}+a_{32})\cos(\eta)\cos(\zeta)\sin(\zeta) + \\ & (a_{13}+a_{31})\sin(\eta)\cos(\zeta)\sin(\zeta), \end{aligned} \tag{3.1.72}$$

$$\begin{aligned} b =\; & -\left[2a_{11}d_1 + (a_{12}+a_{21})d_2 + (a_{13}+a_{31})d_3\right]\sin(\eta)\cos(\zeta) - \\ & \left[2a_{22}d_2 + (a_{12}+a_{21})d_1 + (a_{23}+a_{32})d_3\right]\cos(\eta)\cos(\zeta) - \\ & \left[2a_{33}d_3 + (a_{23}+a_{32})d_2 + (a_{13}+a_{31})d_1\right]\sin(\zeta), \end{aligned} \tag{3.1.73}$$

$$\begin{aligned} c =\; & (d_1)^2 a_{11} + (d_2)^2 a_{22} + (d_3)^2 a_{33} + \\ & d_1 d_2 (a_{12}+a_{21}) + d_2 d_3 (a_{23}+a_{32}) + d_1 d_3 (a_{13}+a_{31}), \end{aligned} \tag{3.1.74}$$

and a_{ij} are the elements of the precision matrix $\mathbf{V}^{-1}$. As with the 2D case, completing the square in the exponential leads to:

$$p_{r,\eta,\zeta}(r,\eta,\zeta) = \frac{r^2\cos(\zeta)}{(2\pi)^{3/2}\left(\det\mathbf{V}\right)^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}}\, e^{-\frac{a}{2}\left(r+\frac{b}{2a}\right)^2}.$$

The density of the azimuth and elevation angles (η, ζ) is then obtained, by marginalizing out the range r:

$$\begin{aligned} p_{\eta,\zeta}(\eta,\zeta) &= \int_0^\infty p_{r,\eta,\zeta}(r,\eta,\zeta)\, dr \\ &= \frac{\cos(\zeta)}{(2\pi)^{3/2}(\det\mathbf{V})^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}} \int_0^\infty r^2\, e^{-\frac{a}{2}\left(r+\frac{b}{2a}\right)^2} dr, \end{aligned} \tag{3.1.75}$$

and carrying out the change of variable $u = \sqrt{a}(r + \frac{b}{2a})$ leading to:

$$\begin{aligned} p_{\eta,\zeta}(\eta,\zeta) &= \frac{\cos(\zeta)}{(2\pi)^{3/2}(\det\mathbf{V})^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}} \int_{b/(2\sqrt{a})}^\infty \left(\frac{u^2}{a^{3/2}} - \frac{ub}{a^2} + \frac{b^2}{4a^{5/2}}\right) e^{-\frac{u^2}{2}}\, du \\ &= \frac{\cos(\zeta)}{(2\pi)^{3/2}(\det\mathbf{V})^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}} \left[\frac{1}{a^{3/2}} I_2\left(-\frac{b}{2\sqrt{a}}\right) - \frac{b}{a^2} I_1\left(-\frac{b}{2\sqrt{a}}\right) + \frac{b^2}{4a^{5/2}} I_0\left(-\frac{b}{2\sqrt{a}}\right)\right] \\ &= \frac{\cos(\zeta)}{(2\pi)^{3/2}(\det\mathbf{V})^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}} \left[-\frac{b}{2a^2}\, e^{-\frac{b^2}{8a}} + \frac{\sqrt{2\pi}}{4a^{5/2}}(4a+b^2) F_{\mathcal{N}}\left(-\frac{b}{2\sqrt{a}}\right)\right], \end{aligned} \tag{3.1.76}$$

where, as in the 2D case, use was made of the previously defined expressions for I_2, I_1 and I_0 in (3.1.43), (3.1.45) and (3.1.44).

Now consider the 'conventional' angular measurement model in 3D space, where the bearings measurement process $\{\mathbf{z}_k\}$ can be expressed as

$$\mathbf{z}_k = h(\mathbf{d}_k) + \mathbf{w}_k, \tag{3.1.77}$$

$$= \begin{bmatrix} \arctan\Big(d_{k1}/d_{k2}\Big) \\ \arctan\left(d_{k3}/\sqrt{(d_{k1})^2 + (d_{k2})^2}\right) \end{bmatrix} + \mathbf{w}_k, \tag{3.1.78}$$

$$= \begin{bmatrix}\theta_k & \alpha_k\end{bmatrix}^{\mathrm{T}} + \mathbf{w}_k \tag{3.1.79}$$

in which the function h converts the true displacement vector $\mathbf{d}_k$ at time k into its azimuth and elevation angles (θ_k, α_k), and $\mathbf{w}_k$ is an additive noise term. It is customary to take $\mathbf{w}_k$ to be a Gaussian variable independent of $\mathbf{d}_k$ with covariance $\sigma^2\mathbf{I}_{2\times 2}$ and with azimuth and elevation components 'wrapped' onto $[-\pi,\ \pi]$ and $[-\pi/2,\ \pi/2]$, respectively.

To illustrate the implications of this measurement noise model, consider figure 3.2

where $\mathbf{p}$ is the projection, onto the unit sphere, of a displacement vector $\mathbf{y}$ whose azimuth and elevation angle vector is $(\theta,\, \alpha)$. Increments $(\delta\theta,\ \delta\alpha)$ due to measurement noise result in arc length increments at $\mathbf{p}$ of $(\cos\alpha\,\delta\theta,\ \delta\alpha)$ along the constant longitudinal and latitudinal contours. We see, qualitatively speaking, that this model has the effect of almost completely suppressing the measurement noise when the true displacement vector is almost vertical (pointing either upwards or downwards). Such a model is unnatural for certain tracking applications, for instance when 2D phased arrays are used to obtain azimuth and elevation measurements. Pathological geometries involving high target elevation can arise, where the usual additive white Gaussian noise model can give a very poor representation of the noise characteristics of the sensor. This problem is, however, less of a concern in applications where the bearings-only measurements are obtained by cameras or forward-looking infrared (FLIR) sensors [9], which typically have fields of view too narrow to give rise to the geometries just described.

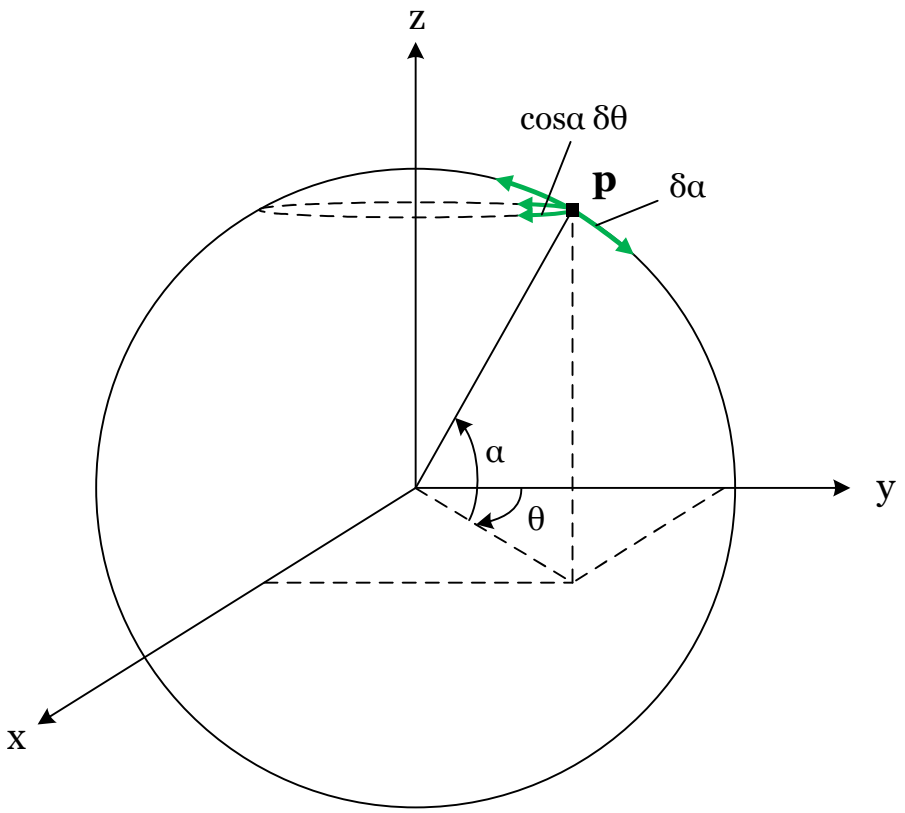


Figure 3.2: Effect of measurement noise, in the standard model, on the deviations of the displacement vector along constant latitudinal and longitudinal lines of the unit sphere.

We propose an alternative model, which associates a Gaussian density with covariance $\sigma^2\mathbf{I}_{2\times 2}$ not to deviations $(\delta\theta, \delta\alpha)$, but to deviations along constant latitudinal and longitudinal contours $(\delta\beta, \delta\gamma)$, respectively, on the unit sphere [2]. These deviations are related according to

$$\delta\beta \;= (\cos\alpha)^{-1}\,\delta\theta \tag{3.1.80}$$

$$\delta\gamma \;= \delta\alpha \tag{3.1.81}$$

where the $(\cos\alpha)^{-1}$ term compensates for the reduction of azimuthal uncertainty as α approaches $\pm\pi/2$. This suggests the following model for the measurement $\mathbf{z}_k$:

$$\mathbf{z}_k \;=\; h(\mathbf{d}_k) + \tilde{\mathbf{w}}_{\mathbf{d}_k}, \tag{3.1.82}$$

where the noise process $\tilde{\mathbf{w}}_{\mathbf{d}_k}$, which is now correlated with $\mathbf{d}_k$, has density

$$p(\tilde{\mathbf{w}}_{\mathbf{d}_k}) \;=\; \mathcal{N}_w\left(\tilde{\mathbf{w}}_{\mathbf{d}_k}; \mathbf{0}, \sigma^2\begin{bmatrix}(\cos\alpha_k)^{-2} & 0\\ 0 & 1\end{bmatrix}\right), \tag{3.1.83}$$

with azimuth and elevation components 'wrapped' and onto $[-\pi,\ \pi]$ and $[-\pi/2,\ \pi/2]$, respectively, and $\alpha_k = \arctan\left(d_{k3}/\sqrt{(d_{k1})^2 + (d_{k2})^2}\right)$.

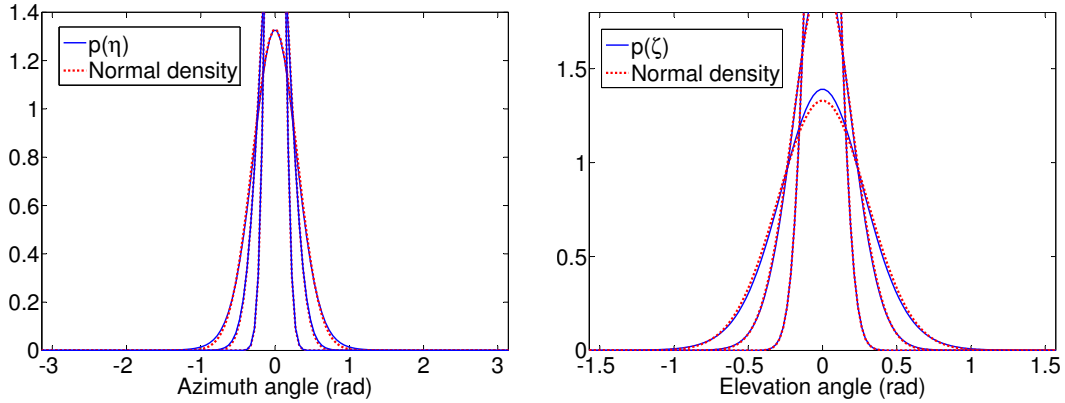
We now investigate how well the modified measurement model 3.1.82 compares to the bearings density $p_{\eta,\zeta}(\eta,\zeta)$ used in the construction of the SRF. The 'wrapped' Gaussian density of the measurement $\mathbf{z}_k$, given $\mathbf{d}_k$, from (3.1.82) is given by:

$$\begin{aligned} p(\mathbf{z}_k) \;=\; &\sum_{i=-\infty}^{\infty}\mathcal{N}(z_{k1}+2\pi i;\,\theta_k, \sigma^2/(\cos\alpha_k)^2)\sum_{j=-\infty}^{\infty}\mathcal{N}(z_{k2}+2\pi j;\,\alpha_k,\sigma^2) + \\ &\sum_{i=-\infty}^{\infty}\mathcal{N}(z_{k1}+\pi+2\pi i;\,\theta_k, \sigma^2/(\cos\alpha_k)^2)\sum_{j=-\infty}^{\infty}\mathcal{N}(\pi - z_{k2}+2\pi j;\,\alpha_k,\sigma^2) \end{aligned} \tag{3.1.84}$$

Figures 3.3, 3.4, 3.5 and 3.6 illustrate the difference between the marginal densities of the azimuth and elevation components of the noise (given the displacement) according

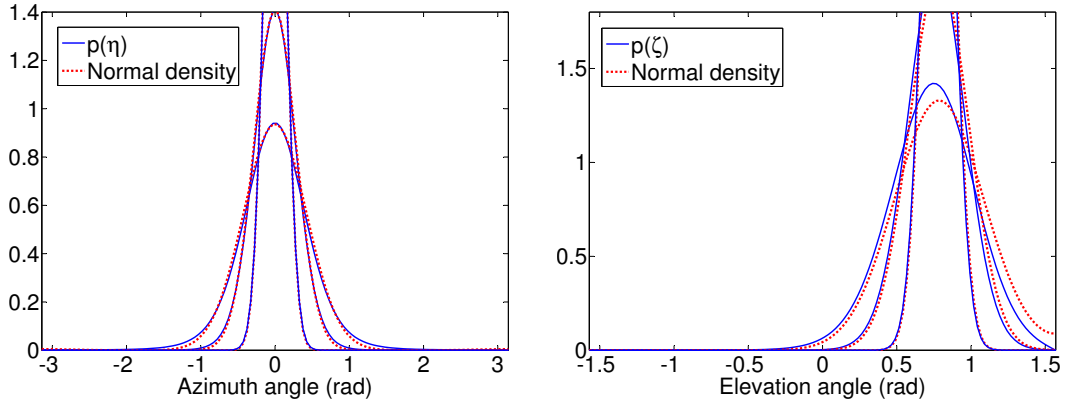
[2] Although theoretically accurate, an in depth analysis of the physics of actual sensors needs to be carried out to assesss the validity of this measurement noise model.

to the two models, for values of σ up to 0.3 and for $\alpha = 0, \frac{\pi}{4}, \frac{\pi}{3}$ and $\frac{\pi}{2}$. We see that for the azimuth component of the measurements, the agreement between the two noise models is very close. We remark that the uniform marginal azimuth densities in figure 3.6 adequately describe the total uncertainty of the azimuth angle of a point exactly above the sensor. The marginal elevation angle densities match as well (although to a slightly lesser extent for high values of α and σ). We note that the singularity at the pole is captured by $p(\zeta)$ in figure 3.6, where the marginal elevation angle density vanishes at $\frac{\pi}{2}$.



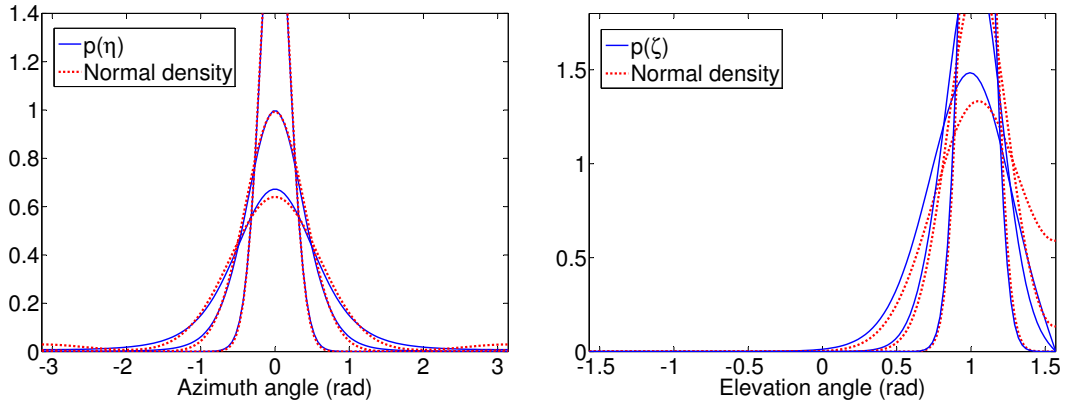
(a) (b)

Figure 3.3: Comparison of marginal azimuth (a) and elevation (b) angle densities with the 'wrapped' normal density for $\alpha = 0$ and $\sigma = 0.1, 0.2, 0.3$



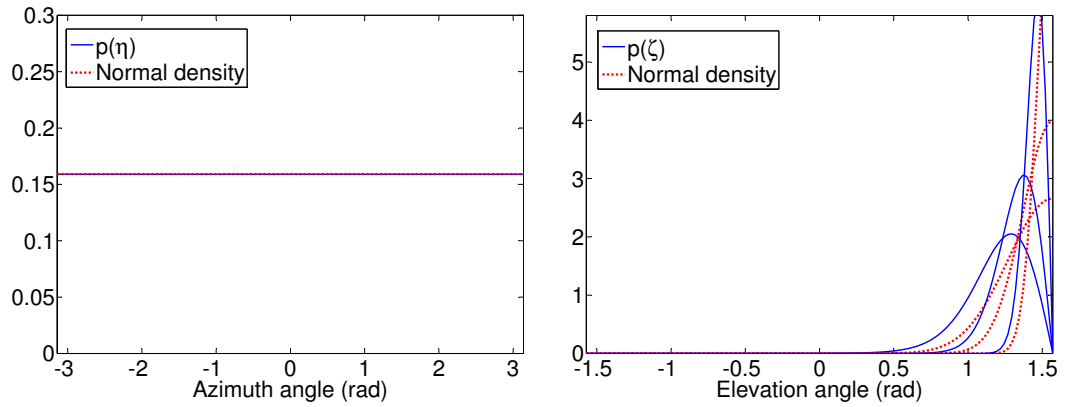
(a) (b)

Figure 3.4: Comparison of marginal azimuth (a) and elevation (b) angle densities with the normal density for $\alpha = \frac{\pi}{4}$ and $\sigma = 0.1, 0.2, 0.3$



(a) (b)

Figure 3.5: Comparison of marginal azimuth (a) and elevation (b) angle densities with the normal density for $\alpha = \frac{\pi}{3}$ and $\sigma = 0.1, 0.2, 0.3$



(a) (b)

Figure 3.6: Comparison of marginal azimuth (a) and elevation (b) angle densities with the normal density for $\alpha = \frac{\pi}{2}$ and $\sigma = 0.1, 0.2, 0.3$

3.1.2 Extension to Data Association

In most practical tracking scenarios, measurements can originate not only from actual targets but also from sources of 'clutter', representing multi-path effects, returns from spurious objects and other artifacts and reflections, generally assumed unrelated with the target. These 'unlabeled' false alarms lead to data association uncertainty and represent a source of nonlinearity in the measurement process. We present next two approaches for the modeling of clutter, which lead to modifications, similar in nature, to standard filtering algorithms and in particular to the SRF.

Probabilistic Data Association

Originally presented in [6], probabilistic data association (PDA) aims to combine contributions from multiple candidate measurements, at each time step k, for the update of the target state estimate. In addition to the usual target state and measurement processes defined in (2.2.1) and (2.2.2), the underlying assumptions of PDA are that at each time step k:

- the density $p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1}) = \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1})$ provides a sufficient statistic of the state trajectory up to the previous time step,
- the available measurements are validated using a gating procedure,
- at most one validated measurement can have originated from the target (allowing for the possibility of less-than-unity probability of detection P_D and for the 'true' measurement falling outside the validation region), the remaining measurements assumed to be clutter,
- clutter measurements are assumed to have uniform i.i.d. distribution (in the measurement space) and Poisson distribution in their number per unit volume (with a given mean λ).

The first step of the PDA algorithm is measurement validation (or gating), which is a crude initial selection process that limits computational complexity. Under the assumption that measurements have a Gaussian distribution, among the N available measurements at time k, the set of $m(k)$ validated measurements is denoted as

$$\boldsymbol{\mathcal{Z}}_k \;=\; \{\mathbf{z}_k^i\}_{i=1}^{m(k)}, \tag{3.1.85}$$

and consists of all measurements $\mathbf{z}_k^j$, for $j = 1, \ldots, m(k)$, such that

$$(\mathbf{z}_k^j - \hat{\mathbf{z}}_{k|k-1})^{\mathrm{T}}\, \mathbf{S}_k^{-1}\, (\mathbf{z}_k^j - \hat{\mathbf{z}}_{k|k-1}) \leq G, \tag{3.1.86}$$

where $\mathbf{S}_k$ is the innovation covariance matrix and $\hat{\mathbf{z}}_{k|k-1}$ is the predicted measurement. Assuming the distribution of clutter measurements is uniform over the surveillance region, we can see that (3.1.86) consists of a likelihood-ratio test, where the so-called gate threshold G can be obtained from tail probability tables of the χ^2 distribution for a given

probability $1 - P_G$ of type I error (i.e. of classifying the 'true' measurement, provided it exists, as originating from clutter). Typically, P_G is fixed to a large value (as high as 0.999) to ensure that the true measurement, if the target is detected, is not discarded. The volume V_k of the validation region defined by (3.1.86) depends, at each time step, on the determinant of $\mathbf{S}_k$ and is therefore variable.

Following the gating procedure, state estimation is carried out by evaluating the posterior densities conditioned on the hypotheses of each validated measurement having originated from the target, and on the null-hypothesis (representing the possibility that none of them is target originated). The resulting $m(k)$-fold Gaussian mixture can be expressed as:

$$\begin{aligned} p(\mathbf{x}_k|\boldsymbol{\mathcal{Z}}_{1:k}) &= \sum_{i=0}^{m(k)} p(\mathbf{x}_k|\epsilon_k^i, \boldsymbol{\mathcal{Z}}_{1:k}) P\{\epsilon_k^i|\boldsymbol{\mathcal{Z}}_{1:k}\} \\ &= \sum_{i=0}^{m(k)} p(\mathbf{x}_k|\epsilon_k^i, \boldsymbol{\mathcal{Z}}_{1:k})\, \beta_k^i, \end{aligned} \tag{3.1.87}$$

where ϵ_k^i, for $i = 0, \ldots, m(k)$, designates the hypothesis that the ith measurement is target originated (with the special case $i = 0$ representing the null-hypothesis), $\beta_k^i \triangleq P\{\epsilon_k^i|\boldsymbol{\mathcal{Z}}_{1:k}\}$ is the conditional probability of that hypothesis and $\boldsymbol{\mathcal{Z}}_{1:k}$ is the accumulation of validated measurements up to time k. This mixture is then reduced to a single Gaussian by moment-matching, ahead of the following iteration. The mean of the overall state estimate is given by

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= E\big[\mathbf{x}_k|\boldsymbol{\mathcal{Z}}_{1:k}\big] \\ &= \sum_{i=0}^{m(k)} E\big[\mathbf{x}_k|\epsilon_k^i, \boldsymbol{\mathcal{Z}}_{1:k}\big]\, \beta_k^i \\ &= \sum_{i=0}^{m(k)} \hat{\mathbf{x}}_{k|k}^i\, \beta_k^i. \end{aligned} \tag{3.1.88}$$

The associated covariance, meanwhile, is expressed as

$$\mathbf{P}_{k|k} = \sum_{i=0}^{m(k)} E\big[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^{\mathrm{T}}|\epsilon_k^i, \boldsymbol{\mathcal{Z}}_{1:k}\big]\, \beta_k^i$$

$$
\begin{aligned}
&= \sum_{i=0}^{m(k)} \beta_k^i \, E\big[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i)^{\mathrm{T}} | \epsilon_k^i, \boldsymbol{\mathcal{Z}}_{1:k}\big] + \sum_{i=0}^{m(k)} \beta_k^i \, (\hat{\mathbf{x}}_{k|k}^i - \hat{\mathbf{x}}_{k|k})(\hat{\mathbf{x}}_{k|k}^i - \hat{\mathbf{x}}_{k|k})^{\mathrm{T}} \\
&= \sum_{i=0}^{m(k)} \beta_k^i \, \mathbf{P}_{k|k}^i + \sum_{i=0}^{m(k)} \beta_k^i \, (\hat{\mathbf{x}}_{k|k}^i - \hat{\mathbf{x}}_{k|k})(\hat{\mathbf{x}}_{k|k}^i - \hat{\mathbf{x}}_{k|k})^{\mathrm{T}}, \qquad (3.1.89)
\end{aligned}
$$

where the first summation is the weighted sum of the conditional covariances (accounting for the covariance 'within' the densities of the mixture) and the second summation is the 'spread of means' (reflecting the covariance 'between' the densities). Clearly, the higher the number $m(k)$ of validated measurements, the more non-negative matrices will enter into the summation describing the overall covariance matrix, leading to increased uncertainty. In both (3.1.88) and (3.1.89), the posterior corresponding to the null hypothesis ($i = 0$) is the one-step prediction with mean $\hat{\mathbf{x}}_{k|k}^0 = \hat{\mathbf{x}}_{k|k-1}$ and covariance $\mathbf{P}_{k|k}^0 = \mathbf{P}_{k|k-1}$.

To calculate the association probabilities β_k^i, we consider the decomposition

$$
\begin{aligned}
\beta_k^i &= P\{\epsilon_k^i | \boldsymbol{\mathcal{Z}}_{1:k}\} \\
&= \frac{1}{c} p(\boldsymbol{\mathcal{Z}}_k | \epsilon_k^i, m(k), \boldsymbol{\mathcal{Z}}_{1:k-1}) \, P\{\epsilon_k^i | m(k), \boldsymbol{\mathcal{Z}}_{1:k-1}\}, \qquad (3.1.90)
\end{aligned}
$$

where c is a normalizing constant. The likelihood of the (validated) measurements is

$$
p(\boldsymbol{\mathcal{Z}}_k | \epsilon_k^i, m(k), \boldsymbol{\mathcal{Z}}_{1:k-1}) = \begin{cases} V_k^{-m(k)+1} P_G^{-1} p(\mathbf{z}_k^i | \boldsymbol{\mathcal{Z}}_{1:k-1}), & i = 1, \ldots, m(k) \\ V_k^{-m(k)}, & i = 0. \end{cases} \qquad (3.1.91)
$$

and the association probability conditioned on the number of validated measurements is

$$
\begin{aligned}
P\{\epsilon_k^i | m(k), \boldsymbol{\mathcal{Z}}_{1:k-1}\} &= P\{\epsilon_k^i | m(k)\} \\
&= \begin{cases} \frac{1}{m(k)} P_D P_G \Big[P_D P_G + (1 - P_D P_G) \frac{\mu_k^F(m(k))}{\mu_k^F(m(k)-1)} \Big]^{-1}, \\ \qquad\qquad i = 1, \ldots, m(k) \\ (1 - P_D P_G) \frac{\mu_k^F(m(k))}{\mu_k^F(m(k)-1)} \Big[P_D P_G + (1 - P_D P_G) \frac{\mu_k^F(m(k))}{\mu_k^F(m(k)-1)} \Big]^{-1}, \\ \qquad\qquad i = 0, \end{cases} \qquad (3.1.92)
\end{aligned}
$$

where $\mu_k^F(m)$ is the probability mass function of the number of clutter measurements in the validation region of volume V_k, given by the poisson distribution

$$\mu_k^F(m) \;=\; e^{-\lambda V_k}\frac{(\lambda V_k)^m}{m!}. \tag{3.1.93}$$

Substitution of (3.1.93) into (3.1.92) results in

$$P\{\epsilon_k^i|m(k)\} \;=\; \begin{cases} P_D P_G\Big[P_D P_G\, m(k) + (1-P_D P_G)\lambda V_k\Big]^{-1}, & i=1,\ldots,m(k) \\ (1-P_D P_G)\lambda V_k\Big[P_D P_G\, m(k) + (1-P_D P_G)\lambda V_k\Big]^{-1}, & i=0, \end{cases} \tag{3.1.94}$$

which can in turn be substituted, along with (3.1.91), into (3.1.90), to yield the final expression for the conditional association probabilities:

$$\beta_k^i \;=\; \begin{cases} \frac{p(\mathbf{z}_k^i|\boldsymbol{\mathcal{Z}}_{1:k-1})}{(1-P_D P_G)P_D^{-1}\lambda+\sum_{j=1}^{m(k)} p(\mathbf{z}_k^j|\boldsymbol{\mathcal{Z}}_{1:k-1})}, & i=1,\ldots,m(k) \\ \frac{(1-P_D P_G)P_D^{-1}\lambda}{(1-P_D P_G)P_D^{-1}\lambda+\sum_{j=1}^{m(k)} p(\mathbf{z}_k^j|\boldsymbol{\mathcal{Z}}_{1:k-1})}, & i=0. \end{cases} \tag{3.1.95}$$

Implementation of the PDA-SRF Implementing the SRF within the PDA framework follows from equations (3.1.88) and (3.1.89) for the mean and covariance of the Gaussian mixture. A Gaussian approximation of each individual posterior $p(\mathbf{x}_k|\epsilon_k^i, \boldsymbol{\mathcal{Z}}_{1:k})$ is estimated by separate SRFs (using the same prior $\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1})$).

The measurement validation procedure is performed in a manner slightly different from the standard method given by (3.1.86). The implicit measurement process in the SRF algorithm is in terms of augmented measurements (representing displacement) rather than bearings measurements (i.e. (θ, α) angles). The resulting measurement likelihood function (3.1.76) is non-Gaussian and is not expressible in closed form (although it is computable, by means of the standard normal cumulative distribution). Consequently, an analytical likelihood-ratio test leading to the ellipsoidal gating of measurements, is not feasible.

An ad-hoc approach to gating augmented measurements would be to generate a Gaussian approximation to the measurement likelihood and carrying out the standard hypothesis test as explained above. The moments of this Gaussian measurement likeli-

hood can be approximated either using a linearized transformation about the predicted bearings measurement (in the same manner as in an EKF), or by means of the unscented transform. However, both these methods introduce errors in the predicted density of the measurements.

A more exact, but computationally onerous, approach can be taken, by evaluating the likelihoods corresponding to all available measurements using the bearings density defined in (3.1.76) (which is parameterized by the mean $\hat{\mathbf{y}}_{k|k-1}$ and covariance $\mathbf{V}_k$ of the *augmented* measurement). Using the fact that (3.1.76) is normalized, a heuristic rejection scheme can be employed, whereby measurements with likelihood below a threshold P_{th} are discarded while those with likelihood greater than P_{th} are validated. A simple strategy to select a value for P_{th} is by recognizing that, for measurements in 3D space, the cumulative density 'rejected' by the thresholding is bounded from above by $2\pi^2 P_{th}$ (the domain of azimuth angles being $[-\pi, \pi]$ and that of elevation angles being $[-\pi/2, \pi/2]$). Equating this upper bound to the cumulative density $1 - P_G$ 'rejected' by the ellipsoidal gating, we arrive at $P_{th} = (2\pi^2)^{-1}(1 - P_G)$. The probabilities β_k^i can then be evaluated as in (3.1.95), substituting P_G with P_{th}.

Single-Measurement Probabilistic Data Association

Having presented the 'general' approach, we now introduce an alternative simplified framework for modelling false alarms and less than unity probability of detection. This framework (appearing in [18,19]) is henceforth used in all simulations, without loss of generality. We stress, however, that rather than providing a highly accurate characterization of the physics of real-world applications, this framework is taken to be a viable, albeit stripped-down substitute, which nonetheless incorporates the salient features of, and related estimation challenges arising from, the previously described more realistic model. Whereas in PDA, multiple (validated) returns are processed at each time step, we consider here a simpler approach whereby at each time step, only one measurement is available, given by:

$$\mathbf{z}_k \;=\; (1 - D_k)\underline{\psi}_k + D_k \mathbf{c}_k, \tag{3.1.96}$$

where the $(d-1)$-vector process $\{\underline{\psi}_k\}$ is the idealized measurement process which would result if there were no clutter, the $(d-1)$-vector process $\{\mathbf{c}_k\}$ is the false measurement of the azimuthal and elevation bearings that arises from clutter and $\{D_k\}$ is a binary process taking the value 0 if the measurement is target originated and 1 if it is due to clutter. The prior probability of clutter p_c is thus defined as

$$P\{D_k = 1\} \;=\; p_c \qquad \text{and} \qquad P\{D_k = 0\} \;=\; 1 - p_c. \tag{3.1.97}$$

It is assumed that clutter is uniformly distributed in all directions in 3D Cartesian space. This is equivalent, from an angular measurement point of view, to assuming that clutter is uniformly distributed on the unit sphere. For given azimuth and elevation angles θ and α, the area δA swept on the unit sphere by angle changes $\delta\theta$ and $\delta\alpha$ can be expressed, using the Jacobian of the spherical to Cartesian coordinate transformation (3.1.33), as $\cos(\alpha)\,\delta\theta\delta\alpha$. It then follows that the clutter density in (θ, α) coordinates is

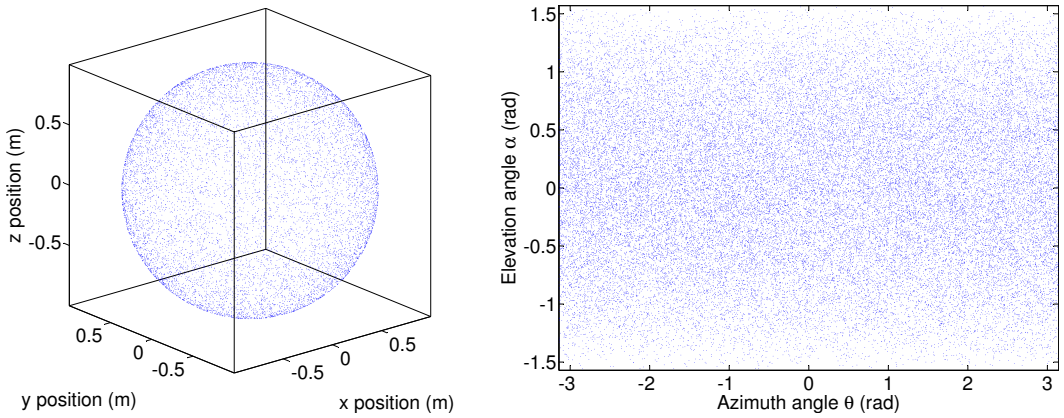
$$\begin{aligned} p(\mathbf{c} = (\theta, \alpha))\,\delta\theta\delta\alpha \;&= (4\pi)^{-1}\delta A \\ &= (4\pi)^{-1}\cos(\alpha)\,\delta\theta\delta\alpha, \end{aligned} \tag{3.1.98}$$

from which we conclude that

$$p(\mathbf{c} = (\theta, \alpha)) \;= (4\pi)^{-1}\cos(\alpha). \tag{3.1.99}$$

This non-uniform density of clutter in the measurement space, in 3D, is at odds with its uniform equivalent, in 2D, for which the corresponding Jacobian term (3.1.32) on the unit circle would have been unity. The correlation between θ and α, introduced in the density (3.1.99), can be interpreted as a necessary scaling ensuring that the density of clutter does not increase as α departs from 0 (i.e. the equator). Figure 3.7a illustrates the uniform distribution of the projection of clutter onto the unit sphere, while figure 3.7b shows the corresponding non-uniform joint distribution of azimuth and elevation angles.

Taking the joint distribution of angular measurements of clutter to be uniform in the measurement space $([-\pi\ \pi], [-\pi/2\ \pi/2])$, as depicted in figure 3.8b, results in a *non-uniform* distribution of its projection onto the unit sphere. This is shown in figure 3.8a,

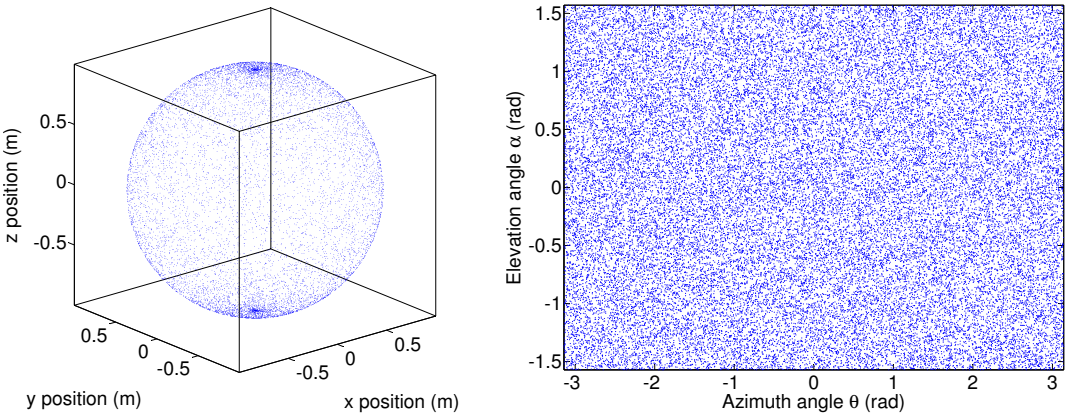


(a) Uniform spacial distribution of clutter projected onto the unit sphere

(b) Non-uniform distribution of clutter in the measurement space

Figure 3.7: Correct modelling of the bearings density of clutter

where the higher probability mass near the poles is a consequence of the uniform (marginal) distribution of elevation angles of clutter, in the measurement space.



(a) Non-uniform spacial distribution of clutter projected onto the unit sphere

(b) Uniform distribution of clutter in the measurement space

Figure 3.8: Incorrect modelling of the bearings density of clutter

Following the clutter measurement model given by (3.1.96), the filtered state pdf can be expressed as:

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{Z}_k) &= p(\mathbf{x}_k|D_k=0,\mathbf{Z}_k)P\{D_k=0|\mathbf{Z}_k\} + p(\mathbf{x}_k|D_k=1,\mathbf{Z}_k)P\{D_k=1|\mathbf{Z}_k\} \\ &= p(\mathbf{x}_k|\mathbf{z}_k=\underline{\psi}_k,\mathbf{Z}_{k-1})P\{D_k=0|\mathbf{Z}_k\} + p(\mathbf{x}_k|\mathbf{Z}_{k-1})P\{D_k=1|\mathbf{Z}_k\}, \quad (3.1.100) \end{aligned}$$

where the probabilities of the events 'no-clutter' and 'clutter', given all measurements up to time k can be expressed using Bayes' rule as

$$\begin{aligned} P\{D_k = 0|\mathbf{Z}_k\} &= c^{-1} p(\mathbf{z}_k|D_k = 0, \mathbf{Z}_{k-1}) P\{D_k = 0|\mathbf{Z}_{k-1}\} \\ &= c^{-1} p(\underline{\psi}_k|\mathbf{Z}_{k-1})(1 - p_c) \qquad (3.1.101) \\ P\{D_k = 1|\mathbf{Z}_k\} &= c^{-1} p(\mathbf{z}_k|D_k = 1, \mathbf{Z}_{k-1}) P\{D_k = 1|\mathbf{Z}_{k-1}\} \\ &= c^{-1} p(\mathbf{c}_k) p_c, \qquad (3.1.102) \end{aligned}$$

and the normalizing constant c is given by

$$\begin{aligned} c &= p(\mathbf{z}_k|\mathbf{Z}_{k-1}) \\ &= p(\underline{\psi}_k|\mathbf{Z}_{k-1})(1 - p_c) + p(\mathbf{c}_k) p_c. \qquad (3.1.103) \end{aligned}$$

This simplified alternative to the PDA framework implicitly combines the possibility of non-detection (i.e. $P_D < 1$) and the presence of clutter within the single quantity p_c. We can see that the filtered pdf (3.1.100) is, similarly to that in the standard PDA algorithm, a Gaussian mixture. However, it is made up of only two densities: the joint posterior density of the state and of the measurement being 'true' and that of the state and of the measurement being clutter. As will be explained in section 4.2, this simpler modelling of clutter allows for a convenient derivation of the posterior Cramer-Rao lower bound for the state estimation error.

Implementation of the 'single-measurement' PDA-SRF The 'single-measurement' PDA-SRF can be readily implemented from equations (3.1.100), (3.1.101) and (3.1.102). The density $p(\underline{\psi}_k|\mathbf{Z}_{k-1})$ can be recognized as the augmented measurement likelihood function, which is given by (3.1.76), and is parameterized by the mean $\hat{\mathbf{y}}_{k|k-1}$ and covariance $\mathbf{V}_k$ of the *augmented* measurement.

3.1.3 Extension to Multiple-Model Maneuvering Target Tracking

As outlined in section 2.4.4, the problem of multiple-model maneuvering target tracking requires suboptimal estimation techniques in order to limit the exponential growth of hypotheses of mode histories. In many applications, the IMM algorithm is used for its

simplicity and its combination of accuracy and computational efficiency. We propose here a different approach, based on work carried out in [64], but adapted for the tracking of maneuvering targets in 3D.

The Shifted Rayleigh Mixture Filter

A direct approach to limiting the number of mixture components representing the state posterior is through the use of 'mixture reduction' algorithms. The shifted Rayleigh mixture filter (SRMF) ([18]) employs the same basic structure as the multiple-model estimator outlined in equations (2.4.29)-(2.4.30), but with a mixture reduction step at the end of each iteration to restrict the number of mixture components to a predefined number N. The posterior state density computed by the SRMF at time $k-1$ can be expressed as:

$$p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1}) \;=\; \sum_{n=1}^{N} w_{k-1}^{n}\,\mathcal{N}(\mathbf{x}_{k-1};\hat{\mathbf{x}}_{k-1|k-1}^{n},\mathbf{P}_{k-1|k-1}^{n}), \tag{3.1.104}$$

where the (positive) weights w_{k-1}^{n} add up to unity. The joint state and mode density is given by:

$$p(\mathbf{x}_{k-1}, r_{k-1}=i|\mathbf{Z}_{k-1}) \;=\; \sum_{n\in\{m:r_{k-1}^{m}=i\}} w_{k-1}^{n}\,\mathcal{N}(\mathbf{x}_{k-1};\hat{\mathbf{x}}_{k-1|k-1}^{n},\mathbf{P}_{k-1|k-1}^{n}), \tag{3.1.105}$$

where the summation over the index n regroups all weighted Gaussian mixture components whose associated mode r_{k-1} is i. This mixture representation is such that the conditional probability of the ith mode being in effect at time $k-1$ is:

$$p(r_{k-1}=i|\mathbf{Z}_{k-1}) \;=\; \sum_{n\in\{m:r_{k-1}^{m}=i\}} w_{k-1}^{n}. \tag{3.1.106}$$

We now introduce, for convenience, a random variable ν taking values in the set $S = \{1,\ldots,N\}$, and which denotes the index of a given mixture component at time $k-1$. The joint posterior density of the state and mode value at time k can then be written as:

$$\begin{aligned} p(\mathbf{x}_k, r_k=j|\mathbf{Z}_k) \;&=\; \sum_{n=1}^{N} p(\mathbf{x}_k, r_k=j, \nu=n|\mathbf{Z}_k) \\ &= \sum_{n=1}^{N} p(\mathbf{x}_k, r_k=j, \nu=n|\mathbf{z}_k) \end{aligned}$$

$$= \sum_{n=1}^{N} p(\mathbf{x}_k|\mathbf{z}_k, r_k = j, \nu = n)\mu_k^{nj}, \tag{3.1.107}$$

where

$$\begin{aligned}
\mu_k^{nj} &= p(r_k = j, \nu = n|\mathbf{z}_k) \\
&= \left(\frac{1}{c}\right) p(\mathbf{z}_k, r_k = j, \nu = n) \\
&= \left(\frac{1}{c}\right) p(\mathbf{z}_k|r_k = j, \nu = n)p(r_k = j|\nu = n)p(\nu = n) \\
&= \left(\frac{1}{c}\right) w_{k-1}^n \pi_{r_{k-1}^n, j}\, p(\mathbf{z}_k|r_k = j, \nu = n)
\end{aligned} \tag{3.1.108}$$

represents the weight of the mixture component representing the transition to the jth mode from the prior component of index n. The normalizing constant c is $p(\mathbf{z}_k)$, while the conditional measurement likelihood is

$$\begin{aligned}
p(\mathbf{z}_k|r_k = j, \nu = n) &= \int p(\mathbf{z}_k, \mathbf{x}_{k-1}|r_k = j, \nu = n)\, d\mathbf{x}_{k-1} \\
&= \int p(\mathbf{z}_k|r_k = j, \mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1}|\nu = n)\, d\mathbf{x}_{k-1} \\
&= \int p(\mathbf{z}_k|r_k = j, \mathbf{x}_{k-1})\, \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^n, \mathbf{P}_{k-1|k-1}^n)\, d\mathbf{x}_{k-1} \\
&= p(\mathbf{z}_k|r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^n, \mathbf{P}_{k-1|k-1}^n)).
\end{aligned} \tag{3.1.109}$$

can be evaluated using (3.1.76), by additionally conditioning on the mode $r_k = j$. The density $p(\mathbf{x}_k|\mathbf{z}_k, r_k = j, \nu = n)$ in (3.1.107) is evaluated using the SRF equations (3.1.5) to (3.1.13) by conditioning on the nth Gaussian state prior and the current mode:

$$\begin{aligned}
p(\mathbf{x}_k|\mathbf{z}_k, r_k = j, \nu = n) &= \int p(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{z}_k, r_k = j, \nu = n)\, d\mathbf{x}_{k-1} \\
&= \int p(\mathbf{x}_k|\mathbf{z}_k, r_k = j, \mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1}|\nu = n)\, d\mathbf{x}_{k-1} \\
&= \int p(\mathbf{x}_k|\mathbf{z}_k, r_k = j, \mathbf{x}_{k-1})\, \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^n, \mathbf{P}_{k-1|k-1}^n)\, d\mathbf{x}_{k-1} \\
&= p(\mathbf{x}_k|\mathbf{z}_k, r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^n, \mathbf{P}_{k-1|k-1}^n)).
\end{aligned} \tag{3.1.110}$$

The filtered density of the state at time k can then finally be written as

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{Z}_k) &= \sum_{j=1}^{M} p(\mathbf{x}_k, r_k = j|\mathbf{Z}_k) \\ &= \sum_{n=1}^{N}\sum_{j=1}^{M} p(\mathbf{x}_k, r_k = j, \nu = n|\mathbf{z}_k) \\ &= \sum_{n=1}^{N}\sum_{j=1}^{M} p(\mathbf{x}_k|\mathbf{z}_k, r_k = j, \nu = n)\, p(r_k = j, \nu = n|\mathbf{z}_k) \\ &= \sum_{n=1}^{N}\sum_{j=1}^{M} p(\mathbf{x}_k|\mathbf{z}_k, r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^n_{k-1|k-1}, \mathbf{P}^n_{k-1|k-1}))\mu^{nj}_k, \end{aligned} \quad (3.1.111)$$

and the corresponding mean and covariance computed as

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \sum_{n=1}^{N}\sum_{j=1}^{M} \mu^{nj}_k\, \hat{\mathbf{x}}^{nj}_{k|k} \\ \mathbf{P}_{k|k} &= \sum_{n=1}^{N}\sum_{j=1}^{M} \mu^{nj}_k \mathbf{P}^{nj}_{k|k} + \sum_{n=1}^{N}\sum_{j=1}^{M} \mu^{nj}_k \left(\hat{\mathbf{x}}^{nj}_{k|k} - \hat{\mathbf{x}}_{k|k}\right)\left(\hat{\mathbf{x}}^{nj}_{k|k} - \hat{\mathbf{x}}_{k|k}\right)^{\mathrm{T}}, \end{aligned} \quad (3.1.112)$$

where $\hat{\mathbf{x}}^{nj}_{k|k}$ and $\mathbf{P}^{nj}_{k|k}$ are the SRF state mean and covariance estimates conditioned on the jth mode hypothesis given the nth Gaussian prior.

Equation (3.1.111) indicates that at each iteration the SRMF outputs from an N-fold Gaussian mixture prior an $M \times N$-fold Gaussian mixture approximation of the posterior. In other words, for each one of the N mode-conditioned priors, every mode-conditioned posterior hypothesis is evaluated (with its corresponding normalized weight μ^{nj}_k). Consequently, reduction of the $M \times N$-fold Gaussian mixture into an N-fold approximation is necessary ahead of the subsequent iteration of the SRMF.

Various schemes exist to maintain the size of a mixture distribution at a constant level. Their common objective is to approximate mixture densities using fewer components, while minimizing the loss of statistical infromation. They range from sophisticated methods aiming to minimize various measures of distance (such as the Integrated Squared Distance [83], the Kullback-Leibler divergence [67] and Mahalanobis distance [68]) between the 'original' and 'reduced' mixtures to simpler 'weight-based' mixture reduction

algorithms. The latter class of methods is based on the assumption that weights represent an adequate measure of information contained by the components. By forgoing the information carried by the statistics of the individual components of the original density, these methods achieve higher computational efficiency, but at the cost of a higher loss of information.

While comparison of these methods is beyond the scope of this thesis, we present next two 'weight-based' mixture reduction algorithms, namely the DEA (detection-estimation algorithm, [78], [79]) and the 'Fearnhead-Clifford' algorithm ([32], [33]).

The Detection-Estimation Algorithm This is a purely deterministic approach to mixture reduction, based on the propagation of the N components with the highest normalized weights. As such, the DEA algorithm can be viewed as a MAP approach, whereby only the components with the maximum a posteriori probabilities get selected. Given an M-fold mixture with weights μ_k^m, the steps in the selection of components to be kept are:

- Re-arrange the set of weights μ_k^m in decreasing order of magnitude - call this set $\bar{\mu}_k^{l_m}$,
- Find the original indices i, for $i = 1, \ldots, N$, of the first N re-arranged weights $\bar{\mu}_k^{l_m}$, such that $i = l_m$
- Select the mixture components with indices i and normalize the N selected weights w_k^i such that $\bar{\mu}_k^i = \frac{\bar{\mu}_k^i}{\sum_{j=1}^N \bar{\mu}_k^j}$

The Fearnhead-Clifford Algorithm The algorithm proposed in [32] and [33] by Fearnhead and Clifford is a stochastic weight-based mixture reduction technique. It has the desirable property of producing an unbiased approximation of a probability mass function μ_k^i of finite support M by a probability mass function w_k^i of support $N < M$, while minimizing the expected L^2 error $E\left(\sum_{i=1}^M (w_k^i - \mu_k^i)^2\right)$ between them. This is achieved through the following steps:

- Calculate the unique solution C_k of: $N = \sum_{i=1}^M \min(C_k \mu_k^i, 1)$,
- Form the sets of mixture component indices: $\mathcal{A}_1 = \{i | \mu_k^i \geq 1/C_k\}$ of size G and $\mathcal{A}_2 = \{i | \mu_k^i < 1/C_k\}$ of size $M - G$,

- Use systematic sampling to form a set $\mathcal{A}_3$ of $N-G$ samples from the set $\mathcal{A}_2$, with weights $\bar{w}_k^i$ normalized such that $\bar{w}_k^i = \frac{\mu_k^i}{\sum_{j\in\mathcal{A}_2}\mu_k^j}$,
- Form the reduced mixture by selecting all G components indexed by $\mathcal{A}_1$, with their original weights, and those indexed by $\mathcal{A}_3$ with weights $\bar{w}_k^i$.

The main characteristic of this algorithm is the 'optimal' combination of deterministic sampling (of all weights exceeding the threshold) and stochastic sampling (of all weights falling below the threshold), resulting in an unbiased selection scheme. The use of systematic sampling, moreover, ensures that no sample from set $\mathcal{A}_2$ is selected more than once, thereby avoiding 'cloning' of samples. The reader is referred to [32], [33] and [64] for a more detailed analysis of the algorithm.

We now present extensions of the SRMF for the treatment of clutter, for both the general and 'single-measurement' models of clutter.

Implementation of the PDA-SRMF

The SRMF can be extended to take into account multiple simultaneous measurements involving clutter. In this context, the state posterior can then be expressed (similarly to (3.1.111)) as:

$$p(\mathbf{x}_k|\boldsymbol{\mathcal{Z}}_k) = \sum_{n=1}^{N}\sum_{j=1}^{M} p(\mathbf{x}_k|\boldsymbol{\mathcal{Z}}_k, r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^n, \mathbf{P}_{k-1|k-1}^n))\mu_k^{nj}, \quad (3.1.113)$$

where the state pdf conditional on the current mode and prior mixture component is evaluated through the PDA-SRF equations and approximated by a matched Gaussian density. More precisely, this density can be written as:

$$\begin{aligned} p(\mathbf{x}_k|\boldsymbol{\mathcal{Z}}_k, r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^n, \mathbf{P}_{k-1|k-1}^n)) = \\ \sum_{i=0}^{m(k)} p(\mathbf{x}_k|\epsilon_k^i, \boldsymbol{\mathcal{Z}}_k, r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^n, \mathbf{P}_{k-1|k-1}^n))\beta_k^{i,(nj)}. \end{aligned} \quad (3.1.114)$$

The measurement association probabilities (conditional on the current mode j and on the prior mixture component of index n) can, in turn, be expressed as:

$$\beta_k^{i,(nj)} = \begin{cases} \frac{\tilde{p}^{nj}(z_k^i)}{(1-P_D P_G)P_D^{-1}\lambda+\sum_{l=1}^{m(k)}\tilde{p}^{nj}(z_k^l)}, & i=1,\ldots,m(k) \\ \frac{(1-P_D P_G)P_D^{-1}\lambda}{(1-P_D P_G)P_D^{-1}\lambda+\sum_{l=1}^{m(k)}\tilde{p}^{nj}(z_k^l)}, & i=0, \end{cases} \tag{3.1.115}$$

using the short-hand notation for the measurement likelihood (conditioned on the current mode j and on the prior mixture component of index n) defined as:

$$\tilde{p}^{nj}(z_k^i) = p(\mathbf{z}_k^i | r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^n, \mathbf{P}_{k-1|k-1}^n)). \tag{3.1.116}$$

and which can be evaluated using the predicted bearing density (3.1.76), by additionally conditioning on the current mode $r_k = j$.

Finally, the mixture weights (representing the joint posterior probabilities of the mode and index of the prior mixture components) can be written as

$$\begin{aligned} \mu_k^{nj} &= \left(\frac{1}{c}\right) w_{k-1}^n \pi_{r_{k-1}^n,j}\, p(\boldsymbol{\mathcal{Z}}_k | r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^n, \mathbf{P}_{k-1|k-1}^n)) \\ &= \left(\frac{1}{c}\right) w_{k-1}^n \pi_{r_{k-1}^n,j} \sum_{i=0}^{m(k)} p(\boldsymbol{\mathcal{Z}}_k | \epsilon_k^i, r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^n, \mathbf{P}_{k-1|k-1}^n))\, P\{\epsilon_k^i | m(k)\} \\ &= \left(\frac{1}{c}\right) w_{k-1}^n \pi_{r_{k-1}^n,j}\, V_k^{-m(k)} \frac{1 - P_D P_G + P_D/\lambda + \sum_{i=1}^{m(k)} \tilde{p}^{nj}(z_k^i)}{1 - P_D P_G + m(k) P_D P_G/(\lambda V_k)}, \end{aligned} \tag{3.1.117}$$

where use was made of relations (3.1.91) and (3.1.94).

Implementation of the 'single-measurement' PDA-SRMF

Clutter, in the 'single-measurement' form, can be accommodated into the SRMF algorithm, in a manner similar to the PDA extension. The state posterior, in this case, can be expressed as in (3.1.111), where the posterior state density conditioned on the current

mode and prior mixture component can be expanded as:

$$\begin{aligned}
&p(\mathbf{x}_k|\mathbf{z}_k, r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^n_{k-1|k-1}, \mathbf{P}^n_{k-1|k-1})) \\
&\quad = p(\mathbf{x}_k|\mathbf{z}_k = \underline{\psi}_k, r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^n_{k-1|k-1}, \mathbf{P}^n_{k-1|k-1}))P\{D_K = 0|\mathbf{Z}_k\}+ \\
&\qquad p(\mathbf{x}_k|r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^n_{k-1|k-1}, \mathbf{P}^n_{k-1|k-1}))P\{D_k = 1|\mathbf{Z}_k\} \\
&\quad = (c^{nj})^{-1}\Big[(1-p_c)\, p(\mathbf{x}_k|\mathbf{z}_k = \underline{\psi}_k, r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^n_{k-1|k-1}, \mathbf{P}^n_{k-1|k-1}))\times \\
&\qquad p(\underline{\psi}_k|r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^n_{k-1|k-1}, \mathbf{P}^n_{k-1|k-1}))+ \\
&\qquad p_c\, p(\mathbf{x}_k|r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^n_{k-1|k-1}, \mathbf{P}^n_{k-1|k-1}))p(\mathbf{c}_k)\Big],
\end{aligned} \tag{3.1.118}$$

where the normalizing constant c^{nj} is

$$\begin{aligned}
c^{nj} &= p(\mathbf{z}_k|r_k = j, \nu = n) \\
&= (1-p_c)\, p(\underline{\psi}_k|r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^n_{k-1|k-1}, \mathbf{P}^n_{k-1|k-1})) + p_c\, p(\mathbf{c}_t).
\end{aligned} \tag{3.1.119}$$

Also, the mixture weights can be expressed as

$$\begin{aligned}
\mu^{nj}_k &= \left(\frac{1}{c}\right) w^n_{k-1}\pi_{r^n_{k-1},j}\, p(\mathbf{z}_k|r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^n_{k-1|k-1}, \mathbf{P}^n_{k-1|k-1})) \\
&= \left(\frac{1}{c}\right) w^n_{k-1}\pi_{r^n_{k-1},j}\Big[(1-p_c)\, p(\underline{\psi}_k|r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^n_{k-1|k-1}, \mathbf{P}^n_{k-1|k-1}))+ \\
&\qquad p_c\, p(\mathbf{c}_k)\Big].
\end{aligned} \tag{3.1.120}$$

The measurement likelihood $p(\underline{\psi}_k|r_k = j, \mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^n_{k-1|k-1}, \mathbf{P}^n_{k-1|k-1}))$ can be evaluated using the predicted bearings density formula (3.1.76), by conditioning on the current mode value, while the density of clutter is given by (3.1.99).

3.2 The Rauch-Tung-Striebel Shifted Rayleigh Smoother

We concentrate in this section on optimal smoothing for the class of systems with linear/Gaussian dynamics. It is closely related to filtering and addresses the problem of refining past state estimates, given measurements accumulated up to the present time. Among its various areas of application, air traffic control and missile tracking are the most notable in the context of target tracking.

The general premise is the evaluation of $\hat{\mathbf{x}}_{k|N}$ and $\mathbf{P}_{k|N}$, given all the measurements from time 0 to time N (with $0 \leq k \leq N$). Three different frameworks for smoothing can be derived:

Fixed point smoothing: $\hat{\mathbf{x}}_{k|N}$ is estimated for fixed k and increasing N,

Fixed lag smoothing: $\hat{\mathbf{x}}_{N-\Delta|N}$ is estimated for fixed Δ and increasing N,

Fixed interval smoothing: $\hat{\mathbf{x}}_{k|N}$ is estimated for fixed N and $k = N, N-1, \cdots, 0$,

We are interested in the problem of fixed interval smoothing, which is the most commonly studied (see [36] for applications of fixed point and fixed lag smoothing). Fixed-interval smoothing is commonly presented, as in [35], as the maximum likelihood combination of independent state estimates from forward and backward filtering passes. This formulation, however, involves inverting the forward system dynamics, to be used in the backward filtering. As pointed out in [55], this does not always result in the correct backward transition. In contrast, Rauch, Tung and Striebel proposed in [62] an equivalent and arguably more intuitive forward/backward smoothing framework which obviates the need to invert system dynamics.

We present here a derivation of the Rauch-Tung-Striebel (RTS) smoothed densities

$$p(\mathbf{x}_k \mid \mathbf{Z}_N) \;=\; \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|1:N}, \mathbf{P}_{k|1:N}), \qquad k = 1, \ldots, N \tag{3.2.1}$$

for the problem of bearings-only tracking, where the state and measurement processes are described by

$$\begin{aligned} \mathbf{x}_k &= \mathbf{F}\mathbf{x}_{k-1} + \mathbf{u}^s_{k-1} + \mathbf{v}_{k-1} \\ \mathbf{d}_k &= \mathbf{H}\mathbf{x}_k + \mathbf{u}^m_k \\ \mathbf{z}_k &= h(\mathbf{d}_k) + \tilde{\mathbf{w}}_{\mathbf{d}_k}, \end{aligned}$$

with $h(\mathbf{d}_k)$ and $\tilde{\mathbf{w}}_{\mathbf{d}_k}$ as defined in (3.1.78) and (3.1.83).

The filtered densities $p(\mathbf{x}_k \mid \mathbf{Z}_k) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k, \mathbf{P}_k)$ are computed for $k = 1, \ldots, N$

using the SRF equations (3.1.5) to (3.1.13). To derive the parameters $\hat{\mathbf{x}}_{k|1:N}$, $\mathbf{P}_{k|1:N}$ of the smoothed densities, we note that:

$$p(\mathbf{x}_k \mid \mathbf{x}_{k+1}, \mathbf{Z}_N) \;=\; p(\mathbf{x}_k \mid \mathbf{x}_{k+1}, \mathbf{Z}_k), \tag{3.2.2}$$

(This follows from the the Markov structure of the model, which implies that

$$p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}, \mathbf{x}_k, \mathbf{Z}_k) = p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}),$$

i.e. $\mathbf{x}_k$ and $\mathbf{Z}_k$ are independent of $\mathbf{z}_{k+1}$, given $\mathbf{x}_{k+1}$.)

As a result, we can write:

$$\begin{aligned} p\left(\mathbf{x}_k \mid \mathbf{Z}_N\right) \;&=\; \int p\left(\mathbf{x}_k \mid \mathbf{x}_{k+1}, \mathbf{Z}_N\right) p\left(\mathbf{x}_{k+1} \mid \mathbf{Z}_N\right)\, d\mathbf{x}_{k+1} \\ &= \int p\left(\mathbf{x}_k \mid \mathbf{x}_{k+1}, \mathbf{Z}_k\right) p\left(\mathbf{x}_{k+1} \mid \mathbf{Z}_N\right)\, d\mathbf{x}_{k+1} \end{aligned} \tag{3.2.3}$$

To calculate $p\left(\mathbf{x}_k \mid \mathbf{x}_{k+1}, \mathbf{Z}_k\right)$ we use the facts that

$$E\left[\mathbf{x}_k \mid \mathbf{Z}_k\right] \;=\; \hat{\mathbf{x}}_{k|k}, \qquad E\left[\mathbf{x}_{k+1} \mid \mathbf{Z}_k\right] \;=\; \mathbf{F}\hat{\mathbf{x}}_{k|k} + \mathbf{u}_k^s$$

and

$$\mathrm{cov}\{\mathbf{x}_k, \mathbf{x}_{k+1} \mid \mathbf{Z}_k\} \;=\; \mathbf{P}_{k|k}\mathbf{F}^{\mathrm{T}}, \qquad \mathrm{cov}\{\mathbf{x}_{k+1} \mid \mathbf{Z}_k\} = \mathbf{F}\mathbf{P}_{k|k}\mathbf{F}^{\mathrm{T}} + \mathbf{Q}^s.$$

It follows from the standard solution to the linear least-squares estimation problem, in which $\mathbf{x}_{k+1}$ is interpreted as a measurement for $\mathbf{x}_k$, that

$$p\left(\mathbf{x}_k \mid \mathbf{x}_{k+1}, \mathbf{z}_k\right) \;=\; \mathcal{N}\left(\mathbf{x}_k; \hat{\mathbf{x}}_k + \tilde{K}_k(\mathbf{x}_{k+1} - (\mathbf{F}\hat{\mathbf{x}}_k + \mathbf{u}_k^s)), \tilde{\mathbf{P}}_k\right), \tag{3.2.4}$$

where

$$\tilde{\mathbf{K}}_k \;= \mathbf{P}_{k|k}\mathbf{F}^{\mathrm{T}}\left[\mathbf{F}\mathbf{P}_{k|k}\mathbf{F}^{\mathrm{T}} + \mathbf{Q}^s\right]^{-1} \tag{3.2.5}$$

$$\tilde{\mathbf{P}}_k \;= \mathbf{P}_{k|k} - \mathbf{P}_{k|k}\mathbf{F}^{\mathrm{T}}\left[\mathbf{F}\mathbf{P}_{k|k}\mathbf{F}^{\mathrm{T}} + \mathbf{Q}^s\right]^{-1}\mathbf{F}\mathbf{P}_{k|k} \tag{3.2.6}$$

From (3.2.3), then

$$\mathcal{N}\left(\mathbf{x}_k; \hat{\mathbf{x}}_{k|1:N}, \mathbf{P}_{k|1:N}\right) = \int \mathcal{N}\left(\mathbf{x}_k; \hat{\mathbf{x}}_k + \tilde{\mathbf{K}}_k(\mathbf{x}_{k+1} - (\mathbf{F}\hat{\mathbf{x}}_k + \mathbf{u}_k^s)), \tilde{\mathbf{P}}_k\right) \times \mathcal{N}\left(\mathbf{x}_{k+1}; \hat{\mathbf{x}}_{k+1|1:N}, \mathbf{P}_{k+1|1:N}\right) d\mathbf{x}_{k+1}. \tag{3.2.7}$$

Using the identity

$$\int \mathcal{N}\left(\mathbf{x}; \mathbf{a} + \mathbf{G}\mathbf{x}', \mathbf{P}_1\right) \times \mathcal{N}\left(\mathbf{x}'; \mathbf{b}, \mathbf{P}_2\right) d\mathbf{x}' = \mathcal{N}\left(\mathbf{x}; \mathbf{a} + \mathbf{G}\mathbf{b},\ \mathbf{P}_1 + \mathbf{G}\mathbf{P}_2\mathbf{G}^{\mathrm{T}}\right),$$

we arrive at the following expressions for the mean and covariance of the smoothed densities:

$$\hat{\mathbf{x}}_{k|1:N} = \hat{\mathbf{x}}_k + \tilde{\mathbf{K}}_k\left(\hat{\mathbf{x}}_{k+1|1:N} - (\mathbf{F}\hat{\mathbf{x}}_k + \mathbf{u}_k^s)\right) \tag{3.2.8}$$

$$\mathbf{P}_{k|1:N} = \tilde{\mathbf{P}}_k + \tilde{\mathbf{K}}_k\mathbf{P}_{k+1|1:N}\tilde{\mathbf{K}}_k^{\mathrm{T}} \tag{3.2.9}$$

which can be evaluated recursively for $k = N-1, \ldots, 1$.

The RTS shifted Rayleigh smoother (RTS-SRS) just described can be readily adapted to handle clutter in the measurements. For both PDA and 'single-measurement PDA' modifications, because the filtered densities are evaluated by moment matching, no change needs to be made to the backward filtering algorithm.

Chapter 4

3D Bearings-Only Tracking Applications

We present in this chapter simulation based assessments of the shifted Rayleigh filter, in relation to competing algorithms, for various 3D tracking scenarios. These have been chosen specifically to challenge tracker robustness and are therefore substantially different from usual 'benign' scenarios that are mostly covered in the literature. More specifically, they aim to test resilience to poor state initialization, sudden changes in bearings measurements (or high bearings-rate), various densities of clutter in the measurements and multiple system dynamic models. Such levels of nonlinearity in the measurements and/or system dynamics are representative of demanding applications such as the tracking of agile military aircraft or missiles in the presence of electronic counter-measures.

4.1 Conduct of Simulations

For each tracking scenario, fair comparison of tracking algorithms requires Monte-Carlo (MC) simulations to be conducted in a systematic fashion. In the following studies, we do so by carrying out simulations runs over different realizations of measurement noise and clutter, while the 'true' state trajectory of the target is held constant. Also, initialization is done using the same procedure for all competing tracking algorithms. Finally, comparisons are made with respect to several measures capturing various aspects of their performance.

4.1.1 Filter Initialization

For fair comparison, all filters are initialized with the same method (see [63]) for each Monte-Carlo simulation run, unless otherwise stated. This is achieved, given a prior on the (relative) target range and possibly on the speed and heading, through a linearized

transformation about the first bearings measurement, which is allowed to be clutter free. More precisely, taking the initial range prior to be $\hat{r}_0 \sim \mathcal{N}(\hat{r}_0; \bar{r}_0, \sigma^2_{r_0})$ and the initial bearings measurement to be

$$\begin{bmatrix} \theta_0 \\ \alpha_0 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \theta_0 \\ \alpha_0 \end{bmatrix}; \begin{bmatrix} \theta_0^{true} \\ \alpha_0^{true} \end{bmatrix}, \begin{bmatrix} \sigma^2_\theta & 0 \\ 0 & \sigma^2_\alpha \end{bmatrix} \right),$$

where $\begin{bmatrix} \theta_0^{true} & \alpha_0^{true} \end{bmatrix}^{\mathrm{T}}$ is the true initial bearings vector, the nonlinear spherical to Cartesian transformation of the target position can be approximated as $\mathbf{x}_0^{pos} \sim \mathcal{N}(\mathbf{x}_0^{pos}; \hat{\mathbf{x}}_{0|0}^{pos}, \mathbf{P}_{0|0}^{pos})$ with

$$\hat{\mathbf{x}}_{0|0}^{pos} = \begin{bmatrix} \hat{x}_{0|0} \\ \hat{y}_{0|0} \\ \hat{z}_{0|0} \end{bmatrix} = \begin{bmatrix} \bar{r}_0 \sin(\theta_0)\cos(\alpha_0) \\ \bar{r}_0 \cos(\theta_0)\cos(\alpha_0) \\ \bar{r}_0 \sin(\alpha_0) \end{bmatrix} + \begin{bmatrix} u_0^{d,1} \\ u_0^{d,2} \\ u_0^{d,3} \end{bmatrix} \tag{4.1.1}$$

and covariance

$$\mathbf{P}_{0|0}^{pos} = \begin{bmatrix} P_{0|0}^{xx} & P_{0|0}^{xy} & P_{0|0}^{xz} \\ P_{0|0}^{yx} & P_{0|0}^{yy} & P_{0|0}^{yz} \\ P_{0|0}^{zx} & P_{0|0}^{zy} & P_{0|0}^{zz} \end{bmatrix} = \nabla g_s(\bar{r}_0, \theta_0, \alpha_0)\, \mathrm{diag}(\sigma^2_{r_0}, \sigma^2_\theta, \sigma^2_\alpha)\, \nabla g_s(\bar{r}_0, \theta_0, \alpha_0)^{\mathrm{T}}, \tag{4.1.2}$$

where $\begin{bmatrix} u_0^{d,1} & u_0^{d,2} & u_0^{d,3} \end{bmatrix}$ is the (known) initial position of the sensor platform, and $\nabla g_s(\bar{r}_0, \theta_0, \alpha_0)$ is the Jacobian of the spherical to Cartesian transformation, given (as in (3.1.33)) by

$$\nabla g_s(\hat{r}_0, \theta_0, \alpha_0) = \begin{bmatrix} \sin(\theta_0)\cos(\alpha_0) & \bar{r}_0 \cos(\theta_0)\cos(\alpha_0) & -\bar{r}_0 \sin(\theta_0)\sin(\alpha_0) \\ \cos(\theta_0)\cos(\alpha_0) & -\bar{r}_0 \sin(\theta_0)\cos(\alpha_0) & -\bar{r}_0 \cos(\theta_0)\sin(\alpha_0) \\ \sin(\alpha_0) & 0 & \bar{r}_0 \cos(\alpha_0) \end{bmatrix}.$$

Additionally, if priors on the target speed s_0, heading c_0 and pitch p_0 are available in the form $\hat{s}_0 \sim N(\hat{s}_0; \bar{s}_0, \sigma^2_{s_0})$ and

$$\begin{bmatrix} c_0 \\ p_0 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} c_0 \\ p_0 \end{bmatrix}; \begin{bmatrix} \bar{c}_0 \\ \bar{p}_0 \end{bmatrix}, \begin{bmatrix} \sigma^2_c & 0 \\ 0 & \sigma^2_p \end{bmatrix} \right),$$

the velocity components of the state vector can be approximated as $\mathbf{x}_0^{vel} \sim \mathcal{N}(\mathbf{x}_0^{vel}; \hat{\mathbf{x}}_{0|0}^{vel}, \mathbf{P}_{0|0}^{vel})$, with

$$\hat{\mathbf{x}}_{0|0}^{vel} = \begin{bmatrix} \hat{\dot{x}}_{0|0} \\ \hat{\dot{y}}_{0|0} \\ \hat{\dot{z}}_{0|0} \end{bmatrix} = \begin{bmatrix} \bar{s}_0 \sin(c_0)\cos(p_0) \\ \bar{s}_0 \cos(c_0)\cos(p_0) \\ \bar{s}_0 \sin(p_0) \end{bmatrix} \tag{4.1.3}$$

and covariance

$$\mathbf{P}_{0|0}^{vel} = \begin{bmatrix} P_{0|0}^{\dot{x}\dot{x}} & P_{0|0}^{\dot{x}\dot{y}} & P_{0|0}^{\dot{x}\dot{z}} \\ P_{0|0}^{\dot{y}\dot{x}} & P_{0|0}^{\dot{y}\dot{y}} & P_{0|0}^{\dot{y}\dot{z}} \\ P_{0|0}^{\dot{z}\dot{x}} & P_{0|0}^{\dot{z}\dot{y}} & P_{0|0}^{\dot{z}\dot{z}} \end{bmatrix} = \nabla g_s(\bar{s}_0, c_0, p_0)\, \mathrm{diag}(\sigma_{s_0}^2, \sigma_c^2, \sigma_p^2)\, \nabla g_s(\bar{s}_0, c_0, p_0)^{\mathrm{T}}, \tag{4.1.4}$$

and

$$\nabla g_s(\bar{s}_0, c_0, p_0) = \begin{bmatrix} \sin(c_0)\cos(p_0) & \bar{s}_0 \cos(c_0)\cos(p_0) & -\bar{s}_0 \sin(c_0)\sin(p_0) \\ \cos(c_0)\cos(p_0) & -\bar{s}_0 \sin(c_0)\cos(p_0) & -\bar{s}_0 \cos(c_0)\sin(p_0) \\ \sin(p_0) & 0 & \bar{s}_0 \cos(p_0) \end{bmatrix}.$$

4.1.2 Comparison Metrics

Three metrics are used to evaluate and compare tracker performance. The first one is the root-mean-square error (RMSE) defined as

$$\mathrm{RMSE}(k) = \sqrt{\frac{1}{N}\sum_{n=1}^{N} \left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k,n}\right)^2},$$

where $\hat{\mathbf{x}}_{k|k,n}$ denotes the state estimate at time k for the nth Monte-Carlo run. Aside from individual state errors, of particular interest are the range and velocity errors which can be easily obtained from this definition. The second metric is the time-averaged root-mean-square error (TARMSE), computed as

$$\mathrm{TARMSE} = \frac{1}{t_{max} - l + 1} \sum_{k=l}^{t_{max}} \sqrt{\frac{1}{N}\sum_{n=1}^{N} \left(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k,n}\right)^2},$$

where l is the time index from which the averaging starts and t_{max} is the simulation time horizon. The final performance measure is the percentage of track divergence. Under the hypothesis of consistent filtering of a linear Gaussian system, the normalized estimation error squared defined as

$$\epsilon_k^2 = (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^{\mathrm{T}} \mathbf{P}_{k|k}^{-1} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})$$

should have a χ^2 distribution with degrees of freedom n_x equal to the dimension of $\mathbf{x}$. In the following simulations, track divergence is defined as the occurrence of either one of the following two hypotheses:

- the RMS range error exceeds a set threshold of D_{th} km for two consecutive time steps, or
- ϵ_k^2 exceeds the 99% probability concentration region of a $\chi_{n_x}^2$ variable for more than 6 time steps.

The number of time steps during which either the range errors exceed the threshold D_{th} or χ^2 goes beyond the 99% probability concentration region of a $\chi_{n_x}^2$ variable, were chosen by trial and error. The values settled on were selected because they represented an acceptable trade-off between stringency and 'tolerance' of transients.

4.2 The Posterior Cramér-Rao Lower Bound

The Cramér-Rao lower bound represents a lower bound on the second-order error of an estimator. Although this does not provide a complete picture of the accuracy of nonlinear filters (as the densities involved in nonlinear filtering have moments higher than the second order), it is nevertheless a commonly used tool for benchmarking purposes. We present here the specific form of the posterior CRLB (PCRLB) which is used throughout the following simulations. We note that the attribute 'posterior' refers to the applicability of the method to cases where the state dynamics are stochastic (i.e. process noise is nonzero). The covariance matrix of an unbiased estimator of the state $\mathbf{x}_k$ of a system is bounded as follows:

$$E\{(\hat{\mathbf{x}}_{k|k} - \mathbf{x}_k)(\hat{\mathbf{x}}_{k|k} - \mathbf{x}_k)^{\mathrm{T}}\} \geq \mathbf{J}_k^{-1}, \tag{4.2.1}$$

where $\mathbf{J}_k$ denotes the Fisher information matrix given by

$$\mathbf{J}_k \;=\; E\{[\nabla_{\mathbf{x}_k} \log p(\mathbf{x}_k, \mathbf{Z}_k)][\nabla_{\mathbf{x}_k} \log p(\mathbf{x}_k, \mathbf{Z}_k)]^{\mathrm{T}}\}, \tag{4.2.2}$$

and the $\nabla_{\mathbf{x}_k}$ operator corresponds to the gradient with respect to $\mathbf{x}_k$. Tichavsky et al. established the following recursion for the computation of the PCRLB [77]:

$$\mathbf{J}_{k+1} \;=\; \mathbf{Q}_k^{-1} + \mathbf{H}_{k+1}^{\mathrm{T}} \mathbf{R}_{k+1}^{-1} \mathbf{H}_{k+1} - \mathbf{Q}_k^{-1} \mathbf{F}_k \left(\mathbf{J}_k + \mathbf{F}_k^{\mathrm{T}} \mathbf{Q}_k^{-1} \mathbf{F}_k\right)^{-1} \mathbf{F}_k^{\mathrm{T}} \mathbf{Q}_k^{-1}, \tag{4.2.3}$$

which can be written, using the matrix inversion lemma, as

$$\mathbf{J}_{k+1} \;=\; \left(\mathbf{Q}_k + \mathbf{F}_k \mathbf{J}_k^{-1} \mathbf{F}_k^{\mathrm{T}}\right)^{-1} + \mathbf{H}_{k+1}^{\mathrm{T}} \mathbf{R}_{k+1}^{-1} \mathbf{H}_{k+1}, \tag{4.2.4}$$

where $\mathbf{F}_k$ is the Jacobian of the nonlinear system equation evaluated at the *true* state $\mathbf{x}_k$, and $\mathbf{H}_{k+1}$ is the jacobian of the nonlinear measurement equation the *true* target state $\mathbf{x}_{k+1}$. At every time instant, the mean-squared error of each state component is bounded by the corresponding diagonal element of $\mathbf{J}_k^{-1}$.

For systems with multiple switching dynamics and clutter in the measurement process, a bound can be computed conditional on a specific mode sequence $\{r_k\}$ and clutter sequence $\{D_k\}$ (using the 'single-measurement' clutter model described in 3.1.2). Making the simplifying assumption that the correct mode sequence $\{r_k^*\} = \{r_1^*, \ldots, r_k^*\}$ and binary valued clutter sequence $\{D_k^*\} = \{D_1^*, \ldots, D_k^*\}$ are known a priori, the PCRLB can be evaluated using a modified form of the recursion (4.2.3):

$$\mathbf{J}_{k+1}^* \;=\; \left(\mathbf{Q}_k(r_k^*) + \mathbf{F}_k(r_k^*) \mathbf{J}_k^{-1} \mathbf{F}_k(r_k^*)^{\mathrm{T}}\right)^{-1} + (1 - D_k^*) \left(\mathbf{H}_{k+1} \mathbf{R}_{k+1}^{-1} \mathbf{H}_{k+1}^{\mathrm{T}}\right). \tag{4.2.5}$$

It should be noted that this is an unfeasibly optimistic bound as it presupposes knowledge of the actual target mode trajectory along with the actual 'detection' and 'miss' sequence, both of which are sources of nonlinearity that need to be estimated by the filter. Nevertheless, the resulting PCRLB can still provide a useful indication of the performance limits of practical filters.

4.3 Single Sensor Tracking of a Non-Maneuvering Target in Clutter

In this section we compare the performance of the shifted Rayleigh filter (labelled 'SRF3C'), for state estimation from 3D bearings-only measurements in the presence of clutter, with that of the EKF, the RPEKF, the UKF and also the SIR particle filter with local EKF linearization (EKPF). We focus on an apparently simple yet highly challenging 3D bearings-only tracking scenario. The problem, which involves a single sensor platform and non-maneuvering target motion, is defined as follows. The sensor platform sets off 4 km above the origin and travels parallel to the xy plane at a constant speed of 130 m/s and a course of $-80°$ for 14 s. It then executes a maneuver and adopts a new course of 153°, which it maintains for the remainder of the simulation period. Meanwhile, the target, initially at an altitude of 6.5 km, a range r_0 of approximately 10.6 km from the sensor platform and a bearing of 29°, executes a descent at a constant speed of 470 m/s, a course of $-149.5°$ and pitch of $-12.3°$, eventually traveling past the sensor platform. The particular geometry of this bearings-only tracking scenario, illustrated in figure 4.1, makes the target range unobservable for the first 14 s, posing a great challenge to filter stability.

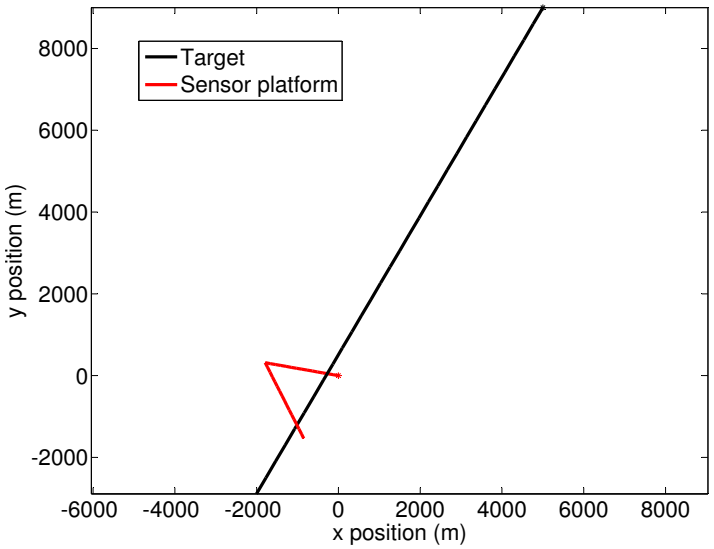


Figure 4.1: Target/sensor platform geometry (projected on the xy plane)

The 6-vector target state comprises the components of the Cartesian position and velocity

along the xyz axes,

$$\mathbf{x}_t = \begin{bmatrix} x_k & \dot{x}_k & y_k & \dot{y}_k & z_k & \dot{z}_k \end{bmatrix}^{\mathrm{T}},$$

and dynamics are those of a 'nearly' constant velocity discrete time model

$$\mathbf{x}_k \;=\; \mathbf{F}\mathbf{x}_{k-1} + \mathbf{u}^s_{k-1} + \mathbf{v}_{k-1}$$

in which

$$F = \mathrm{diag}\left(\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \right),$$

where T is the sampling period in seconds. The deterministic signal $\mathbf{u}^s_k$ is taken to be zero while the movement of the sensor platform is absorbed into the SRF measurement equation

$$\mathbf{d}_k \;=\; \mathbf{H}\mathbf{x}_k + \mathbf{u}^m_k$$

via the signal $\mathbf{u}^m_k$. The covariance matrix of the discretized zero-mean white Gaussian noise process $\mathbf{v}_k$ is expressed as:

$$Q^s = q\ \mathrm{diag}\left(\begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix}, \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix}, \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix} \right),$$

and the system noise parameter q is set to $9.92 \times 10^{-5}\ \mathrm{m}^2/\mathrm{s}^3$. Noisy angle measurements $\mathbf{z}_k$ are received at intervals of $T = 1$ s, with a probability p_c of being clutter. These are generated according to the model

$$\mathbf{z}_k \;=\; (1 - D_k)\boldsymbol{\psi}_k + D_k\mathbf{c}_k,$$

where D_k is a binary process taking the value 0 if the measurement is target originated and 1 if it is clutter, and $\mathbf{c}_k$ is the false azimuth and elevation measurement vector arising from clutter. Finally, $\boldsymbol{\psi}_k$ is the idealized azimuth and elevation measurement vector at time k, which is given, in consistence with the analysis from 3.1.1, by

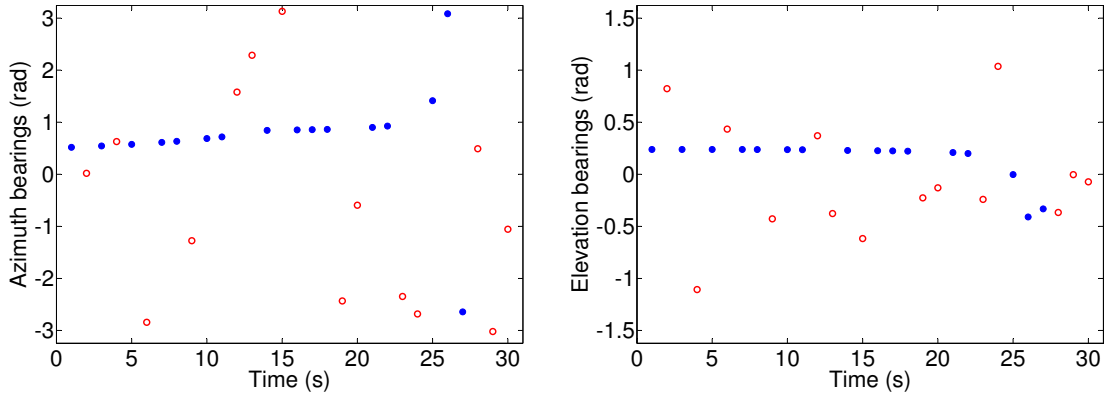
$$\boldsymbol{\psi}_k \;=\; h(\mathbf{d}_k) + \tilde{\mathbf{w}}_k(\mathbf{d}_k),$$

where $\tilde{\mathbf{w}}_k(\mathbf{d}_k)$ is a 2-dimensional zero-mean Gaussian random variable with first and second components 'wrapped' onto $[-\pi,\ \pi]$ and $[-\pi/2,\ \pi/2]$, respectively, and with covariance matrix

$$\mathbf{R}_k(\mathbf{d}_k) \ = \ \sigma^2 \begin{bmatrix} (\cos\alpha_k)^{-2} & 0 \\ 0 & 1 \end{bmatrix},$$

where $\alpha_k = \arctan\left(d_{k3}/\sqrt{(d_{k1})^2 + (d_{k2})^2}\right)$. σ is the scaling factor of the measurement noise covariance and is set to 0.003.

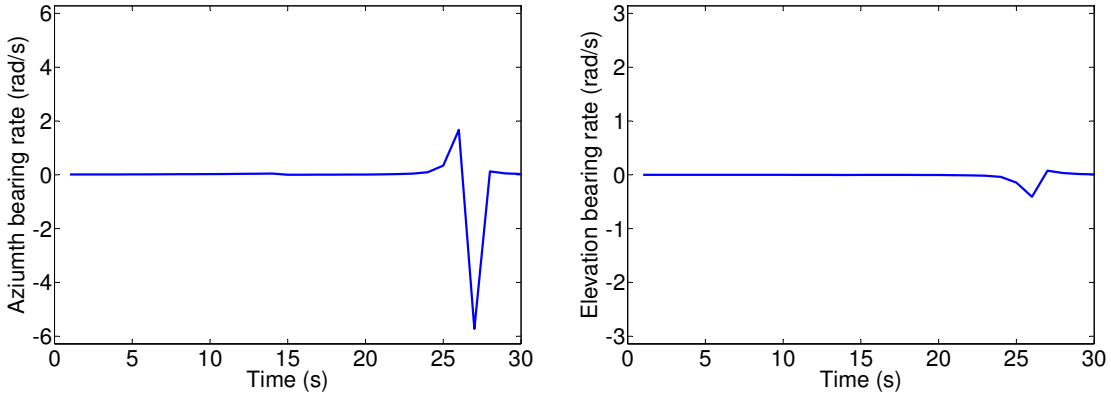
The simulation horizon is 30 s. Figures 4.2a and 4.2b illustrate a sample realization of azimuth and elevation bearings measurements for a probability of clutter $p_c = 0.5$. Figure 4.3a shows the clutter-free azimuth bearing rate, which approaches 2π rad/s in magnitude as the target moves past the sensor platform at $t_k = 27$ s. The corresponding clutter-free elevation bearing-rate, whose magnitude peaks at $\pi/8$ rad/s is shown in figure 4.3b.



(a) (b)

Figure 4.2: Sample realization of azimuth and elevation bearing measurements for $p_c = 0.5$. Filled blue circles represent target originated measurements, red circles designate clutter.

We perform 200 Monte Carlo runs of this scenario over increasing probability of clutter p_c. The initial range prior is taken to be $\hat{r}_{0|0} \sim N(r_0, \sigma^2_{r_0})$, with $\sigma_{r_0} = 5$ km, while the initial speed priors along the xyz axes are given by $\hat{s}_{0|0} \sim N(0, \sigma^2_{s_0})$, where $\sigma_{s_0} = 500$ m/s for the x and y coordinates and $\sigma_{s_0} = 166$ m/s for the z coordinate. The SRF3C, EKF, RPEKF (made up of 6 EKFs covering initial mean range estimates from 1 km to 30 km) and UKF are all implemented using the 'single-measurement PDA' modification described



(a) (b)

Figure 4.3: Azimuth and elevation bearing rates in the absence of clutter, with peak magnitudes of 2π rad/s and $\pi/8$ rad/s, respectively.

in 3.1.2. The EKPF is set to propagate 3000 samples drawn from importance densities generated by separate EKFs (each using the 'single-measurement PDA' modification, to account for clutter in the measurements). [1]

Finally, note that for all filters excluding the SRF3C, the measurement noise covariance $\mathbf{R}_k(\mathbf{d}_k)$, which is correlated with the state $\mathbf{x}_k$, is approximated in terms of the predicted displacement by $\mathbf{R}_k\left(E[\mathbf{d}_{k|k-1}]\right)$. The validity of this construction depends on $E[\mathbf{d}_{k|k-1}]$ being accurate, which is a reasonable assumption to make in most applications.

Figure 4.4 shows the RMS errors of the range estimates for $p_c = 0.5$. The average performances of the EKF and UKF are clearly unacceptable, while the RPEKF shows only slightly improved robustness and accuracy. The overall low precision and frequent track divergence reflected in their RMS range errors point to the inadequacies of tracking algorithms based on simple linearization techniques (whether analytical or statistical, as with the UKF), faced with a challenging scenario involving high clutter probability, high bearings rate and large initial state uncertainty. The SRF3C, on the other hand, comes close on average to achieving the PCRLB and has an RMS error comparable to that of the EKPF. Note that the PCRLB curve lies above the RMS errors of the EKPF for the

[1] The simpler 'bootstrap' filter was initially tested, but performed very poorly; possibly as a result of the high dimensionality of the problem and large initial uncertainty. The low accuracy of the transitional prior, used as proposal density, required a very high number of particles (around 500.000) which came at an unacceptable computational cost. The use of a proposal density closer to the posterior was found to resolve the issue for this scenario.

initial phase of the track during which transient effects are significant, owing to a lack of observability before the abrupt maneuver of the sensor platform at $t_k = 14$ s.

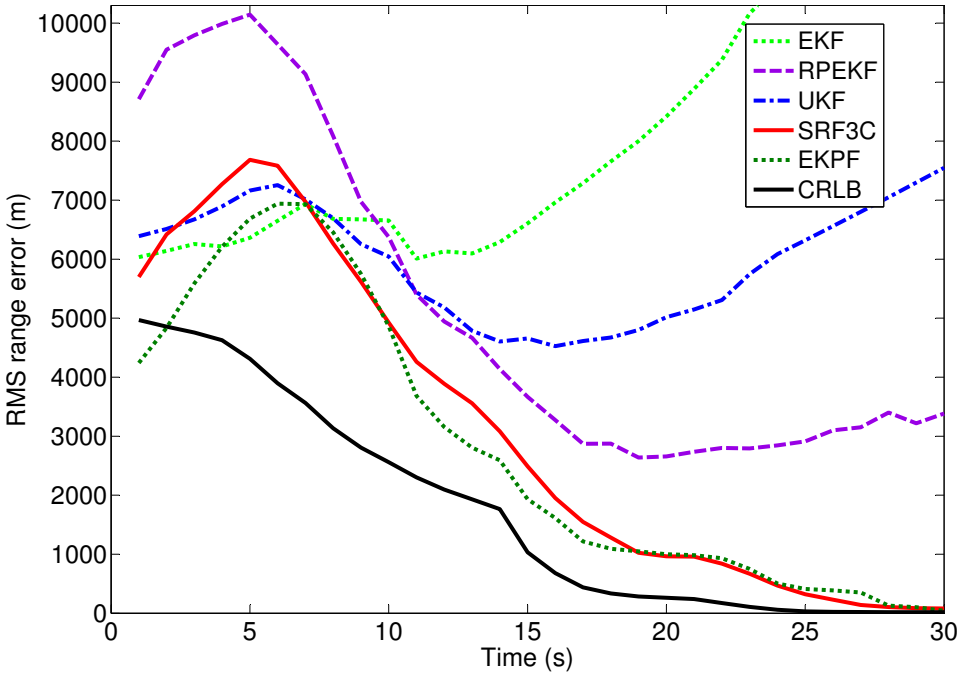


Figure 4.4: RMS error of the target range estimate with $p_c = 0.5$ over 200 Monte Carlo simulations

Figure 4.5 illustrates the percentage of divergent tracks produced by the various filters over the 200 MC runs, with varying levels of clutter probability. The value of D_{th} (indicating the range RMSE above which a track is deemed to have diverged) is set to 6000 m. The SRF3C shows high robustness to clutter with practically no divergent tracks up to a clutter probability of 0.6. On the other hand, the EKF and UKF exhibit frequent breakdowns. The relative robustness of the RPEKF, owing to the use of multiple filters, can also be seen. Finally, the EKPF achieves a level of resilience to clutter comparable to that of the SRF3C.

The TARMS range errors for all the filters are displayed in table 4.1. The time index at which averaging starts is set to $k = 15$ s, which is just after the sensor platform changes course and full observability of the target state is ensured. Because of the high number of lost tracks, the TARMS error results of the EKF (and, for certain values of p_c, those

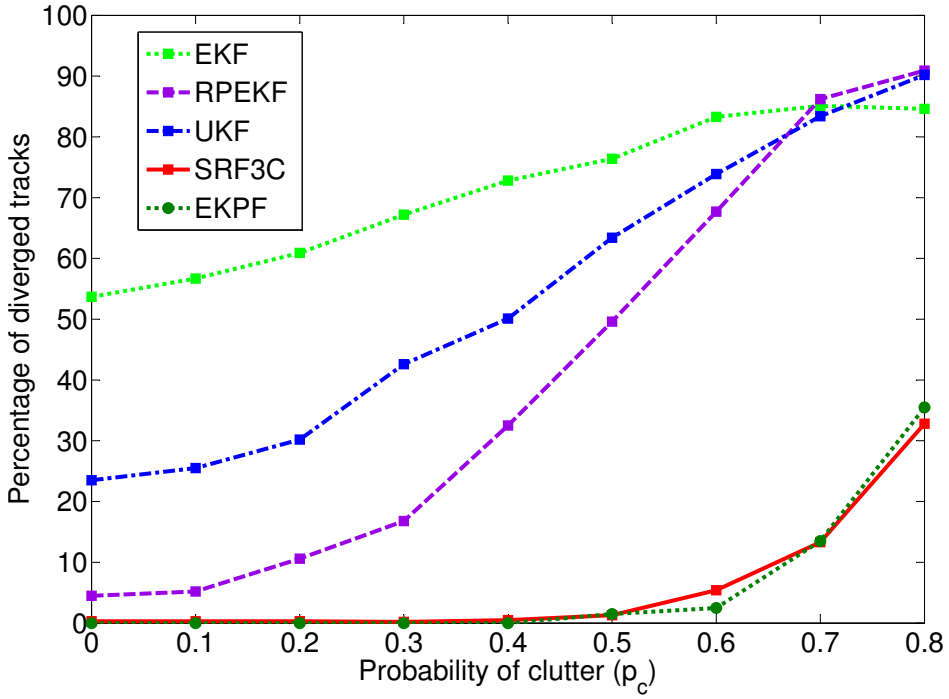


Figure 4.5: Percentage of track divergence versus probability of clutter

of the other filters) are not reported. The RPEKF appears to be competitive for the relatively more benign case where $p_c \leq 0.1$, owing to the aggregation of estimates from parallel EKFs. However, its performance quickly deteriorates as the probability of clutter is increased. The accuracy of the SRF3C is very close to that of the EKPF at all (realistic) levels of clutter and both filters have the lowest tracking error up to a probability of clutter of 0.7. However, tracking becomes erratic beyond this point and all filters exhibit frequent divergence caused by the extreme levels of nonlinearity.

Table 4.1: TARMS range errors (m)

	probability of clutter p_c								
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
EKF	-	-	-	-	-	-	-	-	-
RPEKF	281	332	604	811	2258	3146	4150	-	-
UKF	1674	2643	1686	3039	5058	-	-	-	-
SRF3C	288	360	416	438	629	954	1648	3288	-
EKPF	206	282	315	434	620	950	1620	2947	5581

The relative average computational times (per iteration) of all filters are shown in table

4.2 [2]. With twice the computational requirements of the EKF, the SRF3C is the second least demanding. The EKPF needs by far the highest computational time, at nearly 9000 multiples of that of the EKF.

Table 4.2: Computation times (relative)

	Filter				
	EKF	RPEKF	UKF	SRF3C	EKPF
time	1	5	2.3	2.1	8830

Having discussed the performance benefits of the shifted Rayleigh filter in single sensor tracking scenarios, we turn our attention to the problem of smoothing. We examine, for the scenario presented above, the performance of the fixed-interval RTS shifted Rayleigh smoother of section 3.2 and of the RTS smoother with an RPEKF in the forward pass. The forward filtering pass, for each tracker, takes the form of the SRF3C and RPEKF with the 'single-measurement' PDA modification, while the backward smoothing pass is identical for both trackers and is given by equations (3.2.5), (3.2.6), (3.2.8) and (3.2.9).

Table 4.3 summarizes the range TA-RMSEs of the two smoothers under consideration (labelled RTS-SRS3C and RTS-RPEKS), for increasing probabilities of clutter (unlike the results in table 4.1, the averaging is carried out over all time steps). The RTS-SRS3C is seen to achieve excellent accuracy up to high levels of clutter, while the performance of the RTS-RPEKS quickly degrades with increasing probability clutter. Given the linear/Gaussian dynamics of the target, the accuracy of the smoothed estimates are predominantly determined by the accuracy of the final filtered estimate. We conclude therefore that the superior performance of the RTS-SRS3C is a result of the very high precision (and small associated covariance matrix) ultimately achieved by the SRF3C in the forward pass. Likewise, the limitations of the RTS-RPEKS extend from the poor final estimates of the RPEKF. Note, finally, that results of EKF and UKF based implementations of the RTS smoother were not included because they mostly diverged.

[2] All filters were coded in MATLAB, using vectorization whenever feasible

Table 4.3: TARMS range errors (m)

	probability of clutter p_c								
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
RTS-SRS3C	13	13	14	16	20	40	656	1153	-
RTS-RPEKS	14	20	66	321	1321	2207	2774	4577	-

4.4 The Shifted Rayleigh Particle Filter

We present in this section a performance evaluation of the SRF, when used to generate proposal distributions within a particle filter. The resulting 'shifted Rayleigh particle filter' (SRPF) falls into the category of particle filters using approximations of the true posterior as proposal densities, of which the EKPF and UPF (unscented particle filter) [80] are well known examples. The SRPF, EKPF and UPF algorithms, all share the same Monte-Carlo structure and differ only in the generation of proposal densities. Within these filters, particles are drawn from analytical Gaussian approximations (provided by the SRF, EKF or UKF) of the true posterior, by taking into account the latest observation. Table 4.4 shows a unified representation of one iteration of these filters.

Comparison of the three algorithms is made under the single sensor tracking scenario described in section 4.3 , for a fixed set of 200 runs and $p_c = 0.5$, with all other parameters kept identical. The presence of clutter is handled, for each particle, by computing a proposal density using the 'single-measurement PDA' modification, as in the SRF3C, EKF and UKF of the previous section. All particle filters are initialized with the same N Gaussian priors whose means are drawn from the initializing distribution $\mathcal{N}(\mathbf{x}_{0|0}; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0})$ given in section 4.1.1. Their associated covariance matrices are then set to $c^{-1}\,\mathbf{P}_{0|0}$, where c is a scaling constant. It was found that taking $c < 1$ led to smaller initial transients and faster convergence as a result of tighter initialization of the analytical filters that generate the proposal distributions. The choice $c = 0.01$, used throughout the following simulations, was found to provide adequate performance.

We choose to focus on computational cost in the comparison of the different PF algorithms. Although most performance evaluations in the literature are made for equal particle numbers, our aim here is to demonstrate the possibility of achieving a given level

Table 4.4: PF algorithm using posterior approximation as proposal distribution

1. For $i = 1 : N$
   - Evaluate Gaussian approximation to posterior density: $[\hat{\mathbf{x}}_k^i, \mathbf{P}_k^i] = \text{EKF/UKF/SRF}\left(\hat{\mathbf{x}}_{k-1}^i, \hat{\mathbf{P}}_{k-1}^i, \mathbf{z}_k\right)$
   - Sample $\mathbf{x}_k^i$ from the proposal density $\mathcal{N}(\mathbf{x}_k^i; \hat{\mathbf{x}}_k^i, \mathbf{P}_k^i)$
   - Compute importance weights (as in 2.3.38): $\tilde{w}_k^i = w_{k-1}^i \frac{p(\mathbf{z}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{\mathcal{N}(\mathbf{x}_k^i;\hat{\mathbf{x}}_k^i,\mathbf{P}_k^i)}$,
2. For $i = 1 : N$
   - Normalized weights: $w_k^i = \tilde{w}_k^i / \sum_{i=1}^N \tilde{w}_k^i$
3. Compute effective sample size $\hat{N}_{eff} = \left(\sum_{i=1}^N (w_k^i)^2\right)^{-1}$
4. If $\hat{N}_{eff} < N_{thr}$, resample (using any of the methods in 2.3.2):
   - $\{i^j\}_{j=1}^N = \text{RESAMPLE}\{w_k^i\}_{i=1}^N$
   - For $j = 1 : N$ $[\mathbf{x}_k^j, \mathbf{P}_k^j, w_k^j] = [\mathbf{x}_k^{i^j}, \mathbf{P}_k^{i^j}, 1/N]$

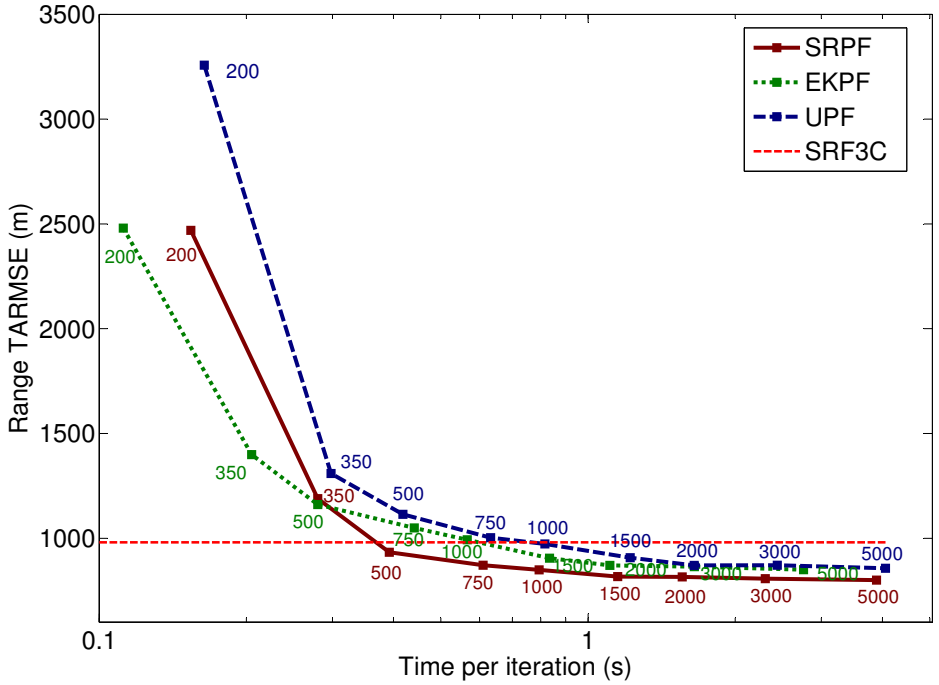


Figure 4.6: Range TARMSE of the target over 10 repetitions of 200 MC runs, with $p_c = 0.5$. Numbers next to curves represent the corresponding number of particles.

of accuracy, at the lowest computational cost. In order to reduce the MC variation of the estimates (resulting from the random sampling in the PFs), we report the average over series 10 MC simulations of the same set of 200 runs, for different numbers of particles. Figure 4.6 shows the range TARMSEs resulting from each PF against computational time, for increasing number of particles N. The range TARMSE of the SRF3C, as computed in the previous section, is also included for comparison purposes. Notice that its value is extended along the x axis for convenient visualization (one iteration of the SRF3C takes 0.5 ms).

The SRPF is seen to outperform the SRF3C with as little as 500 particles, whereas both the EKPF and UPF require more than 1000 particles for the same level of accuracy. Also, this comes at computational requirements respectively 44% and 108% in excess of the SRPF. Moreover, the SRPF achieves asymptotic convergence with 750 particles, against 1500 for the EKPF and UPF.

Even though all three PFs reach similar levels of accuracy for very high numbers

of particle (2000 and beyond), the EKPF and UPF do not exactly match the performance of the SRPF. This is partly due to the occasional divergence encountered by these filters. This is illustrated in figure 4.7, where the percentage divergence of the SRF3C is extended over all values of the x axis. It is evident that although all PFs exhibit monotonically decreasing probability of divergence for increasing number of particles, the SRPF is consistently more robust for equal computational expense.

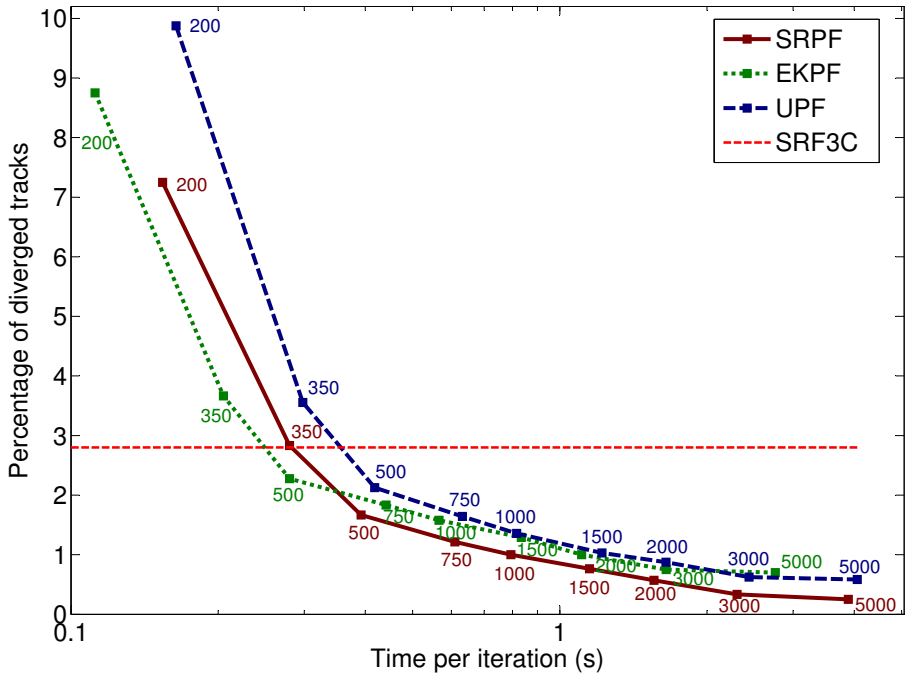


Figure 4.7: Percentage of track divergence over 10 repetitions of 200 MC runs, with $p_c = 0.5$. Numbers next to curves represent the corresponding number of particles.

We conclude that, for the problem of bearings-only tracking, the accuracy of PFs utilizing approximations of the optimal proposal density can be improved on, using the SRF. The performance benefits of the SRF3C over the EKF and UKF can be seen to generalize to sequential Monte Carlo implementations of the same algorithms, although with comparatively smaller improvements.

4.5 Multiple Sensor Tracking of a Maneuvering Target in Clutter

This tracking scenario consists of two spatially separated sensors providing simultaneous azimuth and elevation measurements of a maneuvering target with switching dynamic modes. Sources of nonlinearity are therefore present in both the state process (through the discontinuities introduced by switching modes) and in the measurement process (as a result of the bearings-only measurements). These difficulties are further compounded by the presence of clutter in the measurements from both sensors, high initial state uncertainty and high bearings-rate.

Scenario Definition

We define a 9-dimensional state vector for the target, comprising its Cartesian position, velocity and acceleration components along the xyz axes:

$$\mathbf{x}_t = \left[\begin{array}{ccccccccc} x_k & \dot{x}_k & \ddot{x}_k & y_k & \dot{y}_k & \ddot{y}_k & z_k & \dot{z}_k & \ddot{z}_k \end{array} \right]^{\mathrm{T}}. \tag{4.5.1}$$

Its dynamics are governed by the linear switching model

$$\mathbf{x}_k \;=\; \mathbf{F}(r_k)\mathbf{x}_{k-1} + \mathbf{u}_k^s + \mathbf{v}(r_k), \tag{4.5.2}$$

where, as in 3.1.3, r_k represents the mode in effect during the time interval $(t_{k-1}, t_k]$, $\mathbf{F}(r_k)$ is the mode-dependent system matrix, $\mathbf{u}_{k-1}^s$ is an exogenous state input and $\mathbf{v}(r_k)$ is a mode-dependent zero-mean Gaussian random variable.

The target trajectory is defined by a combination of 3 motion models. Mode 1 is 'nearly' constant velocity motion, used for non-maneuvering segments. Mode 2 describes the onset of a maneuver. It is characterized by a Wiener process acceleration model with exogenous acceleration inputs (injected via the $\mathbf{u}_k^s$ term) to represent abrupt changes in target course. Finally, mode 3 is a 3D 'constant-turn' motion model, as described in 2.4.3. The system and noise covariance matrices are given by

$$\mathbf{F}(r) \;=\; \mathrm{diag}[\mathbf{A}(r), \mathbf{A}(r), \mathbf{A}(r)] \tag{4.5.3}$$

and

$$\mathbf{Q}^s(r) \;=\; \mathrm{diag}[\mathbf{R}(r), \mathbf{R}(r), \mathbf{R}(r)] \tag{4.5.4}$$

where

$$\mathbf{A}(r=1) \;=\; \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \tag{4.5.5}$$

$$\mathbf{A}(r=2) \;=\; \begin{bmatrix} 1 & T & \frac{1}{2}T^2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}, \tag{4.5.6}$$

$$\mathbf{A}(r=3) \;=\; \begin{bmatrix} 1 & \frac{\sin\omega T}{\omega} & \frac{1-\cos\omega T}{\omega^2} \\ 0 & \cos\omega T & \frac{\sin\omega T}{\omega} \\ 0 & -\omega\sin\omega T & \cos\omega T \end{bmatrix}, \tag{4.5.7}$$

and

$$\mathbf{R}(r=1) \;= q_1 \begin{bmatrix} \frac{1}{3}T^3 & \frac{1}{2}T^2 \\ \frac{1}{2}T^2 & T \end{bmatrix}, \tag{4.5.8}$$

$$\mathbf{R}(r=2) \;= q_2 \begin{bmatrix} \frac{1}{20}T^5 & \frac{1}{8}T^4 & \frac{1}{6}T^3 \\ \frac{1}{8}T^4 & \frac{1}{3}T^3 & \frac{1}{2}T^2 \\ \frac{1}{6}T^3 & \frac{1}{2}T^2 & T \end{bmatrix}, \tag{4.5.9}$$

$$\mathbf{R}(r=3) \;= q_3 \begin{bmatrix} \frac{6\omega T-8\sin\omega T+\sin 2\omega T}{4\omega^5} & \frac{2\sin^4(\omega T/2)}{\omega^4} & \frac{-2\omega T+4\sin\omega T-sin2\omega T}{4\omega^3} \\ \frac{2sin^4(\omega T/2)}{\omega^4} & \frac{2\omega T-\sin 2\omega T}{4\omega^3} & \frac{\sin^2\omega T}{2\omega^2} \\ \frac{-2\omega T+4\sin\omega T-\sin 2\omega T}{4\omega^3} & \frac{\sin^2\omega T}{2\omega^2} & \frac{2\omega T+\sin 2\omega T}{4\omega} \end{bmatrix}. \tag{4.5.10}$$

T is the sampling period, set to 1 s, while ω is the turn rate, whose value is fixed at 0.08 rad/s. The intensities q_1, q_2 and q_3 of target perturbations are set to 0 for the purpose of generating the true target track. The mode sequence is:

$\{r_{1:39} = 1,\, r_{40:41} = 2,\, r_{42:59} = 3,\, r_{60:61} = 2,\, r_{62:69} = 3,\, r_{70:109} = 1,\, r_{110:111} = 2, r_{112:129} = 3,$
$r_{130:131} = 2, r_{132:141} = 3, r_{142:160} = 1\}$

and the exogenous input is zero at all times, except when mode 2 is in effect where its values are:

$$\begin{aligned}
\mathbf{u}^s_{40:41} &= \Big[0,\, 0,\, 15,\, 0,\, 0,\, 0,\, 0,\, 0,\, -10\Big]^{\mathrm{T}}, \\
\mathbf{u}^s_{60:61} &= \Big[0,\, 0,\, -20.64,\, 0,\, 0,\, -40,\, 0,\, 0, 13.77\Big]^{\mathrm{T}}, \\
\mathbf{u}^s_{110:111} &= \Big[0,\, 0,\, 10, 0,\, 0,\, 15,\, 0,\, 0,\, -4\Big]^{\mathrm{T}}, \\
\mathbf{u}^s_{130:131} &= \Big[0,\, 0,\, -45, 0,\, 0,\, -30.53,\, 0,\, 0,\, 4.14\Big]^{\mathrm{T}}.
\end{aligned}$$

Two sensors are located on the ground plane at fixed (x, y, z) positions $(\pm 5000, 0, 0)$ and the initial target state is set to $\mathbf{x}_0 = [-3000, 0, 0, 40000, -400, 0, 6500, 0, 0]^{\mathrm{T}}$, i.e. at a range r_0 of about 41.3 km from sensor 1. Figure 4.8 shows the geometry of the tracking scenario, projected onto the xy plane. The target travels at a constant speed and course of 400 m/s and $-\pi$, respectively, and at an altitude of 6500 m for 39 s before executing a diving left turn lasting 19 s. It then levels up and follows a straight trajectory at a new altitude of 2160 m for 39 s. As soon as it passes sensor 1, it executes two abrupt high g left turns, while diving, until an altitude of 290 m is reached. The remainder of the trajectory is non-accelerating motion at a new course of $-\pi/2$ rad.

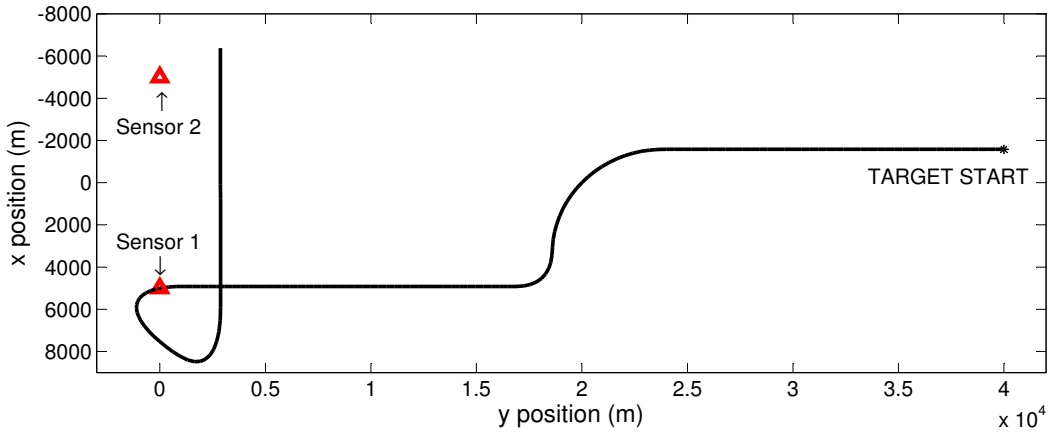


Figure 4.8: Target/sensor geometry (projected on the xy plane)

Figure 4.9 shows the target range from sensor 1, while figure 4.10 depicts the target speed throughout the simulation period. Finally, figure 4.11 shows the magnitude of the target acceleration, which approaches 10 g during the final maneuver.

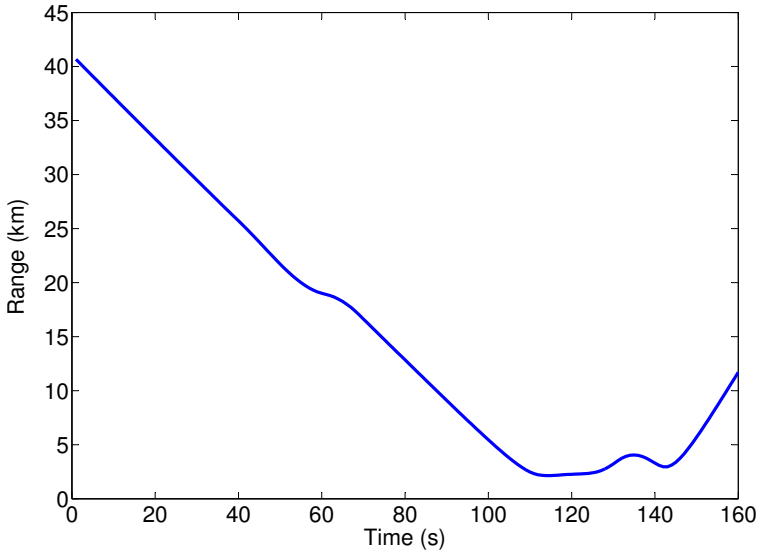


Figure 4.9: Target range from sensor 1

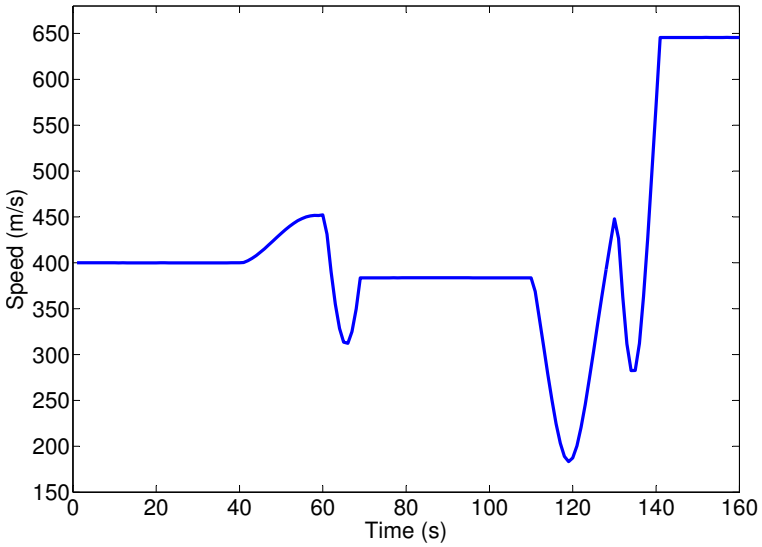


Figure 4.10: Target speed

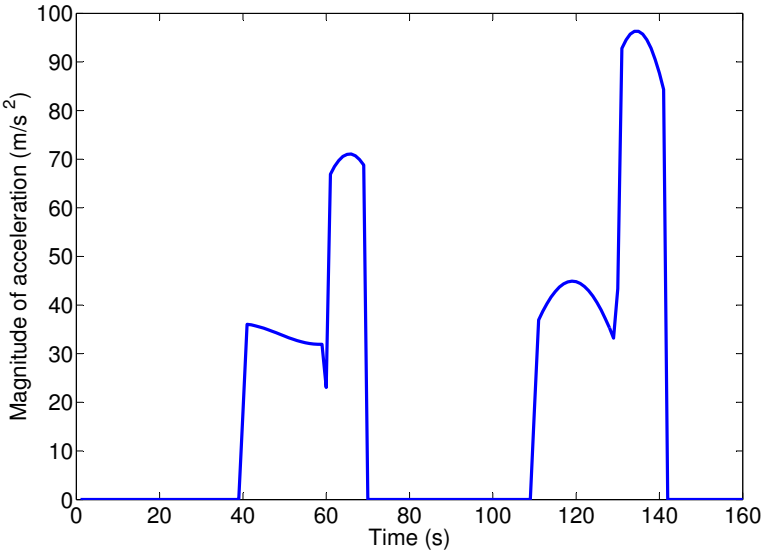


Figure 4.11: Target acceleration

Azimuth and elevation measurements of the target are acquired simultaneously from both sensors at intervals $T = 1$ s and with a probability p_c of being clutter. Defining the displacement $\mathbf{d}_k^l$ at time k from the lth sensor (out of a total of L) to be

$$\mathbf{d}_k^l \;=\; \mathbf{H}\mathbf{x}_k + \mathbf{u}_k^{m,l},$$

(where $\mathbf{u}_k^{m,l}$ denotes the position vector of the lth sensor), the measurements are generated according to the model

$$\mathbf{z}_k^l \;=\; (1 - D_k^l)\boldsymbol{\psi}_k^l + D_k^l \mathbf{c}_k^l,$$

where D_k^l is a binary process taking the value 0 if the measurement is target originated and 1 if it is clutter, and $\mathbf{c}_k^l$ is the false azimuth and elevation measurement vector, arising from clutter, at the lth sensor. $\boldsymbol{\psi}_k^l$ is the idealized azimuth and elevation measurement vector obtained from the lth sensor at time k, which is given, in consistence with the analysis from 3.1.1, by

$$\boldsymbol{\psi}_k^l \;=\; h(\mathbf{d}_k^l) + \tilde{\mathbf{w}}_k^l(\mathbf{d}_k^l).$$

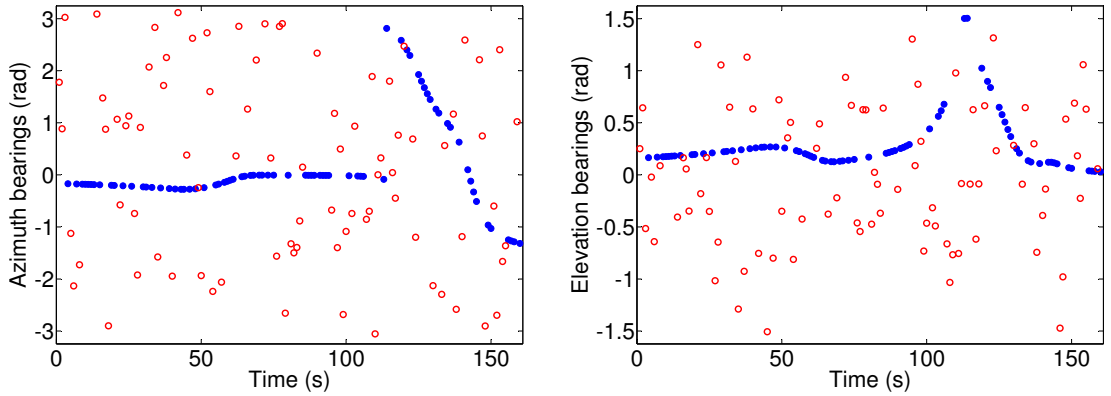
$\tilde{\mathbf{w}}_k^l(\mathbf{d}_k^l)$ is a 2-dimensional zero-mean Gaussian random variable with first and second components 'wrapped' onto $[-\pi,\ \pi]$ and $[-\pi/2,\ \pi/2]$, respectively, and with covariance

matrix

$$\mathbf{R}_k(\mathbf{d}_k^l) \;=\; \sigma^2 \begin{bmatrix} (\cos\alpha_k^l)^{-2} & 0 \\ 0 & 1 \end{bmatrix},$$

where $\alpha_k^l = \arctan\left(d_{k3}^l / \sqrt{(d_{k1}^l)^2 + (d_{k2}^l)^2}\right)$. The scaling factor σ of the measurement noise is set to 0.005.

Figures 4.12a and 4.12b show a representative sequence of azimuth and elevation measurements acquired at sensor 1, with a relatively high probability of clutter of $p_c = 0.5$. The corresponding clutter-free azimuth bearing rate (which almost reaches 2π rad/s at $t_k = 114$ s) and elevation bearing rate are shown in figures 4.13a and 4.13b. Similarly, figures 4.14a and 4.14b illustrate the measurements from sensor 2 and 4.15a and 4.15b show the corresponding clutter-free bearing rates, which are practically constant.



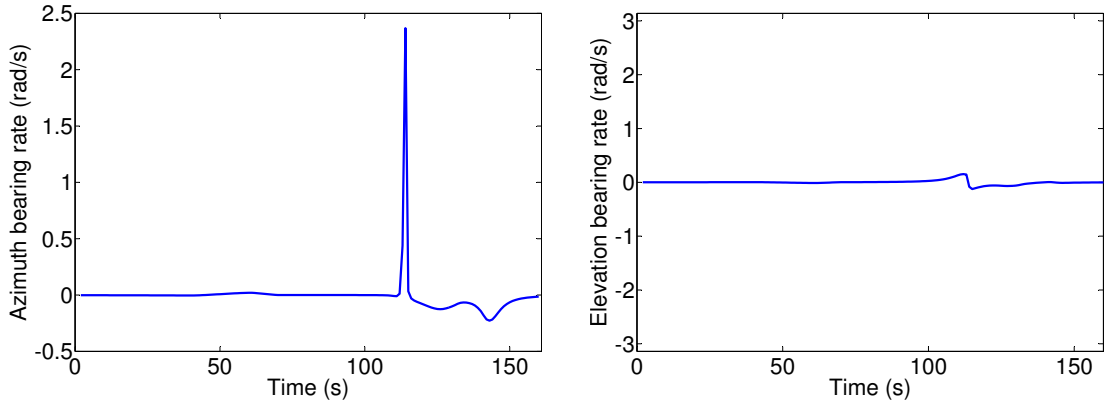
(a) (b)

Figure 4.12: Sample realization of azimuth and elevation bearing measurements at sensor 1 for $p_c = 0.5$. Filled blue circles represent target originated measurements, red circles designate clutter.

Modelling of Target Dynamics

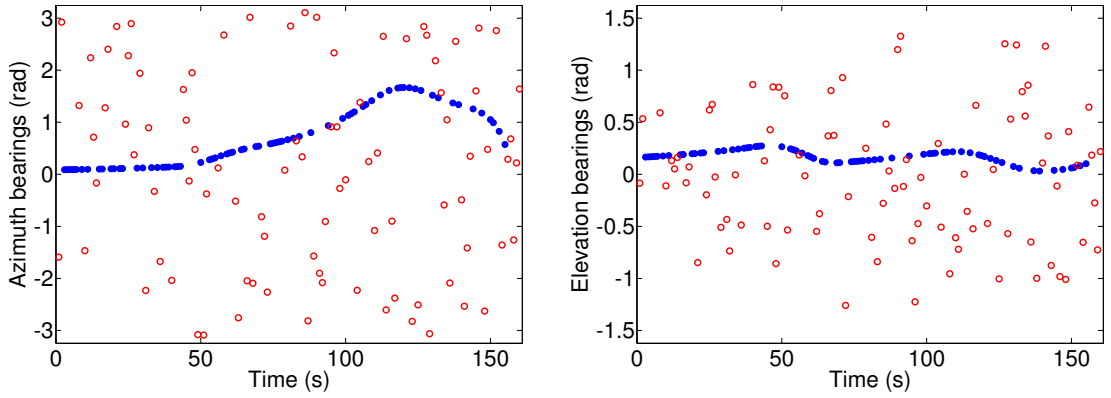
We consider two sets of models for estimating the previously defined trajectory. The first one implies a priori knowledge of the kinematics of the 3D turn model, while the second reflects the more general (and realistic) scenario of unknown maneuver kinematics.

First set of models



(a) (b)

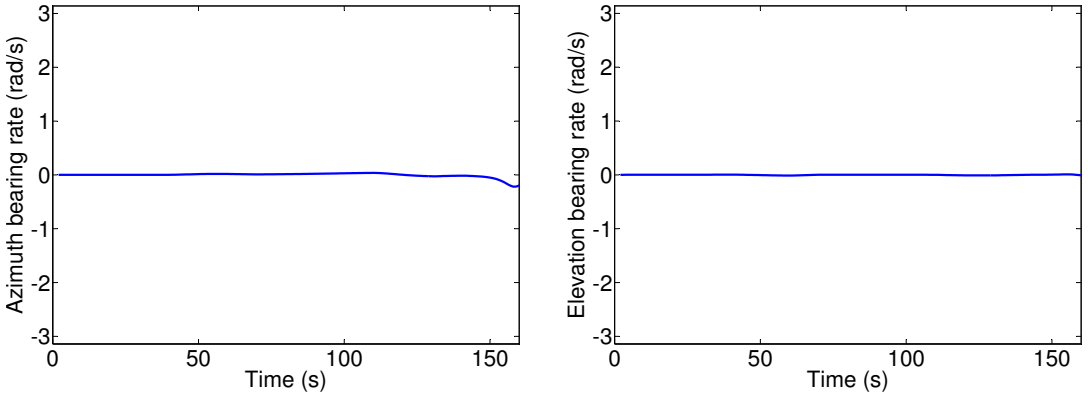
Figure 4.13: Azimuth and elevation bearing rates in the absence of clutter at sensor 1. The former reaches a peak magnitudes of nearly π rad/s.



(a) (b)

Figure 4.14: Sample realization of azimuth and elevation bearing measurements at sensor 2 for $p_c = 0.5$. Filled blue circles represent target originated measurements, red circles designate clutter.

The first set of models is defined as $\mathcal{S}_1 = \{\text{Mode 1}, \text{Mode 2}, \text{Mode 3}\}$, with system matrices as given by (4.5.5)-(4.5.7), and associated noise covariance matrices (4.5.8)-(4.5.10). The intensities q_1 and q_3 of the process noise for Modes 1 and 3 are both set to 9.92^{-5} m^2/s^3. As access to the exogenous acceleration inputs in effect during Mode 2 is unavailable, Mode 2 is implemented using a high value of target perturbation intensity q_2, to cover the large jumps in acceleration which are characteristic of a maneuver onset. The choice of q_2 requires prompt detection of abrupt maneuvers and involves a trade-off between detection speed and transient effects resulting from overshoot. It was found that $q_2 = 40$ m^2/s^3 provided an adequate compromise



(a) (b)

Figure 4.15: Azimuth and elevation bearing rates in the absence of clutter at sensor 2.

between these two objectives.

Second set of models

The second set of models is defined as $\mathcal{S}_2 = \{\text{Mode } 1, \text{Mode } 2, \text{Mode } 2^*\}$. Modes 1 and 2, which account for segments with constant speed and the onset of maneuvers, respectively, are implemented, as above, with $q_1 = 9.92^{-5}\ m^2/s^3$ and $q_2 = 40\ m^2/s^3$. Mode 2*, used for the modelling ongoing maneuvers, is a variant of Mode 2 with a small value q_{2^*}. The choice $q_{2^*} = 1\ m^2/s^3$ was found to strike an appropriate balance between steady state error and ease of discrimination from Mode 1.

The state transition probabilities, which are design parameters, are set to

$$\pi = \begin{bmatrix} 0.98 & 0.02 & 0 \\ 0.25 & 0.50 & 0.25 \\ 0 & 0.02 & 0.98 \end{bmatrix}. \tag{4.5.11}$$

We note that the zero value given to π_{13} reflects the impossibility for the target to switch from non-accelerating motion to 3D 'constant-turn' motion. A transition through Mode 2 is required in order to provide the non-zero acceleration defining the 'maneuver plane' of the 3D 'constant-turn' motion. Similarly, π_{31} is set to 0, effectively forcing a transition from 3D 'constant-turn' motion to non-accelerating motion through Mode 2. This helps to account for the typically large variations in acceleration at the termination of maneuvers. Finally, we remark that the relatively low value of π_{22} reflects the transient nature of Mode

2, which can only be in effect for a short duration of time.

Processing of Measurements

The simultaneous measurements from sensors 1 and 2 are processed sequentially with an artificial time interval of 0 s between them. Although the processing sequence can have an impact on accuracy [3], we set a fixed and arbitrary ordering whereby given an estimate $p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1}^1,\ldots,\mathbf{Z}_{k-1}^L)$ at time $k-1$, a measurement from sensor 1 at time k provides an estimate $p(\mathbf{x}_k|\mathbf{z}_k^1,\mathbf{Z}_{k-1}^1,\ldots,\mathbf{Z}_{k-1}^L)$. The prior ahead of the update from the next sensor measurement (still at time k) is then computed according according to

$$\mathbf{x}_k \;=\; \mathbf{x}_{k-1}, \tag{4.5.12}$$

and the state transition matrix is set to $\pi = \mathbf{I}_{3\times 3}$. This construction effectively fixes the state dynamics during the sequential processing of the simultaneous measurements.

Tracking Algorithms

We assess the performance of the shifted Rayleigh mixture filter for 3D measurements (SRMF3C) with that of the well-known IMM-EKF, the IMM-UKF and also an implementation of the IMM algorithm with model-matched SRF3Cs (labeled IMM-SRF3C). Finally, we include results from a multiple-model particle filter with SRF generated proposal distribution (MM-SRPF). In all simulations, we implemented the SRMF3C using 100 mixture components (reduced at the end of each time step using the Fearnhead-Clifford method of 3.1.3), whereas the sample size of the MM-SRPF is set to $N = 2500$ particles.

200 MC runs of the above tracking scenario are computed, for different values of p_c. Initialization of the filters is carried out according to the method described in section 4.1.1, using the parameters of table 4.5.

[3] The problems of data fusion and sensor management are covered in [7]

Table 4.5: Initializing parameters of target state

Parameter	Value
initial range mean	$\bar{r}_0 = 41.3$ km
initial range standard deviation	$\sigma_{r_0} = 20$ km
initial speed mean	$\bar{s}_0 = 400$ m/s
initial speed standard deviation	$\sigma_{s_0} = 450$ m/s
initial course mean	$\bar{c}_0 = 0$ rad
initial course standard deviation	$\sigma_{c_0} = \pi/\sqrt{12}$ rad
initial pitch mean	$\bar{p}_0 = 0$ rad
initial pitch standard deviation	$\sigma_{p_0} = \pi/\sqrt{80}$ rad

Simulation results and discussion

Figures 4.16 show the range and velocity RMSEs obtained from the filters under consideration, using the first set of models. A high probability $p_c = 0.5$ of clutter was simulated. The PCRLB, as calculated in 4.2 is also included, to give an (optimistic) indication of the lowest achievable RMS error.

We can see that after the initial transients, all filters track the target accurately during its constant velocity legs, up until the second set of maneuvers at $k = 110$. The natural transients starting at the onset of the maneuvers at $k = 40$ and $k = 60$ are acceptable, given the target range and the relatively low parallax in the measurements. However, once the target starts maneuvering around sensor 1 (from $k = 110$ to $k = 141$), the IMM-EKF and IMM-UKF are seen to quickly lose track and diverge. The IMM-SRF3C also produces large errors but manages to avoid catastrophic failure. The SRMF3C, on the other hand, copes very well with the extreme nonlinearities and tracks the target, while achieving a much lower peak range RMSE than the IMM-SRF3C, and matching the level of tracking accuracy of the MM-SRPF.

Comparison of the mode probabilities derived by the various filters in figure 4.17 exposes the shortcomings of the IMM algorithms in relation to the Gaussian mixture representation of the state by the SRMF3C. While all algorithms exhibit apparently similar mode probabilities, the IMM based filters display erratic mode discrimination during the final maneuvering motion phase. The IMM-EKF and IMM-UKF break down at $t_k = 114$, when the azimuth bearing rate (at sensor 1) peaks. This clearly

points to a failure of these algorithms to cope with the measurement nonlinearity (which is compounded by the presence of clutter). As seen in figure 4.16, the IMM-SRF3C copes with the extreme nonlinearity without catastrophical failure and achieves the best tracking accuracy among the IMM based algorithms. However, this comes at the cost of large, and possibly unacceptable, transients. These are caused by the excessive posterior probability accorded to mode 2, upon detection of a maneuver, which results in the state estimate to 'overshoot' because of the high noise intensity of the model. On the other hand, both the SRMF3C and MM-SRPF display accurate mode detection. It can also be seen that upon detection of a maneuver, neither of these algorithms attach too high a probability to mode 2, thereby avoiding high transients in estimation accuracy.

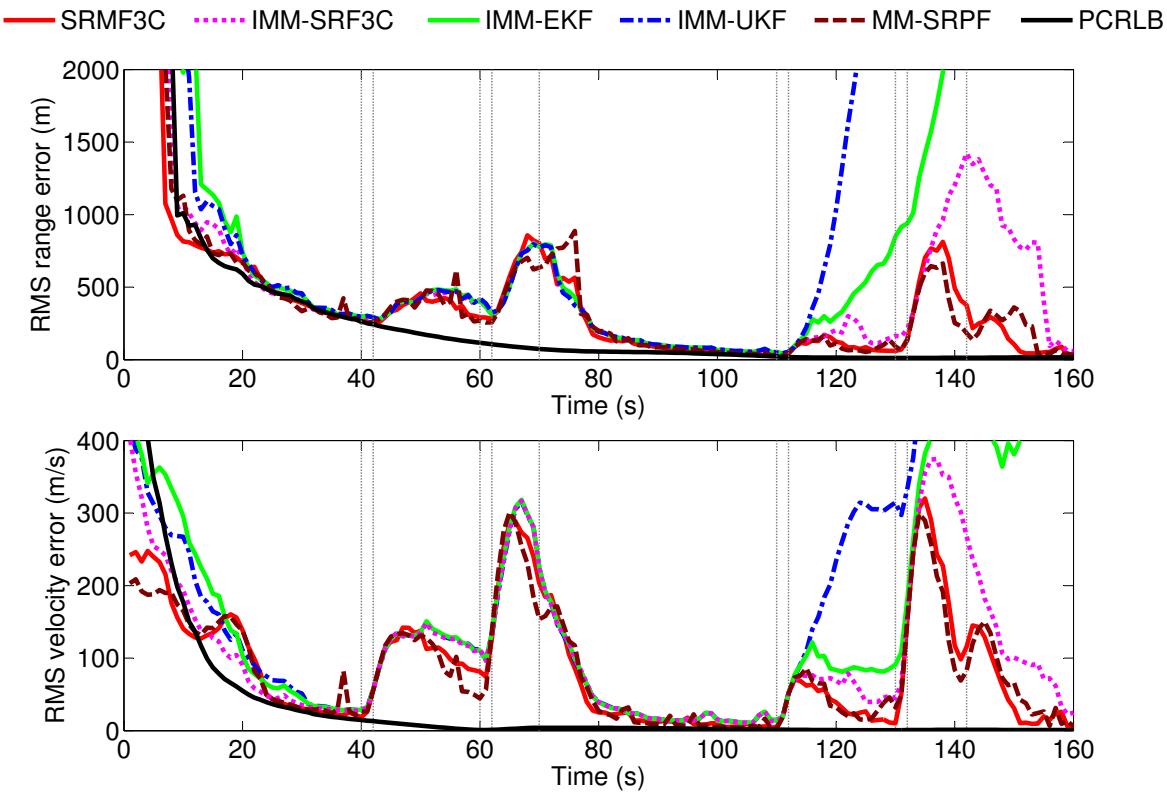


Figure 4.16: Range and velocity RMSEs of the target state estimates for $p_c = 0.5$ using the *first* set of models, over 200 MC runs.

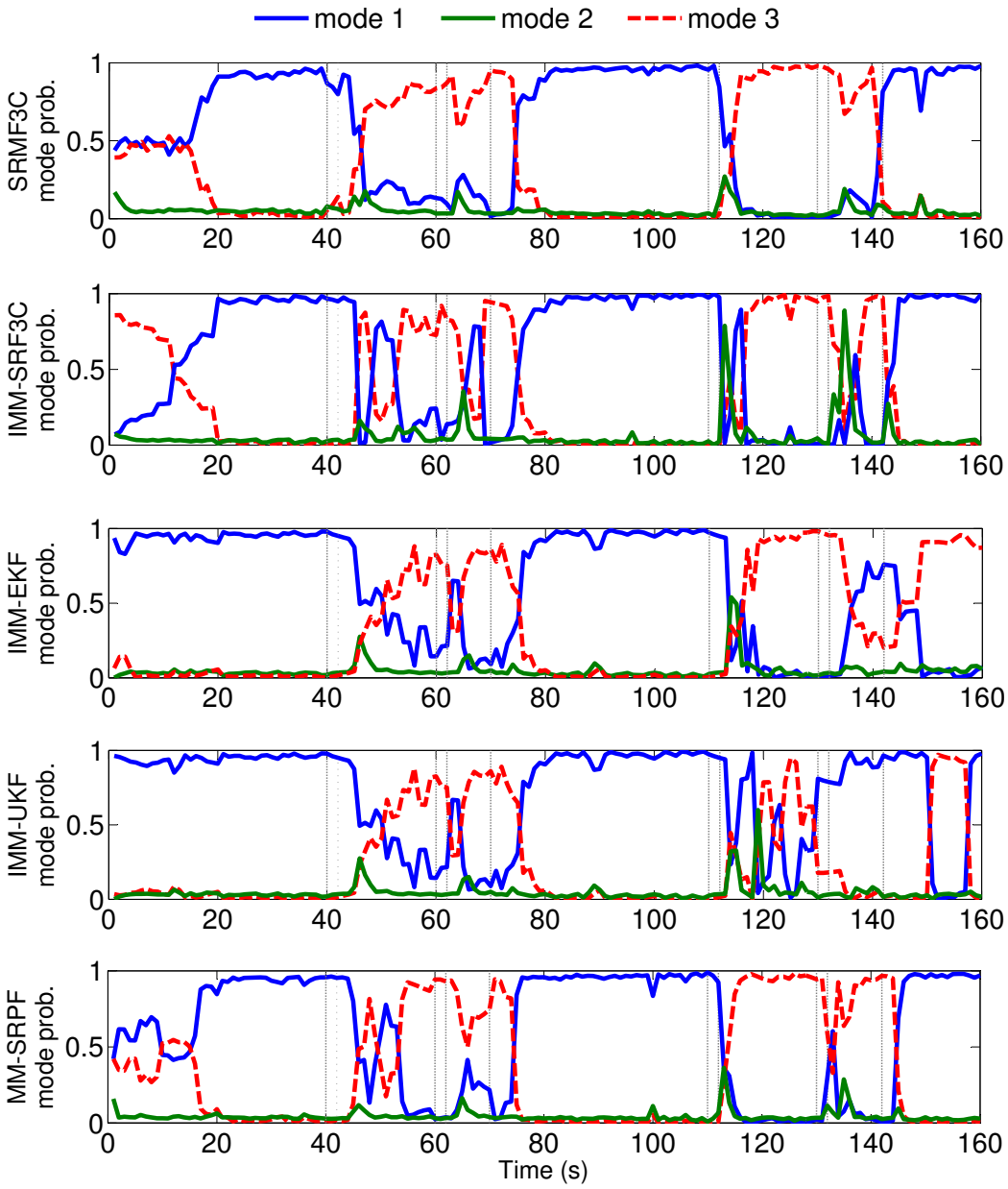


Figure 4.17: Mode probability estimates for $p_c = 0.5$ using the *first* set of models, over 200 MC runs.

The range TARMSEs of the estimates from the different filters are summarized in table 4.6, for probabilities of clutter ranging from $p_c = 0$ to $p_c = 0.7$. Due to the practically consistent failure of the IMM filters for high probabilities of clutter, their errors are not included. We see that the SRMF3C actually outperforms the MM-SRPF for low to moderate levels of clutter, but is surpassed by it for scenarios with extreme probabilities of clutter. For $p_c = 0.5$, the SRMF3C is on average 42% more accurate than the IMM-SRF3C and 68% more accurate than the IMM-EKF.

Table 4.6: TARMS range errors (m) - *first* set of models

	probability of clutter p_c							
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
SRMF3C	124	133	147	169	211	273	393	808
IMM-SRF3C	139	151	172	204	265	469	589	-
IMM-EKF	145	197	217	318	568	855	-	-
IMM-UKF	151	378	729	895	955	-	-	-
MM-SRPF	129	137	153	169	211	275	386	743

The probability of divergence of the 4 filters is illustrated in figure 4.18, against clutter probability. For this scenario, the maximum range RMSE that needed to be exceeded (for 2 consecutive time steps) for a divergence to be recorded was set to $D_{th} = 1500$ m. It can be seen that the SRMF3C and MM-SRPF show the highest robustness to clutter, followed by the IMM-SRF3C and the IMM-EKF. Differences are particularly noticeable for probabilities of clutter above $p_c = 0.5$. Another important point to note is that, although neither the IMM-EKF nor the IMM-UKF shows acceptable accuracy or robustness to clutter, the former is seen to consistently outperform the latter on both fronts. This indicates that despite the theoretically superior accuracy of the unscented transform to that of analytical linearization, the latter can still, in practice, provide better results.

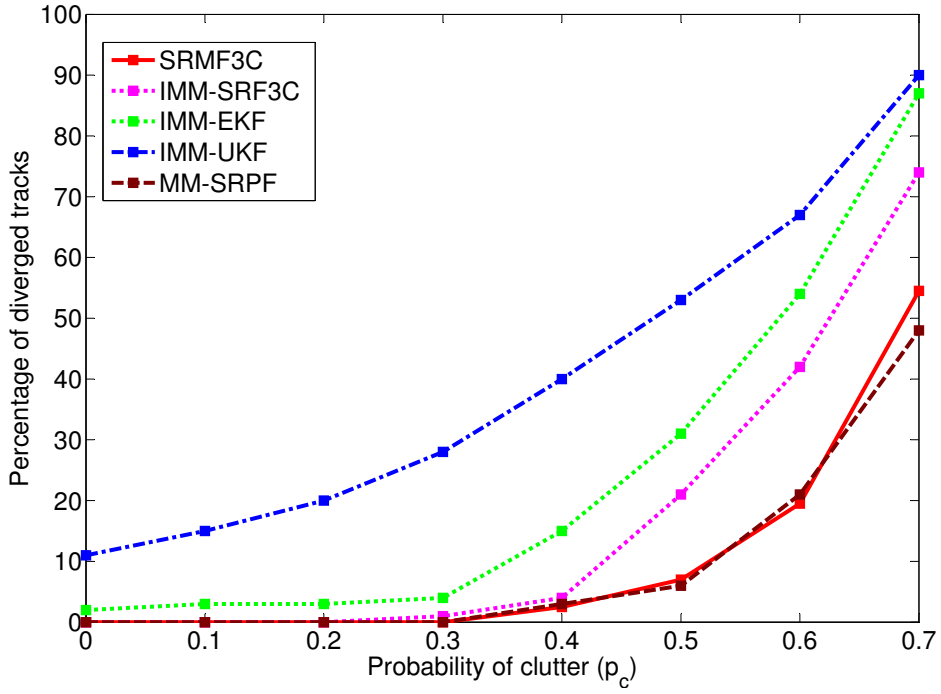


Figure 4.18: Percentage of track divergence against probability of clutter, using the *first* set of models.

The performance of the 4 trackers are now reported, under the same conditions as above, but using the *second* set of models. Figure 4.19 shows the range and velocity RMSEs, while the corresponding mode probability estimates are illustrated in figure 4.20. Table 4.7 summarizes the range TARMSEs, over increasing probability of clutter. Finally, figure 4.21 shows the probability of divergence, against probability of clutter, of all 4 algorithms.

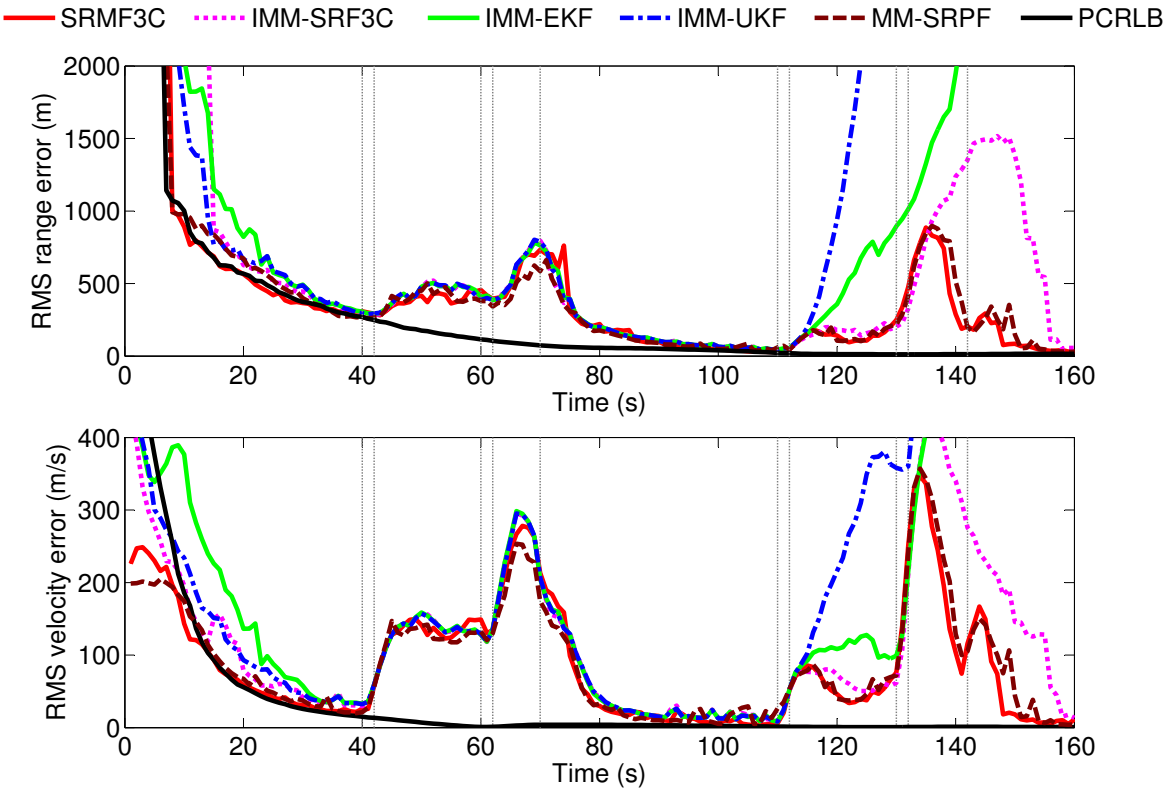


Figure 4.19: Range and velocity RMSEs of the target state estimates for $p_c = 0.5$ using the *second* set of models, over 200 MC runs.

Table 4.7: TARMS range errors (m) - *second* set of models

	probability of clutter p_c							
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
SRMF3C	136	147	168	189	232	312	526	1070
IMM-SRF3C	141	153	177	212	279	513	738	-
IMM-EKF	148	230	388	698	860	-	-	-
IMM-UKF	153	601	900	-	-	-	-	-
MM-SRPF	139	147	165	183	230	303	478	849

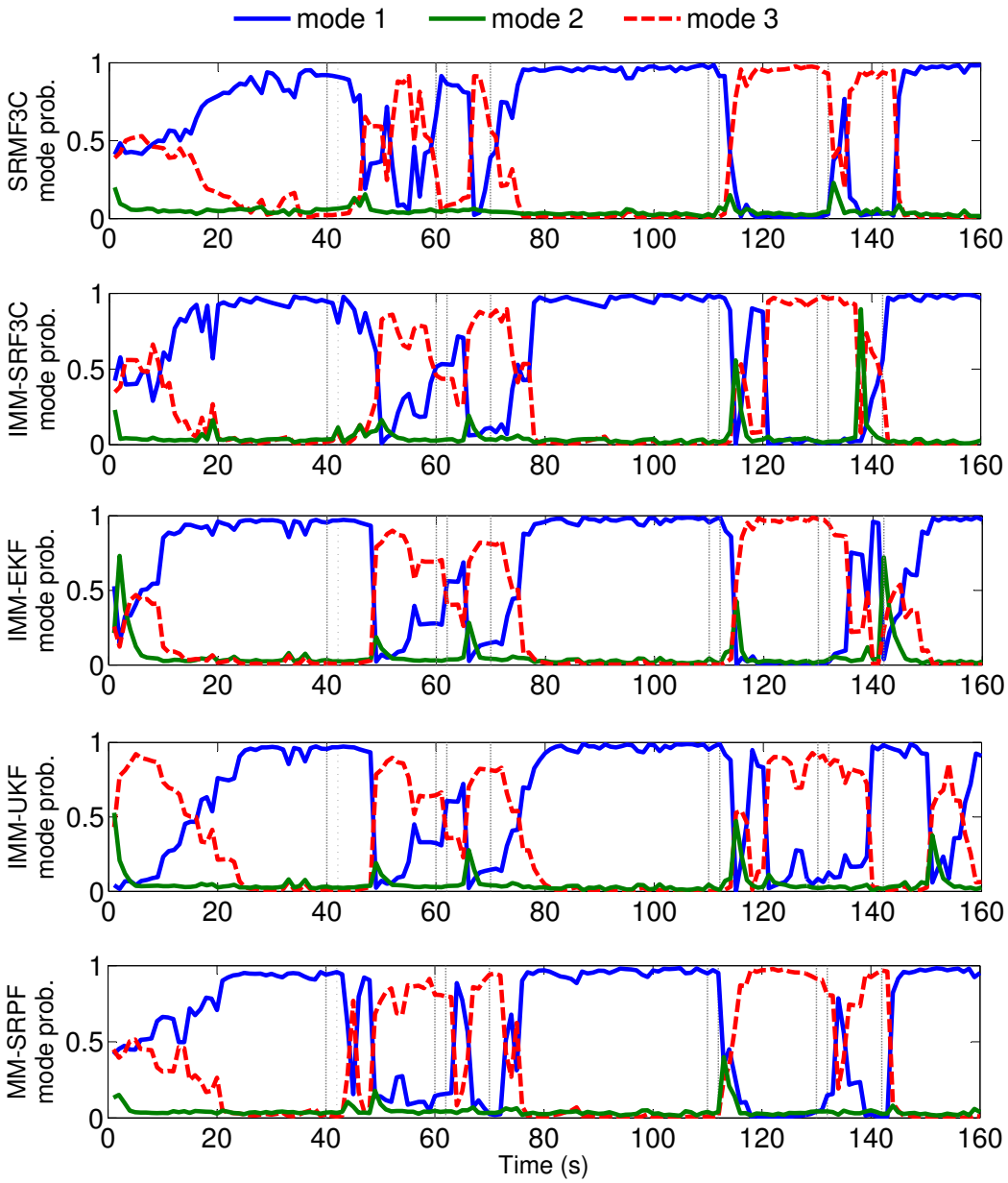


Figure 4.20: Mode probability estimates for $p_c = 0.5$ using the *second* set of models, over 200 MC runs.

The close similarity between the results obtained with the first and second sets of models points to the adequacy of employing approximate dynamic models where knowledge of exact target kinematics cannot realistically be assumed. Interestingly, even though a decline is observed in the performance of all filters, at low levels of p_c, the IMM-SRF3C suffers comparativelly lower losses in accuracy than the SRMF3C and the MM-SRPF.

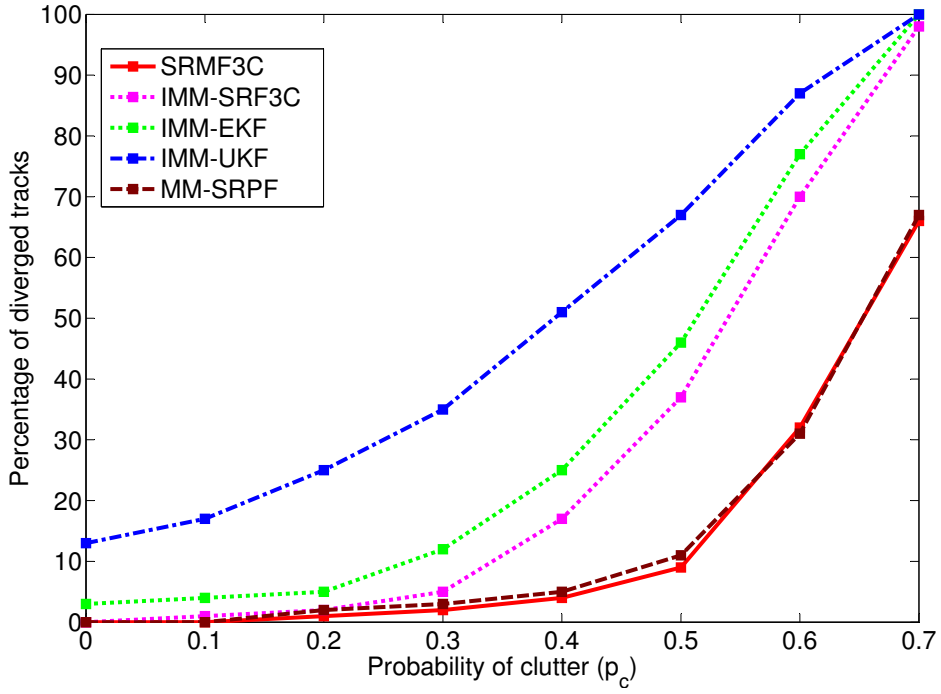


Figure 4.21: Percentage of track divergence against probability of clutter, using the *second* set of models.

In view of the accuracy and robustness of the SRMF3C, and considering the high computational efficiency (shown in table 4.8) in relation to a particle filter of equal accuracy, we can conclude that the SRMF3C is a very competitive algorithm for challenging tracking problems where highly nonlinear target dynamics and high probability of clutter undermine the effectiveness of the traditionally preferred IMM algorithm.

Table 4.8: Computation times (relative)

	Filter				
	IMM-EKF	IMM-SRF3C	IMM-UKF	SRMF3C	MM-SRPF
time	1	1.84	2.63	136	2540

Chapter 5

The Shifted Rayleigh Filter in Bearings-only SLAM

One of the most important and challenging problems in the robotics and computer vision communities is that of Simultaneous Localization and Mapping (SLAM). It is the processes whereby an agent, such as a vehicle or a robot, moves randomly in an unknown environment where it observes features and constructs a map of its surroundings, while simultaneously inferring its own position and orientation within this map. The past twenty years have seen the accumulation of a considerable body of research in this field, ranging from on-line Bayesian estimation-theoretic approaches (initiated by the results of [30], [70] and [69]) and numerical solutions ([75]) to more recently introduced off-line batch methods such as GraphSLAM ([76]). SLAM has a multitude of application areas, including autonomous indoor and outdoor navigation, as well as airborne and submarine exploration ([74], [57]).

By far the most studied and challenging area of the SLAM problem is that of bearings-only SLAM. Whereas the general SLAM problem presumes the availability of range and bearing information of landmarks (measured relative to the sensor position and heading), bearings-only SLAM is limited to measurements of relative bearings of these landmarks. The lack of range information poses a major challenge in the initialization of newly observed landmarks. This is because the non-invertibility of the measurement function results in range ambiguity in individual angular measurements. The high uncertainty in the initial range of landmarks can then violate the local measurement linearity assumption of the EKF algorithm in Cartesian coordinates (which can otherwise be used satisfactorily in range-bearing SLAM systems [25]). Moreover, linearization errors are also known to precipitate the occurrence of inconsistencies ([48], [13]), which

can lead to filter divergence. Consequently, robust and consistent bearings-only SLAM algorithms have been a topic of continued research.

In view of the superior accuracy achieved by the SRF in 'conventional' bearings-only target tracking problems and its robustness to poor target initialization, which are due largely to the absence of linearization errors in the measurement update stage, our aim is to investigate the applicability of the SRF algorithm to SLAM. This chapter is organized as follows. Section 5.1 provides a formal definition of the bearings-only SLAM problem, followed by an overview of one of the main approaches used in the literature. Two new bearings-only SLAM algorithms incorporating the SRF are then proposed in sections 5.3 and 5.4. Finally, section 5.5 provides comparative simulation results of these approaches and of one of the main competing methods.

5.1 Formulation of the Bearings-Only SLAM problem

SLAM is concerned with the *joint* estimation of the pose of the sensing device, referred to henceforth as 'vehicle' (described by its position and orientation), and of a map (made up of fixed landmarks). We denote these quantities by $\mathbf{x}_k^v$ and $\mathbf{m}_k$, composed respectively of the Cartesian position coordinates and heading angle of the vehicle (which can also be interpreted as the direction of its velocity vector), in relation to some arbitrary reference frame, and of the Cartesian position coordinates of the L initialized landmarks in the map. Together, they make up the joint vehicle and map state $\mathbf{x}_k$, given by:

$$\mathbf{x}_k \;=\; \begin{bmatrix} \mathbf{x}_k^v & \mathbf{m}^1 & \dots & \mathbf{m}^L \end{bmatrix}^{\mathrm{T}} \;=\; \begin{bmatrix} x_k^v & y_k^v & \theta_k^v & x^1 & y^1 & \dots & x^L & y^L \end{bmatrix}^{\mathrm{T}}. \qquad (5.1.1)$$

The system equation describing the dynamics of the vehicle (the landmarks being fixed), is given by

$$\mathbf{x}_k^v \;=\; f\left(\mathbf{x}_{k-1}^v, \mathbf{u}_k^s, \mathbf{v}_{k-1}\right), \qquad (5.1.2)$$

where f is a possibly nonlinear function of the previous vehicle state, an exogenous control input $\mathbf{u}_k^s$ applied to the vehicle state during the time interval $[k-1, k)$, and zero-mean white Gaussian noise $\mathbf{v}_{k-1}$ of known covariance $\mathbf{Q}_{k-1}^s$. Most SLAM scenarios assume knowledge of the sequence of *measured* control inputs $\{\mathbf{u}_k^s\}_{k\geq 1}$, composed of

speed and steering components. In situations where such measurements are unavailable, for instance where the sensor is an agile camera with no odometry capabilities, 'nearly' constant velocity or acceleration models have been successfully used in approximating the sensor dynamics ([23], [37]).

The measurements in bearings-only SLAM are defined, through the function h, as

$$\begin{aligned} z_k^l &= h\left(\mathbf{x}_k, l, w_k\right) \\ &= \arctan\left(\frac{y_k^v - y^l}{x_k^v - x^l}\right) - \theta_k^v + w_k, \end{aligned} \tag{5.1.3}$$

where l denotes the index of the observed landmark and w_k is zero-mean white Gaussian noise of known variance σ^2.

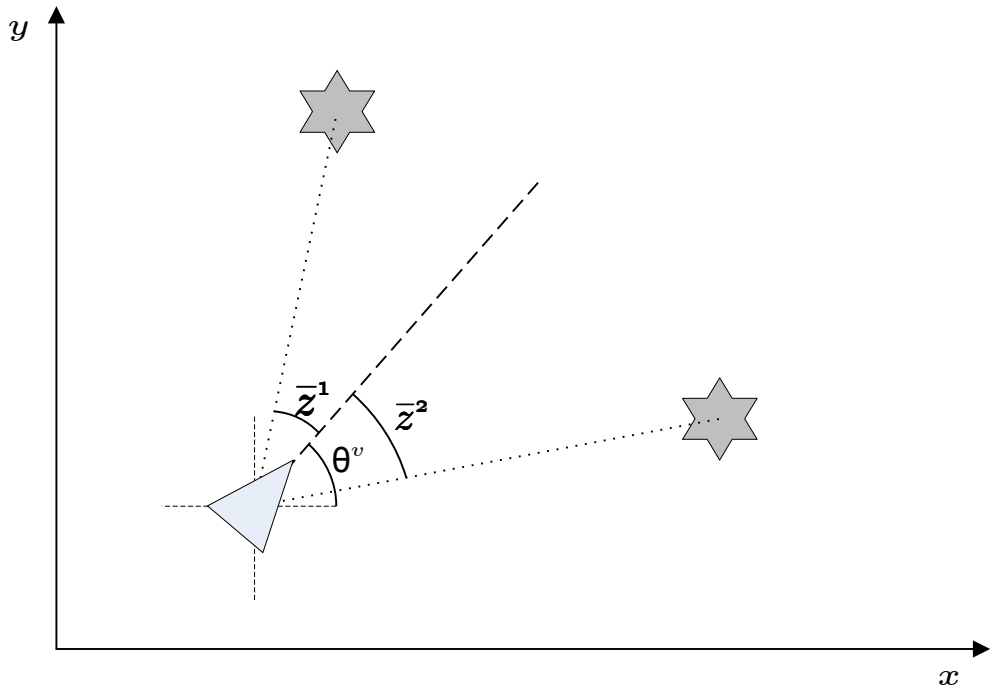


Figure 5.1: Geometry of bearings-only SLAM on the plane. Stars represent detected landmarks, of which bearings measurements $\bar{z}^1$ and $\bar{z}^2$ are acquired by the vehicle (triangle), with respect to its own heading θ^v.

It is important to note that the uncertainty associated with the vehicle pose has a fundamental consequence on the nature of the SLAM problem. Landmark position estimates

will inevitably be correlated with the vehicle pose, since they are observed with respect to the uncertain position and orientation of the vehicle. Furthermore, through their common uncertainty (relative to the vehicle pose), estimates of landmark positions will also be correlated among each other. The effect of this dependency will be that a correction to the estimate of any component of the state vector will lead to a correction to all the other states, as well as an inevitable increase in all correlations [25]. As also established in [14], the evaluation of correlations is therefore critical to the accuracy and consistency of SLAM. This has lead to the well-known Bayesian 'full-covariance' SLAM algorithm, for the estimation of the (assumed Gaussian) joint vehicle and map posterior

$$p(\mathbf{x}_k^v, \mathbf{m} | \mathbf{Z}_{0:k}, \mathbf{u}_{0:k}^s, \mathbf{x}_0), \tag{5.1.4}$$

where $\mathbf{Z}_{0:k}$ is the set of observations of all landmarks up to time k, $\mathbf{u}_{0:k}^s$ is the set of measured control inputs applied to the vehicle state up to time k and $\mathbf{x}_0$ is the initial state. This density is represented by its mean and covariance matrix

$$\hat{\mathbf{x}}_{k|k} = \begin{bmatrix} \hat{\mathbf{x}}_{k|k}^v & \hat{\mathbf{m}}_{k|k}^1 & \dots & \hat{\mathbf{m}}_{k|k}^L \end{bmatrix}^{\mathrm{T}} \tag{5.1.5}$$

$$\mathbf{P}_{k|k} = \begin{bmatrix} \mathbf{P}_{k|k}^{vv} & \mathbf{P}_{k|k}^{v1} & \dots & \mathbf{P}_{k|k}^{vL} \\ \mathbf{P}_{k|k}^{1v} & \mathbf{P}_{k|k}^{11} & \dots & \mathbf{P}_{k|k}^{1L} \\ \mathbf{P}_{k|k}^{2v} & \mathbf{P}_{k|k}^{21} & \dots & \mathbf{P}_{k|k}^{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{k|k}^{Lv} & \mathbf{P}_{k|k}^{L1} & \dots & \mathbf{P}_{k|k}^{LL} \end{bmatrix} \tag{5.1.6}$$

which are augmented, as new landmarks are initialized, with their corresponding mean and covariance estimates.

5.2 Bearings-only EKF SLAM

The EKF is known to provide satisfactory accuracy and consistency in traditional range-bearing SLAM applications ([29]). However, owing to the lack of landmark range observability from individual measurements, direct implementations of the EKF in bearings-only SLAM in most cases leads to catastrophic failure. A widely adopted framework to manage landmark range uncertainty is the so-called 'Inverse Depth

Parametrization' (IDP). The idea, introduced in [16], is to replace the highly nonlinear measurement function (5.1.3) by an equivalent one, expressed in terms of a modified set of state variables, which exhibits a much higher degree of linearity. More specifically, instead of using Cartesian coordinates to represent landmark positions, IDP allows the parametrization of landmark positions using the inverse of their depth (or inverse range) when first detected by the vehicle. Whereas range uncertainty cannot be accurately modelled as Gaussian (unless the landmark is reasonably well localized), uncertainty in inverse depth actually can. As a result, IDP allows the propagation and update of pdfs which can safely be approximated as being Gaussian, in a manner suitable for the linear EKF equations to be used. We provide here only an overview of IDP, of which detailed analyses and results can be found in [16] and [15].

The state vector in IDP is defined as

$$\begin{aligned}\mathbf{x}_k &= \begin{bmatrix} \mathbf{x}_k^v & \mathbf{m}^1 & \dots & \mathbf{m}^L \end{bmatrix}^{\mathrm{T}} \\ &= \begin{bmatrix} x_k^v & y_k^v & \theta_k^v & x_k^{v,1} & y_k^{v,1} & \rho_k^1 & \theta_k^1 & \dots & x_k^{v,L} & y_k^{v,L} & \rho_k^L & \theta_k^L \end{bmatrix}^{\mathrm{T}}, \end{aligned} \tag{5.2.1}$$

where each landmark $\mathbf{m}^l = \psi(x_k^{v,l}, x_k^{v,l}, \rho_k^l, \theta_k^{v,l})$ is represented in terms of the position $(x_k^{v,l}, x_k^{v,l})$ of the vehicle from which the landmark was first observed (i.e. initialized), its inverse depth ρ_k^l and its absolute bearing $\theta_k^l = z_k^l + \theta_k^{v,l}$ at initialization.

The bearings measurements are then expressed, through the function h_{IDP}, as

$$\begin{aligned} z_k^l &= h_{\text{IDP}}\left(\mathbf{x}_k, l, w_k\right) \\ &= \arctan\left(\frac{y_k^{v,l} - y_k^v + \sin(\theta_k^l)(\rho_k^l)^{-1}}{x_k^{v,l} - x_k^v + \cos(\theta_k^l)(\rho_k^l)^{-1}}\right) - \theta_v^k + w_k. \end{aligned} \tag{5.2.2}$$

As detailed in [16], this parametrization permits the estimation of landmarks at depths ranging from near zero to infinity. Moreover, the resulting measurement function is very close to being linear for both low and high parallax (i.e. irrespective of the length of the baseline covered between two successive observations of a landmark). This allows modelling of the inverse depth of a landmark as a Gaussian random variable, and therefore enables the implementation of a standard EKF for state estimation. A major drawback of

IDP, however, is the possibility of landmark depth estimates becoming negative [59]. This issue, which arises from the inability of the EKF to accommodate nonlinear constraints, results in catastrophic failure.

Upon initialization of a landmark, the state vector estimate is augmented, with the IDP representation of the new landmark, as

$$\begin{aligned} \hat{\mathbf{x}}_{k|k}^{\text{new}} &= \begin{bmatrix} \hat{\mathbf{x}}_{k|k} & \hat{\mathbf{m}}_l \end{bmatrix}^{\mathrm{T}} \\ &= \begin{bmatrix} \hat{\mathbf{x}}_{k|k} & \hat{x}_{k|k}^v & \hat{y}_{k|k}^v & \hat{\rho}_{k|k}^l & \hat{\theta}_{k|k}^l \end{bmatrix}^{\mathrm{T}}, \end{aligned} \tag{5.2.3}$$

while the $N \times N$ covariance matrix is extended to

$$\mathbf{P}_{k|k}^{\text{new}} = \mathbf{J} \begin{bmatrix} \mathbf{P}_{k|k} & \mathbf{0}_{N\times 4} \\ \mathbf{0}_{4\times N} & \mathbf{P}_{k|k}^{\text{ext}} \end{bmatrix} \mathbf{J}^{\mathrm{T}} \tag{5.2.4}$$

with

$$\mathbf{P}_{k|k}^{\text{ext}} = \begin{bmatrix} P_{k|k}^{x^v} & P_{k|k}^{x^v\, y^v} & 0 & P_{k|k}^{x^v\, \theta^v} \\ P_{k|k}^{y^v\, x^v} & P_{k|k}^{y^v} & 0 & P_{k|k}^{y^v\, \theta^v} \\ 0 & 0 & \sigma_\rho^2 & 0 \\ P_{k|k}^{\theta^v\, x^v} & P_{k|k}^{\theta^v\, y^v} & 0 & P_{k|k}^{\theta^v} + \sigma^2 \end{bmatrix} \tag{5.2.5}$$

and

$$\mathbf{J} = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{\mathbf{I}_{N\times N}} & \mathbf{0}_{N\times 4} \\ \hline \nabla_{\mathbf{x}^v}\, \psi & \mathbf{0}_{4\times N} & \mathbf{I}_{4\times 4} \end{array}\right], \tag{5.2.6}$$

where $\nabla_{\mathbf{x}^v}\, \psi$ represents the Jacobian of the IDP transformation with respect to the vehicle states.

Although the IDP representation of landmarks can be maintained throughout the entire SLAM process, for reasons of computational efficiency it is advantageous to transform estimates of well localized landmarks into their Cartesian position representation. A test is proposed in [17], by which the linearity of the standard measurement function

(5.1.3) is evaluated. If it is sufficiently high, the landmark estimate is converted from IDP to the more compact xy coding using the relations

$$\hat{\mathbf{m}}_{k|k}^{l,\text{cart}} = g(\hat{\mathbf{x}}_{k|k}, l) = \begin{bmatrix} \hat{x}_{k|k}^v + \cos(\hat{\theta}_k^l)(\hat{\rho}_{k|k}^l)^{-1} \\ \hat{y}_{k|k}^v + \sin(\hat{\theta}_k^l)(\hat{\rho}_{k|k}^l)^{-1} \end{bmatrix} \tag{5.2.7}$$

$$\mathbf{P}_{k|k}^{l,\text{cart}} = \nabla_{x^{\text{cart}}} g(\hat{\mathbf{x}}_{k|k}, l)\, \mathbf{P}_{k|k}^l\, \nabla_{x^{\text{cart}}}^{\text{T}} g(\hat{\mathbf{x}}_{k|k}, l), \tag{5.2.8}$$

where $\nabla_{x^{\text{cart}}} g$ is the Jacobian of the IDP to cartesian transformation of the landmark at time k.

5.3 The quadrature-SRF SLAM algorithm

We propose in this section a new method for bearings-only SLAM, called the quadrature-SRF SLAM algorithm. It is a SRF based moment matching filter, in which the exact first and second moments of the posterior density are numerically evaluated, given a Gaussian prior. In contrast with 'standard' bearings-only target tracking, the measurement equation defining bearings-only SLAM does not allow straightforward implementation of the 'noise before nonlinearity' substitution that makes the elegant evaluation of exact moments possible. More precisely, the incorporation of the angular shift caused by the unknown heading of the vehicle in (5.1.3) within the arctan function is not trivial. As a result, direct evaluation of this quantity by the SRF equations is not possible. We propose therefore an alternative formulation of the moments of the posterior, expressed in terms of the standard SRF equations.

5.3.1 Derivation of the Algorithm

We define for the quadrature-SRF SLAM algorithm a state vector different from the one in (5.1.1), which makes it possible to evaluate the heading angle of the vehicle in terms of the SRF equations. More precisely, we introduce vehicle velocity states $\mathbf{x}_k^v = \begin{bmatrix} \dot{x}_k & \dot{y}_k \end{bmatrix}$ and interpret the heading angle as the direction of this vector. The remaining states $\mathbf{x}_k^p$ are those of the vehicle position and landmark positions. The new N dimensional state vector is thus defined as

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}_k^v & \mathbf{x}_k^p \end{bmatrix}^{\text{T}} = \begin{bmatrix} \dot{x}_k & \dot{y}_k & x_k & y_k & x^1 & y^1 & \dots & x^L & y^L \end{bmatrix}^{\text{T}}. \tag{5.3.1}$$

Now, let $\mathbf{Z}_k$ denote the vector of bearing measurement of all M observed landmarks at time k, and $\mathbf{Z}_k^*$ the corresponding bearing measurement vector in the 'global' frame of reference:

$$\mathbf{Z}_k = \angle(H\mathbf{x}_k^{(p)} + w_k) - \angle(\mathbf{x}_k^v) \tag{5.3.2}$$

$$\mathbf{Z}_k^* = \angle(H\mathbf{x}_k^{(p)} + w_k) \tag{5.3.3}$$

The aim of the estimation problem is to evaluate the filtering density

$$\begin{aligned} p(\mathbf{x}_k \mid \mathbf{Z}_{1:k}) &= \frac{1}{c}\, p(\mathbf{Z}_k \mid \mathbf{x}_k, \mathbf{x}_{1:k-1})\, p(\mathbf{x}_k \mid \mathbf{x}_{1:k-1}) \\ &= \frac{1}{c}\, p(\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v) \mid \mathbf{x}_k)\, p(\mathbf{x}_k \mid \mathbf{x}_{1:k-1}) \end{aligned} \tag{5.3.4}$$

The transition between the two lines above is a result of the equivalence between the measurement likelihoods of $\mathbf{Z}_k^*$ and $\mathbf{Z}_k$ up to a constant angular shift. We note that an expression for the product $p(\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v) \mid \mathbf{x}_k)\, p(\mathbf{x}_k \mid \mathbf{x}_{1:k-1})$ is available, in terms of a Gaussian density with first and second order moments matched to those of a related density, via the Shifted Rayleigh filter equations. Indeed, the simpler filtering problem of estimating the state $\mathbf{x}_k$ given the *absolute* bearings measurements $\mathbf{Z}_{1:k}^*$ can be expressed as

$$p(\mathbf{x}_k \mid \mathbf{Z}_{1:k}^*) \;=\; \frac{p(\mathbf{Z}_k^* \mid \mathbf{x}_k)\, p(\mathbf{x}_k \mid \mathbf{x}_{1:k-1})}{p(\mathbf{Z}_k^* \mid \mathbf{x}_{1:k-1})} \;\approx\; \mathcal{N}\left(\mathbf{x}_k; (\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}), (\mathbf{Z}_k^*)\right), \tag{5.3.5}$$

where $\mathcal{N}\left(\mathbf{x}_k; (\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})(\mathbf{Z}_k^*)\right)$ represents the Gaussian approximation, with mean $\hat{\mathbf{x}}_{k|k}$ and covariance $\mathbf{P}_{k|k}$, of the state $\mathbf{x}$ at time k (evaluated using the SRF equations (3.1.5) - (3.1.13)), given the measurements $\mathbf{Z}_k^*$. Combining equations (5.3.4) and (5.3.5), we find that the density of interest can be expressed as

$$p(\mathbf{x}_k \mid \mathbf{Z}_{1:k}) \;=\; \frac{1}{c}\, p(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \angle(\mathbf{x}_{k|k-1}^v))\, \mathcal{N}\left(\mathbf{x}_k; (\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}), (\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v))\right). \tag{5.3.6}$$

Although this density cannot be written in closed form or in terms of standard library functions, its moments can nonetheless be evaluated numerically. We now detail the steps involved in these computations.

Evaluation of the normalizing constant

We note that by the law of total probability we have

$$\int_{\mathbf{x}_k^v}\int_{\mathbf{x}_k^p} p(\mathbf{x}_k \mid \mathbf{Z}_{1:k})\, d\mathbf{x}_k^v\, d\mathbf{x}_k^p \;=\; 1$$

The constant c can then be evaluated as

$$\begin{aligned}
c \;&=\; \int_{\mathbf{x}_k^v}\int_{\mathbf{x}_k^p} p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \angle(\mathbf{x}_{k|k-1}^v)\right)\, \mathcal{N}\left(\left(\mathbf{x}_k^v, \mathbf{x}_k^p\right); \left(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}\right), \left(\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v)\right)\right)\, d\mathbf{x}_k^v\, d\mathbf{x}_k^p \\
&= \int_{\mathbf{x}_k^v} p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \angle(\mathbf{x}_{k|k-1}^v)\right) \left[\int_{\mathbf{x}_k^p} \mathcal{N}\left(\left(\mathbf{x}_k^p \mid \mathbf{x}_k^v\right); \left(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}\right), \left(\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v)\right)\right)\, d\mathbf{x}_k^p\right] \cdot \\
&\quad \mathcal{N}\left(\mathbf{x}_k^v; \left(\hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^v\right), \left(\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v)\right)\right)\, d\mathbf{x}_k^v \\
&= \int_{\mathbf{x}_k^v} p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \angle(\mathbf{x}_{k|k-1}^v)\right)\, \mathcal{N}\left(\mathbf{x}_k^v; \left(\hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^v\right), \left(\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v)\right)\right)\, d\mathbf{x}_k^v.
\end{aligned}$$

We now map the vehicle velocity vector $\mathbf{x}_k^v$ into its polar form $[s_k, \theta_k]^{\mathrm{T}}$, representing the speed and heading of the vehicle. The previous expression can then be written as

$$\begin{aligned}
c \;&=\; \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \theta_k\right) \cdot \\
&\qquad \left[\int_{s_k=0}^{\infty} s_k\, \mathcal{N}\left(\left(s\cos(\theta_k), s\sin(\theta_k)\right); \left(\hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^v\right), \left(\mathbf{Z}_k^* = \mathbf{Z}_k + \theta\right)\right)\, ds_k\right]\, d\theta_k \\
&= \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \theta_k\right)\, h_1\left(\theta_k, \hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^v\right), \qquad (5.3.7)
\end{aligned}$$

where the function h_1, derived later, can be computed in terms of the error function.

Evaluation of First Moments

We now turn to the evaluation of the first moments of the density $p(\mathbf{x}_k \mid \mathbf{Z}_{1:k})$, which can be written as

$$E\left[\mathbf{x}_k \mid \mathbf{Z}_{1:k}\right] = \begin{bmatrix} E\left[\mathbf{x}_k^v \mid \mathbf{Z}_{1:k}\right] \\ E\left[\mathbf{x}_k^p \mid \mathbf{Z}_{1:k}\right] \end{bmatrix} = \begin{bmatrix} \int_{\mathbf{x}_k} \mathbf{x}_k^v \, p\left(\mathbf{x}_k \mid \mathbf{Z}_{1:k}\right) \, d\mathbf{x}_k \\ \int_{\mathbf{x}_k} \mathbf{x}_k^p \, p\left(\mathbf{x}_k \mid \mathbf{Z}_{1:k}\right) \, d\mathbf{x}_k \end{bmatrix}. \tag{5.3.8}$$

The first integral, representing the mean of the velocities can be expressed as

$$\begin{aligned}
&\int_{\mathbf{x}_k} \mathbf{x}_k^v \, p\left(\mathbf{x}_k \mid \mathbf{Z}_{1:k}\right) \, d\mathbf{x}_k \\
&= \frac{1}{c} \int_{\mathbf{x}_k^v} \int_{\mathbf{x}^p} \mathbf{x}_k^v \, p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \angle(\mathbf{x}_{k|k-1}^v)\right) \mathcal{N}\left((\mathbf{x}^v, \mathbf{x}^p); \left(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}\right), (\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v))\right) \, d\mathbf{x}^v \, d\mathbf{x}^p \\
&= \frac{1}{c} \int_{\mathbf{x}_k^v} \mathbf{x}_k^v \, p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \angle(\mathbf{x}_{k|k-1}^v)\right) \left[\int_{\mathbf{x}_k^p} \mathcal{N}\left(\left(\mathbf{x}_k^p \mid \mathbf{x}_k^v\right); \left(\hat{\mathbf{x}}_{k|k}^p, \mathbf{P}_{k|k}^p\right), (\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v))\right) \, d\mathbf{x}_k^p \right] \cdot \\
&\quad \mathcal{N}\left(\mathbf{x}_k^v; \left(\hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^v\right), (\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v))\right) \, d\mathbf{x}_k^v \\
&= \frac{1}{c} \int_{\mathbf{x}_k^v} \mathbf{x}_k^v \, p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \angle(\mathbf{x}_{k|k-1}^v)\right) \mathcal{N}\left(\mathbf{x}_k^v; \left(\hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^v\right), (\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v))\right) \, d\mathbf{x}_k^v
\end{aligned}$$

Carrying out the same cartesian to polar transformation as before, we get

$$\begin{aligned}
&\int_{\mathbf{x}_k} \mathbf{x}_k^v \, p\left(\mathbf{x}_k \mid \mathbf{Z}_{1:k}\right) \, d\mathbf{x}_k \\
&= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \theta_k\right) \cdot \\
&\quad \left[\int_{s_k=0}^{\infty} \begin{bmatrix} s_k \cos(\theta_k) \\ s_k \sin(\theta_k) \end{bmatrix} s_k \, \mathcal{N}\left((s_k \cos(\theta_k), s_k \sin(\theta_k)); \left(\hat{\mathbf{x}}_k^v, \mathbf{P}_{k|k}^v\right), (\mathbf{Z}_k^* = \mathbf{Z}_k + \theta_k)\right) \, ds_k \right] \, d\theta_k \\
&= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \theta_k\right) \cdot \begin{bmatrix} \cos(\theta_k) \, h_2\left(\theta_k, \mathbf{x}_{k|k}^v, \mathbf{P}_{k|k}^v\right) \\ \sin(\theta_k) \, h_2\left(\theta_k, \mathbf{x}_{k|k}^v, \mathbf{P}_{k|k}^v\right) \end{bmatrix} \, d\theta_k,
\end{aligned} \tag{5.3.9}$$

where the function h_2 derived later is computable in terms of the error function. Going back to (5.3.8), the second integral can similarly be written as

$$\int_{\mathbf{x}_k} \mathbf{x}_k^p \, p\left(\mathbf{x}_k | \mathbf{Z}_{1:k}\right) \, d\mathbf{x}_k$$

$$= \frac{1}{c} \int_{\mathbf{x}_k^v} \int_{\mathbf{x}_k^p} \mathbf{x}_k^p \, p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \angle(\mathbf{x}_{k|k-1}^v)\right) \mathcal{N}\left(\left(\mathbf{x}_k^v, \mathbf{x}_k^p\right); \left(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}\right), \left(\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v)\right)\right) \, d\mathbf{x}_k^v d\mathbf{x}_k^p$$

$$= \frac{1}{c} \int_{\mathbf{x}_k^v} p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \angle(\mathbf{x}_{k|k-1}^v)\right) \left[\int_{\mathbf{x}_k^p} \mathbf{x}_k^p \, \mathcal{N}\left(\left(\mathbf{x}_k^p \mid \mathbf{x}_k^v\right); \left(\hat{\mathbf{x}}_{k|k}^p, \mathbf{P}_{k|k}^p\right) \left(\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v)\right)\right) \, d\mathbf{x}_k^p\right] \cdot$$
$$\mathcal{N}\left(\mathbf{x}_k^v; \left(\hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^v\right) \left(\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v)\right)\right) \, d\mathbf{x}_k^v$$

$$= \frac{1}{c} \int_{\mathbf{x}_k^v} p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \angle(\mathbf{x}_{k|k-1}^v)\right) \left[\hat{\mathbf{x}}_{k|k}^p + \mathbf{P}_{k|k}^{pv} \mathbf{P}_{k|k}^{vv\ -1} \left(\mathbf{x}_k^v - \hat{\mathbf{x}}_{k|k}^v\right)\right] \cdot$$
$$\mathcal{N}\left(\mathbf{x}_k^v; \left(\hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^v\right) \left(\mathbf{Z}_k^* = \mathbf{Z}_k + \angle(\mathbf{x}_k^v)\right)\right) \, d\mathbf{x}_k^v$$

$$= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \theta_k\right) \left[\int_{s_k=0}^{\infty} \left[\hat{\mathbf{x}}_{k|k}^p + \mathbf{P}_{k|k}^{pv} \mathbf{P}_{k|k}^{vv\ -1} \left(\begin{bmatrix} s_k \cos(\theta_k) \\ s_k \sin(\theta_k) \end{bmatrix} - \hat{\mathbf{x}}_{k|k}^v\right)\right] \cdot \right.$$
$$\left. s_k \, \mathcal{N}\left(\left(s_k \cos(\theta_k), s_k \sin(\theta_k)\right); \left(\hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^v\right) \left(\mathbf{Z}_k^* = \mathbf{Z}_k + \theta_k\right)\right) \, ds_k\right] \, d\theta_k,$$

Denoting by α_{ij} the entries of the $M \times 2$ matrix $\mathbf{P}_{k|k}^{pv} \mathbf{P}_{k|k}^{vv\ -1}$, regrouping factors of s_k inside the braces and suppressing the arguments of the normal probability density function (for clarity of notation), we arrive at the following expression:

$$\int_{\mathbf{x}_k} \mathbf{x}_k^p \, p\left(\mathbf{x}_k \mid \mathbf{Z}_{1:k}\right) \, d\mathbf{x}_k$$

$$= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}_{k|k-1}^* = \mathbf{Z}_{k|k-1} + \theta_k\right) \cdot$$
$$\int_{s_k=0}^{\infty} \begin{bmatrix} \left(\alpha_{11} \cos(\theta_k) + \alpha_{12} \sin(\theta_k)\right) s_k^2 \, \mathcal{N} + \left(\hat{\mathbf{x}}_{k|k,1}^p - \alpha_{11} \hat{\mathbf{x}}_{k|k,1}^v - \alpha_{12} \hat{\mathbf{x}}_{k|k,2}^v\right) r \, \mathcal{N} \\ \left(\alpha_{21} \cos(\theta_k) + \alpha_{22} \sin(\theta_k)\right) s_k^2 \, \mathcal{N} + \left(\hat{\mathbf{x}}_{k|k,2}^p - \alpha_{21} \hat{\mathbf{x}}_{k|k,1}^v - \alpha_{22} \hat{\mathbf{x}}_{k|k,2}^v\right) r \, \mathcal{N} \\ \vdots \\ \left(\alpha_{N1} \cos(\theta_k) + \alpha_{N2} \sin(\theta_k)\right) s_k^2 \, \mathcal{N} + \left(\hat{\mathbf{x}}_{k|k,N}^p - \alpha_{N1} \hat{\mathbf{x}}_{k|k,1}^v - \alpha_{N2} \hat{\mathbf{x}}_{k|k,2}^v\right) r \, \mathcal{N} \end{bmatrix} ds_k \, d\theta_k,$$

$$= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}^*_{k|k-1} = \mathbf{Z}_{k|k-1} + \theta_k\right) \cdot$$

$$\begin{bmatrix} (\alpha_{11}\cos(\theta_k) + \alpha_{12}\sin(\theta_k))\, h_2(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) + (\hat{\mathbf{x}}^p_{k|k,1} - \alpha_{11}\hat{\mathbf{x}}^v_{k|k,1} - \alpha_{12}\hat{\mathbf{x}}^v_{k|k,2})\, h_1(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) \\ (\alpha_{21}\cos(\theta_k) + \alpha_{22}\sin(\theta_k))\, h_2(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) + (\hat{\mathbf{x}}^p_{k|k,2} - \alpha_{21}\hat{\mathbf{x}}^v_{k|k,1} - \alpha_{22}\hat{\mathbf{x}}^v_{k|k,2})\, h_1(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) \\ \vdots \\ (\alpha_{N1}\cos(\theta_k) + \alpha_{N2}\sin(\theta_k))\, h_2(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) + (\hat{\mathbf{x}}^p_{k|k,N} - \alpha_{N1}\hat{\mathbf{x}}^v_{k|k,1} - \alpha_{N2}\hat{\mathbf{x}}^v_{k|k,2})\, h_1(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) \end{bmatrix} d\theta_k. \tag{5.3.10}$$

Evaluation of the integrals in (5.3.9) and (5.3.10) yield the first moment of the density $p(\mathbf{x}_k|\mathbf{Z}_{1:k})$. Although not suitable for analytical evaluation, they can be computed by numerical methods.

Evaluation of the Second Moments

The evaluation of second moments is done comparably to that of the first moments. We break down the problem by deriving expressions for $E\left[\mathbf{x}^v_k \mathbf{x}^{v\mathrm{T}}_k \mid \mathbf{Z}_{1:k}\right]$, $E\left[\mathbf{x}^p_k \mathbf{x}^{v\mathrm{T}}_k \mid \mathbf{Z}_{1:k}\right]$ and $E\left[\mathbf{x}^p_k \mathbf{x}^{p\mathrm{T}}_k \mid \mathbf{Z}_{1:k}\right]$.

Similarly to the derivation of (5.3.9) we can express $E\left[\mathbf{x}^v_k \mathbf{x}^{v\mathrm{T}}_k \mid \mathbf{Z}_{1:k}\right]$ as:

$$E\left[\mathbf{x}^v_k \mathbf{x}^{v\mathrm{T}}_k \mid \mathbf{Z}_{1:k}\right]$$

$$= \int_{\mathbf{x}_k} \mathbf{x}^v_k \mathbf{x}^{v\mathrm{T}}_k\, p\left(\mathbf{x}_k|\mathbf{Z}_{1:k}\right)\, d\mathbf{x}_k$$

$$= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}^*_{k|k-1} = \mathbf{Z}_{k|k-1} + \theta_k\right) \left[\int_{s_k=0}^{\infty} \begin{bmatrix} s_k\cos(\theta_k) \\ s_k\sin(\theta_k) \end{bmatrix} \begin{bmatrix} s_k\cos(\theta_k) & s_k\sin(\theta_k) \end{bmatrix} s_k\, \mathcal{N}\, ds_k \right] d\theta_k$$

$$= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}^*_{k|k-1} = \mathbf{Z}_{k|k-1} + \theta_k\right) \left[\int_{s_k=0}^{\infty} \begin{bmatrix} s_k^3\cos^2(\theta_k) & s_k^3\cos(\theta_k)\sin(\theta_k) \\ s_k^3\cos(\theta_k)\sin(\theta_k) & s_k^3\sin^2(\theta_k) \end{bmatrix} \mathcal{N}\, ds_k \right] d\theta_k.$$

Regrouping factors of s_k we arrive at this final expression:

$$E\left[\mathbf{x}^v_k \mathbf{x}^{v\mathrm{T}}_k \mid \mathbf{Z}_{1:k}\right]$$

$$= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}^*_{k|k-1} = \mathbf{Z}_{k|k-1} + \theta_k\right) \cdot \begin{bmatrix} \cos^2(\theta_k) h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}) & \cos(\theta_k)\sin(\theta_k) h_3(\theta, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}) \\ \cos(\theta_k)\sin(\theta_k) h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}) & \sin^2(\theta_k) h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}) \end{bmatrix}, \quad (5.3.11)$$

where again, the function h_3 is shown in the sequel.

With steps akin to those in the evaluation of (5.3.10), $E\left[\mathbf{x}^p_k \mathbf{x}^{v\mathrm{T}}_k \mid \mathbf{Z}_{1:k}\right]$ can be expressed as:

$$\begin{aligned}
&E\left[\mathbf{x}^p_k \mathbf{x}^{v\mathrm{T}}_k \mid \mathbf{Z}_{1:k}\right] \\
&= \int_{\mathbf{x}_k} \mathbf{x}^v_k \mathbf{x}^{p\mathrm{T}}_k \, p\left(\mathbf{x}_k | \mathbf{Z}_{1:k}\right) \, d\mathbf{x}_k \\
&= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}^*_{k|k-1} = \mathbf{Z}_{k|k-1} + \theta_k\right) \left(\int_{s_k=0}^{\infty} \left[\hat{\mathbf{x}}^p_{k|k} + \mathbf{P}^{pv}_{k|k} \mathbf{P}^{vv\,-1}_{k|k} \left(\begin{bmatrix} s_k \cos(\theta_k) \\ s_k \sin(\theta_k) \end{bmatrix} - \hat{\mathbf{x}}^v_{k|k} \right) \right] \cdot \right. \\
&\qquad \left. s_k \begin{bmatrix} s_k \cos(\theta_k) & s_k \sin(\theta_k) \end{bmatrix} \mathcal{N} \, ds_k \right) d\theta_k, \\
&= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}^*_{k|k-1} = \mathbf{Z}_{k|k-1} + \theta_k\right) \cdot \\
&\begin{bmatrix} \kappa_1 \cos(\theta_k)\, h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) & \kappa_1 \sin(\theta_k)\, h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) \\ +\lambda_1 \cos(\theta_k)\, h_2(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) & +\lambda_2 \sin(\theta_k)\sin(\theta_k) \\ \kappa_2 \cos(\theta_k)\, h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) & \kappa_2 \sin(\theta_k)\, h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) \\ +\lambda_2 \cos(\theta_k)\, h_2(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) & +\lambda_2 \sin(\theta_k)\sin(\theta_k) \\ \vdots & \vdots \\ \kappa_M \cos(\theta_k)\, h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) & \kappa_M \sin(\theta_k)\, h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) \\ +\lambda_M \cos(\theta_k)\, h_2(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^v_{k|k}) & +\lambda_M \sin(\theta_k)\sin(\theta_k) \end{bmatrix} d\theta_k, \quad (5.3.12)
\end{aligned}$$

where, for $i = 1, \ldots, N-2$, the factors κ_i and λ_i are given by:

$$\begin{aligned}
\kappa_i &= \alpha_{i1} + \alpha_{i2} \\
\lambda_i &= \hat{\mathbf{x}}^p_{k|k,i} - \alpha_{i1} \hat{\mathbf{x}}^v_{k|k,1} - \alpha_{i2} \hat{\mathbf{x}}^v_{k|k,2}.
\end{aligned}$$

Finally, $E\left[\mathbf{x}_k^p \mathbf{x}_k^{p\mathrm{T}} \mid \mathbf{Z}_{1:k}\right]$ can be evaluated as:

$$
\begin{aligned}
&E\left[\mathbf{x}_k^p \mathbf{x}_k^{p\mathrm{T}} \mid \mathbf{Z}_{1:k}\right] \\
&= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}^*_{k|k-1} = \mathbf{Z}_{k|k-1} + \theta_k\right) \cdot \\
&\left(\int_{s_k=0}^{\infty} \left[\mathbf{P}^{pp}_{k|k} - \mathbf{P}^{pv}_{k|k} {\mathbf{P}^{vv}_{k|k}}^{-1} \mathbf{P}^{vp}_{k|k}\right] + \left[\hat{\mathbf{x}}^p_{k|k} + \mathbf{P}^{pv}_{k|k} {\mathbf{P}^{vv}_{k|k}}^{-1} \left(\begin{bmatrix} s_k \cos(\theta_k) \\ s_k \sin(\theta_k) \end{bmatrix} - \hat{\mathbf{x}}^v_{k|k} \right)\right] \cdot \right. \\
&\left. \left[\hat{\mathbf{x}}^p_{k|k} + \mathbf{P}^{pv}_{k|k} {\mathbf{P}^{vv}_{k|k}}^{-1} \left(\begin{bmatrix} s_k \cos(\theta_k) \\ s_k \sin(\theta_k) \end{bmatrix} - \hat{\mathbf{x}}^v_{k|k} \right)\right]^{\mathrm{T}} s_k \mathcal{N}\, ds_k \right) d\theta_k,
\end{aligned}
$$

which, after lengthy but straightforward manipulation, can be written as

$$
\begin{aligned}
&E\left[\mathbf{x}_k^p \mathbf{x}_k^{p\mathrm{T}} \mid \mathbf{Z}_{1:k}\right] \\
&= \frac{1}{c} \int_{\theta_k=-\pi}^{\pi} p\left(\mathbf{Z}^*_{k|k-1} = \mathbf{Z}_{k|k-1} + \theta_k\right) \cdot \\
&\begin{bmatrix} \ddots & & \\ & \beta_{jm} h_1\left(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}\right) + a_j a_m h_3\left(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}\right) & \\ & + \left(a_j b_m + a_m b_j\right) h_2\left(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}\right) + b_j b_m h_1\left(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}\right) & \\ & & \ddots \end{bmatrix} d\theta_k, \quad (5.3.13)
\end{aligned}
$$

where the factors a_j and b_j are given by:

$$a_j = \alpha_{j1} \cos(\theta_k) + \alpha_{j2} \sin(\theta_k) \quad (5.3.14)$$

$$b_j = \hat{\mathbf{x}}^p_{k|k,j} - \alpha_{j1} \hat{\mathbf{x}}^v_{k|k,1} - \alpha_{j2} \hat{\mathbf{x}}^v_{k|k,2} \quad (5.3.15)$$

and the terms β_{jk} are the entries of the $(N-2) \times (N-2)$ matrix $\mathbf{P}^{pp}_{k|k} - \mathbf{P}^{pv}_{k|k} {\mathbf{P}^{vv}_{k|k}}^{-1} \mathbf{P}^{vp}_{k|k}$.

We can now express the covariance matrix of the density $p(\mathbf{x}_k|\mathbf{Z}_{1:k})$, in terms of

its first two moments, as:

$$\text{Cov}\left[\mathbf{x}_k \mid \mathbf{Z}_{1:k}\right] = \begin{bmatrix} E\left[\mathbf{x}_k^v \mathbf{x}_k^{v\text{T}} \mid \mathbf{Z}_{1:k}\right] & E\left[\mathbf{x}_k^{p\text{T}} \mathbf{x}_k^v \mid \mathbf{Z}_{1:k}\right]^{\text{T}} \\ -E\left[\mathbf{x}_k^v \mid \mathbf{Z}_{1:k}\right] E\left[\mathbf{x}_k^v \mid \mathbf{Z}_{1:k}\right]^{\text{T}} & -E\left[\mathbf{x}_k^v \mid \mathbf{Z}_{1:k}\right] E\left[\mathbf{x}_k^p \mid \mathbf{Z}_{1:k}\right]^{\text{T}} \\ \\ E\left[\mathbf{x}_k^p \mathbf{x}_k^{v\text{T}} \mid \mathbf{Z}_{1:k}\right] & E\left[\mathbf{x}_k^p \mathbf{x}_k^p \mid \mathbf{Z}_{1:k}\right]^{\text{T}} \\ -E\left[\mathbf{x}_k^p \mid \mathbf{Z}_{1:k}\right] E\left[\mathbf{x}_k^v \mid \mathbf{Z}_{1:k}\right]^{\text{T}} & -E\left[\mathbf{x}_k^p \mid \mathbf{Z}_{1:k}\right] E\left[\mathbf{x}_k^p \mid \mathbf{Z}_{1:k}\right]^{\text{T}} \end{bmatrix} \tag{5.3.16}$$

Evaluation of h_1, h_2 and h_3

Computable expressions are now provided for the functions $h_1(\theta_k, \hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^{vv})$, $h_2(\theta_k, \hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^{vv})$ and $h_3(\theta_k, \hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^{vv})$. We recognize that these are in fact the 'zeroth', first and second moments of the 'speed' component of the transformed Gaussian random vector $\mathbf{x}_k^v$.

It turns out that an identical expression to the 'zeroth' order moment has already been evaluated in (3.1.66) as the density of the predicted bearing measurement in 2D. Thus, we have:

$$h_1(\theta_k, \hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^{vv}) \;=\; \frac{1}{2\pi(\det \mathbf{P}_{k|k}^{vv})^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}} \left[\frac{1}{a}\, e^{-\frac{b^2}{8a}} - \frac{\sqrt{2\pi}b}{2a^{3/2}} F_{\mathcal{N}}(-\frac{b}{2\sqrt{a}})\right], \tag{5.3.17}$$

where the coefficients are given by:

$$a \;= a_{11}\sin^2(\theta_k) + a_{22}\cos^2(\theta_k) + (a_{12}+a_{21})\cos(\theta_k)\sin(\theta_k) \tag{5.3.18}$$

$$b \;=\; -2a_{11}x_1^v \sin(\theta_k) - 2a_{22}x_2^v\cos(\theta_k) - (a_{12}+a_{21})(x_1^v\cos(\theta_k) + x_2^v\sin(\theta_k)) \tag{5.3.19}$$

$$c \;= a_{11}(x_1^v)^2 + a_{22}(x_2^v)^2 + (a_{12}+a_{21})x_1^v x_2^v, \tag{5.3.20}$$

and the terms a_{ij} denote the elements of the precision matrix $\mathbf{P}_{k|k}^{vv\,-1}$.

Likewise, the expression for $h_2(\theta_k, \hat{\mathbf{x}}_{k|k}^v, \mathbf{P}_{k|k}^{vv})$ is identical (up to a scaling factor) to

the predicted bearings measurement in 3D (3.1.76), and is given by:

$$h_2(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}) \;=\; \frac{1}{2\pi(\det\mathbf{P}^{vv}_{k|k})^{1/2}}\, e^{\frac{b^2}{8a}-\frac{c}{2}} \left[-\frac{b}{2a^2}\, e^{-\frac{b^2}{8a}} + \frac{\sqrt{2\pi}}{4a^{5/2}}(4a+b^2) F_{\mathcal{N}}(-\frac{b}{2\sqrt{a}}) \right], \tag{5.3.21}$$

with the coefficients a, b and c the same as before.

Finally, an expression for $h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k})$ can be written, similarly to the evaluation of (3.1.66), and with a, b and c as defined previously, as follows:

$$\begin{aligned} h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}) \;&=\; \frac{1}{2\pi(\det\mathbf{P}^{vv}_{k|k})^{1/2}} \int_{s_k=0}^{\infty} s_k^3\, e^{-\frac{1}{2}(s_k^2 a + s_k b + c)}\, ds_k \\ &=\; \frac{e^{\frac{b^2}{8a}-\frac{c}{2}}}{2\pi(\det\mathbf{P}^{vv}_{k|k})^{1/2}} \int_{s_k=0}^{\infty} s_k^3\, e^{-\frac{a}{2}(s_k+\frac{b}{2a})^2}\, ds_k \end{aligned}$$

which reduces, after the change of variable $u = \sqrt{a}(s_k + \frac{b}{2a})$, to:

$$\begin{aligned} h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}) \;&=\; \frac{e^{\frac{b^2}{8a}-\frac{c}{2}}}{2\pi(\det\mathbf{P}^{vv}_{k|k})^{1/2}} \int_{u=b/(2\sqrt{a})}^{\infty} \left(\frac{u^3}{a^{3/2}} - 3\frac{u^2 b}{2a^2} + 3\frac{u b^2}{4a^{5/2}} - \frac{b^3}{8a^3} \right) \frac{e^{-\frac{u^2}{2}}}{\sqrt{a}}\, du \\ &=\; \frac{e^{\frac{b^2}{8a}-\frac{c}{2}}}{2\pi(\det\mathbf{P}^{vv}_{k|k})^{1/2}} \left[\frac{1}{a^2} I_3\left(-\frac{b}{2\sqrt{a}}\right) - \frac{3b}{2a^{5/2}} I_2\left(-\frac{b}{2\sqrt{a}}\right) \right. \\ &\qquad \left. + \frac{3b^2}{4a^3} I_1\left(-\frac{b}{2\sqrt{a}}\right) - \frac{b^3}{8a^{7/2}} I_0\left(-\frac{b}{2\sqrt{a}}\right) \right]. \end{aligned}$$

Evaluating the integrals I_0, I_1, I_2 and I_3 using (3.1.43), (3.1.44) and (3.1.45), we obtain the final expression:

$$h_3(\theta_k, \hat{\mathbf{x}}^v_{k|k}, \mathbf{P}^{vv}_{k|k}) \;=\; \frac{e^{\frac{b^2}{8a}-\frac{c}{2}}}{2\pi(\det\mathbf{P}^{vv}_{k|k})^{1/2}} \left[\left(\frac{b^2}{4a^3} + \frac{2}{a^2} \right) e^{-\frac{b^2}{8a}} - \sqrt{2\pi} \left(\frac{12ab + b^3}{8a^{7/2}} \right) F_{\mathcal{N}}\left(-\frac{b}{2\sqrt{a}}\right) \right] \tag{5.3.22}$$

We have derived explicit formulae for the *exact* mean and covariance of the estimate of the state $\mathbf{x}_k$, assuming a Gaussian prior. These can be used to form a recursive moment-matching algorithm, where at each stage numerical evaluation of the integrals arising in (5.3.7), (5.3.9), (5.3.10), (5.3.11), (5.3.12) and (5.3.13) is carried out.

5.3.2 Landmark Initialization

Whereas EKF based bearings-only SLAM algorithms require special attention to target initialization, we employ for the quadrature-SRF SLAM algorithm a straightforward procedure using linearization of the bearings measurement function together with a normally distributed random initial range. The primary reason for this simple initialization scheme is the high robustness of the SRF to poor target initialization in 'conventional' bearings-only tracking scenarios.

As shown in section 4.1.1 for the case of measurements in 3D space, we take the initial landmark range prior to be $\hat{r}_0 \sim \mathcal{N}(\hat{r}_0; \bar{r}_0, \sigma_{r_0}^2)$ and the initial (relative) landmark bearings measurement to be $z_0 \sim \mathcal{N}(z_0; z_0^{true}, \sigma^2)$, where z_0^{true} is the noiseless relative bearings measurement. The overall state vector can then be augmented with the xy position estimates of the new landmark as:

$$\hat{\mathbf{x}}_{k|k}^{\text{new}} = \begin{bmatrix} \hat{\mathbf{x}}_{k|k} \\ g(\hat{\mathbf{x}}_{k|k}^{v}, \hat{x}_{k|k}, \hat{y}_{k|k}, z_0) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{x}}_{k|k} \\ \bar{r}_0 \cos(z_0 + \hat{\theta}_{k|k}) + \hat{x}_k \\ \bar{r}_0 \sin(z_0 + \hat{\theta}_{k|k}) + \hat{y}_k \end{bmatrix}^{\mathrm{T}}, \tag{5.3.23}$$

where $\hat{\theta}_{k|k} = \arctan\left(\frac{\hat{y}_{k|k}}{\hat{x}_{k|k}}\right)$. The $N \times N$ covariance matrix is extended to

$$\mathbf{P}_{k|k}^{\text{new}} = \mathbf{J} \left[\begin{array}{c|cc} \mathbf{P}_{k|k} & \mathbf{0}_{N\times 2} \\ \hline \mathbf{0}_{2\times N} & \begin{matrix} \sigma_{r_0}^2 & 0 \\ 0 & \sigma^2 \end{matrix} \end{array}\right] \mathbf{J}^{\mathrm{T}} \tag{5.3.24}$$

where

$$\mathbf{J} = \left[\begin{array}{c|cc} \mathbf{I}_{N\times N} & \mathbf{0}_{N\times 2} \\ \hline \nabla_{(x,y)} g(\hat{\mathbf{x}}_{k|k}^{v}, \hat{x}_{k|k}, \hat{y}_{k|k}, z_0) \quad \mathbf{0}_{2\times N} & \begin{matrix} \cos(z_0 + \hat{\theta}_k) & -\bar{r}_0 \sin(z_0 + \hat{\theta}_k) \\ \sin(z_0 + \hat{\theta}_k) & \bar{r}_0 \cos(z_0 + \hat{\theta}_k) \end{matrix} \end{array}\right]. \tag{5.3.25}$$

The choice of $\bar{r}_0$ and $\sigma_{r_0}^2$ is arbitrary and subject to fine tuning. Nevertheless, prior knowledge on the type of environment to be explored (i.e. indoors or outdoors) can help

in the selection of suitable statistics for the range prior.

5.4 The Rao-Blackwellized SRPF-SLAM Algorithm

Rao-Blackwellization is a general algorithmic implementation of the Rao-Blackwell theorem, which informally states that given an estimator $\hat{\theta}$ and a sufficient statistic T of θ, the estimator $\theta^* = E[\hat{\theta}|T]$ has a mean-squared error lower than or equal to that of $\hat{\theta}$. Rao-Blackwellized (or sometimes called 'marginalized') particle filters, where certain components of the state vector are estimated using a PF, while the conditional estimates of the remaining states are evaluated by analytical filters, have been widely studied in the literature ([63], [28], [39]).

This approach has also been implemented in SLAM, where the best known example is the so-called FastSLAM algorithm [73]. Within FastSLAM, all vehicle states are propagated using particles; the conditional estimates of the landmarks then become independent of one another since correlations between landmarks arise only from uncertainty in the vehicle position and/or heading. As a result, correlations between landmarks need not be calculated and each landmark estimate (for a given particle) can be computed with a low dimensional analytical filter (such as an EKF). This method therefore has the advantage of scaling linearly with the number of landmarks. However, FastSLAM is known to suffer from consistency issues related to particle impoverishment [5].

The Rao-Blackwellized SRPF-SLAM algorithm is a Monte-Carlo adaptation of the SRF to the SLAM problem. The direct estimation of the heading angle θ_k being incompatible with the measurement equation of the SRF, its estimation is instead carried out by the RBSRPF SLAM with a Rao-Blackwellized particle filter. Conditional on each particle for the state θ_k, an SRF is used for the joint estimation of the vehicle and landmark positions. A probability mass function of the posterior estimate of θ_k is then computed by assigning to each particle a weight proportional to the measurement likelihood obtained from the associated SRF.

5.4.1 Derivation of the Algorithm

Consider the state vector:

$$\mathbf{x}_k = \begin{bmatrix} \theta_k^v & \mathbf{x}_k^p \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} \theta_k^v & x_k^v & y_k^v & x^1 & y^1 & \dots & x^L & y^L \end{bmatrix}^{\mathrm{T}}, \tag{5.4.1}$$

where, $\mathbf{x}_k^p$ is used to designate the vehicle and landmark positions and θ_k^v is the vehicle heading angle. Our aim is to estimate the density

$$p(\mathbf{x}_k|\mathbf{Z}_{0:k}, \mathbf{u}_{0:k}^s, \mathbf{x}_0),$$

where, as defined before, $\mathbf{Z}_{0:k}$ is the set of observations of all landmarks up to time k, $\mathbf{u}_{0:k}^s$ is the set of measured control inputs applied to the vehicle state up to time k and $\mathbf{x}_0$ is the initial state. Marginalizing out the θ_k^v term, this density can be rewritten as:

$$\begin{aligned} p(\mathbf{x}_k \mid \mathbf{Z}_{0:k}, \mathbf{u}_{0:k}^s, \mathbf{x}_0) &= p(\mathbf{x}_k^p, \theta_k^v \mid \mathbf{Z}_{0:k}, \mathbf{u}_{0:k}^s, \mathbf{x}_0) \\ &= p(\mathbf{x}_k^p, \mid \theta_k^v, \mathbf{Z}_{0:k}, \mathbf{u}_{0:k}^s, \mathbf{x}_0)\, p(\theta_k^v \mid \mathbf{Z}_{0:k}, \mathbf{u}_{0:k}^s, \mathbf{x}_0). \end{aligned} \tag{5.4.2}$$

We notice from the factored representation of (5.4.2) that the SLAM problem can be decomposed into one of joint *vehicle and landmark position estimation (conditioned on the vehicle heading)* and *vehicle heading estimation.* Application of Bayes' rule yields:

$$\begin{aligned} & p(\mathbf{x}_k \mid \mathbf{Z}_{0:k}, \mathbf{u}_{0:k}^s, \mathbf{x}_0) \\ &= \frac{p(\mathbf{Z}_k \mid \mathbf{x}_k^p, \theta_k^v, \mathbf{Z}_{0:k-1}, \mathbf{u}_{0:k}^s, \mathbf{x}_0)\, p(\mathbf{x}_k^p \mid \theta_k^v, \mathbf{Z}_{0:k-1}, \mathbf{u}_{0:k}^s, \mathbf{x}_0)}{p(\mathbf{Z}_k \mid \theta_k^v, \mathbf{Z}_{0:k-1}, \mathbf{u}_{0:k}^s, \mathbf{x}_0)}\, p(\theta_k^v \mid \mathbf{Z}_{0:k}, \mathbf{u}_{0:k}^s, \mathbf{x}_0) \\ &= \frac{p(\mathbf{Z}_k \mid \mathbf{x}_k^p, \theta_k^v)\, p(\mathbf{x}_k^p, \mid \theta_k^v, \mathbf{Z}_{0:k-1}, \mathbf{u}_{0:k}^s, \mathbf{x}_0)}{p(\mathbf{Z}_k \mid \theta_k^v, \mathbf{Z}_{0:k-1})}\, p(\theta_k^v \mid \mathbf{Z}_{0:k}, \mathbf{u}_{0:k}^s, \mathbf{x}_0) \end{aligned} \tag{5.4.3}$$

The ratio of densities in (5.4.3) can be estimated by a bearings-only filter, with the minor addition of a constant 'measurement offset' term θ_k^v. The SRF provides a natural framework for this problem, while the final density in (5.4.3), which characterizes the vehicle heading, can be evaluated by a SIR particle filter. Given a joint gaussian prior for the vehicle and landmark positions vector with mean $\hat{\mathbf{x}}_{k-1|k-1}^p$ and covariance matrix $\mathbf{P}_{k-1|k-1}^p$, one iteration of the algorithm is summarizes in table 5.1.

Table 5.1: One iteration of the RBSRPF SLAM Algorithm

1. For $i = 1 : N$
   - Sample θ_k^{vi} from the proposal density $p(\theta_k^{vi} \mid \theta_{k-1}^{vi}, \mathbf{u}_k^s)$
   - Estimate joint 'vehicle and landmark' positions, conditioned on θ_k^{vi}: $[\hat{\mathbf{x}}_{k|k}^{p^i}, \mathbf{P}_{k|k}^{p^i}] = \text{SRF}\left(\hat{\mathbf{x}}_{k|k-1}^{p^i}, \mathbf{P}_{k|k-1}^{p^i}, \mathbf{Z}_k, \theta_k^{vi}\right)$
   - Compute importance weights (as in 2.3.38): $\tilde{w}_k^i = w_{k-1}^i p(\mathbf{Z}_k \mid \mathbf{x}_k^{p^i}, \theta_k^{vi})$
2. For $i = 1 : N$
   - Normalized weights: $w_k^i = \tilde{w}_k^i / \sum_{i=1}^N \tilde{w}_k^i$
3. Compute effective sample size $\hat{N}_{eff} = \left(\sum_{i=1}^N (w_k^i)^2\right)^{-1}$
4. If $\hat{N}_{eff} < N_{thr}$, resample (using any of the methods in 2.3.2):
   - $\{i^j\}_{j=1}^N = \text{RESAMPLE}\{w_k^i\}_{i=1}^N$
   - For $j = 1 : N$ $[\mathbf{x}_k^{p^j}, \mathbf{P}_k^{p^j}, w_k^j] = [\mathbf{x}_k^{p^{ij}}, \mathbf{P}_k^{p^{ij}}, 1/N]$

The SRF equations implemented within step 1 are as defined in (3.1.5)-(3.1.13). Estimates $\hat{\mathbf{x}}_{k|k}$ and $\mathbf{P}_{k|k}$ of the state vector are computed, prior to resampling, as the mean and covariance of the N weighted pairs $[\hat{\mathbf{x}}_{k|k}^i, \mathbf{P}_{k|k}^{p^i}]$:

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k} &= \sum_{i=1}^N w_k^i \hat{\mathbf{x}}_{k|k}^i \\
\mathbf{P}_{k|k} &= \sum_{i=1}^N w_k^i \mathbf{P}_{k|k}^i + \sum_{i=1}^N w_k^i \left(\hat{\mathbf{x}}_{k|k}^i - \hat{\mathbf{x}}_{k|k}\right)\left(\hat{\mathbf{x}}_{k|k}^i - \hat{\mathbf{x}}_{k|k}\right)^{\mathrm{T}},
\end{aligned}$$

where

$$\mathbf{P}_{k|k}^i = \text{diag}\left(0,\ \mathbf{P}_{k|k}^{p^i}\right).$$

The RBSRPF SLAM algorithm can therefore be interpreted as a weighted sum representation of vehicle and landmark position estimates (with their associated full position covariances), conditioned on different vehicle heading histories. Because the conditional estimation is carried out only with respect to vehicle heading hypotheses (rather than the full vehicle state vector), the remaining uncertainty in vehicle position requires correlation between landmarks to be accounted for. This is a fundamental departure from the Rao-Blackwellization method used in FastSLAM, which allows independent estimation of landmarks.

Landmark initialization is carried out in the same manner as with the quadrature-SRF SLAM algorithm, as described in section 5.3.2.

5.5 Performance of the quadrature-SRF SLAM and RBSRPF SLAM Algorithms

This section provides comparative performance evaluations of the quadrature-SRF SLAM and RBSRPF SLAM algorithms. The simulation scenario which is used consists of an area of 588×735m, populated with 30 uniformly spaced landmarks. A vehicle with a front-facing bearings-only sensor covers a distance of 130m within this space at a constant speed of 13m/s. Legs of constant heading and constant turn-rate are alternated to produce a 'Z' shaped trajectory. In order to test the robustness of the algorithms to the incorporation of information from distant landmarks, the sensor range is set to 500m. The angular sweep is taken to be $\pm 60^\circ$, resulting in an average of 7.87 targets detected at each time step. Finally, the bearings measurements of the landmarks are corrupted with white Gaussian noise of standard deviation 0.5°.

The vehicle motion is governed by the nonlinear continuous-time kinematics:

$$\begin{bmatrix} \dot{x}_t^v \\ \dot{y}_t^v \\ \dot{\theta}_t^v \end{bmatrix} = \begin{bmatrix} V_t^m \cos(\theta_t^v) \\ V_t^m \sin(\theta_t^v) \\ \gamma_t^m \end{bmatrix}, \tag{5.5.1}$$

where $\mathbf{u}_t^s = [V_t^m \quad \gamma_t^m]$ regroups the *measured* speed and steering control inputs. The discretized version of this model can be written as

$$\begin{bmatrix} x_k^v \\ y_k^v \\ \theta_k^v \end{bmatrix} = f(\mathbf{x}_{k-1}^v, \mathbf{u}_k^s) = \begin{bmatrix} x_{k-1}^v + T\, V_k^m \cos(\theta_k^v) \\ y_{k-1}^v + T\, V_k^m \sin(\theta_k^v) \\ \theta_{k-1}^v + T\, \gamma_k^m \end{bmatrix}, \tag{5.5.2}$$

where T is the sampling period, taken to be 1s for our simulations. The 'true' speed and steering inputs are measured with additive white Gaussian noise of standard deviations $\sigma_V = 0.5$m/s and $\sigma_\gamma = 2^\circ$, respectively.

For all algorithms, the prediction stage requires propagating the assumed joint Gaussian pdf of the vehicle states (done here by means of linearization of the nonlinear dynamics) while the landmark states are held fixed. We now outline, for the filters presented so far, the precise way in which these predictions are made, together with their various design parameters.

5.5.1 Implementation of the IDP-EKF SLAM Algorithm

The prediction of the the vehicle states at time k is carried out by propagating the posterior at time $k-1$ through the linearized system dynamics. For the motion model of (5.5.2), the predicted mean is evaluated using the nonlinear dynamics as

$$\begin{bmatrix} \hat{x}_{k|k-1}^v \\ \hat{y}_{k|k-1}^v \\ \hat{\theta}_{k|k-1}^v \end{bmatrix} = \begin{bmatrix} \hat{x}_{k-1}^v \\ \hat{y}_{k-1}^v \\ \hat{\theta}_{k-1}^v \end{bmatrix} + \begin{bmatrix} T\, V_k^m \cos(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m) \\ T\, V_k^m \sin(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m) \\ T\, \gamma_k^m \end{bmatrix}, \tag{5.5.3}$$

while the predicted covariance is computed as

$$\begin{aligned} \mathbf{P}_{k|k-1}^v &= \nabla_{\mathbf{x}^v} f(\mathbf{x}_{k|k-1}, \mathbf{u}_k^s)\, \mathbf{P}_{k-1|k-1}^v\, \nabla_{\mathbf{x}^v} f(\mathbf{x}_{k|k-1}, \mathbf{u}_k^s)^{\mathrm{T}} \\ &\quad + \nabla_{\mathbf{u}^s} f(\mathbf{x}_{k|k-1}, \mathbf{u}_k^s)\, \mathbf{Q}^s\, \nabla_{\mathbf{u}^s} f(\mathbf{x}_{k|k-1}, \mathbf{u}_k^s)^{\mathrm{T}}, \end{aligned} \tag{5.5.4}$$

where

$$\nabla_{\mathbf{x}^v} f(\mathbf{x}_{k|k-1}, \mathbf{u}_k^s) = \begin{bmatrix} 1 & 0 & -T\,V_k^m \sin(\hat{\theta}_{k-1|k-1} + \gamma_k^m) \\ 0 & 1 & T\,V_k^m \cos(\hat{\theta}_{k-1|k-1} + \gamma_k^m) \\ 0 & 0 & 1 \end{bmatrix}, \tag{5.5.5}$$

$$\nabla_{\mathbf{u}^s} f(\mathbf{x}_{k|k-1}, \mathbf{u}_k^s) = \begin{bmatrix} T\,\cos(\hat{\theta}_{k-1|k-1} + \gamma_k^m) & -T\,V_k^m \sin(\hat{\theta}_{k-1|k-1} + \gamma_k^m) \\ T\,\sin(\hat{\theta}_{k-1|k-1} + \gamma_k^m) & T\,V_k^m \cos(\hat{\theta}_{k-1|k-1} + \gamma_k^m) \\ 0 & 0 \end{bmatrix}, \tag{5.5.6}$$

and

$$\mathbf{Q}_s = \begin{bmatrix} \sigma_V^2 & 0 \\ 0 & \sigma_\gamma^2 \end{bmatrix}. \tag{5.5.7}$$

The statistics of the inverse depth ρ_0 of landmarks at initialization are design parameters. It is suggested in [16] that the assumed Gaussian initial inverse depth (with mean $\hat{\rho}_0$ and variance σ_{ρ_0}) should cover ranges from a distance $d_{\min}$ 'close' to the vehicle all the way to infinity within its 95% acceptance region. This leads to (heuristically derived) parameters $\hat{\rho}_0 = (2\,d_{\min})^{-1}$ and $\sigma_{\rho_0} = (4\,d_{\min})^{-1}$. However, for our scenario, using a 'small' value for $d_{\min}$ (i.e. $d_{\min} < 50$m) consistently resulted in landmark depths becoming negative, causing the algorithm to fail (as also noted in [59]). A non-negative depth constraint, as described in [59], was employed, along with a high value of $d_{\min}$ set at 100 m, in order to limit failures caused by negative depth estimates.

5.5.2 Implementation of the quadrature-SRF SLAM Algorithm

The vehicle heading is interpreted in the quadrature-SRF SLAM algorithm as the angle formed by the velocity vectors $\dot{x}$ and $\dot{y}$. Consequently, setting

$$\theta_k^v = \arctan\left(\frac{\dot{y}_k}{\dot{x}_k}\right),$$

a motion model equivalent to (5.5.2) is given by

$$\begin{bmatrix} \dot{x}_k \\ \dot{y}_k \\ x_k \\ y_k \end{bmatrix} = f(\dot{x}_{k-1}, \dot{y}_{k-1}, x_{k-1}, y_{k-1}, \mathbf{u}_k^s) = \begin{bmatrix} T\,V_k^m \cos\left(\theta_{k-1}^v + \gamma_k^m\right) \\ T\,V_k^m \sin\left(\theta_{k-1}^v + \gamma_k^m\right) \\ x_{k-1}^v + T\,V_k^m \cos\left(\theta_{k-1}^v + \gamma_k^m\right) \\ y_{k-1}^v + T\,V_k^m \sin\left(\theta_{k-1}^v + \gamma_k^m\right) \end{bmatrix}. \quad (5.5.8)$$

The predicted vehicle state mean is then

$$\begin{bmatrix} \hat{\dot{x}}_{k|k-1} \\ \hat{\dot{y}}_{k|k-1} \\ \hat{x}_{k|k-1} \\ \hat{y}_{k|k-1} \end{bmatrix} = \begin{bmatrix} T\,V_k^m \cos\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) \\ T\,V_k^m \sin\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) \\ \hat{x}_{k-1|k-1} + T\,V_k^m \cos\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) \\ \hat{y}_{k-1|k-1} + T\,V_k^m \sin\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) \end{bmatrix}. \quad (5.5.9)$$

Introducing the notation $\bar{\mathbf{x}}_k = [\dot{x}_k, \dot{x}_k, x_k, y_k]$ to denote the vehicle states, the predicted vehicle covariance can be expressed, in terms of linearized dynamics, as

$$\begin{aligned} \bar{\mathbf{P}}_{k|k-1} &= \nabla_{\bar{\mathbf{x}}} f(\bar{\mathbf{x}}_{k|k-1}, \mathbf{u}_k^s)\, \bar{\mathbf{P}}_{k-1|k-1}\, \nabla_{\bar{\mathbf{x}}} f(\bar{\mathbf{x}}_{k|k-1}, \mathbf{u}_k^s)^{\mathrm{T}} \\ &\quad + \nabla_{\mathbf{u}^s} f(\bar{\mathbf{x}}_{k|k-1}, \mathbf{u}_k^s)\, \mathbf{Q}^s\, \nabla_{\mathbf{u}^s} f(\bar{\mathbf{x}}_{k|k-1}, \mathbf{u}_k^s)^{\mathrm{T}}, \end{aligned} \quad (5.5.10)$$

with the Jacobians given by:

$$\begin{aligned} &\nabla_{\bar{\mathbf{x}}} f(\bar{\mathbf{x}}_{k-1}, \mathbf{u}_k^s) \\ &= \begin{bmatrix} TV_k^m\,\tilde{y}\,\sin\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) & -TV_k^m\,\tilde{x}\,\sin\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) & 0 & 0 \\ -TV_k^m\,\tilde{y}\,\cos\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) & TV_k^m\,\tilde{x}\,\cos\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) & 0 & 0 \\ 1 + TV_k^m\,\tilde{y}\,\sin\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) & -TV_k^m\,\tilde{x}\,\sin\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) & 1 & 0 \\ -TV_k^m\,\tilde{y}\,\cos\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) & 1 + TV_k^m\,\tilde{x}\,\cos\left(\hat{\theta}_{k-1|k-1}^v + \gamma_k^m\right) & 0 & 1 \end{bmatrix}, \end{aligned} \quad (5.5.11)$$

where

$$\tilde{x} = \frac{\hat{\dot{x}}_{k-1|k-1}}{\hat{\dot{x}}_{k-1|k-1}^2 + \hat{\dot{y}}_{k-1|k-1}^2} \quad \text{and} \quad \tilde{y} = \frac{\hat{\dot{y}}_{k-1|k-1}}{\hat{\dot{x}}_{k-1|k-1}^2 + \hat{\dot{y}}_{k-1|k-1}^2},$$

$$\nabla_{\mathbf{u}^s} f(\bar{\mathbf{x}}_{k|k-1}, \mathbf{u}_k^s) = \begin{bmatrix} T\cos\left(\hat{\theta}^v_{k-1|k-1} + \gamma_k^m\right) & -V_k^m T\sin\left(\hat{\theta}^v_{k-1|k-1} + \gamma_k^m\right) \\ T\sin\left(\hat{\theta}^v_{k-1|k-1} + \gamma_k^m\right) & V_k^m T\cos\left(\hat{\theta}^v_{k-1|k-1} + \gamma_k^m\right) \\ T\cos\left(\hat{\theta}^v_{k-1|k-1} + \gamma_k^m\right) & -V_k^m T\sin\left(\hat{\theta}^v_{k-1|k-1} + \gamma_k^m\right) \\ T\sin\left(\hat{\theta}^v_{k-1|k-1} + \gamma_k^m\right) & V_k^m T\cos\left(\hat{\theta}^v_{k-1|k-1} + \gamma_k^m\right) \end{bmatrix}, \quad (5.5.12)$$

and

$$\mathbf{Q}_s = \begin{bmatrix} \sigma_V^2 & 0 \\ 0 & \sigma_\gamma^2 \end{bmatrix}.$$

The numerical integrations involved in computing the moments of the estimates can be done using a variety of 'off-the-shelf' algorithms. However, because the densities arising in these calculations are unimodal and not excessively 'peaky', the accuracy of the quadrature-SRF SLAM was found to be largely insensitive to the choice of the numerical integration method employed. As a result, this step was implemented using a simple uniform trapezoidal integration scheme using 50 points (accuracy gains proved insignificant for higher numbers of function evaluations). Also, rather than evaluating the integrals over the entire range $[-\pi \quad \pi]$, the densities were (heuristically) truncated to within 6 standard deviations of the predicted heading uncertainty (beyond which limits, the densities were practically zero).

The SRF equations used within the numerical integration of the quadrature-SRF SLAM algorithm are identical to (3.1.5)-(3.1.13). The predicted covariance of the augmented measurement, meanwhile, is computed as in (3.1.58), but with the addition of a translational noise covariance term $\mathbf{Q}_k^{tr}$, representing the uncertainty in the vehicle position. Thus $\mathbf{Q}_k^{tr}$ is set equal to the predicted covariance of the vehicle position.

Finally, the mean and standard deviation of the range of initialized landmarks are design parameters, which are dependent on the maximum range of the sensor and the setting of the SLAM problem (indoor/outdoor). The values $\bar{r}_0 = 350$m and $\sigma_{r_0} = 100$m were found to provide a suitable tradeoff between accurate characterization of targets initialized both near the maximum detectable distance and at closer proximity to the vehicle.

5.5.3 Implementation of the RBSRPF SLAM Algorithm

The RBSRPF SLAM algorithm is set to propagate $N_s = 1000$ particles, with resampling carried out when the effective number of particles falls below $0.5N_s$. The prediction of the *heading conditioned* vehicle position states at time k is done, for each particle, by propagating the posterior at time $k-1$ through the linearized dynamics of the vehicle position. We therefore have:

$$\begin{bmatrix} \hat{x}^{vi}_{k|k-1} \\ \hat{y}^{vi}_{k|k-1} \end{bmatrix} = f(x^{vi}_{k|k-1}, y^{vi}_{k|k-1}, \mathbf{u}^s_k) = \begin{bmatrix} \hat{x}^{vi}_{k-1|k-1} \\ \hat{y}^{vi}_{k-1|k-1} \end{bmatrix} + \begin{bmatrix} T\, V^m_k \cos(\theta^{vi}_k) \\ T\, V^m_k \sin(\theta^{vi}_k) \end{bmatrix}, \quad (5.5.13)$$

where θ^{vi}_k is sampled from the distribution $\mathcal{N}(\theta^{vi}_k; \theta^{vi}_{k-1} + \gamma^m_k, \sigma^2_\gamma)$. The predicted vehicle position covariance, meanwhile, is evaluated as

$$\begin{aligned} \mathbf{P}^v_{k|k-1} &= \nabla_{(x^v,y^v)} f(x^{vi}_{k|k-1}, y^{vi}_{k|k-1}, \mathbf{u}^s_k)\, \mathbf{P}^v_{k-1|k-1}\, \nabla_{(x^v,y^v)} f(x^{vi}_{k|k-1}, y^{vi}_{k|k-1}, \mathbf{u}^s_k)^{\mathrm{T}} \\ &\quad + \nabla_{\mathbf{u}^s} f(x^{vi}_{k|k-1}, y^{vi}_{k|k-1}, \mathbf{u}^s_k)\, \mathbf{Q}^s\, \nabla_{\mathbf{u}^s} f(x^{vi}_{k|k-1}, y^{vi}_{k|k-1}, \mathbf{u}^s_k)^{\mathrm{T}}, \end{aligned} \quad (5.5.14)$$

where

$$\nabla_{(x^v,y^v)} f(x^{vi}_{k|k-1}, y^{vi}_{k|k-1}, \mathbf{u}^s_k) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (5.5.15)$$

$$\nabla_{\mathbf{u}^s} f(x^{vi}_{k|k-1}, y^{vi}_{k|k-1}, \mathbf{u}^s_k) = \begin{bmatrix} T\cos(\theta^{vi}_k) & -T\, V^m_k \sin(\theta^{vi}_k) \\ T\sin(\theta^{vi}_k) & T\, V^m_k \cos(\theta^{vi}_k) \end{bmatrix}, \quad (5.5.16)$$

and

$$\mathbf{Q}_s = \begin{bmatrix} \sigma^2_V & 0 \\ 0 & 0 \end{bmatrix}. \quad (5.5.17)$$

Similarly to the quadrature-SRF SLAM algorithm, the covariance of the augmented measurement noise is supplemented with the predicted uncertainty in the vehicle position $\mathbf{P}^v_{k|k-1}$. Finally, the landmark initialization statistics are taken, as in the quadrature-SRF SLAM, to be $\bar{r}_0 = 350$m and $\sigma_{r_0} = 100$m.

5.5.4 Simulation Results

Figure 5.2 illustrates the disposition of landmarks and the true vehicle path along with their estimates from a representative run of the quadrature-SRF SLAM algorithm. The vehicle trajectory estimated by odometry or 'dead-reckoning' (i.e. computed using only the noisy speed and steering measurements) is also included and is represented by the dashed line.

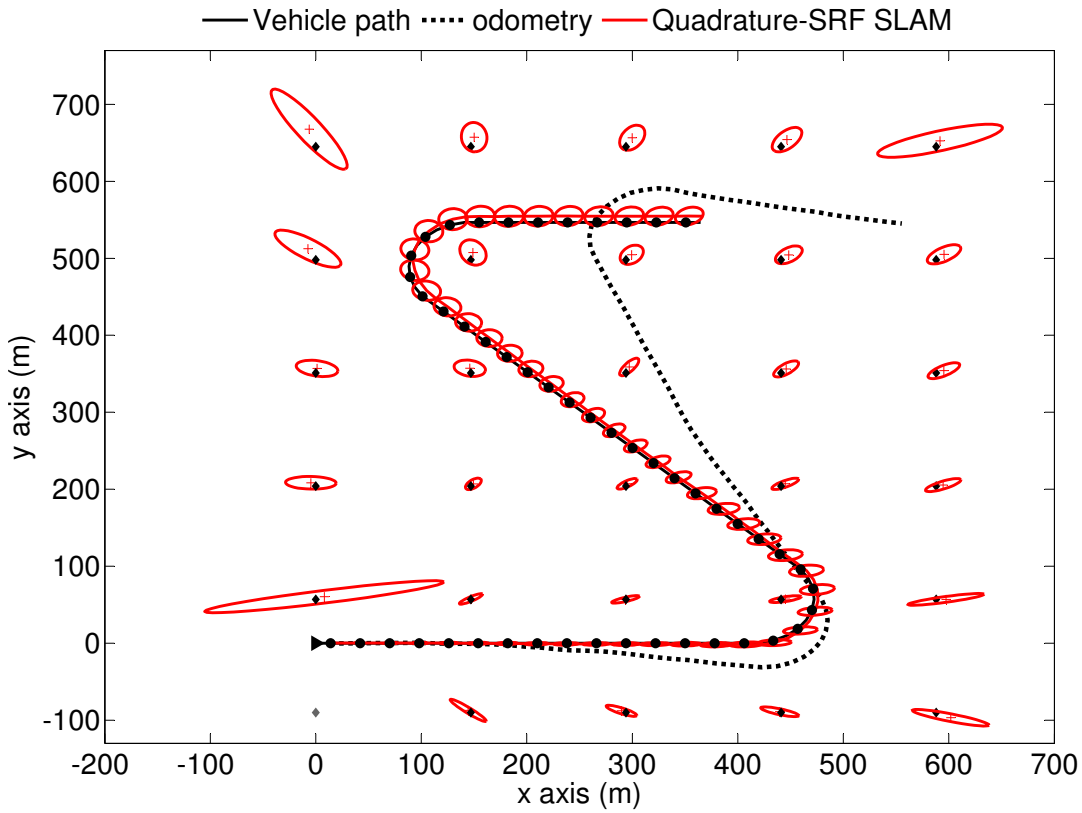


Figure 5.2: Sample run of the quadrature-SRF SLAM algorithm. The 3-sigma confidence regions of the vehicle and landmark position estimates are shown in red. Those of the vehicle position are plotted at every other sample time and corresponding true vehicle positions are shown as black dots along the path.

We simulated the above defined scenarios for each filter until $N_{\text{runs}} = 50$ MC runs were successfully completed. A run was deemed to have failed if the NEES of the vehicle position and heading states, defined as

$$\epsilon_k^2 = (\mathbf{x}_k^v - \hat{\mathbf{x}}_{k|k}^v)^{\mathrm{T}} {\mathbf{P}_{k|k}^v}^{-1} (\mathbf{x}_k^v - \hat{\mathbf{x}}_{k|k}^v), \tag{5.5.18}$$

(with $\mathbf{x}_k^v$, $\hat{\mathbf{x}}_{k|k}^v$ denoting the true and estimated vehicle states, and $\mathbf{P}_{k|k}^v$ denoting the estimated vehicle state covariance) exceeded the 99.7% probability concentration region of a χ_3^2 variable for more than 40 time steps. The 'high' tolerance of this test was set in order to reject only the most extreme outliers and avoid masking the more recurrent failure modes of the different algorithms. Both SRF based algorithms avoided catastrophic failure, while the IDP-EKF SLAM algorithm experienced 16 failures in the course of the MC simulations (i.e. a failure rate of nearly 25%).

The average RMS errors of the vehicle position and heading estimates, as calculated by the 3 algorithms, are shown in figure 5.3. The associated 3-sigma confidence intervals, computed from their estimated covariance matrices, are also included for each plot. We see that the quadrature-SRF SLAM algorithm produces estimates which are within the 99.7% concentration region of their error estimates. They are also close to being zero-mean. The position errors of the RBSRPF SLAM and IDP-EKF SLAM algorithms, meanwhile, do not display signs of sustained consistency. In particular, the IDP-EKF SLAM algorithm provides error estimates which, at certain times, fail to contain the actual RMS errors of the state estimates.

Figure 5.4 shows the average NEES, computed at each time step, for the 3 algorithms. The two-sided 99.7% region of a χ^2 distribution with degrees of freedom $n_x = 150$ (i.e. $3 \times N_{\text{runs}}$) is plotted as a dashed line. This provides a more formal measure of estimator consistency.

The quadrature-SRF SLAM algorithm appears to be mostly consistent, with its average NEES being either contained within or at close proximity of the acceptance bounds. It provides uncertainty estimates which accurately characterize the actual errors made. Consequently, the quadrature-SRF SLAM algorithm can be said to successfully solve the SLAM problem considered in this scenario.

The IDP-EKF SLAM algorithm is seen to lose consistency early on in the simulations. It appears that straight legs of the trajectory, where the same set of landmarks are consecutively observed with little parallax, are particularly problematic. Inconsistency is

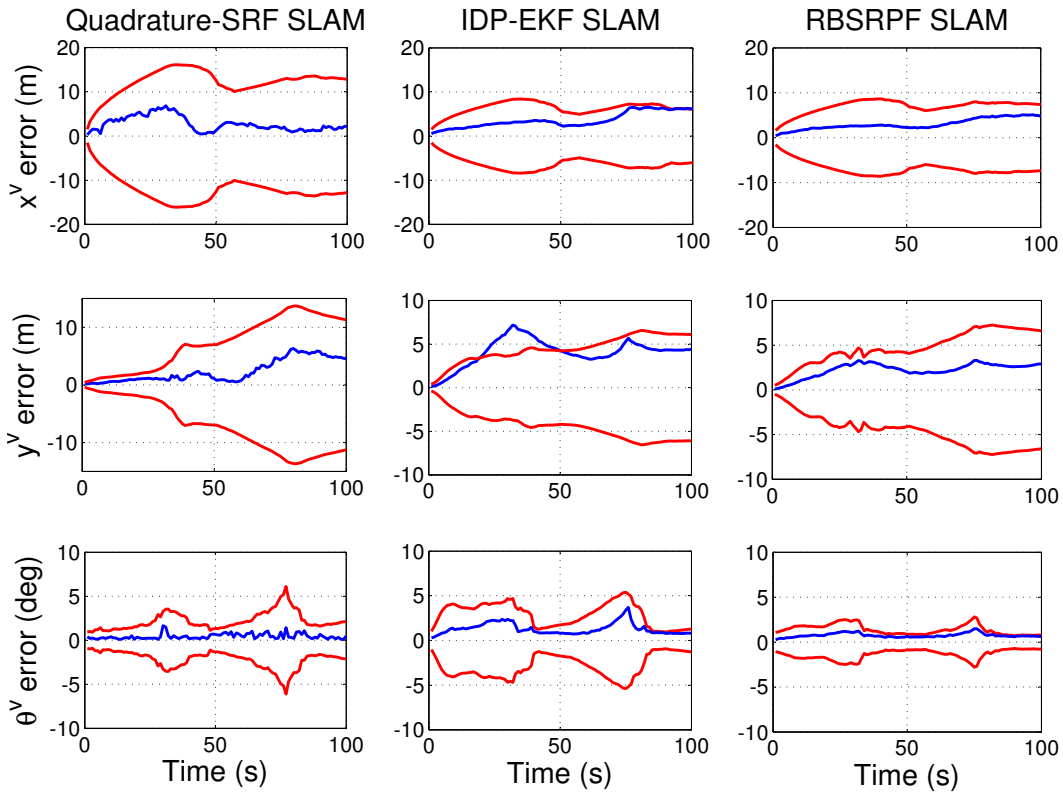


Figure 5.3: Vehicle position and heading errors (red) of the quadrature-SRF SLAM, IDP-EKF SLAM and RBSRPF SLAM algorithms. Corresponding 3-sigma confidence bound estimates are shown as dashed lines

temporarily reduced following turns (at $k = 30$ s and $k = 75$ s), as previously initialized landmarks are again seen after a period of non-detection. However, this does little to offset the eventual and steady drift of estimation errors away from 0. Also, in a self-reinforcing cycle, the initialization of landmarks with respect to erroneously confident estimates of vehicle states exacerbates inconsistency, which gradually feeds through to all state estimates.

The RBSRPF SLAM also returns, on average, inconsistent vehicle state estimates, with the average NEES steadily drifting away from the acceptance bounds. The principal cause for this is believed to be sample impoverishment due to resampling. Every time particles are resampled, diversity among the possible heading histories and their associated relative maps is reduced, leaving instead 'cloned' versions of the same estimates. The effect of this loss of historical information is an unavoidable degradation in the quality of estimates. Moreover, the NEES is seen to occasionally spike. These spikes were found to arise from a certain number of simulation runs, where a high error was made in the

heading estimate as a result of none of the proposed particles being close enough to the true heading. The resulting degeneracy (i.e. low number of effective particles) meant that the heading error estimates (computed as the sample variance of the weighted particles) were overly optimistic, causing the NEES to peak. Nevertheless, these periods of extreme inconsistency were of short duration: usually two proposal and weighting cycles were sufficient to add some measure of diversity to the particles.

Aside from an increase in the number of particles, various schemes (including the use of more sophisticated resampling methods and more accurate proposal distributions) might alleviate or delay the symptoms of inconsistency. However, by failing to address the root of the problem, which is the very use of resampling, they do not represent lasting remedies to eventual inconsistency.

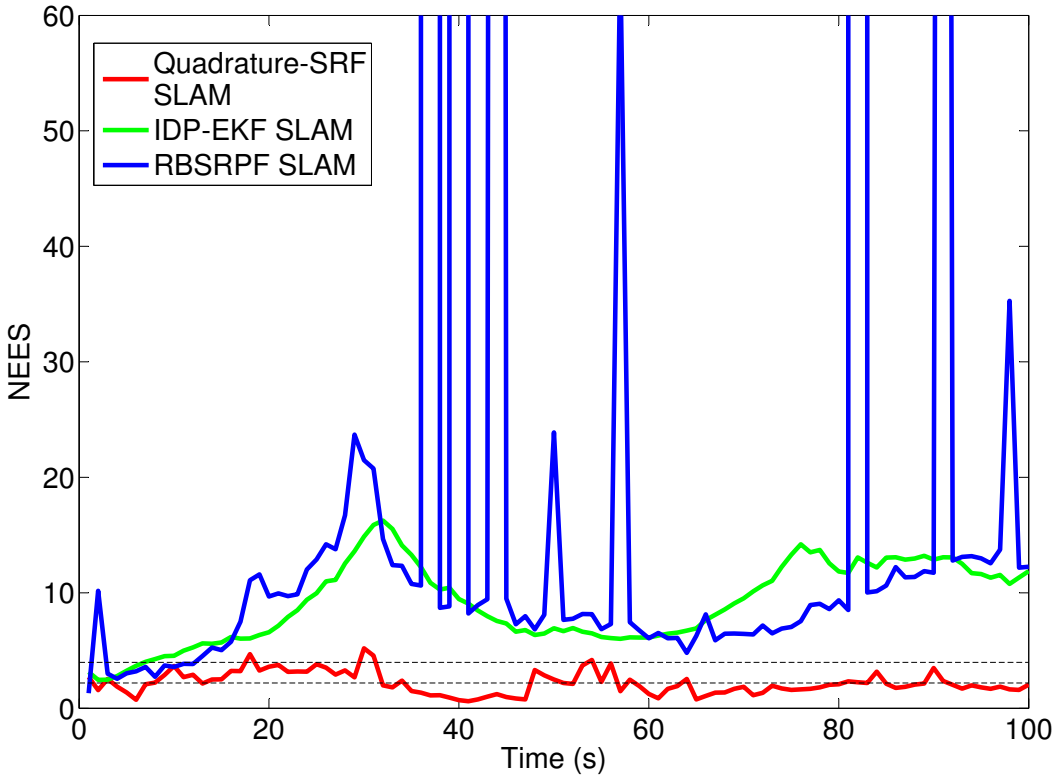


Figure 5.4: Average NEES of vehicle state estimates over 50 MC runs.

The position and heading TARMS errors, as well as the mean computation time per iteration, for the 3 algorithms are shown in table 5.2. It is worth noting, that the

RBSRPF SLAM algorithm produces very accurate point estimates of vehicle position despite its failure in computing correct error estimates. The quadrature-SRF SLAM algorithm, meanwhile achieves satisfactory position accuracy and also the highest heading accuracy. Finally, the IDP-EKF SLAM algorithm is the least accurate of all, with heading estimates on average more than twice as inaccurate as those of the quadrature-SRF SLAM algorithm. The accuracy, consistency and robustness benefits of the quadrature-SRF SLAM algorithm come at a 15 fold increase in computational overhead, relative to the IDP-EKF SLAM algorithm. On the other hand, the 220 fold increase in the computational requirements of the RBSRPF SLAM algorithm, with respect to the IDP-EKF SLAM algorithm, is not justifiable, given its generally poor performance.

Table 5.2: TARMS vehicle position and heading errors and computational cost

	Quadrature-SRF SLAM	IDP-EKF SLAM	RBSRPF SLAM
Position (m)	4.14	4.55	3.78
Heading (deg)	0.438°	1.06°	0.748°
Time/iteration (relative)	15	1	220

Chapter 6

Conclusion

We have investigated in this thesis several important and challenging aspects of dynamic estimation from bearings-only measurements. The first, and principal, part of the work addressed the problem of target tracking in 3D space. Extensions to the shifted Rayleigh filter were proposed, in order to allow implementations involving measurement origin uncertainty (in the form of clutter) and maneuvering targets. The second section of the thesis introduced the problem of bearings-only SLAM and proposed applications of the SRF for this challenging estimation problem. A summary is now provided of the main results, together with concluding remarks and suggestions for future work.

6.1 Achievements

The previously published shifted Rayleigh filter is known to outperform competing analytical filters, for numerous 2D bearings-only tracking scenarios, in terms of accuracy and robustness to ill-conditioned problems. Furthermore, its estimates match the quality of those obtained by particle filters, at a fraction of the computational cost. This is made possible using a moment-matching framework, whereby the exact second order statistics of the filtered posterior are calculated, given a Gaussian prior. This thesis contains analyses and results which extend and supplement the work carried out by J.M.C Clark, R.B. Vinter, M.M. Yaqoob and S.A. Robbiati on the SRF.

The major contribution of our work was the development of accurate and robust tracking algorithms, derived from the SRF, for challenging 3D bearings-only problems characterized by very poor quality observations. Research in the field of 3D bearings-only tracking applications is limited, with the majority of algorithms presented in 2D and generalization to 3D assumed trivial. While this is the case for most generic trackers, the formulation of the SRF requires a special modification for its implementation in

non-trivial 3D tracking scenarios. The probability density of the predicted bearings measurement, interpreted as the 'measurement likelihood', is central to the extension of the SRF to challenging 3D tracking problems. We have proposed a computable formula for this density, which allows data association and maneuver detection in 3D. Consequently, several related algorithms have been proposed, which utilize this formula:

The shifted Rayleigh filter for 3D measurements with clutter (SRF3C) The SRF3C is used for tracking nonmaneuvering targets in the presence of clutter. It was tested in a highly challenging single sensor scenario characterized by high probability of clutter, extreme bearings-rate, low observability and poor initial knowledge of the target position and velocity. The accuracy and robustness of the SRF3C were seen to be on par with those of a sophisticated particle filter using EKF generated proposal distributions, while its computational efficiency was superior to that of the RPEKF and UKF.

The Rauch-Tung-Striebel shifted Rayleigh smoother for 3D measurements with clutter (RTS-SRS3C) The RTS-SRS3C provides a simple and accurate fixed-interval smoothing algorithm for bearings-only measurements with clutter. The forward filtering pass is identical to the SRF3C, while the backward smoothing pass makes use of the convenient Rauch-Tung-Striebel algorithm, which does not require the inversion of system dynamics. The accuracy of the smoothed estimates being highly dependent on the quality of the filtered estimates, the RTS-SRS3C was found, under the previous tracking scenario, to provide results far superior to those of an RPEKF based RTS smoother.

The shifted Rayleigh particle filter (SRPF) The SRPF is a bearings-only particle filtering algorithm, for nonmaneuvering target tracking in the presence of clutter, which approximates the optimal proposal density using the SRF3C. It is structurally similar to the EKPF and UPF, in which proposal densities are generated by EKFs and UKFs, respectively. The performance of the SRPF was evaluated against that of the EKPF and UPF, in the same single-sensor tracking scenario as before. Accuracy and robustness gains were seen to extend from those observed for the SRF3C in relation to the EKF and UKF. Asymptotic convergence, in terms of time-average RMS errors, was achieved with lower computational demands than with the EKPF and UPF. Track loss was also less frequent than with the competing particle filters.

The shifted Rayleigh mixture filter for 3D measurements with clutter (SRMF3C) The SRMF3C is an adaptation of the SRF3C to the problem of multiple model bearings only tracking, in the presence of clutter. It provides an approximation to the true filtered posterior, in the form of a Gaussian mixture with a fixed number of components. A mixture reduction procedure is used at the end of each iteration to limit the complexity of the algorithm. For this purpose, the hybrid deterministic/stochastic weight-based reduction scheme put forward by Fearnhead and Clifford was implemented. The SRMF3C was put to the test in a multiple sensor 3D maneuvering target tracking scenario involving high probability of clutter, poor target state initialization, high-g out-of-plane maneuvers and high bearings-rate. While the IMM-UKF and, to a lesser extent, an IMM implementation of the SRF3C, failed to track the target accurately and reliably, the SRMF3C showed much improved performance and robustness, even for high probabilities of clutter. Moreover, it matched the accuracy and robustness of a high order multiple model implementation of the SRPF at nearly one twentieth of the computational cost. Finally, the SRMF3C was shown to be relatively resilient to model mismatch and therefore suitable in situations where exact target kinematics are unknown.

In the final technical chapter of the thesis, we addressed the problem of bearings-only SLAM problem. We outlined the major difficulties associated with this problem and proposed two new algorithms, based on the SRF, to solve it. The first one is a numerical moment-matching method, which computes estimates of the exact second order statistics of the state vector. The second one is a hybrid Monte Carlo and analytical filter, derived using Rao-Blackwellization of the state vector.

The quadrature shifter Rayleigh filter SLAM algorithm (quadrature-SRF SLAM) Because the analytical extension of the SRF to account for uncertainty in sensor orientation is very difficult - if not impossible -, we introduced an alternative numerical moment-matching algorithm, derived from the SRF, based on evaluations of the exact first and second order moments of the filtered density of interest, given a Gaussian prior. The integrals involved in these calculations require numerical evaluation, which can be done using a range of 1 dimensional numerical integration schemes. A simple trapezoidal method was found to provide adequate accuracy and performance in our simulations. The

initialization of newly detected landmarks, which is the main source of inconsistency and failure in bearings-only SLAM, was implemented by way of a simple linearization of the measurement function. This scheme, which is unsuitable in EKF based SLAM algorithms, was selected in light of the high robustness of the SRF to poor target initialization. Simulation results of a SLAM scenario with 30 landmarks and a long range bearings-only sensor have shown that the quadrature-SRF SLAM algorithm returns accurate and, more importantly, consistent estimates of the vehicle position. On the other hand, the popular 'inverse depth parameterized' EKF suffered from inconsistency and frequent failure.

The Rao-Blackwellized shifted Rayleigh particle filter SLAM algorithm (RBSRPF SLAM) The RBSRPF SLAM algorithm provides a Monte Carlo adaptation of the SRF to bearings-only SLAM. By means of the Rao-Blackwell theorem, the problem is divided into the estimation of the unknown vehicle orientation and the (conditional) estimation of the vehicle and landmark positions. These two steps are carried out, respectively, by a particle filter and the SRF. Although accurate in terms of point estimates, the RBSRPF SLAM algorithm failed to produce consistent estimates in the same scenario used to test the quadrature-SRF SLAM and IDP-EKF SLAM algorithms. The resampling step, in particular, is believed to be at the origin of the gradual deterioration of the quality of estimates.

6.2 Future Work

A number of interesting extensions of the work presented in this thesis are now suggested as possible future research problems.

The results obtained on bearings-only target tracking with clutter can be supplemented by additional simulations incorporating the general 'multiple-measurement' clutter model. The equations for the PDA-SRF and PDA-SRMF algorithms (in 3D space) having already been described, the remaining task is to assess their behaviour in simulations. Results and performance trends are expected to generalize from those obtained using the 'single-measurement' clutter model used for the SRF3C and SRMF3C.

Another extension of the work carried out on bearings-only tracking is that of

multiple maneuvering target tracking. This could be done using the SRMF3C in conjunction with existing data association algorithms such as joint probabilistic data association or by using the SRMF3C to handle both data association and multiple models.

In relation to bearings-only SLAM, the proposed quadrature-SRF SLAM algorithm was shown to provide accurate and consistent estimates under a challenging simulation scenario. The implementation of advanced numerical integration schemes is an interesting topic for future research. Although accuracy gains are expected to be marginal, computational efficiency could be improved by a reduction in the number of function evaluations. The use of submapping techniques could also be looked into, for large scale mapping applications involving hundreds or even thousands to landmarks. Finally, a substantial research topic could be the extension of the quadrature-SRF SLAM algorithm to the problem of '6 degree of freedom' SLAM in 3D. The 2 dimensional integrals arising in the moment calculations are expected to require sophisticated numerical integration methods to enable efficient implementation of the algorithm.

Bibliography

[1] V. J. Aidala, "Kalman filter behavior in bearings-only tracking applications," *IEEE Trans. Aerospace and Electronic Systems*, vol. 15, no. 1, pp. 29 –39, Jan. 1979.

[2] V. J. Aidala and S. E. Hammel, "Utilization of modified polar coordinates for bearings-only tracking," *IEEE Trans. Automat. Contr.*, vol. 28, no. 3, pp. 283 – 294, Mar. 1983.

[3] R. R. Allen and S. S. Blackman, "Implementation of an angle-only tracking filter," in *Proc. SPIE Signal and Data Processing of Small Targets*, vol. 1481, no. 1, Aug. 1991, pp. 292–303.

[4] M. Athans, R. Wishner, and A. Bertolini, "Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements," *IEEE Trans. Automat. Contr.*, vol. 13, no. 5, pp. 504 – 514, Oct. 1968.

[5] T. Bailey, J. Nieto, and E. Nebot, "Consistency of the fastslam algorithm," in *Proc. IEEE Int. Conf. Robotics and Automation*, May 2006, pp. 424 –429.

[6] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," *Automatica*, vol. 11, no. 5, pp. 451 – 460, 1975.

[7] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. Bar-Shalom, Y., 2011.

[8] L. X.-R. Bar-Shalom Y. and K. T., *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*. John Wiley & Sons, 2002.

[9] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, ser. Artech House radar library. Artech House, 1999.

[10] H. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with markovian switching coefficients," *IEEE Trans. Automat. Cont.*, vol. 33, no. 8, pp. 780 –783, Aug. 1988.

[11] B. P. Carlin, N. G. Polson, and D. S. Stoffer, "A Monte Carlo approach to nonnormal and nonlinear state-space modeling," *Journal of the American Statistical Association*, vol. 87, no. 418, pp. 493–500, 1992.

[12] J. Carpenter, P. Clifford, and P. Fearnhead, "Improved particle filter for nonlinear problems," *IEE Proc. Radar, Sonar and Navigation*, vol. 146, no. 1, pp. 2–7, Feb. 1999.

[13] J. A. Castellanos, J. Neira, and J. D. Tardos, "Limits to the consistency of ekf-based slam," Lisbon, Portugal 2004.

[14] J. A. Castellanos, J. D. Tardos, and G. Schmidt, "Building a global map of the environment of a mobile robot: the importance of correlations," in *IEEE Int. Conf. Robotics and Automation*, vol. 2, Apr. 1997, pp. 1053 –1059.

[15] J. Civera, A. J. Davison, and J. M. M. Montiel, "Unified inverse depth parametrization for monocular slam," in *Proc. Robotics: Science and Systems*, 2006.

[16] ——, "Inverse depth parametrization for monocular slam," *IEEE Trans. Robotics*, vol. 24, no. 5, pp. 932–945, 2008.

[17] J. Civera, A. Davison, and J. Montiel, "Inverse depth to depth conversion for monocular slam," in *IEEE Int. conf. Robotics and Automation*, Apr. 2007, pp. 2778 –2783.

[18] J. Clark, S. Robbiati, and R. Vinter, "The shifted rayleigh mixture filter for bearings-only tracking of maneuvering targets," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3218 –3226, Jul. 2007.

[19] J. Clark, R. Vinter, and M. Yaqoob, "The shifted rayleigh filter for bearings only tracking," in *Proc. 8th Int. Conf. Information Fusion*, vol. 1, Jul. 2005, pp. 93–100.

[20] ——, "Shifted rayleigh filter: a new algorithm for bearings-only tracking," *IEEE Trans. Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1373 –1384, Oct. 2007.

[21] W. G. Cochran, *Sampling Techniques*. John Wiley & Sons, 1977.

[22] P. Date and K. Ponomareva, "Linear and non-linear filtering in mathematical finance: a review," *IMA journal of Management and Mathematics*, vol. 22, no. 2, pp. 195–211, Jun. 2010.

[23] A. Davison, I. Reid, N. Molton, and O. Stasse, "Monoslam: real-time single camera slam," *IEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052 –1067, Jun. 2007.

[24] W. F. Denham and S. Pines, "Sequential estimation when measurement function nonlinearity is comparable to measurement error," *AIAA Journal*, vol. 4, pp. 1071–1076, 1966.

[25] M. Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (slam) problem," *IEEE Trans. Robotics and Automation*, vol. 17, no. 3, pp. 229 –241, Jun. 2001.

[26] R. Douc and O. Cappe, "Comparison of resampling schemes for particle filtering," in *Proc. 4th Int. Symp. Image and Signal Processing and Analysis*, Sep. 2005, pp. 64 – 69.

[27] A. Doucet, N. De Freitas, and N. Gordon, Eds., *Sequential Monte Carlo methods in practice*, 2001.

[28] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for bayesian filtering," *Mathematics and Statistics*, vol. 10, no. 3, pp. 197–208, 2000.

[29] H. Durrant-Whyte and T. Bailey, "Simultaneous localisation and mapping (slam): part i the essential algorithms," *IEEE Mag. Robotics and Automation*, vol. 13, no. 2, pp. 1–9, 2006.

[30] H. Durrant-Whyte, "Uncertain geometry in robotics," *IEEE Journ. Robotics and Automation*, vol. 4, no. 1, pp. 23 –31, Feb. 1988.

[31] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of geophysical research*, vol. 99, pp. 10–10, 1994.

[32] P. Fearnhead, "Sequential Monte Carlo methods in filter theory," Ph.D. dissertation, University of Oxford, 1998.

[33] C. P. Fearnhead P., "On-line inference for hidden markov models via particle filters," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 65, pp. 887–899, 2003.

[34] E. Fogel and M. Gavish, "Nth-order dynamics target observability from angle measurements," *IEEE Trans. Aerospace and Electronic Systems*, vol. 24, no. 3, pp. 305 –308, May 1988.

[35] D. Fraser and J. Potter, "The optimum linear smoother as a combination of two optimum linear filters," *IEEE Trans. Automatic Control*, vol. 14, no. 4, pp. 387–390, Aug. 1969.

[36] A. Gelb, *Applied Optimal Estimation.* MIT Press, 1974.

[37] P. Gemeiner, A. Davison, and M. Vincze, "Improving localization robustness in monocular SLAM using a high-speed camera," in *Proc. of Robotics: Science and Systems IV*, Zurich, Switzerland, Jun. 2008.

[38] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," *IEE Proc. Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[39] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P. Nordlund, "Particle filters for positioning, navigation, and tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 425–437, Feb. 2002.

[40] S. Hammel and V. Aidala, "Observability requirements for three-dimensional tracking via angle measurements," *IEEE Trans. Aerospace and Electronic Systems*, vol. 21, no. 2, pp. 200–207, Mar. 1985.

[41] J. D. Hol, T. B. Schon, and F. Gustafsson, "On resampling algorithms for particle filters," in *Proc. IEEE Nonlinear Statistical Signal Processing Workshop*, Sep. 2006, pp. 79–82.

[42] P. Houtekamer and H. Mitchell, "Data assimilation using an ensemble kalman filter technique," *Monthly Weather Review*, vol. 126, no. 3, pp. 796–811, 1998.

[43] S. R. Jammaladak and A. Sengupta, *Topics in Circular Statistics.* World Scientific Publishing Co, 2001.

[44] C. Jauffret and D. Pillon, "Observability in passive target motion analysis," *IEEE Trans. Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1290–1300, Oct. 1996.

[45] A. H. Jazwinski, *Stochastic Processes And Filtering Theory.* Academic Press, 1970.

[46] S. J. Julier and J. K. Uhlmann, "A general method for approximating nonlinear transformations of probability distributions," Robotics Research Group, Department of Engineering Science, University of Oxford, Tech. Rep., 1996.

[47] ——, "A new extension of the kalman filter to nonlinear systems," in *Proc. AeroSense: 11th Int. Symp. Aerospace/Defence Sensing, Simulation and Controls*, 1997, pp. 182–193.

[48] ——, "A counter example to the theory of simultaneous localization and map building," in *Proc. IEEE Int. Conf. Robotics and Automation*, vol. 4, 2001, pp. 4238–4243.

[49] ——, "Unscented filtering and nonlinear estimation," in *Proc. of the IEEE*, vol. 92, no. 3, Mar. 2004, pp. 401–422.

[50] S. J. Julier, J. K. Uhlmann, and H. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Trans. Automatic Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000.

[51] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. AMSE-Journal of Basic Engineering*, vol. 82(D), pp. 35–45, 1960.

[52] R. Karlsson and F. Gustafsson, "Range estimation using angle-only target tracking with particle filters," in *Proc. American Control Conference*, vol. 5, 2001, pp. 3743–3748.

[53] ——, "Recursive bayesian estimation: bearings-only applications," *IEE Proc. Radar, Sonar and Navigation*, vol. 152, no. 5, pp. 305–313, Oct. 2005.

[54] G. Kitagawa, "Monte Carlo filter and smoother for non-gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, 1996.

[55] M. Klaas, M. Briers, N. de Freitas, A. Doucet, S. Maskell, and D. Lang, "Fast particle smoothing: if i had a million particles," in *Proc. Int. Conf. Machine learning*, Jun. 2006, pp. 481–488.

[56] M. Mallick, L. Mihaylova, S. Arulampalam, and Y. Yan, "Angle-only filtering in 3d using modified spherical and log spherical coordinates," in *Proc. Int. Conf. Information Fusion*, Jul. 2011, pp. 1–8.

[57] W. Mohibullah and S. J. Julier, "Bearings-only localisation of targets from low-speed uavs," in *Proc. Int. Conf. Information Fusion*, Jul. 2010, pp. 1–8.

[58] C. Musso, N. Oudjane, and F. Le Gland, "Improving regularized particle filters," in *Sequential Monte Carlo Methods in Practice.* New York: Springer, 2001.

[59] M. P. Parsley and S. J. Julier, "Avoiding negative depth in inverse depth bearing-only slam," in *Proc. IEEE Int. Conf. Intelligent Robots and Systems*, Sep. 2008, pp. 2066–2071.

[60] N. Peach, "Bearings-only tracking using a set of range-parameterised extended kalman filters," *IEE Proc. Control Theory and Applications*, vol. 142, no. 1, pp. 73–80, Jan. 1995.

[61] K. Ponomareva and P. Date, "A new unscented kalman filter with higher order moment-matching," in *Proc. 19th Int. Symp. on Mathematical Theory of Networks and Systems*, Jul. 2010, pp. 1609–1613.

[62] H. E. Rauch, C. T. Striebel, and F. Tung, "Maximum likelihood estimates of linear dynamic systems," *Journal of the American Institute of Aeronautics and Astronautics*, vol. 3, no. 8, pp. 1445–1450, Aug. 1965.

[63] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter - Particle Filters for Tracking Applications.* Artech House, 2004.

[64] S. A. Robbiati, "Sequential gaussian mixture techniques for target tracking applications," Ph.D. dissertation, Imperial College London, 2006.

[65] X. Rong Li and V. Jilkov, "A survey of maneuvering target tracking - part iii: measurement models," vol. 4473, no. 1. SPIE, 2001, pp. 423–446.

[66] ——, "Survey of maneuvering target tracking - part i: Dynamic models," *IEEE Trans. Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, Oct. 2003.

[67] A. Runnalls, "Kullback-leibler approach to gaussian mixture reduction," *IEEE Trans. Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989 –999, Jul. 2007.

[68] D. J. Salmond, "Mixture reduction algorithms for target tracking in clutter," O. E. Drummond, Ed., vol. 1305, no. 1. SPIE, 1990, pp. 434–445.

[69] R. Smith, M. Self, and P. Cheeseman, "Autonomous robot vehicles," I. J. Cox and G. T. Wilfong, Eds. New York, NY, USA: Springer-Verlag New York, Inc., 1990, ch. Estimating uncertain spatial relationships in robotics, pp. 167–193.

[70] R. C. Smith and P. Cheeseman, "On the representation and estimation of spatial uncertainty," *The International Journal of Robotics Research*, vol. 5, no. 4, pp. 56–68, 1986.

[71] D. Stallard, "An angle-only tracking filter in modified spherical coordinates," in *AIAA Guidance, Navigation and Control Conference*, vol. 1, 1987, pp. 542–550.

[72] D. Tenne and T. Singh, "The higher order unscented filter," in *Proc. American Control Conference*, vol. 3, Jun. 2003, pp. 2441–2446.

[73] S. Thrun, M. Montemerlo, D. Koller, B. Wegbreit, J. Nieto, and E. Nebot, "Fastslam: an efficient solution to the simultaneous localization and mapping problem with unknown data association," *Journal of Machine Learning Research*, 2004.

[74] S. Thrun, S. Thayer, W. Whittaker, C. Baker, W. Burgard, D. Ferguson, D. Hahnel, D. Montemerlo, A. Morris, Z. Omohundro, C. Reverte, and W. W, "Autonomous exploration and mapping of abandoned mines," *IEEE Mag. Robotics Automation Magazine*, vol. 11, no. 4, pp. 79–91, Dec. 2004.

[75] S. Thrun, W. Burgard, and D. Fox, "A probabilistic approach to concurrent mapping and localization for mobile robots," *Autonomous Robots*, vol. 5, pp. 253–271, 1998.

[76] S. Thrun and M. Montemerlo, "The graph slam algorithm with applications to large-scale mapping of urban structures," *The International Journal of Robotics Research*, vol. 25, no. 5-6, pp. 403–429, 2006.

[77] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior cramer-rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Processing*, vol. 46, no. 5, pp. 1386–1396, May 1998.

[78] J. K. Tugnait, "Detection and estimation for abruptly changing systems," *Automatica*, vol. 18, no. 5, pp. 607–615, 1982.

[79] J. K. Tugnait and A. H. Haddad, "A detection-estimation scheme for state estimation in switching environments," *Automatica*, vol. 15, no. 4, pp. 477–481, 1979.

[80] R. van der Merwe, A. Doucet, J. F. G. de Freitas, and E. Wan, "The unscented particle filter," Cambridge University Engineering Department, Tech. Rep., 2000.

[81] R. Van der Merwe and E. Wan, "The square-root unscented kalman filter for state and parameter-estimation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 6, 2001, pp. 3461–3464.

[82] E. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Proc. IEEE Symp. Adaptive Systems for Signal Processing, Communications, and Control*, 2000, pp. 153–158.

[83] J. L. Williams and P. S. Maybeck, "Cost-function-based gaussian mixture reduction for target tracking," in *Proc. Int. Conf. Information Fusion*, vol. 2, Jul. 2003, pp. 1047–1054.