Linköping studies in science and technology. Thesis
No. 1772

# Position Estimation in Uncertain Radio Environments and Trajectory Learning

Yuxin Zhao

LINKÖPING UNIVERSITY

This is a Swedish Licentiate's Thesis.

Swedish postgraduate education leads to a Doctor's degree and/or a Licentiate's degree.
A Doctor's Degree comprises 240 ECTS credits (4 years of full-time studies).
A Licentiate's degree comprises 120 ECTS credits,
of which at least 60 ECTS credits constitute a Licentiate's thesis.

Linköping studies in science and technology. Thesis
No. 1772

**Position Estimation in Uncertain Radio Environments and Trajectory Learning**

Yuxin Zhao

*yuxin.zhao@liu.se*
*www.control.isy.liu.se*
*Department of Electrical Engineering*
*Linköping University*
*SE-581 83 Linköping*
*Sweden*

ISBN 978-91-7685-559-1 ISSN 0280-7971



Printed by LiU-Tryck, Linköping, Sweden 2017

*To my husband*

# Abstract

To infer the hidden states from the noisy observations and make predictions based on a set of input states and output observations are two challenging problems in many research areas. Examples of applications many include position estimation from various measurable radio signals in indoor environments, self-navigation for autonomous cars, modeling and predicting of the traffic flows, and flow pattern analysis for crowds of people. In this thesis, we mainly use the Bayesian inference framework for position estimation in an indoor environment, where the radio propagation is uncertain. In Bayesian inference framework, it is usually hard to get analytical solutions. In such cases, we resort to Monte Carlo methods to solve the problem numerically. In addition, we apply Bayesian non-parametric modeling for trajectory learning in sport analytics.

The main contribution of this thesis is to propose sequential Monte Carlo methods, namely particle filtering and smoothing, for a novel indoor positioning framework based on proximity reports. The experiment results have been further compared with theoretical bounds derived for this proximity based positioning system. To improve the performance, Bayesian non-parametric modeling, namely Gaussian processes, have been applied to better model the radio propagation conditions. The position estimates obtained sequentially using filtering and smoothing are further compared with a static solution, which is known as fingerprinting.

Moreover, we propose a trajectory learning framework for flow estimation in sport analytics based on Gaussian processes. To mitigate the computational deficiencies of Gaussian processes, a grid-based on-line algorithm has been adopted for real-time applications. The resulting trajectory modeling for individual athletes can be used for many purposes, such as performance prediction and analysis, health condition monitoring, etc. Furthermore, we aim at modeling the flow of groups of athletes, which could be potentially used for flow pattern recognition, strategy planing, etc.

## Populärvetenskaplig sammanfattning

Att skatta dolda tillstånd från brusiga observationer och göra uppskattningar baserade på en mängd indatatillstånd och få observationer är två utmanande problem inom många forskningsområden. Exempel på tillämpningar inkluderar positionering baserat på olika mätbara radiosignaler i inomhusmiljöer, navigering för självkörande bilar, modellering och skattning av trafikflöden, och flödesmönsteranalyser för folksamlingar. I den här avhandlingen, använder vi huvudsakligen det Bayesianska inferensramverket för positionering i en inomhusmiljö, där utbredning av radio är osäker. När man använder Bayesiansk inferens är det oftast svårt att hitta analytiska lösningar. I sådana fall använder vi Monte Carlometoder för att lösa problemen numeriskt. Dessutom tillämpar vi en variant av Bayesiansk icke-parametrisk modellering vid analys av positionerat data från herrstafetten i längdåkning vid VM i Falun 2015.

Det huvudsakliga bidraget från den här avhandlingen är att förslå sekventiella Monte Carlo-metoder, nämligen partikelfilter och partikelglättare, för ett nytt positioneringsramverk baserat på rapporter av närhet till olika kända positioner. Experimentens resultat har ytterligare jämförts med teoretiska gränser härledda för detta närhetsbaserade positioneringssystem. För att förbättra prestandan så har Bayesiansk icke-parametrisk a metoder baserat på Gaussprocesser använts för att bättre modellera radioutbredning. Slutligen jämförs positionen som är skattad sekventiellt, genom filtrering eller glättning, med positionen från en statisk lösning känd som ett fingeravtryck.

Utöver detta föreslår vi ett ramverk för skattning av trajektorier av flödesskattningar inom sportanalys baserat på Gaussprocesser. För att mildra beräkningsbristerna, har en direkt rutnätsalgoritm anpassats för realtidsanvändning. Den resulterande trajektoriemodellen för individuella atleter har många olika användningsområden, som prestandauppskattningar och analyser, hälsostatusövervakning, osv. Vidare siktar vi på att modellera flödet av grupper av atleter, som skulle kunna användas för flödesmönsterigenkänning, strategiplanering, osv.

# Acknowledgments

Being a PhD is not as simple as I imagined before I started my PhD. However, time flies so fast and now with the help from many people, I have the feeling that gradually I love my life as a PhD student. Especially I feel grateful to be given the opportunity to join Ericsson Research as well as the Automatic Control group at Linköping University. It is sometimes hard to believe that this is already halfway through my whole PhD. But I clearly understand that this cannot be achieved without the help from a lot of you. So now it is a good opportunity for me to thank everyone that helps me since the start of my PhD.

First of all, I would like to say thank you to my supervisors, Fredrik Gunnarsson, Fredrik Gustafsson, and Carsten Fritsche, who gave me great support in the past three years. Thank you for giving me a lot of inspiring ideas when I got stuck in research work. Also, thanks for the great advices considering the possible career paths to the future when I was confused. And please take my apologize if I have troubled you too much with all kinds of questions, a lot of papers to be corrected and many emails/calls when you were on vacation. Particularly, thank you all for giving me confidence and support when I was presenting in front of a lot of audience.

Secondly, I would like to give special thanks to my manager at Ericsson Research, Mehdi Amirijoo, who give my full guidance even before I started my work at Ericsson. Thanks for being so patient when I ask many boring questions. And thank you for making the work environment so friendly and comfortable.

Then, I would like to thank my colleagues and friends at Ericsson and Linköping University. Thanks to Feng Yin, who worked closely with me at the beginning of my PhD. Without your help, I would definitely take much more time to get on the correct research path. Great thanks to all the people at Linlab, you make me feel that we are really a big family. I can remember every Linlab championship, every lunch with tons of questions and all the fun we have had. Parinaz Kasebzadeh and Kamiar Radnosrati, thank you for working with me for many projects. Thanks to Clas Veibäck, who helped me a lot with the thesis template.

Also, I would like to express my thanks to people in TRAX project. Particularly, Lyudmila S Mihaylova, Allan De Freitas, and Olga Isupova, thanks for the research collaborations and great support in all aspects when I was at Sheffield for the secondment. Thank Chao Liu, for the helpful discussions with you for the research collaboration.

There are still a lot more people I want to thank. My family is always here to give me support. Thank my mom for chatting with me when I feel lonely and thanks to my husband, who encouraged me when I was frustrated. Thanks to all the people for making my PhD life colorful and meaningful. There is still a long way to go toward the final destination. But I believe that together with you all, I

will definitely make it.

*Linköping, Sweden, March, 2017*
*Yuxin Zhao*

# Contents

## II Publications

# Notation

### Operators and Symbols

| Notation | Meaning |
|---|---|
| $[\,\cdot\,]^T$ | Vector/matrix transpose |
| $[\,\cdot\,]^{-1}$ | Inverse of a non-singular square matrix |
| $\mathrm{tr}(\,\cdot\,)$ | Trace of a square matrix |
| $\|\,\cdot\,\|$ | Euclidean norm of a vector |
| $\|\cdot\|$ | Cardinality of a set |
| $\mathbb{E}(\,\cdot\,)$ | Statistical expectation |
| $X(\,\cdot\,)^T$ | Short-hand notation for $XX^T$ |
| $\ln(\,\cdot\,)$ | Natural logarithm |
| $\log_{10}(\,\cdot\,)$ | Logarithm to base 10 |
| $\otimes$ | Kronecker product |
| $\nabla_{\boldsymbol{\theta}} = \partial/\partial\boldsymbol{\theta}$ | The gradient operator |
| $\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}^T$ | The Laplace operator |
| $\mathrm{erf}(\,\cdot\,)$ | The standard Gaussian error function. |
| $\mathbf{I}_N$ | Identity matrix of size $N \times N$ |
| $\mathbf{1}$ | A vector of all 1s |
| $\mathbf{0}$ | A vector of all 0s |

### Distributions

| Abbreviation | Meaning |
|---|---|
| $\mathcal{N}(\mu, \sigma^2)$ | Gaussian distribution with mean $\mu$ and variance $\sigma^2$ |
| $Cat(\,\cdot\,)$ | Categorical distribution |

### Abbreviations

| Abbreviation | Meaning |
|---|---|
| BLUE | Best linear unbiased estimator |
| CDF | Cumulative distribution function |
| CI | Confidence interval |
| CMSE | Conditional mean squared error |
| CRB | Cramér-Rao bound |
| FFBSi | Forward filtering backward simulation |
| GP | Gaussian process |
| GPR | Gaussian process regression |
| IID | Independent and identically distributed |
| LS | Least square |
| ML | Maximum likelihood |
| MLE | Maximum likelihood estimator |
| MSE | Mean squared error |
| PDF | Probability density function |
| PF | Particle filter |
| PS | Particle smoother |
| RMSE | Root mean squared error |
| RSS | Received signal strength |
| SIR | Sequential importance resampling |
| SMC | Sequential Monte Carlo |
| SSM | State space model |
| TDOA | Time difference of arrival |
| UE | User equipment |

# 1

# Introduction

In estimation theory, two scientific problems are usually formulated. The first problem relies on the modeling of the system. In such a case, given a set of specific input states $\mathbf{x}$, the output of the system are noisy observations which can be collected or measured. The target is to train a model of the system, which relates the input states with the output observations. The model can be either used to repeat the same experiment or to predict the output observations given a new input state, where the latter case is the focus of this thesis. In the second problem, the true state is usually invisible or unmeasurable. The aim is to infer the true state (or the latent variables in this case) from the noisy observations. This usually consists of a two-steps procedure. In the first step, after the observation data has been collected, a measurement model imputes relationships between latent variables and observations needs to be selected. Then, state inference is performed to estimate the true state from the noisy observations.

To solve the first problem, the process of which is usually known as machine learning or system identification, both parametric and non-parametric modeling can be used. The model is usually explicitly given by specifying a finite number of model parameters in the parametric modeling. For example, the linear regression model in Bishop (2006), autoregressive model and state space model in Ljung (1999). Unlike the parametric modeling, in non-parametric modeling, the number of parameters grows as the number of observations increases. In addition, there is no explicit model form such that the parameters in the non-parametric modeling are only determined by the data, not the model. One example of non-parametric modeling is Gaussian process (GP) Rasmussen and Williams (2006), which will be detailed in section 2.2.

To solve the second problem, as stated previously, a proper measurement model

for the noisy observations should be selected. Then, the inference of latent variables is usually solved in a statistical framework. For the modeling step, it is sometimes easy to choose an accurate model for the observations. For instance, if the observations are the voltage $V$ measured at the two sides of a resistor $R$ and the state is the current $C$ flowing in the resistor, we have $V = R \times C$, according to Ohm's Law.

However, in most cases, it is usually hard to find an exact accurate measurement model. In addition, for most of the time, the observations at the output of the system are disturbed by noise. For instance, let's assume the observations are the received signal strength (RSS) values measured from a reference network node at a set of latent positions. Due to the additive noise, shadowing, multi-path fading, and other effects, there is no explicit expression to impute the relationship between the latent positions and the observed RSS values. In such cases, empirical models have been proposed. For instance, the Hata-Okumura model Hata (1980) and the piecewise linear model Goldsmith (2006). While all those are parametric models, non-parametric models such as Gaussian processes can be used Ferris et al. (2006). Also, the model can be assumed to be known beforehand, otherwise it needs to be determined by solving the first problem at the training phase. The latter case is within the scope of this thesis.

For the inference step, a statistical framework is always used, one of which is the Bayesian inference Box and Tiao (1973). Bayesian inference is used to update the probability for a hypothesis as more evidence or information becomes available. However, a major problem of Bayesian inference is that there are many intractable integrals. Hence, there is usually no closed form expression for these integrals such that in most cases, approximations are needed.

In the 1980s, Monte Carlo methods emerged as a good tool for solving integration problems in Bayesian inference. It is known as one of the computational methods which use statistical simulations to approximate the estimates. Monte Carlo methods are mainly used in three distinct problem classes: optimization, numerical integration, and generating draws from a probability distribution Kroese et al. (2014). Sequential Monte Carlo methods emerged more recent in 1990s, which sequentially takes the observation and updates information about the hidden states Doucet et al. (2001). Here in this thesis, we will mainly use sequential Monte Carlo methods to solve the problem of inferring the latent states from observations. Detailed background and two examples of sequential Monte Carlo methods will be presented in the next chapter.

The main contributions of this thesis are to use different techniques to solve the two problems described above in different applications. In the next following section, two examples of applications are provided to further illustrate the two problems respectively.

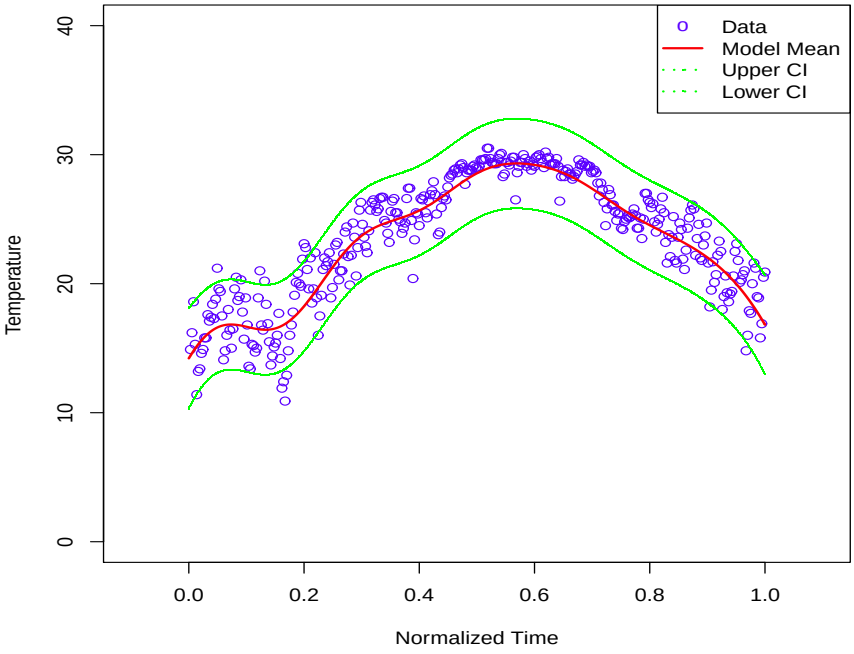


***Figure 1.1:*** *Gaussian process regression for daily temperature in a year Zhao (2016).*

## 1.1 Examples of Applications

So far, a general description of the two problems we are aiming to solve has been given. However, those problems can be encountered in various research areas and applications. To be more concrete, some illustrative examples will be provided to further assist the understanding and to better nail down the scope of this thesis.

We begin with an example of non-parametric modeling in machine learning. Given a dataset which contains one year of the daily temperatures for some place in Japan Zhao (2016), we want to train a non-parametric model which can be used as a reference to predict the trend of temperature for the next year. The dataset has been plotted in Figure 1.1.This simple example shows a typical application of solving the first modeling problem. By using a Gaussian process regression method, we can train a non-parametric model with confidence intervals (CI) as shown in Figure 1.1.

For the second problem of inferring the latent variables from the observations, there are many applications in the target tracking area. One example application

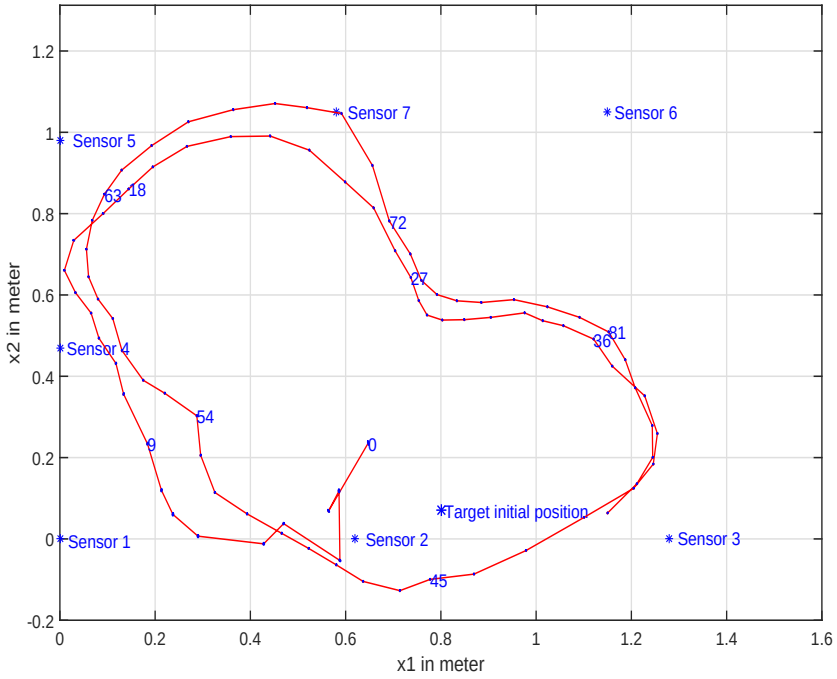


***Figure 1.2:*** *Target tracking in a sensor network Zhao (2015).*

can be to perform positioning of a moving device in a sensor network. For instance, we would like to track a device which carrying a speaker that can make quite distinctive sounds Zhao (2015). Then, a sensor network of microphones can be built up to receive the sound from the speaker. The observations in this case are the time difference of arrivals (TDOA), which can be computed by correlating the received signal at the microphones. The latent variables are the positions of the device carrying the speaker. A deployment of the sensor network with the target we want to track is plotted in Figure 1.2.

Then, the positioning results for the device have been shown in Figure 1.2 by the solid curve. In this example, the aim is to infer the latent positions of the device from the TDOAs, which corresponds to the second problem.

## 1.2 Main Contributions

So far we have seen some examples of the two problems we are aiming to solve. In general, the main contributions of this thesis can be summarized into two parts. The first part is to use sequential Monte Carlo methods together with Gaussian

process modeling for state inference in a novel indoor positioning system. The position estimates are further compared with a static solution, where fingerprinting is applied to find the most likely positions given a set of observations. Finally the last part is to apply non-parametric modeling in sports analytics for trajectory learning. To be more specific, the contributions are detailed as:

- A novel indoor positioning system utilizing proximity based reports has been proposed. Advanced modeling with Gaussian process has be applied to optimize the thresholds which trigger the proximity reports. Sets of optimized thresholds have been obtained by different criteria, which enables a new positioning fashion. (Paper A)

- Corresponding sequential Monte Carlo methods have been proposed for the novel proximity based positioning system to estimate the position of the moving device in such a network. In addition, theoretical positioning performance for proximity based system has been derived. This results in a more simple and efficient way of positioning devices with less signaling and bandwidth, while maintaining adequate positioning accuracy. (Paper B)

- An advanced modeling for proximity reports with Gaussian process has been evaluated with particle filter, one of the most popular sequential Monte Carlo methods. The results well demonstrate that the better modeling of the observations help to improve the positioning accuracy. (Paper C)

- A Gaussian process based algorithm to build receive signal strength database has been proposed with comparison to a Kriging method. Then, a statistical fingerprinting algorithm based on the pre-built RSS database has been designed. The snapshot positions obtained from the fingerprinting algorithm based on both Gaussian process and Kriging are satisfactory as compared to the case where the motion of the device is also considered. (Paper D)

- A novel application of Gaussian processes to trajectory learning in sports area. More precisely, flow models for both single trajectory and multiple trajectories haven been described with various Gaussian processes. Results of the modeling of different trajectories are proved to be accurate and such models provide valuable insights into sport performance analysis, which are usually hard to interpret from the video cameras. (Paper E)

## 1.3 Thesis outline

The thesis consists of two parts. The first part introduces the background material regarding the parametric/non-parametric modeling of system and sequential Monte Carlo methods for state inference. The second part presents the proposed solutions as a collection of peer-reviewed papers.

### 1.3.1 Outline of Part I

In the first part of this thesis, the background material of modeling is presented followed by two examples of models. Then, sequential Monte Carlo methods are introduced with highlights on the particle filter and smoother. Finally, the Cramér-Rao bound is introduced, which provides a theoretical lower limit for the estimator.

### 1.3.2 Outline of Part II

In the second part of this thesis, a collection of edited version of papers will be presented. The summary is given as below.

#### Paper A

Paper A of this thesis is an edited version of,

> F. Yin, Y. Zhao, F. Guannarsson, and F. Gustafsson. Received-signal-strength threshold optimization using Gaussian processes. *Transactions on Signal Processing*, 65(8):2164–2177, 2017.

which is an extension of the two earlier contributions

> F. Yin, Y. Zhao, and F. Gunnarsson. Fundamental bounds on position estimation using proximity reports. In *Proc. IEEE 83rd Vehicular Technology Conference: VTC2016-Spring*, 2016.

> F. Yin, Y. Zhao, and F. Gunnarsson. Proximity report triggering threshold optimization for network-based indoor positioning. In *Proc. Int. Conf. on Information Fusion*, pages 1061–1069, Washington D.C., USA, July 2015.

**Summary:** This paper presents a generic received-signal-strength (RSS) threshold optimization framework for generating informative proximity reports. The proposed framework contains five main building blocks, namely the deployment information, RSS model, positioning metric selection, optimization process and management. Among others, we focus on Gaussian process regression (GPR) based RSS models and positioning metric computation. The optimal RSS threshold is found through minimizing the best achievable localization root-mean-square-error formulated with the aid of fundamental lower bound analysis. Computational complexity is compared for different RSS models and different fundamental lower bounds. The resulting optimal RSS threshold enables enhanced performance of new fashioned low-cost and low-complex proximity report based positioning algorithms. The proposed framework is validated with real measurements collected in an office area where bluetooth-low-energy (BLE) beacons are deployed.

### Paper B

Paper B of this thesis is an edited version of,

> Y. Zhao, C. Fritsche, F. Yin, F. Gunnarsson, and F. Gustafsson. Sequential Monte Carlo methods and theoretical bounds for state inference based on proximity reports. *To be submitted to IEEE Transactions on Vehicular Technology*, 2017a.

which is an extension of the earlier contribution

> Y. Zhao, F. Yin, F. Gunnarsson, M. Amirijoo, E. Özkan, and F. Gustafsson. Partile filtering for positioning based on proximity report. In *Proc. Int. Conf. on Information Fusion*, pages 1046–1052, Washington D.C., USA, July 2015.

**Summary:** In paper A, a framework of optimizing thresholds for converting RSS to proximity reports has been developed. In this paper, we further consider positioning of devices based on a time series of proximity reports, which are generated using the optimized thresholds, from a mobile device to a network node. This corresponds to nonlinear measurements with respect to the device position in relation to the network nodes. Therefore, sequential Monte Carlo methods, namely particle filtering and smoothing, are applicable for positioning. Positioning performance is evaluated in a typical office area with Bluetooth-low-energy beacons deployed for proximity detection and report, and is further compared to parametric Cramér-Rao lower bounds. Accuracy is concluded to vary spatially over the office floor, and in relation to the beacon deployment density.

### Paper C

Paper C of this thesis is an edited version of,

> Y. Zhao, F. Yin, F. Gunnarsson, M. Amirijoo, and G. Hendeby. Gaussian processes for propagation modeling and proximity based indoor positioning. In *Proc. IEEE 83rd Vehicular Technology Conference: VTC-Spring*, 2016a.

**Summary:** In paper B, in order to perform the positioning of device using proximity reports, first we assume a linear log-distance model for the RSS observations, which gives linear relationship between the RSS and the distance of device from the network node in logarithmic scale. Then, the RSS is converted to proximity reports. However, in practice, the modeling of RSS is always considered as a latent and nonlinear observation model. To address these problems, we use one of the powerful tools, namely Gaussian process regression (GPR) for propagation modeling of RSS. This also provides some insights into the spatial correlation of the radio propagation in the considered area. Then, particle filter is combined with GPR to infer the position of the device. Radio propagation modeling and positioning performance are evaluated in a typical office area with BLE beacons deployed for proximity detection and reports. Results show that the positioning accuracy can be improved by using GPR. Accuracy is studied and compared with previous work where linear log-distance model is used.

### Paper D

Paper D of this thesis is an edited version of,

> Y. Zhao, C. Liu, L. S. Mihaylova, and F. Guannarsson. Deterministic Kriging and probabilistic Gaussian processes for RSS reconstruction in indoor localization. *Manuscipt to be submitted to Transactions on Intelligent Transportation Systems*, 2017b.

**Summary:** Paper B and Paper C focus on positioning of a moving device with certain assumed mobility patterns. In practice, there also exist use cases where static positioning is preferred without considering mobility (e.g., the device is not moving or moving without any continuous pattern). In such kind of problems, fingerprinting is usually selected to infer the position of the device. However, the problem associated with fingerprinting method is the collection and maintenance of a relatively large RSS fingerprint database/map. In this work, we propose and compare two algorithms namely, advanced Kriging method and Gaussian process, to reconstruct the RSS map with incomplete training data. To validate the effectiveness of both algorithms, experiments with BLE infrastructure have been conducted. RSS measurements are collected along predefined tracks. Both algorithms are applied to reconstruct the full RSS map within the whole area of interest. Further, statistics about the accuracy of RSS map reconstructed are compared and analyzed. Finally, with the reconstructed complete RSS map, the localization performance using probabilistic fingerprinting method will be evaluated and compared.

### Paper E

Paper E of this thesis is an edited version of,

> Y. Zhao, F. Yin, F. Gunnarsson, F. Hultkratz, and J. Fagerlind. Gaussian processes for flow modeling and prediction of positioned trajectories evaluated with sports data. In *Proc. 19th International Conference on Information Fusion (FUSION)*, pages 1461–1468, July 2016b.

**Summary:** The modeling problem described in previous sections has been occurred also in sports area, which leads to emerging research topics in recent years. In this work, we apply GPR to flow modeling and prediction of athletes in ski races, but the proposed framework can be generally applied to other use cases with device trajectories of positioned data. Some specific aspects can be addressed when the data is periodic, like in sports where the event is split up over multiple laps along a specific track. Flow models of both the individual skier and a cluster of skiers are derived and analyzed. Performance has been evaluated using data from the Falun Nordic World Ski Championships 2015, in particular the Men's cross country $4 \times 10$ km relay. The results show that the flow models vary spatially for different skiers and clusters. We further demonstrate that GPR provides powerful and accurate models for flow prediction.

## 1.4 Publications

Published work of relevance to this thesis are listed below in chronological order.

F. Yin, Y. Zhao, F. Guannarsson, and F. Gustafsson. Received-signal-strength threshold optimization using Gaussian processes. *Transactions on Signal Processing*, 65(8):2164–2177, 2017.

Y. Zhao, F. Yin, F. Gunnarsson, F. Hultkratz, and J. Fagerlind. Gaussian processes for flow modeling and prediction of positioned trajectories evaluated with sports data. In *Proc. 19th International Conference on Information Fusion (FUSION)*, pages 1461–1468, July 2016b.

Y. Zhao, F. Yin, F. Gunnarsson, M. Amirijoo, and G. Hendeby. Gaussian processes for propagation modeling and proximity based indoor positioning. In *Proc. IEEE 83rd Vehicular Technology Conference: VTC-Spring*, 2016a.

F. Yin, Y. Zhao, and F. Gunnarsson. Fundamental bounds on position estimation using proximity reports. In *Proc. IEEE 83rd Vehicular Technology Conference: VTC2016-Spring*, 2016.

F. Yin, Y. Zhao, and F. Gunnarsson. Proximity report triggering threshold optimization for network-based indoor positioning. In *Proc. Int. Conf. on Information Fusion*, pages 1061–1069, Washington D.C., USA, July 2015.

Y. Zhao, F. Yin, F. Gunnarsson, M. Amirijoo, E. Özkan, and F. Gustafsson. Partile filtering for positioning based on proximity report. In *Proc. Int. Conf. on Information Fusion*, pages 1046–1052, Washington D.C., USA, July 2015.

# Part I

# Background

# 2

# Modeling

Considering the two problems formulated in Chapter 1, there exist a plethora of scientific solutions in different research areas. In this chapter, we first present the modeling of systems using both parametric and non-parametric methods, respectively.

In this thesis, we focus on the modeling of the systems of the following general structure

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{e}), \tag{2.1}$$

where $\boldsymbol{x}$ is the state vector, $\boldsymbol{y}$ is the observation, $\boldsymbol{e}$ is the noise and the function $\boldsymbol{f}$ which represents the model. The purpose is to find the model for the function $\boldsymbol{f}$ which can be applied to infer the observations given the state variables. For parametric modeling, various model structures have been proposed in literature. Here one of the most basic structures is the linear regression model, which is given in the following section.

## 2.1 Linear Regression for Parametric Modeling

In linear regression, given a data set in the form of $\{y_i, x_{i,1}, ..., x_{i,p}\}_{i=1}^{n}$, the model structure can be given in the general form

$$y_i = w_0 + \sum_{j=1}^{p} w_j x_{i,j} + e_i. \tag{2.2}$$

In such a linear regression model, the noise factor is typically assumed to be independent and Gaussian distributed with zero mean, i.e., $e_i \sim \mathcal{N}(e_i; 0, \sigma_e^2)$, which

assumes that the noise variance is constant for all the observations. The parameters in this model need to be estimated can be defined as $\boldsymbol{\theta} = \{w_{0:p}, \sigma_e\}$. If we stack all the observations into one vector, then we have

$$\boldsymbol{y} = X\boldsymbol{w} + \boldsymbol{e}, \tag{2.3}$$

where $\boldsymbol{y} = [y_1, ...y_n]^T$ ($(\cdot)^T$ denotes transpose operation), $\boldsymbol{w} = [w_0, ..., w_p]^T$ and

$$X = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix}. \tag{2.4}$$

To train a linear regression model, the aim is to find the proper parameter $\boldsymbol{w}$ based on different criteria. One of the standard criteria is to minimize the squared error between the prediction value and the observed value. The solution based on this criteria is called the least square (LS) method. Then, the parameter can be presented as

$$\hat{\boldsymbol{w}}_{LS} \triangleq \arg\min_{\boldsymbol{w}} \|\boldsymbol{y} - X\boldsymbol{w}\|_2^2, \tag{2.5}$$

where $\|\cdot\|_2$ is the $L_2$-norm. The solution to (2.5) can be obtained in closed-form by taking the derivative of the squared error function w.r.t parameter $\boldsymbol{w}$, which gives Casella and Berger (2002)

$$\hat{\boldsymbol{w}}_{LS} = (X^T X)^{-1}(X^t \boldsymbol{y}), \tag{2.6}$$

and the noise variance is estimated by

$$\hat{\sigma}^2 = \frac{1}{n-p-1} \sum_{i=1}^{n} (y_i - X_i \hat{\boldsymbol{w}}_{LS})^2, \tag{2.7}$$

where $X_i$ denotes row $i$ of the matrix $X$. If further assume the errors are uncorrelated and independent from the states, this LS estimator is known as the best linear unbiased estimator (BLUE) of the estimated parameters. The "best" is in the sense that it gives the lowest variance of the estimate, as compared to other unbiased, linear estimators. Under the assumption that the noise is Gaussian distributed, the LS estimator coincides with the maximum likelihood estimator (MLE).

The linear regression model is one of the most simple model structures. This model will be used later in our Paper A and B. However, in some practical systems, the linear regression model is not sufficient to model the relationship between the observations and the states. Hence, more complex models are required. However, the complexity of the parametric models grow at the expense of increasing the number of parameters to be estimated. In such cases, Bayesian nonparametric modeling has been introduced to give more flexibility as the number

of observations increase Hjort et al. (2010).

In Bayesian non-parametrics, the degree of freedom increase as the number of observations increases. It is a Bayesian model whose parameter space has infinite dimension. In Bayesian statistics, the parameter is modeled as a random variable: the value of the parameter is unknown, and all forms of such uncertainty in Bayesian statistics are expressed as randomness. In another way of interpreting Bayesian non-parametrics is to consider them as infinite stochastic processes. Roughly speaking, a stochastic process is a generalization of a probability distribution to functions Rasmussen and Williams (2006).

The modeling problem formulated above can be summarized as: given a finite set of training data $\mathcal{D}$, we want to find function $f$ that make predictions for all possible input values. Then, we need to make assumptions about the characteristics of the underlying function. In Bayesian non-parametrics, a prior probability is given to every possible function (infinite numbers), where higher probabilities are given to functions that are considered to be more likely. Gaussian process as one of the Bayesian non-parametrics, provides a solution to this problem.

In the following section, we will discuss about one of Bayesian non-parametric models, Gaussian process, which will be used later in our Paper C, D and E.

## 2.2 Gaussian Process for Non-parametric Modeling

Before the introduction to Gaussian process, we need to define some Bayesian terms that will be used later through out this thesis. In Bayesian framework, the *prior* defines the initial belief in the variables $\boldsymbol{x}_i$. The *likelihood* defines the probability of observed variables $y_i$ given $\boldsymbol{x}_i$. Finally, the *posterior* distribution is the distribution of $\boldsymbol{x}_i$ given the observations $y_i$. According to Bayes' theorem, we have

$$p(\boldsymbol{x}_i|y_i) = \frac{p(\boldsymbol{x}_i|y_i)p(\boldsymbol{x}_i)}{p(y_i)} = \frac{p(\boldsymbol{x}_i|y_i)p(\boldsymbol{x}_i)}{\int p(\boldsymbol{x}_i|y_i)p(\boldsymbol{x}_i)d\boldsymbol{x}_i}, i = 1, ..., n, \tag{2.8}$$

where $p(y_i)$ is the marginal likelihood distribution. From 2.8 we can see that the posterior distribution combines all the information from the likelihood and the prior. Stacking all the observations as a vector $\boldsymbol{y}$ and unknown variables as a matrix $X$, we can update the beliefs in the unknown variables through the posterior distribution. However, we are usually primarily interested in the prediction of a new observation $y^*$ at a new unknown variable $\boldsymbol{x}^*$

$$y^*|\boldsymbol{x}^*, X, \boldsymbol{y}. \tag{2.9}$$

Gaussian process is a generalization of the Gaussian probability distribution. In a Gaussian process, every point in some continuous input space is associated with a normally distributed random variable. Moreover, every finite collection of those random variables has a multivariate normal distribution. The distribution of a

Gaussian process is the joint distribution of all those (infinitely many) random variables.

Generally used as a machine learning method, Gaussian process measures the similarity between different points to predict the observation value for new input state. Let's denote the training data set as $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$. Then, the observation can be given by a Gaussian process which is written as

$$y_i = f(\boldsymbol{x}_i) + e_i. \tag{2.10}$$

The noise term is always assumed to be Gaussian distributed with zero mean and variance $\sigma_e^2$. The function $f(\boldsymbol{x}_i)$ is a Gaussian process, which is characterized by its mean function and covariance function. The mean and covariance (kernal) function can be defined as

$$m(\boldsymbol{x}_i) = \mathbb{E}\{f(\boldsymbol{x}_i)\}, \tag{2.11a}$$

$$k\left(\boldsymbol{x}_i, \boldsymbol{x}_i'\right) = \mathbb{E}\left\{[f(\boldsymbol{x}_i) - m(\boldsymbol{x}_i)]\left[f\left(\boldsymbol{x}_i'\right) - m\left(\boldsymbol{x}_i'\right)\right]\right\}. \tag{2.11b}$$

The Gaussian process $f(\boldsymbol{x}_i)$ can be written as

$$f(\boldsymbol{x}_i) \sim \mathcal{GP}\left(m(\boldsymbol{x}_i), k\left(\boldsymbol{x}_i, \boldsymbol{x}_i'\right)\right). \tag{2.12}$$

There are many choices for the kernal function defined in 2.11b. One of the most commonly used one is called squared exponential (SE) kernal, which is formulated as

$$k\left(\boldsymbol{x}_i, \boldsymbol{x}_i'\right) = \sigma^2 \exp(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_i'\|}{2l^2}), \tag{2.13}$$

where $\sigma$ is the function variance, which determines the variation of function values from their means. $l$ is the length scale, which describes the smoothness of a function. Small value means that function values can change quickly, large values characterize functions that change slowly.

In order to show how to update the prior distribution conditioned on observations coming in, illustrations of this process are shown in Figure 2.1. In these three subfigures, the posterior mean and variance of the function have been plotted. As we increase the number of observations, the information we get is also richer, which leads to less uncertainty in the model.

Usually, we are primarily interested in incorporating the knowledge that the training data set provides about the function. Then we make predictions. It is natural to have the joint distribution of the training observations $\boldsymbol{y}$, and the prediction $y^*$ at new test input value $\boldsymbol{x}^*$. From previous statements about Gaussian process, we know that the joint distribution of a collection of Gaussian distributed variables are still following the Gaussian distribution. Hence, we have

$$\begin{bmatrix} \boldsymbol{y} \\ y^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{m}(X) \\ m(\boldsymbol{x}^*) \end{bmatrix}, \begin{bmatrix} K(X,X) + \sigma_e^2 & k(X, \boldsymbol{x}^*) \\ k(\boldsymbol{x}^*, X) & K(\boldsymbol{x}^*, \boldsymbol{x}^*) \end{bmatrix}\right), \tag{2.14}$$

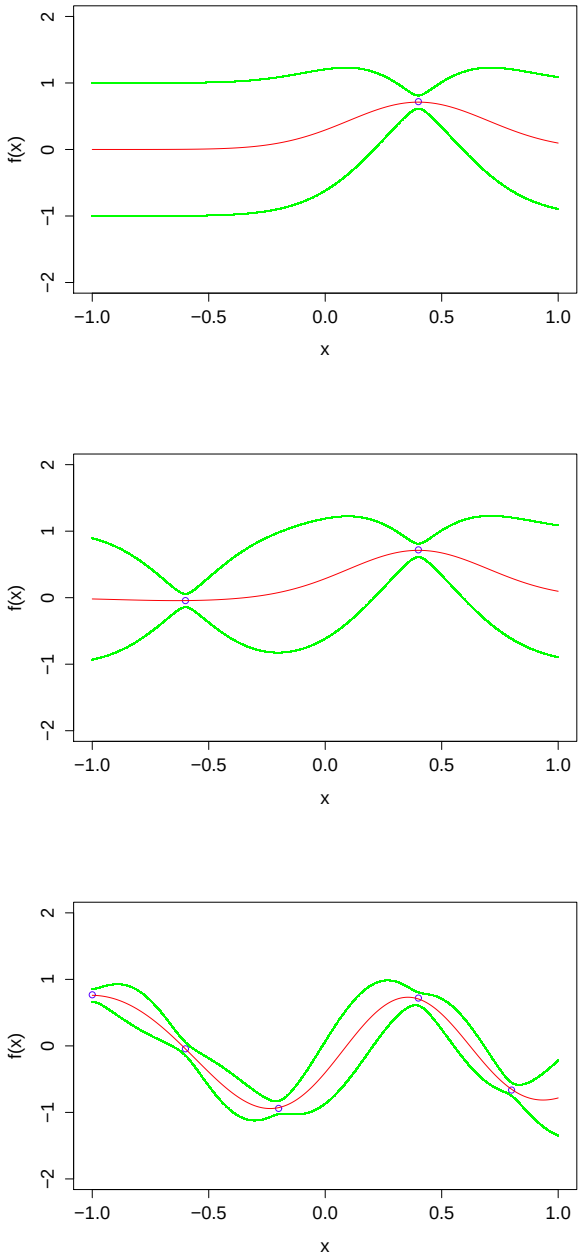


***Figure 2.1:*** *The posterior mean (red) and variance (green) of the function with one, two and five observations.*

where $K(X, X)$ denotes the covariance matrix for the training input data $X$:

$$K(X, X) = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_1, \boldsymbol{x}_n) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_n, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_n, \boldsymbol{x}_n) \end{bmatrix}. \tag{2.15}$$

$k(X, \boldsymbol{x}^*)$ is a symmetric matrix of covariances evaluated at all pairs of training and test data given by $k(X, \boldsymbol{x}^*) = k(\boldsymbol{x}^*, X) = [k(\boldsymbol{x}_1, \boldsymbol{x}^*), k(\boldsymbol{x}_2, \boldsymbol{x}^*), \ldots, k(\boldsymbol{x}_n, \boldsymbol{x}^*)]$. Similarly, $k(\boldsymbol{x}^*, \boldsymbol{x}^*)$ is the covariance of $\boldsymbol{x}^*$. Then, the prediction distribution given all the training data and the test point $\boldsymbol{x}^*$ can be obtained by marginalizing the joint Gaussian distribution on the training data, which gives

$$y^* | \boldsymbol{x}^*, X, \boldsymbol{y} \sim \mathcal{N}\left(\bar{\mu}^*, \bar{k}^*\right), \tag{2.16}$$

where

$$\bar{\mu}^* = m(\boldsymbol{x}^*) + k(\boldsymbol{x}^*, X)[K(X, X) + \sigma_e^2 I]^{-1}(\boldsymbol{y} - \boldsymbol{m}(X)), \tag{2.17a}$$

$$\bar{k}^* = k(\boldsymbol{x}^*, \boldsymbol{x}^*) - k(\boldsymbol{x}^*, X)[K(X, X) + \sigma_e^2]^{-1} k(X, \boldsymbol{x}^*) + \sigma_e^2 I. \tag{2.17b}$$

Detailed derivations can be found in Rasmussen and Williams (2006, Apendix 2). If we take a closer look at the predictive distribution given by (2.17), then the mean of the prediction is a linear combination of observations $\boldsymbol{y}$. It is also noted that the variance in (2.17) does not depend on the observations, which only depends on the input variables. This is a propertyof the Gaussian distribution.

It is useful also to introduce the term *marginal likelihood*, which is $p(\boldsymbol{y}|X)$ at this point. This can be obtained directly by observing that $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{m}(X), K(X, X) + \sigma_e^2 I)$. Then the logarithm of the marginal likelihood is given by

$$\begin{aligned} \log p(\boldsymbol{y}|X) = & \\ & -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{m}(X))^T (K(X, X) + \sigma_e^2 I)^{-1} (\boldsymbol{y} - \boldsymbol{m}(X)) \\ & -\frac{1}{2} \log |K(X, X) + \sigma_e^2 I| - \frac{n}{2} \log 2\pi. \end{aligned} \tag{2.18}$$

We see from the kernal definition that there are parameters in the kernal function which are needed to be determined. In order to distinguish between the free parameters in parametric modeling, here we call these *hyperparameters*. For instance, in the SE kernal function given previously in this section, there are two hyperparameters, namely $\sigma$ and $l$, which determine the variance and the smoothness of the function, respectively. The effects of varying the two hyperparameters can be significant. This leads to a model selection problem where in this case is to determine the proper hyperparameters for a specific kernal function. Usually, the optimal hyperparameters can be obtained by maximizing the marginal likelihood function given in (2.18). It is noted that the determined hyperparameters for the Gaussian process is closely related to the training data set.

So far we have introduced some general aspects regarding Gaussian process. From what has been stated, we can see that it is a very powerful tool of non-parametric modeling. The model is fully determined by the training data set rather than the structure of the model, which can be quite useful in the cases when we can hardly tell the model structure from the training data. However, if we examine again the prediction equations in (2.17), there the main computation complexity relies on the inverting of the matrix $K(X, X) + \sigma_e^2 I$, which size is proportional to the size of the training data. The computation complexity for inverting $n$ by $n$ matrix is then $\mathcal{O}(n^3)$. In most modeling cases, the more training data, the better the model fit results. However, in this case of GP, the computation complexity has also been increased. The studies of reducing the complexity while still keeping the efficiency of GP have been arise in recent years. One way to avoid inverting the big covariance matrix is to use a grid based on-line Gaussian process, which is derived in Huber (2014). We also present a very practical use case with grid based on-line GP in our Paper E and A.

So far, we have discussed about the modeling of observations. As stated previously, the measurement model can be either assumed to be known or can be trained using parametric/non-parametric modeling. After the model has been selected and well trained, the next step is to infer the true state from the noisy observation, which will be detailed in the next chapter.

# 3

# Bayesian Filtering and Smoothing

Before we introduce the sequential Monte Carlo methods, some basics about Bayesian inference will be provided. Bayesian inference is a framework of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more information becomes available. Bayesian filtering and smoothing are well known methods to perform Bayesian inference. In Bayesian filtering and smoothing, the posterior distribution of the hidden states is recursively updated given the noisy observations. There are many applications that make use of Bayesian filtering and smoothing, among which are target tracking, navigation, image processing and so on.

Before introduce the Bayesian filtering and smoothing, we first formulated the inference problem in a Bayesian way. Given a time series of noisy observations $\boldsymbol{y}_{1:T} \triangleq \{\boldsymbol{y}_k\}_{k=1}^{T}$, we would like to infer the hidden states variables $\boldsymbol{x}_{0:T} \triangleq \{\boldsymbol{x}_k\}_{k=0}^{T}$. If interpreted in the Bayesian way, we would like to compute the posterior joint distribution of all states given all the observations. A straightforward way to get the posterior distribution by applying Bayes' theorem is given by

$$p(\boldsymbol{x}_{0:T}|\boldsymbol{y}_{1:T}) = \frac{p(\boldsymbol{y}_{1:T}|\boldsymbol{x}_{0:T})p(\boldsymbol{x}_{0:T})}{p(\boldsymbol{y}_{1:T})}, \tag{3.1}$$

where $p(\boldsymbol{x}_{0:T})$ is the prior distribution, $p(\boldsymbol{y}_{1:T}|\boldsymbol{x}_{0:T})$ is the likelihood distribution, and $p(\boldsymbol{y}_{1:T}) = \int p(\boldsymbol{y}_{1:T}|\boldsymbol{x}_{0:T})p(\boldsymbol{x}_{0:T})d\boldsymbol{x}_{0:T}$ is usually a constant normalization factor.

Although it looks very straightforward to compute the posterior distribution, it is always hard. Since when there is a new observation coming in, the posterior distribution for all the states needs to be recomputed. This is a typical problem in the case which is considered in this thesis, where we want to get the best estimate immediately after we have obtained a new observation. As the number of

time steps increases, the computation will become intractable, since the computation complexity increases as the dimensionality of the posterior distribution gets larger. This is a very tough problem especially in real-time applications.

Then, if we relax the computation a bit, that is to say, instead of computing the full posterior distribution, we can compute the marginal distributions of the states. In order to get to this marginal distribution, we have to assume the following system models in a probabilistic way Jazwinski (1970).

$$\boldsymbol{x}_0 \sim p(\boldsymbol{x}_0), \tag{3.2a}$$

$$\boldsymbol{x}_k \sim p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}), \tag{3.2b}$$

$$\boldsymbol{y}_k \sim p(\boldsymbol{y}_k|\boldsymbol{x}_k). \tag{3.2c}$$

Here

- $p(\boldsymbol{x}_0)$ is the prior distribution of $\boldsymbol{x}_0$.
- $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ specifies the dynamic model (indicating how $\boldsymbol{x}_k$ evolves over time) of $\boldsymbol{x}_k$, which also has the assumption that $\boldsymbol{x}_k$ is a Markov process, such that the state at time $k$ only depends on the state at time $k-1$.
- $p(\boldsymbol{y}_k|\boldsymbol{x}_k)$ is the likelihood distribution which is specified by the measurement model as given in (2.1).

The model structure given above is also called the state space model Ljung (1999). With the model structure defined, we can formulate the following Bayesian filtering and smoothing problems by computing corresponding marginal distributions.

- *Bayesian filtering problem:* we are aiming at computing the marginal distributions of the current state $\boldsymbol{x}_k$ given the current and previous measurements $\boldsymbol{y}_{1:k}$.
$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}), k = 1, ..., T. \tag{3.3}$$
- *Bayesian prediction problem:* we are aiming at computing the marginal distributions of the future state $\boldsymbol{x}_{k+m}$, $m$ steps after the current time step $k$:
$$p(\boldsymbol{x}_{k+m}|\boldsymbol{y}_{1:k}), k = 1, ..., T, m = 1, 2, .... \tag{3.4}$$
- *Bayesian smoothing problem:* we are aiming at computing the marginal distributions of the state $\boldsymbol{x}_k$ given a certain interval of measurements $\boldsymbol{y}_{1:T}$ with $k < T$:
$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:T}), k = 1, ..., T. \tag{3.5}$$

In this thesis, we are focusing on the filtering and smoothing problems.

In order to solve the filtering and smoothing problems, many algorithms have been developed for state space models of various forms. Some example algorithms may include:

- *Kalman filter (KF) Kalman (1960):* it provides a closed form solution to the linear Gaussian filtering problem. This is under the assumption that both the dynamic and measurement model are linear and Gaussian distributed, such that the posterior distribution is exactly Gaussian and no numerical approximations are needed.

- *Rauch–Tung–Striebel smoother Rauch et al. (1965):* this is the corresponding closed form smoother for linear Gaussian state model.

- *Sequential Monte Carlo (SMC) filter and smoother:* namely particle filter and smoother. Since Bayesian optimal filtering and smoothing equations are generally computationally intractable, SMC is one of the numberical methods that represents the posterior distribution as a weighted set of Monte Carlo samples.

In what follows, the SMC methods applied in this thesis, will be introduced.

## 3.1 Sequential Monte Carlo Methods for State Inference

The main problem of Bayesian inference is to obtain an estimation of the state $\boldsymbol{x}$, which can be reduced to the computation of the expectations over the posterior distribution:

$$\mathbb{E}\{\boldsymbol{x}|\boldsymbol{y}_{1:T}\} = \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}_{1:T}) d\boldsymbol{x}. \tag{3.6}$$

However, for most of the time, the computation of such integral is intractable, where numerical approximations are needed. *Monte Carlo methods* are a class of methods for computing the integrals of the form (3.6). They are methods to approximate the closed form expression of statistical terms by drawing samples from the distribution and using sample average to estimate the expectation. For instance, a typical Monte Carlo approximation to (3.6) can be given as: draw $N$ independent random samples from the distribution $\boldsymbol{x}^{(i)} \sim p(\boldsymbol{x}|\boldsymbol{y}_{1:T})$, for $i = 1, ..., N$. Then the expectation is estimated as

$$\mathbb{E}\{\boldsymbol{x}|\boldsymbol{y}_{1:T}\} \approx \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}^{(i)}. \tag{3.7}$$

It seems to be promising to draw samples from the distribution $p(\boldsymbol{x}|\boldsymbol{y}_{1:T})$. However, it is usually not possible due to the complicated functional form of the distribution. In order to solve this problem, an approximated distribution is used, namely importance distribution $\pi(\boldsymbol{x}|\boldsymbol{y}_{1:T})$, from which we can easily draw samples. With importance distribution, the Monte Carlo approximation to the expec-

tation can be reformulated as

$$\begin{aligned}
\mathbb{E}\{\boldsymbol{x}|\boldsymbol{y}_{1:T}\} &= \int p(\boldsymbol{x}|\boldsymbol{y}_{1:T})\boldsymbol{x}d\boldsymbol{x} \\
&= \frac{\int p(\boldsymbol{y}_{1:T}|\boldsymbol{x})\boldsymbol{x}p(\boldsymbol{x})d\boldsymbol{x}}{\int p(\boldsymbol{y}_{1:T}|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}} \\
&= \frac{\int \left[\frac{p(\boldsymbol{y}_{1:T}|\boldsymbol{x})\boldsymbol{x}p(\boldsymbol{x})}{\pi(\boldsymbol{x}|\boldsymbol{y}_{1:T})}\right]\pi(\boldsymbol{x}|\boldsymbol{y}_{1:T})d\boldsymbol{x}}{\int \left[\frac{p(\boldsymbol{y}_{1:T}|\boldsymbol{x})p(\boldsymbol{x})}{\pi(\boldsymbol{x}|\boldsymbol{y}_{1:T})}\right]\pi(\boldsymbol{x}|\boldsymbol{y}_{1:T})d\boldsymbol{x}} \\
&\approx \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{p(\boldsymbol{y}_{1:T}|\boldsymbol{x}^{(i)})p(\boldsymbol{x}^{(i)})}{\pi(\boldsymbol{x}^{(i)}|\boldsymbol{y}_{1:T})}\boldsymbol{x}^{(i)}}{\frac{1}{N}\sum_{j=1}^{N}\frac{p(\boldsymbol{y}_{1:T}|\boldsymbol{x}^{(j)})p(\boldsymbol{x}^{(j)})}{\pi(\boldsymbol{x}^{(j)}|\boldsymbol{y}_{1:T})}} \\
&= \sum_{i=1}^{N} w^{(i)}\boldsymbol{x}^{(i)}. \qquad (3.8)
\end{aligned}$$

In sequential Monte Carlo method, observations are taken into consideration sequentially and the states are updated sequentially, where we have $\mathbb{E}\{\boldsymbol{x}_k|\boldsymbol{y}_{1:k}\} \approx \sum_{i=1}^{N} w_k^{(i)}\boldsymbol{x}_k^{(i)}$. Then, considering the posterior distribution of states $\boldsymbol{x}_{0:k}$ given the obsearvations $\boldsymbol{y}_{1:k}$, using the Markov properties of the state space model, we have:

$$\begin{aligned}
p(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k}) &\propto p(\boldsymbol{y}_k|\boldsymbol{x}_{0:k},\boldsymbol{y}_{1:k-1})p(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k-1}) \\
&= p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1},\boldsymbol{y}_{1:k-1})p(\boldsymbol{x}_{0:k-1}|\boldsymbol{y}_{1:k-1}) \\
&= p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{x}_{0:k-1}|\boldsymbol{y}_{1:k-1}). \qquad (3.9)
\end{aligned}$$

If we draw samples from the the importance distribution $\boldsymbol{x}_{0:k} \sim \pi(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k})$, then, the weights at time $k$, is given as

$$w_k^{(i)} \propto \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)})p(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})p(\boldsymbol{x}_{0:k-1}^{(i)}|\boldsymbol{y}_{1:k-1})}{\pi(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k})}. \qquad (3.10)$$

Since the importance distribution can be obtained recursively, we have

$$\pi(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k}) = \pi(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1},\boldsymbol{y}_{1:k})\pi(\boldsymbol{x}_{0:k-1}|\boldsymbol{y}_{1:k-1}). \qquad (3.11)$$

The weights can be updated as

$$w_k^{(i)} \propto \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)})p(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})}{\pi(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{0:k-1},\boldsymbol{y}_{1:k})}\frac{p(\boldsymbol{x}_{0:k-1}^{(i)}|\boldsymbol{y}_{1:k-1})}{\pi(\boldsymbol{x}_{0:k-1}^{(i)}|\boldsymbol{y}_{1:k-1})}. \qquad (3.12)$$

Now assume that we have already drawn samples $\boldsymbol{x}_{k-1}^{(i)}$ from the importance distribution $\pi(\boldsymbol{x}_{0:k-1}|\boldsymbol{y}_{1:k-1})$ and computed the corresponding weights $w_{k-1}^{(i)}$, we can draw samples $\boldsymbol{x}_{0:k}^{(i)}$ from $\pi(\boldsymbol{x}_{0:k}|\boldsymbol{y}_{1:k})$ by drawing new state samples at step $k$ as

$\boldsymbol{x}_k^{(i)} \sim \pi(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}^{(i)}, \boldsymbol{y}_{1:k})$. Then, the weights can be computed recursively as

$$w_k^{(i)} \propto \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)})p(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})}{\pi(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{0:k-1}, \boldsymbol{y}_{1:k})} w_{k-1}^{(i)}, \tag{3.13}$$

where $w_{k-1}^{(i)} \propto \frac{p(\boldsymbol{x}_{0:k-1}^{(i)}|\boldsymbol{y}_{1:k-1})}{\pi(\boldsymbol{x}_{0:k-1}^{(i)}|\boldsymbol{y}_{1:k-1})}$. SMC methods are also known as particle filtering and smoothing, which will be introduced in the following subsections.

### 3.1.1 Particle Filter

A general recursive Bayesian filtering procedure can be summarized as:

- *Initialization:* start the recursion from the prior distribution $p(\boldsymbol{x}_0)$.
- *Time update:* predict the state $\boldsymbol{x}_k$ at time $k$ given the dynamic model $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$, which is given by

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{1:k-1})d\boldsymbol{x}_{k-1}. \tag{3.14}$$

- *Measurement update:* given the measurement $\boldsymbol{y}_k$ at time step $k$, the posterior distribution of the state $\boldsymbol{x}_k$ can be computed by Bayes' theorem

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) = \frac{1}{C_k} p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}), \tag{3.15}$$

  where $C_k$ is the normalization constant, which is given by

$$C_k = \int p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1})d\boldsymbol{x}_k. \tag{3.16}$$

With the definition of importance sampling, sequential Monte Carlo approximation, and the state space model as specified in (3.2), we can have a general particle filtering (also known as sequential importance resampling (SIR) filter) procedure.

- *Initialization:* draw $N$ samples $\boldsymbol{x}_0^{(i)}$ from the prior distribution $p(\boldsymbol{x}_0)$.

$$\boldsymbol{x}_0^{(i)} \sim p(\boldsymbol{x}_0), i = 1, ..., N. \tag{3.17}$$

  Assign equal weight to each sample with $w_0^{(i)} = \frac{1}{N}$, for $i = 1, ..., N$.

- For each time step $k = 1, ..., T$, do the following:

  1. *Time update:* draw samples $\boldsymbol{x}_k^{(i)}$ from the importance distribution

$$\boldsymbol{x}_k^{(i)} \sim \pi(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{y}_{1:k}), i = 1, ..., N. \tag{3.18}$$

2. *Weights update:* compute the new weight at time $k$ as

$$w_k^{(i)} \propto \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)})p(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})}{\pi(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{0:k-1}, \boldsymbol{y}_{1:k})} w_{k-1}^{(i)}. \tag{3.19}$$

Normalize the weights to sum to 1

$$w_k^{(i)} = \frac{w_k^{(i)}}{\sum_{i=1}^{N} w_k^{(i)}}. \tag{3.20}$$

3. The estimate of state at $k$ is given as

$$\hat{\boldsymbol{x}}_k \approx \sum_{i=1}^{N} w_k^{(i)} \boldsymbol{x}_k^{(i)}. \tag{3.21}$$

4. *Resampling:* if the number of effective particles are too low, i.e., $N_{eff} = \frac{1}{\sum_{i=1}^{N}\left(w_k^{(i)}\right)^2} < N_{th}$, perform resampling as
   (a) Interpret each weight $w_k^{(i)}$ as the probability of obtaining the sample index $i$ in the set $\left\{\boldsymbol{x}_k^{(i)} : i = 1, ..., N\right\}$.
   (b) Draw $N$ samples from that discrete distribution and replace the old sample set with this new one.
   (c) Set all weights equal to $\frac{1}{N}$.

Here the resampling step duplicates the particles with large weights and discards the particles with relatively small weights. This has to be done to focus the particle filters to the more relevant part of the state space, i,e,. to states that are with high probabilities. It is also noted that resampling is not performed every step, but only when the number of effective particles becomes too small. The effective number of particles can be estimated from the variance of the particle weights Liu and Chen (1995) as given in the resampling algorithm given above. There are many ways of doing resampling. The resampling method described here is one example which is known as *multinomial resampling*, which will be used later in this thesis. For a comprehensive comparison between different resampling methods, the reader can refer to Douc (2005).

The performance of the particle filter depends on the selection of importance distribution $\pi(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{y}_{1:k})$, and The importance distribution should be in such a functional form that we can easily draw samples from it and that it is possible to evaluate the probability densities of the sample points. The optimal importance distribution in terms of variance is given as Doucet et al. (2001)

$$\pi(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{y}_{1:k}) = p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_k). \tag{3.22}$$

However, it is generally hard to sample from this distribution. Here in this thesis, the conditional prior of the state $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ is used as the importance distribution for simplicity but at the cost of losing useful information in $\boldsymbol{y}_k$. This will result in the *bootstrap filter* given in Gordon et al. (1993), where the corresponding weights update becomes

$$w_k^{(i)} \propto w_{k-1}^{(i)} p(\boldsymbol{y}_k|\boldsymbol{x}_k^{(i)}). \tag{3.23}$$

So far in this thesis we have introduced a general particle filtering algorithm which uses the observations obtained before and at the current step for computing the best possible estimate of the current state (and possibly future states). However, sometimes it is also of interest to estimate states for each time step conditional on all the measurements that we have obtained. This problem can be solved with Bayesian smoothing.

### 3.1.2 Particle Smoother

The purpose of Bayesian smoothing is to compute the marginal posterior distribution of the state $\boldsymbol{x}_k$ at time $k$, given observations up to time step $T$, where $T > k$:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:T}). \tag{3.24}$$

As can be see from (3.24), the Bayesian smoother uses also the future observations for computing the estimates. The Bayesian smoothing equations give a general procedure to compute the state estimates given future observations in a backward recursive way. The smoothed distribution of $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:T})$ for any $k < T$ can be computed as Kitagawa (1994)

$$p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k}) = \int p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) d\boldsymbol{x}_k, \tag{3.25a}$$

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:T}) = p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \int \left[ \frac{p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:T})}{p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k})} \right] d\boldsymbol{x}_{k+1}, \tag{3.25b}$$

where $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$ is the filtering distribution at time $k$, and $p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k})$ is the predicted distribution at time $k+1$.

Similarly to the particle filtering introduced in previously, the particle smoothing algorithm is aiming to solve the smoothing problems in non-linear/non-Gaussian state space models. SIR particle smoother is corresponding to the SIR filtering algorithm in the previous subsection. It is based on direct use of the SIR filtering results for smoothing. In filtering, we approximate the posterior distribution for all states. However, only the current state $\boldsymbol{x}_k^{(i)}$ at time $k$ is kept, while the history states $\boldsymbol{x}_{0:k-1}^{(i)}$ have been discarded.

An approximation to the smoothing distribution can be obtained by keeping the full histories. To get the smoothing solution, we also need to resample the state histories, not only the current states, to prevent the resampling from breaking

the state histories. The resulting algorithm is similar to the particle filtering algorithm, except that we store all the particle histories $\left\{\boldsymbol{x}_{0:k}^{(i)}, w_{0:k}^{(i)}\right\}, i = 1, ..., N$. The approximation to the smoothed posterior distribution at time step $k$, given the observations up to the time step $T > k$, can be given as

$$\hat{\boldsymbol{x}}_k|\boldsymbol{y}_{1:T} = \sum_{i=1}^{N} w_T^{(i)} \boldsymbol{x}_k^{(i)}, \tag{3.26}$$

where $\boldsymbol{x}_k^{(i)}$ is the $k$th compoenent in the particle histories $\boldsymbol{x}_{0:T}^{(i)}$. This SIR smoothing algorithm is straightforward and simple to understand. However, the estimation performance will be significantly degraded when $T \gg k$ Kitagawa (1996).

Instead of simply storing the particle histories in the SIR smoother, the forward filtering backward-simulation (FFBSi) particle smoother in Godsill et al. (2004)is based on the reuse of filtering results, and propagate individual trajectories backwards from the last time step $T$ to the start time. It can be summarized as

1. From particle filtering, store the set of weighted particles as $\left\{x_k^{(i)}, w_k^{(i)}, i = 1, ..., N, k = 1, ..., T\right\}$.

2. For time $T$, choose $\tilde{\boldsymbol{x}}_T = \boldsymbol{x}_T^{(i)}$ with probability $w_T^{(i)}$.

3. For time $k = T - 1, ..., 0$, do the following:

   (a) Update the weights according to

   $$w_{k|k+1}^{(i)} \propto w_k^{(i)} p(\tilde{\boldsymbol{x}}_{k+1}|\boldsymbol{x}_k^{(i)}). \tag{3.27}$$

   (b) Choose backward state $\tilde{\boldsymbol{x}}_k = \boldsymbol{x}_k^{(i)}$ with probability $w_{k|k+1}^{(i)}$.

Repeat the above procedure for $M$ times, then we get $\tilde{\boldsymbol{x}}_{0:T}^m$ for $m = 1, ..., M$. The smoothing estimation can be approximated as:

$$\hat{\boldsymbol{x}}_k|\boldsymbol{y}_{1:T} \approx \frac{1}{M} \sum_{m=1}^{M} \tilde{\boldsymbol{x}}_k^m, k = 0, ..., T. \tag{3.28}$$

This algorithm will be adopted to our specific state space models in detail later in Paper B.

# 4

# The Cramér-Rao Bound

Estimations of unknowns are usually not optimal, no matter in a static case or a dynamic case where filtering and smoothing are applied. In order to measure how far away the estimation performance is from the theoretical limits, Cramér-Rao bound is often applied as one benchmark to bound the covariance of an estimator.

The calculation of CRB is beneficial in many aspects. First of all, before we design any estimation algorithm, the CRB can be computed to give the best achievable performance of the estimator. Further more, we can compare different estimators by looking into the closeness to the CRB. Then, we aim at improving the estimator if it is far away from the CRB.

Generally speaking, given an estimator of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}} = [\hat{\theta_1}, \hat{\theta_2}, ..., \hat{\theta_p}]^T$ of dimension $p$ from an observed vector $\boldsymbol{y}$, the CRB provides a lower bound on the covariance of the estimator under certain regularity conditions. The following relationship can be given for the mean squared error (MSE) matrix of any estimator

$$\mathcal{M}(\hat{\boldsymbol{\theta}}(\boldsymbol{y})) \triangleq \mathbb{E}\left\{\left[\hat{\boldsymbol{\theta}}(\boldsymbol{y}) - \boldsymbol{\theta}\right]\left[\hat{\boldsymbol{\theta}}(\boldsymbol{y}) - \boldsymbol{\theta}\right]^T\right\} \geq P^{CRB}, \tag{4.1}$$

where the operator $\geq$ means that the difference $\mathcal{M}(\hat{\boldsymbol{\theta}}(\boldsymbol{y})) - P^{CRB}$ is positive semidefinite.

As mentioned previously, the estimator can either be obtained in a static way (i.e., there is no dynamic change in the estimator) or a dynamic way (i.e., with filtering and smoothing, where the dynamic model of the estimator is also considered). Correspondingly, we have CRBs for both static estimator and dynamic estimator, which will be introduced in the following sections, respectively.

## 4.1 Cramér-Rao Bound for Static Estimator

For the static case, assume the estimator $\hat{\boldsymbol{\theta}}$ and observation $\boldsymbol{y}$ is related by the probability density function $p(\boldsymbol{y}, \boldsymbol{\theta})$. The regularity for the CRB to hold is as follow Van Trees (1968):

- The function $p(\boldsymbol{y}, \boldsymbol{\theta})$ is always differentiable with respect to $\boldsymbol{\theta}$. And $\nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{y}, \boldsymbol{\theta})$ is finite.
- The term $\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \ln p(\boldsymbol{y}, \boldsymbol{\theta})$ exists and is finite.
- The operation of integration with respect to $\boldsymbol{y}$ and differentiation with respect to $\boldsymbol{\theta}$ can be interchanged in the expectation of $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$

$$\nabla_{\boldsymbol{\theta}} \left[ \int \hat{\boldsymbol{\theta}}(\boldsymbol{y}) p(\boldsymbol{y}, \boldsymbol{\theta}) d\boldsymbol{y} \right] = \int \hat{\boldsymbol{\theta}}(\boldsymbol{y}) \left[ \nabla_{\boldsymbol{\theta}} p(\boldsymbol{y}, \boldsymbol{\theta}) \right] d\boldsymbol{y}. \tag{4.2}$$

Then, the CRB is calculated as Ibargimov and Hasminskii (1981)

$$P^{sCRB} = \boldsymbol{b}(\boldsymbol{\theta})(\boldsymbol{b}(\boldsymbol{\theta}))^T + \Upsilon(\boldsymbol{\theta}) J(\boldsymbol{\theta})^{-1} \Upsilon(\boldsymbol{\theta})^T, \tag{4.3}$$

where $\boldsymbol{b}(\boldsymbol{\theta}) = \mathcal{E}\left\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\}$ is the bias, and

$$\Upsilon(\boldsymbol{\theta}) = \left( \frac{\partial (\boldsymbol{b}(\boldsymbol{\theta}) + \boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} \right), \tag{4.4a}$$

$$J(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}} \left\{ \left( \frac{\partial \ln p(\boldsymbol{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) (\cdot)^T \right\}. \tag{4.4b}$$

Here $J(\boldsymbol{\theta})$ is called the information matrix and $\Upsilon(\boldsymbol{\theta})$ is the translation matrix.

The static CRB will be used as a comparison for the static position estimator in Paper A.

## 4.2 Cramér-Rao Bound for Dynamic Estimator

With filtering and smoothing, the dynamic model of the states are also considered, which gives the link between the previous time step $k-1$ and current time $k$. This additional information will change the calculation of CRB in dynamic case.

Usually for the state space model given in 3, we can generate random models for the state trajectories. The posterior CRB provide a lower bound that is averaged over all possible state trajectories. For a specific trajectory, we have the parametric CRB, which can be used as a lower bound conditioned on this specific state trajectory.

Similar to the filtering case, where the state estimates are updated sequentially as more observations coming in, the CRB for filtering can be updated recursively. In the sequel, we will introduce both the posterior and parametric CRB for Bayesian filtering and smoothing, respectively.

The posterior CRB for filtering can be recursively computed in the following steps, which are derived in Bergman (1999); Van Trees (1968). At each time step $k$, the posterior CRB is given as the inverse of information matrix $P_{k|k} = (J_{k|k})^{-1}$, which is computed as

$$J_{k|k} = Q_{k-1} + R_k - S_{k-1}^T (J_{k-1|k-1} + V_{k-1})^{-1} S_{k-1}, \tag{4.5}$$

where

$$Q_k = \mathbb{E}_{\boldsymbol{x}_k} \left\{ -\Delta_{\boldsymbol{x}_{k+1}}^{\boldsymbol{x}_{k+1}} \ln p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) \right\}, \tag{4.6a}$$

$$R_k = \mathbb{E}_{\boldsymbol{x}_k, \boldsymbol{y}_k} \left\{ -\Delta_{\boldsymbol{x}_k}^{\boldsymbol{x}_k} \ln p(\boldsymbol{y}_k|\boldsymbol{x}_k) \right\}, \tag{4.6b}$$

$$S_k = \mathbb{E}_{\boldsymbol{x}_k} \left\{ -\Delta_{\boldsymbol{x}_k}^{\boldsymbol{x}_{k+1}} \ln p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) \right\}, \tag{4.6c}$$

$$V_k = \mathbb{E}_{\boldsymbol{x}_k} \left\{ -\Delta_{\boldsymbol{x}_k}^{\boldsymbol{x}_k} \ln p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) \right\}. \tag{4.6d}$$

The posterior CRB for smoothing can be obtained by inverting the smoothing information matrix $J_{k|T}$. The smoothing information matrix is computed by a forward-backward recursion, which is given by first compute the forward recursion as

$$J_{k+1} = Q_k - S_k^T \left( J_k + R_k + V_k \right)^{-1} S_k, k = 0, 1, ..., T. \tag{4.7}$$

Then, choose the smoothing information matrix at time $T$ as $J_{T|T} = J_T$, and propagate backward as

$$J_{k|T}^{-1} = J_{k|k}^{-1} + J_{k|k}^{-1} S_k J_{k+1|T}^{-1} S_k^T J_{k|k}^{-1}, k = T-1, ..., 1, 0, \tag{4.8}$$

where $J_{k|k}$ is the filtering information matrix at time $k$.

So far we have introduced the posterior CRB. However, in certain scenarios, to compute the bound it is required to form Monte Carlo averages by simulating state trajectories from the state-space model (in the non-linear case as shown in Paper B). In such cases, the generated state trajectories may not be practical. Or sometimes it is impossible to generate state trajectories with certain constraints. Thus, we can resort to the parametric CRB, which is conditioned on a deterministic state trajectory $\left\{ \boldsymbol{x}_k^* \right\}_{k=0:T}$.

Correspondingly, the parametric CRB for filtering is derived in Bergman (1999); Fritsche et al. (2016), which can be summarized as: for time $k+1$, the parametric CRB is computed from the previous time step $k$

$$P_{k+1|k+1} = D_k - D_k H_k^T \left( H_k D_k H_k^T + R_k \right)^{-1} H_k D_k, \tag{4.9}$$

where $D_k = F_k P_{k|k} F_k^T$, and

$$F_k = \nabla_{\boldsymbol{x}_k} p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)|_{\boldsymbol{x}_k^*}, \tag{4.10a}$$

$$H_k^T R_k^{-1} H_k = \mathbb{E}_{\boldsymbol{x}_k, \boldsymbol{y}_k} \left\{ -\Delta_{\boldsymbol{x}_k}^{\boldsymbol{x}_k} \ln p(\boldsymbol{y}_k|\boldsymbol{x}_k) \right\}|_{\boldsymbol{x}_k^*}. \tag{4.10b}$$

With additive Gaussian distributed measurement noise $\boldsymbol{e}_k \sim \mathcal{N}(0, \Sigma_{\boldsymbol{e}_k})$ as given in (2.1), and the measurement model can be written as

$$\boldsymbol{y}_k = f_k(\boldsymbol{x}_k) + \boldsymbol{e}_k. \tag{4.11}$$

The matrix $H_k$ and $R_k$ can be selected as

$$H_k = \left[ \nabla_{\boldsymbol{x}_k} f_k^T(\boldsymbol{x}_k) \right]^T, \tag{4.12a}$$

$$R_k = \Sigma_{\boldsymbol{e}_k}. \tag{4.12b}$$

The parametric CRB for smoothing is derived in details in Fritsche and Orguner (2016).

It should be noted that for linear Gaussian state space models, the derivation of CRBs is straightforward and simple. However, In the case investigated in this thesis, the system model is highly non-linear and Monte Carlo approximations are needed to compute the parametric CRB for both filtering and smoothing, which will be detailed in Paper B.

# 5

# Concluding Remarks

So far we have described two main problems to be solved in general systems. The first problem is to model the observations as a function of input variables, where both parametric and non-parametric modeling can be used. To be more specific, we give two examples of modeling methods, namely the linear regression method for parametric modeling and Gaussian process for non-parametric modeling.

Another problem focused in this thesis is the state inference problem given a set of collected observations. Typically to solve such kind of problems, Bayesian inference framework will be used. However, numerical approximations to the integrals in Bayesian inference are usually applied in cases where there are no analytical solutions to the Bayesian equations. One of such numerical solutions is known as sequential Monte Carlo methods, concretely particle filtering and smoothing.

The rest of the thesis will make use of all the tools in various applications to solve the two main problems described in previous chapters.

## 5.1 Summary of Contribution

The contribution of this thesis is to adopt the sequential Monte Carlo methods in a novel positioning framework. Gaussian process is combined with sequential Monte Carlo methods as well as static fingerprinting algorithms to improve position accuracy with different observations. In addition, we apply Gaussian process for flow modeling and prediction in trajectory learning with sport data, which may lead to new fashions in sport data analytics.

Paper A presents a generic received-signal-strength (RSS) threshold optimization

framework for generating informative proximity reports. The proposed framework contains five main building blocks, namely the deployment information, RSS model, positioning metric selection, optimization process and management. Among others, we focus on Gaussian process regression (GPR) based RSS models and positioning metric computation. The optimal RSS threshold is found through minimizing the best achievable localization root-mean-square-error formulated with the aid of fundamental lower bound analysis. Computational complexity is compared for different RSS models and different fundamental lower bounds. The resulting optimal RSS threshold enables enhanced performance of new fashioned low-cost and low-complex proximity report based positioning algorithms. The proposed framework is validated with real measurements collected in an office area where bluetooth-low-energy (BLE) beacons are deployed.

Based on Paper A, in Paper B, we further consider positioning of devices based on a time series of proximity reports, which are generated using the optimized thresholds, from a mobile device to a network node. This corresponds to nonlinear measurements with respect to the device position in relation to the network nodes. Therefore, sequential Monte Carlo methods, namely particle filtering and smoothing, are applicable for positioning. Positioning performance is evaluated in a typical office area with Bluetooth-low-energy beacons deployed for proximity detection and report, and is further compared to parametric Cramér-Rao lower bounds. Accuracy is concluded to vary spatially over the office floor, and in relation to the beacon deployment density.

Paper C proposes an improved model for RSS measurements by applying GP. Then, the RSS measurements are further converted to proximity for positioning purpose. The use of GP provides some insights into the spatial correlation of the radio propagation in the considered area. Then, particle filter is combined with GP to infer the position of the device. Radio propagation modeling and positioning performance are evaluated in a typical office area with BLE beacons deployed for proximity detection and reports. Results show that the positioning accuracy can be improved by using GP. Accuracy is studied and compared with previous work in Paper B where linear log-distance model is used.

Unlike Paper B and C, where the focus is on positioning of a moving device with certain assumed mobility patterns, Paper D proposes a fingerprint based position estimation framework with GP. This is due to the fact that in practice, there also exist use cases where static positioning is preferred without considering mobility (e.g., the device is not moving or moving without any continuous pattern). In such kind of problems, fingerprinting is usually selected to infer the position of the device. However, the problem associated with fingerprinting method is the collection and maintenance of a relatively large RSS fingerprint database/map. In this work, we propose and compare two algorithms namely, advanced Kriging method and Gaussian process, to reconstruct the RSS map with incomplete training data. To validate the effectiveness of both algorithms, experiments with BLE infrastructure have been conducted. RSS measurements are collected along

predefined tracks. Both algorithms are applied to reconstruct the full RSS map within the whole area of interest. Further, statistics about the accuracy of RSS map reconstructed are compared and analyzed. Finally, with the reconstructed complete RSS map, the localization performance using probabilistic fingerprinting method will be evaluated and compared.

Paper E aims to solve the modeling problems emerging in the sport analytics area. To be more specific, in this work, we apply GPs to flow modeling and prediction of athletes in ski races, but the proposed framework can be generally applied to other use cases with device trajectories of positioned data. Some specific aspects can be addressed when the data is periodic, like in sports where the event is split up over multiple laps along a specific track. Flow models of both the individual skier and a cluster of skiers are derived and analyzed. Performance has been evaluated using data from the Falun Nordic World Ski Championships 2015, in particular the Men's cross country $4 \times 10$ km relay. The results show that the flow models vary spatially for different skiers and clusters. We further demonstrate that GPR provides powerful and accurate models for flow prediction.

## 5.2 Some Insights into Future Work

As proved in Paper C and Paper D, Gaussian processes provide powerful tools for non-parametric modeling of the RSS measurements. However, there is no theoretical analysis to demonstrate how much improvement we can obtain by utilizing the GP model. In literature, Cramér-Rao bounds are usually used as lower limits for the accuracy of estimators. Hence, it is natural to derive CRB for an estimator based on GP assisted filtering and smoothing.

Further more, there are more to explore in estimating flows based on trajectories from positioned devices. One aspect would be analyze the flow of public transportation vehicles based on positions obtained through GPS. Such kind of analysis may be beneficial for the smart plan of traffic, intelligent traffic control and traffic guidance.

Another aspect that attract much of our attention is the area of sports analytics, since in nowadays, most statistics in sport activities are mainly based on camera captured videos. In the near future, it will become more and more common for the athletes to wear sensors which can measure and transmit useful information for further analysis. From such kind of data, a lot of metrics can be studied and analyzed, which can be beneficial for both the coaches and the public audience. This will definitely lead to more research topics that are within our interest, for instance, the dynamic modeling of the motion patterns of athletes, performance prediction based on history data, health condition monitoring, and so on.

# Bibliography


3GPP1. Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification. TS 36.331, 3rd Generation Partnership Project (3GPP). URL http://www.3gpp.org/DynaReport/36331.htm.

3GPP2. Radio Resource Control (RRC); Protocol specification. TS 25.331, 3rd Generation Partnership Project (3GPP). URL http://www.3gpp.org/DynaReport/25331.htm.

J. S. Abel. A bound on mean-square-estimate error. *IEEE Trans. Information Theory*, 39(5):1675–1680, Sep. 1993. ISSN 0018-9448.

L. N. Atallah, J.-P. Barbot, and P. Larzabal. From Chapman-Robbins bound towards Barankin bound in threshold behaviour prediction. *Electronics Letters*, 40(4):279–280, Feb. 2004. ISSN 0013-5194.

Y. Bar-Shalom, T. Kirubarajan, and X.-Rong Li. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., New York, NY, USA, 2002. ISBN 0471221279.

N. Bergman. *Recursive Bayesian Estimation Navigation and Tracking Applications*. PhD thesis, Linköping University, 1999.

C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.

G. C. Bohling. Introduction to geostatistics and Variogram analysis. Lecture Notes, 2005. URL http://people.ku.edu/{~}gbohling/cpe940/Variograms.pdf.

G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, New York, USA, 1973.

M. Briers, A. Doucet, and S. Maskell. Smoothing algorithms for state–space models. *Annals of the Institute of Statistical Mathematics*, 62(1): 61, 2009. ISSN 1572-9052. doi: 10.1007/s10463-009-0236-2. URL http://dx.doi.org/10.1007/s10463-009-0236-2.

L. Bruno, M. Khider, and P. Robertson. On-line training of the path-loss model in Bayesian WLAN indoor positioning. In *International Conference on Indoor Positioning and Indoor Navigation*, pages 1–9, Oct 2013. doi: 10.1109/IPIN.2013.6817897.

M. Bshara, U. Orguner, F. Gustafsson, and L. Van Biesen. Fingerprinting localization in wireless networks based on received-signal-strength measurements: A case study on WiMAX networks. *IEEE Trans. Vehicular Technology*, 59(1): 283–294, Jan. 2010. ISSN 0018-9545. doi: 10.1109/TVT.2009.2030504.

N. Bulusu, J. Heidemann, and D. Estrin. GPS-less low-cost outdoor localization for very small devices. *IEEE Personal Communications*, 7(5):28–34, Oct. 2000. ISSN 1070-9916.

T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *Wireless Communications & Mobile Computing (WCMC): Special Issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, 2: 483–502, 2002.

G. Casella and R. L. Berger. *Statistical inference*. Duxbury advanced series. B. Cole, Pacific Grove (Calif.), 2002. ISBN 0-495-39187-5. URL `http://opac.inria.fr/record=b1134456`.

C. B. Chang and K. P. Dunn. *Applied State Estimation and Association*. MIT press, 2016.

A. Chen, C. Harko, D. Lambert, and P. Whiting. An algorithm for fast, model-free tracking indoors. *SIGMOBILE Mob. Comput. Commun. Rev.*, 11(3):48–58, July 2007. ISSN 1559-1662.

M. P. Deisenroth and J. W. Ng. Distributed Gaussian processes. *arXiv preprint arXiv:1502.02843*, 2015.

R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch. Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G. *IEEE Signal Process. Mag.*, 31(6):102–112, Nov. 2014. ISSN 1053-5888.

R. Douc. Comparison of resampling schemes for particle filtering. In *In 4th International Symposium on Image and Signal Processing and Analysis (ISPA*, pages 64–69, 2005.

A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later, 2009.

A. Doucet, N. de Freitas, and N. Gordon. *An Introduction to Sequential Monte Carlo Methods*, pages 3–14. Springer New York, New York, NY, 2001. ISBN 978-1-4757-3437-9. doi: 10.1007/978-1-4757-3437-9_1. URL `http://dx.doi.org/10.1007/978-1-4757-3437-9_1`.

Z. Duan, V. P. Jilkov, and X. R. Li. State estimation with quantized measurements: Approximate MMSE approach. In *11th International Conference on Information Fusion*, pages 1–6, June 2008.

D. Duvenaud, J. R. Lloyd, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani. Structure discovery in nonparametric regression through compositional kernel search. *arXiv:1302.4922v4*, 2013.

D. Ellis, E. Sommerlade, and I. Reid. Modelling pedestrian trajectory patterns with Gaussian processes. In *Proc. of IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1229–1234, Sep. 2009. doi: 10.1109/ICCVW.2009.5457470.

F. Evennou, F. Marx, and E. Novakov. Map-aided indoor mobile positioning system using particle filter. In *Proc. Int. Conf. Wireless Communications and Networking*, volume 4, pages 2490–2494, March 2005.

B. Ferris, D. Hähnel, and D. Fox. Gaussian processes for signal strength-based location estimation. In *In Proc. of Robotics: Science and Systems (RSS)*, pages 1–8, Philadelphia, USA, August 2006.

C. Fritsche and U. Orguner. On parametric smoothing Cramér-Rao bounds. Other academic LiTH-ISY-R, 3100, Linköping University, Department of Electrical Engineering, Sweden, 2016.

C. Fritsche, U. Orguner, and F. Gustafsson. On parametric lower bounds for discrete-time filtering. In *2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 4338–4342, 2016.

G. Giorgetti, S. K. S. Gupta, and G. Manes. Optimal RSS threshold selection in connectivity-based localization schemes. In *Proc. Int. Symp. on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 220–228, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-235-1.

S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing for nonlinear time series. *Journal of the American Statistical Association*, 99(465):156–168, 2004. doi: 10.1198/016214504000000151. URL `http://dx.doi.org/10.1198/016214504000000151`.

A. Goldsmith. *Wireless Communications*. Cambridge University Press, New York, NY, USA, 2006.

N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F - Radar and Signal Processing*, 140(2):107–113, April 1993. ISSN 0956-375X. doi: 10.1049/ip-f-2.1993.0015.

M. Gudmundson. Correlation model for shadow fading in mobile radio systems. *Electronics Letters*, 27(23):2145–2146, Nov. 1991. ISSN 0013-5194. doi: 10.1049/el:19911328.

F. Gustafsson. *Statistical Sensor Fusion*. Studentlitteratur, Lund, Sweden, 2nd edition, 2012.

F. Gustafsson and R. Karlsson. Generating dithering noise for maximum likelihood estimation from quantized data. *Automatica*, 49(2):554–560, Feb. 2013. ISSN 0005-1098.

F. Gustafsson, F. Gunnarsson, Niclas Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund. Particle filters for positioning, navigation, and tracking. *IEEE Transactions on Signal Processing*, 50(2):425–437, Feb. 2002. ISSN 1053-587X. doi: 10.1109/78.978396.

M. Hata. Empirical formula for propagation loss in land mobile radio services. *IEEE Transactions on Vehicular Technology*, 29(3):317–325, Aug 1980. ISSN 0018-9545. doi: 10.1109/T-VT.1980.23859.

T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher. Range-free localization schemes for large scale sensor networks. In *Proc. of Annual International Conference on Mobile Computing and Networking*, pages 81–95, New York, NY, USA, 2003. ACM. ISBN 1-58113-753-2.

R. J. Herring. *Real-Time Traffic Modeling and Estimation with Streaming Probe Data using Machine Learning*. PhD thesis, UC Berkeley: Industrial Engineering & Operations Research, 2010.

N. Hjort, C. Holmes, P. Mueller, and S. Walker. *Bayesian Nonparametrics: Principles and Practice*. Cambridge University Press, Cambridge, UK, 2010.

J. D. Hol, T. B. Schon, and F. Gustafsson. On resampling algorithms for particle filters. In *2006 IEEE Nonlinear Statistical Signal Processing Workshop*, pages 79–82, Sept. 2006. doi: 10.1109/NSSPW.2006.4378824.

M. F. Huber. Recursive Gaussian process regression. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3362–3366, May 2013.

M. F Huber. Recursive Gaussian process: On-line regression and learning. *Pattern Recognition Letters*, 45:85–91, 2014. ISSN 0167-8655.

I. A. Ibargimov and R. Z. Hasminskii. *Statistical Estimation. Asymptotic Theory*. Springer-Verlag, Berlin, 1981.

S. S. Jan, S. J. Yeh, and Y. Liu. Received signal strength database interpolation by kriging for a Wi-Fi indoor positioning system. *Sensors(Basel, Switzerland)*, 15 (9):21377–21393, 2015.

A. H. Jazwinski. *Stochastic processes and filtering theory*. Mathematics in science and engineering. Academic press, New York, 1970. ISBN 0-12-381550-9. URL `http://opac.inria.fr/record=b1078357`. UKM.

D. Jin, F. Yin, C. Fritsche, A. M. Zoubir, and F. Gustafsson. Dithering in quantized RSS based localization. In *Proc. Int. Conf. Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2015.

J. Joseph, F. Doshi-Velez, S. Huang, and N. Roy. A Bayesian nonparametric approach to modeling motion patterns. *Auton. Robots*, 31(4):383–400, Nov. 2011. ISSN 0929-5593.

R. E. Kalman. A new approach to linear filtering and prediction problems. *ASME Journal of Basic Engineering*, 1960.

R. Karlsson and F. Gustafsson. Filtering and estimation for quantized sensor information. In *13th European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey, Sept. 2005.

G. Kitagawa. The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother. *Annals of the Institute of Statistical Mathematics*, 46(4):605–623, 1994. ISSN 1572-9052. doi: 10.1007/BF00773470. URL `http://dx.doi.org/10.1007/BF00773470`.

G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996. ISSN 10618600. URL `http://www.jstor.org/stable/1390750`.

E. D. Klenske, M. N. Zeilinger, B. Scholkopf, and P. Hennig. Gaussian process-based predictive control for periodic error correction. *IEEE Transactions on Control Systems Technology*, 24(1):110–121, Jan 2016. ISSN 1063-6536. doi: 10.1109/TCST.2015.2420629.

J. Ko and D. Fox. GP-BayesFilters: Bayesian filtering using Gaussian process prediction and observation models. In *Proc. Int. Conf. on Intelligent Robots and Systems (IROS),*, pages 3471–3476, Nice, France, Sep. 2008. doi: 10.1109/IROS.2008.4651188.

A. Konak. Estimating path loss in wireless local area networks using ordinary Kriging. In *Proc. 2010 Winter Simul. Conf.*, pages 2888–2896, Baltimore, MD, 2010.

D. P. Kroese, T. Brereton, T. Taimre, and Z. I. Botev. Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(6):386–392, 2014. ISSN 1939-0068. doi: 10.1002/wics.1314. URL `http://dx.doi.org/10.1002/wics.1314`.

B. Li, Y. Wang, H. K. Lee, A. Dempster, and C. Rizos. Method for yielding a database of location fingerprints in WLAN. *IEE Proc. Commun.*, 152(5):580–586, 2005.

J. Lin, M. Vlachos, E. Keogh, and D. Gunopulos. *Advances in Database Technology: 9th International Conference on Extending Database Technology*, chapter Iterative Incremental Clustering of Time Series, pages 106–122. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. ISBN 978-3-540-24741-8.

C. Liu, A. Kiring, N. Salman, and L. Mihaylova. A Kriging algorithm for location fingerprinting based on received signal strength. In *Proc. 10th Work. Sens. Data Fusion*, Bonn, Germany, 2015.

J. S. Liu and R. Chen. Blind deconvolution via sequential imputations. *Journal of the American Statistical Association*, 90(430):567–576, 1995. doi: 10.1080/01621459.1995.10476549.

L. Ljung. *System Identification (2Nd Ed.): Theory for the User*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1999. ISBN 0-13-656695-2.

D. J. C. MacKay. Gaussian processes – a replacement for supervised neural networks?, 1997.

A. Manirabona and L. C. Fourati. A Kriged fingerprinting for wireless body area network indoor localization. *Wirel. Pers. Commun.*, 80(4):1501–1515, 2015. ISSN 09296212. doi: 10.1007/s11277-014-2095-2.

T. L. Marzetta. Computing the Barankin bound, by solving an unconstrained quadratic optimization problem. In *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages 3829–3832, Apr. 1997.

N. A. Maung Maung and M. Kawai. Performance improvement of connectivity-based localization using iterative learning. In *in Proc. of Int. Conf. Localization and GNSS (ICL-GNSS)*, pages 1–6, Helsinki, Finland, June 2014. doi: 10.1109/ICL-GNSS.2014.6934177.

A. Mousa, Y. Dama, M. Najjar, and B. Alsayeh. Optimizing outdoor propagation model based on measurements for multiple RF cell. *International Journal of Computer Applications*, 60(5):5–10, December 2012. Full text available.

N. Patwari and A. O. Hero, III. Using proximity and quantized RSS for sensor localization in wireless networks. In *Proc. ACM Int. Conf. on Wireless Sensor Networks and Applications*, pages 20–29, New York, NY, USA, 2003. ACM. ISBN 1-58113-764-8.

J. Quiñonero Candela and C. E. Rasmussen. A unifying view of sparse approximate Gaussian process regression. *J. Mach. Learn. Res.*, 6:1939–1959, Dec. 2005. ISSN 1532-4435.

C. E. Rasmussen and H. Nickisch. Gaussian processes for machine learning (GPML) toolbox. *J. Mach. Learn. Res.*, 11:3011–3015, Dec. 2010. ISSN 1532-4435.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, USA, 2006.

H. E. Rauch, C. T. Striebel, and F. Tung. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3:1445–1450, August 1965. doi: 10.2514/3.3166.

S. Raza, P. Misra, T. He, and T. Voigt. Bluetooth smart: An enabling technology for the Internet of Things. In *IEEE Int. Conf. on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 155–162, Abu Dhabi, UAE, Oct. 2015. doi: 10.1109/WiMOB.2015.7347955.

S. Reece, R. Mann, I. Rezek, and S. Roberts. Gaussian process segmentation of co-moving animals. *Proc. AIP Conference*, 1305(1):430–437, 2011. doi: http://dx.doi.org/10.1063/1.3573650.

P. Richter and M. Toledano-Ayala. Revisiting Gaussian process regression modeling for localization in wireless sensor networks. *Sensors (Basel)*, 15(9):22587–22615, Sep. 2015.

N. Salman, N. Alsindi, L. Mihaylova, and A. H. Kemp. Super resolution WiFi indoor localization and tracking. In *2014 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–5, Oct 2014. doi: 10.1109/SDF.2014.6954723.

S. Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press., 2013.

S. Särkkä, A. Solin, and J. Hartikainen. Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through Kalman filtering. *IEEE Signal Processing Magazine*, 30(4): 51–61, July 2013. ISSN 1053-5888. doi: 10.1109/MSP.2013.2246292.

A. Schwaighofer, M. Grigoras, V. Tresp, and C. Hoffmann. GPPS: A Gaussian process positioning system for cellular networks. In *Advances in Neural Information Processing Systems (NIPS)*. MIT Press, 2004.

Y. Shen, A. Y. Ng, and M. Seeger. Fast Gaussian process regression using KD-trees. In *Proc. Advances in Neural Information Processing Systems*. MIT Press, 2006.

B. L. Smith and M. J. Demetsky. Short-term traffic flow prediction: neural network approach. *Transportation Research Record*, (1453), 1994.

E. Snelson and Z. Ghahramani. Sparse Gaussian processes using pseudo-inputs. In *Proc. Int. Conf. Advances in Neural Information Processing Systems*, pages 1257–1264. MIT press, 2006.

J. Talvitie, M. Renfors, and E. S. Lohan. Distance-based interpolation and extrapolation methods for RSS-based localization with indoor wireless signals. *IEEE Transactions on Vehicular Technology*, 64(4):1340–1353, April 2015. ISSN 0018-9545. doi: 10.1109/TVT.2015.2397598.

P. Tichavsky, C. H. Muravchik, and A. Nehorai. Posterior Cramer-Rao bounds for discrete-time nonlinear filtering. *IEEE Transactions on Signal Processing*, 46 (5):1386–1396, May 1998. ISSN 1053-587X. doi: 10.1109/78.668800.

M. K. Titsias. Variational learning of inducing variables in sparse Gaussian processes. In *Proc. Artificial Intelligence and Statistics*, pages 567–574, 2009.

TRAX. European Union FP7 Tracking in Complex Sensor Systems (TRAX) project. https://www.trax.utwente.nl/, 2014.

H. L. Van Trees. *Detection, estimation, and modulation theory 1*. Wiley, 1 edition, September 1968. ISBN 0471095176.

R. Want, A. Hopper, V. Falcão, and J. Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, 10(1):91–102, Jan. 1992. ISSN 1046-8188. doi: 10.1145/128756.12875.

D. Wu and L. Xia. Reference device-assisted adaptive location fingerprinting. *Sensors*, 16(6):802, 2016. ISSN 1424-8220. doi: 10.3390/s16060802. URL `http://www.mdpi.com/1424-8220/16/6/802`.

H. Wymeersch, S. Marano, W. M. Gifford, and M. Z. Win. A machine learning approach to ranging error mitigation for UWB localization. *IEEE Transactions on Communications*, 60(6):1719–1728, June 2012. ISSN 0090-6778. doi: 10.1109/TCOMM.2012.042712.110035.

Y. Xie, K. Zhao, Y. Sun, and D. Chen. Gaussian processes for short-term traffic volume forecasting. *Transportation Research Record: Journal of the Transportation Research Board*, 2165:69–78, 2010.

K. Yedavalli, B. Krishnamachari, S. Ravula, and B. Srinivasan. Ecolocation: a sequence based technique for RF localization in wireless sensor networks. In *Proc. Information Processing in Sensor Networks (IPSN)*, pages 285–292, Los Angeles, California, USA, April 2005.

F. Yin. *Robust Wireless Localization in Harsh Mixed Line-of-Sight/Non-Line-of-Sight Environments*. PhD thesis, Technische Universitaet Darmstadt, Darmstadt, Germany, Aug. 2014.

F. Yin, Y. Zhao, and F. Gunnarsson. Proximity report triggering threshold optimization for network-based indoor positioning. In *Proc. Int. Conf. on Information Fusion*, pages 1061–1069, Washington D.C., USA, July 2015.

F. Yin, Y. Zhao, and F. Gunnarsson. Fundamental bounds on position estimation using proximity reports. In *Proc. IEEE 83rd Vehicular Technology Conference: VTC2016-Spring*, 2016.

F. Yin, Y. Zhao, F. Guannarsson, and F. Gustafsson. Received-signal-strength threshold optimization using Gaussian processes. *Transactions on Signal Processing*, 65(8):2164–2177, 2017.

Y. Zhang and Y. Xie. Forecasting of short-term freeway volume with v-support vector machines. *Transportation Research Record: Journal of the Transportation Research Board*, (2024):92–99, 2008.

Y. Zhao. *Lab Report: Localization Using a Microphone Network, TSRT14 Sensor Fusion*, 2015.

Y. Zhao. *Lab Report: Gaussian Process Regression, 732A52 Introduction to Machine Learning*, 2016.

Y. Zhao, F. Yin, F. Gunnarsson, M. Amirijoo, E. Özkan, and F. Gustafsson. Partile filtering for positioning based on proximity report. In *Proc. Int. Conf. on Information Fusion*, pages 1046–1052, Washington D.C., USA, July 2015.

Y. Zhao, F. Yin, F. Gunnarsson, M. Amirijoo, and G. Hendeby. Gaussian processes for propagation modeling and proximity based indoor positioning. In *Proc. IEEE 83rd Vehicular Technology Conference: VTC-Spring*, 2016a.

Y. Zhao, F. Yin, F. Gunnarsson, F. Hultkratz, and J. Fagerlind. Gaussian processes for flow modeling and prediction of positioned trajectories evaluated with sports data. In *Proc. 19th International Conference on Information Fusion (FUSION)*, pages 1461–1468, July 2016b.

Y. Zhao, C. Fritsche, F. Yin, F. Gunnarsson, and F. Gustafsson. Sequential Monte Carlo methods and theoretical bounds for state inference based on proximity reports. *To be submitted to IEEE Transactions on Vehicular Technology*, 2017a.

Y. Zhao, C. Liu, L. S. Mihaylova, and F. Guannarsson. Deterministic Kriging and probabilistic Gaussian processes for RSS reconstruction in indoor localization. *Manuscipt to be submitted to Transactions on Intelligent Transportation Systems*, 2017b.

L. Zuo, R. Niu, and P. K. Varshney. A sensor selection approach for target tracking in sensor networks with quantized measurements. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2521–2524, March 2008. doi: 10.1109/ICASSP.2008.4518161.

# Part II

# Publications

**Licentiate Theses**
**Division of Automatic Control**
**Linköping University**

**P. Andersson:** Adaptive Forgetting through Multiple Models and Adaptive Control of Car Dynamics. Thesis No. 15, 1983.

**B. Wahlberg:** On Model Simplification in System Identification. Thesis No. 47, 1985.

**A. Isaksson:** Identification of Time Varying Systems and Applications of System Identification to Signal Processing. Thesis No. 75, 1986.

**G. Malmberg:** A Study of Adaptive Control Missiles. Thesis No. 76, 1986.

**S. Gunnarsson:** On the Mean Square Error of Transfer Function Estimates with Applications to Control. Thesis No. 90, 1986.

**M. Viberg:** On the Adaptive Array Problem. Thesis No. 117, 1987.

**K. Ståhl:** On the Frequency Domain Analysis of Nonlinear Systems. Thesis No. 137, 1988.

**A. Skeppstedt:** Construction of Composite Models from Large Data-Sets. Thesis No. 149, 1988.

**P. A. J. Nagy:** MaMiS: A Programming Environment for Numeric/Symbolic Data Processing. Thesis No. 153, 1988.

**K. Forsman:** Applications of Constructive Algebra to Control Problems. Thesis No. 231, 1990.

**I. Klein:** Planning for a Class of Sequential Control Problems. Thesis No. 234, 1990.

**F. Gustafsson:** Optimal Segmentation of Linear Regression Parameters. Thesis No. 246, 1990.

**H. Hjalmarsson:** On Estimation of Model Quality in System Identification. Thesis No. 251, 1990.

**S. Andersson:** Sensor Array Processing; Application to Mobile Communication Systems and Dimension Reduction. Thesis No. 255, 1990.

**K. Wang Chen:** Observability and Invertibility of Nonlinear Systems: A Differential Algebraic Approach. Thesis No. 282, 1991.

**J. Sjöberg:** Regularization Issues in Neural Network Models of Dynamical Systems. Thesis No. 366, 1993.

**P. Pucar:** Segmentation of Laser Range Radar Images Using Hidden Markov Field Models. Thesis No. 403, 1993.

**H. Fortell:** Volterra and Algebraic Approaches to the Zero Dynamics. Thesis No. 438, 1994.

**T. McKelvey:** On State-Space Models in System Identification. Thesis No. 447, 1994.

**T. Andersson:** Concepts and Algorithms for Non-Linear System Identifiability. Thesis No. 448, 1994.

**P. Lindskog:** Algorithms and Tools for System Identification Using Prior Knowledge. Thesis No. 456, 1994.

**J. Plantin:** Algebraic Methods for Verification and Control of Discrete Event Dynamic Systems. Thesis No. 501, 1995.

**J. Gunnarsson:** On Modeling of Discrete Event Dynamic Systems, Using Symbolic Algebraic Methods. Thesis No. 502, 1995.

**A. Ericsson:** Fast Power Control to Counteract Rayleigh Fading in Cellular Radio Systems. Thesis No. 527, 1995.

**M. Jirstrand:** Algebraic Methods for Modeling and Design in Control. Thesis No. 540, 1996.

**K. Edström:** Simulation of Mode Switching Systems Using Switched Bond Graphs. Thesis No. 586, 1996.

**J. Palmqvist:** On Integrity Monitoring of Integrated Navigation Systems. Thesis No. 600, 1997.

**A. Stenman:** Just-in-Time Models with Applications to Dynamical Systems. Thesis No. 601, 1997.

**M. Andersson:** Experimental Design and Updating of Finite Element Models. Thesis No. 611, 1997.

**U. Forssell:** Properties and Usage of Closed-Loop Identification Methods. Thesis No. 641, 1997.

**M. Larsson:** On Modeling and Diagnosis of Discrete Event Dynamic systems. Thesis No. 648, 1997.

**N. Bergman:** Bayesian Inference in Terrain Navigation. Thesis No. 649, 1997.

**V. Einarsson:** On Verification of Switched Systems Using Abstractions. Thesis No. 705, 1998.

**J. Blom, F. Gunnarsson:** Power Control in Cellular Radio Systems. Thesis No. 706, 1998.

**P. Spångéus:** Hybrid Control using LP and LMI methods – Some Applications. Thesis No. 724, 1998.

**M. Norrlöf:** On Analysis and Implementation of Iterative Learning Control. Thesis No. 727, 1998.

**A. Hagenblad:** Aspects of the Identification of Wiener Models. Thesis No. 793, 1999.

**F. Tjärnström:** Quality Estimation of Approximate Models. Thesis No. 810, 2000.

**C. Carlsson:** Vehicle Size and Orientation Estimation Using Geometric Fitting. Thesis No. 840, 2000.

**J. Löfberg:** Linear Model Predictive Control: Stability and Robustness. Thesis No. 866, 2001.

**O. Härkegård:** Flight Control Design Using Backstepping. Thesis No. 875, 2001.

**J. Elbornsson:** Equalization of Distortion in A/D Converters. Thesis No. 883, 2001.

**J. Roll:** Robust Verification and Identification of Piecewise Affine Systems. Thesis No. 899, 2001.

**I. Lind:** Regressor Selection in System Identification using ANOVA. Thesis No. 921, 2001.

**R. Karlsson:** Simulation Based Methods for Target Tracking. Thesis No. 930, 2002.

**P.-J. Nordlund:** Sequential Monte Carlo Filters and Integrated Navigation. Thesis No. 945, 2002.

**M. Östring:** Identification, Diagnosis, and Control of a Flexible Robot Arm. Thesis No. 948, 2002.

**C. Olsson:** Active Engine Vibration Isolation using Feedback Control. Thesis No. 968, 2002.

**J. Jansson:** Tracking and Decision Making for Automotive Collision Avoidance. Thesis No. 965, 2002.

**N. Persson:** Event Based Sampling with Application to Spectral Estimation. Thesis No. 981, 2002.

**D. Lindgren:** Subspace Selection Techniques for Classification Problems. Thesis No. 995, 2002.

**E. Geijer Lundin:** Uplink Load in CDMA Cellular Systems. Thesis No. 1045, 2003.

**M. Enqvist:** Some Results on Linear Models of Nonlinear Systems. Thesis No. 1046, 2003.

**T. Schön:** On Computational Methods for Nonlinear Estimation. Thesis No. 1047, 2003.

**F. Gunnarsson:** On Modeling and Control of Network Queue Dynamics. Thesis No. 1048, 2003.

**S. Björklund:** A Survey and Comparison of Time-Delay Estimation Methods in Linear Systems. Thesis No. 1061, 2003.

**M. Gerdin:** Parameter Estimation in Linear Descriptor Systems. Thesis No. 1085, 2004.

**A. Eidehall:** An Automotive Lane Guidance System. Thesis No. 1122, 2004.

**E. Wernholt:** On Multivariable and Nonlinear Identification of Industrial Robots. Thesis No. 1131, 2004.

**J. Gillberg:** Methods for Frequency Domain Estimation of Continuous-Time Models. Thesis No. 1133, 2004.

**G. Hendeby:** Fundamental Estimation and Detection Limits in Linear Non-Gaussian Systems. Thesis No. 1199, 2005.

**D. Axehill:** Applications of Integer Quadratic Programming in Control and Communication. Thesis No. 1218, 2005.

**J. Sjöberg:** Some Results On Optimal Control for Nonlinear Descriptor Systems. Thesis No. 1227, 2006.

**D. Törnqvist:** Statistical Fault Detection with Applications to IMU Disturbances. Thesis No. 1258, 2006.

**H. Tidefelt:** Structural algorithms and perturbations in differential-algebraic equations. Thesis No. 1318, 2007.

**S. Moberg:** On Modeling and Control of Flexible Manipulators. Thesis No. 1336, 2007.

**J. Wallén:** On Kinematic Modelling and Iterative Learning Control of Industrial Robots. Thesis No. 1343, 2008.

**J. Harju Johansson:** A Structure Utilizing Inexact Primal-Dual Interior-Point Method for Analysis of Linear Differential Inclusions. Thesis No. 1367, 2008.

**J. D. Hol:** Pose Estimation and Calibration Algorithms for Vision and Inertial Sensors. Thesis No. 1370, 2008.

**H. Ohlsson:** Regression on Manifolds with Implications for System Identification. Thesis No. 1382, 2008.

**D. Ankelhed:** On low order controller synthesis using rational constraints. Thesis No. 1398, 2009.

**P. Skoglar:** Planning Methods for Aerial Exploration and Ground Target Tracking. Thesis No. 1420, 2009.

**C. Lundquist:** Automotive Sensor Fusion for Situation Awareness. Thesis No. 1422, 2009.

**C. Lyzell:** Initialization Methods for System Identification. Thesis No. 1426, 2009.

**R. Falkeborn:** Structure exploitation in semidefinite programming for control. Thesis No. 1430, 2010.

**D. Petersson:** Nonlinear Optimization Approaches to $\mathcal{H}_2$-Norm Based LPV Modelling and Control. Thesis No. 1453, 2010.

**Z. Sjanic:** Navigation and SAR Auto-focusing in a Sensor Fusion Framework. Thesis No. 1464, 2011.

**K. Granström:** Loop detection and extended target tracking using laser data. Thesis No. 1465, 2011.

**J. Callmer:** Topics in Localization and Mapping. Thesis No. 1489, 2011.

**F. Lindsten:** Rao-Blackwellised particle methods for inference and identification. Thesis No. 1480, 2011.

**M. Skoglund:** Visual Inertial Navigation and Calibration. Thesis No. 1500, 2011.

**S. Khoshfetrat Pakazad:** Topics in Robustness Analysis. Thesis No. 1512, 2011.

**P. Axelsson:** On Sensor Fusion Applied to Industrial Manipulators. Thesis No. 1511, 2011.

**A. Carvalho Bittencourt:** On Modeling and Diagnosis of Friction and Wear in Industrial Robots. Thesis No. 1516, 2012.

**P. Rosander:** Averaging level control in the presence of frequent inlet flow upsets. Thesis No. 1527, 2012.

**N. Wahlström:** Localization using Magnetometers and Light Sensors. Thesis No. 1581, 2013.

**R. Larsson:** System Identification of Flight Mechanical Characteristics. Thesis No. 1599, 2013.

**Y. Jung:** Estimation of Inverse Models Applied to Power Amplifier Predistortion. Thesis No. 1605, 2013.

**M. Syldatk:** On Calibration of Ground Sensor Networks. Thesis No. 1611, 2013.

**M. Roth:** Kalman Filters for Nonlinear Systems and Heavy-Tailed Noise. Thesis No. 1613, 2013.

**D. Simon:** Model Predictive Control in Flight Control Design — Stability and Reference Tracking. Thesis No. 1642, 2014.

**J. Dahlin:** Sequential Monte Carlo for inference in nonlinear state space models. Thesis No. 1652, 2014.

**M. Kok:** Probabilistic modeling for positioning applications using inertial sensors. Thesis No. 1656, 2014.

**J. Linder:** Graybox Modelling of Ships Using Indirect Input Measurements. Thesis No. 1681, 2014.

**G. Mathai:** Direction of Arrival Estimation of Wideband Acoustic Wavefields in a Passive Sensing Environment. Thesis No. 1721, 2015.

**I. Nielsen:** On Structure Exploiting Numerical Algorithms for Model Predictive Control. Thesis No. 1727, 2015.

**C. Veibäck:** Tracking of Animals Using Airborne Cameras. Thesis No. 1761, 2016.

**N. Evestedt:** Sampling Based Motion Planning for Heavy Duty Autonomous Vehicles. Thesis No. 1762, 2016.