THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Tracking Mobile Phones in Urban Areas

Christer Åkerblom

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg, Sweden
Göteborg, September 2000

Tracking Mobil Phones in Urban Areas
CHRISTER ÅKERBLOM

ISSN 0347-2809/NO 2000:57
Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31-772 1000

Matematiskt centrum
Göteborg, September 2000

Abstract

In this report we consider the problem of position tracking of mobile phones. The Kalman filter, with extensions, is a popular solution to this type of filtering problems, and we review a recent article by Hellebrandt and Mathar [19] as an example of a mobile phone tracking application of the Kalman filter.

The particle filters has recently received much interest in nonlinear and non Gaussian filtering applications. These Monte Carlo based methods provide a non-parametric approximation to the distribution of the filter state, conditional on the observations. We give an introduction to the particle filter and propose it as an alternative to the Kalman filter in positioning of mobile phones.

A simple model of the signal strength fading between mobiles and microcells in an urban area is developed, and used in a simulation to estimate the position of a mobile phone using the received signal strength as observation. The performance of the Kalman filter and the particle filter is investigated and the particle filter shows superior performance compared to the Kalman filter both concerning tracking accuracy and global positioning ability. These features makes the particle filter very attractive for mobile positioning and tracking applications.

Finally, we discuss possible extensions to the simple models of movement used in the simulations and the combination of different positioning methods to obtain higher accuracy and reliability in the position estimates.

Keywords: Mobile positioning, tracking, Kalman filter, particle filter, nonlinear filtering, density estimation, signal model.

MSC 2000 subject classifications: 62M20, 62p30

Preface

The Report

This licentiate thesis is a part of an ongoing research project about statistical methods in telecommunication at the Stochastic Centre and the Department of Mathematical Statistics, Chalmers University of Technology and Göteborg University.

During this thesis project I have been supported by the Swedish Foundation for Strategic Research (SSF) through the Stochastic Network in the National Network in Applied Mathematics (NTM).

The thesis is also the final part of the ECMI (European Consortium for Mathematics in Industry) five semester, post-graduate program in Industrial Mathematics. The ECMI program includes a block of core courses covering several areas of applied mathematics and a block of specialised courses in an area of interest. The final part of the program is a project devoted to an industry related problem and I have been working with Ericsson Radio Systems AB in Kista. This thesis summarises the mathematical results from the project.

Acknowledgements

I wish to express my gratitude to several people:

First, I would like to thank my supervisor Mats Rudemo for suggesting an interesting problem to me, for encouragement and for supporting me in my many changes of direction during the work.

Magnus Almgren at Ericsson Radio Systems AB has given the project the touch of reality that I hope it contains, but should not be blamed when it is lacking since many, but not all, of his opinions have been included. He has also learned me that writing MATLAB code is a good way to communicate, even if we have had our share of language related problems.

I would also like to thank Henrik Nyberg, also at Ericsson Radio Systems AB, who initiated the contact with Ericsson which has been very fruitful for me. He has also always been interested in, and willing to discuss the thesis.

I spent three months during my studies at BioSS (Biomathematics & Statistics Scotland) in beautiful Edinburgh, Scotland. I would like to thank my host

Dr. Chris Glasbey and the rest off the staff at BioSS for giving me such a pleasant stay.

Finally, I would like to thank my fiancée Johanna and my big family for continuous support and encouragement during my studies.

List of Acronyms

Telecommunication is certainly an area of engineering where abbreviations and acronyms are frequently used. All such words used in the thesis are introduced in the text, but for reference they are also presented in this list.

AOA	Angle of Arrival
BTS	Base Transceiver Station
FCC	Federal Commission of Communication
GPS	Global Positioning System
GSM	Groupe Spécial Mobile or Global System for Mobile communication
LOS	Line of Sight
LSC	Location Service Client
MLS	Mobile Location Solution
Mobile	Mobile phone
MPC	Mobile Positioning Centre
NLOS	Non Line of Sight
PDF	Probability Density Function
PSTN	Public Switched Telephone Network
RAA	Resource Allocation Algorithm
SIR	Signal to Interference ratio
TDOA	Time Difference of Arrival
TOA	Time of Arrival

Contents

Chapter 1

Introduction

1.1 Problem Introduction

Tracking of mobile phones has recently received a lot of attention in both media and engineering science. The possibility to make reliable position estimates opens doors to many interesting new services that can be offered to the user by combining a positioning tool with position specific information. Position dependent information available directly in the mobile phone is thus a hot issue nowadays.

Reliable position estimates require models of user mobility, but more important is a good model of the signal propagation and hence much work is put into this kind of modelling. Several different methods have been proposed for location estimation using only the information available in the received signal.

There are some constraints to have in mind when designing a position tracking algorithm for mobile phones. The tracking must be done in real time and should not require too expensive hardware changes in the system. It must also be easy to incorporate new information such as spatial changes in the area, calling activity in different areas or other relevant information.

1.2 Ericsson's Solution

Ericsson AB mobile positioning service is called mobile location solution (MLS) and is a framework where there is room for different positioning techniques, making it possible to satisfy different information quality demands and to be flexible to new positioning techniques. Ericsson has developed the MLS framework to meet the rapidly changing demand of products and services within positioning. The design of the system allows development of applications independently of the mobile positioning system. MLS consists of three subsystems [32].

1. The positioning subsystem consists of a variety of different positioning methods such as:

(a) a network based positioning system, which only uses information available within the cellular network, *i.e.* time difference between signals, signal strength, signal angle of arrival or combinations thereof.

(b) a network assisted GPS[1] system, which uses the cellular network information to improve the GPS estimates.

(c) other techniques. Here is room for other future or non-standardised methods such as the SIM card toll-kit.

2. The mobile positioning centre (MPC) is a gateway subsystem in MLS which retrieves data from the positioning subsystem and converts it to information for the location service client (LSC). The MPC makes it possible to monitor the usage of position information hence allowing operators to charge for them.

3. The LSC subsystem contains applications. There are both internal applications like emergency services and external applications that could be provided by the system operators or other application developers.

In this thesis we will deal with problems and methods that concern part 1a of the positioning subsystem described above.

1.3 Outline of the Report

In Chapter 2, we give a brief introduction to basic concepts in wireless communications. Abbreviations are frequently used in the area of mobile communication and some of them will be defined in this chapter. Different methods for mobile positioning are presented in Chapter 3. Chapter 4 reviews basic filtering theory and especially the discrete time Kalman filter which has been used in several studies of mobile location tracking.

Chapter 5 is devoted to the study of particle filters. First we give an introduction to the SIR filter and then other kinds of particle filters are presented. Finally, some improvements and implementation issues are discussed.

In Chapter 6 a signal model, that emphasises the shifts between line of sight and non line of sight conditions, is developed for a Manhattan like area. The performance of the particle filter is compared with the Kalman filter in a simulation using the signal strength as observations. In the simulations very simple models for mobile movement is used but in Chapter 7 we discuss more complex models and also give some hints on how to incorporate, and take advantage of, different kinds of information in the system. Finally, a summary will conclude the report in Chapter 8.

---

[1] Global Positioning System

Chapter 2

Concepts in Mobile Communication

In this chapter we review some of the basic concepts in wireless communication that are used later in the thesis. For the interested reader the books [1] and [30] give a much more complete introduction to the subject compared to the short survey presented here.

2.1 Some Basic Concepts

A communication system where the number of users and their locations are not known *a priori* is often referred to as a wireless network. In Figure 2.1 a schematic picture of a GSM wireless network is given. The network can be divided into two parts; a fixed network and a wireless system. The fixed network provides connections, by cable or microwave link, between the base transceiver stations (BTS) and connects the wireless network to other networks as the ordinary public switched telephone network (PSTN) and the Internet. The wireless system provides connections to the mobile phones moving in the area covered by the BTS. For notational convenience we will call the mobile phones mobiles from now on.

The area around a BTS where the conditions are favourable enough to maintain communication between the mobile and the BTS is called the coverage area of the BTS. The coverage area is often highly non-regular and might contain holes due to complex signal propagation conditions in the area. This fact causes problems for the system designer whose goal is to minimise the number of BTS needed to offer adequate service in the area. The union of the coverage areas must thus equal the service area[1] and in reality there must also be some overlapping. However, if some areas are more frequently visited by mobiles than others, it is rather the probability that a randomly chosen user can be provided

[1] The service area is the area where the operator wants to maintain service.

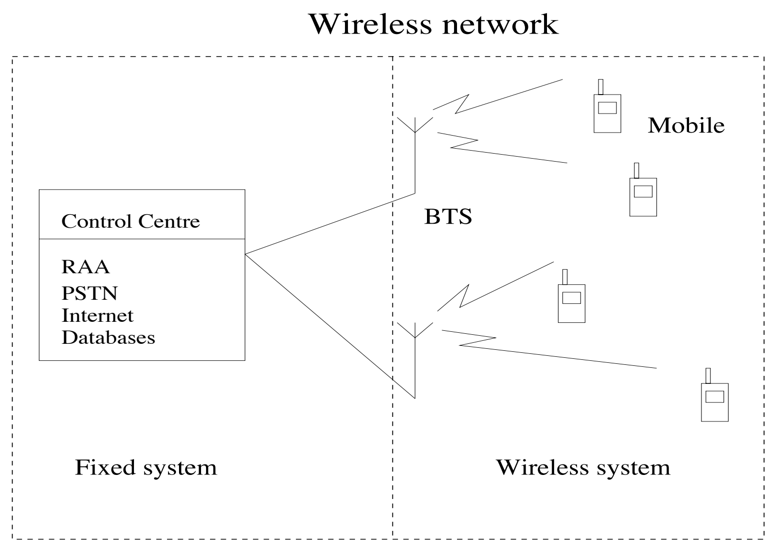


Figure 2.1: *A schematic picture of a wireless network. The mobile wireless network consists of two systems; a fixed and a wireless system.*

adequate service that should be optimised, since the number of channels at each BTS is limited. The latter approach require good models of both calling intensities, movement and data traffic which makes it an even harder problem.

The traditional picture of a mobile wireless system is a plane covered by hexagonal cells as in Figure 2.2. A BTS is in the centre of each cell and the cell borders mark the points where the communication shifts to be more favourable within the neighbouring cell than in the present. It is important to remember that the hexagonal cells give a very idealised model of the signal propagation since the true coverage areas are highly non-regular, but nevertheless they provide a useful picture to keep in mind when thinking about wireless networks. The process when the mobile changes from one BTS to another is called hand-over and is done automatically in the GSM (Global System for Mobile communications) system and can be triggered by a number of events. For example, when there is a major difference in received signal power between different BTS a hand-over is made if there is a free channel available. The hand-over is mostly a problem for fast moving mobiles but since it takes up important system resources, such as time and bandwidth, the number of hand-overs should be minimised.

In the early days of wireless communication the systems were often range limited and the noise was a major concern. Today the number of users in the systems is so large compared with the available bandwidth that the channels used at one BTS must be reused at another, and thus interference between the signals using the same channels will be a major concern. The signal to

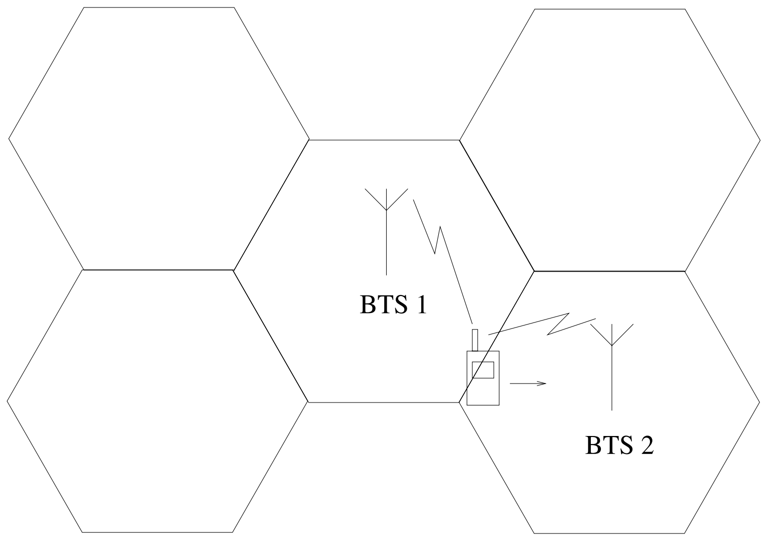


Figure 2.2: *The hexagonal cell model. A hand-over takes place between BTS 1 and BTS 2. The hexagonal cell model is natural when the signal power is strictly decreasing with distance, but that is seldom the case in a real application.*

interference ratio (SIR) is defined as the ratio of received signal power and received interference power and is a relevant parameter to study when designing the system. An interesting fact is that fast decay of signal power increases the SIR and makes it possible to reuse the frequency spectrum more often which will allow a higher user intensity. The smallest distance between two BTS using the same channel is called the reuse distance of the channel.

An important part of the system is the resource allocation algorithm (RAA) which describes how access ports, channels and power are assigned. A good allocation algorithm will assign links between the mobiles and the BTS that meet the SIR requirements for as many mobiles as possible. The SIR must be high enough both in the up-link (mobile to BTS) and in the down-link (BTS to mobile). The largest amount of mobiles that can be handled by the system is called the system capacity. Unfortunately it is not easy to find an allocation scheme maximising the system capacity at each instant. This problem is hard and there is no exact efficient solution available today, as far as the author knows. A naive complete-search-approach takes too much time since the number of feasible allocations grows exponentially with the number of mobiles and hence the algorithms in work today often use some kind of heuristics.

2.2 Performance Measures

The system capacity, introduced in the last section, is a random variable since it is dependent on the number of mobiles and their positions, which naturally are random. The classical performance measure in teletraffic theory is instead the maximum rate of calls for which the probability that a newly arrived request is denied access is kept below a certain level. This probability is called the blocking probability. The blocking probability is often too complex to use in mobile communication since it ideally should deal with phenomena like hand-over and lost calls due to mobility, which would require a very complex model. Another approach, described in [1], that is a little less complex, is to minimise the assignment failure rate ν defined as

$$\nu = \frac{E(Z)}{E(M)},$$

where E is the expectation, M is the number of active calls and $Z = M - Y$ where Y is the number of adequately served calls. For large $E(M)$ the assignment failure rate is a good approximation of the probability that a randomly chosen active mobile at some instant is not provided a useful channel. The instantaneous capacity $\omega^*(\nu_0)$ of a wireless system is defined as the maximum traffic rate ω for which the assignment failure rate ν can be kept below a certain level ν_0,

$$\omega^*(\nu_0) = \{\max \omega : \nu \leq \nu_0\}.$$

To get analytically tractable models a Poisson process is often used to model the rate of incoming service requests and the call lengths are supposed to be exponentially distributed. These models are today being questioned. The possibility to send different kinds of data through the wireless network using different kinds of protocols sometimes make the Poisson assumptions strange and raise many interesting questions.

Baccelli and Zuyev have also studied models based on Poisson assumptions using point processes and stochastic geometry as tools. They show in [3] that several performance evaluation problems based on those models can be posed and solved by computing the expectation of certain functionals of the point processes analytically. The Poisson process is used both for modelling position of BTS, road system and traffic intensity. The mobiles can be in two states; think mode and active mode and the changes are driven by a two state Markov chain. All other characteristics are then expressed as functionals of these point processes and depend thus only on the parameters in their distributions. For example, the distribution of the number of active mobiles in a base station cell and the rate of mobiles crossing a certain cell border is calculated and analytical results are given.

The kind of performance measures described above are valuable for an understanding of what factors that are of importance for the capacity of a system and what parameters that influence the system in general, but can of course be rather crude since they rely on assumptions and models that are not always fulfilled.

Chapter 3

Mobile Positioning and Tracking

To meet the requirements of standardisation regarding mobile positioning, but also to offer new commercial services, a variety of positioning methods have been developed. This section describes why information about mobile position will be important in the near future, and the methods that are used or are under investigation for obtaining this information.

3.1 Exploiting Mobile Position Information

Mobile positioning has recently received increased attention in media. Emergency situations are frequently a reason for interest in positioning of mobiles. Since there has been an explosive growth of the number mobile phone users during the nineties, it is almost always possible to get in contact with help services very fast whenever and wherever there has been an accident. However, it is not always the case that the caller can inform about location, due to shock or other reasons. In fact there is indication [31] that as much as 25% of the mobile phone users do not know their position in an emergency situation and hence there is a need to automatically position mobile calls to improve security.

There are of course a lot of other reasons to develop a positioning system. We have organised the major driving forces behind a positioning device for the mobile cellular network into three groups: legal aspects, commercial possibilities and system improvements.

3.1.1 Legal Aspects

Legal aspects have been one of the main driving forces behind the positioning standardisation. In the USA the Federal Communications Commission (FCC) have with help from other organisations representing different emergency centres, including fire brigades and hospitals, formulated the requirements in the standard E-911. In the first phase of the standard the cellular systems must be able to transmit the ID-number and the cell number of the caller to the emergency service. This is not yet fully working in all states. In phase two, with deadline October 2001, the systems should give an estimate of the mobile position (longitude and latitude) with an accuracy of 125 metres in 67% of the cases. No such methods are implemented today, but that has not stopped FCC to ask for comments on tightened accuracy requirements and estimates of vertical position.

3.1.2 System Improvement

There is also a large gain for the system providers and system operators in knowing the positions of the mobiles. If a model of the cell structure is available, bad communication conditions can be predicted and a hand-over could be done in good time without loss of information. The number of hand-overs can also be reduced which saves valuable time.

It will be easier to construct algorithms that maximises the number of users in the system when the positions are known. We can for example squeeze in another user in a full cell by observing that a current user in the cell is moving into a neighbouring cell, giving the channel to the new user and providing a new channel in the neighbouring cell to the current user. Another possibility is to assign channels with different reuse distance to decrease the reuse distance between channels. A mobile close to the BTS will be assigned a channel with a small reuse distance and a mobile close to the border of the cell will be assigned a channel with high reuse distance.

Once position can be predicted with high enough accuracy, the mobiles can be used as data collectors for the system providers. This information should be valuable for deciding when and where new investments in the system are of interest.

3.1.3 Commercial Possibilities

Even if safety is the primary motivation for mobile phone tracking there are also a lot of possibilities for applications of more commercial kind. Some examples of them are listed below.

1. Personal tracking.
2. Navigation assistance.
3. Position dependent billing.
4. Mobile yellow pages.
5. Position dependent advertising.
6. Time table information.

7. Many more.

Considering the possibilities with positioning and mobile Internet the system operators will surely have an interesting future with many new business opportunities.

3.2 Positioning Methods

The different positioning methods can be divided into two groups: network based solutions and terminal based solutions, depending on if the position estimate computations take place in the fixed BTS network or in the mobile unit. The BTS network can offer more computer power, but a mobile unit would increase personal identity security and decrease the network load. We will not consider this question in this short survey over different positioning methods. For a comprehensive overview of different location estimate methods see [31].

A lot of simulations and measurements have been done to compare the different positioning methods, but they often give contradictory results. This is expected since the signal propagation is sensitive to the surrounding environment. However, excluding GPS (Global Positioning System), the methods based on signal propagation time are today the most reliable methods.

3.2.1 Angle of Arrival (AOA)

By measuring the angle of arrival of the transmitted signal at two or more different BTS, it is possible to find the position of the mobile by triangulation. Antenna array BTS is especially suitable for calculating the angle of arrival; see [25] for an overview of antenna array techniques. The AOA method could be useful in the future since antenna arrays are planned for future mobile networks, where their primary task is to provide directional transmission to improve the capacity of the network. The literature gives contradictory results about performance of the AOA method. For example, Owen *et al.* [29] gives negative results for urban location estimation but a company in the US claim on their homepage that their solution fulfils the requirements of FCC, see [34].

3.2.2 GPS

The GPS system is commonly used for navigation purposes today, but the GSM system can give additional information to the GPS receivers to obtain better coverage and accuracy. These methods are very accurate and positioning can be done within metres precision depending on the type of GPS and the surrounding environment. The interest has yet been quite low in this kind of solution from the mobile community since the GPS receivers are relative expensive, require line of sight (LOS) conditions and can not penetrate buildings well enough for indoor positioning. The power consumption in the mobile will also increase which requires more powerful batteries. This might be a problem since there is a trend of almost ever decreasing sizes of the mobile phones.

3.2.3 Time of Arrival (TOA) and Time Difference of Arrival (TDOA)

The propagation time from mobile to BTS (or vice versa) is the most commonly used variable for position estimation today. In the TOA method the propagation time is calculated by letting the mobile bounce the signal back to the BTS. This duplex signalling is not efficient, and is the major drawback with the TOA method. In [13, 14] different algorithms for TOA position location is compared in a simulation.

The most popular candidate for future position systems seems to be TDOA. The method measures the relative arrival time from one mobile at three BTS at the same time (or vice versa). This requires exact synchronisation between the BTS. The position estimate will be given by the intersection of two hyperboloids and the solution to the equation system thus have to be found by some kind of iterative method.

3.2.4 Signal Strength Analysis

Triangulation using received signal strength requires either a very good fading model or empirical data of the signal strength sampled from all interesting positions in the area. Both approaches are difficult, but there are benefits using a signal strength analysis compared with other methods; the signal strength information can be very useful for the system operator. Empirical models have been derived by Hata in [16], but these models are of global character and does not describe the local fading in a urban area well enough for position location purposes. However, much effort are put into making good signal models using methods like ray-tracing, geographical information systems and diffraction models, which might make this approach more attractive in the future. Since the signal model is only calculated once it is not a time critical step in the algorithms.

In [18, 19] signal strength analysis is used for mobile tracking. Highly accurate results are reported, but in Chapter 6 we find that the choice of signal strength decay model is crucial for their results and in more complex situations the performance is much worse.

3.3 Problems in Location Estimating

Location estimating using the mobile network is very convenient since it takes advantage of the existing cellular network structure and only requires the signal as input. Unfortunately it also inherits the disadvantages imposed by the design of the network.

In most of the techniques presented above two or more non-serving BTS are involved in the location procedure which might cause problems due to low received signal strength at a remote BTS; especially when the mobile is close to the serving BTS and thus might automatically be forced to transceive at a lower power level. This is called the hearability problem. The fast decay of the signal power is due to the non line of sight (NLOS) conditions and is something that is highly appreciated for interference reduction making the channel reuse distance smaller, but it makes triangulation harder.

Another important problem with NLOS conditions is that signals in urban areas tend to propagate along the streets and timing, signal strength and angle of arrival will differ from what can be expected from a LOS model. The received signal also consists of several copies of the signal but with different time delays, magnitude and phase. This phenomenon, when the signal interferes with itself, is called multi-path propagation and is important to include in a signal propagation model. TDOA reduces this problem since it measures the time of arrival difference between different BTS and the errors are then hoped to cancel, but for more accurate results this effect must be included in the models. For the signal strength this will result in fast fading, which can be described as the rapid changes of the signal strength due to interference. The signal strength also heavily depends on how the user directs the antenna and the very local surroundings. These two phenomena reduces the applicability of the signal strength analysis.

The geographical positions of the BTS can also be important, especially in rural areas or along highways where the BTS tend to be aligned which will reduce the accuracy of the triangulation.

3.4 A Tracking Algorithm

In this section we introduce a position tracking algorithm proposed recently by Hellebrandt and Mathar [19] as an example of an algorithm for mobile positioning. This method will also be used in the simulations in Chapter 6.

3.4.1 Model and Algorithm

The idea is to find the least squares estimate of the position using the received signal strength from a number of neighbouring BTS and use it as an observation input to a tracking filter. Of course, the least squares problems will be difficult to solve since the complex signal landscape will produce a lot of local minima, but if a good initial estimate is given some faith can be put into the next step estimate and the least squares minimisation restricted to a smaller area. This approach is investigated in [18] and [19], where the latter reference uses a Kalman filter and the former uses linear regression. The Kalman filter proves to be the more successful approach according to the authors [19].

The method is not restricted to tracking based on signal strength but can be used for any of the proposed positioning methods in Chapter 3. We introduce some notation to have a closer look at the prediction algorithm.

Let n be the number of BTS in the area of interest $A \subset \mathbb{R}^2$ which is bounded, and let $s_k(i)$ be the signal power received at BTS $i = 1, \dots, n$ at discrete time instants $k \in \mathbb{N}$. Further, we assume that we know the average signal strength

$r_k(i, \mathbf{z})$ measured at time k at BTS i from a mobile in position $\mathbf{z} \in A$. The least squares estimate[1] is then given by

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z} \in A} \sum_{i=1}^{n} (r_k(i, \mathbf{z}) - s_k(i))^2 \tag{3.1}$$

which gives an estimate of the position $\mathbf{z}$ at time t_k. The least squares problem (3.1) is difficult to solve since a large number of local minima are expected. But since the complexity of the signal landscape gives us no hope in modelling the signal strength decay with high accuracy it is useful to simplify the least squares problem. The minimisation problem could be reduced to an exhaustive search only over a local area grid $G \subset A$, given that the initial estimate is reliable. The size of the local area is determined by an upper limit of the velocity of the mobile in the area times the sampling time and the size of the grid is determined by the accuracy requirements and the signal model. We will come back to this problem in Chapter 6.

The Kalman filter in [19] consists of a linear dynamical system given by

$$\mathbf{X}_{k+1} = \Phi \mathbf{X}_k + \Gamma \mathbf{W}_k, \tag{3.2}$$

where

$$\mathbf{X}_k = \begin{pmatrix} X_k^1 \\ \dot{X}_k^1 \\ X_k^2 \\ \dot{X}_k^2 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 0 & 0 \\ \Delta & 0 \\ 0 & 0 \\ 0 & \Delta \end{pmatrix}, \quad \Phi = \begin{pmatrix} 1 & h & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & h \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.3}$$

and $\mathbf{W}_k$ are 2×1 random vectors assumed to be i.i.d. $\mathcal{N}(\mathbf{0}, \mathbf{C})$. An observation model is also presented to complete the Kalman filter. The observations from the least squares estimates (3.1) are modelled by another linear system,

$$\mathbf{Y}_k = \mathbf{M}\mathbf{X}_k + \mathbf{V}_k, \tag{3.4}$$

where

$$\mathbf{Y}_k = \begin{pmatrix} Y_k^1 \\ Y_k^2 \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{3.5}$$

and $\mathbf{V}_k$ are 2×1 random vectors again assumed to be independent normally distributed $\mathcal{N}(\mathbf{0}, \mathbf{R})$.

The movement of the mobile is thus modelled as white Gaussian noise acceleration and the observations are the true position in additive white Gaussian noise.

[1] Under assumptions of Gaussian distributions with equal variances this is the maximum *a posteriori* estimate.

3.4.2 Remarks

The least-squares approach followed by a Kalman filter has several good qualities. It is fast and possible to implement for real time usage, because of the local search and the recursive structure of the Kalman filter. It is also possible to use theoretical signal propagation models as well as signal strength measurements as data in the algorithm.

Two major disadvantages with the algorithm is that it is very dependent on a good initial position estimate and that it is restricted to linear models with additive Gaussian noise, see Chapter 4. This will cause problems as we will see later on.

Chapter 4

State Estimation Theory

In this chapter we will review some basic results from filter theory. The general filtering problem is first introduced and then the necessary results for an analytical solution is presented. The well known Kalman filter solution of the linear filtering problem with additive Gaussian noise is also reviewed as an important special case of the general filtering problem.

For a more detailed study with proofs we refer to the good introductory books on filtering by Jazwinsky [22], McGarty [26] and to Durrett [12] for an introduction to conditional means and martingales, which are important concepts in filtering.

4.1 The General Filtering Problem

The ultimate goal of practically all filtering and tracking algorithms is to determine the probability density function (PDF) of the objects state vector, given the observations up to the present time point[1]. Since this PDF contains all available statistical information it is the complete solution to the filtering problem.

To make the filtering procedure effective it is important to have a good model of the system. Two main restrictions must then be met by the model: first the model must give an acceptable description of the physical situation, second it must be mathematically tractable. The latter demand has been taken care of in the filtering problem by a Markovian approach. The PDF for the next state of the system is thus only dependent on the present distribution, and does not depend on its history. The Markovian framework is general enough to model many interesting physical processes and will be used exclusively in this thesis.

We first define the continuous version of the general filtering problem. Suppose the state vector $X_t \in \mathbb{R}^n$ at time t of a dynamical system is given by the

[1]The problems of smoothing and deconvolution will not be considered here, but they fit into the same framework.

stochastic differential equation

$$dX_t = f(X_t, t)dt + \sigma_X(X_t, t)du_t, \tag{4.1}$$

where $f : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$, $\sigma_X : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^{n\times p}$ and $(u_t)_{t\geq 0}$ is a p-dimensional stochastic process with independent increments. Usually further restrictions are put on the functions f and σ_X to assure existence and uniqueness of a solution to the stochastic differential equation, but we refer to [28].

Also the observations Y_t are continuous and given by

$$dY_t = h(X_t, t)dt + \sigma_Y(X_t, t)dv_t, \tag{4.2}$$

where $h : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^m$, $\sigma_Y : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^{m\times r}$ and $(v_t)_{t\geq 0}$ is a r-dimensional independent increment process, also independent of $(u_t)_{t\geq 0}$. Once again we refer to [28] for constraints on h and σ_Y for existence and uniqueness.

The filtering problem is now to find the best estimate $\hat{X}_t$ of the state vector X_t given the observations $\{Y_s,\ 0 \leq s \leq t\}$ satisfying (4.2). In other words: Find $\hat{X}_t$ that is measurable w.r.t. $\mathcal{Y}_t = \sigma(\{Y_s : 0 \leq s \leq t\})$ such that

$$E[|X_t - \hat{X}_t|^2] = \inf\{E[|X_t - Z|^2] : Z \in L^2(P),\ Z \text{ is } \mathcal{Y}_t \text{ measurable}\}, \tag{4.3}$$

where P is the measure of the probability space $(\Omega, \mathcal{A}, P)$.

In some situations a sampling procedure is imposed and then a discrete model is more natural[2],

$$X_{k+1} = f(X_k, k) + \sigma_X(X(k), k)u_k, \tag{4.4}$$

where X_k is a $n \times 1$ state vector at time t_k, $(n_k)_{k\geq 0}$ is a white noise, zero mean sequence with known distribution and independent of current and past states and $f_k : \mathbb{R}^n \times \mathbb{N} \to \mathbb{R}^n$ is the state transition function. The observations Y_k of the system are given by

$$Y_k = h(X_k, k) + \sigma_Y(X_k, k)v_k, \tag{4.5}$$

where Y_k is a $m \times 1$ vector; $(v_k)_{k\geq 0}$ is a zero mean, white noise sequence with known distribution, independent of current and past states and also independent of the system noise $(n_k)_{k\geq 0}$; $h : \mathbb{R}^n \times \mathbb{N} \to \mathbb{R}^m$ is the observation function. This situation will be our main focus in the rest of the thesis. The filtering problem is the same as before, but with the exception that $\mathcal{Y}_k = \sigma(Y_1, \ldots, Y_k)$ in condition (4.3). If the discrete model arrives from a discretisation of the continuous version, some care has to be taken in the discretisation stage.

The general filtering problem could of course be made even more general since we have assumed additive noise in our models. We would then have arrived at the following system

$$\begin{cases} X_{k+1} &= f(X_k, k, n_k) \\ Y_k &= h(X_k, k, v_k) \end{cases}$$

[2] All procedures implemented in a computer certainly must be discrete since a computer operates sequentially in time and memory is finite.

We will not consider that generalisation here since it makes it hard to develop general analytical results, but we come back to this kind of generalisations when we study approximative methods in Chapter 5.

4.2 Optimal Estimates

Consider random variables defined on the probability space $(\Omega, \mathcal{A}_0, P)$ and in particular those belonging to the Hilbert space $L^2(\mathcal{A}_0) = \{Y \in \mathcal{A}_0 : EY^2 < \infty\}$, equipped with the scalar product $E(XY)$, where E is the expectation operator and $X, Y \in L^2(\mathcal{A}_0)$. Let $\mathcal{A}$ be a sub σ-algebra such that $\mathcal{A} \subset \mathcal{A}_0$ and hence, $L^2(\mathcal{A})$ will be a closed subspace of $L^2(\mathcal{A}_0)$. Hilbert space theory then gives existence and uniqueness of a pair of orthogonal projections $\mathcal{P}$ and $\mathcal{Q}$,

$$\begin{aligned}\mathcal{P} &: L^2(\mathcal{A}_0) \to L^2(\mathcal{A}),\\ \mathcal{Q} &: L^2(\mathcal{A}_0) \to L^2(\mathcal{A})^{\perp},\end{aligned}$$

such that any $X \in L^2(\mathcal{A}_0)$ can be decomposed as

$$X = \mathcal{P}(X) + \mathcal{Q}(X). \tag{4.6}$$

The orthogonal projection $\mathcal{P}(X)$ of X on $\mathcal{A}$ is the stochastic variable $Y \in \mathcal{A}$ that minimises the error norm $||X - Y||^2 = E(X - Y)^2$ and is in this sense the optimal estimator of X in $\mathcal{A}$. The following theorem determines the orthogonal projection operator as the conditional mean of X given $\mathcal{A}$, see for example [12].

Theorem 1 *Let $X \in L^2(\mathcal{A}_0)$, then the conditional mean $E(X|\mathcal{A})$ is the stochastic variable $Y \in L^2(\mathcal{A})$ that minimises $E(X - Y)^2$.*

Different kind of subspaces could be of interest for projection, but generally we want them to be spanned by a set of functionals $G : \mathbb{R}^n \to \mathbb{R}$ of the stochastic observation functions $Y_1, \ldots, Y_n$ that are measurable on $\mathcal{A}$ and square integrable. We give two examples of natural subspaces to be more concrete.

Example 1 *Let $\mathcal{N}$ be the Hilbert space of all square integrable, measurable functions of the observations $Y(s_1, \omega), \ldots, Y(s_n, \omega)$, $s_i \in [t_0, t]$, $i = 1, \ldots n$ or limits of such. $\mathcal{N}$ equipped with the usual inner product is a closed subspace of $L^2(\mathcal{A}_0)$ and a unique orthogonal projection $\mathcal{P} : L^2(\mathcal{A}_0) \to \mathcal{N}$ exists.*

Example 2 *Let $\mathcal{L}$ be the Hilbert spaces of all random variables that are finite, linear combinations of $Y(s_i, \omega)$, $s_i \in [t_0, t]$, $i = 1, \ldots n$ or limits of such. Again a unique orthogonal projection $\mathcal{P} : L^2(\mathcal{A}_0) \to \mathcal{L}$ exists.*

Clearly we have that $\mathcal{L} \subset \mathcal{N} \subset L^2(\mathcal{A}_0)$ and hence the projection on $\mathcal{N}$ will result in a smaller error norm than the projection on $\mathcal{L}$. On the other hand, in Section 4.4 we will review an analytical solution to the filtering problem on $\mathcal{L}$ while there in general exists no analytical solution on $\mathcal{N}$.

If continuous measurements are available some care has to be taken when studying the conditional mean. If the observation process $Y(s, \omega)$ is assumed to be a separable process it is sufficient to study the process on an countable number of points rather than on the uncountable number of points in the interval. Let $\mathcal{C}_i$ be the σ-algebra generated by the random variables $Y(t_1, \omega), \dots, Y(t_i, \omega)$ and observe that if $f : \mathcal{A}_0 \to \mathbb{R}$ then $E(f|\mathcal{C}_i)$ is a martingale. The following martingale convergence theorem is then available.

Theorem 2 *Suppose $X \in L^1$ and $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \dots$ be an increasing sequence of σ-algebras and let $\mathcal{C}_\infty$ be the smallest σ-algebra such that $\bigcup_{n=1}^\infty \mathcal{C}_n \subset \mathcal{C}_\infty$. Then*

$$\lim_{n\to\infty} E(X|\mathcal{C}_n) = E(X|\mathcal{C}_\infty) \quad \textit{a.s. and in } L^1.$$

The conditional mean, and thus the orthogonal projection in the case of continuous measurements of a separable process, is thus well defined as a limiting procedure of the case of finite observations.

4.3 Propagating Probability Densities

In the previous section we concluded that the conditional mean $E(X(t)|\mathcal{A})$ was the optimal estimate (minimises the mean square error $E(X-Y)^2$) and since

$$E(X(t)|\mathcal{A}) = \int x p_X(x, t|\mathcal{A}) dx,$$

the main concern for this section is to determine the conditional probability density $p_X(x, t|\mathcal{A})$ of the process X at every time t.

There is both a differential and integral approach to propagation of the probability density. The differential approach leads to the Fokker-Planck equations, which is a parabolic partial differential equation. In this section we will concentrate on the integral approach since it is a natural starting point for the development of the type of approximative methods that will be the main subject in Chapter 5. Since we consider applications where a sampling procedure is natural we also restrict the presentation to discrete time which makes the problem easier.

Using Bayes' theorem it is possible to develop a recursive procedure for propagating the conditional density. This process proceeds through repeated applications of prediction and correction. The prediction step uses the dynamical system (4.4) to predict the next state $\hat{X}_{k+1}$ using the best estimate $\hat{X}_k$ at time k. The update stage uses the latest information Y_{k+1} satisfying (4.5) to modify the prediction step. The procedure is called recursive Bayesian estimation.

Let $p(x_k|y_\mathbf{k})$ denote[3] the probability density function for the stochastic state variable X at time k given the finite sequence of observations $Y_\mathbf{k}$, where the boldfaced $\mathbf{k}$ indicates the sequence $Y_\mathbf{k} = (Y_i)_{i=1}^k$. We further assume that $(X_k)_{k\geq 0}$ is Markovian, that we have known parametric forms of the densities $p(y_k|x_k)$, $p(x_{k+1}|x_\mathbf{k}) = p(x_{k+1}|x_k)$ and the initial state density $p(x_0)$. The filtering can

[3]We wright $p(x_k|y_\mathbf{k})$ instead of $p_{X_k|Y_\mathbf{k}}(x_k|y_\mathbf{k})$ for notational convenience.

now be viewed as a two-stage recursive procedure. First the current density at time k is propagated one time step using only observations available at time k,

$$p(x_{k+1}|y_{\mathbf{k}}) = \int p(x_{k+1}|x_k)dP(x_k|y_{\mathbf{k}}). \tag{4.7}$$

Second, we make an observation Y_{k+1} of the system at time $k+1$ which can be included using Bayes' theorem

$$p(x_{k+1}|y_{\mathbf{k+1}}) = \frac{p(y_{k+1}|x_{k+1})p(x_{k+1}|y_{\mathbf{k}})}{p(y_{k+1}|y_{\mathbf{k}})}, \tag{4.8}$$

where

$$p(y_{k+1}|y_{\mathbf{k}}) = \int p(y_{k+1}|x_{k+1})dP(x_{k+1}|y_{\mathbf{k}}). \tag{4.9}$$

Unfortunately these integrals can seldom be evaluated analytically and hence we are restricted to numerical approximations, which will be the main subject in the next chapter. But there is an important special case when it is possible to find an analytical solution to the filtering problem which is presented in the following section.

It is important to note that if X_k is a finite set of known discrete points the problems with the integrals disappear and we get

$$p(x_{k+1}|y_{\mathbf{k}}) = \sum_{x_k} p(x_{k+1}|x_k)p(x_k|y_{\mathbf{k}}), \tag{4.10}$$

and

$$p(x_{k+1}|y_{\mathbf{k+1}}) = \frac{p(y_{k+1}|x_{k+1})p(x_{k+1}|y_{\mathbf{k}})}{\sum_{x_{k+1}} p(y_{k+1}|x_{k+1})p(x_{k+1}|y_{\mathbf{k}})}. \tag{4.11}$$

In the first step all the points are weighted by the transition probability and in the second step each step is weighted by the likelihood.

4.4 Discrete Time Kalman Filter

In the special case with linear model and Gaussian distributions it is possible to derive an analytical solution to the filtering problem. In his paper [23] Kalman uses the concept of orthogonal projections to derive the minimum mean square linear filter for the state of a dynamical system. We will follow this work here.

Consider again the linear, discrete time, dynamic model

$$\begin{cases} X_{k+1} &= \Phi(k+1|k)X_k + \Gamma_k u_k \\ Y_k &= M_k X_k + v_k \end{cases} \tag{4.12}$$

where X_k is a n-vector, Y_k is a p-vector and $(u_k)_{k\geq 0}$, $(v_k)_{k\geq 0}$ are independent sequences of Gaussian stochastic variables of n- respectively p-vectors. Further,

M_k and Γ_k are $p \times n$ respectively $n \times n$ matrices with real-valued, non-random elements. The state transition function $\Phi(k|l)$ is a real valued, $n \times n$ matrix that describes the deterministic transition of the state from time l to time k, where $k > l$.

Let $\mathcal{H}$ be the space of random n-vectors where the components have finite second order moment and let $\mathcal{H}$ be equipped with a scalar product,

$$\langle u, v \rangle = E(u^T v),$$

and norm

$$||u||^2 = \langle u, u \rangle.$$

Define the linear subspace $\mathcal{Y}_k \subset \mathcal{H}$ by

$$\mathcal{Y}_k = \{z : z = \sum_{i=1}^{k} A(i) Y_i\}, \quad \forall A(i) \in \mathbb{R}^{p \times n}. \tag{4.13}$$

The filtering problem is now to find an estimate $\hat{X}_k \in \mathcal{Y}_k$ of X_k that minimises the expected loss $E(X_k - \hat{X}_k)^2$.

Since $\mathcal{H}$ is a Hilbert space, the results from Section 4.2 applies. But we will not use the concept of conditional PDF; we follow Kalman and work with orthogonal projections on the linear subspace $\mathcal{Y}_k$. We know from Section 4.2 that the orthogonal projection of X_k on $\mathcal{Y}_k$ is the optimal estimate when we use a quadratic loss function. Thus, the optimal estimate is the same as in the Gaussian case even if we are dealing with non Gaussian noise. This is important to note since it is then possible to generalise the model (4.12). The assumption of Gaussian noise in (4.12) is convenient since the conditional mean will be normal for all times due to the linear models. The filter thus only has to propagate the mean and the variance of the distribution which makes it a finite dimensional problem.

The next step is to express the orthogonal projection as a recursive equation and in terms of our dynamical system. Let the best estimate of X_k given the information at time l be denoted $\hat{X}_k^l$ and define $\bar{X}_k^l$ by

$$X_k = \hat{X}_k^l + \bar{X}_k^l,$$

where $\hat{X}_k^l \in \mathcal{Y}_l$ and $\bar{X}_k^l \in \mathcal{Y}_l^{\perp}$. This decomposition exists and is unique according to (4.6).

The predicted estimate of X_{k+1} given the information at time k is presented by the following lemma.

Lemma 1 *The predicted optimal estimate of X_{k+1}, given the data at time k, is*

$$\hat{X}_{k+1}^k = \Phi(k+1|k) \hat{X}_k^k \tag{4.14}$$

This follows since $\Phi(k+1|k)\hat{X}_k^k \in \mathcal{Y}_k$ by definition (4.13) and $E((\hat{X}_{k+1}^{k+1} - \Phi(k+1|k)\hat{X}_k^k)^T Y) = 0$, $\forall Y \in \mathcal{Y}_k$.

Using the same kind of orthogonality arguments as above it can be shown that when the observation Y_{k+1} becomes available the orthogonal projection on $\mathcal{Y}_{k+1}$ is given by (4.14) and a linear correction term. The discrete time Kalman filter can then be summarised as:

Theorem 2 *The minimum variance estimate for the discrete system (4.12) can be stated as a difference equation for the conditional mean and the error covariance. Between observations we have*

$$\begin{aligned} \hat{X}_{k+1}^k &= \Phi(k+1|k)\hat{X}_k^k, \\ P_{k+1}^k &= \Phi(k+1|k)P_k^k\Phi^T(k+1|k) + \Gamma_k Q_{k+1}\Gamma_k^T, \end{aligned}$$

and at observations

$$\begin{aligned} \hat{X}_k^k &= \hat{X}_k^{k-1} + K_k(Y_k - M_k\hat{X}_k^{k-1}), \\ P_k^k &= P_k^{k-1} - K_k M_k P_k^{k-1}, \end{aligned}$$

where

$$K_k = P_k^{k-1} M_k^T (M_k P_k^{k-1} M_k^T + R_k)^{-1}$$

is the Kalman gain.

Chapter 5

Approximative Solution - Particle Filters

In Chapter 4 we studied how to obtain the conditional density as a function of time and a recursive, but infinite dimensional, solution was obtained. Practical filter implementation in a computer calls for methods that are finite dimensional and hence there is a need for approximative methods. In this section we will study particle filters as a solution to this problem. These are Monte Carlo methods and give a non-parametric approximation to the conditional distribution.

5.1 Overview of Different Approximative Solutions

As already mentioned there is in general no analytical solution available for nonlinear estimation. The extended Kalman filter (EKF) is probably the most frequently used filter under these circumstances, see for example [27]. In the EKF the system of equations is linearised around the the most recently predicted state and an ordinary Kalman filter can be applied. Unfortunately, for many systems the EKF will often diverge which has led to a number of modifications of the EKF based on improved linearisation techniques or coordinate transformations.

Several other methods have also been proposed in the literature. Numerical integration over a fixed grid of points in the state space has been proposed in [24]. The Gaussian sum filter is investigated in [2, 20] and methods which approximates the first two moments of the density is proposed in [33].

These methods are usually far from optimal in high noise environments, when the observations are highly uncertain or the nonlinear equations are not smooth enough. Furthermore, in [7] it is concluded that none of the methods is general applicable on a large enough class of problems, in the perspective of applications, without much work with tuning of parameters to take account of the problem specific features. Also, the updating stage often requires a formidable computational overhead which can be crucial in real time applications.

Recently there has been much work concerned with methods in which the required probability density is approximated with a set of random samples that propagate through the state space. These methods have various names depending on the area of application, such as: Bayesian bootstrap filters, condensation algorithm, Monte Carlo importance re-sampling filters and Monte Carlo particle filters. In this thesis we will simply call them particle filters. The particle filters have been found efficient for nonlinear filtering problems and do not require the same amount of parameter tuning as the previous approximative methods.

The particle filter has been suggested by various authors independently. One of the earliest and most well known methods is the SIR (sampling importance re-sampling) filter presented by Gordon, Salmond and Smith in [15] and by Del Moral, Rigal, Noyer and Salut in [11] independently.

5.2 Basic SIR Particle Filter Algorithm

The basic SIR point filter algorithm is intuitive and easy to understand even if the theoretical justifications can be quite involved, see [10, 9] which seems to be the first rigoroous convergence result. Consider again the system

$$X_{k+1} = f(X_k, k) + n_k, \tag{5.1}$$
$$Y_k = h(X_k, k) + v_k, \tag{5.2}$$

where X_k is the state vector of the dynamical system and Y_k is the observation at time $k \in \mathbb{N}$. Without loss of generalisation the state vector is assumed to be two dimensional in the figures to make visualisation easier.

Assume that that the conditional PDF $p(x_k|y_\mathbf{k})$ is known and we want to approximate $p(x_{k+1}|y_\mathbf{k+1})$. The basic idea in the SIR filter is to provide a Monte Carlo approximation of $p(x_k|y_\mathbf{k})$ called a random measure and then to propagate and update it using the Bayesian technique described in Section 4.3 to obtain an approximation of $p(x_{k+1}|y_\mathbf{k+1})$.

A random measure $(X^e(i), q(i))_{i=1}^N$ is defined by a set of random variables $(X^e(i))_{i=1}^N$ called the support and weights $(q(i))_{i=1}^N$ which sums to 1. The random variables $(X_k^e(i))_{i=1}^N$ will also be called particles or exploring particles. Following the basic idea we approximate $p(x_k|y_\mathbf{k})$ with a random measure where the support are i.i.d. random variables $(X_k^e(i))_{i=1}^N$, $X_k^e(i) \sim p(x_k|y_\mathbf{k})$ and $q_k(i) = \frac{1}{N}$. The basic SIR particle filter algorithm can now be formulated:

1. **Initialisation**. The particles $X_0^e(1) = x_0^e(1), \ldots, X_0^e(N) = x_0^e(N)$ are sampled from $p(x_0)$ and are all given uniform weight $q_0(i) = \frac{1}{N}$.
2. **Prediction**. Assume the random measure $(X_k^e(i), q_k(i) = \frac{1}{N})_{i=1}^N$ at time k is known (see Figure 5.1(a)). Propagate the particles according to the given dynamical system (see Figure 5.1(b)):

$$\tilde{X}_{k+1}^e(i) = f(X_k^e(i), k) + n_k.$$

The PDF $p(x_{k+1}|y_\mathbf{k})$ is then approximated with the random measure $(\tilde{X}^e_{k+1}(i), q_k(i) = \frac{1}{N})_{i=1}^N$ (see Figure 5.1(c)).

3. **Correction.** The observation at time $k+1$ becomes available and the weights are updated according to (see Figure 5.1(d)):

$$\tilde{q}_{k+1}(i) = \frac{p(y_{k+1}|\tilde{x}^e_{k+1}(i))}{\sum_{j=1}^N p(y_{k+1}|\tilde{x}^e_{k+1}(j))}. \tag{5.3}$$

4. **Estimation.** The conditional mean is approximated by the sum of the weighted particles,

$$E(X_{k+1}|Y_1 = y_1, \dots, Y_{k+1} = y_{k+1}) \approx \sum_{j=1}^N \tilde{q}_{k+1}(j)\tilde{x}^e_{k+1}(j). \tag{5.4}$$

5. **Re-sampling.** Re-sample N times from the discrete probability distribution given by $(\tilde{X}^e_{k+1}(i), \tilde{q}_{k+1}(i))_{i=1}^N$ gives the random measure $(X^e_{k+1}(i), q_{k+1}(i) = \frac{1}{N})_{i=1}^N$.

6. **Loop.** Return to step 2.

The initialisation- and the prediction-step only require sampling and not evaluation from the distributions $p(x_0)$ and $p(x_{k+1}|x_k)$, which makes it possible to generalise (5.1) to

$$X_{k+1} = f(X_k, k, n_k). \tag{5.5}$$

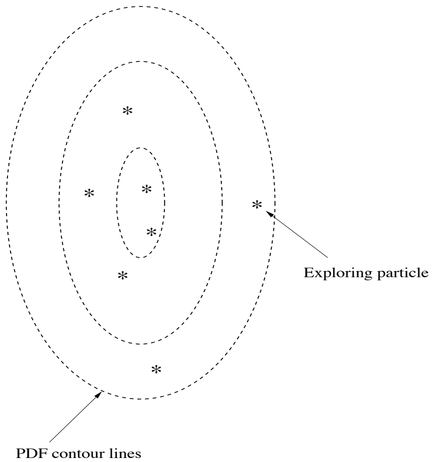
The correction step requires that $p(y_k|x_k)$ is evaluable for every y_k and x_k. Note that the position estimation is done before the re-sampling stage not to introduce an unnecessary source of noise, see [7]. As $N \to \infty$ the approximation in (5.4) converges in probability to the conditional expectation, which is proved in [10].

5.3 Particle Filters

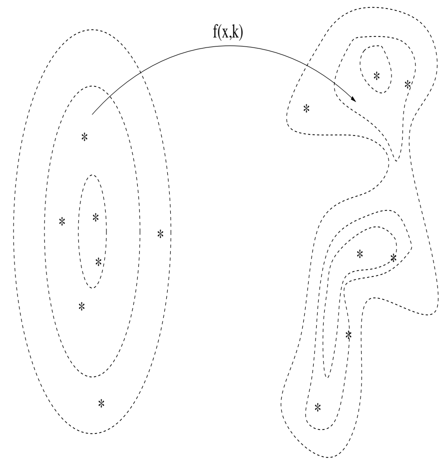
In this section we describe the particle filter algorithm in a more general way, which will result in a number of different particle filters.

Particle filters can be defined as the class of Monte Carlo filters which recursively approximate the distribution of the random variable $X_k|Y_\mathbf{k}$ by the random measure $(X^e_k(i), q_k(i))_{i=1}^N$, that is

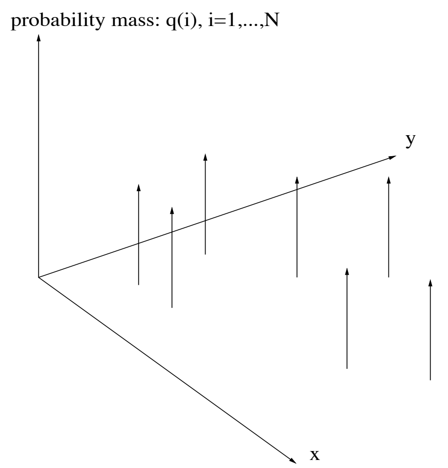
$$P(X_k \in dx_k|Y_\mathbf{k}) \approx \sum_{i=1}^N \delta_{X^e_k(i)} q_k(i) \tag{5.6}$$



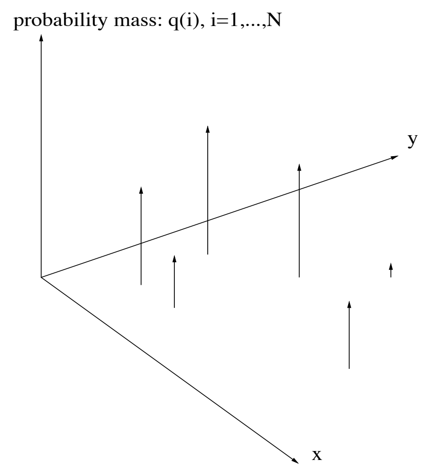
(a) Assume that $p(x_k|y_{\mathbf{k}})$ is known and let the exploring particles $X_k^e(i)$ (marked * in the figure) be a random sample from this distribution.



(b) The particles are propagated one step by the dynamical system (5.1).



(c) The PDF $p(x_{k+1}|y_{\mathbf{k}})$ is approximated by the random measure $(\tilde{X}_k^e(i), q_k(i) = \frac{1}{N})_{i=1}^N$



(d) The observation $Y_{k+1} = y_{k+1}$ at time $k+1$ is used to re-weight the particles using the likelihood (5.3).

Figure 5.1: *One step in the SIR filtering algorithm. The number of particles in the figures is small and not representative for a real filtering situation.*

where $\delta_X(\cdot)$ is the point measure with support in X. This is an approximation in the sense that the random variables $X_k^e(i)$ with associated weights $q_k(i)$ are chosen such that

$$\int g(x_k) \sum_{i=1}^{N} \delta_{X_k^e(i)}(dx_k) q_k(i) = \sum_{i=1}^{N} g(X_k^e(i)) q_k(i) \longrightarrow E(|g(X_k)||Y_{\mathbf{k}}), \tag{5.7}$$

in probability as $N \to \infty$, for measurable functions g where $E(|g(X_k)||Y_{\mathbf{k}}) < \infty$ and $E(|g(X_k)|^2|Y_{\mathbf{k}}) \leq C < \infty$.

The recursion starts with the Bayesian prediction of the conditional distribution described in (4.7)

$$p(x_{k+1}|y_{\mathbf{k}}) = \int p(x_{k+1}|x_k) dP(x_k|y_{\mathbf{k}}), \tag{5.8}$$

and applying the random measure gives

$$p(x_{k+1}|y_{\mathbf{k}}) \approx \hat{p}(x_{k+1}|y_{\mathbf{k}}) = \sum_{i=1}^{N} p(x_{k+1}|x_k^e(i)) q_k(i). \tag{5.9}$$

The second step in the recursion is the correction step described in (4.8)

$$p(x_{k+1}|y_{\mathbf{k+1}}) \propto p(y_{k+1}|x_{k+1}) p(x_{k+1}|y_{\mathbf{k}}), \tag{5.10}$$

and inserting (5.9) gives the following approximation of $p(x_{k+1}|y_{\mathbf{k+1}})$:

$$\hat{p}(x_{k+1}|y_{\mathbf{k+1}}) \propto p(y_{k+1}|x_{k+1}) \sum_{i=1}^{N} p(x_{k+1}|x_k^e(i)) q_k(i). \tag{5.11}$$

The particle filter now concludes the filtering step by approximating (5.11) by a random measure $(X_{k+1}^e(i), q_{k+1}(i))_{i=1}^N$. There are several ways to sample from $\hat{p}(x_{k+1}|y_{\mathbf{k+1}})$ which result in different type of point filters.

5.3.1 SIR Sampling

One of the most frequently used sampling methods is SIR, where a random measure with uniform weights $(X_k^e(i), q_k(i))_{i=1}^N$ is used to approximate $p(x_k|y_{\mathbf{k}})$. For each of the exploring particles $(X_k^e(i))_{i=1}^N$ a sample is drawn from $p(x_{k+1}|x_k^e)$ giving new particles $(\tilde{X}_{k+1}^e(i))_{i=1}^N$ with associated weights $\tilde{q}_{k+1}(i)$

$$w_{k+1}(j) = p(y_{k+1}|\tilde{x}_{k+1}^e(j)), \quad \tilde{q}_{k+1}(j) = \frac{w_{k+1}(j)}{\sum_{i=1}^N w_{k+1}(i)}, \quad j = 1, \ldots, N. \tag{5.12}$$

This new random measure is then re-sampled to obtain $(X_{k+1}^e(i), q_{k+1}(i) = \frac{1}{N})_{i=1}^N$. The re-sampling stage is important since it redistributes the particles to areas in the state space with high probability.

5.3.2 Stratified Sampling

In the basic SIR filter one particle is sampled from each one of the sub-densities (or strata) in (5.11). An improved SIR filter is proposed in [7] where stratified sampling is used for variance reduction. Suppose that a PDF consists of N strata

$$p(x) = \sum_{i=1}^{N} \beta_i p_i(x). \tag{5.13}$$

Let M_i be the number of samples from $p_i(x)$, where $N = \sum_{i=1}^{N} M_i$ is the total number of samples drawn from $p(x)$. According to sampling theory there is a most efficient way to sample from $p(x)$ in order to reduce the variance in the estimate of $\int g(x)p(x)dx$: let $M_i \propto \beta_i \sigma_i$ where σ^2 is the variance of $g(x)$ under $p_i(x)$. In most problems, σ_i is unknown and instead $M_i \propto \beta_i$ is chosen. This method can be used in the SIR filter since (5.11) is a PDF of type (5.13) with

$$\beta_i = \frac{\int p(x_{k+1}|x_k^e)p(y_{k+1}|x_{k+1})dx_{k+1}}{\sum_{i=1}^{N} \int p(x_{k+1}|x_k^e)p(y_{k+1}|x_{k+1})dx_{k+1}}, \tag{5.14}$$

and

$$p_i(x) = \frac{p(x_{k+1}|x_k^e)p(y_{k+1}|x_{k+1})}{\int p(x_{k+1}|x_k^e)p(y_{k+1}|x_{k+1})dx_{k+1}}. \tag{5.15}$$

The authors report significant improvements using stratified sampling on the bearings-only-problem. Unfortunately, β_i and $p_i(x)$ is seldom available and importance sampling is necessary. Note that choosing one sample from each strata, as in the basic SIR filter, is also a form of stratified sampling.

5.3.3 Random Sampling

An alternative sampling method is rejection sampling, which is based on simulating from $p(x_k|x_{k-1}^e)$ and accepting particles with probability

$$p = \frac{p(y_k|x_k)}{f(x)}, \tag{5.16}$$

where $f(x) \geq p(y_k|x_k)$ is the comparison function. Since we do not know $p(y_k|x_k)$ a natural choice is

$$f(x) = p(y_k|x_{max}), \quad x_{max} = \arg\max_{x_k} p(y_k|x_k). \tag{5.17}$$

One major problem with rejection sampling is that x_{max} can be hard to compute for high dimensional problems. Another alternative which also produces random samples is Markov Chain Monte Carlo (MCMC) sampling.

Both of the alternative methods can be very time consuming if there are a lot of rejections, and the author's opinion is that the SIR sampling method is better suited for real time sampling.

5.3.4 Branching Particle Filters

In this section we will describe a generalisation to the SIR particle were the particles are allowed to branch in every step. In the SIR filter every particle at some instant propagates one step through the state space and are then re-sampled according to how well they match the observation. The branching process introduces a more complex interaction between the particles to avoid exploration of uninteresting regions of the state space. The branching particle filters differ from the SIR filter at two points.

Prediction. For each of the exploring particles $(X_k^e(i))_{i=1}^N$:

1. Initialise M auxiliary particles $(X_k^{a,j}(i))_{j=1}^M$ with position $X_k^{a,j}(i) = X_k^e(i)$ and weight $q_k^{a,j}(i) = q_k(i)$ for $j = 1, \ldots, M$.
2. Propagate the auxiliary particles one step according to the dynamical system (5.1) to get $(X_{k+1}^{a,j}(i))_{j=1}^M$.

Correction. For each of the exploring particles $(X_k^e(i))_{i=1}^N$:

1. Choose one of the subsystem of particles $(X_{k+1}^{a,j}(i))_{j=1}^M$ at random with probability
$$\frac{\sum_{j=1}^M (p(y_{k+1}|x_{k+1}^{a,j}(i))}{\sum_{i=1}^N \sum_{j=1}^M p(y_{k+1}|x_{k+1}^{a,j}(i))}. \tag{5.18}$$
2. Choose one of the auxiliary particles $X_{k+1}^{a,j}(i)$, $j = 1, \ldots, M$ in the chosen subsystem with probability
$$\frac{p(y_{k+1}|x_{k+1}^{a,j}(i)}{\sum_{j=1}^M p(y_{k+1}|x_{k+1}^{a,j}(i))}. \tag{5.19}$$

A generalisation of the particle filter can be done by allowing the auxiliary particles to propagate R steps before they are given weights corresponding to how well their trajectories fits the observations.

Note that we now have introduced two more parameters, M and R, that should be set according to some optimality criteria which probably will depend on the system equations and thus makes the filter more problem specific.

5.4 Time Complexity of SIR Algorithm

Increasing the number of particles N in the filter will improve the estimates but of course also increase the computing time. In time critical applications there will always be a compromise between accuracy and time efficiency and it is thus interesting to study the time complexity of one step in the SIR particle filter algorithm.

If all function evaluations are assumed to take constant time both the propagation step and the update stage in the algorithm will have a time complexity $O(N)$ in a single processor architecture[1]. But the re-sampling stage requires some thought to be done in an efficient way. The naive approach is to simulate standard uniform variables $(u_i)_{i=1}^N$ and then use a binary search to find which particle, $(x_i)_{i=1}^N$ with corresponding weights $(q_i)_{i=1}^N$, to choose in the i:th stage. That is, choose particle with index j if

$$Q_{j-1} < u_i \leq Q_j$$

where $Q_j = \sum_{k=0}^{j} q_k$ and $Q_0 = 0$. Since the binary search procedure have time complexity $O(\log N)$ the whole sampling procedure have time complexity $O(N \log N)$. In [7] an algorithm which results in an $O(N)$ sampling is presented. Simulate $N+1$ exponentially distributed random variables $e_0, \dots, e_N$ and then calculate the empirical cumulative distribution function (CDF) $E_j = \sum_{i=0}^{j} e_i$. The two discrete CDF:s E_j and Q_j are then merged together as described in Algorithm 1.

Algorithm 1 *($O(N)$ sampling)*

```
i=0; j=0;
while j ≤ N loop
    if Q_j E_N > E_i then
        i=i+1; output x_j
    else
        j=j+1;
    end if;
end loop;
```

We can conclude that the one step in the particle filter can be done with time complexity $O(N)$. Of course, in a time critical situation as mobile tracking the true computational time is more important, but the $O(N)$ time complexity shows that the computational burden will not explode in a situation which requires a larger amount of particles.

5.5 The Number of Particles in the Filter

The possibly large amount of particles needed for a certain level of accuracy can be very critical in a real time filtering application even if the algorithm has linear time complexity. The convergence results presented so far are asymptotic results, and give no hint of how to choose the number of particles in the filter. However, it is clear that the number of particles N is dependent on a number

[1] The structure of the propagation stage in the SIR filter is well suited for parallelisation.

of factors. The dimension of the state vector is important since the components are often correlated which should require an increased number of particles to resolve the dependencies. In the case of independent components, N does not increase with the dimension of the state vector. The SIR algorithm converges for every time k, but if the convergence is not uniform in time the number of particles can be expected to increase. In [10] a regularization procedure is introduced that give uniform convergence. Finally, N will also depend on the prior $p(x_k|y_{\mathbf{k-1}})$ and the likelihood $p(y_k|x_k)$. If the variance of the likelihood is small compared to the variance of the prior, *i.e.* the region of the state space where the observations are probable are small compared to the region where the prior is significant, almost all of the exploring particles will receive a very small weight. If the likelihood falls into a region of low prior density and hence containing only a few exploring particles, these particles will have large weight and all the others very small weight and hence a small chance of being re-sampled. The approximation of the conditional distribution might then collapse and only be representated by a single value. This phenomena is often called sampling impoverishment in the particle filter literature.

We have presented particle filters using a constant number of particles, but that is of course not necessary. Intuitively, it should be wise to have a large number of particles when the variance of the conditional distribution is large and then reduce it when it becomes smaller. This can be seen from the Chebychev bound

$$P(|\frac{1}{N}\sum_{i=1}^{N} X_i - \mu| < \epsilon) \leq \frac{C}{N\epsilon^2}, \tag{5.20}$$

where X_i are uncorrelated random variables with $E(X_i) = \mu$ and $var(X_i) \leq C$. By estimating the variance of the distribution it would be possible to adaptively change the number of particles, but such estimates can not be based on the exploring particles.

In [7] a simple method for determining the filter size is proposed that is based on running the filter independently a number of times on a known trajectory and then compare the variance of the estimated conditional mean within and between the replicates. But the number of particles should of course be determined from the observations of its performance on the whole population of trajectories. However, this method is still useful since it quite often is possible to distinguish a few examples that should be hard to filter and study the performance on these.

Chapter 6

Simulation of Tracking in an Urban Area

In this chapter we will compare the performance of the Kalman filter and the particle filter in a simulated urban environment. The Kalman filter has been proposed for mobile tracking in [19], but depends heavily on linear models for optimality in the mean square sense (see Chapter 4).

A mobile moving in an urban environment will experience frequent shifts between LOS and NLOS conditions, communicating with a BTS. This will result in a highly nonlinear decay of the signal strength. The particle filter turns out to be the better algorithm under these conditions.

6.1 Simulation Model

In the simulations, signal strength is used as input to the tracking algorithms, following the approach in [19]. One can argue that a TDOA based input would have been more realistic since it has been proposed for many systems that are under development. But an advantage is that the signal strength contains useful information for the system operator and more advanced and realistic signal models are under development. However, the position tracking methods that are investigated in this thesis are of course applicable even in a TDOA approach and a possible scenario could be to use signal strength models as a complement to methods based on TOA, TDOA or AOA, (see Chapter 7).

In the simulations presented below we assume that it is possible to make a model of the signal strength decay for every BTS. The area is discretised into pixels and for every BTS and every pixel the model gives a value of the signal strength. This approach is realistic today, since the development of the database systems has increased the speed of information retrieval. However, it is not straightforward to discretise the signal strength decay model, since we must deal with the problem of fast fading and other disturbances, see Chapter 3.

6.1.1 Topology

The two dimensional topology used in the simulations is presented in Figure 6.1. This is of course a very simplified model of a city since we do not include any altitude information, but it includes the shifts between LOS and NLOS conditions, which is our main modelling concern in this simulation. We will call it a Manhattan area because of the perpendicular street crossings.

The area has been divided into uniform pixels which have been included in a small part of Figure 6.1. The uniform pixel size is not a restriction, but makes the simulations more convenient. The pixel size is set to 10×10 metres and the total area is 100×100 pixels or 1000×1000 metres.

Five BTS are present in the area and they are marked with a '*' in Figure 6.1. Note that a mobile moving in the area will experience both LOS conditions and NLOS conditions.

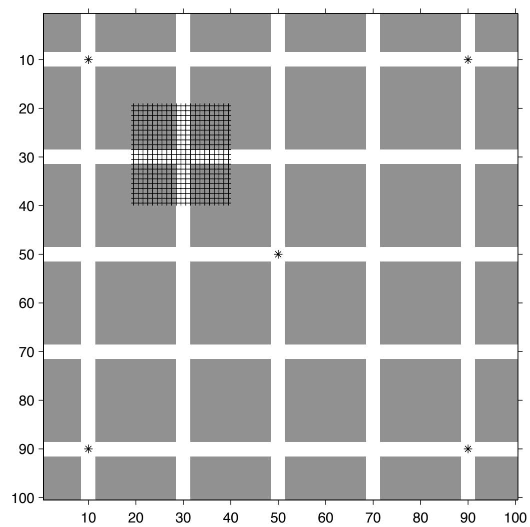


Figure 6.1: *City topology. The grey pixels represent buildings and the white pixels represent streets. The positions of the five BTS are marked with a '*' in the figure. The individual pixels are plotted in a small part of the area.*

6.1.2 Signal Model

The signal model determines the piecewise constant signal strength at each pixel, for every BTS. The model can be derived from measurements, some theoretical model or a combination of both. The theoretical models have improved a lot and will probably be relevant alternatives to measurements in the future. An overview of different models for signal propagation loss prediction is given in

[17], but here we restrict to a special class of models, since it will be enough for our purposes.

The Okumura-Hata formula [16] has been used for signal strength modelling in a number studies, see for example [18]. It is an empirical formula given by

$$P_r = P_t - G_D - G_A - G_F \quad [dB], \tag{6.1}$$

where P_r and P_t is the received respectively transmitted signal strength, G_A is the antenna gain, G_F is the fading gain which is assumed to be $\mathcal{N}(0, \sigma^2)$ distributed and G_D is the distance gain. The attenuation of the signal due to distance is modelled as

$$G_D = R_0 + 10\alpha \log(d) \quad [dB], \tag{6.2}$$

where R_0 is a constant, α is the attenuation factor and d is the Euclidean distance between BTS and mobile. For LOS with no multipath or reflections we have the well known quadratic decay of the signal strength and hence $\alpha = 2$. The Okumura-Hata model is of global character and it is really only applicable if the antenna height is higher than the surrounding rooftops and the distance d is over one kilometre [17]. Thus it will not describe the situation in a city well enough for position prediction and a more local model must be used if high accuracy predictions should be possible.

When modelling radio-wave propagation in urban areas it is important to distinguish macrocells from microcells. The antenna in a macrocell is often located on the top of tall buildings, giving them many metres clearance above the urban skyline. The signal thus passes above the surrounding buildings and enters the streets via reflections and diffraction. Macrocells used in relatively open urban areas have been modelled by Okumura-Hata formula and by multiple diffraction models, see [4] for an example. In microcells the antennas are often mounted on the side of buildings, under the urban skyline. The signals main propagation is around buildings and along streets, so under these conditions the streets can be modelled as waveguides connected via street crossings [21]. In this simulation we study microcells in the Manhattan area.

In the simulations a simplified version of the model described in [5] is used. It concentrates on the shifts between LOS and NLOS conditions, but does not include important characteristics like dual slope behaviour. We will only consider isotropic antennas ($G_A = 0$). The received signal strength P_r at the BTS is modelled by

$$P_r = P_t - 10\alpha \log(\frac{d_n}{K}) + \mathcal{N}(0, \sigma^2) \quad [dB], \tag{6.3}$$

where P_t is the transmitted signal strength, K is a constant, α is the attenuation factor and d_n is a NLOS distance which tries to incorporate the behaviour of the received signal strength under NLOS conditions in the model. The NLOS distance is defined by the recursion

$$\begin{cases} d_n = d_{n-1} + C_n + \beta_n D_n & n \neq 0 \\ d_0 = 0 & n = 0 \end{cases} \tag{6.4}$$

where n is the number of 90° turns, C_n is a variable modelling the loss due to direction change, β_n is a multiplying factor and D_n is the Euclidean distance between turn $n-1$ and turn n. Since we have discretised the bounded Manhattan area in Figure 6.1, d_n can only take a finite number of values.

The signal model given by (6.3) for a BTS at position $(10, 10)$ in the Manhattan area is shown in Figure 6.2, where the rapid decay of the received signal strength with distance, at NLOS can be seen. It is apparent that the received signal strength from a mobile inside a building is considerably lower than from outside. Berg [6] reports measurements where differences inside buildings compared to the level outside is from 30 dB up to 55 dB, which is used in the model.

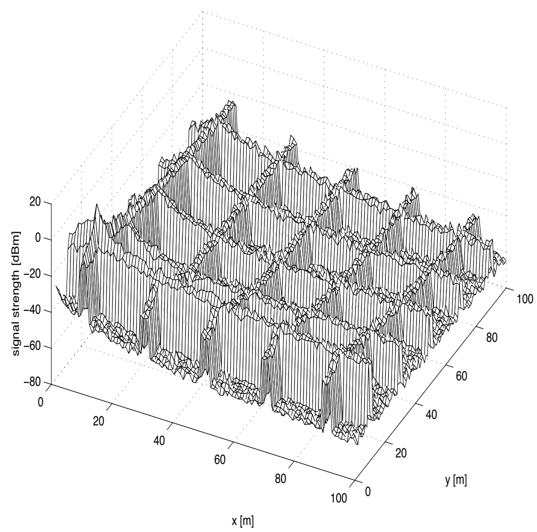
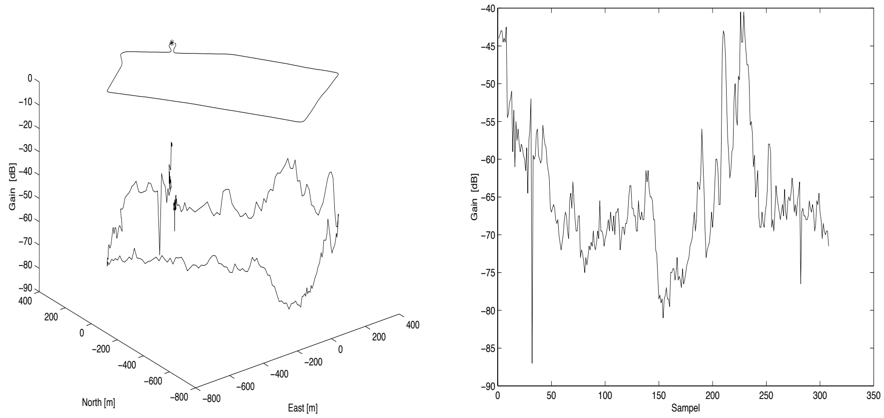


Figure 6.2: *Signal strength landscape from BTS* $(10, 10)$*. Note the rapid decay of the signal strength when there is NLOS conditions.*

Data from mobile signal strength measurements in an (sub-)urban area [35] is shown in Figure 6.3. A mobile is moving along a path shown in Figure 6.3(a) and the received signal strength has been sampled. In Figure 6.3(a) the signal strength is plotted as a function of spatial coordinates and in Figure 6.3(b) as a function of time (or sample). Some of the shifting of the signal strength in Figure 6.3 is possible to relate to shifting LOS and NLOS conditions given the map in Figure 6.4, but much better models are possible if altitude information is available.

The fading characteristic of the measurements can be compared with the characteristics of the model in Figure 6.7. The figures indicate that the model is not unrealistic containing the rapid shifts of signal strength between LOS and NLOS conditions, but this is of course a hand waving argument.



(a) Signal strength as a function of position.

(b) Signal strength as a function of sample.

Figure 6.3: *Measurements of received signal strength [db] from mobile moving in urban area with shifting LOS and NLOS conditions. Data from Ericsson [35].*

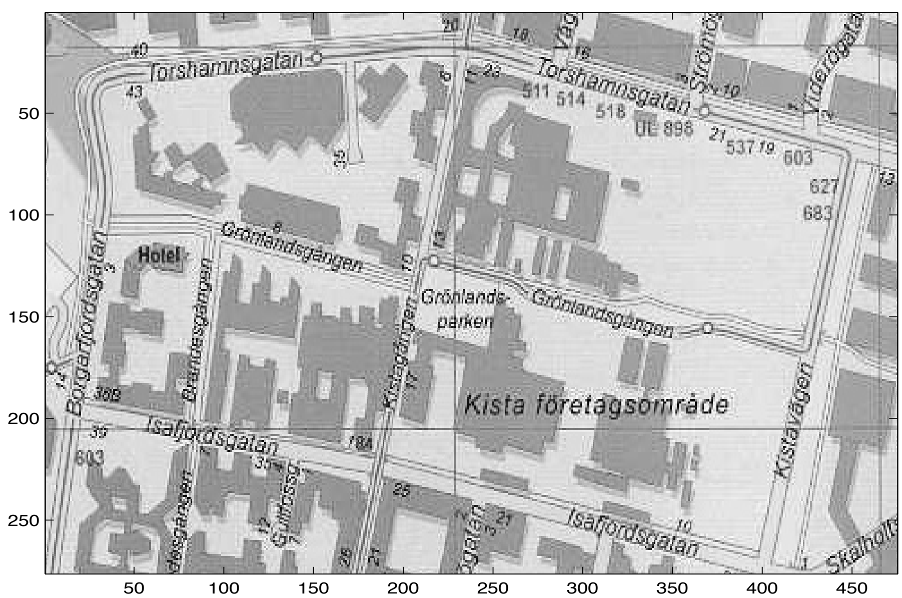


Figure 6.4: *A map over the Kista area where the measurements have been made. The BTS is mounted close to the start and of the movement. It should be possible to recognise the mobile route by comparing the map with Figure 6.3(a).*

Data from slowly moving mobiles [35] has been used to estimate the standard deviation σ of the fast fading. We found that $\hat{\sigma} \approx 6$ $[dB]$, but the estimate is suspected to depend heavily on the local environment.

6.1.3 Position Tracking Methods

The performance of the Kalman filter described in [18] and in Chapter 3 is compared to the basic SIR particle filter described in Section 5.2.

We will use two different state propagation models for the dynamic system in the Kalman filter case. The model in the filter called Kalman1 includes only the space coordinates in the state vector

$$\begin{pmatrix} X_{k+1}^1 \\ X_{k+1}^2 \end{pmatrix} = \begin{pmatrix} X_k^1 \\ X_k^2 \end{pmatrix} + \mathbf{W}_k, \tag{6.5}$$

where $\mathbf{W}_k \sim \mathcal{N}([0,0]^T, diag(6,6))$. In Kalman2 the velocity coordinates are included

$$\begin{pmatrix} X_{k+1}^1 \\ \dot{X}_{k+1}^1 \\ X_{k+1}^2 \\ \dot{X}_{k+1}^2 \end{pmatrix} = \begin{pmatrix} 1 & h & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & h \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_k^1 \\ \dot{X}_k^1 \\ X_k^2 \\ \dot{X}_k^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{W}_k, \tag{6.6}$$

where $\mathbf{W}_k \sim \mathcal{N}([0,0]^T, diag(1,1))$.

The particle filter does not include any velocity coordinates in the state vector as a first approximation and the velocity components are, as in the Kalman1 case, modelled as white noise. If the exploring particles at time k is $X_k^e(i)$, the exploring particles at time $k+1$ is a random sample from rectangular uniform distributions $p(x_{k+1}|x_k^e(i))$ with centre points $X_k^e(i) = x_k^e(i)$. Figure 6.5 shows the distribution used in the simulations, with a variance comparable with Kalman1.

The models for mobile movements used in these simulations are as simple as possible, but more advanced models are proposed in Chapter 7.

The observation model in the SIR case is a straightforward use of the signal model (6.3), where the received signal strength at the five BTS in the Manhattan area are assumed to be independent. This is a faster and more direct model than the observation model (3.4) in the Kalman filter, where the local least squares estimate of the position is treated as the observation in order to find the observation matrix M_k. The least square estimate has been calculated over a 8×8 pixel grid, centred around the previous position estimate, and the observation noise is $\mathcal{N}([0,0]^T, diag(3,3))$. Both the the size of the least squares grid and the variance of the observation should ideally be a function of the signal noise, but are assumed to be constant.

The models in the simulations are thus not exactly equivalent which reflects the fact that the particle filter is more general than the Kalman filter and does not need to be tuned to the specific problem to such an extent as the Kalman filter.

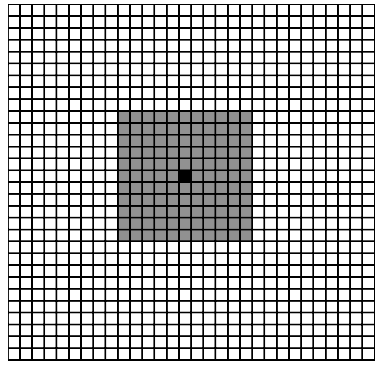


Figure 6.5: *Propagation model for the particle filter. The position of a particle at time k is marked with a black pixel. The propagation of this particle is a random sample from the uniform distribution over the rectangle marked by the grey pixels.*

6.2 Simulation 1 - Tracking Properties

In this simulation we investigate what effects the nonlinear signal landscape will have on the tracking algorithms. The initial position is supposed to be known, and only the tracking performance is studied.

6.2.1 Scenario

A mobile moves in the Manhattan area from left to right as presented in Figure 6.6. The speed is one pixel/second which corresponds to 36 km/hour. The signal strength sampling rate is one sample/second which is rather low[1]. A higher sampling rate would of course improve the estimates.

The received signal strength from one of the BTS as a function of time is presented in Figure 6.7. There are two peaks in the received signal strength corresponding to LOS conditions at the two street crossings during the movement. The transceiving BTS position is $(50, 50)$ in Figure 6.6, which can easily be seen by comparing the plot of the received signals strength with the movement.

[1]In the GSM system a test signal, that can be received by all the BTS is sent out every 0.48 seconds.

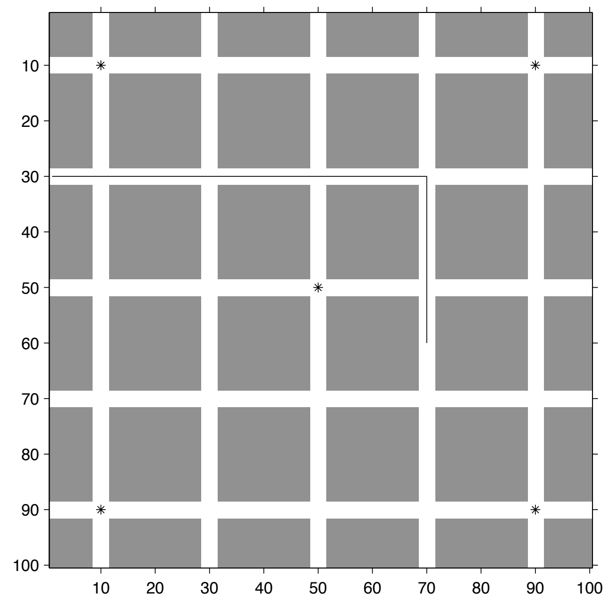


Figure 6.6: *The mobile moves from position* $(1, 30)$ *to position* $(70, 60)$ *following the line.*

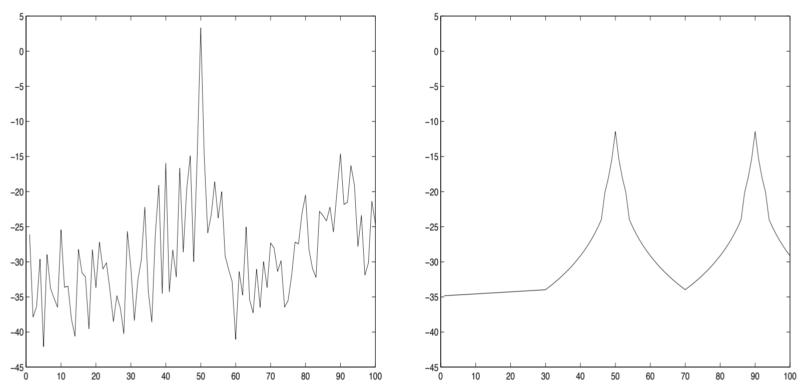


Figure 6.7: *The left plot is the noisy* ($\sigma = 6$) *received signal strength [dB] for the BTS in position (50,50) as a function of time, and the right plot is the undisturbed signal.*

The initial distribution for Kalman2 is $\mathcal{N}(\mu_2, \mathbf{C}_2)$ where

$$\mu_2 = \begin{pmatrix} 1 \\ 0 \\ 30 \\ 0 \end{pmatrix}, \quad \mathbf{C}_2 = \begin{pmatrix} 9 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}, \tag{6.7}$$

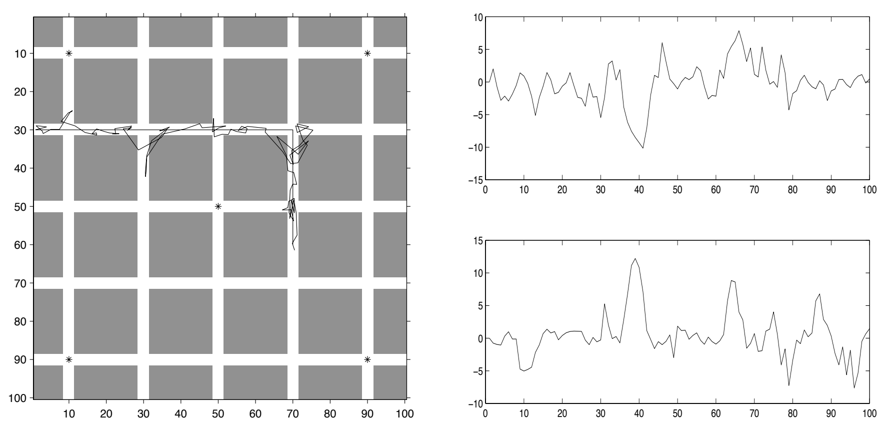
and $\mathcal{N}(\mu_1, \mathbf{C}_1)$ for Kalman1 and the particle filter where

$$\mu_1 = \begin{pmatrix} 1 \\ 30 \end{pmatrix}, \quad \mathbf{C}_1 = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}. \tag{6.8}$$

The filters are initiated with the true initial positions which is very advantageous in nonlinear tracking, but this is of course a rare event in a mobile tracking application.

6.2.2 Results

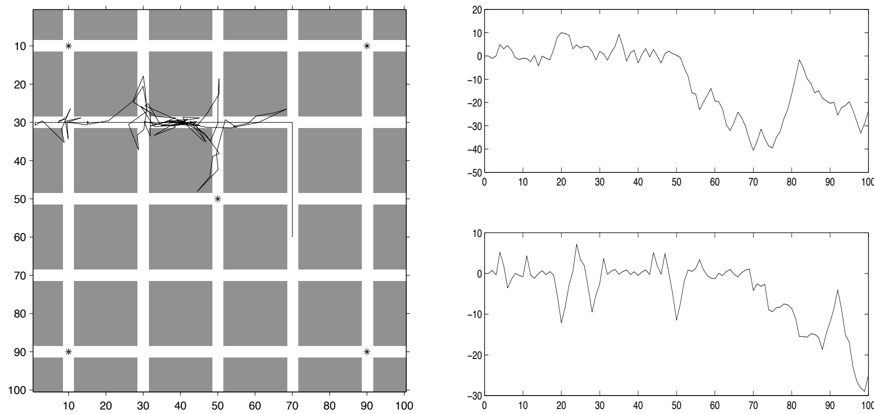
If exact information about initial position is available the Kalman filters work satisfactory when low power noise is added to the signal, see Figure 6.8. But as the noise level is raised the Kalman filters loose accuracy and often get stuck close to the initial point as in Figure 6.9.



(a) (b)

Figure 6.8: *For low noise levels the Kalman filter works satisfactory. (a) Positions predicted by the Kalman filter and (b) error in the x- and y-direction in pixel units ($\sigma = 2$).*

Inspired by the FCC standard (presented in Chapter 3) we define a performance measure to compare the particle filter and the Kalman filters in the



(a) (b)

Figure 6.9: *For higher noise levels the Kalman filter often get stuck close to the initial point. (a) Positions predicted by the Kalman filter and (b) error in the x- and y-direction in pixel units ($\sigma = 6$).*

simulation. Let T_s denote the number of estimated trajectories in the simulation, where the Euclidean distance between the position at time k and the true position at this time does not exceed 200 metres for any k in the sampling interval, and let the tracking success ratio R_s be defined by

$$R_s = \frac{T_s}{S}, \tag{6.9}$$

where S is the total number of simulations. In Figure 6.10 we present the result of $S = 100$ simulations for each noise level σ. The particle filter consists of $N = 200$ particles in these simulations, but with a larger number of particles the result of course becomes better.

6.3 Simulation 2 - Initial Conditions

In Simulation 1 we started the filters in the true position, but in a mobile tracking application this is seldom the case. An important feature of a good tracking filter is thus the adaptability to inaccurate initial conditions. In this simulation we will compare the performance of the filters when the initial distribution is biased.

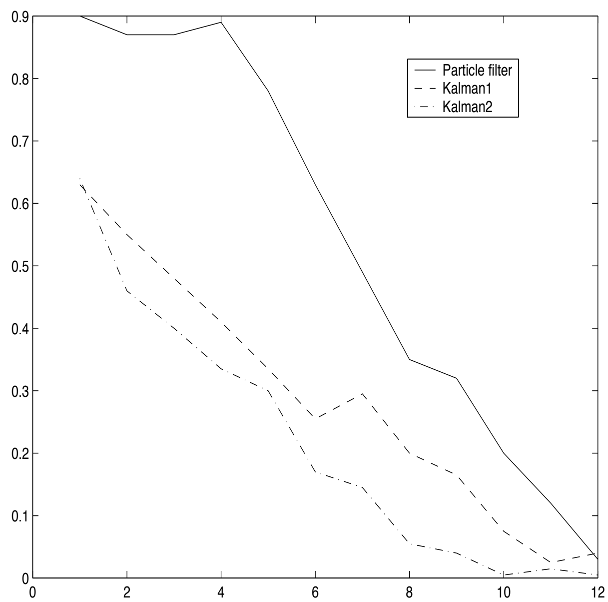


Figure 6.10: *The success ratio plotted for* $\sigma = 1, \ldots, 12$. *The particle filter performs better than the Kalman filters for all* σ.

6.3.1 Scenario

A mobile is stationary in the Manhattan area at position $(30, 70)$, see Figure 6.11, and the signal strength sampling rate is again one sample per second. The initial distribution for Kalman2 is assumed to be $\mathcal{N}(\mu_2, \mathbf{C}_2)$ where

$$\mu_2 = \begin{pmatrix} 30 \\ 0 \\ 30 \\ 0 \end{pmatrix}, \quad \mathbf{C}_2 = \begin{pmatrix} 225 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 225 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}. \tag{6.10}$$

The initial distribution for Kalman1 is assumed to be $\mathcal{N}(\mu_1, \mathbf{C}_1)$ where

$$\mu_1 = \begin{pmatrix} 30 \\ 30 \end{pmatrix}, \quad \mathbf{C}_1 = \begin{pmatrix} 225 & 0 \\ 0 & 225 \end{pmatrix}. \tag{6.11}$$

In Figure 6.11 circles which contains 39% and 86% of the probability mass in the space domain is marked. The true position is thus not very probable. The particle filter is initialised by sampling N particles from $\mathcal{N}(\mu_1, \mathbf{C}_1)$.

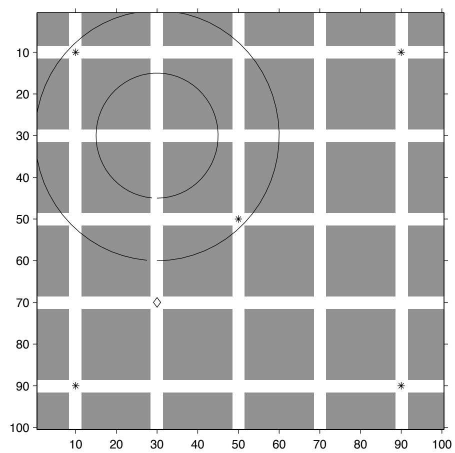


Figure 6.11: *The mobile is stationary in the position marked with a diamond and the initial Gaussian distribution is biased. The circles contains* 39% *and* 86% *of the probability mass in the space domain.*

6.3.2 Results

In this simulation we want to investigate if the filters locate the mobile and how long time it takes. Hence we define a convergence ratio R_c by

$$R_c(k) = \frac{T_c(k)}{S}, \tag{6.12}$$

where $T_c(k)$ is the number of independent simulations where the estimated position are within 50 metres from the true position at time k, and S is the total number of simulations. The results of $S = 100$ simulations for each algorithm are presented in Figure 6.12.

6.4 Discussion and Conclusions

It is clear from the results of the simulation that the particle filter is superior to the Kalman filter given the signal model and signal environment used in the simulations. We will summarise the simulations by emphasising some important factors that distinguish the filters.

6.4.1 Representation of Distributions

The Kalman filter only propagate the mean and the covariance of the conditional distribution assuming it is Gaussian which might be highly misleading. The particle filter propagates an approximation of the whole distribution and is

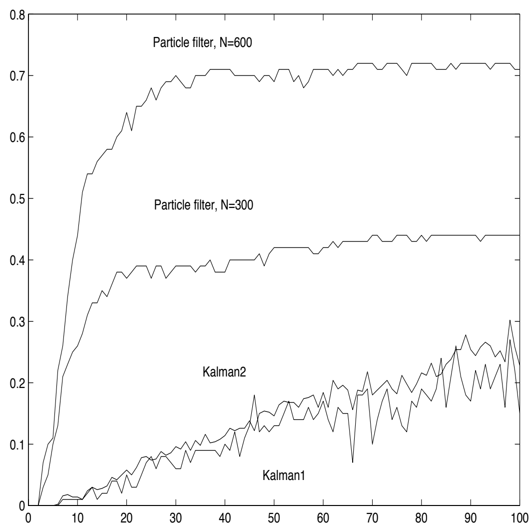


Figure 6.12: *The convergence ratio $T_c(k)$ as a function of time (sample) k. The particle filters show a higher convergence ratio than the Kalman filter for both $N = 300$ and $N = 600$.*

thus not restricted to Gaussian distributions as seen in Figure 6.13 where the distribution adapts to the street structure.

In Figure 6.14 the distribution is bimodal due to symmetry at the middle BTS, which is the most informative since it is the nearest and have LOS conditions. Since the Kalman filter assumes an unimodal Gaussian distribution the risk that it makes a erroneous choice in a street crossing substantial.

In a mobile positioning application it is important to use all the information available in the conditional distribution and not only the expectation. The true position in Figure 6.14 is around $(50, 30)$ but the conditional distribution also have a second mode around $(30, 50)$. Only presenting the mean in a situation like this is a waste of information and can also be totally misleading, even if it is the optimal minimum variance estimate. Taking the whole conditional distribution under consideration gives much more information and should be used by the positioning system.

6.4.2 Initial Estimates

The dependence on a good initial estimate is crucial for the Kalman filter. But in reality good initial estimates are seldom available, and often the uncertainty can be very large. In Simulation 2 we simulated a situation where the initial distribution was misleading (see Figure 6.11). Since the mean value of the distribution was on the wrong side of the LOS street $(\cdot, 50)$ for BTS $(50, 50)$, it is difficult for the Kalman filter to pass this street and the filter often gets stuck.

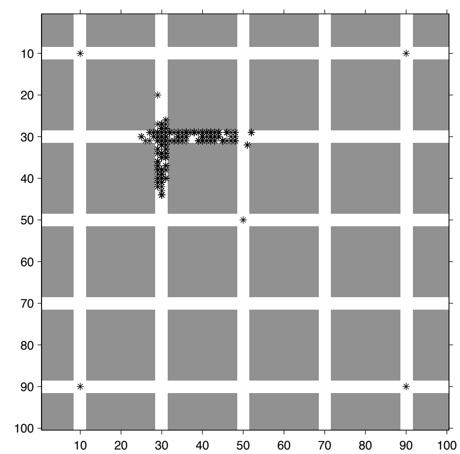


Figure 6.13: *Approximation of* $p(x_k|y_{\mathbf{k}})$ *with the particle filter. The conditional distribution is centred around the true position* $(40, 30)$ *and adapts well to the street crossing.*

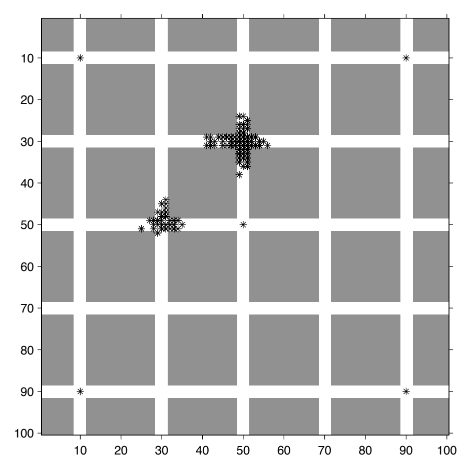


Figure 6.14: *Approximation of* $p(x_k|y_{\mathbf{k}})$ *with the particle filter. The conditional distribution is bimodal since the signal model is symmetric for the BTS in position (50,50). The true position is* $(50, 30)$.

The particle filter is better suited for this situation because the random sample from the initial distribution will, with probability depending on the number of particles, include some particles close to or on the right side of the LOS street for BTS $(50, 50)$. This is a very useful feature, that is of great importance for global positioning problems. In [8] for example, the SIR filter is used to track a mobile service robot in a museum under very uncertain initial conditions, with successful outcome.

6.4.3 Nonlinear Signal Landscape

An important condition for the optimality of the Kalman filter is a linear model. This is not the case in an urban area and not in the signal model used in the simulation, as shown in Figure 6.2. The effect on the Kalman filter can be seen in Figure 6.8 where the predicted positions often changes from inside a building to outside. This is due to the erroneous assumption of linear signal landscape. If the observation does not match the predicted position the Kalman gain matrix will modify the position under linear model assumption. In a nonlinear signal landscape the result can be totally misleading.

Chapter 7

Enhanced Models

7.1 Model Extensions

The model is one of the main components in a successful tracking procedure. We have already mentioned that it is important that it does not only give a good description of the physical situation but also that it is mathematical tractable. The model must also provide a good balance between the complexity and the possibility to make inference from the observations. These obvious reflections are important to have in mind when developing models.

A natural extension to the simple propagation models used in the simulations is to let the model of the mobile movement change with position and time. For example, the models should of course be different if the mobile is moving on the city square or on a train out from the city; rush hour traffic models should differ from more low traffic models and so on. The dependence on state and time in the model is straightforward to implement in the particle filter. This also holds for the Kalman filter but with the usual restriction to linear models.

It could also be fruitful to include the velocity in the state vector also in the particle filter. The velocity information can be used for hand over predictions, which can increase the system capacity. It is straightforward to use a continuous velocity distribution but it might also be relevant to distinguish between a finite number of states. For instance, the mobiles movements can be classified as: stationary, walking, driving a car in different speeds depending on road or using the subway. Increasing the dimension of the state vector by including velocity components will increase the number of particles needed to approximate the distribution in most cases, due to new correlation structures.

7.2 Combinations of Positioning Methods

Different positioning methods all have their pros and cons. The signal strength measurements that we have used as input in the simulations suffer from fast fading. There are also many different factors that influences the signal strength but are impossible to include in a model, such as if the mobile is held close to the head or in the air, if the mobile is in a crowded place or not, etc.

TDOA, TOA and AOA are not dependent on the signal strength explicitly but they all suffer from multipath effects even if the errors might cancel out to a certain degree assuming similar NLOS conditions to each BTS in the TDOA case. One major problem in common for all of the methods is the hearability since is certainly not always possible to detect a signal at three or more BTS at the same time. In Figure 7.1 we observe how the initial PDF is spread out if only one isotropic BTS detects the signal in the TOA case. A signal strength model combined with the TOA method could have resolved this problem better depending on the symmetry conditions in the signal strength model. Note that the particle filter is useful for positioning under these conditions.

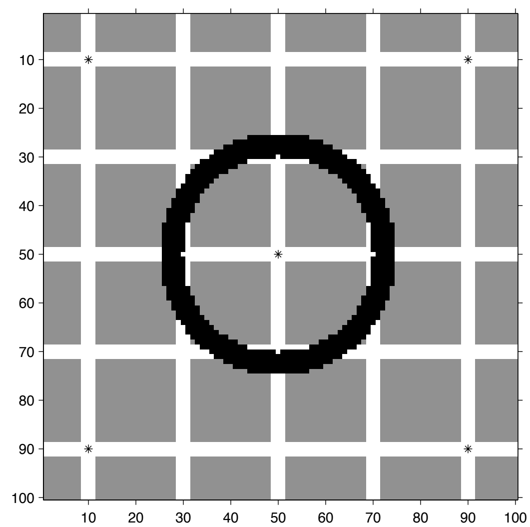


Figure 7.1: *Initial distribution for the position of a mobile when only one isotropic BTS detects the signal in the TOA case.*

In a close future there will also be a demand for positioning in the height dimension, which calls for improved techniques. A combination of the methods seems very appealing since it would both increase the accuracy in positioning and also give information about future trends in signal strength for the mobile at different BTS. An open question is how to combine the observations, which is much an engineering problem. However, the particle filter gives much freedom for different solutions and they should all be straightforward to implement.

Chapter 8

Summary

To conclude the thesis, we briefly review the thesis and comment on the results from the simulations.

Recently there has been much interest in the area of mobile positioning, only using information from received signals. In Chapter 3 we reviewed some of the different methods used in positioning. A TDOA approach seems to be the most promising, but a combination with methods like TOA, AOA and signal strength should be preferable since it can reduce hearability problems, and also provide useful additional information to the system operator. If the signal strength analysis shall be a useful tool in the future, better signal propagation models need to be developed. A combination of theoretical models and measurements could be fruitful.

Since the mobiles are not stationary, and a recursive estimation procedure is preferable, the Kalman filter has been proposed as a solution in a recent study [19]. However, the Kalman filter is restricted to linear models and Gaussian noise and a more general but approximative class of filters has been investigated. This class is called particle filters, and in Chapter 5 an introduction to the particle filters is given. In contrast to other approximative methods a particle filter does not need any problem specific adjustments, it can be applied to a wide class of nonlinear filtering problems.

In Chapter 6 we compared the performance of the Kalman filter and the SIR particle filter on tracking in a simulated Manhattan area. A nonlinear signal strength propagation model based on [4, 5] has been developed and used in the simulations. Simulation 1 compared the ability to track the moving mobile, given the true initial position, under different noise levels. Simulation 2 investigated how the filter managed to track a stationary mobile when the initial distribution was biased. The particle filter was superior compared to the Kalman filter in both simulations.

The particle filters seem like a very relevant tool for mobile tracking mainly because of four reasons:

1. They are general enough to handle a large class of filtering problems.

2. They deal with uncertain initial conditions in a robust way.
3. They are intuitive and easy to implement in a computer and to extend as the models are improved.
4. They contain more information than methods that only propagates the first moments of the conditional distribution.

There are several questions that remain unanswered and are open for further research. The actual running time has not been investigated since no time optimised implementation has been done; it would probably be efficient to implement the SIR filter for parallelisation. The particle filter converges to the conditional mean as the number of particles goes to infinity, but how many particles are needed to reach a certain accuracy at a certain time point and how will the number of points depend on the dimension of the state vector? The latter question will of course depend on the correlation between the components in the state vector. Future research could also involve the problem of both tracking the mobile and updating the signal model at the same time. A model of how small changes in the position and in the signal model changes the positioning error must then be developed.

Bibliography

[1] L. Ahlin, J. Zander. *Principles of wireless communications*, Studentlitteratur, 1998.

[2] D. L. Alspach, H. W. Sørenson. Non-linear Bayesian estimation using Gaussian sum approximation. *IEEE Trans. Auto. Control*, 17, pp. 439-447, 1972.

[3] F. Bacceli, S. Zuyev. Stochastic geometry models of mobile communication networks. In *Frontiers in queueing. Models and Applications in Science and Engineering*, Eds. J.H. Dshalalow, CRC Press, pp. 227-244, 1997.

[4] J. E. Berg, H. Holmquist. An FFT Multiple Half-Screen Diffraction Model. *Vehic. Tech. Conference, 1994 IEEE 44th*, pp. 195-199, 1995.

[5] J. E. Berg. A Recursive Method For Street Microcell Path Loss Calculations. *Personal, Indoor and Mobile Radio Communications, 1995. Wireless: Merging onto the Information Superhighway. Sixth IEEE International Symposium on:*, Vol. 1, pp. 140-143, 1995.

[6] J. E. Berg. Building penetration loss at 1700 MHz along line of sight street microcells. *Personal, Indoor and Mobile Radio Communications, 1992. Proceedings. Third IEEE International Symposium on:*, pp. 86-87, 1992.

[7] J. Carpenter, P. Clifford, P. Fearnhead. Improved particle filter for nonlinear problems. *IEE Proc.-Radar, Sonar Navig*, Vol. 146, No. 1, February 1999.

[8] F. Dellaert, D. Fox, W. Burgard, S. Thrun. Monte Carlo Localization for Mobile Robots. *Proceedings on the 1999 IEEE Int. Conf. on Robotics & Automation*, Detroit, Michigan, 1999.

[9] P. Del Moral. Non-linear filtering: interacting particle solution. *Markov Processes and Related Fields*, Vol. 2, pp. 555-579, 1996.

[10] P. Del Moral. A Uniform Convergence Theorem for the Numerical Solving of the Nonlinear Filtering Problem. *J. Appl. Prob.* 35, pp. 873-884, 1998.

[11] P. Del Moral, G. Rigal, J.C. Noyer, G. Salut. Traitement non-linéare du signal par reseau particulaire: application radar. *Quatorziéme colloque GRETSI*, Juan les Pins, 1993.

[12] R. Durrett. *Probability: Theory and examples.* Duxbury Press, 1996.

[13] S. Fisher, H. Koorapaty, E. Larsson, A. Kangas. System performance evaluation of mobile positioning methods. *Vehicular Technology Conference, 1999 IEEE 49th*, Vol. 3, pp. 1962-1966, 1999.

[14] S. Fisher, H. Grubeck, A. Kangas, H. Koorapaty, E. Larsson, P. Lundqvist. Time of arrival estimation of narrowband TDMA signals for mobile positioning. *Personel, Indoor and Mobile Communications, The Ninth IEEE International Symposium on:*, Vol. 1, pp. 451-455, 1998.

[15] N.J. Gordon, D.J. Salmond, A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE-Proceedings-F 140*, pp. 107-113, 1993.

[16] M. Hata. Empirical Formula for Propagation Loss in Land Mobile Radio Services. *IEEE Trans. on Vehicular Technology*, Vol. 29, No. 3, pp. 317-325, August 1980.

[17] M. Hata. Propagation Loss Prediction Models for Land Mobile Communications. 1998 International Conference on *Microwave and Millimeter Wave Technology Proceedings, 1998* , pp. 15-18, 1998.

[18] M. Hellebrandt, R. Mathar, M. Sheibenbogen, Estimating Position and Velocity of Mobiles in a Cellular Radio Network, *IEEE Trans. on Vehicular Technology*, Vol. 46, No. 1, pp. 65-71, February 1997.

[19] M. Hellebrandt, R. Mathar, Location tracking of Mobiles in a Cellular Radio Networks, *IEEE Trans. on Vehicular Technology*, Vol. 48, No. 5, pp. 1558-1562, February 1999.

[20] K. Ito, K. Xiong. Gaussian Filters for Nonlinear Filtering Problems. *IEEE Transaction on Automatic Control*, Vol. 45, No. 5, 2000.

[21] R. Jakoby, U. Liebenow. Modelling of Radiowave Propagation in Microcells. *Antennas and Propagation 4-7 April 1995*, IEEE Conference Publication No. 407, 1995.

[22] A.H. Jazwinsky, *Stochastic processes and filtering theory*, Academic Press, 1970.

[23] R.E. Kalman, A New Approach to Linear Filtering and Prediction Problems, *Trans. ASME, Ser. D: J. Basic Eng. 82*, pp. 35-45, 1960.

[24] S. C. Kramer, H. W. Sørenson. Recursive Bayesian estimation using piecewise constant approximations. *Automatica*, 24, pp. 789-801, 1988.

[25] H. Krim, M. Viberg. Two Decades of Array Signal Processing - The Parametric Approach. *IEEE Signal Processing Magazine*, pp. 67-94, July 1996.

[26] T.P. McGarty. *Stochastic Systems and State Estimation*, Wiley-Interscience, 1974.

[27] J. R. Moon, C. F. Stevens. An approximate linearisation approach to bearings-only tracking. *IEE Target Tracking and Data Fusion: Colloquium digest 96/253*, pp. 8/1-8/14, 1996.

[28] B. Øksendal. *Stochastic Differential Equations - an introduction with applications.* Springer 1998.

[29] R. Owen, L. Lopez. Experimental analysis of the use of angle of arrival at an adaptive antenna array for location estimation. Presented at *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, Vol. 2, pp. 607-611, 1998.

[30] J.G. Proakis. *Digital communications.* McGraw-Hill International Editions, 1995.

[31] T. S. Rappaport, J. H. Reed and B. D. Woerner. Position Location Using Wireless Communication on Highways of the Future. *IEEE Communications Magazine*, pp. 33-41, October 1996.

[32] G. Swedberg. Ericsson's mobile location system. *Ericsson Review*, No. 04, 1999.

[33] M. West, P. J. Harrison, H. S. Migon. Dynamic generalised linear models and Bayesian forecasting, with discussion. *J. Am. Stat. Assoc.*, 80, pp. 73-97, 1985.

[34] Homepage: www.ksix.com

[35] Data från Magnus Almgren, Ericsson Radio Systems AB, Stockholm 1999.