TAMPERE UNIVERSITY OF TECHNOLOGY

Henri Nurminen

Position estimation using RSS measurements with unknown measurement model parameters

Master of Science Thesis

Subject approved by the Department Council on 4.4.2012

Examiners: Prof. Robert Piché (TUT)

D.Tech. Simo Ali-Löytty (TUT)

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Science and Engineering
NURMINEN, HENRI JAAKKO JULIUS: Position estimation using RSS measurements with unknown measurement model parameters
Master of Science Thesis, 54 pages, 10 Appendix pages
December 2012
Major: Mathematics
Examiners: Prof. Robert Piché and D.Tech. Simo Ali-Löytty
Keywords: indoor positioning; outdoor positioning; cellular networks; WLAN; received signal strength; path loss model; statistical estimation

The availability and performance of satellite-based navigation systems are the weakest in urban areas and indoor spaces, where the user density would be high. In these environments alternative low-cost positioning techniques are needed. This thesis considers positioning using received signal strength (RSS) measurements of terrestrial wireless networks.

No prior knowledge of the considered wireless networks is assumed in this thesis, but only a simplified statistical path loss model for signal propagation. The model parameters are estimated for each base station of the network separately using pre-collected learning data. The method is based on Bayesian estimation theory that characterizes the precision of the parameter estimates, which is an essential feature. Three Bayesian position estimation methods are proposed in this thesis. Two versions of each are compared: one uses point estimates for the model parameters and assumes them to be accurate, whereas the other takes the finite parameter precisions into account.

Real-data tests are accomplished using cellular networks in outdoor and wireless local area networks (WLAN) in indoor spaces. The tests indicate that taking the finite parameter precisions into account improves positioning accuracy and especially makes error estimation more realistic. Furthermore, RSS-based methods outperform the method that uses only the list of observed base stations and no RSS information. The advantages of parametric methods compared with the k-nearest neighbour method, which can be regarded as the state-of-the-art positioning method, are also shown.

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Teknis-luonnontieteellinen koulutusohjelma
NURMINEN, HENRI JAAKKO JULIUS: Paikan estimointi käyttäen signaalinvoimakkuusmittauksia tuntemattomilla malliparametreilla
Diplomityö, 54 sivua, 10 liitesivua
Joulukuu 2012
Pääaine: Matematiikka
Tarkastajat: professori Robert Piché ja TkT Simo Ali-Löytty
Avainsanat: sisätilapaikannus; ulkopaikannus; matkapuhelinverkko; langaton lähiverkko; signaalinvoimakkuusmittaus; vaimenemismalli; tilastollinen estimointi

Satelliitteihin perustuvien paikannusjärjestelmien suorituskyky on heikoimmillaan kaupunkialueilla ja sisätiloissa, joissa käyttäjätiheys olisi suuri. Näissä ympäristöissä on tarvetta vaihtoehtoisille matalan kustannuksen paikannusmenetelmille. Tämä työ käsittelee paikannusmenetelmiä, jotka perustuvat maanpäällisten langattomien verkkojen signaalinvoimakkuuksien mittaamiseen.

Käsiteltävistä langattomista verkoista ei oleteta mitään esitietoja vaan vain yksinkertainen tilastollinen vaimenemismalli signaalin etenemiselle. Mallin parametrit estimoidaan verkon kullekin tukiasemalle erikseen käyttäen etukäteen kerättyä opetusaineistoa. Menetelmä perustuu bayesläiseen estimointiteoriaan, joka mallintaa myös parametriestimaattien tarkkuutta, mikä on keskeinen ominaisuus. Paikan estimointiin esitetään kolme bayesläistä menetelmää. Kustakin vertailussa on kaksi versiota: toinen käyttää mallin parametreille piste-estimaatteja ja olettaa ne tarkoiksi, kun taas toinen huomioi parametriestimaattien tarkkuudet.

Menetelmiä testataan todellisella mittausaineistolla käyttäen matkapuhelinverkkoja ulkopaikannuksessa ja langattomia lähiverkkoja sisätilapaikannuksessa. Testien perusteella äärellisten parametritarkkuuksien huomioiminen parantaa paikannustarkkuutta ja erityisesti virhearvioinnin realistisuutta. Lisäksi signaalinvoimakkuusmittauksia käyttävät menetelmät toimivat paremmin kuin menetelmä, joka käyttää vain tietoa kuulluista tukiasemista ilman signaalinvoimakkuusinformaatiota. Tässä työssä osoitetaan myös parametristen menetelmien edut verrattuna k:n lähimmän naapurin menetelmään, jota voidaan pitää kehittyneimpänä ratkaisuna esitettyyn paikannusongelmaan.

Preface

I wrote this Master of Science Thesis at the Department of Mathematics at Tampere University of Technology. The topic of the thesis arose from the research that I have been conducting for the Personal Positioning Algorithms Research Group during the last two and a half years.

I thank my examiners D.Tech. Simo Ali-Löytty and Prof. Robert Piché for inspiring research topics and enlightening feedback. I am also thankful to my colleagues in the Positioning Group, especially to Philipp Müller and Laura Wirola. Furthermore, I thank Jukka Talvitie for cooperation. The project in which I have been working is funded by Nokia Inc., and I am thankful to D.Tech. Lauri Wirola and D.Tech. Jari Syrjärinne from Nokia for ideas, comments and advice in technological questions. Shweta Shrestha, Toni Fadjukoff and Miila Martikainen are acknowledged for collecting measurements that were used in this work.

For support in education and life I thank my fellow-students in "Math Corner" and all the other people who belong to my life. I want to express my special gratitude to my family in Vammala and to my wife Sini.

"For the Lord gives wisdom; from his mouth come knowledge and understanding."

Tampere, 20th November 2012

Henri Nurminen
Kanjoninkatu 17 B 31
33720 Tampere

Contents

Abbreviations

2D	two-dimensional
3D	three-dimensional
95% err.	95% quantile of errors
"acc"	algorithm assuming point estimates of PL parameters accurate
AOA	angle of arrival
AP	access point
BS	base station
CA	coverage area
CKF	Cubature Kalman filter
Cons.	(95%) consistency of a Bayesian estimate
EKF	Extended Kalman filter
EKF2	Second order extended Kalman filter
FP	fingerprint
GMF	Gaussian mixture filter
GN	Gauss–Newton
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
ID	identification number
IEKF	Iterated extended Kalman filter
KF	Kalman filter
MAP	*maximum a posteriori*, mode of the posterior density
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
Mean	mean error
Med	median error
MH	Metropolis–Hastings
"N"	algorithm using normal prior for PL parameters
pdf	probability density function
PL	path loss
PKF	Positioning Kalman filter

RM	radiomap
RSCP	Received Signal Code Power
RSS	received signal strength
RX	signal receiver
TDOA	time difference of arrival
TOA	time of arrival
TX	signal transmitter
UE	user equipment
UKF	Unscented Kalman filter
WCDMA	Wideband Code Division Multiple Access
WLAN	wireless local area network

Symbols

Matrices are denoted with unitalicized uppercase letters. Vectors and scalars are not distinguished. Random variables are typed with boldface font, their realizations with normal font.

$\propto$	is proportional, equal up to a multiplying constant
$\propto\!\!\!\!\sim$	is approximately proportional
$x \ll y$	x is much less than y
$x := y$	value y is assigned into variable x (in algorithm listings)
$x \leftarrow \mathrm{D}$	random number is generated from probability distribution D and assigned to variable x
$x \in A$	x is an element of set A
$A \ni x$	set A includes element x
$A \subset B$	set A is a subset of set B
$\mathbb{N}$	set of natural numbers (without zero)
$\mathbb{N}_0$	set of natural numbers and zero
$\mathbb{Z}$	set of integers
$\mathbb{R}$	set of real numbers
$\mathbb{R}^+$	set of positive real numbers
$\operatorname{argmin}_x f(x)$	value x that minimizes function f
$\operatorname{argmin}_{x \in A} f(x)$	value x that minimizes function f in the set A
$f : A \to B$	function f with domain A and codomain B
$\operatorname{supp}(f)$	support of function f
$x \mapsto f(x)$	function that maps x to $f(x)$
$\int$	integral
$\frac{\mathrm{d}f}{\mathrm{d}x},\ f'$	derivative of function f, derivative vector of function f (row vector), Jacobian matrix of f
$\frac{\partial f}{\partial x}$	partial derivative (matrix) of function f with respect to variable x
$\log_a(x)$	a-based logarithm of x
$\ln(x)$	natural logarithm of x

$\|x\|$	Euclidean norm of vector x
A^{T}	transpose of matrix A
A^{-1}	inverse of matrix A
$\mathrm{A}^{\dagger}$	Moore–Penrose pseudoinverse of matrix A
I	identity matrix
O	zero matrix
$\mathcal{N}(\mathrm{A})$	kernel of matrix A
rank(A)	rank of matrix A
det(A)	determinant of matrix A
blkdiag(A, B)	block diagonal matrix
$\|\mathrm{A}\|$	Frobenius norm of matrix A
$\{x_k\},\ \{x_k\}_{i=0}^{\infty}$	sequence of $x_0, x_1, x_2, \ldots$
$\{x_k\}_{i=0}^{N}$	finite sequence of $x_0, x_1, x_2, \ldots, x_N$
$\{\boldsymbol{x}_k\}$	stochastic process of $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$
$\sum_{k=0}^{N} x_i$	the sum $x_0 + x_1 + x_2 + \cdots + x_N$
$\boldsymbol{x}$	the state of the system
$\boldsymbol{x}_k$	time-variant state components of the dynamical system at time instant t_k
$\hat{x}$	prior mean of the state
$\hat{\Sigma}_{\boldsymbol{x}}$	prior covariance matrix of the state
$\hat{x}^{+}$	posterior mean of the state
$\hat{\Sigma}^{+}$	posterior covariance matrix of the state
$\boldsymbol{a}$	unknown model parameters, time-invariant state components
$\hat{a}$	prior mean of the unknown model parameters
$\hat{\Sigma}_{\boldsymbol{a}}$	prior covariance matrix of the unknown model parameters
$\boldsymbol{\xi}$	complete state, $\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{x}^{\mathrm{T}} & \boldsymbol{a}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$
$\boldsymbol{w}$	process noise (vector)
Q	covariance matrix of process noise
Φ	state transition matrix, objective function of the Gauss–Newton (GN) algorithm
$\boldsymbol{y}$	measurement vector, received signal strength (RSS) measurement
h	measurement model function
H	measurement model matrix, Jacobian matrix of the measurement model function h
$\boldsymbol{v}$	measurement noise, shadowing term in the path loss model
R	covariance matrix of measurement noise
$\mathrm{P}(A \mid B)$	probability of A given B
$p(x \mid \boldsymbol{y} = y)$	conditional probability density function (pdf) of random variable $\boldsymbol{x}$ given $\boldsymbol{y} = y$

$p(x \mid y)$	short-hand notation of the previous one
$\mathcal{L}(x)$	likelihood function of random variable $\boldsymbol{x}$
$\mathrm{E}(\boldsymbol{x})$	expectation value of random variable $\boldsymbol{x}$
$\mathrm{var}(\boldsymbol{x})$	variance or covariance matrix of random variable $\boldsymbol{x}$
$\mathrm{E}(\boldsymbol{x} \mid y)$	conditional expectation value of random variable $\boldsymbol{x}$ given $\boldsymbol{y} = y$
$\mathrm{var}(\boldsymbol{x} \mid y)$	conditional variance or covariance matrix of random variable $\boldsymbol{x}$ given $\boldsymbol{y} = y$
$\mathrm{N}(m, \mathrm{P})$	(multivariate) normal distribution with mean m and covariance matrix P
$\mathrm{N}(x \mid m, \mathrm{P})$	pdf of the (multivariate) normal distribution with mean m and covariance matrix P
$\mathrm{Unif}(x, y)$	uniform distribution in interval (x, y)
π	target distribution of the Metropolis–Hastings (MH) algorithm
$r(x, y)$	MH ratio with current state x and proposal value y
N_b	length of burn-in period in MH algorithm
J_k	Jacobian matrix of the measurement model function at kth GN iteration
$(\Delta x)_k$	GN step at kth iteration
α_k	GN damping coefficient at kth iteration
$m_k^-,\ \mathrm{P}_k^-$	mean and covariance matrix of the Kalman filter prediction step
K_k	Kalman gain
$m_k^+,\ \mathrm{P}_k^+$	mean and covariance matrix of the Kalman filter update step
$P(r)$	signal power at distance r from the transmitter measured at some reference surface
P_RX	signal power received by the receiver (RX)
A	apparent transmitter power
n	path loss exponent, attenuation factor
m	position of the transmitter
σ	(in path loss model context) standard deviation of shadowing
N_ℓ	number of fingerprints for one base station (BS)
A^-, n^-, m^-	prior mean values for PL parameters
$\sigma_{\boldsymbol{A}}^{-2}, \sigma_n^{-2}, \sigma_m^{-2}$	prior mean values for PL parameters
$\hat{A}, \hat{n}, \hat{m}$	estimated mean values for PL parameters
$\hat{\sigma}^2_{\boldsymbol{A}}, \hat{\sigma}^2_{\boldsymbol{n}}, \hat{\Sigma}_{\boldsymbol{A},\boldsymbol{n}}, \hat{\Sigma}_{\boldsymbol{m}}$	estimated variances or covariance matrices for PL parameters
p	user position
$\hat{p}, \hat{\Sigma}_{\boldsymbol{p}}$	prior mean and covariance matrix for user position
$N_{\boldsymbol{y}}$	number of received RSS measurements (positioning phase)
$\mathrm{P}_{\boldsymbol{p}}, \mathrm{P}_{\boldsymbol{m}}$	covariance matrices of the proposal distribution of user position and BS position in MH algorithm

Chapter 1

Introduction

The commercial significance of location information and navigating methods has been growing rapidly due to the upsurge in the number of mobile applications that are based on awareness of the user's position. The process of finding one's location is called positioning in this thesis. Many positioning applications use information from a global navigation satellite system (GNSS) such as the Global Positioning System (GPS). Nowadays it is generally considered that satellite-based outdoor positioning is a solved problem at least in sparsely populated areas.

However, there are numerous use cases where a low cost positioning method that does not utilize satellite-based information can be preferable. There are also cases where satellite-based information is completely unavailable. For instance, densely built urban areas and indoor spaces may be completely or partially shadowed from radio signals transmitted by satellites. The user density might, however, be highest in these environments. The technology for receiving and interpreting a GNSS measurement is already considered inexpensive, but the energy consumption that the GNSS connection requires can be too high for many use cases. Satellite-based methods can also be supported by some terrestrial source, for example in decreasing the time to first fix value of the GNSS receiver.

Recently, an increasing amount of research interest has been focused on terrestrial sources of location information, such as cellular networks, WLANs (wireless local

area networks) and television and radio networks. In addition to different network sources, inertia sensors, barometers and compasses, for example, could also be used for positioning. Algorithms that combine pieces of location information that originate from various sources are called hybrid positioning algorithms. This thesis concentrates on positioning with wireless networks, but the possibilities of combining additional information into the presented algorithms are discussed.

A network measurement that is used for positioning includes typically the identification number (ID) of the transmitting base station (BS) and other information on the signal, which is an electromagnetic wave. The receiver is a mobile terminal that is called the user equipment (UE) in this thesis. This thesis studies fingerprinting, which is a technique where a positioning measurement is compared with a database of similar measurements (radiomap, RM) that have been collected beforehand. It is then probable that the UE is located in the area where the measurement resembles the RM measurements. This thesis considers mainly the measurements that are based on measuring the intensity of the signal that the UE receives. RSS is the received signal strength, and path loss model is a model for signal strength attenuation.

In the literature, there are numerous studies on location estimation with terrestrial wireless networks and fingerprinting methods, among others [6, 22, 34, 15]. There are also commercial implementations, for example [41, 11, 31].

The usability of many terrestrial location information sources is limited by noisiness of measurements. Noise can be highly dependent on e.g. the topography of the surrounding terrain or buildings. It can vary by season, by the time of day or even by the way the UE is carried. Thus, the noise tends to be too complex to be modeled accurately. Stochastic models are typically used in modeling the measurement processes; the position cannot be calculated accurately but it has to be estimated. For general usability and especially for combining measurements of different kind, it is crucial to have information on the accuracy of the position estimate.

Bayesian statistics is a collection of statistical principles and methods in which randomness is seen to describe the missing information. Therefore, all the quantities are represented by probability distributions, i.e. random variables, and whenever the exact value of a quantity is unknown, the distribution is assumed to have a nonzero variance. Thus, the concept of error estimate has a well-defined and intuitive probabilistic interpretation in this approach. Bayesian statistics has proved to be a successful approach in various fields of engineering. In this thesis, not only the measurements but also the user position is modeled as a random variable; every possible position is linked with a number that describes the probability of the UE to be located in that particular spot.

In this thesis the form of the RM is parametric, so the learning data are summarized by a relatively small number of statistics that are estimated using the data. These

statistics are then used in positioning as measurement model parameters. Since the parameter values are, however, estimated from the learning data rather than some physical or otherwise known constants, they are dynamical and imprecise in nature. This feature is emphasized by the fact that positioning is done using the existing infrastructure; no reconfiguring of the existing communication networks is assumed in this thesis. Therefore, in this thesis the measurement model parameters are also random variables that have have nonzero variance. This is shown to have a significant effect on the estimation of the user position and its uncertainty.

The strength of the Bayesian methodology becomes apparent, when there are several measurements whose information contents are to be merged to obtain the optimal estimate. These measurements may have completely different statistical models. Examples of different measurement models are a direct position measurement with multivariate Gaussian noise, a distance measurement with univariate Gaussian noise, a velocity measurement and a direction measurement.

Another typical example of measurement combination is a time series, which may contain more information on the position at a certain time instant than only the measurement of this single time instant. This leads to a state space model that consists of motion and measurement models, both being stochastic models. Bayesian statistics provides a very general framework for the time series problem and state space model. A recursive algorithm that estimates the state given the previous state and the new measurements is called a filter.

The theoretical scope of this thesis is to study the use of parametric estimation methods and unknown measurement model parameters in positioning. Concerning the application, RSS-based positioning with wireless networks, this thesis also presents methods for using RSS measurements in positioning and evaluates their efficiency with real data tests.

Chapter 2 of this thesis provides mathematical background for three mathematical methods that are typical in Bayesian estimation: grid, Metropolis–Hastings sampler (MH) and Gauss–Newton algorithm (GN). The fundamental theory of Bayesian time series estimation involving unknown static parameters is presented in Chapter 3.

In Chapter 4 the presented algorithms are formulated for the localization problem. A GN-based algorithm is presented for dynamic estimation of the model parameters and their precisions for each BS. The input is a learning data set collected at known positions. Furthermore, methods for positioning with the estimated model parameters are presented. The proposed positioning algorithms are MH sampler and GN algorithm. The grid method, in which the probability density values are computed in a predefined set of positions, is used as a reference algorithm.

In Chapter 5 the performance of each method is evaluated by collecting test sets of data from both outdoor and indoor environments. For each of the methods, two

versions are compared: The first one uses point estimates for the model parameters and assumes them to be accurate. The second version assumes the parameters to follow specified probabilistic model in which uncertainty is involved in the parameter values. The advantage of RSS measurements is measured by comparing the results with a cell-ID-based method that does not use RSS but relies only on the list of observed BSs [19]. In the outdoor tests, the used network is a WCDMA (Wideband Code Division Multiple Access) cellular network in Tampere, Finland [45, Ch. 3]. In the indoor tests, the network of WLAN access points inside one campus building is used. In the indoor case, the measurements are filtered to achieve reliable results. Chapter 6 presents the conclusions of the thesis and some ideas for future research.

Chapter 2

Static estimation

This section considers Bayesian inference of a system that is observed by noisy measurements. In the Bayesian philosophy, the estimation theory is based on two random variables: the state that characterizes the system and the measurement that represents the observable properties of the system. In this thesis, the state is divided into two components, $\boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{x}}}$ being the target of estimation and $\boldsymbol{a} \in \mathbb{R}^{N_{\boldsymbol{a}}}$ including system parameters whose values are uninteresting but whose uncertainty has to be taken into account in the estimation process. Measurement vector $\boldsymbol{y} \in \mathbb{R}^{N_{\boldsymbol{y}}}$ includes the data that are received and that are related to the state through a specified measurement model function. The aim of the estimation process is to infer the pdf (probability density function) of the state given a realization of the measurement vector $p(x \mid \boldsymbol{y} = y)$, where the argument of the pdf indicates the random variable whose distribution the pdf represents. In this thesis, this conditional pdf is denoted shortly with $p(x \mid y)$.

In Bayesian statistical terminology, the prior distribution is the pdf of the state which comprises all the information on the state that exists before the actual measurement is received. The joint prior distribution of $\boldsymbol{x}$ and $\boldsymbol{a}$ is denoted by $p(x, a)$. In this thesis, the state components $\boldsymbol{x}$ and $\boldsymbol{a}$ are assumed independent *a priori*. The measurements are included using the measurement model $p(y \mid \boldsymbol{x} = x, \boldsymbol{a} = a)$, in short $p(y \mid x, a)$. When this expression is treated as a function of the state, it is called the likelihood function and denoted with $\mathcal{L}(x, a) = p(y \mid \boldsymbol{x} = x, \boldsymbol{a} = a) = p(y \mid x, a)$.

The pdf $p(x, a \mid y)$ is called the posterior of the (complete) state. Posterior pdf comprises all the information of the state given the prior information and the measurement information. The mutual relation of these functions is expressed in the following theorem that is very important in the Baysian statistical theory.

Theorem 2.1. ***(Bayes' rule)***

$$p(x, a \mid y) = \frac{p(y \mid x, a)p(x, a)}{\iint p(y \mid x, a)p(x, a)\,\mathrm{d}a\,\mathrm{d}x} \tag{2.1}$$

Proof. By the definition of conditional pdf

$$p(x, a \mid y) = \frac{p(x, a, y)}{p(y)} = \frac{p(y \mid x, a)p(x, a)}{p(y)},$$

which is again obtained using the definition of conditional pdf. Thus, since

$$p(y) = \iint p(x, a, y)\,\mathrm{d}a\,\mathrm{d}x = \iint p(y \mid x, a)p(x, a)\,\mathrm{d}a\,\mathrm{d}x,$$

the relation of the theorem holds. □

The Bayes' rule is conventionally expressed in the form $p(x, a \mid y) \propto p(y \mid x, a)p(x, a)$, where symbol $\propto$ indicates equality except for a multiplier that is constant with respect to the posterior density variables x and a.

The following theorem gives the theoretical background for combining several measurements recursively. According to it, the posterior distribution can be used as a prior, when a new measurement is received and the posterior is updated.

Theorem 2.2. ***(Updating the posterior)*** *If random variables $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are conditionally independent given $\boldsymbol{x}$ and $\boldsymbol{a}$, then*

$$p(x, a \mid y_1, y_2) = \frac{p(y_2 \mid x, a)p(x, a \mid y_1)}{\int p(y_2 \mid x, a)p(x, a \mid y_1)\,\mathrm{d}a\,\mathrm{d}x}. \tag{2.2}$$

Proof. By Theorem 2.1

$$\begin{aligned} p(x, a \mid y_1, y_2) &\propto p(y_2 \mid x, a, y_1)p(x, a \mid y_1) \\ &= p(y_2 \mid x, a)p(x, a \mid y_1), \end{aligned}$$

where the last equality follows from the conditional independence of $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ given $\boldsymbol{x}$ and $\boldsymbol{a}$. Since the denominator in (2.1) is by definition the normalization constant of the pdf, this proves the statement. □

A very commonly used measurement model is the model with Gaussian prior, nonlinear measurement function and additive Gaussian measurement errors. For this model, the measurement equation can be expressed as

$$\boldsymbol{y} = h(\boldsymbol{x}, \boldsymbol{a}) + \boldsymbol{v}, \tag{2.3}$$

where h is possibly nonlinear, Borel measurable measurement function, and $\boldsymbol{v} \sim \mathrm{N}(0, \mathrm{R})$ is zero-mean Gaussian random variable that is independent of the state components $\boldsymbol{x}$ and $\boldsymbol{a}$. Formulated with probability distributions, the measurement model is

$$\begin{aligned} p(x, a) &= \mathrm{N}\left(x \mid \hat{x}, \hat{\Sigma}_{\boldsymbol{x}}\right) \cdot \mathrm{N}\left(a \mid \hat{a}, \hat{\Sigma}_{\boldsymbol{a}}\right) \\ p(y \mid x, a) &= \mathrm{N}\left(y \mid h(x, a), \mathrm{R}\right) \end{aligned} \tag{2.4}$$

The estimation methods that are described in this chapter are the grid algorithm, Metropolis–Hastings sampler and Gauss–Newton algorithm. Regarding the Gauss–Newton algorithm, the results of this thesis could be extended to Student-t distributed measurement noises [30]. Other classes of measurement noise distributions are usually tractable only by the grid method or by Monte Carlo methods, to which class the Metropolis–Hastings algorithm belongs.

2.1 Grid algorithm

The mathematically simplest but generally computationally heaviest estimation algorithm is the grid method. In this method the estimation space is a bounded subset of the complete state-space that is assumed to include almost all posterior probability mass. The grid is a set of regularly located points with a predefined density [40]. The posterior pdf for the state components $\boldsymbol{x}$ in the model (2.4) is thus obtained using the Bayes' rule (2.1) and marginalisation of the full posterior:

$$\begin{aligned} p(x \mid y) &= \int p(x, a \mid y) \,\mathrm{d}a \\ &\propto \int p(y \mid x, a) p(x, a) \,\mathrm{d}a = \int p(y \mid a, x) p(x) p(a) \,\mathrm{d}a. \end{aligned} \tag{2.5}$$

The normalisation constant does not have to be computed, since it is the same for all the grid points. If the marginalisation integral cannot be evaluated analytically, i.e. if there is no closed elementary function representation for it, the integral can be approximated with some numerical methods [28]. This thesis uses the Monte Carlo integration which is a numerical integration method that is based on generating pseudo-random numbers [18, Ch. 4].

Grid-based methods tend to be computationally inefficient especially if the state-space is multidimensional or the grid area cannot be bounded tightly enough. Furthermore, the grid area should not be bounded too tightly, since this could distort the estimate. [40]

2.2 Metropolis–Hastings algorithm

One solution for computational inefficiency problems of grid solvers is the class of Monte Carlo (MC) algorithms. They are algorithms that use pseudo-random numbers. Pseudo-random numbers are generated by computers using deterministic algorithms that have been shown to produce sequences that follow the specified distribution. There are sophisticated algorithms for generating sets of pseudo-random numbers from the uniform distribution between 0 and 1, and for some of the most common probability distributions, such as the multivariate normal distribution, there are simple analytical formulas for converting these sets to follow other distributions. [18, Ch. 3]

Markov chain Monte Carlo methods (MCMC) are Monte Carlo algorithms that generate realizations of a certain type of random variable sequences called Markov processes (See Definition 3.2). The Markov process simulated by the MCMC method should characterize the problem in a certain sense. The Metropolis–Hastings (MH) sampling algorithm is an MCMC algorithm that can, at least in principle, generate a random sample from any probability distribution. Unlike some other MC algorithms, the MH algorithm generates equally weighted samples, so the locations of the samples follow the target distribution as such. [24]

In this context the posterior pdf (or any probability distribution from which the samples are drawn) is called the target distribution and denoted with $\pi : \mathbb{R}^{N_{\boldsymbol{x}}} \to \mathbb{R} : \pi(x)$. The support of a function f is defined to be the set $\operatorname{supp}(f) = \{x \in \mathcal{D}(f) \mid f(x) \neq 0\}$. Let the proposal distribution of the MH algorithm be a probability distribution $x \mapsto q(x \mid y)$ for which $\operatorname{supp}(\pi) \subset \operatorname{supp}(q(\cdot \mid y))$ holds and from which it is straightforward to generate pseudo-random numbers for all $y \in \operatorname{supp}(\pi)$.

In the MH algorithm a chain of numbers is generated, and the elements of the chain are called states. MH algorithm consists of the following steps: Assume that the current state of the Markov process is $x \in \operatorname{supp}(\pi)$. Proposal value x' is generated from the proposal distribution $q(\cdot \mid x)$. Let the MH ratio be

$$r(x, x') = \frac{\pi(x')q(x \mid x')}{\pi(x)q(x' \mid x)}.$$

A pseudo-random number u is generated from the standard uniform distribution $\operatorname{Unif}(0, 1)$. If $r(x, x') > u$ holds, the proposal value is accepted and the value of the

next state of the chain is set to be the proposal value. In case the proposal value is rejected, the value of the next state is set to equal the current state. This iteration is repeated until some termination criterion is met.

Note that it is only the ratio of the target pdf values that is concerned, so the target distribution function may be unnormalized. This is a notable property, since posterior pdfs in Bayesian statistics tend to be so complicated that analytical normalization would be impossible and numerical normalization would be at least as demanding a task as the actual estimation.

Since the first values of the computed chain are highly dependent on the given initial value, a number of them is conventionally ignored in the estimation. These states are called the burn-in period, and the length of the burn-in period is denoted with N_b. This cannot be determined exactly, but it is usually found out by eye by examining the sampling history plots produced by the algorithm. The MH algorithm is described in detail in Algorithm 2.1. In the algorithm listing, the estimated mean of the target distribution is estimated by $\hat{x}^+$ and the estimated covariance matrix by $\hat{\Sigma}_{\boldsymbol{x}}^+$.

Algorithm 2.1 Metropolis–Hastings algorithm

1. Let x_0 be the initial sample and set $k := 1$. Let N be the sample size.
2. Generate $x'_k \leftarrow q(\cdot \mid x_{k-1})$.
3. Compute
$$r(x_{k-1}, x'_k) := \frac{\pi(x'_k)\ q(x_{k-1} \mid x'_k)}{\pi(x_{k-1})\ q(x'_k \mid x_{k-1})}.$$
Generate $u \leftarrow \mathrm{Unif}\,(0,1)$. If $r(x_{k-1}, x'_k) > u$, set $x_k := x'_k$. Otherwise, set $x_k := x_{k-1}$.
4. If $k < N$, set $k := k + 1$ and go to step 2. Otherwise, set
$$\hat{x}^+ := \tfrac{1}{N-N_\mathrm{b}+1} \sum_{k=N_\mathrm{b}}^{N} x_k\ , \quad \hat{\Sigma}_{\boldsymbol{x}}^+ := \tfrac{1}{N-N_\mathrm{b}+1} \sum_{k=N_\mathrm{b}}^{N} (x_k - \hat{x})(x_k - \hat{x})^\mathrm{T},$$
where N_b is the length of the burn-in period.

The theorem that guarantees asymptotic convergence of random sequences to the true moments of target distributions is called the law of large numbers [29, Ch. 2.3]. The Markov processes related to the MH algorithms are autocorrelated, so they do not fulfill the assumptions of the most well-known law of large numbers. However, a similar result can be proved to this specific group of Markov processes.

Theorem 2.3. ***(Law of large numbers for MH algorithm)*** *Let $\{\boldsymbol{x}_k\}$ be a Markov process simulated by the MH algorithm with π as a target distribution. Then for every π-integrable Borel-measurable function $f : \mathbb{R}^{N_{\boldsymbol{x}}} \to \mathbb{R}^{N_f}$*

$$\forall x_0 \in \mathbb{R}^{N_{\boldsymbol{x}}} : \ \mathrm{P}\left(\lim_{n\to\infty} \tfrac{1}{n+1} \sum_{i=0}^{n} f(\boldsymbol{x}_k) = \int f(x)\pi(x)\,\mathrm{d}x \mid \boldsymbol{x}_0 = x_0\right) = 1, \tag{2.6}$$

holds, that is, the arithmetic mean of the function values converges almost surely to the π-integral of the function.

Proof. Omitted. See [18, Ch. 11]. □

Let us now consider the model (2.4). Let the complete state be $\boldsymbol{\xi} = \begin{bmatrix}\boldsymbol{x}^{\mathrm{T}} & \boldsymbol{a}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$. Then the target distribution is the posterior pdf $\pi(\xi) = \pi(x, a) \propto p(y \mid x, a) \cdot p(x, a)$. The marginal distribution of $\boldsymbol{x}$ can be extracted by considering only the component sequence $\{x_k\}_{k=0}^{N}$ of the sequence $\{\xi_k\}_{k=0}^{N}$.

In Figure 2.1 there is an example of positioning with an MH sample with flat prior. The measurements are noisy distance measurements. The 68% uncertainty ellipse is based on the Gaussianity approximation of the posterior. Since the initial position is relatively far away from the likelihood peaks, the first samples are far from the true likelihood. However, they do not influence the estimation, since the burn-in samples are discarded.

2.3 Gauss–Newton algorithm

The least squares problem is formulated as finding

$$\operatorname*{argmin}_{x} \Phi(x), \tag{2.7}$$

where $\Phi : \mathbb{R}^{N_x} \to \mathbb{R}^{+}$ is a function of form $\Phi(x) = \frac{1}{2}\left\|f(x)\right\|^2 = \frac{1}{2}\sum_{i=1}^{N_f} f_i(x)^2$, where $f : \mathbb{R}^{N_x} \to \mathbb{R}^{N_f}$, $N_f \geq N_x$ is a known function. If f is an affine function, that is, it can be presented in form $f(x) = \mathrm{A}x + b$, where A is a matrix and b a vector, the optimization problem is a linear least squares problem. Otherwise it is a nonlinear least squares problem.

Gauss–Newton algorithm (GN), also known as the Iterative Reweighted Least Squares method is an iterative optimization method for nonlinear least squares estimation problems. With suitable measurement models, iterative state estimation methods can be as accurate as any closed form solution but simpler and easier to implement [39].

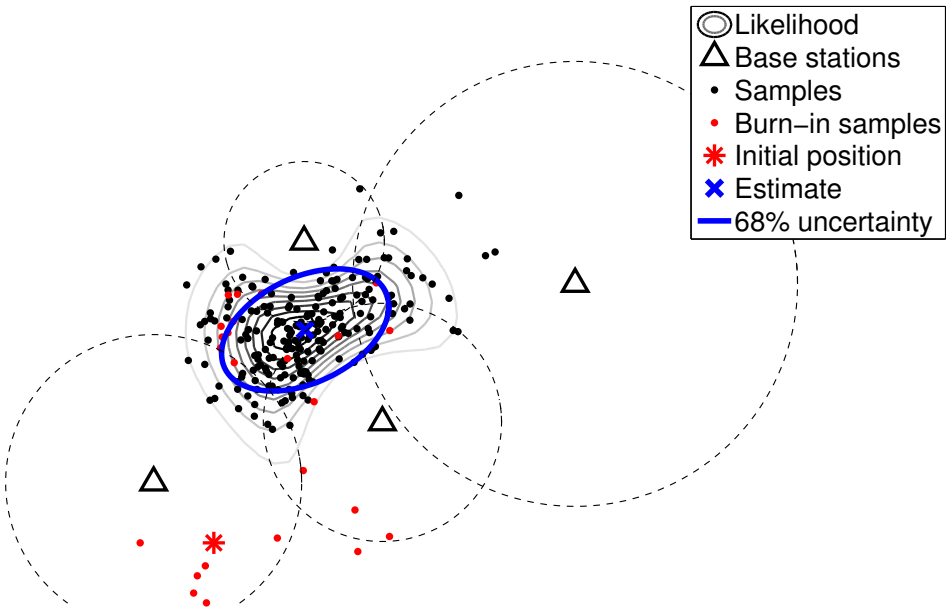

Figure 2.1: *MH sample of four noisy distance measurements. There are 1000 samples of which 50 samples form the burn-in that is discarded from the estimation. Compare with Figure 2.2 in which the same likelihood is estimated with the same initial position using the Gauss–Newton algorithm.*

The derivation of the GN method is based on iterative linearization and analytical solving of the linearized problem. The linearization approximation of the objective function at point x_0 is

$$f(x) \approx f(x_0) + \mathrm{J}(x_0)(x - x_0), \tag{2.8}$$

where $\mathrm{J} = \frac{\mathrm{d}f}{\mathrm{d}x}$ is the derivative matrix of function f also known as the Jacobian matrix. Thus, the problem can be approximated by the linear least squares problem

$$\begin{aligned} \operatorname*{argmin}_x \Phi(x) &\approx \operatorname*{argmin}_x \frac{1}{2} \left\| f(x_0) + \mathrm{J}_0(x - x_0) \right\|^2 \\ &= \operatorname*{argmin}_x \frac{1}{2} \left\| \mathrm{J}_0 x - (\mathrm{J}_0 x_0 - f(x_0)) \right\|^2, \end{aligned} \tag{2.9}$$

where x_0 is a linearization point and $\mathrm{J}_0 = \mathrm{J}(x_0)$. The linearized minimization problem can be solved using the following theorem.

Theorem 2.4. ***(Linear least squares)*** *Assume that $x \in \mathbb{R}^n$ and assume that $y \in \mathbb{R}^m$ is a known vector with $m \geq n$ and that $\mathrm{A} \in \mathbb{R}^{m \times n}$ is a known matrix. Then there exists a solution to the problem*

$$\operatorname*{argmin}_x \frac{1}{2} \left\| \mathrm{A}x - y \right\|^2. \tag{2.10}$$

It is unique if and only if $\operatorname{rank}(\mathrm{A}) = n$. *Furthermore, if the solution exists and is unique, it is*

$$\underset{x}{\operatorname{argmin}} \frac{1}{2} \|\mathrm{A}x - y\|^2 = (\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}}y. \tag{2.11}$$

Proof. Let us first reformulate the expression to be minimized:

$$\|\mathrm{A}x - y\|^2 = \left\|\mathrm{A}x - \mathrm{A}\mathrm{A}^{\dagger}y + \mathrm{A}\mathrm{A}^{\dagger}y - y\right\|^2 = \left\|(\mathrm{A}x - \mathrm{A}\mathrm{A}^{\dagger}y) + (\mathrm{A}\mathrm{A}^{\dagger}y - y)\right\|^2,$$

where $\mathrm{A}^{\dagger}$ is the Moore–Penrose pseudo-inverse of matrix A. The two summation terms are orthogonal, since

$$\begin{aligned}
(\mathrm{A}x - \mathrm{A}\mathrm{A}^{\dagger}y)^{\mathrm{T}}(\mathrm{A}\mathrm{A}^{\dagger}y - y) &= (x - \mathrm{A}^{\dagger}y)^{\mathrm{T}}\mathrm{A}^{\mathrm{T}}(\mathrm{A}\mathrm{A}^{\dagger} - \mathrm{I})y \\
&\overset{*}{=} (x - \mathrm{A}^{\dagger}y)^{\mathrm{T}}\mathrm{A}^{\mathrm{T}}(\mathrm{A}\mathrm{A}^{\dagger} - \mathrm{I})^{\mathrm{T}}y \\
&= (x - \mathrm{A}^{\dagger}y)^{\mathrm{T}}(\mathrm{A}\mathrm{A}^{\dagger}\mathrm{A} - \mathrm{A})^{\mathrm{T}}y \\
&\overset{*}{=} (x - \mathrm{A}^{\dagger}y)^{\mathrm{T}}\mathrm{O}^{\mathrm{T}}y \\
&= 0
\end{aligned}$$

where the equalities marked with an asterisk are based on the characteristic properties of Moore–Penrose pseudo-inverse. Hence, by the Pythagorean theorem

$$\|\mathrm{A}x - y\|^2 = \left\|\mathrm{A}x - \mathrm{A}\mathrm{A}^{\dagger}y\right\|^2 + \left\|\mathrm{A}\mathrm{A}^{\dagger}y - y\right\|^2$$

holds. Thus, the point $x^* = \mathrm{A}^{\dagger}y$ minimizes the objective function so the solution to the linear least squares problem exists. Furthermore, every x_v^* that satisfies

$$x_v^* = \mathrm{A}^{\dagger}y + v,\ v \in \mathcal{N}(\mathrm{A}),$$

where $\mathcal{N}(\mathrm{A})$ is the kernel of matrix A, minimizes the objective function, so x^* is a unique minimizer if and only if $\mathcal{N}(\mathrm{A}) = \{0\}$. By basic linear algebra, $\mathcal{N}(\mathrm{A}) = \{0\}$ if and only if $\operatorname{rank}(\mathrm{A}) = n$.

It remains to show that if $\operatorname{rank}(\mathrm{A}) = n$, then $\mathrm{A}^{\dagger} = (\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}}$. Since, $\operatorname{rank}(\mathrm{A}^{\mathrm{T}}\mathrm{A}) = \operatorname{rank}(\mathrm{A}) = n$, the inverse $(\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}$ exists. Furthermore, the following four properties, which characterise the Moore–Penrose pseudo-inverse of the matrix A hold:

1. $\mathrm{A}\left((\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}}\right)\mathrm{A} = \mathrm{A}(\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}(\mathrm{A}^{\mathrm{T}}\mathrm{A}) = \mathrm{A}$
2. $\left((\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}}\right)\mathrm{A}\left((\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}}\right) = (\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}(\mathrm{A}^{\mathrm{T}}\mathrm{A})(\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}} = (\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}}$
3. $\left(\mathrm{A}\left((\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}}\right)\right)^{\mathrm{T}} = \mathrm{A}((\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1})^{\mathrm{T}}\mathrm{A}^{\mathrm{T}} = \mathrm{A}(\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}}$, since $(\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}$ is symmetric because $\mathrm{A}^{\mathrm{T}}\mathrm{A}$ is symmetric
4. $\left((\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}}\right)\mathrm{A} = \mathrm{I}$, which is symmetric

Thus, $\mathrm{A}^\dagger = (\mathrm{A}^\mathrm{T}\mathrm{A})^{-1}\mathrm{A}^\mathrm{T}$, if $\operatorname{rank}(\mathrm{A}) = n$. □

By Theorem 2.4, a solution to the linearized problem (2.9) exists and is unique if and only if the J_0 is full-rank, and then the solution is

$$\begin{aligned}\operatorname*{argmin}_x \frac{1}{2}\left\|f(x_0) + \mathrm{J}_0(x - x_0)\right\|^2 &= (\mathrm{J}_0{}^\mathrm{T}\mathrm{J}_0)^{-1}\mathrm{J}_0{}^\mathrm{T}(\mathrm{J}_0 x_0 - f(x_0)) \\ &= (\mathrm{J}_0{}^\mathrm{T}\mathrm{J}_0)^{-1}(\mathrm{J}_0{}^\mathrm{T}\mathrm{J}_0)x_0 - (\mathrm{J}_0{}^\mathrm{T}\mathrm{J}_0)^{-1}\mathrm{J}_0{}^\mathrm{T} f(x_0) \\ &= x_0 - (\mathrm{J}_0{}^\mathrm{T}\mathrm{J}_0)^{-1}\mathrm{J}_0{}^\mathrm{T} f(x_0).\end{aligned} \tag{2.12}$$

In the GN algorithm, after computing the optimum of the linearized system, the system is linearized again in the solution point and the relinearized system is solved. This iteration is continued until the solution stabilizes, that is, the step size $\Delta x = -(\mathrm{J}^\mathrm{T}\mathrm{J})^{-1}\mathrm{J}^\mathrm{T} f(x)$ becomes small enough. In practice the iteration is usually terminated, when the norm of the step size is under some threshold or when the number of iterations becomes too high.

The convergence of the standard GN iteration is not guaranteed [9, Ch. 9.2.]. To ensure convergence to a minimum at least locally the step size must be adapted so that the value of the objective function decreases rapidly enough. This algorithm is called the damped GN algorithm, and it is presented in detail in Algorithm 2.2. With certain conditions, damping guarantees that the algorithm converges to a local minimum if the initialization point is close enough. This statement is proved by Lemma 2.5 and Theorem 2.6. Lemma 2.5 and Theorem 2.6 use the convergence results of Appendix C that are based on the ideas of [8, Ch. 1.2].

Algorithm 2.2 Damped Gauss–Newton algorithm

1. Let x_0 be the initial value and set $k := 0$. Set the configuration parameters $s > 0$ and $\delta > 0$.
2. Compute the Jacobian matrix $\mathrm{J}_k := \frac{\mathrm{d}f}{\mathrm{d}x}(x_k)$.
3. Set $(\Delta x)_k := -(\mathrm{J}_k{}^\mathrm{T}\mathrm{J}_k)^{-1}\mathrm{J}_k{}^\mathrm{T} f(x_k)$.
4. Set $\alpha_k := \operatorname*{argmin}_{\alpha\in(0,s]} \left\|f(x_k + \alpha\cdot(\Delta x)_k)\right\|$.
5. Set $x_{k+1} := x_k + \alpha_k \cdot (\Delta x)_k$.
6. If $\|x_{k+1} - x_k\| < \delta$, stop. Otherwise, set $k := k + 1$ and go to step 2.

Lemma 2.5. *Assume that the function f is continuously differentiable and for the Jacobian matrix J the set*

$$\{\sigma \in \mathbb{R}_0^+ \mid \sigma \text{ is a singular value of } \mathrm{J}(x) \text{ for some } x \in \mathbb{R}^{N_x}\} \tag{2.13}$$

is bounded above and bounded away from zero. Then the following properties hold for the damped GN algorithm:

1. *Objective function Φ is continuously differentiable.*
2. *Gauss–Newton step $(\Delta x)_k$ is of form $-\mathrm{D}_k\Phi'(x_k)^{\mathrm{T}}$ where the set*

   $$\{\sigma \in \mathbb{R}_0^+ \mid \sigma \text{ is an eigenvalue of } \mathrm{D}_k \text{ for some } k \in \mathbb{N}_0\}$$

   is bounded above and bounded away from zero.
3. *The sequence $\{\alpha_k\}$ is bounded.*
4. *There exists a constant $c \in \mathbb{R}^+$ such that $\|(\Delta x)_k\| \leq c\,\|\Phi'(x_k)\|$ for all indices k.*
5. *Inequality $\Phi(x_{k+1}) \leq \Phi(x_k)$ holds for all indices k.*

Proof. The statements are proved one by one.

1. As a composition of functions f and $\frac{1}{2}\|\cdot\|^2$, which are both continuously differentiable, the objective function Φ is continuously differentiable.
2. By the chain rule of differentiation

   $$\Phi'(x) = \frac{1}{2} \cdot 2\,\|f(x)\|\,\frac{\mathrm{d}\,\|f(x)\|}{\mathrm{d}x} = \|f(x)\| \cdot \frac{f(x)^{\mathrm{T}}}{\|f(x)\|}\frac{\mathrm{d}f(x)}{\mathrm{d}x} = f(x)^{\mathrm{T}}\mathrm{J}(x).$$

   Hence,

   $$(\Delta x)_k = -({\mathrm{J}_k}^{\mathrm{T}}\mathrm{J}_k)^{-1}\Phi'(x_k)^{\mathrm{T}}.$$

   Since the singular values of each J_k are bounded away from zero, matrix ${\mathrm{J}_k}^{\mathrm{T}}\mathrm{J}_k$ is invertible, and matrix $({\mathrm{J}_k}^{\mathrm{T}}\mathrm{J}_k)^{-1}$ is positive definite. From the assumption (2.13) it follows that the set

   $$\{\sigma \in \mathbb{R}_0^+ \mid \sigma \text{ is an eigenvalue of } ({\mathrm{J}_k}^{\mathrm{T}}\mathrm{J}_k)^{-1} \text{ for some } k \in \mathbb{N}_0\}$$

   is bounded above and bounded away from zero.
3. This follows straight from the definition of α_k in the damped GN algorithm.

4. Since the set

$$\{\sigma \in \mathbb{R}_0^+ \mid \sigma \text{ is an eigenvalue of } (\mathrm{J}_k{}^{\mathrm{T}}\mathrm{J}_k)^{-1} \text{ for some } k \in \mathbb{N}_0\}$$

is bounded above, the set of matrix norms $\{\|(\mathrm{J}_k{}^{\mathrm{T}}\mathrm{J}_k)^{-1}\| \mid k \in \mathbb{N}_0\}$ is also bounded above. Thus,

$$\begin{aligned}\|(\Delta x)_k\| &= \left\|(\mathrm{J}_k{}^{\mathrm{T}}\mathrm{J}_k)^{-1}\mathrm{J}_k{}^{\mathrm{T}} f(x_k)\right\| \\ &= \left\|(\mathrm{J}_k{}^{\mathrm{T}}\mathrm{J}_k)^{-1}\Phi'(x_k)^{\mathrm{T}}\right\| \\ &\leq \left\|(\mathrm{J}_k{}^{\mathrm{T}}\mathrm{J}_k)^{-1}\right\| \|\Phi'(x_k)\| \\ &\leq c\,\|\Phi'(x_k)\|\,,\end{aligned}$$

where $\|\cdot\|$ is the 2-norm of vector or matrix depending on the context and c is a finite real number that does not depend on the index k.

5. By step 2 and by Lemma C.3 the GN step $(\Delta x)_k$ is a descent direction, so the statement follows from Lemma C.2.

□

Theorem 2.6. ***(Local convergence of damped GN)*** *Assume that the function f is continuously differentiable and for the Jacobian matrix $\mathrm{J}(x)$ the set*

$$\{\sigma \in \mathbb{R}_0^+ \mid \sigma \textit{ is a singular value of } \mathrm{J}(x) \textit{ for some } x \in \mathbb{R}^{N_x}\}$$

is bounded above and bounded away from zero. Let x^* be a local minimum of the objective function $\Phi = \frac{1}{2}\|f(\cdot)\|^2$. Then there exists an open set $S \ni x^*$ such that if $x_0 \in S$, the damped Gauss–Newton algorithm converges to x^*.

Proof. By items 1 and 2 of Lemma 2.5 the damped GN algorithm fulfills the assumptions of Proposition C.4. Thus, if a subsequence of the GN sequence converges, it converges to a stationary point of Φ.

Furthermore, by the aforementioned property and by items 3, 4 and 5 of Lemma 2.5 the damped GN algorithm fulfills the assumptions of Proposition C.5. This proves the statement. □

For simplicity, the exact line search is not used in this thesis. Instead, the damping coefficient α is chosen to be the largest element of the sequence $\{2^{-i}\}_{i=0}^{\infty}$ that ensures the objective function value to decrease.

Let us now consider the model (2.4), denoting the complete state with $\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{a} \end{bmatrix}$. The GN optimization can now be applied for finding an estimate for the MAP (*maximum a posteriori*, posterior mode) value

$$\begin{aligned}
&\underset{\xi}{\operatorname{argmax}}\ p(\xi \mid y) \\
=&\underset{\xi}{\operatorname{argmax}}\ p(\xi)p(y \mid \xi) \\
=&\underset{\xi}{\operatorname{argmax}}\ \left(\mathrm{N}\left(\xi \mid \hat{\xi}, \hat{\Sigma}_{\boldsymbol{\xi}}\right) \cdot \mathrm{N}\left(y \mid h(\xi), \mathrm{R}\right)\right) \\
=&\underset{\xi}{\operatorname{argmin}}\ \left((\xi - \hat{\xi})^{\mathrm{T}} \hat{\Sigma}_{\boldsymbol{\xi}}^{-1} (\xi - \hat{\xi}) + (y - h(\xi))^{\mathrm{T}} \mathrm{R}^{-1} (y - h(\xi))\right) \\
=&\underset{\xi}{\operatorname{argmin}}\ \left\| \begin{bmatrix} \sqrt{\hat{\Sigma}_{\boldsymbol{\xi}}^{-1}}(\xi - \hat{\xi}) \\ \sqrt{\mathrm{R}^{-1}}(y - h(\xi)) \end{bmatrix} \right\|^2,
\end{aligned}$$

where $\sqrt{\mathrm{A}}$ denotes a matrix for which $\sqrt{\mathrm{A}}^{\mathrm{T}}\sqrt{\mathrm{A}} = \mathrm{A}$. Thus, the optimization function and the corresponding Jacobian matrix are

$$f(\xi) = \begin{bmatrix} \sqrt{\hat{\Sigma}_{\boldsymbol{\xi}}^{-1}}(\xi - \hat{\xi}) \\ \sqrt{\mathrm{R}^{-1}}(y - h(\xi)) \end{bmatrix}, \quad \mathrm{J}(\xi) = \begin{bmatrix} \sqrt{\hat{\Sigma}_{\boldsymbol{\xi}}^{-1}} \\ -\sqrt{\mathrm{R}^{-1}}\mathrm{H}(\xi) \end{bmatrix},$$

where $\mathrm{H}(\xi)$ is the Jacobian matrix of the measurement model function h. Thus, one GN step is

$$\begin{aligned}
(\Delta\xi)_k &= -\left(\mathrm{J}_k{}^{\mathrm{T}}\mathrm{J}_k\right)^{-1}\mathrm{J}_k{}^{\mathrm{T}} f(\xi_k) \\
&= -\left(\hat{\Sigma}_{\boldsymbol{x}}^{-1} + \mathrm{H}(\xi_k)^{\mathrm{T}}\mathrm{R}^{-1}\mathrm{H}(\xi_k)\right)^{-1} \cdot \left(\hat{\Sigma}_{\boldsymbol{\xi}}^{-1}(\xi_k - \hat{\xi}) + \mathrm{H}(\xi_k)^{\mathrm{T}}\mathrm{R}^{-1}(h(\xi_k) - y)\right).
\end{aligned}$$

The covariance matrix of the posterior can be approximated by linearizing the system at the MAP-solution of the GN algorithm. Note, however, that the approximation is valid only if the measurement equation is only slightly nonlinear. The covariance matrix of the linearized system can be determined using Lemma B.3 of the product of Gaussian distributions:

$$\operatorname{var}(\boldsymbol{\xi} \mid y) \approx \left(\hat{\Sigma}_{\boldsymbol{\xi}}^{-1} + \mathrm{H}^{\mathrm{T}}\mathrm{R}^{-1}\mathrm{H}\right)^{-1},$$

where H is the Jacobian of the measurement model function at the linearization point.

The linearization approximation also implies that the posterior is a Gaussian distribution, for which the MAP value equals the expectation value. Therefore, for

almost linear measurement equations the posterior approximation returned by the GN algorithm is

$$p(\xi \mid y) \approx \mathrm{N}\left(\xi \mid \xi_k, \left(\hat{\Sigma}_{\boldsymbol{\xi}}^{-1} + \mathrm{H}(\xi_k)^{\mathrm{T}}\mathrm{R}^{-1}\mathrm{H}(\xi_k)\right)^{-1}\right), \tag{2.14}$$

where ξ_k is final state estimate of the GN algorithm. Since the marginal distributions of a multivariate normal distribution are normal distributions with the corresponding components of the mean and covariance matrix of the multivariate distribution as the parameters, the marginal posterior for any set of state components can be computed.

Figure 2.2 presents the damped GN sequence of the measurement setup that is the same as in Figure 2.1. The 68% uncertainty ellipse is based on the Gaussianity approximation of the posterior.

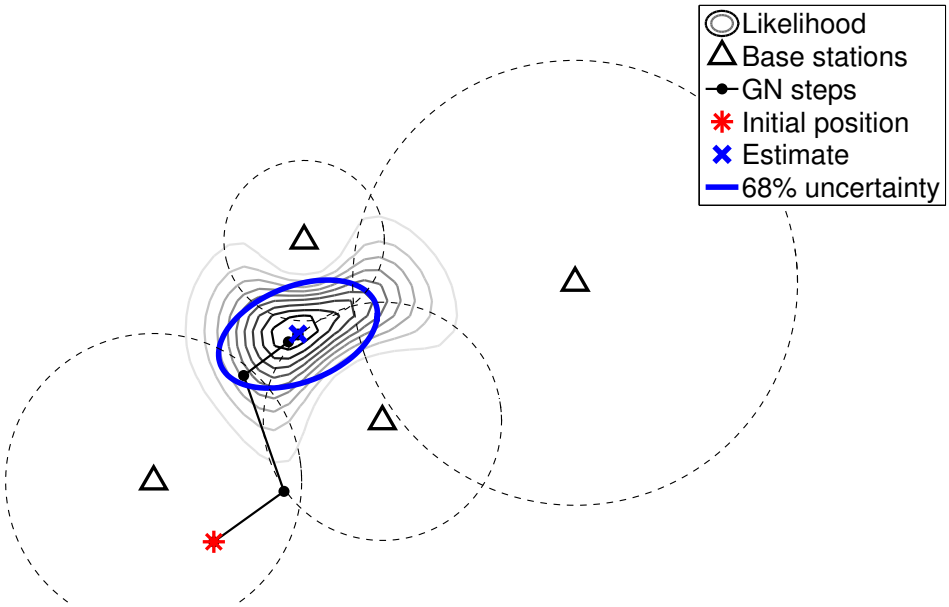

Figure 2.2: *GN steps of four noisy distance measurements. Compare with Figure 2.1 in which the same likelihood is estimated with the Metropolis–Hastings algorithm.*

Chapter 3

Time-series estimation

In the theory of statistical time series estimation, the unknown properties that characterize the system are called the state of the system. The state may be time-variant, and the time-evolution of the state is modeled statistically by a stochastic motion model. The state is measured by a noisy measurement process, and the information contained by the measurements should be fused using the motion model. The aim of Bayesian time series estimation is to infer the posterior distribution of the state for each time instant given a time series of measurements. This thesis assumes that the state includes time-invariant components, the estimation of which may require specific procedures.

To begin with, let us define some basic mathematical concepts involved in time series estimation. It is assumed that the reader is knowledgeable about the basic concepts of probability theory.

Definition 3.1. (Stochastic Process) Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and T a set of parameter values that are called time instants. A stochastic process is a mapping $\boldsymbol{x} : \Omega \times T \to \mathbb{R}^n$ such that $\boldsymbol{x}(\cdot, t)$ is random variable for every fixed $t \in T$. Furthermore, if the parameter set T is enumerable and its elements are denoted by t_k where $k \in \mathcal{I} \subset \mathbb{Z}$, the sequence $(\boldsymbol{x}(\cdot, t_k))_{k \in \mathcal{I}}$ is a discrete stochastic process and its elements are denoted by $\boldsymbol{x}_k = \boldsymbol{x}(\cdot, t_k)$. ■

A stochastic process is called a Markov process, if the distribution of the current state depends on the earlier state realizations only through the previous state realization.

Definition 3.2. (Markov Process) A discrete stochastic process $\{\boldsymbol{x}_k\}$ is a Markov process, if

$$\forall k \in \mathbb{N} : p(x_k \mid x_{0,\ldots,k-1}) = p(x_k \mid x_{k-1}).$$

■

A white process is a Markov process whose current state is independent of all the earlier states, that is, for each time step the conditional distribution equals the marginal distribution.

Definition 3.3. (White process) A discrete stochastic process $\{\boldsymbol{x}_k\}$ is a white process, if

$$\forall k \in \mathbb{N} : p(x_k \mid x_{0,\ldots,k-1}) = p(x_k).$$

■

It can also be proved that an element of a Markov process depends on the future only through the next element and an element of a white process is independent of the future.

Let us now consider a system whose hidden state can be divided into two components: dynamic and static. Let us denote the dynamic component with $\boldsymbol{x}_k \in \mathbb{R}^{N_{\boldsymbol{x}}}$. It may be time-varying, and it is the actual target of time-series estimation. The static component $\boldsymbol{a}_k \in \mathbb{R}^{N_{\boldsymbol{a}}}$ is assumed time-invariant, and it contains system parameters whose actual values are uninteresting similarly to the state component $\boldsymbol{a}$ in Chapter 2.

Because the state component $\boldsymbol{x}_k$ may be time-varying, this section introduces a statistical motion model for the state. Since the state is unknown, the Bayesian estimation theory considers it a random variable at each time instant. To allow computation with digital computers, the time space has to be discrete, but in the real positioning system the state evolution is typically a continuous process. However, certain continuous processes can be discretized so that the estimation can be performed discretely. [16, Ch. 7.2],[1, Ch. 2.2.5]. Thus, the state is modeled as a discrete stochastic process.

Let us denote the initial dynamical state with $\boldsymbol{x}_0$. In this thesis, it is assumed that the state evolution follows the discrete stochastic difference equation

$$\boldsymbol{x}_{k+1} = f_k(\boldsymbol{x}_k, \boldsymbol{w}_k), \tag{3.1}$$

which is independent of the static part of the state $\boldsymbol{a}_k$. The stochastic process $\boldsymbol{w}_k \sim \mathrm{N}(0, \mathrm{Q}_k)$ is called the process noise term, which is assumed to be a white process and independent of the initial dynamical state $\boldsymbol{x}_0$. It can be proved that the process $\{\boldsymbol{x}_k\}$ is thus a Markov process [16, Ch. 3.9]. This class of models is extensively covered by the literature and has numerous physical applications [16, Ch. 3.7]. Markovian models also allow recursive time series estimation, where only the latest state estimate and the latest measurement are needed to estimate the current state.

In this thesis, the system consists of a physical person moving in a Cartesian coordinate system without nonlinear constraints. Therefore, the motion model is assumed to be linear and the process noise is assumed to be additive. Thus, the motion model can be formulated

$$\boldsymbol{x}_{k+1} = \Phi_k \boldsymbol{x}_k + \boldsymbol{w}_k, \tag{3.2}$$

where matrix $\Phi_k \in \mathbb{R}^{N_{\boldsymbol{x}} \times N_{\boldsymbol{x}}}$ is called the state transition matrix. The system is observed by measurement process that follows the model of (2.3). Thus, the system follows the Gaussian state-space model of the form

$$\begin{aligned} \boldsymbol{x}_0 &\sim \mathrm{N}(m_0, \mathrm{P}_0) \\ \boldsymbol{a}_0 &\sim \mathrm{N}\left(\hat{a}, \hat{\Sigma}_{\boldsymbol{a}}\right) \\ \boldsymbol{x}_{k+1} &= \Phi_k \boldsymbol{x}_k + \boldsymbol{w}_k, \;\; \boldsymbol{w}_k \sim \mathrm{N}(0, \mathrm{Q}_k) \\ \boldsymbol{a}_{k+1} &= \boldsymbol{a}_k \\ \boldsymbol{y}_k &= h_k(\boldsymbol{x}_k, \boldsymbol{a}_k) + \boldsymbol{v}_k, \;\; \boldsymbol{v}_k \sim \mathrm{N}(0, \mathrm{R}_k), \end{aligned} \tag{3.3}$$

where $\{\boldsymbol{w}_k\}$ and $\{\boldsymbol{v}_k\}$ are white processes and independent of each other and independent of $\boldsymbol{x}_0$ and $\boldsymbol{a}_0$. It can be proved that the stochastic process $\{\left[\boldsymbol{x}_k{}^{\mathrm{T}}, \boldsymbol{a}_k{}^{\mathrm{T}}, \boldsymbol{y}_{k+1}{}^{\mathrm{T}}\right]^{\mathrm{T}}\}$ is also a Markov process [16, Ch. 3.9]. Since $\boldsymbol{a}$ is a random constant, it is denoted without the subscript in the next section.

3.1 Optimal Bayesian filtering equations with static state components

In the Bayesian estimation theory, optimal filtering means a recursive algorithm for solving the posterior distribution of the state given the all the measurements up to that time instant $p(x_k \mid y_{1:k})$. The filter implementations are usually two-phased:

The prediction step predicts the state using the posterior of the latest time instant and the motion model. The update step modifies the predicted estimate based on the newest measurement.

The Chapman–Kolmogorov equation for Markov processes is the mathematical background for the prediction step, thus allowing recursive state updates.

Theorem 3.4. ***(Chapman–Kolmogorov equation)*** *If the motion model is independent of* $\boldsymbol{a}$ *and the stochastic process* $\{\left[\boldsymbol{x}_k{}^{\mathrm{T}}, \boldsymbol{a}_k{}^{\mathrm{T}}, \boldsymbol{y}_{k+1}{}^{\mathrm{T}}\right]^{\mathrm{T}}\}$ *is a Markov process, the equation*

$$p(x_{k+1}, a \mid y_{1:k}) = \int p(x_{k+1} \mid x_k) p(x_k, a \mid y_{1:k}) \,\mathrm{d}x_k. \tag{3.4}$$

holds.

Proof. By the definition of conditional probability

$$\begin{aligned} p(x_{k+1}, x_k, a \mid y_{1:k}) &= p(x_{k+1} \mid x_k, a, y_{1:k}) p(x_k, a \mid y_{1:k}) \\ &= p(x_{k+1} \mid x_k) p(x_k, a \mid y_{1:k}), \end{aligned}$$

where the second equality follows from the independence of the motion model from $\boldsymbol{a}$ and from the Markov property of the sequence $\{\left[\boldsymbol{x}_k{}^{\mathrm{T}}, \boldsymbol{a}_k{}^{\mathrm{T}}, \boldsymbol{y}_{k+1}{}^{\mathrm{T}}\right]^{\mathrm{T}}\}$. Thus,

$$\begin{aligned} p(x_{k+1}, a \mid y_{1:k}) &= \int p(x_{k+1}, x_k, a \mid y_{1:k}) \,\mathrm{d}x_k \\ &= \int p(x_{k+1} \mid x_k) p(x_k, a \mid y_{1:k}) \,\mathrm{d}x_k. \end{aligned}$$

□

The following theorem is Equation (3.4) for the linear Gaussian special case.

Theorem 3.5. ***(Filter prediction for linear Gaussian model)*** *Given*

$$p(x_k \mid y_{1:k}) = \mathrm{N}\left(x_k \mid m_k^+, \mathrm{P}_k^+\right), \; p(x_{k+1} \mid x_k) = \mathrm{N}\left(x_{k+1} \mid \Phi_k x_k, \mathrm{Q}_k\right)$$

the following holds:

$$p(x_{k+1} \mid y_{1:k}) = \mathrm{N}\left(x_{k+1} \mid \Phi_k m_k^+, \Phi_k \mathrm{P}_k^+ \Phi_k{}^{\mathrm{T}} + \mathrm{Q}_k\right). \tag{3.5}$$

Proof. By Lemma B.2,

$$\begin{aligned} p(x_{k+1} \mid x_k) p(x_k \mid y_{1:k}) &= p(x_{k+1} \mid x_k, y_{1:k}) p(x_k \mid y_{1:k}) \\ &= \mathrm{N}\left(\begin{bmatrix} x_k \\ x_{k+1} \end{bmatrix} \mid \begin{bmatrix} m_k^+ \\ \Phi_k m_k^+ \end{bmatrix}, \begin{bmatrix} \mathrm{P}_k^+ & \mathrm{P}_k^k \Phi_k{}^{\mathrm{T}} \\ \Phi_k \mathrm{P}_k^+ & \Phi_k \mathrm{P}_k^+ \Phi_k{}^{\mathrm{T}} + \mathrm{Q}_k \end{bmatrix}\right) \end{aligned}$$

holds. Thus, by Theorem 3.4 and the marginalisation rule of multivariate Gaussian distributions

$$\begin{aligned} p(x_{k+1} \mid y_{1:k}) &= \int p(x_{k+1} \mid x_k) p(x_k \mid y_{1:k}) \, \mathrm{d}x_k \\ &= \mathrm{N}\left(x_{k+1} \mid \Phi_k m_k^+, \Phi_k \mathrm{P}_k^+ {\Phi_k}^{\mathrm{T}} + \mathrm{Q}_k\right), \end{aligned}$$

which proves the statement. □

The following theorem justifies the update step of the Bayesian filter.

Theorem 3.6. ***(State update)*** *For the model 3.3 the equation*

$$p(x_k, a \mid y_{1:k}) \propto p(y_k \mid x_k, a) \; p(x_k, a \mid y_{1:k-1}) \tag{3.6}$$

holds.

Proof. Since the measurement noise $\{\boldsymbol{v}_k\}$ is a white process, $\boldsymbol{y}_k$ is independent of $\boldsymbol{y}_{1:k-1}$ given the state. Thus, the statement is true by Theorem 2.2. □

In the application of this thesis, the use case is that the estimate of the static state $\boldsymbol{a}$ is not updated online. Formally this means that the posterior distribution of static state $\boldsymbol{a}$ is approximated to remain unaffected by the received measurements, that is $p(a \mid x_k, y_{1:k-1}) \approx p(a)$. In this case, the estimation can be performed only for the dynamic components $\boldsymbol{x}_k$:

$$p(x_k \mid y_{1:k}) \propto \int p(y_k \mid x_k, a) \; p(a) \; p(x_k \mid y_{1:k-1}) \, \mathrm{d}a. \tag{3.7}$$

This will simplify the estimation notably especially if $\boldsymbol{a}$ has high dimensionality or the full conditional distribution of the static variables $p(a \mid x_k, y_{1:k-1})$ is not analytically tractable.

3.2 Kalman filter and its extensions

In case the measurement model function h_k is linear and independent from the static state components $\boldsymbol{a}$, that is, there exists a matrix H_k such that $h_k(x, a) = \mathrm{H}_k x$, the mean and covariance of the posterior state of the model (3.3) can be solved analytically. This celebrated solution was originally published by Rudolf Emil Kálmán in 1960 [17]. His algorithm, commonly called the Kalman filter (KF), is given in Algorithm 3.1. The first two rows inside the while loop form the prediction step of the filter, which is based on the result of Theorem 3.5.

Algorithm 3.1 Kalman filter

Let mean and covariance matrix of the initial prior be m_0 and P_0 respectively.

$m_0^+ := m_0$
$\mathrm{P}_0^+ := \mathrm{P}_0$
while There are measurements **do**
  $m_k^- := \Phi_k m_{k-1}^+$
  $\mathrm{P}_k^- := \Phi_k \mathrm{P}_{k-1}^+ {\Phi_k}^{\mathrm{T}} + \mathrm{Q}_k$
  $\mathrm{K}_k := \mathrm{P}_k^- {\mathrm{H}_k}^{\mathrm{T}} (\mathrm{H}_k \mathrm{P}_k^- {\mathrm{H}_k}^{\mathrm{T}} + \mathrm{R}_k)^{-1}$
  $m_k^+ := m_k^- + \mathrm{K}_k (y_k - \mathrm{H}_k m_k^-)$
  $\mathrm{P}_k^+ := (\mathrm{I} - \mathrm{K}_k \mathrm{H}_k) \mathrm{P}_k^-$
end while

The posterior is $p(x_k \mid y_{1:k}) = \mathrm{N}\left(x_k \mid m_k^+, \mathrm{P}_k^+\right)$.

The high usability of the KF is, among others, due to its certain favourable statistical properties which are not restricted in Gaussian error models and easiness of its implementation [5, Ch. 5]. However, it is applicable only for linear motion and measurement models, which is a significant limitation. Therefore, several approximative Kalman filter based methods have been developed for nonlinear problems. The simplest of them is the (first order) Extended Kalman filter (EKF). In the EKF, each distribution is approximated by normal distribution with the covariance matrix that is computed using the linearized model.

With highly nonlinear models, local linearization is not adequate approximation: for example multiple density peaks or asymmetries of peak shapes are omitted [4]. Nonlocal approximations can be made among others by approximating the mean and covariance of the distributions and assuming normality, such as in Unscented Kalman filter (UKF) and Cubature Kalman filter (CKF) [36], or by using the weighted sum of unimodal pdfs as the pdf, such as in Gaussian mixture filter (GMF) [3]. Local approximations can also be improved using higher order derivatives of the motion and measurement models (Second order extended Kalman filter, EKF2 [7, p. 384]. The problem with EKF2 is that computing Hessian matrices analytically tends to be challenging or completely intractable.

The idea of the Iterated extended Kalman filter (IEKF) is to improve linearization approximation by choosing the mode of the distribution as linearization point. In IEKF this is performed linearizing the model iteratively at the solution of the latest linearized model. Thus, the Gauss–Newton (GN) algorithm is performed at every time instant. IEKF is presented in Algorithm 3.2.

The IEKF can also be modified by replacing the GN with any other static positioning algorithm that takes prior mean and covariance matrix as input and returns posterior mean and covariance matrix. In this technique, the prior can be used for regularizing the convergence of the static method and for excluding impossible optima. However, since the likelihood is assumed independent of the prior information in the Bayes' rule, the prior should not influence the accomplished measurement

Algorithm 3.2 Iterated extended Kalman filter

$m_0^+ := m_0,\ \mathrm{P}_0^+ := \mathrm{P}_0,\ k := 1$
while There are measurements **do**
  $m_k^- := \Phi_k m_{k-1}^+$
  $\mathrm{P}_k^- := \Phi_k \mathrm{P}_{k-1}^+ {\Phi_k}^{\mathrm{T}} + \mathrm{Q}_k$
  Perform the damped GN algorithm (Algorithm 2.2) with prior mean m_k^- and covariance matrix P_k^-. Denote the output mean and covariance matrix with m_k^+ and P_k^+
  $k := k + 1$
end while
m_k^+ and P_k^+ are the mean and covariance matrix of the approximative posterior.

model approximations too dramatically. For instance, the estimation area of an estimation grid should not be confined by the prior.

Chapter 4

Positioning using RSS measurements

This section describes methods for location estimation using signal reception reports that a mobile user equipment (UE) gives based on signals received from wireless networks. The process includes two phases: learning phase and positioning phase.

In the learning phase, a pre-collected set of data is used for learning the characteristics of each base station's signal coverage. The learning measurements are in many contexts called fingerprints (FP). The database of FPs is maintained in an external server, so learning can be performed as an offline process, i.e. computations do not have to be done as real-time processing and their computational efficiency is not critical in terms of user experience.

The suggested learning method is a parametric estimation method, so the information contained in the data is summarized by a relatively small number of parameters whose values are estimated. The method has been designed so that the estimated parameter values can be updated recursively, which is useful, since new measurements will be assimilated continually. This is essential for the system's positioning performance, since the signal characteristics are not static properties; environmental effects that influence on signal propagation can vary dramatically over time, and the structure of the network may not be time invariant either.

In this thesis, positioning is an online procedure, so the tracking is made in a real-time manner as the user is using the terminal. There are potentially several ways to exploit signal reception reports for positioning, and the choice of positioning method obviously determines the form of the BS database. The mere list of the BSs whose signal is observed at a certain location is used e.g. in [21, 20]. Distances between BSs and UE can be estimated using Time of Arrival (TOA) information provided that the clocks of BSs and UE are synchronized or that their clock biases are known accurately. For using Time Difference of Arrival (TDOA) it is sufficient that only the BSs' clock biases are known. Angle of Arrival (AOA) indicates the direction of the UE with respect to the BS, and it is also usable for positioning if available. [45, Ch. 2.3], [37].

This thesis mainly considers the received signal strength (RSS) indicators that measure the intensity of the electromagnetic wave that the mobile terminal receives. Using the signal characteristics estimated in the learning phase, RSS measurement can be transformed into a distance measurement that has some statistical variance. The signal characteristics estimates are a result of statistical estimation procedure so they also have finite precisions, and one of the key topics of this thesis is the significance of this uncertainty on how realistic the statistical description of the system is.

4.1 Path loss model

A path loss (PL) model is a model for the propagation of (electromagnetic) signal in space. This section introduces a simple statistical PL model for an attenuating radio signal. A standard assumption is that the signal attenuates logarithmically in terms of the distance between the BS and the UE. This tradition was introduced by the famous Okumura–Hata models [27, 14].

A PL model typically contains several parameters that have to be determined before the positioning phase. One conventional way for tuning the parameters is to collect large amount of data by accomplishing extensive measurement campaigns. If the model contains enough BS-specific variables whose values are known for each BS, it may be possible to formulate a model whose parameter values can be tuned generically for all the BSs, at least with some restrictions. The Okumura–Hata models are an example of this methodology. Since the attenuation characteristics tend to change as a function of the distance, some authors tune several PL models for a BS and choose the correct one based on the RSS value [10].

The methodology adopted by this thesis is dynamic statistical estimation based on the Bayesian philosophy. Thus, the distribution of the model parameters is estimated statistically from the learning data, and it is updated if new learning data is added to the database. This kind of approach has been tested by estimating the parameters

generically such as in [33, 23], or by estimating the parameters separately for each BS such as in [25]. In this thesis, the parameters are estimated independently for each BS in the learning data. The locations of the BSs are not assumed known, but they are also estimated based on the learning data.

Path loss parameter and location estimation is described in this section only for a single BS, and this procedure is then applied to each BS. The model is very general and does not contain variables concerning e.g. frequency or antenna structure of the BS. The assumption that the parameters of different BSs are statistically independent may result in some information losses, but it will simplify the form of the BS database that is created and reduce the number of recorded statistics.

4.1.1 Derivation of the path loss model

The signal strength as a function of distance r from the BS is defined as

$$y(r) = 10 \log_{10} \left(P(r) \right),$$

where $P(r)$ stands for signal power (in milliwatts) measured at some reference surface. The received signal strength (RSS) is the signal strength measured by the receiver (RX)

$$y = 10 \log_{10} \left(P_{\mathrm{RX}} \right),$$

where $P_{\mathrm{RX}} = P(r_{\mathrm{RX}})$, where r_{RX} is the distance between BS and RX. Due to the logarithmic scale, the unit of RSS measurement is dBm (decibels referenced to milliwatt). In a line-of-sight case the intensity of a received radio signal follows the inverse-square law

$$P(r) \propto \frac{1}{r^2},$$

where r is the distance from the BS. Thus, the received signal power can be expressed as

$$P_{\mathrm{RX}} = \frac{r^2}{r_0^2} P(r_0),$$

where r_0 is a specified reference distance. The RSS measurement becomes

$$\begin{aligned} y &= 10 \log_{10} \left(\frac{r^2}{r_0^2} P(r_0) \right) \\ &= 10 \log_{10} \left(P(r_0) \right) - 20 \log_{10} \left(\frac{r}{r_0} \right) \\ &= y(r_0) - 20 \log_{10} \left(\frac{r}{r_0} \right). \end{aligned}$$

Conventionally, the reference distance is chosen to be 1 m and the signal strength at 1 m is called the apparent transmitter power and denoted with A.

The line-of-sight assumption means that there should be no obstacles between BS and RX. In practice, however, signal attenuation tends to be more rapid than this model predicts, and this effect depends on the propagation environment. Therefore, a BS-specific attenuation constant is introduced. The standard solution is to define the path loss exponent n, for which the relation

$$P(r) \propto \frac{1}{r^n}$$

holds. Because all disturbances of the free space assumption lead to larger n, the inequality $n \geq 2$ generally holds. This leads to the equation

$$y = A - 10n \log_{10}(r).$$

In each environment there tend to be attenuating factors that are very case-specific, direction dependent, time-variant or otherwise unreasonable to be modeled accurately. For instance, ground surface fluctuations, buildings, heavy traffic and weather conditions outdoors and walls, floors and furniture indoors could act as such factors. These factors together form the so called shadowing phenomenon (shadow fading, slow fading). Thus, the path loss model is

$$\boldsymbol{y} = A - 10n \log_{10}(r) + \boldsymbol{v}, \tag{4.1}$$

where $\boldsymbol{v}$ is the random variable that models the shadowing. It is usually assumed that $\boldsymbol{v}$ follows zero-mean normal distribution with environment-specific standard deviation σ [35, 38]. If the path loss parameters as well as the BS position $\boldsymbol{m}$ and the UE position $\boldsymbol{p}$ are unknown, they are modeled as random variables:

$$\boldsymbol{y} = \boldsymbol{A} - 10\boldsymbol{n} \log_{10} \|\boldsymbol{m} - \boldsymbol{p}\| + \boldsymbol{v}. \tag{4.2}$$

4.1.2 Parameter learning

In this section, a method for estimating the PL parameters of a BS is presented. Assume that there is a database of measurements each containing the true position, the list of heard BS-ID's and the observed RSS levels. This database has been collected beforehand, and it is assumed that the position fixes are accurate compared with the accuracy of RSS-based positioning, for example GNSS measurements outdoors.

Let us now consider the observations of one certain BS in the database. Each measurement with index $i \in \{1, \ldots, N_\ell\}$ contains (accurate) position p_i and measured RSS level y_i. Let us denote the BS position with m. Using the measurement model

(4.1) and a Gaussian prior for all the parameters, the structure of the model is similar to that of (2.4) so that

$$\begin{aligned} \hat{x} &= \begin{bmatrix} A^- & n^- & m^{-\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \\ \hat{\Sigma}_{\boldsymbol{x}} &= \operatorname{blkdiag}\left(\sigma_{\boldsymbol{A}}^{-2}, \sigma_{\boldsymbol{n}}^{-2}, \Sigma_{\boldsymbol{m}}^{-}\right) \\ h(A,n,m) &= \begin{bmatrix} h_1(A,n,m) \\ \vdots \\ h_{N_\ell}(A,n,m) \end{bmatrix} \\ \mathrm{R} &= \sigma^2 \cdot \mathrm{I}_{N_\ell \times N_\ell}, \end{aligned}$$

where the measurement model functions are

$$h_i(A,n,m) = A - 10n\log_{10}(\|p_i - m\|).$$

The measurement model function is continuously differentiable everywhere outside the set of the measurement points, and the partial derivatives of the measurement function are derived in Section 4.2.3. The Jacobian matrix of the measurement function h is

$$\mathrm{H} = \begin{bmatrix} 1 & -10\log_{10}(\|m - p_1\|) & -\frac{10}{\ln(10)} n \frac{(m-p_1)^{\mathrm{T}}}{\|m-p_1\|^2} \\ \vdots & \vdots & \vdots \\ 1 & -10\log_{10}(\|m - p_{N_\ell}\|) & -\frac{10}{\ln(10)} n \frac{(m-p_{N_\ell})^{\mathrm{T}}}{\|m-p_{N_\ell}\|^2} \end{bmatrix}. \tag{4.3}$$

Thus, the inference can be made using the GN algorithm.

All the quantities are given Gaussian priors to improve convergence properties of the Gauss–Newton algorithm and to ensure the convergence to a physically sensible set of values. The priors are to be almost uninformative, i.e. Gaussian distributions with so large variance that the influence on the optimum is negligible. A suitable initial value for the BS position is the position of the strongest observed measurement. Initial values for A and n can be chosen more arbitrarily from the valid ranges, since the distribution is can be supposed to be unimodal enough, if the number of data points is large.

The algorithm also returns an approximation for the covariance matrix of each quantity. Consequently, it is potentially possible to distinguish between trustworthy and untrustworthy path loss models. Note that to simplify the analysis it is assumed in this thesis that the correlations between the PL parameters and the BS coordinates are so small that they can be neglected.

Due to Theorem 2.2 the Bayesian algorithms are usually recursive: the obtained posterior can be used as a prior as new measurements are received.

In an ideal case, the BS position is unambiguously indicated by the strongest RSS values. If there are several peaks in a practical situation, the GN algorithm may not handle them satisfactorily. In these cases, the BS positions could be estimated in grid and the rest of the parameters could be estimated analytically for each grid point. However, the empirical tests showed that the influence of these effects on positioning performance is usually negligible and the GN algorithm is reliable enough.

In Figures 4.1 and 4.2 there are examples of real-data-based RSS powermaps that are computed by interpolating between measurement points. Coverage areas are the 68% probability ellipses of the normal distribution fitted in the data [21, 20]. Figure 4.1 presents the power map of a BS of cellular network, and in Figure 4.2 there is the powermap of a wireless local area network (WLAN) access point (AP). Note that in the WLAN power map the influence of the building's floor plan is clearly visible, and the AP is not truly omnidirectional.

4.2 Static positioning

The main purpose of this section is to present implementable algorithms for estimating the mean and covariance matrix of the user position's posterior distribution. The detailed algorithm listings are presented in Appendix A.

The RSS model might result in multiple almost equally weighted maximum likelihood solutions for the position. In the Bayesian methodology, a prior distribution, which may also contain positioning information, is potentially able to make difference between the likelihood peaks. Moreover, the prior may improve the convergence properties of iterative algorithms. For mathematical convenience, a Gaussian prior distribution for the user's position $p(p) = \mathrm{N}\left(p \mid \hat{p}, \hat{\Sigma}_{\boldsymbol{p}}\right)$ is assumed.

The prior can possibly reflect location information from other sources, or in case of time-series filtering, the filter prediction functions as the prior mean $\hat{p}$ and covariance $\hat{\Sigma}_{\boldsymbol{p}}$. If no prior information exists, e.g. the arithmetic mean of BS positions can be used as the prior mean. In this case the covariance matrix should be chosen to be large so that the influence of the prior on the posterior is negligible.

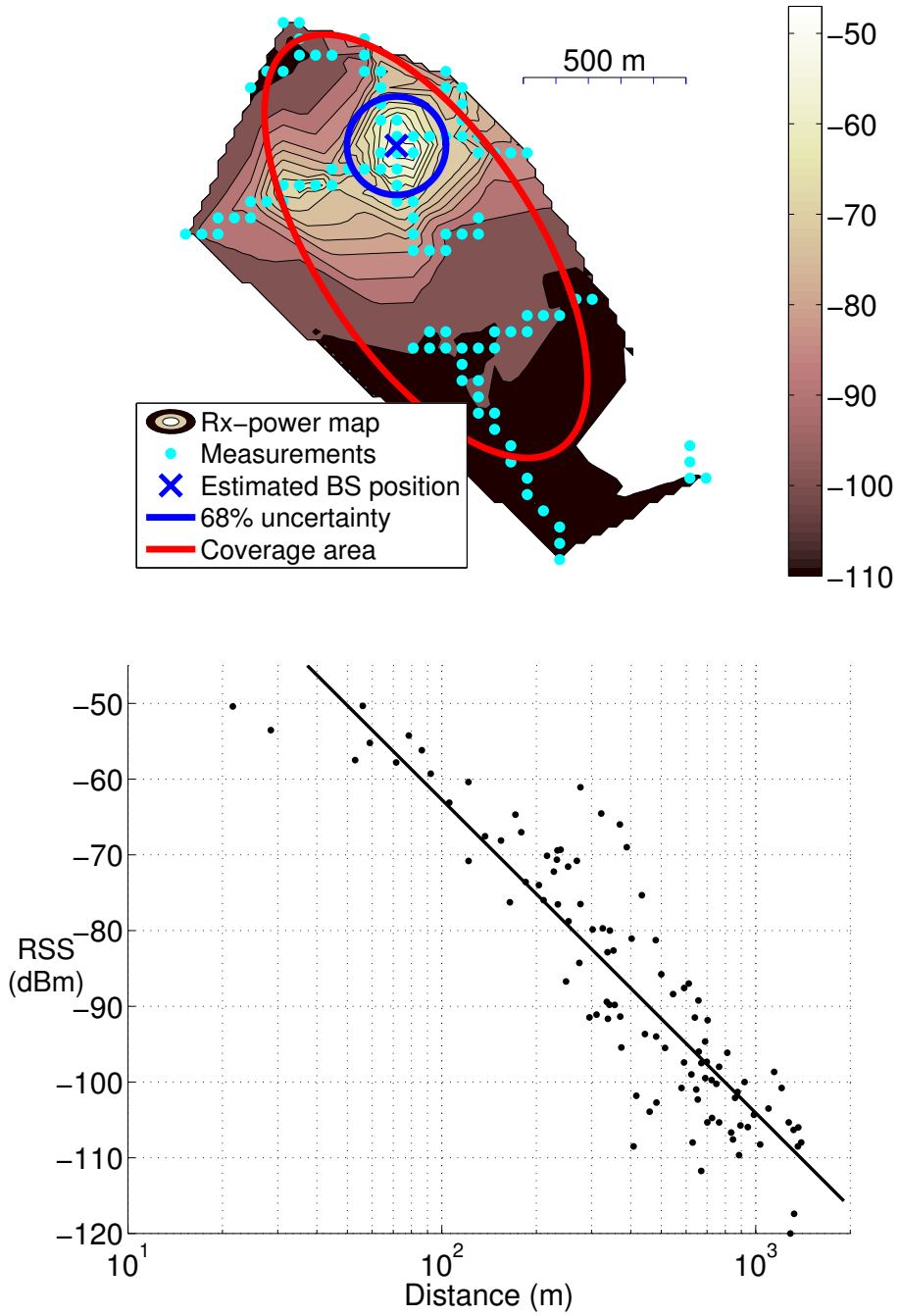

Figure 4.1: *Interpolated power map of a BS of a cellular network with the estimates for BS position and coverage area and the path loss curve of the same BS. The path loss model is linear for A and n, so finding them for a given BS position is the problem of fitting a line to the semilogarithmic distance–RSS plot.*

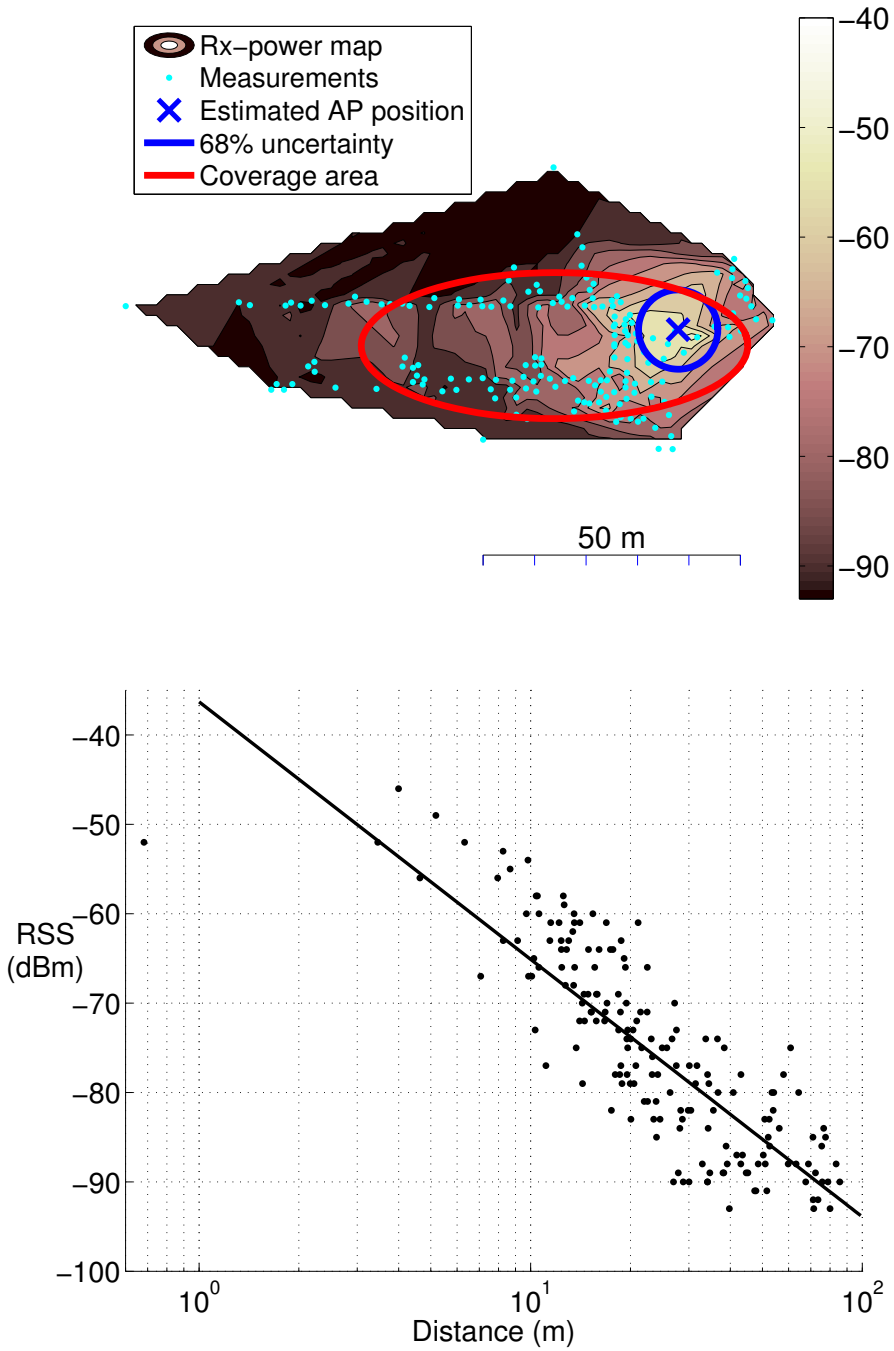

Figure 4.2: *Interpolated power map of a WLAN AP with the estimates for AP position and coverage area and the path loss curve of the same AP.*

4.2.1 Grid algorithm for RSS measurements

Let p be user position, A apparent TX power, n path loss exponent and m base station position. By the given independence assumptions, the full posterior distribution can be expressed as

$$\begin{aligned} p(p, A, n, m \mid y) \propto & \mathrm{N}\left(y \mid A - 10n\log_{10}(\|m-p\|)), \sigma^2\right) \\ & \cdot \mathrm{N}\left(p \mid \hat{p}, \hat{\Sigma}_{\boldsymbol{p}}\right) \cdot \mathrm{N}\left(\begin{bmatrix} A \\ n \end{bmatrix} \mid \begin{bmatrix} \hat{A} \\ \hat{n} \end{bmatrix}, \hat{\Sigma}_{\boldsymbol{A},\boldsymbol{n}}\right) \\ & \cdot \mathrm{N}\left(m \mid \hat{m}, \hat{\Sigma}_{\boldsymbol{m}}\right). \end{aligned} \tag{4.4}$$

The posterior distribution for position p is obtained by marginalising out the nuisance parameters:

$$\begin{aligned} p(p \mid y_{1:N_{\boldsymbol{y}}}) \propto & \mathrm{N}\left(p \mid \hat{p}, \hat{\Sigma}_{\boldsymbol{p}}\right) \cdot \prod_{i=1}^{N_{\boldsymbol{y}}} \iiint \mathrm{N}\left(y_i \mid A_i - 10n_i \log_{10} \|m_i - p\|, \sigma^2\right) \\ & \cdot \mathrm{N}\left(\begin{bmatrix} A_i \\ n_i \end{bmatrix} \mid \begin{bmatrix} \hat{A}_i \\ \hat{n}_i \end{bmatrix}, \hat{\Sigma}_{\boldsymbol{A}_i,\boldsymbol{n}_i}\right) \cdot \mathrm{N}\left(m_i \mid \hat{m}_i, \hat{\Sigma}_{\boldsymbol{m}_i}\right) \mathrm{d}A_i \ \mathrm{d}n_i \ \mathrm{d}m_i. \end{aligned}$$

This can be computed by sampling A_i, n_i and m_i, $i = 1, \ldots, N_{\boldsymbol{y}}$ from their prior distribution and setting

$$p(p \mid y_{1:N_{\boldsymbol{y}}}) \propto\!\!\!\propto \mathrm{N}\left(p \mid \hat{p}, \hat{\Sigma}_{\boldsymbol{p}}\right) \prod_{i=1}^{N_{\boldsymbol{y}}} \frac{1}{N} \sum_{j=1}^{N} \mathrm{N}\left(y_i \mid A_{i,(j)} - 10n_{i,(j)} \log_{10} \left\|m_{i,(j)} - p\right\|, \sigma^2\right),$$

where N is the sample size parameter and where i is the index of the measurement and j the index of the sample. Symbol $\propto\!\!\!\propto$ denotes approximate proportionality.

For simplicity, this thesis uses rectangular grid with constant step sizes. The determination of the grid coverage is loosely based on the position prior distribution. However, one should be careful with this technique, since the likelihood should not depend on the prior.

The most crucial implementation issues are the Monte Carlo sample size parameter N as well as grid size and density. Note that combining the likelihoods of different BSs is done in logarithmic space to avoid numerical underflows. The grid method is described in detail in Algorithm A.1.

4.2.2 Metropolis–Hastings sampler for RSS measurements

To reduce computational complexity and Monte Carlo variance, PL parameters A and n are marginalized out analytically. This can be done using Lemma B.3 as follows

$$\begin{aligned}
&p(p, m_{1:N_y} \mid y_{1:N_y}) \\
&\propto \iint p(y_{1:N_y} \mid p, A_{1:N_y}, n_{1:N_y}, m_{1:N_y})\; p(A_{1:N_y}, n_{1:N_y})\, \mathrm{d}A_{1:N_y}\, \mathrm{d}n_{1:N_y} \cdot p(p) \cdot p(m_{1:N_y}) \\
&= p(p) \cdot \prod_{i=1}^{N_y} p(m_i) \cdot \iint \mathrm{N}\left(y_i \mid \mathrm{B}_i \begin{bmatrix} A_i \\ n_i \end{bmatrix}, \sigma^2\right) \cdot \mathrm{N}\left(\begin{bmatrix} A_i \\ n_i \end{bmatrix} \mid \begin{bmatrix} \hat{A}_i \\ \hat{n}_i \end{bmatrix}, \hat{\Sigma}_{\boldsymbol{A}_i, \boldsymbol{n}_i}\right) \mathrm{d}A_i\, \mathrm{d}n_i \\
&= p(p) \cdot \prod_{i=1}^{N_y} p(m_i) \cdot \mathrm{N}\left(y_i \mid \mathrm{B}_i \begin{bmatrix} \hat{A}_i \\ \hat{n}_i \end{bmatrix}, \mathrm{B}_i \hat{\Sigma}_{\boldsymbol{A}_i, \boldsymbol{n}_i} \mathrm{B}_i{}^{\mathrm{T}} + \sigma^2\right),
\end{aligned} \tag{4.5}$$

where

$$\mathrm{B}_i = \begin{bmatrix} 1 & -10 \log_{10} \|m_i - p\| \end{bmatrix}$$

The MH algorithm for RSS measurements is presented in Algorithm A.2, in which the sample indices are typed in parentheses for clarity.

In the implementation phase, great care must be taken when setting the proposal distributions to make the algorithm converge in a computationally feasible number of iterations. To avoid numerical underflows, the numerator and denominator of the MH ratio should be computed in logarithmic scale. For convenience, the proposal distributions are chosen to be multivariate normal with the latest accepted value as the mean and positive definite matrices tuned from the prior covariances of $\boldsymbol{p}$ and $\boldsymbol{m}_i$ as covariance matrices. The covariance matrices of the proposal distributions are denoted with $\mathrm{P}_{\boldsymbol{p}}$ and $\mathrm{P}_{\boldsymbol{m}_i}$.

4.2.3 Gauss–Newton algorithm for RSS measurements

An example of iterative state estimation method is the Gauss–Newton algorithm (GN) that is described is Section 2.3. In this section the GN method is formulated for the position estimation problem with the presented path loss model.

For formulating the Regularised Gauss–Newton algorithm, the partial derivatives of the measurement function h are calculated as follows:

$$\begin{aligned}
\frac{\partial h_i}{\partial p} &= -\frac{10n_i}{\ln(10)} \frac{1}{\|m_i - p\|} \cdot \frac{\partial \|m_i - p\|}{\partial p} \\
&= -\frac{10n_i}{\ln(10)} \frac{1}{\|m_i - p\|} \cdot \frac{(m_i - p)^{\mathrm{T}}}{\|m_i - p\|} \cdot \frac{\partial (m_i - p)}{\partial p} = \frac{10n_i}{\ln(10)} \frac{(m_i - p)^{\mathrm{T}}}{\|m_i - p\|^2}, \\
\frac{\partial h_i}{\partial A_i} &= 1, \\
\frac{\partial h_i}{\partial n_i} &= -10 \log_{10}(\|m_i - p\|), \\
\frac{\partial h_i}{\partial m_i} &= -\frac{10n_i}{\ln(10)} \frac{1}{\|m_i - p\|} \cdot \frac{(m_i - p)^{\mathrm{T}}}{\|m_i - p\|} \cdot \frac{\partial (m_i - p)}{\partial m_i} = -\frac{10n_i}{\ln(10)} \frac{(m_i - p)^{\mathrm{T}}}{\|m_i - p\|^2}.
\end{aligned}$$

The remaining partial derivatives are zeros. Since all the partial derivatives are continuous, the measurement model function is continuously differentiable whenever $m_i \neq p$ holds for each i. In case this condition does not hold at some stage of the GN algorithm, a small perturbation can be added to the position estimate.

The prior covariance matrix is always full-rank, so the singular values of the Jacobian are bounded away from zero, and provided that the perturbation steps are done, the singular values of the Jacobian are bounded above. Thus, the model fulfills the assumptions of Theorem 2.6 provided that the simplified line search algorithm is efficient enough. Theorem 2.6 implies that the algorithm converges locally to a stationary point of the posterior. Note that the goodness of the prior distribution and the initial point of the iteration are crucial in forcing the algorithm to converge to the true MAP (*maximum a posteriori*, posterior mode).

The GN algorithm for positioning with RSS measurements is presented in Algorithm A.3. The complete state containing both user position and all the PL parameters is denoted with $\boldsymbol{\xi}$. The output of the algorithm contains estimates for the MAP and the covariance matrix of the posterior of the linearized model.

4.3 Time series positioning

In this thesis the algorithm used for time-series estimation is an approximative version of the Bayesian filter. Each of the presented positioning algorithms return the mean and covariance matrix of the posterior density. The posterior of the position is then approximated be a normal distribution, so by Theorem 3.5 the prediction step can be done using the prediction step of the Kalman filter (KF). The predicted estimate is again a Gaussian, so it can be used as a prior for the presented positioning algorithms. This algorithm is thus the modified-IEKF procedure that is described in Section 3.2.

The fact that the shadowing terms of successive RSS measurements are correlated is omitted in this thesis [13],[19, Ch. 4]. The measurement interval is several seconds, which diminishes the correlations to some extent compared to cases with higher measurement frequencies.

Within the considered application, the model parameter distributions are not expected to be updated online. Thus, the system parameter distributions are not included in the filter state, so the update step is based on Equation (3.7).

Another filtering alternative would be to use the static estimate as a direct position measurement in the Kalman filter. In this method, the prior should not be used in the static estimation except maybe for determining the initial state. This approach is commonly known as the Positioning Kalman filter (PKF) [4]. In this thesis, the algorithm described in the previous paragraphs is implemented, since an accurate prior may help avoid divergence of the static positioning methods and accelerate their convergence. On the other hand, if the motion model was chosen to contain derivatives of position, such as velocity or acceleration, the PKF would be the most flexible scheme.

The main reason for including a Monte Carlo positioning method into this thesis is the class of particle filter (PF) algorithms [24]. Particle filters are Monte Carlo algorithms that can, at least in principle, handle practically any motion and measurement model. In positioning, measurement models that are typically handled by particle filters are for example map constraints and sensor measurements of different kind. Especially in the field of indoor positioning, this approach has been a subject of increasing research interest [12, 43, 44].

There also exist algorithms for time-series estimation with static motion models, that is, motion models with zero variance. In the case of KF extensions, the model parameters could be included in the filter state with static motion model, since the positive-definiteness of the measurement error covariance matrix ensures existence of the matrix inverse that appears in the KF formulas. This kind of static motion model may, however, result in problems with numerical stability, since the variances decrease at every measurement time step. PF algorithms can be used with static motion models only within special conditions [42]. Otherwise, heuristic procedures such as adding small artificial noise must be adopted to prevent degeneration of the sample set to only one or few different values.

Chapter 5

Testing and comparison

5.1 RSS likelihood

Figure 5.1 illustrates the influence of the uncertain parameters on the likelihood of the UE location. In the example case the parameter statistics are $\begin{bmatrix} \hat{A} \\ \hat{n} \end{bmatrix} = \begin{bmatrix} 2 \\ 3.2 \end{bmatrix}$, $\hat{\Sigma}_{\boldsymbol{A},\boldsymbol{n}} = \begin{bmatrix} 100 & 3.5 \\ 3.5 & 0.15 \end{bmatrix}$, $\hat{m} = 0$ and $\hat{\Sigma}_{\boldsymbol{m}} = 2 \cdot 10^4 \cdot \mathrm{I}$. These values are based on our experimental knowledge of the Finnish cellular network. The likelihood plots presented in the figures are computed using the grid algorithm with very sense grid. The left column illustrates the likelihoods of the model that takes the parameter uncertainties into account, and the right column shows the likelihoods assumed that the path loss parameter values are accurate. The RSS values corresponding to the likelihoods are -50 (on the left), -75 and $-90\,\mathrm{dBm}$. The shadowing standard deviation is $\sigma = 6\,\mathrm{dBm}$. It can be seen that with strong signals the RSS likelihood is unimodal or almost unimodal when the parameter uncertainties are taken into account.

In the case of radially symmetric BS position distribution, the posterior density depends only on the distance from the mean of the BS position estimate. Figure 5.2 illustrates the likelihoods of the UE position as a function of this distance. They

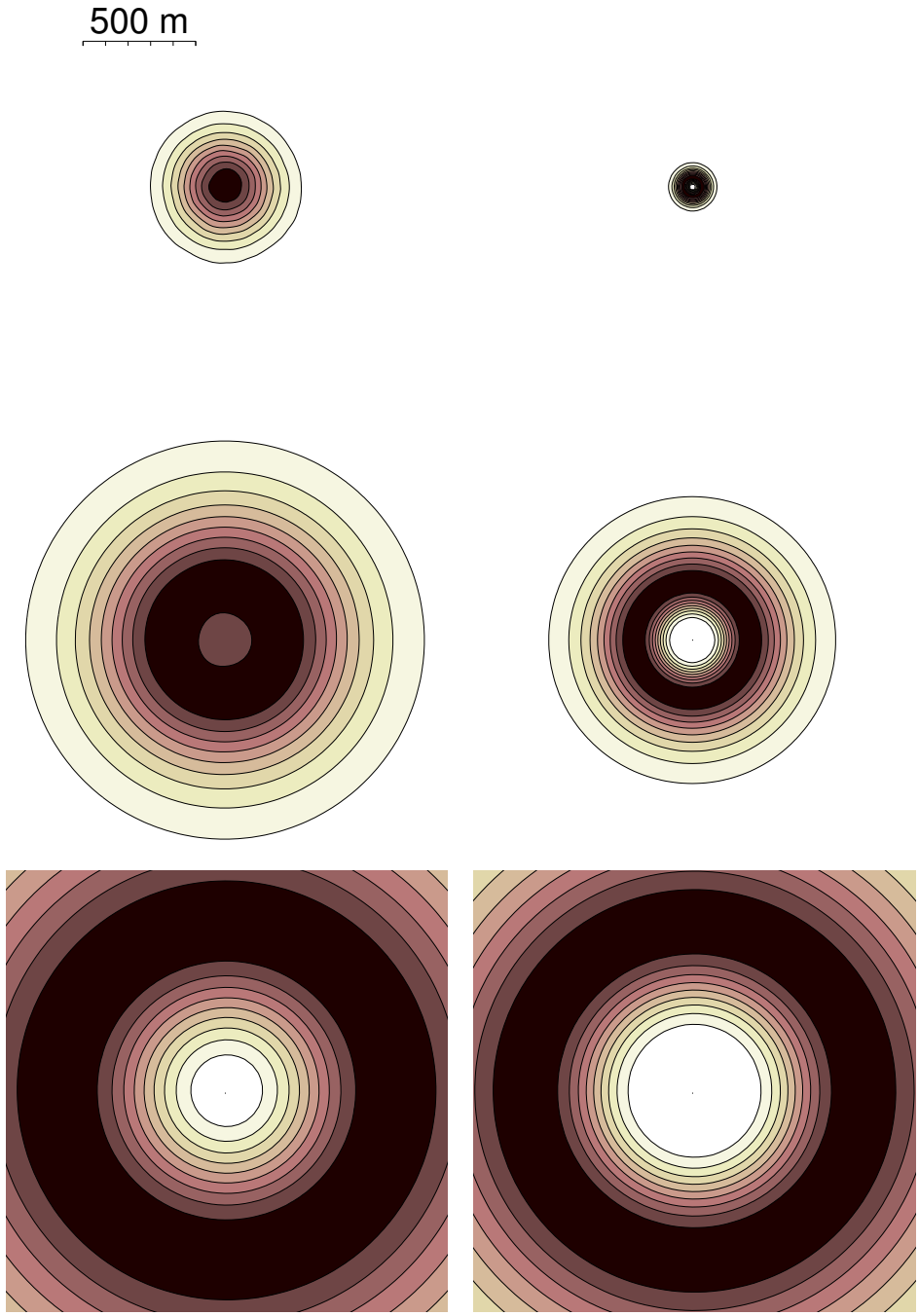

Figure 5.1: *The likelihoods of measurements* -50, -75 *and* -90 *dBm (from the top row downwards). In the left column, parameter uncertainties have been taken into account. The likelihoods have been normalized so that the maximal value of each likelihood is one, and each plot uses the same colour map.*

have been computed using standard Monte Carlo integration and normalized so that the maximal likelihood value is one. Curve "N" represents algorithms that assume PL parameters to be normally distributed *a priori*, and curve "acc" the algorithms that assume that the parameters are known accurately. Fig. 5.2 shows that the tail of the "N" curve are considerably heavier.

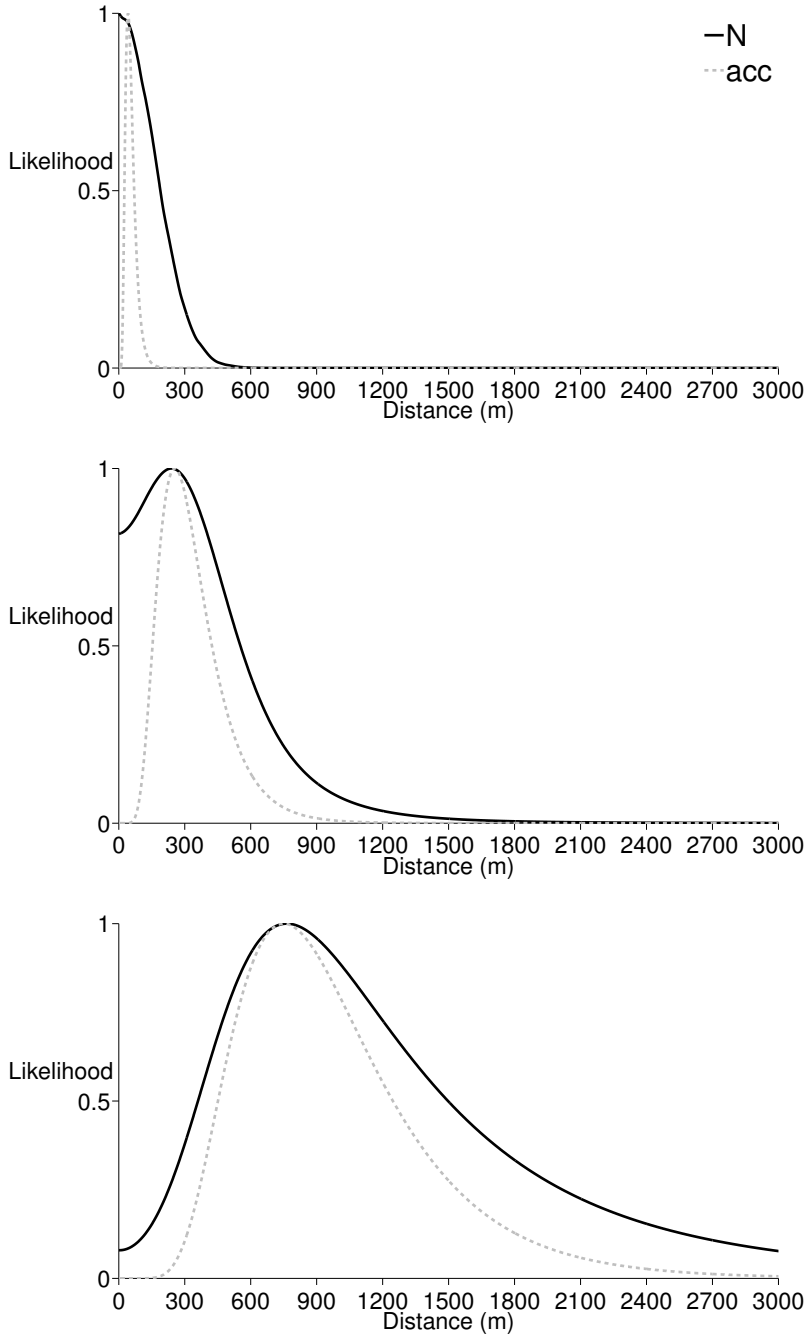

Figure 5.2: *Likelihoods of measurements* -50, -75 *and* -90 *dBm as a function of the distance from the mean of the BS position estimate. Curve "N" represents algorithms that assume PL parameters to be normally distributed a priori, and curve "acc" the algorithm that assumes that the parameters are known accurately.*

In Figure 5.3 the likelihood of two RSS measurements of −80 dBm is presented. The PL parameters are similar to the ones in Figure 5.1, and the distance between BS's is 400 meters. If the parameter uncertainties are not taken into account (on the right), the support of the likelihood consists of two separate parts, whereas in the left figure there is significant amount of likelihood mass also in the BS positions' surroundings.

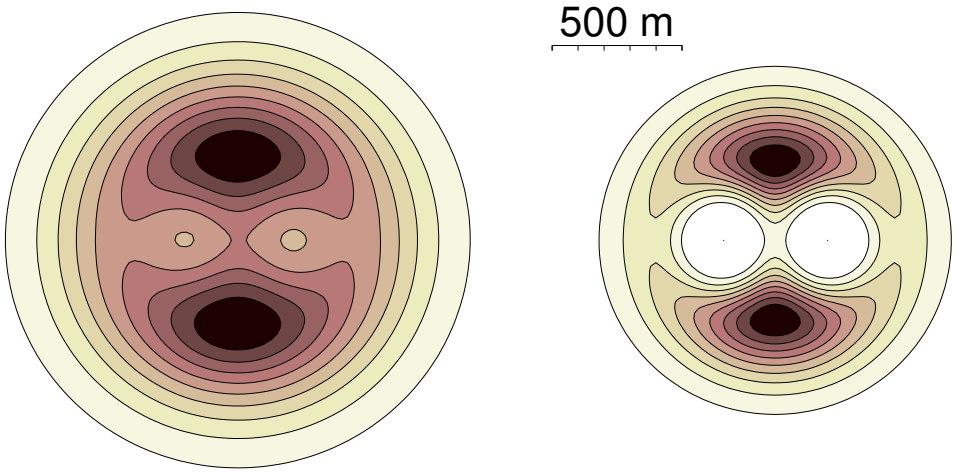

Figure 5.3: *The combined likelihood of two BS's with signal strengths* −80 *dBm. On the left, parameter uncertainties have been taken into account. The likelihoods have been normalized so that the maximal value of each likelihood is one, and both the plots use the same colour map.*

5.2 Outdoor tests with cellular data

5.2.1 Experiment setup

A measurement campaign was accomplished to evaluate the performance of different algorithms in a real use case. First, a large set of fingerprints was collected from a WCDMA cellular network in Tampere urban area, Finland for learning the radiomap. The measured RSS values are based on the measured Received Signal Code Power (RSCP) indicator reported by the user equipment. The coverage area of each BS was estimated by fitting a normal distribution to the data [21, 20]. Furthermore, path loss model parameters were estimated using the method that was presented in Section 4.1.2 of this thesis.

The measurement data of cellular networks tends to be spatially correlated [13]. In order to reduce the effect of correlations, the measurements are mapped to a grid of pre-specified points before the estimation process. In this setup, the RSS value of a

grid point is set to be the mean of the RSSs observed in the proximity of the grid point.

2D-projection effects in the proximity of the BS antenna are taken into account by increasing the covariance matrix of the BS position artificially with a diagonal constant matrix. This reflects also errors that stem from GPS errors in the learning data and measurement error correlations due to environmental effects.

There are three separate outdoor test tracks. The first track (Hervanta) was collected by a pedestrian in a densely populated urban/suburban area. In the second track (Lukonmäki) the measurer rode a bicycle with a low velocity in a suburban area. The third track (Linnainmaa) is a higher velocity suburban bicycle case. In all the cases the true user positions were tracked down using conventional GPS positioning. By plotting the GPS solutions on the map, it was confirmed that the GPS error on the area is small compared to the cellular positioning accuracy.

In all the test tracks, the prior distribution is computed in each estimation point using the estimated coverage areas of the BS's that are observed at the point concerned. The prior is the product distribution of Gaussian coverage areas, so it is also a Gaussian distribution.

5.2.2 Results and discussion

The results of the real-data outdoor tests are in Table 5.1. Abbreviation "N" stands for the algorithms that assume the path loss parameters to be normally distributed *a priori* whereas "acc" indicates that the parameter values are assumed accurate. "CA" refers to the method in which only the product of coverage areas is used. For the Lukonmäki case, both the GPS solution and the estimated track are plotted on the map in Figure 5.4.

The positioning error at one time step is the Euclidean distance of the position estimate and the corresponding reference location. Columns "Mean", "Med" and "95% err." are mean error, median error and empirical 95% percentile of errors in meters. "Time" is the the average running time of our MATLAB implementation in seconds. Note that the codes are not highly optimized so the running time values have to be considered only roughly indicative. The times are also highly dependent on the chosen configuration parameters. Column "Cons." displays the 95% consistency that was determined using Gaussian consistency test [7, p. 235] with risk level 5%. The solver is deemed to be consistent at a certain time step, if the true position is within the 95%-ellipse of the posterior distribution, assuming normality of the posterior. The closer "Cons." is to 95%, the more realistic the covariance matrix estimation is.

Table 5.1: *Results of static cellular positioning with real data.*

Solver	Mean (m)	Med (m)	95% err. (m)	Cons. (%)	Time (s)
Hervanta					
grid, N	251	203	433	95	90.3
grid, acc	255	222	446	82	36.5
MH, N	257	211	482	82	42.9
MH, acc	267	228	477	75	14.1
GN, N	255	208	450	88	0.4
GN, acc	259	224	461	67	0.4
CA	260	213	458	99	0.3
Lukonmäki					
grid, N	223	159	449	96	100.0
grid, acc	224	159	522	81	53.1
MH, N	225	153	590	87	66.9
MH, acc	234	156	574	74	21.7
GN, N	226	158	536	90	0.6
GN, acc	237	173	548	64	0.5
CA	258	200	575	96	0.3
Linnainmaa					
grid, N	195	157	418	96	129.4
grid, acc	195	160	402	90	49.8
MH, N	209	170	500	91	45.2
MH, acc	204	167	476	86	13.9
GN, N	185	155	413	94	0.4
GN, acc	188	158	409	83	0.3
CA	268	207	741	99	0.2

From the figures of Table 5.1 it can be seen that taking the parameter uncertainties into account improves the consistency remarkably for all the estimation methods. Accurate covariance matrix estimation is crucial especially when location information from other sources is combined with RSS measurements or when positioning is done with Bayesian time-series filters [4].

Presented "N" algorithms seem to outperform "acc" algorithms slightly in the positioning accuracy. In the Hervanta case it is questionable whether the RSS measurements should be used at all, if the uncertainties are not in the model. Some filtering results are also provided in Table 5.2. The used filtering algorithm is described in Section 4.3. In the Lukonmäki case, this highly simplified filtering method seems to reduce mean errors by at least 10%, and performance differences between "N" and "acc" algorithms are somewhat clearer than in the static results in Table 5.1. The filtering algorithm's tendency to make the appearance of the estimated track smoother and less jumpy is illustrated by Figure 5.4.

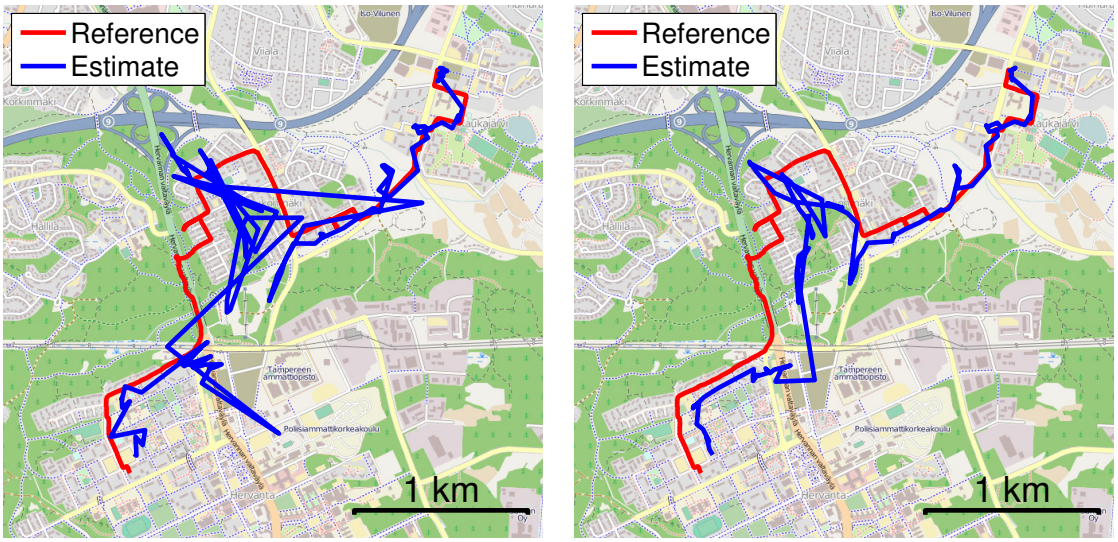

Figure 5.4: *The GPS track (Reference) and the track estimated by the Gauss–Newton algorithm (Estimate) for the Lukonmäki test track. The static solution is on the left, and the filtering solution on the right.*

Among the three estimation methods, grid and MH sampler approach the exact Bayesian posterior distribution. The grid gives the precise posterior values in the grid points assuming that the Monte Carlo integration's accuracy is adequate. The MH sampler converges theoretically to the true posterior as the sample size parameter N approaches infinity. In practice, the rate of convergence in MH algorithms is highly dependent on the form and parameters of the proposal distributions. With the chosen configuration the method usually fails to compete with the grid especially in consistency. However, as discussed in Section 4.3, the Monte Carlo framework is a flexible and efficient tool especially in time-series analysis of highly nonlinear or non-Gaussian measurements.

Table 5.2: *Cellular positioning results using time-series filtering. Both coverage areas and path loss models are used in the estimation. The track is the same as "Lukonmäki" in Table 5.1.*

Solver	Mean (m)	Med (m)	95% err. (m)	Cons. (%)
Lukonmäki				
grid, N	193	137	493	86
grid, acc	198	151	503	74
MH, N	198	148	490	81
MH, acc	202	155	493	67
GN, N	194	138	501	85
GN, acc	203	161	510	64
CA	218	205	469	90

The Gauss–Newton method lacks global convergence properties and the covariance matrix estimate is based on linearized model and has thus a less clear Bayesian interpretation. Indeed, the real data tests show that the algorithm's convergence is more sensitive to the quality of the prior distribution. However, the presented results are comparable with those of the other methods, and the GN is clearly the computationally lightest one of the presented PL algorithms, and the most applicable one in many real-time solutions.

5.3 Indoor tests with WLAN data

5.3.1 Experiment setup

A large set of WLAN fingerprints was collected in public indoor spaces in the city of Tampere, Finland for learning the radiomap. The test cases presented in this thesis are located in a building at Tampere University of Technology campus area. The test track and most of the learning data have been collected indoors. The test track consists of several parts measured at different floors of the same building. The measurement device is a tablet computer, and the reference locations were set manually on the floor plan figure. Each floor has a separate radiomap, and the correct floor is assumed known in both learning and positioning phases; that is, no floor detection is implemented.

Similarly to the outdoor case, the effects due to measurement error correlations are taken into account heuristically by increasing the covariance matrix of the AP position artificially with a small diagonal constant matrix. The matrix has been configured observing the accuracy and consistency of the algorithm in various tracks.

In the indoor cases, the presented RSS-based methods are compared with both coverage area (CA) method and the weighted k-nearest neighbour (WKNN) method, which can be regarded as the state-of-the-art solution to the RSS-positioning problem [15]. In the the WKNN method, the measurements are not compressed into parametric form, i.e. no statistical assumptions are made of the measurement model. Instead, the whole measurement database is stored in the memory, and the positioning measurement is compared with every learning measurement. In this thesis, the used method is the weighted 3-neighbour method, in which the difference of the measurement to each database point is computed using the Euclidean norm of RSS differences, and the location estimate is set to the mean value of three closest database points. The WKNN estimates are not filtered in this paper.

5.3.2 Results and discussion

In outdoor cases, the coverage areas were used as a prior to regulate the convergence of the RSS-based algorithms. In indoor spaces, however, the performance of the coverage area algorithm may not be adequate for this purpose. Since, in addition to that, indoor positioning data tend to be relatively noisy, Bayesian time-series filters are tested. They provide a regularizing prior and hinder too large jumps of the position estimate. Table 5.3 presents the positioning results of each positioning algorithm with different priors. "CA" means that the coverage area solution is used as a prior for the RSS solution. Columns "Mean", "Med" and "95% err." are mean error, median error and empirical 95% percentile of errors in meters. Column "Cons." is the 95% consistency that is defined in Section 5.2.2.

Table 5.3: *Comparison of different prior construction methods in indoor positioning. In "CA" algorithms, the coverage area estimate is included in the prior for RSS methods. In "filter" methods, a Bayesian time-series filter is used.*

Solver	Mean (m)	Med (m)	95% err (m)	Cons (%)
CA & grid, static	7.3	7.0	15.0	16
CA & grid, filter	7.3	6.5	14.7	13
grid, filter	7.0	5.4	18.6	87
CA & MH, static	7.7	7.0	15.4	7
CA & MH, filter	7.4	6.7	14.7	13
MH, filter	7.5	6.1	20.2	84
CA & GN, static	7.4	7.1	15.4	16
CA & GN, filter	7.3	6.6	14.6	13
GN, filter	7.6	6.1	20.9	84

Based on Table 5.3, the algorithms with CA prior have serious consistency problems, which is probably due to dependency of adjacent coverage areas. Thus, the method that uses the filter prediction as the prior and no CA information is adopted for the comparison purposes. Note that in the used filter, the PL parameter estimates are not updated online, but the same prior for the parameters is used for all time instants. The used filtering algorithm is explained in Section 4.3.

Figure 5.5 shows position solutions for a part of the test track given by both versions of the GN algorithm. The results of the real-data indoor tests are in Table 5.4 and Table 5.5. In Table 5.5 only 20% of the fingerprints in the learning data have been used for 75% of the APs. Abbreviation "N" stands for the algorithms that assume the PL parameters to be normally distributed *a priori* whereas "acc" indicates that the parameter values are assumed known. "CA" refers to the product of coverage areas. "WKNN" is the (weighted) 3-nearest neighbour method with the Euclidean distance. "Time" is the the average running time of our MATLAB implementation in seconds.

Table 5.4: *Filtering results for the real data tests for the indoor case. The complete learning data set used.*

Solver	Mean (m)	Med (m)	95% err (m)	Cons (%)	Time (s)
grid, N	7.0	5.4	18.6	87	112.7
grid, acc	7.2	5.4	19.8	80	43.0
MH, N	7.5	6.1	20.2	84	68.2
MH, acc	7.4	6.0	20.1	79	15.5
GN, N	7.6	6.1	20.9	84	0.4
GN, acc	7.7	6.3	23.5	53	0.2
CA	8.7	7.6	16.3	13	0.2
WKNN	5.6	5.2	12.9		30.6

Table 5.5: *Results for the real data tests for the indoor case. For 75% of the APs nine out of ten location reports have been removed artificially. These APs and the left-out points were chosen randomly.*

Solver	Mean (m)	Med (m)	95% err (m)	Cons (%)
grid, N	7.0	5.8	15.6	93
grid, acc	7.1	6.1	17.7	76
MH, N	7.6	6.8	17.0	84
MH, acc	7.3	6.1	18.2	71
GN, N	7.0	6.0	15.8	93
GN, acc	7.6	6.2	20.0	54
CA	9.9	9.1	20.6	10
WKNN	10.4	7.9	26.5	

In terms of the error statistics presented in Table 5.4, the proposed RSS methods seem to perform better than the coverage area solution but slightly worse than the WKNN solution in positioning accuracy. Note, however, that both Gauss–Newton solutions are computationally much more efficient and the requirements for the database are much lower for the parametric algorithms, since only the PL parameter estimates and their variances have to be stored for each AP instead of all the measurement points. Moreover, pruning the database influences the fingerprint solution much more than the parametric methods.

The parameter uncertainties seem to be essential from the consistency's viewpoint also in the indoor case. In practice WLAN positioning in indoor spaces is complemented by additional, more refined sources such as map information, inertial navigation systems or Bluetooth. When several types of measurements are combined, it is crucial to be knowledgeable of the accuracy of each measurement, and so the improvement in consistency is a good reason for taking the parameter uncertainties into account in indoor positioning.

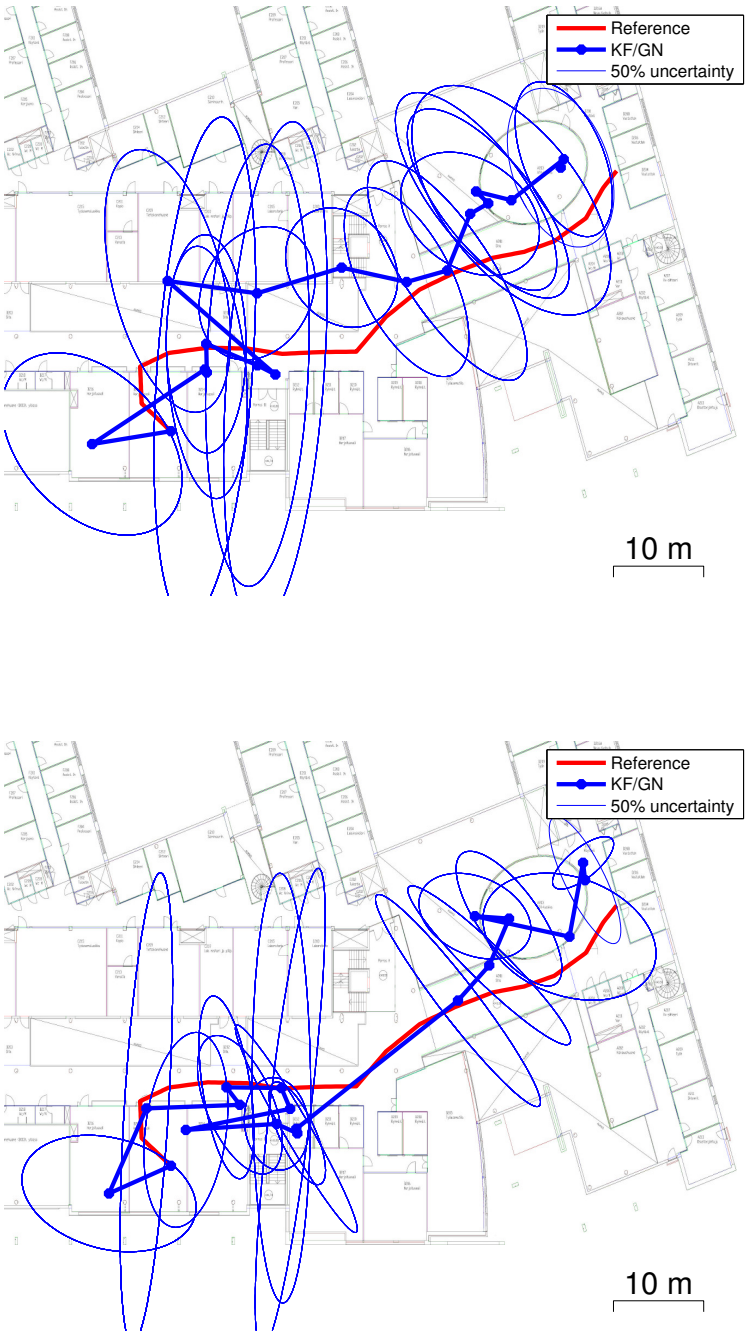

Figure 5.5: *Part of the indoor test track with the GN method. In the upper figure parameter uncertainties have been taken into account.*

Similarly to the outdoor case, the presented "N" algorithms seem to outperform "acc" algorithms slightly in the positioning accuracy too. However, the performance differences are somewhat clearer in Table 5.5, where the learning data sets of some the APs have been pruned. Thus, it seems that PL parameter uncertainties should

be taken into account especially if some of the APs are likely to be badly mapped. This might be the case e.g. if there are newly added APs or if the area as a whole is inadequately covered by the database.

Chapter 6

Conclusions

This thesis presented three statistical positioning methods that use measurement model with unknown system constants. The performance of the methods was evaluated collecting several test sets of real data from wireless networks in outdoor (WCDMA cellular network) and indoor (WLAN) environments. The used measurements were received signal strength (RSS) measurements, and the unknown system constants were base station -specific path loss model parameters, whose prior distributions were estimated from learning data using the Gauss–Newton algorithm.

The real-data tests showed that RSS-based path loss methods outperform the coverage area method that does not use RSS measurements but is only based on the list of observed BSs. This holds for both cellular–outdoor and WLAN–indoor cases. In the indoor scenario, the path loss methods were also shown to be comparable in accuracy with the nonparametric k-nearest neighbour method, which compares the measurement with the complete fingerprint database. Furthermore, the performance of the k-nearest neighbour method is more sensitive to inadequately mapped APs.

It was also shown that taking the parameter uncertainties into account in the positioning phase improves consistency of error estimates compared to the methods where the path loss parameters are assumed to be known accurately. This means that the positioning error estimates are more realistic when the uncertainties are

taken into account. Consistency improvement was clear in both outdoor and indoor scenarios. In positioning accuracy, the versions that take the parameter uncertainties into account seemed to perform equally or mostly slightly better than the ones that assume known parameters.

The performance differences were emphasized favourably to the methods that take the parameter uncertainties into account if the database coverage is inadequate. This was an expectable relation, since the parameter uncertainties measure the goodness of the PL parameter fit and adequacy of the learning data; if the fit is less reliable for one BS than for the others, the larger uncertainties automatically give less weight to the observations from this BS.

In the tests presented in this thesis, taking the parameter uncertainties into account appeared computationally relatively demanding. However, if needed, the cost could be cut down without deteriorating the performance too much by using suitable approximations and simplifications, such as [26, 32]. The crucial idea in this thesis is that the uncertainties of unknown parameters affect the estimation even though we are not interested in the actual parameter values.

The consistency improvement provided by parameter uncertainty modeling is especially significant, when different measurements are combined, since the information contents of two noisy measurements cannot be merged unless there is reliable knowledge of the accuracy of both measurements. In practice, wireless network -based positioning is typically complemented by additional, more refined techniques such as map information or inertial navigation systems. This holds especially in indoor environments where line-of-sight conditions are challenging and accuracy requirements high.

Future topics include studying more refined filtering techniques for the RSS measurements with unknown measurement model parameters, such as particle filtering. Filtering methods for models with correlated measurement noise could also be studied, since the shadowing term of RSS measurements is known to be spatially correlated.

In Bayesian philosophy, it is straightforward to combine other measurements with the presented positioning algorithms, at least in principle. Adding maps and inertia-based information into outdoor and indoor positioning and showing the influence of the parameter uncertainties in a hybrid positioning system is a topic for future research. Additional future topic is expanding the presented methods into 3D position space especially in indoor spaces, in which floor detection is an interesting and essential part of navigation.

Bibliography

[1] Simo Ali-Löytty. Kalmanin suodatin ja sen laajennukset paikannuksessa. M.Sc. thesis, Tampere University of Technology, 2004. URL http://URN.fi/URN:NBN:fi:tty-201012141388.

[2] Simo Ali-Löytty. On the Convergence of the Gaussian Mixture Filter. Department of Mathematics. Research report 89, Tampere University of Technology, 2008. URL http://URN.fi/URN:NBN:fi:tty-2011041510704.

[3] Simo Ali-Löytty. *Gaussian Mixture Filters in Hybrid Positioning.* PhD thesis, Tampere University of Technology, August 2009. URL http://URN.fi/URN:NBN:fi:tty-200905191055.

[4] Simo Ali-Löytty, Niilo Sirola, and Robert Piché. Consistency of three Kalman filter extensions in hybrid navigation. In *Proceedings of The European Navigation Conference GNSS 2005*, Munich, Germany, July 2005.

[5] Brian D. O. Anderson and John B. Moore. *Optimal Filtering.* Electrical Engineering Series. Prentice Hall, 1979.

[6] Paramvir Bahl and Venkata N. Padmanabhan. RADAR: An In-Building RF-based User Location and Tracking System. *Proceedings of the 19th International conference on Computer Communications (InfoCom). IEEE*, 2:775–784, March 2000.

[7] Yaakov Bar-Shalom, Rong X. Li, and Thiagalingam Kirubarajan. *Estimation with Applications to Tracking and Navigation, Theory Algorithms and Software.* John Wiley & Sons, 2001.

[8] Dimitri P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 1995.

[9] Åke Björck. *Numerical methods for least squares problems.* Society for Industrial and Applied Mathematics, 1996.

[10] Atreyi Bose and Chuan Heng Foh. A practical path loss model for indoor WiFi positioning enhancement. In *Information, Communications Signal Processing, 2007 6th International Conference on*, pages 1–5, December 2007.

[11] Ekahau. [online; referred 29.10.2012]. URL http://www.ekahau.com.

[12] Frédéric Evennou, Marx Francois, and Emil Novakov. Map-aided indoor mobile positioning system using particle filter. *Wireless Communications and Networking Conference, IEEE*, 4:2490 – 2494, 2005.

[13] Mikael Gudmundson. Correlation model for shadow fading in mobile radio systems. *Electronics Letters*, 27(23):2145–2146, November 1991.

[14] Masaharu Hata. Empirical formula for propagation loss in land mobile radio services. *IEEE Transactions on Vehicular Technology*, VT-29(3):317–325, 8 1980.

[15] Ville Honkavirta, Tommi Perälä, Simo Ali-Löytty, and Robert Piché. A Comparative Survey of WLAN Location Fingerprinting Methods. In *Proceedings of the 6th Workshop on Positioning, Navigation and Communication 2009 (WPNC'09)*, pages 243–251, 2009.

[16] Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory*, volume 64 of *Mathematics in Science and Engineering*. Academic Press, 1970.

[17] Rudolph Emil Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[18] Petri Koistinen. Monte Carlo methods with an emphasis on Bayesian computation. Lecture notes, 2010.

[19] Laura Koski. Positioning with Bayesian coverage area estimates and location fingerprints. Master's thesis, University of Tampere, March 2010. URL `http://tutkielmat.uta.fi/pdf/gradu04187.pdf`.

[20] Laura Koski, Tommi Perälä, and Robert Piché. Indoor positioning using WLAN coverage area estimates. In *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, September 2010.

[21] Laura Koski, Robert Piché, Ville Kaseva, Simo Ali-Löytty, and Marko Hännikäinen. Positioning with coverage area estimates generated from location fingerprints. In *Proceedings of the 7th Workshop on Positioning, Navigation and Communication 2010 (WPNC'10)*, pages 99–106, March 2010.

[22] Heikki Laitinen, Jaakko Lähteenmäki, and Tero Nordström. Database Correlation Method for GSM Location. In *Proceedings of the IEEE 53rd Vehicular Technology Conference*, volume 4, pages 2504–2508, May 2001.

[23] Xinrong Li. RSS-Based Location Estimation with Unknown Pathloss Model. *IEEE Transactions on Wireless Communications*, 5(12):3626–3633, December 2006. ISSN 1536-1276. doi: 10.1109/TWC.2006.256985.

[24] Jun S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, 2001.

[25] Santiago Mazuelas, Francisco A. Lago, David González, Alfonso Bahillo, Juan Blas, Patricia Fernandez, Ruben M. Lorenzo, and Evaristo J. Abril. Dynamic estimation of optimum path loss model in a RSS positioning. In *Position, Location and Navigation Symposium, 2008 IEEE/ION*, pages 679–684, May 2008.

[26] Philipp Müller, Simo Ali-Löytty, Marzieh Dashti, Henri Nurminen, and Robert Piché. Gaussian Mixture Filter Allowing Negative Weights and its Application to Positioning Using Signal Strength Measurements. In *Proceedings of the 9th Workshop on Positioning, Navigation and Communication 2012 (WPNC'12)*, March 2012.

[27] Yoshihisa Okumura, Eiji Ohmori, Tomihiko Kawano, and Kaneharu Fukuda. Field strength and its variability in uhf and vhf land-mobile radio service. *Review of the Electrical Communication Laboratory*, 16:825–873, September-October 1968.

[28] Henri Pesonen. Numerical integration in Bayesian positioning. M.Sc. thesis, Tampere University of Technology, June 2006.

[29] Robert Piché. Stochastic processes. Lecture notes, August 2012. URL `http://URN.fi/URN:NBN:fi:tty-201012021377`.

[30] Robert Piché, Simo Särkkä, and Jouni Hartikainen. Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate student-t distribution. In *2012 IEEE International Workshop on Machine Learning for Signal Processing*, September 2012.

[31] Polaris Wireless. [online; referred 29.10.2012]. URL `http://www.polariswireless.com`.

[32] Matti Raitoharju, Simo Ali-Löytty, Robert Piché, and Marzieh Dashti. Positioning with Multilevel Coverage Area Models. In *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, November 2012.

[33] Javier Rodas and Carlos J. Escudero. Dynamic path-loss estimation using a particle filter. *IJCSI International Journal of Computer Science Issues*, 7(4), May 2010.

[34] Teemu Roos, Petri Myllymäki, Henry Tirri, Pauli Misikangas, and Juha Sievänen. A Probabilistic Approach to WLAN User Location Estimation. *International Journal of Wireless Information Networks*, 9(3):155–164, July 2002.

[35] Jari Salo, Lasse Vuokko, Hassan M. El-Sallabi, and Pertti Vainikainen. An additive model as a physical basis for shadow fading. *IEEE Transactions on Vehicular Technology*, 56(1):13–26, January 2007. ISSN 0018-9545. doi: 10.1109/TVT.2006.883797.

[36] Simo Särkkä. Bayesian Estimation of Time-varying Systems: Discrete-Time Systems. Lecture notes, Aalto University, 2012.

[37] Ali H. Sayed, Alireza Tarighat, and Nima Khajehnouri. Network-based wireless location: challenges faced in developing techniques for accurate wireless location information. *IEEE Signal Processing Magazine*, 22(4):24–40, July 2005.

[38] Shweta Shrestha, Elina Laitinen, Jukka Talvitie, and Elena Simona Lohan. RSSI channel effects in cellular and WLAN positioning. In *Proceedings of the 9th Workshop on Positioning, Navigation and Communication 2012 (WPNC'12)*, March 2012.

[39] Niilo Sirola. Closed-form algorithms in mobile positioning: Myths and misconceptions. In *Proceedings of the 7th Workshop on Positioning, Navigation and Communication 2010 (WPNC'10)*, pages 38–44, Dresden Germany, March 2010.

[40] Niilo Sirola and Simo Ali-Löytty. Local positioning with parallelepiped moving grid. In *Proceedings of 3rd Workshop on Positioning, Navigation and Communication 2006 (WPNC'06)*, pages 179–188, Hannover, March 16th 2006.

[41] Skyhook. [online; referred 29.10.2012]. URL `http://www.skyhookwireless.com`.

[42] Geir Storvik. Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, 50(2):281 –289, February 2002.

[43] Hui Wang, Henning Lenz, Andrei Szabo, Joachim Bamberger, and Uwe D. Hanebeck. WLAN-Based Pedestrian Tracking Using Particle Filters and Low-Cost MEMS Sensors. In *Proceedings of the 4th Workshop on Positioning, Navigation and Communication (WPNC'07)*, 2007.

[44] Widyawan, Martin Klepal, and Stéphane Beauregard. A novel backtracking particle filter for pattern matching indoor localization. In *MELT '08: Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments*, pages 79–84, New York, NY, USA, 2008. ACM.

[45] Lauri Wirola. *Studies on location technology standards evolution in wireless networks*. PhD thesis, Tampere University of Technology, 2010. URL `http://URN.fi/URN:NBN:fi:tty-201002121065`.

Appendix A

Algorithms for RSS positioning

Algorithm A.1 Grid for positioning with RSS measurements

1. Set a grid $\{p_m \in \mathbb{R}^2 \mid m \in \{1, \ldots, N_m\}\}$ that is assumed to cover almost all posterior probability mass.

2. For each detected base station $i = 1, \ldots, N_y$, draw

$$\begin{bmatrix} A_{i,(k)} \\ n_{i,(k)} \end{bmatrix} \leftarrow \mathrm{N}\left(\begin{bmatrix} \hat{A}_i \\ \hat{n}_i \end{bmatrix}, \hat{\Sigma}_{\boldsymbol{A}_i, \boldsymbol{n}_i} \right)$$
$$m_{i,(k)} \leftarrow \mathrm{N}\left(\hat{m}_i, \hat{\Sigma}_{\boldsymbol{m}_i} \right)$$

   for $k = 1, \ldots, N$.

3. At each grid point p_m compute for each base station $i = 1, \ldots, N_y$ and for each sample $k = 1, \ldots, N$

$$I_{i,m,(k)} := \mathrm{N}\left(y_i \mid A_{i,(k)} - 10 n_{i,(k)} \log_{10}\left(\left\| m_{i,(k)} - p_m \right\| \right), \sigma^2 \right),$$

   and $I_{i,m} := \frac{1}{N} \sum_{k=1}^{N} I_{i,m,(k)}$. Then set

$$\ell_m := \ln\left(\mathrm{N}\left(p_m \mid \hat{p}, \hat{\Sigma}_{\boldsymbol{p}} \right) \right) + \sum_{i=1}^{N_y} \ln(I_{i,m}), \quad L_m := \exp(\ell_m).$$

4. Normalize the grid to get a set of weights $w_m = \frac{L_m}{\sum_{m=1}^{N_m} L_m}$ and compute mean and covariance matrix estimates

$$\hat{p}^+ = \sum_{m=1}^{N_m} w_m p_m \ , \quad \hat{\Sigma}_{\boldsymbol{p}}^+ = \sum_{m=1}^{N_m} w_m (p_m - \hat{p}^+)(p_m - \hat{p}^+)^{\mathrm{T}}.$$

Algorithm A.2 Metropolis–Hastings algorithm for RSS measurements

1. Set $p_{(0)} := \hat{p}$, $A_{i,(0)} := \hat{A}_i$, $n_{i,(0)} := \hat{n}_i$ and $m_{i,(0)} := \hat{m}_i$ for $i = 1, \ldots, N_y$. Set $\rho_{(0)}$ using the formula in step 3. Set $k := 1$.

2. Generate $p'_{(k)} \leftarrow \mathrm{N}\left(p_{(k-1)}, \mathrm{P}_{\boldsymbol{p}}\right)$, and for each BS $i = 1, \ldots, N_y$, generate $m'_{i,(k)} \leftarrow \mathrm{N}\left(m_{i,(k)}, \mathrm{P}_{\boldsymbol{m}_i}\right)$.

3. For each $i = 1, \ldots, N_y$, compute $\mathrm{B}_{i,(k)} = \begin{bmatrix} 1 & -10 \log_{10}\left(\left\| m'_{i,(k)} - p'_{(k)} \right\|\right) \end{bmatrix}$ and

$$\rho'_{(k)} := -\frac{1}{2}(p'_{(k)} - \hat{p})^{\mathrm{T}} \hat{\Sigma}_{\boldsymbol{p}}^{-1} (p'_{(k)} - \hat{p}) + \sum_{i=1}^{N_y} \Big[-\frac{1}{2}(m'_{i,(k)} - \hat{m}_i)^{\mathrm{T}} \hat{\Sigma}_{\boldsymbol{m}_i}^{-1} (m'_{i,(k)} - \hat{m}_i) + \ln\left(\mathrm{N}\left(y_i \mid \hat{A}_i - 10\hat{n}_i \log_{10}\left\| m'_{i,(k)} - p'_{(k)} \right\|, \mathrm{B}_{i,(k)} \hat{\Sigma}_{\boldsymbol{A}_i, \boldsymbol{n}_i} {\mathrm{B}_{i,(k)}}^{\mathrm{T}} + \sigma^2\right)\right)\Big]$$

4. Set $r := \exp(\rho'_{(k)} - \rho_{(k-1)})$. Generate $u \leftarrow \mathrm{Unif}\,(0, 1)$. Compute
   if $r > u$ **then**
      for $i = 1 : N_y$ **do**
         $m_{i,(k)} := m'_{i,(k)}$
      end for
      $p_{(k)} := p'_{(k)}$, $\rho_{(k)} := \rho'_{(k)}$
   else
      for $i = 1 : N_y$ **do**
         $m_{i,(k)} := m_{i,(k-1)}$
      end for
      $p_{(k)} := p_{(k-1)}$, $\rho_{(k)} := \rho_{(k-1)}$
   end if

5. Set $k := k + 1$. If $k < N$, go to step 2. Otherwise, set

$$\hat{p}^{+} := \frac{1}{N - N_{\mathrm{b}} + 1} \sum_{k=N_{\mathrm{b}}}^{N} p_{(k)} \ , \quad \hat{\Sigma}_{\boldsymbol{p}}^{+} := \frac{1}{N - N_{\mathrm{b}} + 1} \sum_{k=N_{\mathrm{b}}}^{N} (p_{(k)} - \hat{p}^{+})(p_{(k)} - \hat{p}^{+})^{\mathrm{T}},$$

   where N_{b} is the length of the burn-in period.

Algorithm A.3 Gauss Newton algorithm for positioning with RSS measurements

1. Choose the stopping tolerance δ and maximal iteration number $k_{\max}$. Let

$$\hat{\Sigma}_{\boldsymbol{\xi}} = \text{blkdiag}\left(\hat{\Sigma}_{\boldsymbol{p}}, \hat{\Sigma}_{\boldsymbol{A}_1,\boldsymbol{n}_1}, \hat{\Sigma}_{\boldsymbol{m}_1}, \ldots, \hat{\Sigma}_{\boldsymbol{A}_{N_{\boldsymbol{y}}},\boldsymbol{n}_{N_{\boldsymbol{y}}}}, \hat{\Sigma}_{\boldsymbol{m}_{N_{\boldsymbol{y}}}}\right)$$

and

$$\hat{\xi} = \begin{bmatrix} \hat{p}^{\mathrm{T}} & \hat{A}_1 & \hat{n}_1 & \hat{m}_1{}^{\mathrm{T}} & \ldots & \hat{A}_{N_{\boldsymbol{y}}} & \hat{n}_{N_{\boldsymbol{y}}} & \hat{m}_{N_{\boldsymbol{y}}}{}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$$

be the prior covariance and mean. Let the initial state be $\xi_0 := \hat{\xi}$. Set $k := 0$. The objective function is denoted with

$$\theta(\xi) := (\xi - \hat{\xi})^{\mathrm{T}} \hat{\Sigma}_{\boldsymbol{\xi}}{}^{-1} (\xi - \hat{\xi}) + \frac{1}{\sigma^2} \sum_{i=1}^{N_{\boldsymbol{y}}} (h_i(\xi) - y_i)^2.$$

2. Compute the Jacobian matrix

$$\mathrm{J}_k = \begin{bmatrix} \frac{\partial h_1}{\partial p} & \frac{\partial h_1}{\partial A_1} & \frac{\partial h_1}{\partial n_1} & \frac{\partial h_1}{\partial m_1} & 0_{4N_{\boldsymbol{y}}-4}{}^{\mathrm{T}} \\ \vdots & & \ddots & & \vdots \\ \frac{\partial h_{N_{\boldsymbol{y}}}}{\partial p} & 0_{4N_{\boldsymbol{y}}-4}{}^{\mathrm{T}} & \frac{\partial h_{N_{\boldsymbol{y}}}}{\partial A_{N_{\boldsymbol{y}}}} & \frac{\partial h_{N_{\boldsymbol{y}}}}{\partial n_{N_{\boldsymbol{y}}}} & \frac{\partial h_{N_{\boldsymbol{y}}}}{\partial m_{N_{\boldsymbol{y}}}} \end{bmatrix}$$

3. Set

$$\Delta\xi_k := -\left(\hat{\Sigma}_{\boldsymbol{\xi}}^{-1} + \frac{1}{\sigma^2} \mathrm{J}_k{}^{\mathrm{T}} \mathrm{J}_k\right)^{-1} \cdot \left(\hat{\Sigma}_{\boldsymbol{\xi}}^{-1} (\xi_k - \hat{\xi}) + \frac{1}{\sigma^2} \mathrm{J}_k{}^{\mathrm{T}} (h(\xi_k) - y)\right).$$

4. Perform the damping of step length:

   $\alpha := 1$
   while $\|\theta(\xi_k + \alpha\Delta\xi_k)\| \geq \|\theta(\xi_k)\|$ and $\alpha > \alpha_0$ **do**
   $\alpha := \frac{\alpha}{2}$
   end while

   where $\alpha_0 \ll 2$ is a configuration parameter. Set $\xi_{k+1} := \xi_k + \alpha\Delta\xi_k$.

5. If stopping condition $\|\Delta\xi_k\| < \delta$ is not satisfied and $k \leq k_{\max}$, set $k := k + 1$ and go to Step 2. Otherwise compute $\mathrm{P} := \left(\hat{\Sigma}_{\boldsymbol{\xi}}^{-1} + \frac{1}{\sigma^2} \mathrm{J}_k{}^{\mathrm{T}} \mathrm{J}_k\right)^{-1}$ and set the state estimate

$$\hat{p}^+ := \xi_{k+1,1:2}, \quad \hat{\Sigma}_{\boldsymbol{p}}^+ := \mathrm{P}_{1:2,1:2}$$

Appendix B

Properties of Gaussian distribution

Definition B.1. (Non-degenerate multivariate normal distribution) Let $\mu \in \mathbb{R}^d$ be a vector and $\Sigma \in \mathbb{R}^{d\times d}$ be a symmetric positive definite matrix. A random variable $\boldsymbol{x} \in \mathbb{R}^d$ follows non-degenerate multivariate normal distribution with mean μ and covariance matrix Σ if its probability density function (pdf) is

$$p(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right). \tag{B.1}$$

Then the pdf of the distribution is denoted with $p(x) = \mathrm{N}\left(x \mid \mu, \Sigma\right)$. ■

Lemma B.2. *Assume that*

$$p(x) = \mathrm{N}\left(x \mid m^-, \mathrm{P}^-\right)$$
$$p(y \mid x) = \mathrm{N}\left(y \mid \mathrm{H}x, \mathrm{R}\right).$$

Then

$$p(x,y) = p(x)\cdot p(y \mid x) = \mathrm{N}\left(\begin{bmatrix} x \\ y \end{bmatrix} \mid \begin{bmatrix} m^- \\ \mathrm{H}m^- \end{bmatrix}, \begin{bmatrix} \mathrm{P}^- & \mathrm{P}^-\mathrm{H}^{\mathrm{T}} \\ \mathrm{HP}^- & \mathrm{HP}^-\mathrm{H}^{\mathrm{T}} + \mathrm{R} \end{bmatrix}\right).$$

Proof. Omitted. See [2]. □

Lemma B.3. *Assume that*

$$\begin{aligned} p(x) &= \mathrm{N}\left(x \mid m^-, \mathrm{P}^-\right) \\ p(y \mid x) &= \mathrm{N}\left(y \mid \mathrm{H}x, \mathrm{R}\right). \end{aligned}$$

Then

$$p(x \mid y) \propto p(x) \cdot p(y \mid x) = c \cdot \mathrm{N}\left(x \mid m^+, \mathrm{P}^+\right),$$

where

$$\begin{aligned} \mathrm{P}^+ &= \left(\mathrm{P}^{-{}^{-1}} + \mathrm{H}^{\mathrm{T}}\mathrm{R}^{-1}\mathrm{H}\right)^{-1} = \left(\mathrm{I} - \mathrm{P}^-\mathrm{H}^{\mathrm{T}}\mathrm{S}^{-1}\mathrm{H}\right)\mathrm{P}^- \\ m^+ &= \mathrm{P}^+\left(\mathrm{P}^{-{}^{-1}}m^- + \mathrm{H}^{\mathrm{T}}\mathrm{R}^{-1}y\right) = m^- + \mathrm{P}^-\mathrm{H}^{\mathrm{T}}\mathrm{S}^{-1}(y - \mathrm{H}m^-) \\ c &= \mathrm{N}\left(y \mid \mathrm{H}m^-, \mathrm{S}\right), \end{aligned}$$

where $\mathrm{S} = \mathrm{H}\mathrm{P}^-\mathrm{H}^{\mathrm{T}} + \mathrm{R}$.

Proof. Omitted. See [2]. □

Appendix C

Convergence results for gradient methods

Gradient methods are optimization methods that are based on moving towards the optimum point stepwise and the step direction determined by the gradient (derivative) of the objective function. This section proves the key convergence results for the general form of gradient method. Further discussion of the topic is provided e.g. in the book [8]. The proofs of this section are mainly based on ideas presented in [8, Ch. 1.2].

Definition C.1. (Descent direction) Let $\Phi : \mathbb{R}^{N_x} \to \mathbb{R}_0^+$ be a nonnegative function and $x \in \mathbb{R}^{N_x}$ a vector. Vector $d \in \mathbb{R}^{N_x}$ is a descent direction, if

$$\Phi'(x)d < 0.$$

■

Lemma C.2. *Let $c \in (0, 1)$ be a constant and $\Phi : \mathbb{R}^{N_x} \to \mathbb{R}_0^+$ a continuously differentiable function. For any vector $x \in \mathbb{R}^{N_x}$ and descent direction $d_x \in \mathbb{R}^{N_x}$ there exists $s \in \mathbb{R}^+$ so that*

$$\forall \alpha \in (0, s] : \ \Phi(x) - \Phi(x + \alpha d_x) \geq -c\alpha\Phi'(x)d_x.$$

Proof. By the definition of multivariate derivative, the following relation holds for any $x \in \mathbb{R}^{N_x}$:

$$\Phi(x + \alpha d_x) - \Phi(x) = \alpha \Phi'(x) d_x + \alpha \epsilon_x(\alpha),$$

where $\epsilon_x(\alpha) \to 0$ as $\alpha \to 0$. Since $\epsilon_x(\alpha) \to 0$ as $\alpha \to 0$ and $\Phi'(x)d_x < 0$, there exists $s \in \mathbb{R}^+$ so that $|\epsilon_x(\alpha)| \leq -(1-c)\Phi'(x)d_x$ for each $\alpha \in (0, s]$. Thus, the inequality

$$\begin{aligned} \Phi(x + \alpha d_x) - \Phi(x) &= \alpha \Phi'(x) d_x + \alpha \epsilon_x(\alpha) \\ &\leq \alpha \Phi'(x) d_x + \alpha \left| \epsilon_x(\alpha) \right| \\ &\leq \alpha (\Phi'(x) d_x - (1-c)\Phi'(x) d_x) \\ &= c\alpha \Phi'(x) d_x \end{aligned}$$

holds for each $\alpha \in (0, s]$. This proves the statement. □

Lemma C.3. *Let Φ be a continuously differentiable function and $x \in \mathbb{R}^{N_x}$ a point such that $\Phi'(x) \neq 0$. Then, a vector of form $d = -\mathrm{D}\Phi'(x)^{\mathrm{T}}$, where D is a positive definite matrix, is a descent direction for function Φ and point x.*

Proof.

$$\Phi'(x) d = -\Phi'(x) \mathrm{D} \Phi'(x) < 0,$$

since matrix D is positive definite. □

Proposition C.4. *Let $\Phi : \mathbb{R}^{N_x} \to \mathbb{R}_0^+$ be a continuously differentiable function. Let $\{x_k\}$ be a sequence generated by the recursion*

$$x_{k+1} = x_k + \alpha_k d_k.$$

Directions $\{d_k\}$ are defined by $d_k = -\mathrm{D}_k \Phi'(x_k)$ with symmetric positive definite matrix D_k such that

$$\{\sigma \in \mathbb{R}_0^+ \mid \sigma \text{ is an eigenvalue of } \mathrm{D}_k \text{ for some } k \in \mathbb{N}_0\}$$

is bounded above and bounded away from zero, and

$$\alpha_k = \operatorname*{argmin}_{\alpha \in (0,s]} \Phi(x_k + \alpha d_k).$$

for some $s \in \mathbb{R}^+$. Then if a subsequence of $\{x_k\}$ converges, it converges to a stationary point of Φ.

Proof. Let us prove that the negation of the statement implies contradiction. Assume that $\bar{x}$ is a limit point of $\{x_k\}$ such that $\Phi'(\bar{x}) \neq 0^{\mathrm{T}}$.

Since the eigenvalues of the matrix D_k are positive and bounded away from zero, there exists a constant $c \in \mathbb{R}^+$ such that $\min(\mathrm{eig}(\mathrm{D}_k)) \geq c$ for all k. Hence, the inequalities

$$\Phi'(x)d_k = -\Phi'(x_k)^{\mathrm{T}}\mathrm{D}_k\Phi'(x_k) \leq -\min(\mathrm{eig}(\mathrm{D}_k))\left\|\Phi'(x_k)\right\|^2 \leq -c\left\|\Phi'(x_k)\right\|^2 \quad \text{(C.1)}$$

holds for each k. Since now $\lim_{k\to\infty}\Phi'(x_k) = \Phi'(\bar{x}) \neq 0^{\mathrm{T}}$, the expression $\Phi'(x_k)d_k$ is bounded away from zero from some index onwards. Without loss of generality, only these tails of the sequences are considered in the rest of the proof.

Let the sequence $\{\tilde{\alpha}_k\} \in \mathbb{R}$ be chosen so that for all k

$$\tilde{\alpha}_k = \max\left\{\alpha \in (0, s] \mid \Phi(x_k) - \Phi(x_k + \alpha d_k) \geq -\frac{1}{2}\alpha\Phi'(x_k)d_k\right\}.$$

By Lemma C.3 direction d_k is a descent direction, so by Lemma C.2 this sequence is well-defined. Since α_k maximizes $\Phi(x_k) - \Phi(x_{k+1})$, relation $\Phi(x_k) - \Phi(x_{k+1}) \geq \Phi(x_k) - \Phi(x_k + \tilde{\alpha}_k d_k)$ holds, so

$$\Phi(x_k) - \Phi(x_{k+1}) \geq -\frac{1}{2}\tilde{\alpha}_k\Phi'(x_k)d_k$$

holds. Since $x_k \to \bar{x}$ and Φ is continuous, $\Phi(x_k) \to \Phi(\bar{x})$. Thus, $\Phi(x_k) - \Phi(x_{k+1}) \to 0$, so $-\frac{1}{2}\tilde{\alpha}_k\Phi'(x_k)d_k \to 0$, so because $\{\Phi'(x_k)d_k\}$ is bounded away from zero, $\tilde{\alpha}_k \to 0$ holds. Hence, $\tilde{\alpha}_k \leq \frac{s}{2}$ from some index onwards, so the inequality

$$\Phi(x_k) - \Phi(x_k + 2\tilde{\alpha}_k d_k) < -\tilde{\alpha}\Phi'(x_k)d_k$$

holds. By the mean value theorem there exist $\xi_k \in [0, \tilde{\alpha}_k]$ such that

$$\Phi(x_k + 2\tilde{\alpha}_k d_k) - \Phi(x_k) = \Phi'(x + \xi_k d_k) \cdot 2\tilde{\alpha}_k d_k,$$

which implies that

$$\Phi'(x_k + \xi_k d_k)d_k > \frac{1}{2}\Phi'(x_k)d_k.$$

Since $\{d_k\}$ is a bounded sequence, it has a convergent subsequence $\{d_{n_k}\}$, the limit of which is denoted by $\bar{d}$, and since $0 \leq \xi_k \leq \tilde{\alpha}_k$, $\xi_k \to 0$ holds. Hence, because

$$\Phi'(x_{n_k} + \xi_{n_k} d_{n_k})d_{n_k} > \frac{1}{2}\Phi'(x_{n_k})d_{n_k}$$

and because Φ' is continuous,

$$\Phi'(\bar{x} + 0 \cdot \bar{d})\bar{d} \geq \frac{1}{2}\Phi'(\bar{x})\bar{d}$$

holds. Thus, the inequality

$$\Phi'(\bar{x})\bar{d} \geq 0 \tag{C.2}$$

is true.

However, by Eq. C.1, the inequality

$$\Phi'(\bar{x})\bar{d}_k = \lim_{k\to\infty} \Phi'(x_{n_k})d_{n_k} \leq -c \lim_{k\to\infty} \|\Phi'(x_{n_k})\|^2 < 0 \tag{C.3}$$

holds. Inequalities (C.2) and (C.3) contradict, so $\Phi'(\bar{x}) = 0^{\mathrm{T}}$. Thus, the statement is true. □

Proposition C.5. *Let $\Phi : \mathbb{R}^{N_x} \to \mathbb{R}_0^+$ be a continuously differentiable function. Let $\{x_k\}$ be a sequence generated by the recursion*

$$x_{k+1} = x_k + \alpha_k d_k,$$

where $\{\alpha_k\}$ is a bounded sequence and for which there exists a constant $c \in \mathbb{R}^+$ such that

$$\forall k \in \mathbb{N}_0 : \ \|d_k\| \leq c \|\Phi'(x_k)\| .$$

Furthermore, assume that $\Phi(x_{k+1}) \leq \Phi(x_k)$ holds for all indices k and that every convergent subsequence of $\{x_k\}$ converges to a stationary point of Φ. Let x^* be a local minimum of Φ and the only stationary point of Φ within some open set.

Then there exists an open set S such that if $x_K \in S$ for some index K, then $x_k \in S$ for every $k \geq K$ and

$$\lim_{k\to\infty} x_k = x^*.$$

Proof. By the assumption we can choose a number $\rho \in \mathbb{R}^+$ such that x^* is the only stationary point of Φ in the set $\{x \mid \|x - x^*\| < 2\rho\}$ and that

$$\forall x \in \mathbb{R}^{N_x} : \ \|x - x^*\| < \rho \Rightarrow \Phi(x^*) < \Phi(x).$$

Let s be an upper bound of the sequence $\{\alpha_k\}$. Since Φ' is continuos, $\delta \in (0, \frac{\rho}{2}]$ can be chosen such that the relation

$$\forall x \in \mathbb{R}^{N_x} : \ \|x - x^*\| < \delta \Rightarrow \|\Phi'(x)\| = \|\Phi'(x) - \Phi'(x^*)\| < \frac{\rho}{2sc}$$

holds. Furthermore, let

$$C = \min_{\{x|\delta \leq \|x-x^*\| \leq \rho\}} \Phi(x)$$

and

$$S = \{x \mid \|x - x^*\| < \rho,\ \Phi(x) < C\}.$$

Let us now assume that $x_k \in S$ for some $k \in \mathbb{N}_0$. If $\|x_k - x^*\| \geq \delta$ held, by the definition of C either $\|x_k - x^*\| > \rho$ or $\Phi(x_k) \geq C$ would hold, which would contradict with the assumption that $x_k \in S$. Thus, $\|x_k - x^*\| < \delta$ holds, so

$$\begin{aligned}
\|x_{k+1} - x^*\| &= \|(x_k + \alpha_k d_k) - x^*\| \\
&= \|(x_k - x^*) + \alpha_k d_k\| \\
&\leq \|x_k - x^*\| + \|\alpha_k d_k\| \\
&= \|x_k - x^*\| + \alpha_k \|d_k\| \\
&< \delta + sc\,\|\Phi'(x_k)\| \\
&< \frac{\rho}{2} + sc \cdot \frac{\rho}{2sc} = \rho
\end{aligned}$$

holds. Furthermore, since by assumption $\Phi(x_{k+1}) \leq \Phi(x_k)$,

$$\Phi(x_{k+1}) \leq \Phi(x_k) < C$$

holds, so $x_{k+1} \in S$. Hence, by the induction principle, assumption $x_K \in S$ implies $x_k \in S$ for every $k \geq K$.

Let $\bar{S}$ be the closure of the set S. Since $\bar{S}$ is closed and bounded, it is sequentially compact, so by the Bolzano–Weierstrass theorem, the sequence $\{x_k\}$ has at least one convergent subsequence, which converges towards a point in $\bar{S}$. Since $\bar{S} \subset \{x \mid \|x - x^*\| < 2\rho\}$ holds, x^* is the only stationary point of Φ within $\bar{S}$. Since by assumption the limit of each convergent subsequence of $\{x_k\}$ is a stationary points of Φ and x^* is the only stationary point of Φ within $\bar{S}$, all the convergent subsequences of $\{x_k\}$ converge to x^*.

If now $\{x_k\}$ did not converge to x^*, there would be a subsequence $\{x_{n_k}\}$ such that for some $\epsilon_0 \in \mathbb{R}^+$ inequality $\|x_{n_k} - x^*\| \geq \epsilon_0$ holds for all k. By the Bolzano–Weierstrass theorem, this subsequence would have a convergent subsequence, which would not converge to x^*, since all its elements have the distance of at least ϵ_0 to x^*. This contradicts with the assumption that every convergent subsequence of $\{x_k\}$ converges to x^*. Thus, $\lim_{k\to\infty} x_k = x^*$. □