Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Radio Engineering

Positioning and Data Fusion Algorithms for Radionavigation Systems

Doctoral Thesis

Václav Navrátil

Prague, June 2019

Ph.D. Programme:	P2612 – Electrical Engineering and Information Technology
Branch of Study:	2601V010 – Radioelectronics

Supervisor:	prof. Ing. František Vejražka, CSc.
Supervisor-Specialist:	doc. Ing. Josef Dobeš, CSc.

Abstract

The doctoral thesis reviews various possibilities of radio signal exploitation for positioning purpose and presents the respective theoretical accuracy bounds as well as their inherent limitations and drawbacks. Applications of the basic principles are shown on the real systems, such as GNSS or UWB localization systems.

The positioning is always performed on the basis of several observables; such problems are typically nonlinear. Within the thesis, several iterative algorithms that are able to solve the "navigation equations" are provided and applied on real problems. Methods of integration of several radionavigation systems and possibly other systems (such as inertial navigation) are described as well. The thesis is focused on the Kalman filtering approaches, in particular.

The author's innovative contribution can be seen mainly in development of a precise wireless synchronization algorithm for UWB localization networks, investigation of stochastic properties for the asymmetric double-sided two-way ranging estimator, development of a soft-constrained error-state Kalman filter for GNSS/UWB/IMU integrated system, and evaluation of the performance bound of antenna-pattern-based angle-of-arrival estimator under Rician channel conditions.

Keywords

radionavigation, integrated navigation, Kalman filtering, GNSS, UWB

Abstrakt

Tato disertační práce shrnuje několik způsobů využití radiových signálů pro účely určování polohy, popisuje jejich teoretické meze přesnosti, přirozená omezení a nedostatky. Aplikace těchto základních principů jsou demonstrovány na reálných systémech jako např. GNSS nebo UWB lokalizačních systémech.

K určování polohy je vždy využito měřených parametrů radiových signálů, jejichž vztah k samotné informaci o poloze je obvykle nelineární. V rámci práce je popsáno několik iterativních algoritmů, které jsou vhodné k řešení nelineárních soustav „navigačních rovnic.“ Některé z algoritmů jsou aplikovány na reálné problémy. Jsou také popsány metody integrace několika radionavigačních systémů, případně dalších systémů (např. inerciální navigace). Práce se zaměřuje zejména na metody Kálmánovy filtrace.

Inovativní přínos autora lze spatřovat zejména ve vývoji algoritmu pro přesnou synchronizaci UWB lokalizačních sítí, odvození stochastických parametrů asymetrického dvojstranného obousměrného odhadu vzdálenosti (ADS-TWR), vývoje chybového Kálmánova filtru s měkkými omezujícími podmínkami pro integrovaný GNSS/UWB/IMU systém a odvození dolní meze přesnosti odhadu směru příchodu signálu založeného na tvaru směrové charakteristiky antény.

Klíčová slova

radionavigace, integrovaná navigace, Kálmánova filtrace, GNSS, UWB

Acknowledgements

I would like to thank my friends, colleagues and professors from the Czech Technical University; in particular, to my supervisors prof. František Vejražka and doc. Josef Dobeš for the guidance during the process of writing this thesis, to Josef Krška and Zdeněk Fiala for the help with testing the UWB-related algorithms, and to Rostislav Karásek for the collaboration on the DVB-T localization. Further, I would like to express my gratitude to the SPIN Laboratory team at The Ohio State University (Zoltán Koppányi, Andy Xu, Dorota Grejner-Brzezinska and Charles Toth) for the opportunity to work on the constrained sensor fusion algorithm.

My deepest thanks go to my family, my parents Dana and Petr and my brother Michal, and to Eva, for the infinite patience, support, motivation and also for the effort to rectify the use of English language within the thesis.

Contents

Annotation, Keywords (Anotace, Klíčová slova) i

Acknowledgements v

Contents vii

1. Introduction 1

2. Radionavigation Systems 7
2.1. Elementary Principles of Radio Positioning . . . 7
2.1.1. Signal strength . . . 7
2.1.2. Signal delay . . . 12
2.1.3. Doppler-effect based positioning . . . 23
2.1.4. Angle of arrival . . . 26
2.2. Global Navigation Satellite Systems . . . 37
2.2.1. Observables, errors and limitations . . . 37
2.2.2. GNSS positioning methods . . . 50
2.2.3. Overview of the available systems . . . 62
2.3. Ultra-Wide Band Positioning . . . 68
2.3.1. Available positioning methods . . . 69
2.3.2. Error sources . . . 73
2.3.3. Brief summary of UWB localization . . . 79
2.4. LF and VLF systems . . . 79
2.4.1. Omega . . . 79
2.4.2. Loran family . . . 80
2.5. Local terrestrial navigation systems . . . 80
2.5.1. Locata . . . 81
2.5.2. NextNav . . . 81
2.6. Signals of Opportunity . . . 82
2.6.1. Basic principle of SoO processing . . . 82
2.6.2. Examples of SoO use . . . 85

3. Inertial Navigation 87
3.1. Coordinate Frames and Attitude Parametrization . . . 88
3.1.1. Attitude parametrization and rotation formalisms . . . 88
3.1.2. Coordinate frames . . . 95
3.2. Strapdown IMU mMechanization . . . 99
3.2.1. Attitude update . . . 100
3.2.2. Velocity update . . . 101
3.2.3. Position update . . . 104
3.2.4. Initialization . . . 105

3.3. Effects of the IMU Errors . . . 106
3.3.1. IMU measurement errors . . . 106
3.3.2. Bias and random walk propagation . . . 107

4. Positioning Solution and Data Fusion 111
4.1. Iterative Solutions of Positioning Equations . . . 111
4.1.1. Newton-Raphson algorithm . . . 112
4.1.2. Gradient descent algorithm . . . 113
4.1.3. Levenberg-Marquardt algorithm . . . 114
4.1.4. Termination conditions . . . 115
4.1.5. Application on the navigation problems . . . 116
4.1.6. Estimate covariance and dilution of precision . . . 120
4.1.7. Pseudomeasurement as a soft equality constraint . . . 122
4.2. Kalman Filtering and Data Fusion . . . 124
4.2.1. Linear Kalman filter . . . 124
4.2.2. Exended Kalman filter . . . 126
4.2.3. Unscented Kalman filter . . . 128
4.2.4. Constraining the Kalman filter . . . 131
4.2.5. Kalman smoother . . . 133
4.3. Alternative methods of data fusion . . . 134

5. Conclusion 135

Appendices:

A. Bias and Variance of the Asymmetric Double-sided TWR 141
A.1. Approximation of the Estimator Performance Parameters . . . 141
A.2. Evaluation for the Particular Estimators . . . 144
A.3. Remarks . . . 159
A.4. The Complete Jacobian and Hessian Matrices . . . 160

B. Examples of TDoA Positioning 163
B.1. DVB-T Opportunistic Localization . . . 163
B.1.1. Obtaining the TDoA measurements . . . 163
B.1.2. Applying the Newton-Raphson algorithm . . . 163
B.1.3. Experimental results . . . 165
B.2. Ultra-Wide Band System . . . 167
B.2.1. General definition of the TDoA measurements . . . 167
B.2.2. Applying the (un)constrained weighted solver . . . 168
B.2.3. Experimental results . . . 169

C. Dual Application of Linear KF in UWB systems 175
C.1. Synchronization . . . 175
C.1.1. Two-node synchronization . . . 175
C.1.2. Chained synchronization . . . 181
C.1.3. Remarks . . . 183
C.2. Position Filtering . . . 185

D. Application of Constrained EKF for UXO Mapping **191**
D.1. The IMU Mechanization . . . 192
D.2. The Error-state Kalman Filter . . . 193
D.3. Motion Constraints . . . 195
D.4. Forward-backward Solution Combination . . . 198
D.5. Experimental Results . . . 199
D.6. Remarks . . . 201

Bibliography **205**

Lists:
List of Acronyms . . . 219
List of Symbols and Operators . . . 222
List of Figures . . . 225
List of Tables . . . 227
List of Author's Publications and Citations . . . 229

1. Introduction

Before the aims of the thesis are presented, let us take a closer look at how the word *navigation* is understood today. This will provide a justification of the work that will be described within this thesis.

Firstly, the etymology of *navigation* is investigated. In the original sense the verb *to navigate* means to sail, since it comes from a Latin verb *navigare* that is a composite of noun *navis* – ship and verb *agere* – to drive. Clearly, the meaning of this world has changed throughout the centuries. Navigation is no longer limited to maritime voyage, but it rather describes an effort to guide someone or something from point A to point B.

The navigation (in the modern sense) is quite a complex set of problems. Therefore it is practical to divide it into three inter-connected tasks:

Positioning (Sensing) – The first task is to determine the current position of the user or the platform. In various applications velocity and attitude and precise time are also determined. Moreover, other parameters are useful to be sensed, for instance wind components for aircraft.

Route Planning – During this phase a route to the destination is prepared, for certain applications the destination is chosen at this point. The route can be planned with respect to various aspects, e.g. timeliness, economy, distance, risk, etc. A map and other datasets are required to plan the route effectively.

Steering (Actuation) – Naturally, the last point is affecting the movement of the navigated platform in order to follow the planned trajectory with certain parameters. This task is still left to the user (driver, pilot, etc.) commonly.

These tasks are to be iterated in a loop-like manner to successfully navigate the platform. As an example, we may take an aircraft flying towards a beacon using automatic direction finder (ADF) under crosswind conditions. From the ADF, the pilot knows the azimuth towards a non-directional beacon (NDB) and the aircraft's attitude – this is the *sensing* phase. Then, the direct path to the NDB is chosen, which is a trivial example of route *planning*. Consequently, the pilot uses the yoke and pedals to *actuate* the control surfaces to steer the aircraft heading to match the direction towards the NDB. Nonetheless, the aircraft is pushed sideways by the wind and therefore, the azimuth towards changes as can be seen in the Fig. 1.1a. The pilot has to react to this change, otherwise the aircraft would miss the beacon; the navigational sensing-planning-actuation loop has to be closed.

Intuitively, the success of the navigation is dependent on the available data. For instance, when airspeed and crosswind component are known, it is possible to mitigate the wind effect. The heading is corrected by a proper crab angle to fly against the wind in order to stay on the linear trajectory, see Fig. 1.1b.

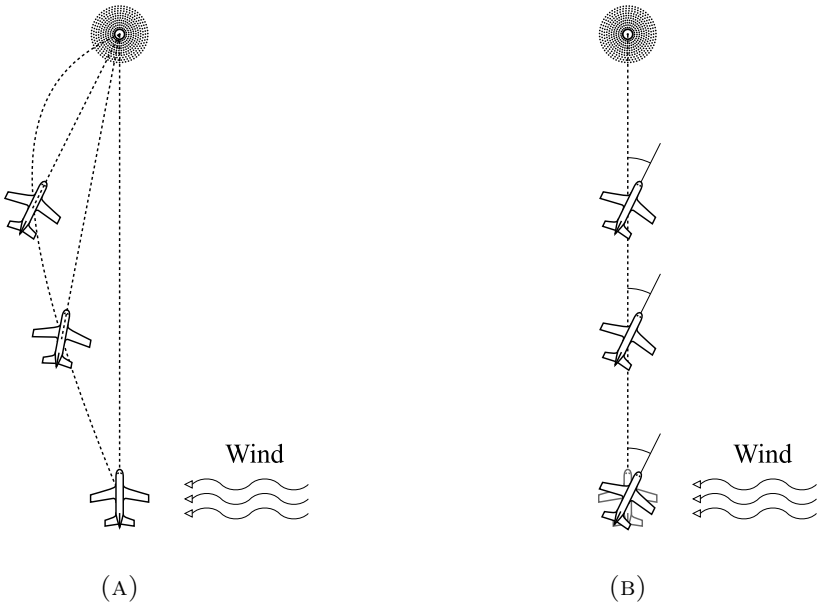


FIGURE 1.1.: Aircraft navigated towards a beacon.

This thesis does not target the navigation in its entirety, but instead it focuses on one of the three inherent tasks – the positioning. In particular, the thesis aims on the radio-based positioning. Various radionavigation systems have been developed since the discovery of radio waves in the late 19^{th} century. Of course, many of the systems are obsolete and replaced by modern systems, but the physical principles used for position determination remain the same. Even though the modern systems use advanced signals and smart methods of their processing, neither of them can provide a reliable and ubiquitous service in a global scale. Therefore, part of the thesis is devoted to methods, which can be used for fusion of multiple radionavigation or alternative systems.

Thesis organization

The text of the thesis targets the positioning from a rather theoretical standpoint; the practical application of the described principles is provided in the appendices.

Chapter 2 begins with the review of the basic principles of the radionavigation, their theoretical performance bounds (Cramer-Rao lower bound, CRLB) are presented. For some of the methods the bounds were already available in literature (e.g. [1, 2]), however, in certain cases the CRLB was derived by the author.

It is worth pointing out that CRLB for the antenna-pattern based angle of arrival measurement (AoA, Sec. 2.1.4–I) was evaluated not only for a simple log-normal fading channel (which is commonly used) model, but for the Rician channel model as well. A prior application of the latter of the models to the signal-power based AoA accuracy bound evaluation is not known to the author. Such model is able to provide the results for the signals that are weak and thus near the noise floor, power-wise.

A very detailed discussion is devoted to the signal-delay based localization (Section 2.1.2), as this is currently the most widely used principle of obtaining the position estimates. It is used extensively in the author's work as well.

One of the intentions was to show and exploit the link between the particular types of measurements – ToA (time of arrival), TDoA (time difference of arrival) and TWR (two-way ranging). The three methods can be interpreted as different approaches of dealing with the nuisance parameters – the clock offsets. In the ToA and TDoA cases the localizing infrastructure is synchronized, therefore only the infrastructure-to-user clock offset has to be dealt with. It can be eliminated by means of measurement differentiation (TDoA), when the time information is not relevant, or estimated from multiple measurements (ToA); in both cases, unidirectional communication is sufficient.

The TWR is dealing with a non-synchronized pairs of nodes, thus it has to eliminate (or estimate) the clock offset in each measurement pair by means of bi-directional timestamped communication. The performance of several methods is investigated, both in terms of possible accuracy (CRLB) and immunity to clock drift in the measurement nodes. Particularly, a novel Asymmetric Double-Sided TWR (ADS-TWR) proposed in [3] is thoroughly examined. The method promises excellent immunity to clock drift, nonetheless, range estimate may be biased due to its nonlinear nature, yet the estimator variance has not been assessed theoretically. In order to avoid breaking the flow of the thesis text, Appendix A is designated for the detailed ADS-TWR analysis.[1]

The connection of the delay-based positioning to the Doppler-based position and velocity estimation is shown and certain limitations are presented. In Section 2.1.4–II the results of the delay-based methods are further extended to evaluate the accuracy of AoA estimation by means of antenna arrays. It is worth remarking that the difference between narrowband (signal phase) and wideband (signal delay) approaches is highlighted.

The rest of Chapter 2 briefly introduces a couple of different navigation systems, the respective underlying radio-positioning principles (which were already described in the beginning of the chapter) are pointed out. The substantial parts are devoted to the global navigation satellite systems GNSS (using ToA method), which are inarguably the most extensively used radio-positioning systems of today. The section reviews various methods of GNSS positioning, their generally anticipated performances are compared. More importantly, the fundamental limits, such as low-power of the received signals and the requirement of direct lines of sight between receiver and satellites, is pointed out.

The section concerning the ultra-wide band UWB positioning systems compares the application of TWR and TDoA positioning methods. It is shown in the thesis that UWB systems are able to provide outstanding accuracy due to their signal characteristics regardless the method used. Analysis of both, random and systematic, errors in the UWB localization is provided within Section 2.3. The methods of their suppression or mitigation are included as well.

[1] The results of the ADS-TWR bias and variance analysis are to be published as [4].

The later approach is able to provide substantially more frequent position fixes, however, it requires precise synchronization of the UWB infrastructure in order to be able to perform TDoA measurements. A novel algorithm for synchronization of the UWB infrastructure nodes, which was developed by the author, is described in detail within Appendix C.1.[2] Additionally, the solution of a set of TDoA equations is presented in Appendix B.2.

At the end of the second chapter, the exploitation of Signals of Opportunity (SoO) for navigation is mentioned.[3] The TDoA approach is applicable when utilizing signals of a single frequency network of DVB-T transmitters. The application of iterative methods for solving this particular TDoA problem is provided in Appendix B.1.

The purpose of Chapter 3 is to provide a basic knowledge about processing of inertial measurements in order to obtain position, velocity and attitude estimates. The background from this chapter is vastly exploited within Appendix D, where an experimental integrated localization system for an UXO mapping platform is described. Briefly, the effects of common inertial sensor errors are shown on the results of a simulation.

Chapter 4 introduces in its first part several iterative methods suitable for solving navigation problems in a least-squares manner. A pseudomeasurement approach of soft equality constraint inclusion into the least-squares solver was described in Section 4.1.7. The practical application is demonstrated in the appendices. The Newton-Raphson method was used for solving the TDoA problem in DVB-T based positioning (Appendix B.1); application of constrained Levenberg-Marquardt algorithm is available in Appendix B.2, where the set of TDoA equations for an UWB localizing system is solved.

The second part of Chapter 4 focuses on Kalman filters and their use in state estimation and fusion of multiple navigation systems into an integrated system. The basic equations for linear, extended and unscented Kalman filters are presented and commented in order to provide the reader with useful clues for application of KFs in the navigation-related applications. An effort is made to clearly state the limitations of the various approaches. The differences in handling of the nonlinear models by the EKF and UKF are shown. Two examples of a linear Kalman filters are provided in Appendix C; the first use is the estimation of the clock state of UWB infrastructure node, which is the critical part of the chained wireless synchronization algorithm developed by the author.

Section 4.2.4 introduces a pseudomeasurement approach of incorporation soft equality constraints into the Kalman filters. The advantage of this approach can be seen in the ease of implementation and modifications; the state constraints may be defined in a rather straightforward way. The application of such soft constraints applied on a EKF is available in Appendix D.

The "linear" introduction of the thesis structure does not seem sufficient, because it does not express the interconnections between the sections from different chapters. Therefore, a mindmap visualization of the important sections of the thesis is provided as Figure 1.2.

[2] The algorithm and its results were already published in author's work [5, 6].

[3] It is worth remarking that the methods of SoO processing are similar to passive radar signal processing in many aspects.

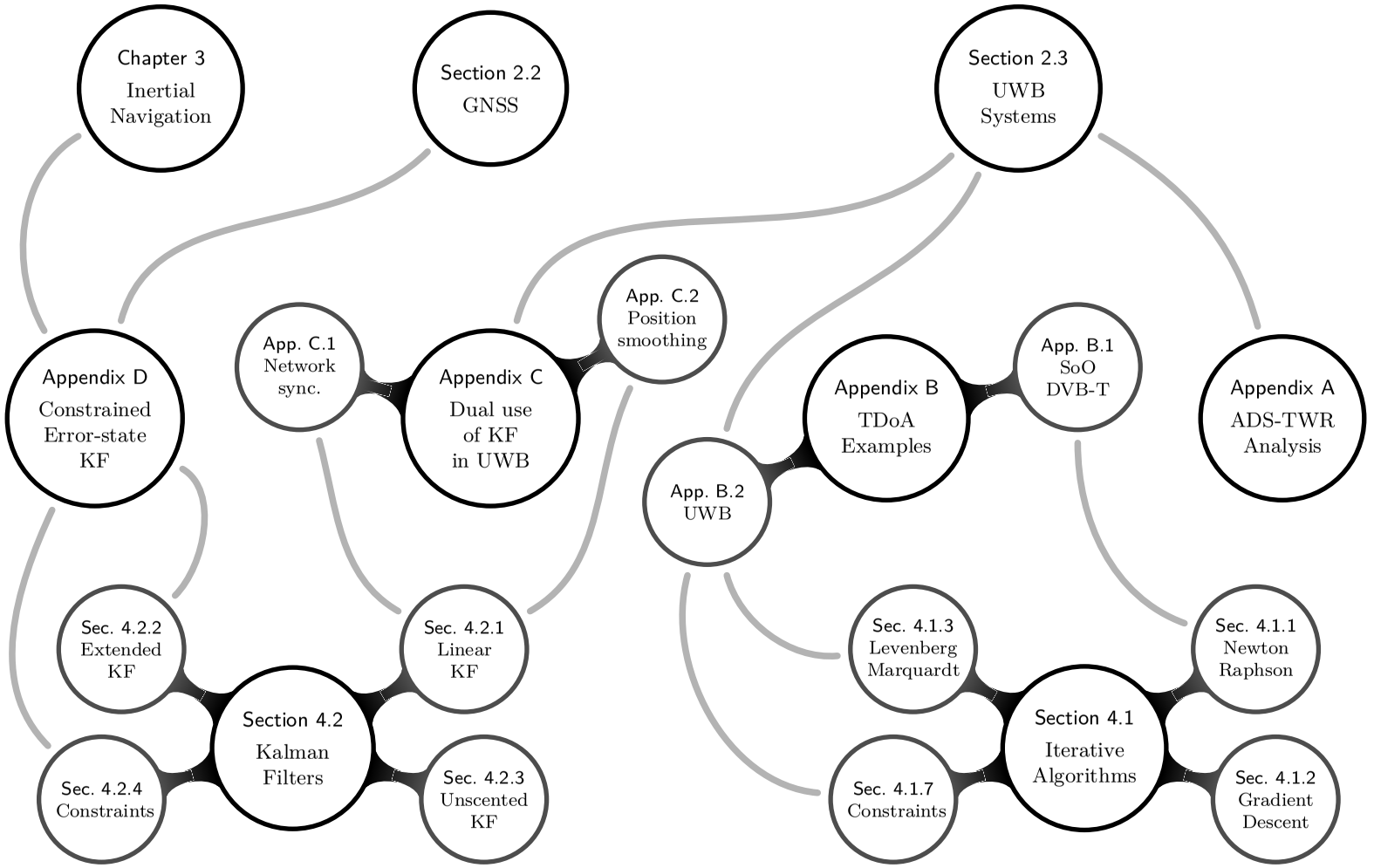


FIGURE 1.2.: Mindmap of the thesis contents.

2. Radionavigation Systems

This chapter provides an overview of the systems for terrestrial radionavigation. Certainly, it is not an exhaustive list and description of all the principles and systems, since this would be out of the scope of the thesis; hence, this chapter is intended to serve as a guide that briefly summarizes the most important aspects of the different approaches to radionavigation and also the particular systems.

In the first section we will focus on the principles, the latter sections provide insight into the particular systems, or their groups that exploit some of the principles mentioned in the former one. An emphasis will be put on the ways of exploitation of various signal characteristics and the respective consequences on the signal performance.

2.1. Elementary Principles of Radio Positioning

Only a few basic principles of radio positioning exist – they are based on exploiting of signal characteristics that differ spatially in a determined way. Most generally, three very essential phenomena are of interest for positioning: the decay of the electromagnetic wave power in free space, the finite velocity of propagation of electromagnetic waves, and the Doppler effect.

2.1.1. Signal strength

I. Physical principle and possible accuracy

The first effect on the waves to exploit (or not to exploit) for the purpose of positioning is the signal power. Consider an isotropic point radiator of electromagnetic waves placed in the isotropic free-space lossless environment. The waveplanes emitted would be spherical in such case. Also, the infinitesimal energy that is carried in a single waveplane is constant due to the lossless environment, however, the area of the waveplane grows quadratically as the wave propagates further from its origin. Thus, the power density decreases quadratically.

It is possible to derive the Friis transmission equation for wireless link power budget under ideal conditions

$$P_{\mathrm{Rx}} = P_{\mathrm{Tx}} G_{\mathrm{Tx}} G_{\mathrm{Rx}} \left(\frac{\lambda}{4\pi r} \right)^2, \tag{2.1}$$

where P_{Rx} and P_{Tx} denote received and transmitted power, G_{Tx} and G_{Rx} are the respective antenna gains. Symbol λ denotes the wavelength of the signal and r is the distance between transmitter and receiver. The squared term in the bracket is commonly referred as free-space loss (FSL). One may come to an invalid conclusion that there is a simple determined relation between the range and the signal power. The

Friis equation is a good approximation for free space, as would be for example in space. Under terrestrial conditions, the environment exhibits lossy atmosphere and numerous obstacles such as buildings, vehicles, land, water surface, *et cetera*. As anticipated, the obstacles affect the propagation of the electromagnetic waves – they cause attenuation, reflection and diffraction of the waves. Those scattered waves are then combined at certain points in a constructive or destructive way.

Nonetheless, it is impossible to describe the environment completely, moreover it is subject to rapid changes. The propagation can therefore be characterized in a probabilistic manner, based on empirical data. One of the simplest empirical models used for propagation in real environment, the one-slope model, follows

$$\begin{aligned} P(r) &= P(r_0)\left(\frac{r}{r_0}\right)^{-n} \\ P_{\mathrm{dB}}(r) &= P_{\mathrm{dB}}(r_0) - 10n\log_{10}\left(\frac{r}{r_0}\right), \end{aligned} \tag{2.2}$$

where $P(r)$ is signal power received at a distance r, r_0 is a short distance where the received power is assumed to be $P(r_0)$, and finally n denotes the path loss exponent. The second line of the equation above describes the same model in the decibel scale. By comparison with the free-space loss bracket in (2.1) it can be seen that for the ideal free space propagation the n exponent is equal to 2; however in the real environment it can be both higher and lower.[4]

For the sake of evaluation of the lower bound of distance measurement – the Cramér–Rao lower bound (CRLB) – a model has to be constructed. For simplicity, let us model large-scale signal fading only; a log-normal random distribution is considered. We neglect the contribution of thermal noise, it is assumed that its influence is suppressed by averaging of the measured value. It is convenient to use the second line expression of the one-slope model (2.2), since the log-normal distribution becomes a normal distribution in the decibel scale. Let us denote the randomly faded signal power at the distance r as $\tilde{P}(r)$ in the linear scale and as $\tilde{P}_{\mathrm{dB}}(r)$ in the decibel scale. Then the model yields

$$\tilde{P}_{\mathrm{dB}}(r) = 10\log_{10}\tilde{P}(r) = \mathcal{N}\left(P_{\mathrm{dB}}(r),\, \sigma^2_{\mathrm{lf}}\right) = P_{\mathrm{dB}}(r) + \mathcal{N}\left(0,\, \sigma^2_{\mathrm{lf}}\right), \tag{2.3}$$

where $\mathcal{N}\left(P_{\mathrm{dB}}(r),\, \sigma^2_{\mathrm{fade}}\right)$ denotes the normal (Gaussian) distribution with mean corresponding to the one-slope model and variance σ^2_{lf} equal to an empirical value for given environment.[5] For instance, the values used for UWB channel models (2–10 GHz frequency band) are presented within Table 2.1.

In order to obtain maximal possible accuracy of an unbiased range estimator that is based on signal power measurements, we have to investigate the corresponding Cramér–Rao lower bound (CRLB). Generally, this bound gives us the minimum possible vari-

[4] The path loss exponent value lower than 2 means that the signal decays slower than predicted by the Friis transmission equation, which may appear as a non-physical behavior – the energy in the spherical waveplane would grow with the distance from the wave origin. It has to be considered that the model is empirical and based on measured data. The spherical wave can not be assumed anymore in the environment with obstacles. Because the propagation of the wave is limited (between floor and ceiling in the office environment, for instance), the wave energy decays in a less-than-quadratic manner.

[5] While the exponent is dimensionless, the unit of the fading variance should be dB^2.

ance of an unbiased estimate of a certain parameter [1]. For a single-variate estimator it follows

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{\boldsymbol{\mathcal{I}}(\theta)}, \tag{2.4}$$

where θ is the parameter and $\hat{\theta}$ its estimate. Symbol $\boldsymbol{\mathcal{I}}(\theta)$ denotes the Fisher information that is for the single-variate case equal to

$$\boldsymbol{\mathcal{I}}(\theta) = \mathrm{E}\left[\left(\frac{\partial \Lambda(\theta|X)}{\partial \theta}\right)^2\right], \tag{2.5}$$

where $\Lambda(\theta|X)$ is the likelihood function of the parameter θ with respect to the random observation X.

Here, the CRLB that assumes the one-slope propagation model will be derived; however, it is possible to construct the CRLB for different model, e.g. Rician fading [8]. Firstly, we have to derive the likelihood function $\mathcal{L}$ of the distance parameter r. Let us consider the reference distance unitary $r_0 = 1\,\mathrm{m}$ and denote the reference power by P_{dBref}. Denoting probability density function (PDF) by $\mathrm{p}(\cdot)$ and natural exponential $\exp(\cdot)$ it can be written

$$\begin{aligned}\mathcal{L}(r|\tilde{P}_{\mathrm{dB}}) &= \mathrm{p}(P_{\mathrm{dB}}|r) \\ &= \frac{1}{\sqrt{2\pi\sigma_{\mathrm{lf}}^2}} \exp\left(-\frac{\left(\tilde{P}_{\mathrm{dB}} - P_{\mathrm{dBref}} + 10n\log_{10} r\right)^2}{2\sigma_{\mathrm{lf}}^2}\right).\end{aligned} \tag{2.6}$$

It is necessary to derive the log-likelihood, which is the natural logarithm $\ln(\cdot)$ of the likelihood function:

$$\begin{aligned}\Lambda(r|\tilde{P}_{\mathrm{dB}}) &= \ln \mathcal{L}(r|\tilde{P}_{\mathrm{dB}}) \\ &= \ln\left(\frac{1}{\sqrt{2\pi\sigma_{\mathrm{lf}}^2}}\right) - \frac{\left(\tilde{P}_{\mathrm{dB}} - P_{\mathrm{dBref}} + 10n\log_{10} r\right)^2}{2\sigma_{\mathrm{lf}}^2}.\end{aligned} \tag{2.7}$$

Table 2.1.: Common values of channel model coefficients [2, 7]

Environment	**Path loss exponent** n	**Fading variance** σ_{lf}^2
Ideal	2.00	0.00
Residental LOS	1.79	2.22
Residental NLOS	4.58	3.51
Office LOS	1.63	1.90
Office NLOS	3.07	3.90
Outdoor LOS	1.76	0.83
Outdoor NLOS	2.50	2.00
Industrial LOS	1.20	6.00
Industrial NLOS	2.15	6.00

Further, the log-likelihood is to be differentiated with respect to the parameter r; by means of the chain rule we get the result

$$\begin{aligned}\frac{\partial \Lambda(r|P_{\mathrm{dB}})}{\partial r} &= \frac{-2}{2\sigma_{\mathrm{lf}}^2}\left(\tilde{P}_{\mathrm{dB}} - P_{\mathrm{dBref}} + 10n\log_{10}r\right)\frac{\partial}{\partial r}\left(\tilde{P}_{\mathrm{dB}} - P_{\mathrm{dBref}} + 10n\log_{10}r\right)\\ &= \frac{-10n}{r\sigma_{\mathrm{lf}}^2\ln(10)}\left(\tilde{P}_{\mathrm{dB}} - P_{\mathrm{dBref}} + 10n\log_{10}r\right). \end{aligned} \tag{2.8}$$

By comparing with the last expression of the one-slope model (2.2) it is clear that the term within the brackets is a normally distributed variable with variance σ_{lf}^2 and zero mean. Therefore, the Fisher information for this particular case is

$$\begin{aligned}\boldsymbol{\mathcal{I}}(r) &= \mathrm{E}\left[\left(\frac{-10n}{r\sigma_{\mathrm{lf}}^2\ln(10)}\mathcal{N}(0,\,\sigma_{\mathrm{lf}}^2)\right)^2\right]\\ &= \left(\frac{-10n}{r\sigma_{\mathrm{lf}}^2\ln(10)}\right)^2\mathrm{E}\left[\left(\mathcal{N}(0,\,\sigma_{\mathrm{lf}}^2)\right)^2\right] = \left(\frac{10n}{r\sigma_{\mathrm{lf}}\ln(10)}\right)^2. \end{aligned} \tag{2.9}$$

We justify the modifications above by the facts that the second moment of the zero-mean Gaussian distribution $\mathrm{E}\left[\left(\mathcal{N}(0,\,\sigma^2)\right)^2\right] = \sigma^2$ and that the minus sign cancels out due to the squaring. Consequently, it is straightforward that the CRLB of range measurement based on signal power measurement is

$$\sqrt{\mathrm{var}(\hat{r})} \geq \frac{\ln(10)}{10}\frac{r\,\sigma_{\mathrm{lf}}}{n} \tag{2.10}$$

when expressed in the convenient scale of standard deviation.

By substituting UWB channel parameter values from Tab. 2.1 into (2.10) we may obtain the dependence of CRLB on the true range, see Fig. 2.1. Obviously, the bound grows linearly with the true distance. In an intuitive way the slope grows with the empirical large scale signal fading variance; in contrast, the path-loss exponent reduces the slope. Although this may appear strange at first, the quicker power decay implies that the measured value changes are larger with the same perturbation of range, an therefore the influence of the fading is suppressed. Nonetheless, this model neglects the existence of the noise floor, which implies that beyond a certain range only the power of the noise of the receiver and environment is measured. It is worth noting that the CRLB for the ideal conditions corresponding to the Friis transmission equation is equal to zero due to its purely deterministic nature, which is expressed by the fading variance set to zero.

Clearly, a method of ranging based on power measurement cannot benefit from the signal structure. Moreover, the fundamental limit of accuracy depends on the environment factors only, the signal parameters such as transmitted power do not influence the CRLB.[6] Also, it will be shown in the following sections that other methods of range or position estimation overperform the signal-power approach.

[6] Should the real noise floor be assumed, the transmitted power would matter, since it extends the range and increases the power margin above the noise floor.

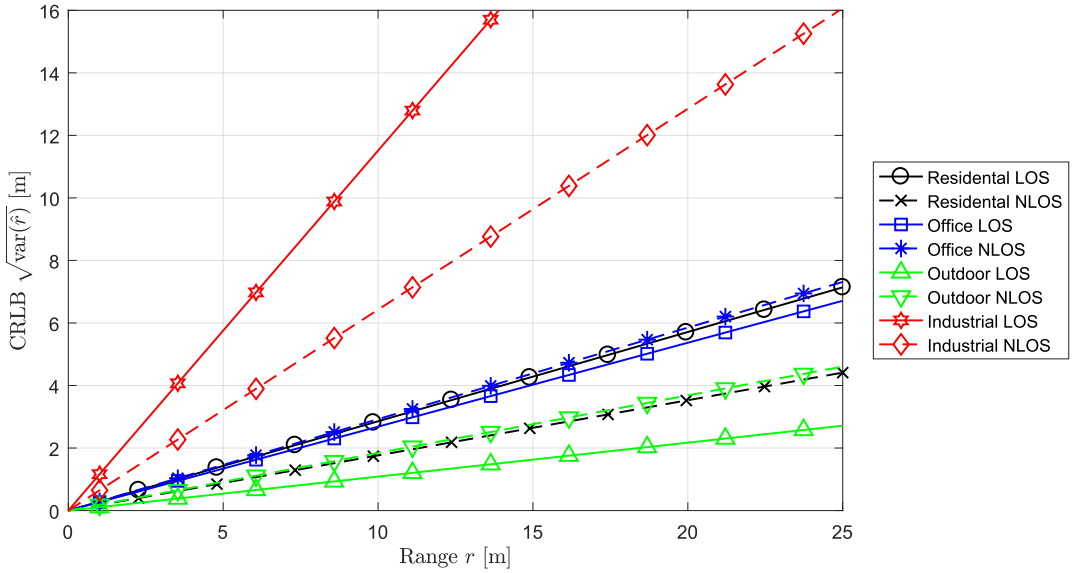


FIGURE 2.1.: Cramér–Rao lower bound of range estimate based on signal power measurement in different environments.

However, the received signal power from multiple radio sources can be exploited in a different way. A map of signal power "fingerprints" can be constructed for a certain area, and later the user can be located by comparing actual received power data with the available radio map. The map can be pre-surveyed, or simultaneous localization and mapping (SLAM) algorithms can be employed in order to construct the map along with the localization. Nonetheless, neither of these approaches is in the scope of the thesis.

From the geometrical point of view, we may obtain a set of ranges between the user equipment and a couple of other radio nodes. The equation for the range between the user equipment and the radio node with index i is

$$r_{\mathrm{u}i} = \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_i\| \tag{2.11}$$

$$= \begin{cases} \sqrt{(r_{\mathrm{u};x} - r_{i;x})^2 + (r_{\mathrm{u};y} - r_{i;y})^2} & \text{in two dimensions,} \\ \sqrt{(r_{\mathrm{u};x} - r_{i;x})^2 + (r_{\mathrm{u};y} - r_{i;y})^2 + (r_{\mathrm{u};z} - r_{i;z})^2} & \text{in three dimensions,} \end{cases} \tag{2.12}$$

where the position vectors $\boldsymbol{r}_{\mathrm{u}}$, $\boldsymbol{r}_i$, which respect a Cartesian system, have two or three elements, i.e. $r_{\mathrm{u};x}$, $r_{\mathrm{u};y}$ and $r_{\mathrm{u};z}$. Provided that the position vectors of the non-user nodes $\boldsymbol{r}_i$ are known, it is required to have at least the same number of measurements as is the dimensionality of the position vector in order to be able to solve for user position vector $\boldsymbol{r}_{\mathrm{u}}$. The methods for solving such systems of nonlinear equations, and remarks on solution availability and quality are provided in Chapter 4.

2.1.2. Signal delay

I. Physical principle and possible accuracy

The other phenomena that is widely exploited by positioning systems is the delay caused by the finite velocity of propagation of the electromagnetic waves. This velocity is known as the speed of light and in vacuum it is equal to $c_0 = 299\,792\,458\,\mathrm{ms}^{-1}$. The propagation delay for an arbitrary environment can be described by integrating the generally non-constant refractive index n along the path l from a transmitter to a receiver (denoted by Tx and Rx) [9]:

$$\tau = \int_{\mathrm{Tx}}^{\mathrm{Rx}} \frac{n(l)}{c_0} \mathrm{d}l. \tag{2.13}$$

The value of the refractive index is determined by relative permitivity (dielectric constant) ϵ_r and relative permeability μ_r; however, for non-magnetic materials it is almost equal to 1 and therefore the approximation

$$n = \sqrt{\epsilon_r \mu_r} \approx \sqrt{\epsilon_r} \tag{2.14}$$

is valid. The refractive index of air is higher than and almost equal to 1. Such approximation is applicable for short range applications, anyway, for long range or satellite systems the delay caused by non-unitary n has to be accounted for and corrected, mostly by means of empirical models (see e.g. section 2.2.2–I).

Regardless the impact of the varying refractive index, the fundamental limit of the ranging or delay measurement accuracy is the CRLB of the actual measurement. The derivation of the CRLB is available e.g. in [1] Assuming the model of a channel with AWGN and delay only, the received signal $s_{\mathrm{Rx}}(t)$ becomes the replica of the transmitted signal $s_{\mathrm{Tx}}(t)$ with the noise $w(t)$:

$$s_{\mathrm{Rx}}(t) = s_{\mathrm{Tx}}(t - \tau) + w(t). \tag{2.15}$$

Note that no attenuation of the signal is modeled, since its effect can be emulated by increasing the noise variance, i.e. decreasing SNR:

$$\sqrt{\mathrm{var}(\hat{\tau})} \geq \frac{1}{2\pi \sqrt{\frac{2\mathcal{E}}{N_0}} B_{\mathrm{RMS}}} = \frac{1}{2\pi \sqrt{\mathrm{SNR}}\, B_{\mathrm{RMS}}} \tag{2.16}$$

$$\sqrt{\mathrm{var}(\hat{r})} \geq \frac{c}{2\pi \sqrt{\frac{2\mathcal{E}}{N_0}} B_{\mathrm{RMS}}} = \frac{c}{2\pi \sqrt{\mathrm{SNR}}\, B_{\mathrm{RMS}}}. \tag{2.17}$$

where $\mathcal{E}$ denotes signal energy, $N_0/2$ is the constant double-sided power spectral density (PSD) of the Gaussian noise and B_{RMS} is the RMS bandwidth. The second line is the same CRLB scaled by propagation velocity c, in order to show the minimum possible variance of obtaining range estimate $\hat{r}$. First, let us show the definitions of the measures mentioned above. The signal energy is equal to

$$\mathcal{E} = \int_{-\infty}^{+\infty} |s(t)|^2 \mathrm{d}t. \tag{2.18}$$

Note that $\frac{2\mathcal{E}}{N_0}$ can be substituted by SNR under assumption sufficiently long observation time $2T$, since the white noise is strictly ergodic:

$$\mathrm{SNR} = \frac{\frac{1}{2T}\,\mathcal{E}}{\frac{1}{2T}\int_{-T}^{T}|w(t)|^2\mathrm{d}t} = \frac{\frac{1}{2T}\,\mathcal{E}}{\frac{1}{2T}\mathrm{E}\left[|w(t)|^2\right]} = \frac{2\mathcal{E}}{N_0}, \tag{2.19}$$

where $E[\cdot]$ is the expected value operator. According to [10] the RMS bandwidth is defined as following:

$$B_{\mathrm{RMS}} = \sqrt{\frac{\int_{-\infty}^{+\infty} f^2\,|S(f)|^2\,\mathrm{d}f}{\int_{-\infty}^{+\infty}|S(f)|^2\,\mathrm{d}f}} = \sqrt{\frac{\int_{-\infty}^{+\infty} f^2\mathcal{S}(f)\,\mathrm{d}f}{\int_{-\infty}^{+\infty}\mathcal{S}(f)\,\mathrm{d}f}}. \tag{2.20}$$

Here, the symbol f denotes frequency (in Hertz units) and $X(f)$ is the signal spectrum.[7] Note that $|S(f)|^2$ is the PSD of the signal, which is denoted by $\mathcal{S}(f)$. Often, the RMS bandwidth is approximated by the 3 dB bandwidth of the signal. By exploiting the properties of the Fourier transform [1], the RMS bandwidth can be computed also from the time domain representation of the signal $x(t)$:[8]

$$B_{\mathrm{RMS}} = \frac{1}{2\pi}\sqrt{\frac{\int_{-\infty}^{+\infty}\left(\frac{\mathrm{d}s(t)}{\mathrm{d}t}\right)^2\mathrm{d}t}{\int_{-\infty}^{+\infty}s(t)^2\mathrm{d}t}}. \tag{2.21}$$

Getting back to the delay CRLB (2.16); it is clearly observable that the key to the accuracy is the RMS bandwidth, since it influences the denominator of the formula in a linear manner. This means that minimal possible standard deviation is inversely proportional to the bandwidth; similarly, variance is inversely proportional to the squared bandwidth. Therefore, it is desirable to use signals with bandwidth as wide as possible.

It is possible to approach this recommendation from another point of view. In order to precisely measure the delay, a very sharp and narrow peak of the signal auto-correlation function (ACF) is needed. The ACF represents the mutual power between a signal and its time-shifted replica, therefore a narrow peak implies good delay resolution. This is indeed consistent with the requirement on wide bandwidth, because PSD of a signal is the Fourier transform of the ACF of the same signal, and consequently, wide PSD corresponds to narrow ACF, and vice versa.

Consider an example of rectangular-pulsed signal, whose is inversely proportional to the duration of the pulse. This straightforward shape is not convenient, since the side-lobes of the rectangular-pulse spectrum decay very slowly.[9] Also, in order to conserve the constant energy within the pulse, the peak power would have to be extremely high. Luckily, techniques of modulation and coding that combine long duration of the signal with wide bandwidth are available. These are often referred as pulse compression techniques in radar technology; in communication and navigation they are denoted as the

[7] Sometimes, the RMS bandwidth is defined in radians. In such case there is a factor $(2\pi)^2$ before the f^2 in the numerator of the equation (2.20).

[8] Should the RMS bandwidth be defined in radians, the $1/2\pi$ factor would be omitted.

[9] The PSD of a rectangular pulse of length T_p and energy $\mathcal{E}_p$ is a squared sinc function $sinc(2f)^2$.

spread spectrum techniques. One may take linear-frequency modulated (LFM) radar pulses, Barker codes, Gold codes or BOC modulation as examples.[10]

The other way of enhancing the ranging accuracy is increasing of the SNR (or $^{2\mathcal{E}}/_{N_0}$). The very straightforward way of boosting up the signal energy is increasing the transmitted power. This, however, cannot be always realized practically due to limited capabilities of the available hardware and emission limits. The energy can also be increased by extending the signal in the time domain; under assumption of constant signal power the energy grows linearly with the duration of the signal. Consequently, the minimum ranging standard deviation decreases with the square root of signal duration. It is also possible to slightly reduce the noise PSD by improving the receiver sensitivity and noise figure.

II. Two-way ranging

In a default scenario, there is no time synchronization between the transmitters and receivers of the ranging signals. Both, receiver and transmitter are able to provide timestamping of the signal transmission/reception, however, in their own, local timescale. For simplicity of the following explanation, let us neglect the Doppler effect and oscillator imperfections – all local oscillators are considered to have zero frequency offset. Under such conditions only a constant offset between the receiver and transmitter does exist.

Imagine that we would like to know the geometric distance between nodes A and B. Thus, a signal is sent from node A to node B, and the transmission and reception are timestamped. A superscript in square brackets will be used to denote the timescale in which a timestamp was taken. The reception timestamp in the receiving (B) node is

$$t_{\mathrm{RxB}}^{[\mathrm{B}]} = t_{\mathrm{TxA}}^{[\mathrm{A}]} + \frac{1}{c_0} r_{\mathrm{AB}} + \delta t^{[\mathrm{AB}]}, \tag{2.22}$$

where $t_{\mathrm{Tx}}^{[\mathrm{A}]}$ is the transmission timestamp in the A's timescale, r_{AB} denotes the geometric range between the nodes, and $\delta t^{[\mathrm{AB}]}$ is the offset of the A and B timescale (positive when A is delayed w.r.t. B). It is obvious that without prior knowledge of the offset term it is not possible to determine the range. Luckily, it is possible to send another signal from B to A, and get following equation:

$$t_{\mathrm{RxA}}^{[\mathrm{A}]} = t_{\mathrm{TxB}}^{[\mathrm{B}]} + \frac{1}{c_0} r_{\mathrm{BA}} + \delta t^{[\mathrm{BA}]}. \tag{2.23}$$

Since $r_{\mathrm{AB}} = r_{\mathrm{BA}}$ and $\delta t^{[\mathrm{AB}]} = -\delta t^{[\mathrm{BA}]}$, it is possible to solve the very simple set of linear equations of the two measurements and obtain the range and the timescale offset, which is a nuisance parameter in this case. It is straightforward that

$$r_{\mathrm{AB}} = \frac{c_0}{2}\left(\left(t_{\mathrm{RxA}}^{[\mathrm{A}]} - t_{\mathrm{TxA}}^{[\mathrm{A}]}\right) - \left(t_{\mathrm{RxB}}^{[\mathrm{B}]} - t_{\mathrm{TxB}}^{[\mathrm{B}]}\right)\right) \tag{2.24}$$

$$= \frac{c_0}{2}\left(\tau_A^{[\mathrm{A}]} - \tau_B^{[\mathrm{B}]}\right) \tag{2.25}$$

[10] While LFM and barker codes are used in radar field, Gold codes and BOC are exploited by GNSS.

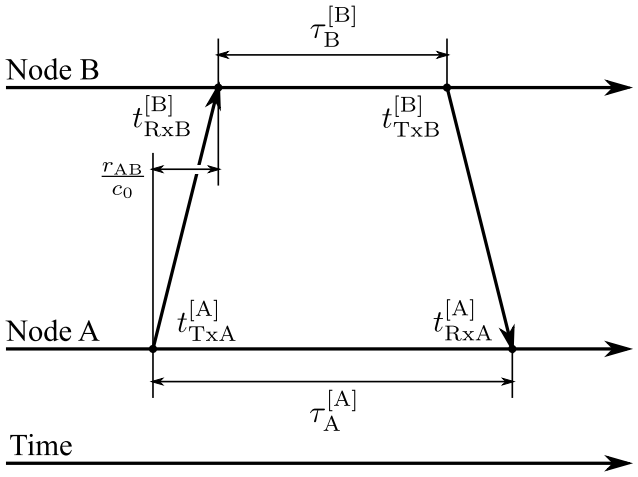


FIGURE 2.2.: Single-sided two-way ranging timing diagram.

This approach is called "single-sided two-way ranging" (SS-TWR), because it can be seen as a single request – response negotiation, where delays on the requester and responder side ($\tau_A^{[\mathrm{A}]}$ and $\tau_B^{[\mathrm{B}]}$) are measured. A typical timing diagram of the negotiation is in Figure 2.2;[11] note that the response delays and the propagation delay are not necessarily to scale.

The CRLB result (2.17) can be used to estimate the CRLB for the SS-TWR ranging approach. The transmission times are precisely known, and the two measurements of the time of reception are independent. Also, we can assume the channel to be reciprocal, and therefore we may only double the variance of the delay measurement. In order to convert the delay measurements to range in the SS-TWR ranging manner, the $c_0/2$ factor is to be used. Thus, the CRLB is

$$\sqrt{\mathrm{var}(\hat{r}_{\mathrm{AB}})} \geq c_0 \frac{\sqrt{2}}{2\sqrt{\frac{2\mathcal{E}}{N_0}} B_{\mathrm{RMS}}} = c_0 \frac{\sqrt{2}}{2\sqrt{\mathrm{SNR}} B_{\mathrm{RMS}}}. \tag{2.26}$$

Now, it is time to assume oscillator frequency offset, which is often referenced as clock drift. We may assume that temporal reading of the drifting $t^{[\mathrm{A}]}$ clock at true time t to be

$$t^{[\mathrm{A}]}(t) = t^{[\mathrm{A}]}(t_0) + t\Big(1 + \delta\nu^{[\mathrm{A}]}\Big), \tag{2.27}$$

where $\delta\nu^{[\mathrm{A}]}$ is the relative clock drift. Implicitly, the time differences in the one-way ranging equations are corrupted in the following way:

$$\breve{\tau}_A^{[\mathrm{A}]} = \tau_A + \delta\nu^{[\mathrm{A}]}\,\tau_A. \tag{2.28}$$

Note that the breve above the symbol denotes the corrupted value and that the same applies for node B. Since the clock offset is irrelevant for the delay measurement within

[11] It is worth mentioning that absolute time information (e.g. timestamp) is denoted by t, and time intervals (e.g. delays) are denoted by τ.

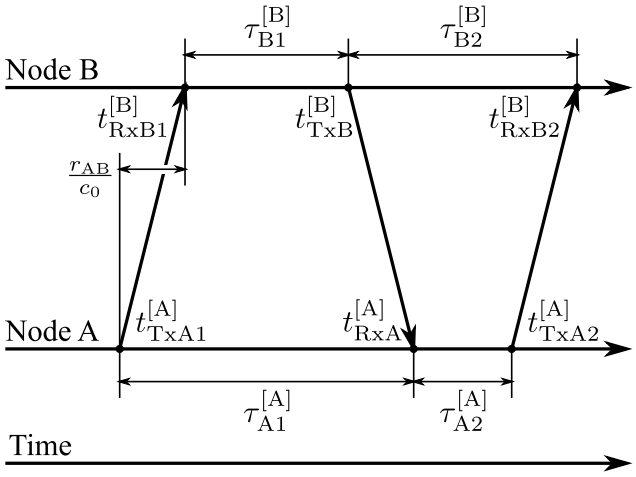


FIGURE 2.3.: Double-sided two-way ranging timing diagram.

a single node, the corresponding superfix is omitted. The error induced by the oscillator frequency offset is then

$$\breve{r}_{\mathrm{AB}} - r_{\mathrm{AB}} = \frac{c_0}{2}(\breve{\tau}^{[\mathrm{A}]}_A - \tau_A - \breve{\tau}^{[\mathrm{B}]}_B + \tau_B) \tag{2.29}$$

$$= \frac{c_0}{2}(\delta\nu^{[\mathrm{A}]}\,\tau_A - \delta\nu^{[\mathrm{B}]}\,\tau_B) \tag{2.30}$$

$$= \frac{c_0}{2}\left(\delta\nu^{[\mathrm{A}]}\frac{2r_{\mathrm{AB}}}{c_0} + \left(\delta\nu^{[\mathrm{A}]} - \delta\nu^{[\mathrm{B}]}\right)\tau_B\right). \tag{2.31}$$

It is worth to investigate when the error becomes zero; the assumption that both true range and responder delay (τ_B) are positive is understandable. Apart from the case when both frequency offsets are zero, the term (2.29) is null in the unlikely coincidence when equation

$$\frac{2r_{\mathrm{AB}}}{c_0} = \tau_B\frac{\delta\nu^{[\mathrm{A}]} - \delta\nu^{[\mathrm{B}]}}{\delta\nu^{[\mathrm{A}]}} \tag{2.32}$$

is satisfied. From this exercise it is clear that single-sided two-way ranging is prone to clock-drift-related errors.

Such errors may be compensated by adding the third signal that is transmitted from node A and received by node B, as is described on a timing diagram in Figure 2.3. This approach is called double-sided two-way ranging (DS-TWR) the very basic equation for the range estimation can be written in a couple of different ways:

$$r_{\mathrm{AB}} = \frac{c_0}{4}\left(\quad \tau^{[\mathrm{A}]}_{\mathrm{A1}} \quad - \quad \tau^{[\mathrm{B}]}_{\mathrm{B1}} \quad + \quad \tau^{[\mathrm{B}]}_{\mathrm{B2}} \quad - \quad \tau^{[\mathrm{A}]}_{\mathrm{A2}} \quad\right) \tag{2.33}$$

$$= \frac{c_0}{4}\left(\left(t^{[\mathrm{A}]}_{\mathrm{RxA}} - t^{[\mathrm{A}]}_{\mathrm{TxA1}}\right) - \left(t^{[\mathrm{B}]}_{\mathrm{TxB}} - t^{[\mathrm{B}]}_{\mathrm{RxB1}}\right) + \left(t^{[\mathrm{B}]}_{\mathrm{RxB2}} - t^{[\mathrm{B}]}_{\mathrm{TxB}}\right) - \left(t^{[\mathrm{A}]}_{\mathrm{TxA2}} - t^{[\mathrm{A}]}_{\mathrm{RxA}}\right)\right) \tag{2.34}$$

$$= \frac{c_0}{4}\left(t^{[\mathrm{B}]}_{\mathrm{RxB2}} + t^{[\mathrm{B}]}_{\mathrm{RxB1}} - t^{[\mathrm{A}]}_{\mathrm{TxA2}} - t^{[\mathrm{A}]}_{\mathrm{TxA1}} + 2\left(t^{[\mathrm{A}]}_{\mathrm{RxA}} - t^{[\mathrm{B}]}_{\mathrm{TxB}}\right)\right). \tag{2.35}$$

We may modify the CRLB result (2.17) in a similar manner, as we did for the single-sided ranging. In the DS-TWR, there are three independent measurements of the time of reception, which we can assume to have equal variance. Since one of the measurements is doubled in the equation, we have to multiply its variance by four; when the other two measurements are taken into account, we get a factor of six. The multiplying factor from delay to range representation is $\frac{c_0}{4}$ in this case, so the expression of CRLB of the DS-TWR in the standard deviation scale becomes

$$\sqrt{\mathrm{var}(\hat{r}_{\mathrm{AB}})} \geq c_0 \frac{\sqrt{6}}{4\sqrt{\frac{2\mathcal{E}}{N_0}} B_{\mathrm{RMS}}} = c_0 \frac{\sqrt{6}}{4\sqrt{\mathrm{SNR}} B_{\mathrm{RMS}}}. \tag{2.36}$$

Under the assumption of corruption of the measurements by the oscillator frequency offset (2.28) the corrupted version of range estimate becomes

$$\breve{r}_{\mathrm{AB}} = \frac{c_0}{4}\left(\breve{\tau}_{\mathrm{A1}}^{[\mathrm{A}]} - \breve{\tau}_{\mathrm{B1}}^{[\mathrm{B}]} + \breve{\tau}_{\mathrm{B2}}^{[\mathrm{B}]} - \breve{\tau}_{\mathrm{A2}}^{[\mathrm{A}]}\right) \tag{2.37}$$

$$= \frac{c_0}{4}\left(\left(1+\delta\nu^{[\mathrm{A}]}\right)\left(\tau_{\mathrm{A1}}^{[\mathrm{A}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\right) + \left(1+\delta\nu^{[\mathrm{B}]}\right)\left(\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{B1}}^{[\mathrm{B}]}\right)\right). \tag{2.38}$$

Taking into account that $c_0(\tau_{\mathrm{A1}} - \tau_{\mathrm{B1}}) = 2r_{\mathrm{AB}}$ and $c_0(\tau_{\mathrm{B2}} - \tau_{\mathrm{A2}}) = 2r_{\mathrm{AB}}$ it is straightforward to find and modify the equation for the error to the following form:

$$\breve{r}_{\mathrm{AB}} - r_{\mathrm{AB}} = \left(\delta\nu^{[\mathrm{A}]} + \delta\nu^{[\mathrm{B}]}\right)\frac{r_{\mathrm{AB}}}{2} + \frac{c_0}{4}\left(\delta\nu^{[\mathrm{A}]} - \delta\nu^{[\mathrm{B}]}\right)(\tau_{\mathrm{B1}} - \tau_{\mathrm{A2}}). \tag{2.39}$$

The former part of the expression in the large bracket says that part of the error is proportional to the true value of the measured range. Since the frequency offsets are likely to be very small (in order of 10^{-5} or smaller even for low-cost oscillators), this contribution is not significant for distances up to a few kilometers. The latter part says that a part of the error is proportional to the difference of the reply times of the nodes. Since the factor is quarter of the speed of light, the error may become significant even for a difference of several tens of microseconds. In order to minimize the latter source of error, the symmetric double-sided two-way ranging (SDS-TWR) is often applied.

Moreover, there is a patented [3, 11] formula that relaxes the requirement on the equality of the reply delays on both nodes. For the sake of completeness we will derive the corresponding formula here, too.

It is convenient to consider ideal devices with no oscillator frequency offset for the purpose of the formula derivation, and to find the impact of the offsets later. Clearly, the double-sided two-way ranging can be broken into two single-sided ones as follows:

$$r_{\mathrm{AB}} = \frac{c_0}{2}\left(\tau_{\mathrm{A1}}^{[\mathrm{A}]} - \tau_{\mathrm{B1}}^{[\mathrm{B}]}\right) \tag{2.40}$$

$$r_{\mathrm{AB}} = \frac{c_0}{2}\left(\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\right). \tag{2.41}$$

Then the two equations are reorganized so that the delays measured on the requester are on the left-hand side of the equations and the other terms on the right-hand side:

$$\tau_{\mathrm{A1}}^{[\mathrm{A}]} = \frac{2\,r_{\mathrm{AB}}}{c_0} + \tau_{\mathrm{B1}}^{[\mathrm{B}]} \tag{2.42}$$

$$\tau_{\mathrm{B2}}^{[\mathrm{B}]} = \frac{2\,r_{\mathrm{AB}}}{c_0} + \tau_{\mathrm{A2}}^{[\mathrm{A}]}. \tag{2.43}$$

Followingly, the product of the requester delays is computed and the result is rewritten as:

$$\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} = \frac{4\,r_{\mathrm{AB}}\,r_{\mathrm{AB}}}{c_0^2} + \frac{2\,r_{\mathrm{AB}}}{c_0}\tau_{\mathrm{A2}}^{[\mathrm{A}]} + \frac{2\,r_{\mathrm{AB}}}{c_0}\tau_{\mathrm{B1}}^{[\mathrm{B}]} + \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]} \tag{2.44}$$

$$\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]} = \frac{2\,r_{\mathrm{AB}}}{c_0}\left(\frac{2\,r_{\mathrm{AB}}}{c_0} + \tau_{\mathrm{A2}}^{[\mathrm{A}]} + \tau_{\mathrm{B1}}^{[\mathrm{B}]}\right). \tag{2.45}$$

Then, one of the terms (2.42), (2.43) is substituted into (2.45), resulting in

$$\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]} = \frac{2\,r_{\mathrm{AB}}}{c_0}\left(\tau_{\mathrm{A1}}^{[\mathrm{A}]} + \tau_{\mathrm{A2}}^{[\mathrm{A}]}\right) \tag{2.46}$$

$$\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]} = \frac{2\,r_{\mathrm{AB}}}{c_0}\left(\tau_{\mathrm{B2}}^{[\mathrm{B}]} + \tau_{\mathrm{B1}}^{[\mathrm{B}]}\right) \tag{2.47}$$

By means of a simple modification and combination of the equations above we may obtain three variants of the range estimate from the same measurements.

$$r_{\mathrm{AB},2} = \frac{c_0}{2}\,\frac{\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]}}{\tau_{\mathrm{A1}}^{[\mathrm{A}]} + \tau_{\mathrm{A2}}^{[\mathrm{A}]}} \tag{2.48}$$

$$r_{\mathrm{AB},1} = \frac{c_0}{2}\,\frac{\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]}}{\tau_{\mathrm{B2}}^{[\mathrm{B}]} + \tau_{\mathrm{B1}}^{[\mathrm{B}]}} \tag{2.49}$$

$$r_{\mathrm{AB},3} = c_0\,\frac{\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]}}{\tau_{\mathrm{A1}}^{[\mathrm{A}]} + \tau_{\mathrm{A2}}^{[\mathrm{A}]} + \tau_{\mathrm{B2}}^{[\mathrm{B}]} + \tau_{\mathrm{B1}}^{[\mathrm{B}]}} \tag{2.50}$$

Now, we can find the impact of oscillator frequency offset on the range estimates obtained using these equations. It is straightforward to substitute the corrupted versions of the respective delay measurements, according to model from (2.28), and modify the expressions to the following form:

$$\breve{r}_{\mathrm{AB},2} = \frac{\left(1+\delta\nu^{[\mathrm{A}]}\right)\left(1+\delta\nu^{[\mathrm{B}]}\right)}{\left(1+\delta\nu^{[\mathrm{B}]}\right)}\,\frac{c_0}{2}\,\frac{\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]}}{\tau_{\mathrm{A1}}^{[\mathrm{A}]} + \tau_{\mathrm{A2}}^{[\mathrm{A}]}} = \left(1+\delta\nu^{[\mathrm{A}]}\right) r_{\mathrm{AB}} \tag{2.51}$$

$$\breve{r}_{\mathrm{AB},1} = \frac{\left(1+\delta\nu^{[\mathrm{A}]}\right)\left(1+\delta\nu^{[\mathrm{B}]}\right)}{\left(1+\delta\nu^{[\mathrm{B}]}\right)}\,\frac{c_0}{2}\,\frac{\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]}}{\tau_{\mathrm{B2}}^{[\mathrm{B}]} + \tau_{\mathrm{B1}}^{[\mathrm{B}]}} = \left(1+\delta\nu^{[\mathrm{B}]}\right) r_{\mathrm{AB}} \tag{2.52}$$

$$\begin{aligned}
\breve{r}_{\mathrm{AB},3} &= c_0\,\frac{\left(1+\delta\nu^{[\mathrm{A}]}\right)\left(1+\delta\nu^{[\mathrm{B}]}\right)\left(\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]}\right)}{\left(1+\delta\nu^{[\mathrm{A}]}\right)\left(\tau_{\mathrm{A1}}^{[\mathrm{A}]} + \tau_{\mathrm{A2}}^{[\mathrm{A}]}\right) + \left(1+\delta\nu^{[\mathrm{B}]}\right)\left(\tau_{\mathrm{B1}}^{[\mathrm{B}]} + \tau_{\mathrm{B2}}^{[\mathrm{B}]}\right)} \\
&= \frac{\left(1+\delta\nu^{[\mathrm{A}]}\right)\left(1+\delta\nu^{[\mathrm{B}]}\right)}{\left(1+\delta\nu^{[\mathrm{A}]}\right) + \left(1+\delta\nu^{[\mathrm{B}]}\right)}\,2\,\frac{c_0}{2}\,\frac{\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]}}{2\left(\tau_{\mathrm{B2}}^{[\mathrm{B}]} + \tau_{\mathrm{B1}}^{[\mathrm{B}]}\right)} \\
&= \frac{\left(1+\delta\nu^{[\mathrm{A}]}\right)\left(1+\delta\nu^{[\mathrm{B}]}\right)}{2+\delta\nu^{[\mathrm{A}]}+\delta\nu^{[\mathrm{B}]}}\,2\,r_{\mathrm{AB}}.
\end{aligned} \tag{2.53}$$

In order to obtain the error value, the true value has to be subtracted from the corrupted value. The trio of the error equations yields

$$\breve{r}_{\mathrm{AB},2} - r_{\mathrm{AB}} = \delta\nu^{[\mathrm{A}]} r_{\mathrm{AB}} \tag{2.54}$$

$$\breve{r}_{\mathrm{AB},1} - r_{\mathrm{AB}} = \delta\nu^{[\mathrm{B}]} r_{\mathrm{AB}} \tag{2.55}$$

$$\breve{r}_{\mathrm{AB},3} - r_{\mathrm{AB}} = \frac{2\,\delta\nu^{[\mathrm{A}]}\delta\nu^{[\mathrm{B}]} + \delta\nu^{[\mathrm{A}]} + \delta\nu^{[\mathrm{B}]}}{2 + \delta\nu^{[\mathrm{A}]} + \delta\nu^{[\mathrm{B}]}}\, r_{\mathrm{AB}} \approx \frac{\delta\nu^{[\mathrm{A}]} + \delta\nu^{[\mathrm{B}]}}{2}\, r_{\mathrm{AB}} \tag{2.56}$$

The approximation in (2.56) is valid for typical conditions, since both relative frequency offsets $|\delta\nu^{[\mathrm{A}]}|$, $|\delta\nu^{[\mathrm{B}]}| \ll 1$, i.e. even for low-cost crystal oscillators it is in the order of 10^{-5} (tens of ppm) or lower.

Obviously, in all three cases the error is proportional to the true range between the two nodes of interest r_{AB}, and more importantly, it does not depend on the true delay values (denoted by τ) at all. The former two range estimate equations, i.e (2.48) and (2.49), have the advantage that the errors are prevailed by one of the oscillators only; the frequency offset of the other one does not manifest in the ranging error, see results (2.54) and (2.55). This behavior might be useful when the oscillator performance in one of the nodes is superior to the other. The error in the latter case, i.e according to (2.48), is determined by the average of the two frequency offsets, which is coincidentally the same value that is achieved by SDS-TWR provided the equality condition on response delays is satisfied, see comparison of (2.56) and (2.39).

The ADS-TWR algorithm is clearly nonlinear; consequently, it is not inherently unbiased and its variance cannot be evaluated in a straightforward manner. The detailed analysis of the bias and variance of the ADS-TWR estimators (2.48) to (2.50), which is supported by Monte-Carlo simulations, is provided in Appendix A. It is shown that the estimators are unbiased or negligibly biased under realistic conditions.[12] The variance of the ADS-TWR is very similar to the variance of the SDS-TWR. Hence, the algorithm can be considered safe to be used instead of the SDS-TWR. No accuracy penalty is brought by the insensitivity to clock drift.

From the geometrical standpoint, we do not have to care about the method for obtaining the range measurements. The equation for the range between the user equipment and a radio node with index i is the same as in the case of range estimates that are based on signal power from page 11:

$$r_{\mathrm{u}i} = \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_i\| \,. \tag{2.11}$$

Again, we refer to Chapter 4 for the solution of the set of range equations and related discussion.

III. Time of Arrival

Consider a case when the infrastructure of the localization system shares the same timescale. To achieve this, the infrastructure has to be synchronized accurately.

Without the loss of generality we may adopt a scenario where the infrastructure nodes of the radio positioning network transmit signals at a certain time and that

[12] The bias of ADS-TWR is a few orders of magnitude lower than its variance.

the located node receives those signals. The same mathematical derivation would take place if the situation was reciprocal, i.e. the localized node was transmitting and infrastructure nodes were receiving. The common system timescale is denoted by [S] superfix. The timescale of the localized node is denoted by [U] superfix, since the localized node is usually the user equipment. Therefore, we can write for the user-node reception time of the signal i-th infrastructure node

$$t_{\mathrm{Rx,i}}^{[\mathrm{U}]} = t_{\mathrm{Tx,i}}^{[\mathrm{S}]} + \frac{1}{c_0} r_{\mathrm{u,i}} + \delta t^{[\mathrm{SU}]}, \tag{2.57}$$

where $t_{\mathrm{Tx}}^{[\mathrm{S}]}$ is the transmission time in the system timescale, and $r_{\mathrm{u,i}}$ is the geometric range between the i-th infrastructure node and the user node. The timescale offset term $\delta t^{[\mathrm{SU}]}$ is also present in the equation.

Usually, but not always, the difference of the reception and transmission time is taken as the time of arrival (ToA) measurement. The value is scaled by the speed of light and called pseudorange (especially in the GNSS field).

$$\rho_{\mathrm{i}} = c_0 \left(t_{\mathrm{Rx,i}}^{[\mathrm{U}]} - t_{\mathrm{Tx,i}}^{[\mathrm{S}]} \right) = c_0 \left(\frac{1}{c_0} r_{\mathrm{u,i}} + \delta t^{[\mathrm{SU}]} \right), \tag{2.58}$$

The pseudorange is equal to the true range corrupted by the scaled unknown timescale offset. However, the infrastructure nodes are synchronized and therefore all the pseudorange measurements are corrupted by the same value of the timescale offset. In the GNSS field, for instance, the offset is usually scaled by the speed of light and therefore expressed in the terms of length instead of duration. Such scaled value is commonly called bias and denoted by the letter b. Then, the pseudorange can be expressed as

$$\rho_{\mathrm{i}} = \| \boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{i}} \| + b, \tag{2.59}$$

where $\boldsymbol{r}_{\mathrm{u}}$ and $\boldsymbol{r}_{\mathrm{i}}$ are the position vectors of the localized node and the i-th infrastructure node, respectively. Apparently, the variance of the pseudorange measurement is equal to the variance of the time of reception, considering that time of transmission is known precisely.

Provided that the infrastructure position vectors are known, there are only a few unknown parameters: two or three coordinates of the localized node position vector and the bias. The same variables are present independently on the index of the infrastructure node. Consequently, a set of at least three or four pseudorange equations, depending on the position vector dimensionality, is required for position and bias determination. The proper investigation of the characteristics and methods of solving such sets of equations are provided in Chapter 4.

Since we already know the CRLB for a plain delay measurement, derivation of CRLB for ToA is apparent. We have assumed that time of transmission is perfectly known and equal for all the signals, therefore, only the time of reception is measured. The communication is only from the infrastructure to the user, so it becomes evident that the CRLB of ToA follows exactly the equation (2.17).

It is also convenient to derive the equation for the rate of change of the pseudorange:

$$\frac{\mathrm{d}\rho_\mathrm{i}}{\mathrm{d}t} = \frac{\|\boldsymbol{r}_\mathrm{u} - \boldsymbol{r}_\mathrm{i}\|}{\mathrm{d}t} + \frac{\mathrm{d}b}{\mathrm{d}t} \tag{2.60}$$

$$\frac{\mathrm{d}\rho_\mathrm{i}}{\mathrm{d}t} = \frac{(\boldsymbol{r}_\mathrm{u} - \boldsymbol{r}_\mathrm{i}) \cdot \left(\frac{\boldsymbol{r}_\mathrm{u}}{\mathrm{d}t} - \frac{\boldsymbol{r}_\mathrm{i}}{\mathrm{d}t}\right)}{\|\boldsymbol{r}_\mathrm{u} - \boldsymbol{r}_\mathrm{i}\|} + \frac{\mathrm{d}b}{\mathrm{d}t}, \tag{2.61}$$

When the simplified notation with the time derivative is expressed by the dot accent, the equation yields

$$\dot{\rho}_\mathrm{i} = \mathbb{1}_{\mathrm{u,i}} \cdot (\dot{\boldsymbol{r}}_\mathrm{u} - \dot{\boldsymbol{r}}_\mathrm{i}) + \dot{b} \tag{2.62}$$

$$\dot{\rho}_\mathrm{i} = \mathbb{1}_{\mathrm{u,i}} \cdot (\boldsymbol{v}_\mathrm{u} - \boldsymbol{v}_\mathrm{i}) + \dot{b}, \tag{2.63}$$

where v denotes the velocity vector and $\mathbb{1}_{\mathrm{u,i}}$ is the unitary direction vector pointing from i-th infrastructure node towards the localized (user) node. Vectors v denote the velocity vectors. From the latter expression it is obvious that pseudorange rate is equal to the radial component of the relative velocity of the nodes, of course corrupted by the rate of bias, i.e. clock frequency mismatch. It is worth noting that radial velocity component is proportional to the Doppler frequency shift of the signal.

IV. Time Difference of Arrival

As well as the ToA approach, the Time Difference of Arrival (TDoA) method of positioning requires synchronized infrastructure. The localized equipment does not share the timescale with the infrastructure; its clock suffers from bias w.r.t. the system timescale.

Nonetheless, the clock bias term that corrupts the range measurements (the pseudorange measurement is taken) is constant for all measurements. Therefore, consider the difference of two pseudorange measurements, denoted by $d_{\mathrm{i,j}}$:

$$\begin{aligned} d_{\mathrm{i,j}} &= \rho_\mathrm{i} - \rho_\mathrm{j} \\ &= \|\boldsymbol{r}_\mathrm{u} - \boldsymbol{r}_\mathrm{i}\| + b - (\|\boldsymbol{r}_\mathrm{u} - \boldsymbol{r}_\mathrm{j}\| + b) \\ &= \|\boldsymbol{r}_\mathrm{u} - \boldsymbol{r}_\mathrm{i}\| - \|\boldsymbol{r}_\mathrm{u} - \boldsymbol{r}_\mathrm{j}\| \,. \end{aligned} \tag{2.64}$$

The bias term cancels out and the raw difference of geometric ranges between the localized equipment and the infrastructure nodes remains, as anticipated. The difference of the pseudorange has to be investigated also from the time of reception and transmission measurements in the respective timescales:

$$\begin{aligned} d_{\mathrm{i,j}} &= c_0 \left[\left(t^{[\mathrm{U}]}_{\mathrm{Rx,i}} - t^{[\mathrm{S}]}_{\mathrm{Tx,i}} \right) - \left(t^{[\mathrm{U}]}_{\mathrm{Rx,j}} - t^{[\mathrm{S}]}_{\mathrm{Tx,j}} \right) \right] \\ &= c_0 \left[\left(t^{[\mathrm{U}]}_{\mathrm{Rx,i}} - t^{[\mathrm{U}]}_{\mathrm{Rx,j}} \right) - \left(t^{[\mathrm{S}]}_{\mathrm{Tx,i}} - t^{[\mathrm{S}]}_{\mathrm{Tx,j}} \right) \right] . \end{aligned} \tag{2.65}$$

From the expression on the second line it is apparent that only the time difference of signal reception $\left(t^{[\mathrm{U}]}_{\mathrm{Rx,i}} - t^{[\mathrm{U}]}_{\mathrm{Rx,j}} \right)$ has to be measured, therefore the offset of the timescale does not matter at all. The term $\left(t^{[\mathrm{S}]}_{\mathrm{Tx,i}} - t^{[\mathrm{S}]}_{\mathrm{Tx,j}} \right)$ is the difference in time of transmission in the system timescale. This has to be zero (when the time of transmission is

synchronized) or set to a predefined value, which is known to the localized equipment (for instance in time-multiplexed systems).

In order to estimate the TDoA CRLB, we will use the previous results. It does not matter whether we consider the TDoA measurement as a difference of pseudoranges or times of reception. In both cases, the difference of the two stochastically independent values is taken as a result. Therefore,

$$\operatorname{var} d_{\mathrm{i,j}} = \operatorname{var} \rho_{\mathrm{i}} + \operatorname{var} \rho_{\mathrm{j}} = \operatorname{var} t^{[\mathrm{U}]}_{\mathrm{Rx,i}} + \operatorname{var} t^{[\mathrm{U}]}_{\mathrm{Rx,j}}. \tag{2.66}$$

If all the pseudoranges (or times of arrival) are assumed to have equal CRLB, the expression can be simplified to

$$\operatorname{var} d = 2 \operatorname{var} \rho \qquad = 2 \operatorname{var} t^{[\mathrm{U}]}_{\mathrm{Rx}} \tag{2.67}$$

$$\sqrt{\operatorname{var} d} = \sqrt{2}\sqrt{\operatorname{var} \rho} = \sqrt{2}\sqrt{\operatorname{var} t^{[\mathrm{U}]}_{\mathrm{Rx}}}. \tag{2.68}$$

Note that in the simplified case the minimal variance (the CRLB) is doubled, and the corresponding standard deviation is $\sqrt{2}$ times the value from eq. (2.17).

Practical examples of TDoA positioning can be found in Appendix B. In particular, Section B.1 describes an opportunistic localization system that exploits synchronized, stationary DVB-T transmitters and measures TDoA of signals from those transmitters.

It is worth noting that it is possible to operate a TDoA system in a reverse manner, i.e. the localized equipment transmits a message that is captured by multiple synchronized receivers. In such case the time of transmission is inherently constant, since a single message is received by multiple receivers. Consequently, the TDoA measurement scaled by speed of light can be for this particular case written as

$$\begin{aligned} d_{\mathrm{i,j}} &= c_0 \left(t^{[\mathrm{S}]}_{\mathrm{Rx,i}} - t^{[\mathrm{S}]}_{\mathrm{Rx,j}} \right) - \left(t^{[\mathrm{U}]}_{\mathrm{Tx}} - t^{[\mathrm{U}]}_{\mathrm{Tx}} \right) \\ &= c_0 \left(t^{[\mathrm{S}]}_{\mathrm{Rx,i}} - t^{[\mathrm{S}]}_{\mathrm{Rx,j}} \right). \end{aligned} \tag{2.69}$$

An example that exploits such reversed principle is described in Section B.2 of Appendix B.

Now we will find the time derivative of the TDoA measurement, as we did for the ToA in the previous section.

$$\frac{\mathrm{d}d_{\mathrm{i,j}}}{\mathrm{d}t} = \mathbb{1}_{\mathrm{u,i}} \cdot (\boldsymbol{v}_{\mathrm{u}} - \boldsymbol{v}_{\mathrm{i}}) - \mathbb{1}_{\mathrm{u,j}} \cdot (\boldsymbol{v}_{\mathrm{u}} - \boldsymbol{v}_{\mathrm{j}}) \tag{2.70}$$

It is worth comparing this result with the result for the ToA case. Due to the differentiation the clock frequency mismatch term is not present in the equation. This result was expectable, since the differentiation of the measurements is a linear operation. Provided that the infrastructure is stationary, i.e. $\boldsymbol{v}_{\mathrm{i}} = 0$ and $\boldsymbol{v}_{\mathrm{j}} = 0$, we may write a simplified equation

$$\frac{\mathrm{d}d_{\mathrm{i,j}}}{\mathrm{d}t} = (\mathbb{1}_{\mathrm{u,i}} - \mathbb{1}_{\mathrm{u,j}}) \cdot \boldsymbol{v}_{\mathrm{u}} \tag{2.71}$$

Using the mathematical description of the Doppler effect[13] it is easy to show that the time-derivative of the TDoA measurement is proportional to the difference of the

[13]The Doppler effect is described in detail within Section 2.1.3.

Doppler shifts of the received signals:

$$\frac{\mathrm{d}d_{\mathrm{i,j}}}{\mathrm{d}t} = \frac{c_0}{f_{\mathrm{nom}}} \left(f_{\mathrm{D,i}} - f_{\mathrm{D,j}} \right), \tag{2.72}$$

where f_{nom} is the nominal transmission frequency. Understandably, this kind of measurements is often called Frequency Difference of Arrival (FDoA).

2.1.3. Doppler-effect based positioning

The Doppler effect (non-relativistic) is a wave frequency shift due to the relative motion of the wave transmitter and receiver. The frequency on the receiver can be described in a single dimension by equation

$$f_{\mathrm{Rx}} = \frac{c_0 + v_{\mathrm{Rx}}}{c_0 + v_{\mathrm{Tx}}} f_{\mathrm{Tx}} \tag{2.73}$$

Assuming that the velocities of receiver and transmitter are substantially smaller than the wave propagation velocity c_0, the equation may be approximated by a simpler form to

$$f_{\mathrm{Rx}} = \left(1 + \frac{v_{\mathrm{Rx}} - v_{\mathrm{Tx}}}{c_0 + v_{\mathrm{Tx}}} \right) f_{\mathrm{Tx}} \approx \left(1 + \frac{v_{\mathrm{Rx}} - v_{\mathrm{Tx}}}{c_0} \right) f_{\mathrm{Tx}} \tag{2.74}$$

This form of the description is quite convenient, because it renders the Doppler shift additive to the transmitted frequency.

The situation in multiple dimensions is rather similar, however, it is necessary to extract the radial component from the relative velocity vector. This can be done rather easily by performing a dot product with the unitary direction vector from transmitter to receiver $\mathbb{1}_{\mathrm{Rx,Tx}}$.

$$f_{\mathrm{Rx}} \approx \left(1 + (\boldsymbol{v}_{\mathrm{Rx}} - \boldsymbol{v}_{\mathrm{Tx}}) \cdot \mathbb{1}_{\mathrm{Rx,Tx}} \right) \frac{f_{\mathrm{Tx}}}{c_0} \tag{2.75}$$

For the sake of simplicity we will denote the Doppler frequency shift $f_{\mathrm{Rx}} - f_{\mathrm{Rx}}$ as f_{D}. Consider that a receiver is capable of integrating the Doppler shift f_{D}, i.e. accumulating the Doppler phase φ_{D}:

$$\begin{aligned} \varphi_{\mathrm{D}}(t_{\mathrm{i}}, t_{\mathrm{i+1}}) &= 2\pi \int_{t_{\mathrm{i}}}^{\mathrm{i+1}} f_{\mathrm{D}}(t) \mathrm{d}t \\ &= 2\pi \frac{f_{\mathrm{Tx}}}{c_0} \int_{t_{\mathrm{i}}}^{\mathrm{i+1}} (\boldsymbol{v}_{\mathrm{Rx}} - \boldsymbol{v}_{\mathrm{Tx}}) \cdot \mathbb{1}_{\mathrm{Rx,Tx}} \, \mathrm{d}t \end{aligned} \tag{2.76}$$

Clearly, the antiderivative (primitive function) of the integrated function is the difference of the range between the transmitter and receiver. One of justifications of this statement can be seen in equations (2.60) to (2.63), where the derivative of the range is found. Therefore the integral, according to the fundamental theorem of calculus, is equal to the difference of the ranges at the temporal boundaries of the observation, i.e. ranges at the times t_{i} and $t_{\mathrm{i+1}}$:

$$\varphi_{\mathrm{D}}(t_{\mathrm{i}}, t_{\mathrm{i+1}}) = 2\pi \frac{f_{\mathrm{Tx}}}{c_0} \Big(\left\| \boldsymbol{r}_{\mathrm{Rx}}(t_{\mathrm{i+1}}) - \boldsymbol{r}_{\mathrm{Tx}}(t_{\mathrm{i+1}}) \right\| - \left\| \boldsymbol{r}_{\mathrm{Rx}}(t_{\mathrm{i}}) - \boldsymbol{r}_{\mathrm{Tx}}(t_{\mathrm{i}}) \right\| \Big). \tag{2.77}$$

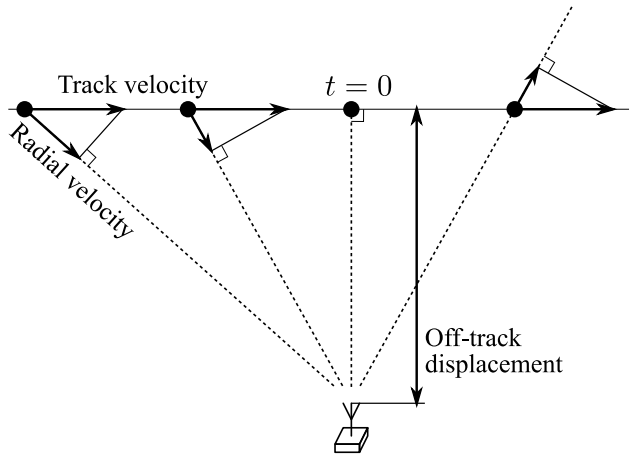


FIGURE 2.4.: Geometry of Doppler effect in two dimensions.

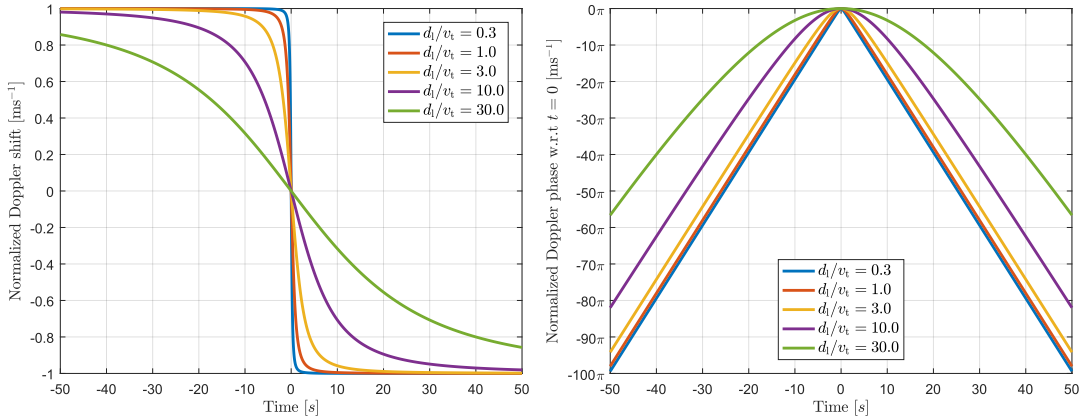
If the position of the transmitter is known at any time, and if the receiver is stationary,[14]only two or three integrated Doppler phase measurements φ_{D} are required, depending on whether 2D or 3D localization is performed. The advantage of such approach is that the offset of the receiver and transmitter timescales does not affect the measurements at all. Should the measurement be corrupted by an unknown oscillator frequency offset, there will be additional nuisance parameter to be estimated δf_{Rx} in the equation for the accumulated Doppler phase. Assuming the frequency offset constant through all measurement epochs the equation yields:

$$\varphi_{\mathrm{D}}(t_{\mathrm{i}}, t_{\mathrm{i+1}}) = 2\pi \frac{f_{\mathrm{Tx}}}{c_0} \Big(\big\| \boldsymbol{r}_{\mathrm{Rx}}(t_{\mathrm{i+1}}) - \boldsymbol{r}_{\mathrm{Tx}}(t_{\mathrm{i+1}}) \big\| - \big\| \boldsymbol{r}_{\mathrm{Rx}}(t_{\mathrm{i}}) - \boldsymbol{r}_{\mathrm{Tx}}(t_{\mathrm{i}}) \big\| + \delta f_{\mathrm{Rx}} (t_{\mathrm{i+1}} - t_{\mathrm{i}}) \Big). \tag{2.78}$$

Let us analyze the two-dimensional case which is captured in Figure 2.4. We will assume a transmitter moving on a straight track with a constant track velocity v_t, and a stationary receiver that is displaced laterally from the axis of movement; in the following equations the lateral displacement is denoted by symbol d_{l}. The radial velocity, which is the determining factor for the resulting Doppler shift, is obtained by a simple projection of the track velocity onto the axis connecting transmitter and receiver. Obviously, the situation is axially symmetric (w.r.t. axis of movement), and therefore inherently ambiguous.

In Figure 2.5a a parametric plot of the normalized Doppler shift is plotted for this scenario. The Doppler shift value is divided by the $(v_{\mathrm{t}}\, f_{\mathrm{Tx}}/c_0)$ factor, therefore the normalized value equal to one corresponds to the maximum achievable Doppler shift with the chosen track velocity. That would correspond to a case where receiver is on the track and the radial velocity is the same as track velocity. The horizontal axis displays a time relative to the moment of the transmitter passing by the receiver (i.e. the moment with the zero Doppler shift). The parameter of the curves was chosen

[14]Actually, the requirement on stationary receiver is not strict, only the relation between the user position vectors at the times t_{i}, $t_{\mathrm{i+1}}$, $t_{\mathrm{i+2}}$, ... has to be known.



(A) Frequency shift. (B) Accumulated phase w.r.t. passing point.

FIGURE 2.5.: Parametric plots of Doppler values for linear transmitter trajectory and stationary receiver.

to be the ratio of the track velocity and lateral displacement. The expression for the Doppler shift in the analyzed scenario was manipulated in the following way for the purpose of the plot:

$$f_\mathrm{D}(t) = \frac{f_\mathrm{Tx}}{c_0}\, v_\mathrm{t} \frac{t\, v_\mathrm{t}}{\sqrt{(t\, v_\mathrm{t})^2 + {d_\mathrm{l}}^2}} \tag{2.79}$$

$$\frac{f_\mathrm{D}(t)\, c_0}{v_\mathrm{t}\, f_\mathrm{Tx}} = \frac{t}{\sqrt{t^2 + \left(\frac{d_\mathrm{l}}{v_\mathrm{t}}\right)^2}} \tag{2.80}$$

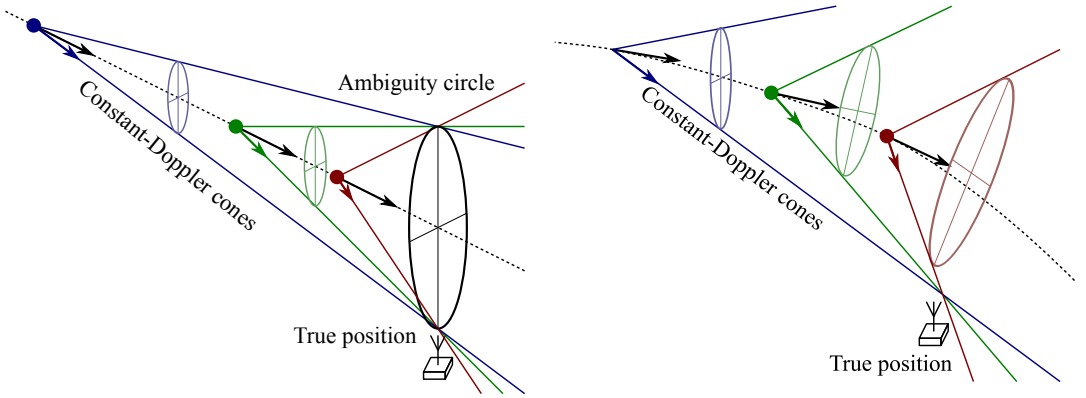
Similarly an equation for the expression accumulated phase can be evaluated with respect to the passing point, i.e. $t = 0$, and normalized in the same way

$$\varphi_\mathrm{D}(0,\, t) = 2\pi \left(d_\mathrm{l} - \sqrt{(t\, v_\mathrm{t})^2 + {d_\mathrm{l}}^2} \right) \tag{2.81}$$

$$\frac{\varphi_\mathrm{D}(0,\, t)\, c_0}{v_\mathrm{t}\, f_\mathrm{Tx}} = 2\pi \left(\frac{d_\mathrm{l}}{v_\mathrm{t}} - \sqrt{t^2 + \left(\frac{d_\mathrm{l}}{v_\mathrm{t}}\right)^2} \right) \tag{2.82}$$

It should be noted that the constant oscillator frequency offset would add a constant bias to the frequency measurement and a linear function to the phase result.

Due to the symmetry of the scenario the receiver position cannot be determined unambiguously. In the plane, the constant-Doppler shape is formed by two axially-symmetric semi-axes with a common origin. One of the possible ambiguity resolution approaches may be making the transmitter trajectory curved; consequently, the Doppler shift profiles on the left and right sides of the trajectory differ. The axial symmetry of each constant-Doppler semi-axis pair still holds, yet the axis of symmetry changes due to the curvature of the trajectory. In such scenario of 2-D positioning a single transmitter observed for a suitable time interval should provide enough information that is needed for unambiguous positioning.



(A) Linear trajectory (B) Curved trajectory

FIGURE 2.6.: Visualization of the constant-Doppler cones.

When generalized into three dimensions, the situation is still axially symmetric, therefore the surface with a constant Doppler shift is an infinite cone – the surface originates from rotation of the transmitter-to-receiver semi-axis around the track axis. Should the trajectory be linear, the intersection of the consecutive constant Doppler cones would be a circle, see Figure 2.6a. As well as in the two-dimensional case, the ambiguity can be resolved by means of curving the transmitter trajectory. However, in order to fully resolve the ambiguity, the trajectory must not lie in a plane; otherwise the symmetry of the scenario would be degraded from axial to planar, and the ambiguity space would only be reduced from a circle to two points. The case where the cones intersect unambiguously is depicted in Figure 2.6b.

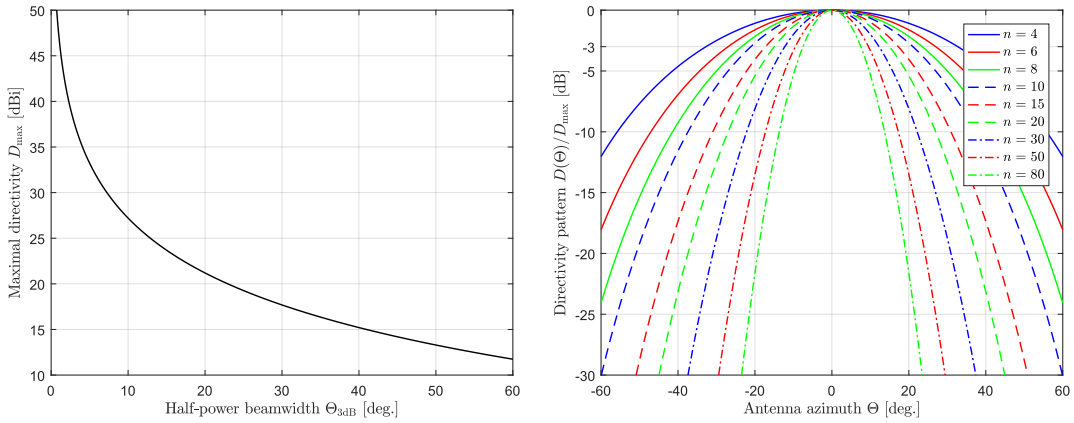
The Doppler effect may be also exploited in a ToA or a TDoA system to estimate the velocity vector of a localized equipment. Typically, the delay measurement (TDoAs or pseudoranges are used as a primary source for position information, the Doppler shift information is required for joint position and velocity information. See equations (2.63) and (2.71) for Doppler frequency shift in the ToA and TDoA scenarios.

2.1.4. Angle of arrival

This section elaborates on direction of the signal arrival measurement, typically denoted as angle of arrival (AoA). Firstly, the methods that exploit the directional characteristics of the antennae will be examined; those are herein called signal power approaches, since it is the signal characteristic under investigation. Secondly, the methods that exploit antenna arrays and phase (or delay) information from the array elements will be described.

I. Signal power approaches

The intuitive way of determining AoA is steering a directional antenna towards the signal origin, i.e. finding the direction with the highest received signal power. This is quite inconvenient, since the main lobe maximum is typically rather flat and therefore the angular accuracy is compromised. This can be illustrated by approximate relation



(A) Directivity vs. half-power beamw. (B) Main lobe central part

FIGURE 2.7.: Approximation of antenna patterns

of the maximal antenna directivity [15] $D_{\max}$ and the half-power beamwidth[16] $\Theta_{3\mathrm{dB}}$ that is available in [12].

$$D_{\max} \approx 2 \left(1 - \cos \frac{\Theta_{3\mathrm{dB}}}{2}\right)^{-1} \approx \frac{16}{(\Theta_{3\mathrm{dB}})^2} \tag{2.83}$$

Note that the latter expression requires the beamwidth expressed in radians. The relation is visualized in Fig. 2.7a for beamwidths up to 60°. Obviously, more than 25 dB directionality is needed in order to achieve beamwidths of a few degrees. However, such requirement implies electrically spacious (i.e. w.r.t signal wavelength) antenna structures, consequently, the antennae may be inconveniently large for frequencies up to several GHz.

It is convenient to approximate the shape of the main lobe central area by

$$\frac{D(\beta)}{D_{\max}} = \begin{cases} \cos^n \beta & \text{for } |\beta| \leq \frac{\pi}{2} \\ 0 & \text{for } |\beta| > \frac{\pi}{2} \end{cases}, \tag{2.84}$$

where β is the bearing angle and n is a beam shape coefficient. According to [13], this approximation is valid for numerous antenna patterns. The approximate gains and half-power beamwidths for the particular values of the n coefficient are provided in Table 2.2; Figure 2.7b presents the approximate shapes of the main lobe central area.

[15] The directivity is a ratio of radiation intensity (in a certain direction) of the antenna and the intensity radiated by the isotropic radiator. The maximal directivity is often denoted as the "antenna directivity" parameter. Gain antenna is the directivity of the antenna minus losses; for a lossless antenna directivity is equal to gain.

[16] The half-power beamwidth of antenna is the angle between the points of directivity (gain) pattern, where the value is half of the maximal directivity (gain). Half of the power corresponds to −3 dB.

For the purpose of AoA we are not interested in the absolute value of the gain, rather the angular variation of the gain is exploited. The received power yields

$$P(\beta) = P(0)\frac{D(\beta)}{D_{\max}} \tag{2.85}$$

$$P_{\mathrm{dB}}(\beta) = P_{\mathrm{dB}}(0) + 10n\,\log_{10}(\cos\beta). \tag{2.86}$$

For simplicity we will assume log-normal distribution of the received power; i.e. channel with large scale fading and no noise floor. The same assumption has been taken in case of signal-power-based ranging in Section 2.1.1. Consequently, the power in the decibel scale is Gaussian distributed as follows:

$$\tilde{P}_{\mathrm{dB}}(\beta) = \mathcal{N}\left(P_{\mathrm{dB}}(\beta),\,\sigma^2_{\mathrm{lf}}\right) = P_{\mathrm{dB}}(\beta) + \mathcal{N}\left(0,\,\sigma^2_{\mathrm{lf}}\right). \tag{2.87}$$

In order to obtain the CRLB, the likelihood function of the bearing angle β has to be formulated:

$$\begin{aligned}\mathcal{L}(\beta|\tilde{P}_{\mathrm{dB}}) &= \mathrm{p}(P_{\mathrm{dB}}|\beta)\\ &= \frac{1}{\sqrt{2\pi\sigma^2_{\mathrm{lf}}}}\exp\left(-\frac{\left(\tilde{P}_{\mathrm{dB}} - P_{\mathrm{dB}}(0) - 10n\,\log_{10}(\cos\beta)\right)^2}{2\sigma^2_{\mathrm{lf}}}\right).\end{aligned} \tag{2.88}$$

The log-likelihood follows

$$\begin{aligned}\Lambda(\beta|\tilde{P}_{\mathrm{dB}}) &= \ln\mathcal{L}(\beta|\tilde{P}_{\mathrm{dB}})\\ &= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2_{\mathrm{lf}}}}\right) - \frac{\left(\tilde{P}_{\mathrm{dB}} - P_{\mathrm{dB}}(0) - 10n\,\log_{10}(\cos\beta)\right)^2}{2\sigma^2_{\mathrm{lf}}},\end{aligned} \tag{2.89}$$

and its derivative with respect to the parameter θ is simplified to the form

$$\begin{aligned}\frac{\partial\Lambda(\beta|P_{\mathrm{dB}})}{\partial\beta} &= 10n\sin\beta\,\frac{\tilde{P}_{\mathrm{dB}} - P_{\mathrm{dB}}(0) + \frac{10n\,\ln(\cos\beta)}{\ln 10}}{\sigma^2_{\mathrm{lf}}\ln 10\,\cos(\beta)}\\ &= \frac{10n\tan\beta}{\sigma^2_{\mathrm{lf}}\,\ln 10}\left(\tilde{P}_{\mathrm{dB}} - P_{\mathrm{dB}}(0) + 10n\,\log_{10}(\cos\beta)\right).\end{aligned} \tag{2.90}$$

TABLE 2.2.: Approx. beamwidth and directivity for $\cos^n$ main lobe shape

Shape coeff. n **[-]**	**Half-power beamw.** $\beta_{3\mathrm{dB}}$ **[deg.]**	**Max. directivity** $D_{\max}$ **[dB]**
4	65.4	11.0
6	53.9	12.6
10	42.1	14.8
15	34.5	16.5
25	26.8	18.7
50	19.0	21.6

Clearly, the bracket corresponds to the log-normally (in linear scale) or normally (in decibel scale) distributed variable with zero mean and variance equal to the variance used in the large scale signal fading model σ_{lf}^2. The typical values of the variance for certain scenarios areas are available in Table 2.1 on page 9. The Fisher information is obtained as

$$\begin{aligned}\boldsymbol{\mathcal{I}}(\beta) &= \mathrm{E}\left[\left(\frac{10n\tan\beta}{\sigma_{\text{lf}}^2\,\ln 10}\mathcal{N}(0,\,\sigma_{\text{lf}}^2)\right)^2\right] \\ &= \frac{100n^2\,\tan^2\beta}{\sigma_{\text{lf}}^2\,\ln^2 10}.\end{aligned} \tag{2.91}$$

By means of fraction inversion and optional square-root the CRLB for $\cos^n(\beta)$ main lobe shape under log-normal fading is obtained.

$$\begin{aligned}\sqrt{\mathrm{var}(\hat{\beta})} &\geq \frac{\sigma_{\text{lf}}\,\ln 10}{10\,n|\tan\beta|} \\ \sqrt{\mathrm{var}(\hat{\beta})} &\geq \frac{\ln 10}{10\,n}\sigma_{\text{lf}}\,|\cot\beta|.\end{aligned} \tag{2.92}$$

Due to the nature of the cotangent function, the standard deviation of the bearing angle estimate $\hat{\beta}$ grows to infinity for true value of $\beta \to 0$. The reason for such gradual reduction of accuracy is the flatness of the main lobe directivity (gain) pattern in the vicinity of the maximum. Consequently, steering the antenna towards the maximum power is inaccurate and impractical.

On the other hand, the nulls of the antenna characteristics are usually very steep, and (theoretically) infinitely deep. In such case the antenna would be steered towards the minimum of the received power. For simplicity, let us assume infinitesimally small radiating loop, which is sometimes called Hertzian loop. We consider that the axis of the loop is horizontally oriented and aligned with bearing $\theta = 0°$. According to [12], the normalized power pattern o is of $\sin^2$ shape, therefore the received power yields

$$P(\beta) = P(0)\frac{D(\beta)}{D_{\max}} = P(0)\sin^2\beta \tag{2.93}$$

$$P_{\text{dB}}(\beta) = P_{\text{dB}}(0) + 20\,\log_{10}(\sin\beta). \tag{2.94}$$

Under assumption of the log-normal fading we may obtain the partial derivative of the log-likelihood of the bearing angle in the following form:

$$\begin{aligned}\frac{\partial\Lambda(r|P_{\text{dB}})}{\partial r} &= -20\cos\beta\,\frac{\tilde{P}_{\text{dB}} - P_{\text{dB}}(0) + \frac{10\,\ln(\sin^2\beta)}{\ln 10}}{\sigma_{\text{lf}}^2\ln 10\,\sin(\beta)} \\ &= \frac{-20\cot\beta}{\sigma_{\text{lf}}^2\,\ln 10}\,\left(\tilde{P}_{\text{dB}} - P_{\text{dB}}(0) + 10\,\log_{10}(\sin^2\beta)\right).\end{aligned} \tag{2.95}$$

Then, it is straightforward to obtain the CRLB value by the same approach as was used in the power-maximum search. The CRLB follows

$$\sqrt{\mathrm{var}(\hat{\beta})} \geq \frac{\ln 10}{20}\sigma_{\text{lf}}\,|\tan\beta|. \tag{2.96}$$

This result suggests infinite accuracy in the vicinity of zero bearing parameter, which is practically impossible. The physical discrepancy comes from the assumption of log-normal power distribution. Consequently, the model behaves the same regardless the absolute received power; it does not matter whether it is +150 dBm or −150 dBm, because the noise floor is neglected. Therefore, the log-normal power model is suitable for high values of SNR only.

In Figure 2.8a (page 33) the CRLBs are plotted for several values of the large-scale fading coefficients (see Table 2.1 on page 9). The red curves depict the bearing standard deviation bound near the maximum of a $\cos^4$-shaped main lobe of the antenna pattern, the blue curves present the bound near the null of a $\sin^2$-shaped pattern. The result suggests that under large scale signal fading the exploitation of the pattern null outperforms the exploitation of pattern maximum. It is worth noting that for more directive antennae the maximum of the pattern is more pointy, the n coefficient grows, and the CRLB is consequently lower. Still, it diverges to infinity near the maximum due to the cotangent shape of the CRLB. On the contrary, a relatively simple dipole or loop antenna is sufficient for achieving approximately the $\sin^2$-shaped pattern.

For the low values of SNR a different model has to be assumed. Especially when exploiting the null of the antenna pattern, the proximity of the noise floor may play more important role than the fading. It is known that the amplitude distribution of a radio signal envelope under white noise conditions is Rician, and converges to Rayleigh distribution when the signal component disappears. The power of the noisy RF signal therefore follows a non-central χ^2 distribution with two degrees of freedom; in the absence of the useful signal it converges to central χ^2 distribution. In general, the PDF of the non-central χ^2 distribution of variable X follows

$$_{\text{n.c.}}\chi^2_k(\upsilon) = \frac{1}{2} e^{\frac{X+\upsilon}{-2}} \left(\frac{X}{\upsilon}\right)^{\frac{k-2}{4}} I_{\frac{k-2}{2}}\left(\sqrt{\upsilon X}\right), \tag{2.97}$$

where υ is the non-centrality parameter, k is the count of degrees of freedom, and I_m denotes a modified Bessel function of the first kind and order m. For our case with two degrees of freedom we may use formula

$$_{\text{n.c.}}\chi^2_2(\upsilon) = \frac{1}{2} e^{\frac{X+\upsilon}{-2}} I_0\left(\sqrt{\upsilon X}\right). \tag{2.98}$$

Provided that the power of the noise is constant, the received signal power is proportional to the received signal SNR. We may describe relation of the SNR on reception with the bearing β by a simple function, let us denote it $\mathrm{f}(\beta)$. For the maximum search we are looking for the shape of the main lobe shape and therefore the received SNR is

$$\mathrm{SNR} = \mathrm{f}(\beta) = \mathrm{SNR}_{\max} \cos^n \beta, \tag{2.99}$$

where $\mathrm{SNR}_{\max}$ is the maximal possible received SNR. In contrast, the pattern shape for the minimum power search is

$$\mathrm{SNR} = \mathrm{f}(\beta) = \mathrm{SNR}_{\max} \sin^2 \beta. \tag{2.100}$$

Note that the same main antenna pattern shape approximations are used as were used in the log-normal distribution case.

TABLE 2.3.: Directivity pattern shape functions and their derivatives

Type	$\mathrm{f}(\beta)/\mathrm{SNR}_{\max}$	$\mathrm{f}'(\beta)/\mathrm{SNR}_{\max}$	$(\mathrm{f}'(\beta)/\mathrm{SNR}_{\max})^2$
Peak	$\cos^n(\beta)$	$-n\sin(\beta)\cos^{n-1}(\beta)$	$n^2\sin^2(\beta)\cos^{2n-2}(\beta)$
Null	$\sin^2(\beta)$	$2\sin(\beta)\cos(\beta)$	$\sin^2(2\beta)$

Note: The pattern shape functions were normalized by $\mathrm{SNR}_{\max}$ for the sake of table clarity. Non-normalized versions have to be used in the CRLB derivation.

It may be easily derived that the measured $\tilde{\mathrm{SNR}}$ (note the tilde that distinguishes the noisy version) will have non-central χ^2_2 distribution with the non-centrality parameter equal to the true SNR, which depends on the bearing. By substituting pattern the shape function into the χ^2_2 we can get the formula for the log-likelihood of bearing β w.r.t. measured $\tilde{\mathrm{SNR}}$:

$$\begin{aligned}\Lambda(\beta|\tilde{\mathrm{SNR}}) &= \ln\mathcal{L}(\beta|\tilde{\mathrm{SNR}}) = \ln p(\tilde{\mathrm{SNR}}|\beta)\\ &= \ln\left({}_{\mathrm{n.c.}}\chi^2_2(\mathrm{f}(\beta))\right)\\ &= \ln\frac{1}{2} - \frac{\mathrm{f}(\beta)+\tilde{\mathrm{SNR}}}{2} + \ln I_0\left(\sqrt{\mathrm{f}(\beta)\tilde{\mathrm{SNR}}}\right).\end{aligned} \tag{2.101}$$

The partial derivative w.r.t. to the parameter is found in the general form by using the chain rule.

$$\frac{\partial\Lambda(\beta|\tilde{\mathrm{SNR}})}{\partial\beta} = \frac{\mathrm{f}'(\beta)}{2}\left(\sqrt{\frac{\tilde{\mathrm{SNR}}}{\mathrm{f}(\beta)}}\frac{I_1\left(\sqrt{\mathrm{f}(\beta)\tilde{\mathrm{SNR}}}\right)}{I_0\left(\sqrt{\mathrm{f}(\beta)\tilde{\mathrm{SNR}}}\right)} - 1\right) \tag{2.102}$$

Here, the $\mathrm{f}'(\beta)$ denotes the first derivative of $\mathrm{f}(\beta)$, of course with respect to β.

The Fisher information is then found by a simple substitution of the results above. The expectation operation is performed from the definition, i.e. by integration of the function multiplied by the PDF.

$$\begin{aligned}\boldsymbol{\mathcal{I}}(\beta) &= \mathrm{E}\left[\left(\frac{\partial\Lambda(\beta|\tilde{\mathrm{SNR}})}{\partial\beta}\right)^2\right] = \int_0^\infty\left(\frac{\partial\Lambda(\beta|\tilde{\mathrm{SNR}})}{\partial\beta}\right)^2 p(\tilde{\mathrm{SNR}}|\beta)\,\mathrm{d}\,\tilde{\mathrm{SNR}}\\ &= \int_0^\infty\frac{\mathrm{f}'(\beta)^2}{8}\left(\sqrt{\frac{\tilde{\mathrm{SNR}}}{\mathrm{f}(\beta)}}\frac{I_1\left(\sqrt{\mathrm{f}(\beta)\tilde{\mathrm{SNR}}}\right)}{I_0\left(\sqrt{\mathrm{f}(\beta)\tilde{\mathrm{SNR}}}\right)} - 1\right)^2 e^{-\frac{\mathrm{f}(\beta)+\tilde{\mathrm{SNR}}}{2}} I_0\left(\sqrt{\mathrm{f}(\beta)\tilde{\mathrm{SNR}}}\right)\mathrm{d}\,\tilde{\mathrm{SNR}}\end{aligned} \tag{2.103}$$

For the sake of completeness, the simplified expressions for the derivative and the squared derivative of $\mathrm{f}(\beta)$ are provided in Table 2.3.

The Fisher information formula cannot be evaluated in a closed form, nonetheless, the integral can be evaluated in a numerical way. We have used MATLAB `vpaintegral` function, which uses variable-precision arithmetic to approximate the result; the relative accuracy was set to 10^{-20}. Ramos–Llorden suggests that the upper integration

bound should be 50f(β) and the lower integration bound is always zero [14], however, this was found not being applicable across the couple of orders of magnitude of the non-centrality parameter of the χ^2 distribution. The upper (infinite) bound was set to $\mathrm{f}(\beta) + 20\sqrt{\mathrm{f}(\beta)}$ or 100 – whichever is greater – in order to maintain numerical stability of the computations; the lower bound was set to $\mathrm{f}(\beta) - 20\sqrt{\mathrm{f}(\beta)}$ or zero, should the value be negative.[17] Such approach secures integration interval wide enough to provide reasonable accuracy while avoiding unnecessary integration of "almost-zero" portions of the integrand.

The CRLB of the bearing angle estimate $\hat{\beta}$ is obtained by scalar inversion of the Fisher information $\boldsymbol{\mathcal{I}}(\beta)$. The numerical evaluation was performed for $\cos^4$ and $\sin^2$ shape functions and $\mathrm{SNR}_{\max}$ values from 5 to 50 dB (from $10^{0.5}$ to 10^5 in the linear scale). The results are presented in Figure 2.8b. Clearly, in both cases the lower bound grows to infinity for $\beta \to 0$. The red and blue meshes depict the CRLBs for the $\cos^4$-shaped and $\sin^2$-shaped patterns, respectively.

In the vicinity of the $\cos^4$ antenna pattern maximum, the steepness (derivative) of the shape function converges to zero in a rather smooth way. Consequently, the CRLB grows steadily to the infinity, even for the high values of $\mathrm{SNR}_{\max}$. When the $\mathrm{SNR}_{\max}$ is low, the bearing estimate accuracy is deteriorated as the measurement is noisy.

For the $\sin^2$ pattern null shape, in contrast, the bearing estimate $\hat{\beta}$ accuracy under high $\mathrm{SNR}_{\max}$ conditions improves towards the lower values of the true bearing value β. Near the zero bearing value the power of the incoming signal is suppressed to the noise floor level, and consequently the CRLB grows towards infinity. As the $\mathrm{SNR}_{\max}$ decreases, the accuracy degrades further from the zero bearing, because larger section of the pattern null is "hidden" under the noise floor. Additionally, the steepness of the null shape decreases with the growing bearing angle and makes the transition between the noise-induced inaccuracy and null-steepness-induced inaccuracy smoother.

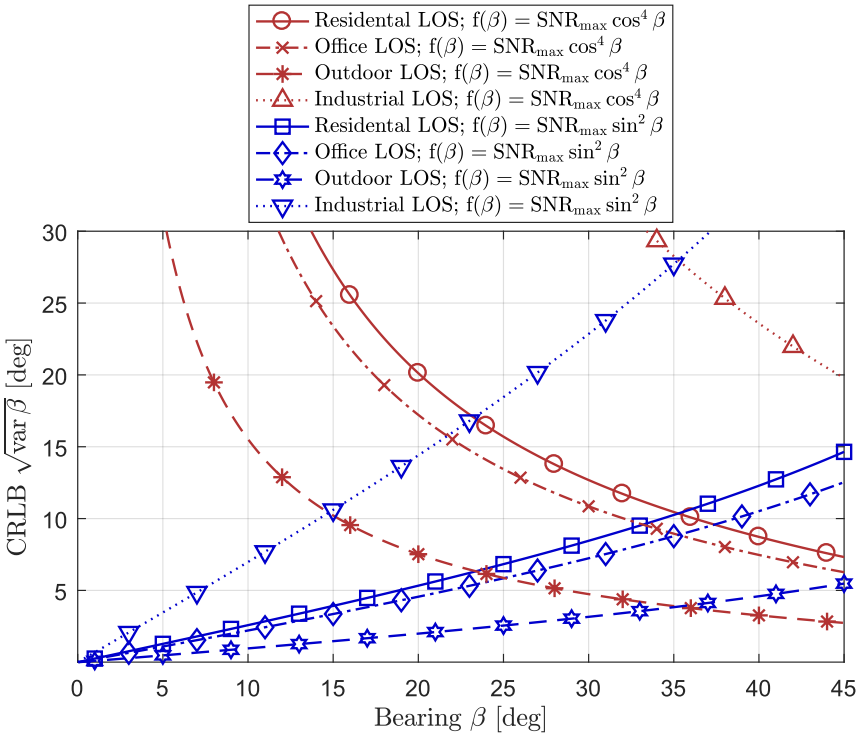
To conclude this section, we can state that exploitation of the null is favorable to using maximum of the antenna pattern, since it provides better accuracy in general. The statement is supported by the evaluation of CRLBs for two received signal models: with log-normal distribution of the power and with non-central χ^2 distribution of the signal-to-noise ratio.

[17] For low values of the non-centrality parameter of the non-central χ^2 distribution converges to central χ^2 distribution. For χ^2_2 (two degrees of freedom) the PDF values beyond 100 are almost zero (less than 10^{-22}), since it follows the function $e^{-x/2}/2$.

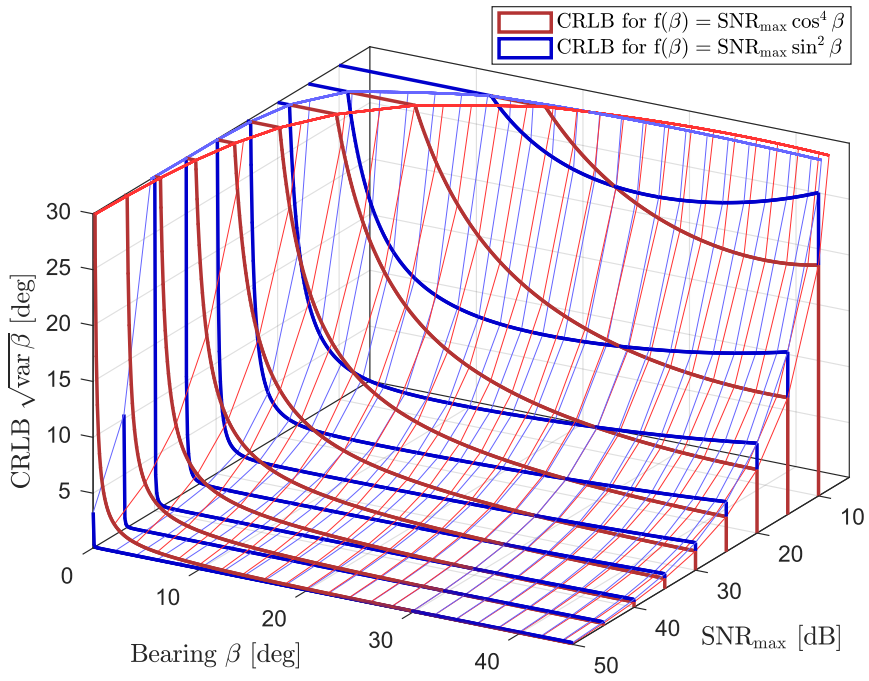
For higher values of the non-centrality parameter, the non-central χ^2 may be approximated by a normal distribution. According to [15], if a random variable X follows ${}_{\text{n.c.}}\chi^2_k(\upsilon)$ then

$$\frac{X - k - \upsilon}{\sqrt{2k + 4\upsilon}} \to \mathcal{N}(0,\, 1) \quad \text{for} \quad k \to +\infty \quad \text{or} \quad \upsilon \to +\infty.$$

Considering $k = 2$ and $\upsilon \gg k$, the distribution of X may be approximated by $\mathcal{N}(\upsilon,\, 4\upsilon)$. Since the Gaussian PDF is sufficiently decayed at 10σ point above the mean (approx. 10^{-22} w.r.t. to the maximum of the PDF), we may set the integration bound to this point, i.e. $\upsilon + 10\sqrt{4\upsilon}$. The same applies for the lower bound, i.e. lower bound may be set to 10σ point below the mean, i.e. $\upsilon - 10\sqrt{4\upsilon}$. Note that the negative values of the lower bound (when $\upsilon < 400$) have to be clipped to zero. By substituting $\mathrm{f}(\beta)$ for υ we obtain the bound values mentioned above.



(A) Log-normal power distribution



(B) Non-central χ^2 SNR distribution

FIGURE 2.8.: CRLB of signal-power based AoA.

II. Delay and phase approach

When an antenna array[18] is used, the signal path length may slightly differ for the individual elements. As a consequence, the signal phase or delay may be different at the elements of the array. For simplicity, let us assume a simple, linear array consisting of M isotropic elements with a regular element spacing d. We will use letter i to index the elements, the reference element has $i = 0$. The bearing angle β will be defined as the angle from the axis of the array.

In general, the bearings β_i and the ranges to the signal source r_i differ for the different elements, as can be seen in Figure 2.9a. However, the source is assumed to be far from the array (condition $r_0 \gg d$ holds) and therefore the bearings can be assumed equal for all elements, i.e.

$$\beta = \beta_0 \approxeq \beta_1 \approxeq \ldots \approxeq \beta_\mathrm{i} \approxeq \ldots \approxeq \beta_\mathrm{M}. \tag{2.104}$$

Since the paths of the signal to the individual elements are considered perpendicular near the array, the difference of the ranges to the signal source depends on the bearing in the following way:

$$\Delta r_\mathrm{i} = i\, d\, \cos\beta, \tag{2.105}$$

see Figure 2.9b for the graphical representation. Due to the path-length difference, the signal arrives to the other antennae later or earlier, depending on the bearing.

Most commonly, a narrowband signal is assumed and most of the array dimensions are expressed as the multiples of the carrier wavelength λ. In the classical antenna array theory this is rather useful, because the same methods may be applied regardless the frequency, wavelength and absolute dimensions. Under the narrowband assumption, the delay of the signal is perceived as a phase-shift, since a wavelength-long signal path corresponds to a 360° shift.

Kay [1] derives the CRLB for narrowband signals, where the dependence on the wavelength is inherently included:

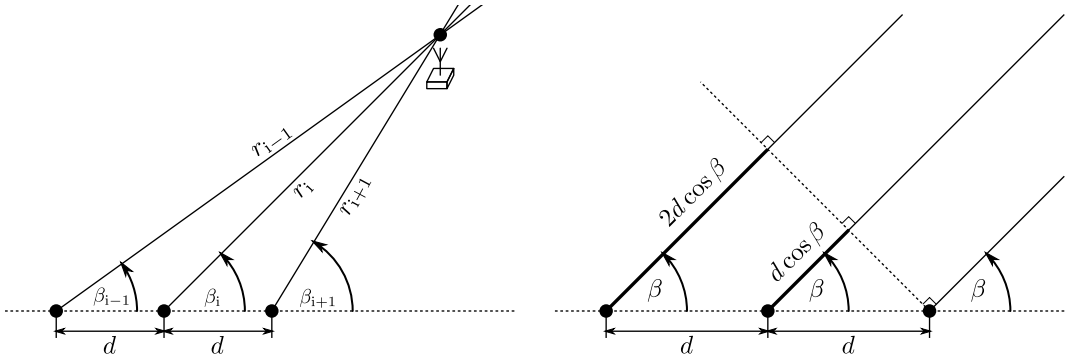
$$\begin{aligned}
\sqrt{\operatorname{var}\hat{\beta}_\mathrm{nb}} &\geq \frac{\sqrt{3}}{\frac{l}{\lambda}\pi\sqrt{\mathrm{SNR}\,\frac{M(M+1)}{M-1}}\,|\sin\beta|} \\
\sqrt{\operatorname{var}\hat{\beta}_\mathrm{nb}} &\geq \frac{c_0\sqrt{3}}{f\,d\,\pi\sqrt{\mathrm{SNR}\,M(M^2-1)}\,|\sin\beta|}.
\end{aligned} \tag{2.106}$$

It is worth noting that both expressions are equivalent due to the link between wavelength and frequency.

Nonetheless, the narrowband assumption is inapplicable in certain cases; the corresponding Cramér–Rao lower bound would be inaccurate. For instance, the UWB signals often exhibit relative bandwidth more than 20 % and absolute bandwidth of a few gigahertz. Owing to the persisting validity of the signal reception principles

[18] Usually, the antenna *arrays* consist of a several *elements* – typically a simple antennae. The relative element position and the feeding amplitude and phase of the elements is tuned to provide desired gain pattern. This point of view is widely used in the antenna theory, see e.g. [12, chap. 22–23].

Herein the same principles are exploited in the opposite way; the phase of the received signal at the individual elements with known relative position is used to estimate the angle of signal arrival, i.e. the bearing of the source.



(A) Non-simplified description (B) Simplification from distant signal source

FIGURE 2.9.: Schematic description of the principal assumption on a linear array.

described in the beginning of this section and within Figures 2.9, the signal delay on the individual elements can be still described by simple trigonometric relations. The physical dimensions cannot be interpreted as λ-multiples, since it is preposterous to characterize the wideband signals by a single wavelength.

According to [2, 16, 17], the CRLB for the array-based AoA may be expressed in two equivalent forms (2.107) and (2.108):

$$\operatorname{var}\hat{\beta}_{\mathrm{wb}} \geq \frac{6{c_0}^2\mathcal{N}_0}{d^2\int|s'(t)|^2\,\mathrm{d}t\,M(M^2-1)\sin^2\beta} \tag{2.107}$$

$$\operatorname{var}\hat{\beta}_{\mathrm{wb}} \geq \frac{6{c_0}^2}{d^2\,(B_{\mathrm{RMS}}\,2\pi)^2\,\mathrm{SNR}\,M(M^2-1)\,\sin^2\beta}$$

$$\sqrt{\operatorname{var}\hat{\beta}_{\mathrm{wb}}} \geq \frac{\sqrt{3}\,c_0}{B_{\mathrm{RMS}}\,d\,\pi\sqrt{\mathrm{SNR}\,M(M^2-1)}\,|\sin\beta|}. \tag{2.108}$$

The symbol B_{RMS} denotes the RMS bandwidth, as was defined by (2.20); The SNR is equal to $2\mathcal{E}/N_0$. It is worth noting that the two expressions (2.107) and (2.108) are related through the Parseval's identity. The comparison of the narrowband and wideband CRLB of the AoA is provided in [17]. Both CRLBs are almost equivalent for narrow bandwidths, however, the wideband CRLB is lower than its narrowband approximation for wider bandwidths.

As an example of the CRLB we will provide a result for a baseband root-raised cosine pulse, whose power spectrum is defined as [18][19]

$$\mathcal{S}_{\mathrm{RRC}}(f) = \begin{cases} T_{\mathrm{p}} & \text{for} \quad |f| < \frac{1-\zeta}{2T_{\mathrm{p}}} \\ \frac{T_{\mathrm{p}}}{2}\left(1+\cos\left(\frac{\pi T_{\mathrm{p}}}{\zeta}\left(|f|-\frac{1-\zeta}{T_{\mathrm{p}}}\right)\right)\right) & \text{for} \quad \frac{1-\zeta}{2T_{\mathrm{p}}} < |f| < \frac{1+\zeta}{2T_{\mathrm{p}}} \\ 0 & \text{for} \quad |f| > \frac{1+\zeta}{2T_{\mathrm{p}}}, \end{cases} \tag{2.109}$$

where T_{p} is the pulse duration and ζ is the roll-off factor of the raised cosine shape.[20] The roll-off factor has to be chosen from interval from 0 to 1, where 0 corresponds to

[19] The amplitude spectrum of the root-raised cosine pulse is a square root of (2.109).

[20] Commonly, the roll-off factor is denoted by symbol β, however, ζ was used here in order to avoid confusion with the bearing angle.

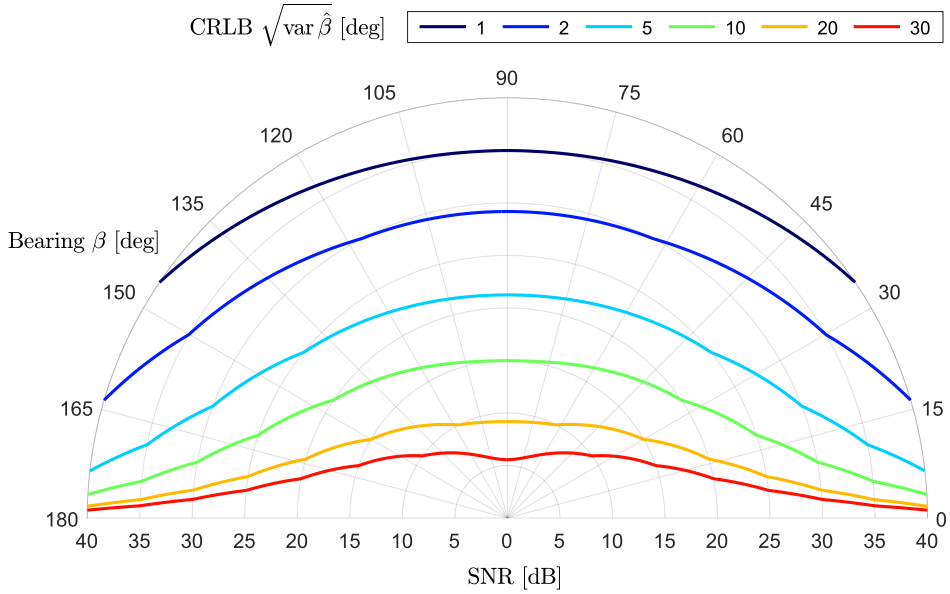


FIGURE 2.10.: The wideband CRLB of delay-based AoA for raised-cosine pulse with duration $T_\mathrm{p} = 2\,\mathrm{ns}$. The linear array consists of $M = 5$ elements equally spaced by $d = 10\,\mathrm{cm}$.

fully-rectangular shape and 1 to fully-cosine shape. For our purpose it is set to $\zeta = 0.5$; under such assumption the $B_\mathrm{RMS} \approx 0.31/T_\mathrm{p}$.[21] This example was chosen, since it is the reference pulse shape for the HRP (high rate pulse repetition frequency) UWB physical layer definition in the IEEE 802.15.4 standard [19, p. 464-466]. In Figure 2.10 we provide the evaluation of the CRLB for a root-raised-cosine-shaped pulse with 2 ns duration and a five-element array. Clearly, the bearing estimate accuracy deteriorates for the true bearing values around 0° and 180°. It is worth mentioning that because the RMS bandwidth is inversely proportional to the pulse duration T_p, the CRLB

$$\sqrt{\text{var}\,\hat{\beta}_\mathrm{wb}} \propto T_\mathrm{p}. \tag{2.110}$$

Consequently, the CRLB for different pulse durations may be obtained by means of a simple linear scaling.

[21] The two-sided null-to-null bandwidth of the root-raised cosine pulse is obviously $(1+\zeta)/T_\mathrm{p}$. The RMS bandwidth for the baseband representation of the pulse follows

$$B_\mathrm{RMS} = \frac{\sqrt{3\pi^2\zeta^2 - 24\zeta^2 + \pi^2}}{2\sqrt{3}\pi T_\mathrm{p}}$$

which is a monotonically rising function from approximately $0.29/T_\mathrm{p}$ at $\zeta = 0$ through $0.31/T_\mathrm{p}$ at $\zeta = 0.5$ to $0.36/T_\mathrm{p}$ at $\zeta = 1$.

2.2. Global Navigation Satellite Systems

The GNSS are the a first example of the radionavigation systems by numerous reasons. Indisputably, they are the most widely used system that provides position, velocity and time information PVT to the user. In simplicity, each GNSS satellite transmits a unique ranging signal, the transmission of the signal is synchronized to the main system time. The signals are received by the user equipment ant the time of arrival (ToA) is measured for each signal. The positions of the satellites are well known at all time epochs, and this information is available to the user through a navigation data message that is transmitted within the satellite signal.

In general, all the systems consist of three segments: Control, Space and User. Our description flows from the system side to the user. The control system is in duty of tracking and maintaining the constellation of the satellites, i.e. monitoring, estimation and prediction satellite orbital parameters, state of their atomic clocks w.r.t. the system clock, modeling of ionospheric and tropospheric effects, *et cetera*, and uploading the relevant data to the respective satellites via a microwave up-link channel. For that purpose, a few interconnected ground-based stations are deployed around the globe.

The space segment consists of satellites, typically in Medium Earth Orbit (MEO) with highly stable clocks. Each satellite transmits one or more ranging signals. The signals vary by their structure, carrier frequency or encryption, i.e. availability to wide public. Commonly, the satellites transmit almost exclusively in the L-band, which is a specific subset of the UHF-band,[22] exceptions exist, however.

The user segment is a receiver of one or more of the GNSS signals. From the measurements taken on the signals, so called *observables*, the PVT solution is obtained.

One of the clear advantages of the GNSS is that the user equipment is passive, i.e. does not have to transmit anything in order to obtain the position estimate. For users it has several important consequences; firstly, the number of users is unlimited. The signal is available in space and any number of receivers can "listen" to it. Secondly, position information is available only to the user, unless shared by other means – the user cannot be tracked by the GNSS system itself, since it does not provide the user *uplink* channel. Thirdly, the equipment may be completely silent, and therefore cannot be located by another system; this is particularly important for the military users, since it does not allow enemy to localize the user, who tries to obtain the position.

In the following section we will briefly summarize the measurements taken by a GNSS receiver. The limitations and errors of such measurements will be reviewed. A section describing possible exploitation of such measurements follows. An overview of the currently available GNSS is provided as well.

2.2.1. Observables, errors and limitations

As users, we only can affect the processing done in the user segment, i.e. the receiver. Therefore we find useful to describe how the GNSS observables are obtained, and what effects they suffer from. This knowledge is fundamental for understanding the strengths and weaknesses of the GNSS.

[22] The L-Band covers frequencies from 1 to 2 GHz [20], the GNSS use carrier frequencies in the bands from approximately 1150 to 1300 MHz and from 1550 to 1600 MHz.

Within this work, the description of the GNSS is somehow truncated and simplified in some aspects. An excellent guide to GPS and GNSS systems may be found in books [9, 21, 22].

I. Signal Structure and Receiver Observables

A typical GNSS signal, such as GPS's C/A code, consists of three components – the carrier wave is spreaded by the pseudo-random ranging code and then modulated by the data message. The signal is continuously transmitted by the satellites. In the following mathematical description $c(t)$ and $d(t)$ are the continuous-time representation of the ranging code and data message.

$$s_C A = \sin(2\pi f_L 1) c(t) d(t) \tag{2.111}$$

$$c(t) = \sum_k \mathrm{rect}(\frac{t}{T_{\mathrm{chip}}} - k - \frac{1}{2}) c[k] \tag{2.112}$$

$$d(t) = \sum_l \mathrm{rect}(\frac{t}{T_{\mathrm{sym}}} - l - \frac{1}{2}) d[k] \tag{2.113}$$

Let us consider the legacy C/A code as a representative example. The GPS L1 carrier frequency is 1575 MHz, and it is a pure sine wave. The ranging code is a sequence of chips with very sharp and unambiguous correlation function that allows tracking of the signal delay. The sequence is pseudorandom and therefore it is inherently periodic. In case of the C/A code it is a BPSK-modulated 1023-chip long Gold sequence with a period 1 ms, i.e. chip rate is 1.023 Mchips/s and $t_{\mathrm{chip}} = 1/1.023 \cdot 10^6$ s. The representation of the sequence (in the discrete time) $c[k]$ is a periodic sequence of ± 1s. In the continuous time version of the ranging code $c(t)$, the rectangular pulse shape is chosen. Note that the definition of the rectangular function follows[23]

$$\mathrm{rect}(t) = \begin{cases} 1 & \text{for} \quad |t| < \frac{1}{2} \\ \frac{1}{2} & \text{for} \quad |t| = \frac{1}{2} \\ 0 & \text{for} \quad |t| > \frac{1}{2}. \end{cases} \tag{2.114}$$

In the equation (2.112) it is shifted so that it begins at kT_{chip} and has duration of T_{chip}. Theoretically, an arbitrary pulse shape can be used for a linear modulation such as BPSK or any M-PSK. The rectangular shape is quite straightforward and has a few other advantages, however. First, in its ideal form, it does not suffer from inter-symbol interference; second, any M-PSK symbols with rectangular shape satisfy constant-envelope criterion; third, it has a sharp auto-correlation function.

Each of the satellites has a unique "variant" of the pseudo-random code,[24] in case of GPS the variant is determined by the PRN number. The cross-correlation between the ranging codes used by different satellites is minimized by design for all mutual delays. Therefore, all the satellites may transmit at the same carrier frequency without

[23] The assignment of the value 1/2 in the rect(·) function discontinuities is favorable, since the Fourier series converges to this point for the periodic rectangular function, for instance sign(sin(t)).

[24] Russian system GLONASS is an exception from this point of view.

interfering – such technique of sharing the radio space is called CDMA (code-division multiple-access).[25]

The data message is BPSK-modulated as well, with a 50 Hz rate, though. Consequently, the polarity of the ranging code may be inverted according to the data message content every $T_{\text{sym}} = 20\,\text{ms}$, i.e. every 20 periods of the ranging Gold sequence. The bits of the data message are organized in frames, subframes, *et cetera*, include predefined preambles and reserved bits; however, this organization differs from system to system and is not important for the principal description of the GNSS operation. The navigation message contains

Timestamp – a time of transmission of a certain point in the navigation message. Since the navigation message is aligned with the ranging code (e.g. 20 periods of Gold code in one data symbol for GPS C/A) so it serves also as a timestamp of a certain period of the ranging code.

Ephemeris – the data that is needed to compute satellites position and velocity at any time. Keplerian parameters of the orbit with perturbations or coefficients of differential equations describing the satellite motion may be included.

Model coefficients – corrections for the pre-defined models of the satellite hardware and especially clock imperfections, or propagation effects.

Status – satellite configuration (i.e. which other signal or its component is active), satellite health, expected pseudorange accuracy, *et cetera*.

Almanac – information about the other satellites in the constellation, such as less accurate ephemeris data and health.

Age and validity of the data – typically the time of ephemeris issue and almanac issue, alongside with the interval of validity (usually several hours).

It has to be noted that the purpose of the list above is to give a brief idea of the navigation message contents. The exact information for the various GNSS can be found in the respective Interface Control Documents (ICDs).

As the signal propagates from a satellite to the user it is delayed and attenuated. In addition, due to the relative movement of the user and the satellite (the satellite orbits rotating Earth) the signal is frequency-shifted by a significant amount. Naturally, the values of delay, Doppler shift and signal power depend on the geometrical constellation of the receiver and a single satellite.

Frequency Locked Loops (FLL) and Phase Locked Loops (PLL) are used to track the carrier wave – its Doppler-shifted frequency and carrier phase. The receiver exploits a Delay Locked Loop (DLL) in order to track the delay (or phase) of the periodic ranging code. Since the satellites, i.e. transmitters of the signals, are synchronized, we may denote the delays of the received ranging sequences w.r.t. the receiver clock as the

[25] In order to satisfy the non-interference criterion of the signals, they have to be *orthogonal*. In the CDMA case the ranging/spreading code sequences are chosen so that the mutual energy of any two arbitrarily shifted sequences is negligible to the signal energy itself, i.e. that the arbitrarily shifted signals are are *almost* orthogonal. The other means to achieve the orthogonality is the FDMA (frequency-division MA), since the harmonic signals with different frequencies are orthogonal, or TDMA (time-division MA), the orthogonality is forced by having only one signal non-zero at a time. For more detailed information see e.g. [18, chap. 15].

pseudoranges. Additionally, the receiver also estimates the signal quality, typically in sense of carrier-to-noise-ratio C/N_0. Let us now review those observables in detail:

Pseudorange – an unambiguous measure of range between satellite and user, however, corrupted by the unknown bias of the user clock and the system clock.

Due to the periodicity of the ranging code the code phase is ambiguous – in case of GPS C/A it is 1 ms, i.e. 300 km. However, the navigation message structure provides resolution of this ambiguity, because it contains timestamps aligned with the ranging code and the number of periods passed from the time reference is known to the receiver.

The accuracy of the measurement depends on the structure of the ranging code (especially the RMS bandwidth) and noise/interference conditions. Also, the quality depends on the factors affecting the signal generation on satellite and propagation effects. See the following sections for more details.

The precision, however, may be on a few-decimeter level, of course under favorable noise conditions [21]. Not only the signal quality, but the structure and parameters of the DLL such as loop bandwidth, loop-filter order, correlator spacing, *et cetera* affect the precision of the pseudorange estimate. The mathematical description of the DLL performance prediction is provided e.g. in [21, 22]. The pseudorange measurements are considered noisier than the other observables [9], nonetheless, their unambiguous nature is advantageous.

It is important to emphasize that the precision of the pseudorange measurement is often better than the accuracy, which is affected by various errors.[26]

Doppler shift – is an unambiguous measurement, as well. It is proportional to the radial velocity of the orbiting satellite and the possibly non-stationary user in the vicinity of the rotating Earth's surface. The motion of the satellite is known and therefore the Doppler shifts may be used alongside the pseudoranges in joint position and velocity estimation. Because the measurement originates from the DLL/PLL tracking subsystem, the measurement is rather precise. Typically, the error is bellow 10 Hz and may drop even below Hertz level under favorable for high C/N_0 ratios.

Carrier Phase – sometimes also denoted as the *integrated Doppler* is an extraordinarily precise measurement, however, it suffers from a fundamental ambiguity originating from the sinusoidal carrier wave periodicity. It is an accumulated phase of the incoming carrier wave w.r.t the reference frequency, i.e. when the Doppler shift is positive, the carrier phase grows and *vice versa*.

The precision is typically a small fraction of the wavelength of the signal, which means millimeter-to-centimeter RMS error. The instantaneous value of the carrier phase is not useful information, however, it may be utilized when observed continuously for a sufficiently long time.[27]

[26] Precision and accuracy should not be interchanged; it is possible to have a very precise, but inaccurate measurements. Despite low random error, the measurement may be biased.

[27] This phenomena can be imagined on an example of two tape measures stretched between the satellite and receiver: The pseudorange is read from a measure with thick, blurry, sparsely spaced, but *labeled* ticks, whilst the carrier phase is obtained from a precise measure with sharp, densely spaced, but *unlabeled* ticks.

Without the ambiguity resolution the carrier phase can be used to smooth the noisy pseudorange measurements [9]. The principle of *carrier smoothing* exploits the shape similarity of the time series of pseudorange and carrier phase measurements. See Section 2.2.2–I for more information.

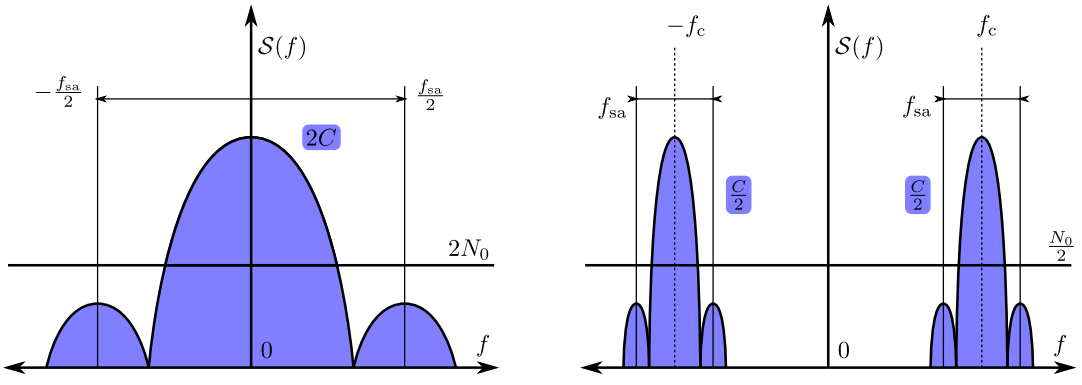
The continuity of the phase measurement is vitally important, especially for the techniques that do resolve the ambiguity. When the phase-lock is lost, the ambiguity resolution for that particular satellite becomes invalid; the ambiguity has to be resolved again when new phase-lock occurs. The same holds for cycle-slip events – they invalidate the ambiguity resolution. Receivers usually provide a status of the PLL, it is usually denoted as LLI (Loss-of-Lock Indicator). For the methods like the carrier smoothing it only means missing data.

Power – receivers also measure the carrier-to-noise ratio C/N_0 or signal-to noise ratio (SNR). To the author's best knowledge, this value is not used to estimate PVT directly, however, it is used as a measure the signal quality, and more importantly, the quality of the measurements taken on the signal. Sometimes, the satellite data is excluded (or less trusted) based on a combination of C/N_0 and elevation criteria. The generally preferred representation is C/N_0 ratio, rather than SNR. The reason is that the C/N_0 (usually expressed in dB/Hz), i.e. the unmodulated carrier power[28] versus noise spectral density, does not change with receiver bandwidth. On the contrary, the SNR (usually expressed in dB) is the ratio of the useful signal power integrated in the receiver bandwidth versus the noise power integrated in the receiver bandwidth. Fig. 2.11 illustrates the difference between the two approaches for real, bandpass signals (i.e. as the signals that propagate through the media), and for a generally complex baseband representation signal (i.e as the signal is processed in the receiver). In the SNR case, both, the signal and the noise power may not be integrated in its entirety due to the limited bandwidth; the power of noise is proportional to the bandwidth.

Many receivers allow logging of the observables for the purpose of post processing, or to allow user to apply custom PVT estimation algorithms. Although the protocol of communication with the receiver is proprietary, some receivers may provide log-file output in a standardized format for the sets of GNSS observables; it is called RINEX – The Receiver Independent Exchange Format [23]. The format definition is available from the International GNSS Service (IGS) website; the standard is evolving as do the GNSS and the receiver technology. The last available version at the time of writing of this thesis was 3.03, which supports observables from all systems mentioned in the oncoming sections.[29] For the purpose of the real-time distribution of the observables (for instance for DGPS or RTK, see Sec. 2.2.2) the standard of the RTCM SC-104 (Special committee fort Differential GNSS Standards) [29]. These standards are usu-

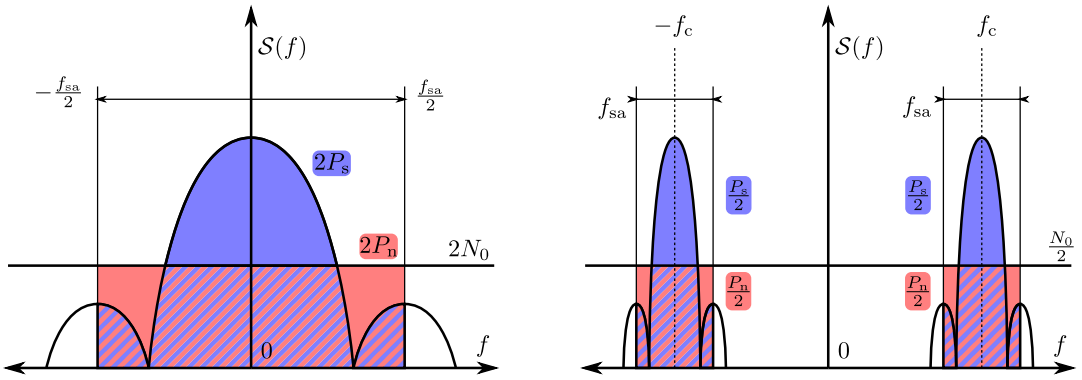
[28] Since the modulation is considered normalized, the modulated signal power is the same.

[29] There are several other formats for sharing GNSS-related data, some of them have common name "roots" with RINEX. From the "human-readable" (based ASCII coded text files) it is for instance IONEX – for ionospheric data grids [24], SINEX – for station solution data (position, velocity, tropospheric parameters, *et cetera*) [25], sp3 – Standard Product 3 (version c, provides GPS and GLONASS orbit solutions) [26], ANTEX – antenna calibration files [27]. A "machine-readable" (binary) format, which should provide the functionality of the formats mentioned above, exists and has a foreseeable name BINEX (Binary Exchange) [28].



(A) C/N_0 of baseband signal

(B) C/N_0 of real, bandpass signal



(C) SNR of baseband signal

(D) SNR of real, bandpass signal

FIGURE 2.11.: Schematic description of the principal assumption on a linear array.

ally denoted only as the RTCM standards, the latest version is RTCM 3.3, yet the RTCM 2.3 is still widely used.[30]

II. Errors and Compensation

The accuracy of the GNSS observables is deteriorated through a number of effects. Not only the propagation delay takes part in the error budget, but also the accuracy of satellite position and time reference is a determining factor for GNSS measurement accuracy, and consequently the positioning accuracy as well. In the following paragraphs the non-negligible errors will be summarized, and for some of them the methods for their suppression or elimination are mentioned. The description of the errors will follow the information flow typical for the GNSS, from the Control and Satellite segment through the non-ideal propagation media to the User segment.

Clocks

The satellite clocks are the keystone of any GNSS. Even if the satellite clocks were ideal, they would be affected by relativistic effects. Two major relativistic effects are

[30] Both versions are maintained by the Committee as the "current" standards; the two versions are not compatible [30].

usually mentioned with regard to the space-vehicle (SV) clock. The first one assumes a perfectly circular orbit; due to this "static" relativistic effect a constant positive frequency shift occurs on the MEO satellites [21].[31] The shift is compensated by intentional detuning of the SV atomic clocks. For example, the 10.23 MHz clocks of the GPS satellites are detuned by $-4.567\,\mathrm{mHz}$, i.e. by a relative factor of $\approx 4.5 \cdot 10^{-10}$.

The second one is deduced from the ellipticity of the orbit, and may be compensated when ephemeris (i.e. orbital parameters), or satellite position and velocity vectors are known. This effect has to be compensated in the receiver; the value of the correction can be expressed by two ways – from the Keplerian parameters or directly from the instantaneous position and velocity vectors of the satellite (denoted by $\boldsymbol{r}$ and $\boldsymbol{v}$):

$$\delta t_r = \frac{-2\sqrt{\mu}}{c_0^2}\sqrt{A}e \sin E_k = -2\frac{\boldsymbol{r}\cdot\boldsymbol{v}}{c_0^2} \tag{2.115}$$

Symbols μ, A and E_k denote the value of Earth's universal gravitational parameter,[32] length of the orbits main semi-axis, eccentricity and eccentric anomaly,[33] respectively. The advantage of the latter expression is that it can be used even when the Keplerian parameters are not available, e.g. for GLONASS or precise ephemeris products, which may be distributed by other means.

Each navigation satellite usually carries multiple redundant high-performance clocks (with stability better than 10^{-13}, typically); only a few technologies were used as a GNSS satellite reference clocks: Cesium clock, RAFS (Rubidium Atomic Frequency Standard), PHM (Passive Hydrogen Maser). An extensive research effort has been put into developing, characterizing and the space clocks, for instance, a comparison of performance of various types of clocks is available in [31–33]. Although extremely precise, the clocks are not perfect, and moreover, the accuracy required for the positioning still exceeds the performance available from the raw clocks.

The SV clocks are not completely synchronous with the system master time (e.g. the GPST (GPS time) in case of GPS, GST in case of Galileo, *et cetera*), still they are required to be almost synchronous. For instance, for GPS the SV clock should be always within approximately $\pm 1\,\mathrm{ms}$ w.r.t. the system time.[34] The SV clocks are monitored by the Control segment and correction terms are uploaded to the satellites periodically. Polynomial approximation is used for describing the clock error. For example GPS uses a second-order polynomial; the 0^{th}-order coefficient corresponds to SV clock bias, 1^{st}- and 2^{nd}-order terms correspond to SV clock drift (frequency offset) and clock drift rate (frequency drift), respectively. Source [9] claims that the accuracy of the corrected clock is 5 to 10 ns, when relativistic effects are compensated and polynomial clock correction applied.

It is necessary to mention that the clock corrections broadcasted by the SV are in fact predictions. However, the IGS uses a network of monitoring stations to estimate

[31] For satellites bellow approx. 9500 km orbit radius, the frequency shift would be negative.

[32] The value to be used for the particular system is available in the corresponding ICD; for instance, GPS uses $\mu = 3.986005 \cdot 10^{14}\,\mathrm{m^3 s^{-2}}$.

[33] Those parameters are either included in satellite ephemeris, or solved from ephemeris and accurate time information.

[34] The $\pm 1-$ms range is based on information from [9], as well as on the maximal (minimal) value of the corresponding 0^{th}-order clock correction coefficient in the GPS navigation message [34].

the precise *a posteriori* satellite clock state. Actually, the comparison of the broadcast clock corrections with the precise ones provided by IGS shows that the RMS clock error is 3 to 4 ns, whereas the standard deviation (clock bias is excluded in such metric) is around 1 ns, typically [35].

Satellite hardware

Neither satellite hardware is perfect. Probably the most relevant issue is a different group delay of the individual frequencies. This delay difference is surveyed by a GNSS control segment and included in the ephemeris (e.g. GPS provides group delay differential of L1 & L2 bands in the legacy C/A code message [34]).

In general, a bias may be present between any two codes transmitted by the satellite, regardless whether they share the carrier wave frequency or not. These systematic errors are called differential code biases (DCB) and they are monitored and published by certain analysis groups [24, 36, 37]. It should be noted that the DCBs are negligible for the common GNSS applications; principally, they are irrelevant for single-frequency single-code receivers. Nonetheless, accurate DCB values are required for some high-precision navigation applications and special applications.

Satellite position and attitude

As well as the satellite clock corrections, the ephemeris broadcasted by the satellite is a prediction, which is valid over a time interval that is a few hours long (depending on the particular GNSS). Undoubtedly, there will be a disagreement between the predicted and the real satellite position, regardless the method of orbit description used. Such orbit errors may exhibit the RMS magnitude of a few meters [21]; for a common GNSS user, there is no way of correcting this error. The *a posteriori* estimated orbits, as are provided for instance by the IGS or NGA, provide an accuracy of a few centimeters [38].

Moreover, for certain high-accuracy methods it is not sufficient to take into account that the signals appear to be transmitted from a virtual point – the antenna phase center, which is generally displaced from the satellites center-of-mass (or the ephemeris reference point); the absolute displacement is inherently dependent on the satellite attitude. The data is publicly available from multiple sources, e.g. [39–41].

The phase-windup effect might be worth of interest for non-differential precise applications [42]. Due to the circular polarization of the signal, any mutual rotation of satellite and receiver antennae around the line-of-sight axis induces a change of the phase proportional to the angle of physical rotation.

Signal propagation

One would expect that the propagation of the electro-magnetic signal from a satellite to a receiver would follow straight line. This would happen only if the propagation media was homogeneous, e.g. the path from satellite to transmitter would be entirely in vacuum. In such case the signal would be delayed, Doppler-shifted and attenuated[35] at the receiver.

[35] There would be no attenuation, or losses, in the propagation media, but the intensity of the wave (power per unit surface) decreases with the square of the distance from the transmitter, since the same amount of power has to be distributed over the "inflating" spherical wavefront. See Friis transmission equation (2.1), page 7.

Indeed, the signal on its way towards the receiver encounters the atmosphere. Therefore, the trajectory is bent according to the refractive index profile of the atmosphere, since it has to obey *Fermat's principle* [12], i.e. the signal follows the path that minimizes the propagation delay. The mechanics of the propagation delay were already mentioned in Section 2.1.2; in this context, see particularly eq. (2.13).

As the signal of the GNSS has to propagate through the layers of atmosphere, it gets attenuated, delayed and distorted which induces errors in the pseudorange measurements. The ionosphere and troposphere are the strongest contributors to the overall errors, luckily, the effect of those atmospheric layers can be compensated by means of empiric models.

Troposphere Obviously, the troposphere is an inhomogeneous, lossy medium. Commonly (not always, however), its refractive index decreases with altitude. Most importantly, for our purpose, the troposphere is non-dispersive, i.e. the refractive index, and therefore the related delay, is frequency-independent.

The length of the signal trajectory through troposphere is dependent on the elevation angle, indeed. Should the atmosphere be modeled as an uniform shell of the Earth with a 6 km thickness (height above surface), the troposphere path length for near-nadir elevation could be more than 250 km. Thus, the complexity of the troposphere is more significant on the measurements from the satellites visible under low elevation angles, where the conditions may vary substantially along the signal path.

Dry and wet atmosphere is often distinguished [21]; the former includes gases, whilst the latter is formed by water vapor. The wet component is less uniform and more difficult to predict, nevertheless, it is responsible for approximately 10 % of the tropospheric effect on GNSS measurements.

To the advantage of GNSS, the attenuation by both, atmospheric gasses [43] and precipitation [44], is almost negligible in the 1 to 2 GHz band for high-elevation satellites. For the low-elevation satellites the attenuation itself is not likely to cause noteworthy problems. On the other hand, user should avoid using the nadir satellites, since most the other effects, such as refraction, ionosphere, multipath, *et cetera* are enhanced.

Naturally, the signal is delayed in the troposphere and thus, the pseudorange measurements are prolonged. The excess range varies from approximately 2.5 m at zenith to 25 m at 5° elevation [21]. There are several models i.e. Saastamoinen, Hopfield, *et cetera* [45] for estimating the total tropospheric delay w.r.t. satellite elevation and local meteorological parameters.[36] Typically, a mapping function is used to transform the zenith delays to the total slant delay (often wet and dry delays are mapped separately). Source [21] provides summary of a number of models, the widely used ones (e.g. Saastamoinen) provide centimeter-level accuracy for elevations above approximately 10°, which is sufficient for most users.

[36] The actual meteorological conditions do not have to be sensed by the receiver. It is convenient to use either standard atmosphere assumption, or averages for given latitude, local time and season.

Ionosphere In contrast to troposphere, ionosphere is dispersive, and consequently, the signals on different carrier frequencies encounter generally different ionospheric delays. Actually, the ionosphere affects the signal propagation in a fairly complex manner [46]; in addition to delay the attenuation, refraction, dispersion and Faraday rotation are involved. Fortunately, the magnitude of these effects is proportional to the total electron content (TEC)[37] integrated along the signal trajectory, and inversely proportional to the square of frequency. The TEC is strongly dependent on solar activity, i.e. it exhibits diurnal periodicity),[38] and also is dependent on the season and phase of solar cycle.

The attenuation of the ionosphere is insignificant for frequencies above 70 MHz, the dispersion-induced distortion of the ranging code (i.e. distortion induced by the group delay difference within the signal bandwidth) is not likely to become significant as well [46]. The Faraday rotation does not affect the circularly polarized GNSS signals at all [21, pp. 485–516].

The ionospheric delay is the only effect that needs to be targeted. According to [47], 99.9 % of pseudorange and carrier phase delays depend on the inverse square of the frequency. Thus, such model is adequate for most of the applications. Methods of modeling the higher-order ionospheric terms are provided e.g. in [48]. It is important to remark that while the envelope (i.e. the modulation function, the ranging code) propagates with a *group* velocity, which is slower than the nominal transmission speed, the carrier phase travels with *phase* velocity, which is faster than the nominal transmission speed.

Consequently, there is ionospheric delay affecting the pseudorange measurements and ionospheric advance affecting the carrier phase measurements.According to [47, 48], the magnitude of these two terms is the same when higher-order ionospheric terms are neglected; the delays are then equal to [9, 21, 22]

$$\tau_\rho^I = -\tau_c^I = -40.309\,\frac{\mathrm{STEC}}{f^2}, \tag{2.116}$$

where τ_ρ^I and τ_c^I denote the pseudorange and carrier additional delays that are caused by ionosphere.

Due to the strong dependency on the inverse frequency square, it is possible to almost mitigate the ionospheric delay by means of a special combination of two measurements performed on different carrier frequencies. The ionosphere–free (or iono–free) combination is a simple linear combination of the measurement on two frequencies, which has the coefficients carefully chosen. For both pseudorange

[37] TEC is an integral of number of electron within $1\,\mathrm{m}^2$ cross-section column along the path. Commonly, VTEC (vertical TEC) or only TEC is used for the electron content along the zenith path, STEC (slanted TEC) is used to denote the electron content along the path with arbitrary elevation angle. It may be expressed in the units of electron count per unit surface [$\mathrm{e}^-/\mathrm{m}^2$] or TECU (TEC unit); $1\,\mathrm{TECU} = 10^{16}\,\mathrm{e}^-/\mathrm{m}^2$.

[38] When extremely simplified, TEC is higher during day and lower at night.

and carrier phase measured on frequencies f_1 and f_2 it yields:

$$\rho_{\text{iono}-\text{free}} = \frac{f_1^2 \rho_{f1} - f_2^2 \rho_{f2}}{f_1^2 - f_2^2} \tag{2.117}$$

$$\phi_{\text{iono}-\text{free}} = \frac{f_1^2 \phi_{f1} - f_2^2 \phi_{f2}}{f_1^2 - f_2^2}. \tag{2.118}$$

Naturally, the combination "virtual" carrier frequency is higher, so the combination "virtual" wavelength shortens. Moreover, the random errors that are independent on the two frequencies are scaled and accumulated

$$\sigma_{\rho,\,\text{iono}-\text{free}} = \sqrt{\left(\frac{f_1^2 \sigma_{\rho,\,f1}}{f_1^2 - f_2^2}\right)^2 + \left(\frac{f_2^2 \sigma_{\rho,\,f2}}{f_1^2 - f_2^2}\right)^2} \tag{2.119}$$

For instance, if f_1 and f_2 are chosen to be the L1 and L2 frequencies of the GPS, under assumption of the same pseudorange noise standard deviation on L1 and L2, the iono–free combination deviation is 3-times higher than the one of the individual pseudoranges.[39]

The single-frequency GNSS receivers (almost all non-professional ones) have to use a model to correct the pseudorange for the ionospheric delay. The models evolved throughout the years of operation and further development of the GNSS. Naturally, there is a trade-off between the accuracy of the model and amount of data to be transmitted from space vehicles to the user. For instance, GPS exploits Klobuchar model [21], which uses a parametrized cosine model of the diurnal TEC development.[40] More sophisticated and data-throughput-demanding models are used in the newer systems, such as NeQuick2 model for Galileo. The space-based augmentation systems (SBAS) deliver more accurate ionosphere modeling, e.g. TEC grids. Generally, all the models have to be provided with approximate user and satellite positions, in order to estimate the trajectory through the ionosphere, or the IPP (ionospheric pierce point) and elevation if the model simplifies the ionosphere to a thin layer.

Still, it is barely possible to model and predict the ionosphere behavior during the periods of rapid ionospheric activity – e.g. magnetic storms. There are certain areas, where irregularities in the ionosphere appear more likely. These are especially two sub-equatorial bands (commonly during the evening in local time), where the irregularities cause amplitude scintillations (more than 20 dB was observed) and phase scintillations (rapid change of phase that may cause loss-of-lock of the PLL on receivers with narrow bandwidth)[21].

[39] Naturally, the iono-free combination is not the only possible linear combination of the measurements, in addition, three (or more) signals may be combined. For instance, [49–51] thoroughly investigate on this subject.

[40] According to the ICD [34], the Klobuchar model requires only eight 8-bit parameters to be delivered to the user via satellite links and still is claimed to reduce the ionospheric delay by more than 50 % (RMS) around the globe. The model also simplifies the ionosphere to a single layer at 350 km height.

Multipath When the electromagnetic wave encounters an ideally conductive planar surface in a perpendicular direction, it reflects and changes its polarization to the opposite one (i.e. inverts polarity when linearly polarized, switches the sense of rotation when circularly polarized). If the conductor is lossy, the reflection is attenuated accordingly.

On a planar interface of two dielectrics the wave is partially reflected and partially transmitted further into the second dielectric. For oblique incidences of the wave the transmission and reflection follows Snel's laws [12]. These effects can be generally described by the transmission and reflection coefficients, which may be further interpreted as transmission loss or reflection loss.[41] It is important to note that there is no actual power loss at the media interface, the power of the incident wave is divided between the transmitted and reflected waves. However, the wave may be attenuated in the lossy media.

As a consequence, all objects surrounding the GNSS receiver (ground, mountains, buildings, vehicles, *et cetera*) reflect the signals to the receiver antenna; some of the reflections are stronger (e.g. from metallic objects), some weaker (e.g. from non-conductive objects). The reflected signals are denoted as the multipath signals and are generally considered nuisance in radionavigation. From the signal-power perspective, the multipath is often observed as a fast fading process, especially for moving receivers.

The direct signal always arrives first, since the paths of the reflected signals are always longer and thus the signal is more delayed. Nonetheless, all the signals are summed physically at the antenna and the signal and its correlation with the ideal replica becomes distorted. Should the relative delay of the multipath signal be short enough to affect the main peak of the correlation function, the pseudorange measurement becomes biased. The polarity of the multipath-induced bias depends on the polarity of the disturbing multipath signal at the antenna, i.e. constructive combination (same polarity) of the signals lengthens the pseudorange reading, destructive combination (opposite polarity) shortens the pseudorange reading.

The reflections that do not affect the main lobe do not cause problems in the signal tracking, however may be harmful for the acquisition process. Nonetheless, severity of the multipath may become higher, when the direct path is heavily attenuated or completely unavailable – in such case the first reflected signal may be interpreted by the receiver as the direct one. Such misinterpretation will certainly cause a significant systematic error in the pseudorange value.

The multipath effects may be limited by several precautions or processing options. First, the shape of signal autocorrelation function, especially the "sharpness" of the main lobe is the determining factor of envelope of the possible multipath-error values – the narrow main lobe reduces the maximal reflected

[41] The coefficients are the ratio of the electric field amplitude of the outcoming waves w.r.t. the incoming wave. It is negative when the polarity of wave is inversed and its absolute value cannot be higher than unity. The loss is ratio of powers, usually expressed in decibels, therefore the polarity information is lost in such interpretation.

signal delay that may affect the tracking. For instance, the envelopes for GPS C/A signal are available in [21, chap. 14] Some components may be successfully reduced by the antenna characteristics. For instance, antenna with high cross-polarization discrimination may be used. Such device will naturally suppress the signals reflected with the opposite polarization from the conductive structures.[42] Moreover, reflections may be reduced by the gain pattern of the antenna, since most of the multipath signals exhibit low elevation angle of incidence (e.g. ground or vehicle reflections). Narrow-spacing ranging-code correlator [9] or an advanced architecture of the ranging code tracking loops, i.e. high resolution correlator [52] or MEDLL (multipath-estimating DLL) [53–55] help to reduce the multipath-related ranging error.

It is known that signals from low-elevation satellites tend to be affected by multipath, since for instance the path-length difference of ground-reflected signal is usually small, reflections even from relatively low surrounding objects are likely. Therefore, elevation masks may be applied, i.e. satellites with elevation lower than certain limit are still tracked, but excluded from the position solution.[43] If a single multipath-affected measurement is used in an over-determined (more than four pseudoranges are available) position estimation solution, it may exhibit excessively high residual that may indicate a multipath-related problem.

The signal propagation is also affected by vegetation canopy. The representative data such as probabilistic description of amplitude fading w.r.t. elevation angle is provided in [21, chap. 15]. Nonetheless, the severity of the foliage-related fading depends heavily on the kind of vegetation canopy (tree species, height, density of trees) and season. In some cases, the GNSS positioning service may become disrupted or completely unavailable due to the attenuation.

Receiver

The receiver contributes to the overall error of the PVT estimate through multiple effects.

First, the signal is received by a non-ideal antenna. The signals coming from different directions are summed in the phase center. As well as in the satellite case, its position is frequency and incident angle dependent; nonetheless, for standard application it can be neglected, since the variations are not critical at the higher elevation angles.

Then, the signal propagates through a cable to the receiver itself and consequently, the signal is delayed and attenuated. It is necessary to note that the cable delay has to be compensated in timing applications, because the GNSS positioning algorithms solve for the PVT at the point of signal summation.[44]

The receiver itself is not perfect, it may suffer from inter-frequency biases or DCBs. Both can be either measured and at least partially compensated by the manufacturer within the receiver software, or they can be estimated externally [56].

[42] Naturally, the signal may be reflected more than once. Since the sense of the circular polarization changes with each reflection, the multipath signals with even reflection count are not discriminated.

[43] Elevation masking is common especially for precise GNSS measurements, e.g. GNSS reference stations or geodetic surveys.

[44] Since the cable-delay is common to all signals, it appears to be a part of the bias between the system and receiver clock. Thus, position and velocity estimates remain unaffected.

The ranging accuracy depends mostly on the performance of the code and phase tracking loops, the DLL and PLL. There are several demands on the loops that are conflicting. For instance, the immunity to noise can be improved by narrowing the receiver bandwidth. On the other hand, narrow loop bandwidth is impractical for high dynamic applications and requires more precise acquisition of signals. A number of architectures of the loops does exist, moreover, there is a plenty of parameters that can be tuned. Detailed description of the actual receiver tracking hardware and software solutions and their behavior is beyond the scope of the thesis; a deep insight can be found in several books, e.g. [9, 10, 21, 22, 57, 58].

Generally, in common applications the receiver should not contribute to the overall error by more than 20 cm [21]. Professional, high-end, receivers promise precision of pseudorange measurements below 10 cm (RMS) and carrier phase below 1 mm (RMS), see e.g. [59, 60].

III. Fundamental Limitations

It is obvious that several conditions have to be met in order to allow GNSS to provide a correct PVT estimate. The most critical is the requirement on direct line-of-sight availability from satellites to receiver. Even if the line-of-sight is unavailable, the signal may be received via a reflected path – however, such pseudorange (or carrier phase) measurement is biased and may corrupt the position estimate. This implies that user should expect compromised availability and accuracy in environment with frequent line-of-sight obstructions, such as urban canyons, narrow valleys, forests, indoors, *et cetera*.

The power of the GNSS signal at the vicinity of Earth's surface is comparable to the noise power level. Not only the GNSS signals may become unavailable even when attenuated slightly (e.g. cellular communication or television and radio broadcasting would deliver their service in sufficient quality in the same conditions), but their low power-levels it render them vulnerable to jamming. For the user experience it does not matter, whether the jamming is intentional or accidental, in both cases relatively low-power jammer is able to disrupt the GNSS service. Even a strong geomagnetic storm may hinder successful positioning. It is also possible to spoof the receiver, i.e. provide fake signals to the receiver and mislead the PVT estimate [61].

Although numerous countermeasures to jamming and spoofing do exist, they are not always implemented in the commercial receivers. And even though the receiver technology has been certainly improved since the early years of GPS, the physical principles and the implied limits remain the same. The GNSS indisputably became the ultimate tool for navigation and timing in a global frame. Nonetheless, it is not flawless and therefore it cannot be relied on carelessly.

2.2.2. GNSS positioning methods

As multiple GNSS observables with different qualities and hindrances may be available, several methods with different characteristics of the PVT outputs can be used. Below, four kinds methods are summarized.

I. Standard code-based positioning

First, we will focus on the basic, stand-alone, code-based (or pseudorange-based) positioning. This method is the simplest one and does not require any information from external sources, i.e. the GNSS navigation message broadcast provides all necessary parameters.

For the sake of simplicity, we will consider that for M satellites in view we have M pseudorange measurements available.[45] On the first look, the positioning is very straightforward. The position of all the satellites (denoted by $\boldsymbol{r}_\mathrm{i}$) is known from the navigation message, pseudorange correction models are known as well. Thus, it is possible to obtain user position $\boldsymbol{r}_\mathbf{u}$ and receiver clock bias b estimates by solving a set of pseudorange equations (2.59), which were derived in Section 2.1.2–III.

$$\rho_\mathrm{i} = \|\boldsymbol{r}_\mathrm{u} - \boldsymbol{r}_\mathrm{i}\| + b, \tag{2.59}$$

Such set of nonlinear equations may be easily solved by one of the iterative methods presented in Chapter 4.

The reality, however, is not that simple. The pseudorange measurements are taken at exactly the same moment, and referenced to the receiver time – which should be coarsely aligned with the system/satellite time by means of the navigation message. It is obvious that the receiver position estimate will be valid for the moment of pseudorange measurement.

By the same principle, the satellite position at the time of transmission is the determining one for the positioning. Nonetheless, the propagation delay of the signal may be substantially different for each of the satellites, and consequently, the time of transmission of the signals is different.[46] The propagation delay is proportional to the true range between the user and the satellite – so the position of the user is required in order to compute it. The same problem occurs with the application of the tropospheric and ionospheric models, since they are both user-position-dependent.

In order to overcome this issue, the delay is guessed, as well as the corrections to the pseudorange from the models. The set of pseudorange equations is solved, and rather inaccurate estimate of user position is obtained. Based on this inaccurate estimate it is possible to determine the propagation delays, as well as the pseudorange corrections, and then solve the set of equations again in order to obtain the position estimate with better accuracy. Typically, only a few iterations are enough to reach the final accuracy, which is determined by the errors described in the previous section.

By such approach a single-point solution may be obtained; naturally, the result can be used as the initial guess for the next measurement epoch. It may be convenient to exploit the dynamic properties of the positioned platform, i.e. to provide some connection between the measurement epochs. This can be effectively done by using a

[45] Such condition is typical not only to the single-frequency GNSS receivers that are usually found in consumer-grade electronics (e.g. smartphones, car navigation, sport trackers, *et cetera*). For example, professional dual-frequency receiver may use ionosphere-free combination, which results in a single pseudorange per satellite as well.

[46] The propagation delay from MEO GNSS satellites varies approximately from 64 to 100 ms, depending on the orbit radius and elevation of the satellite w.r.t. user. The shortest propagation delay for a particular satellite is in user's zenith.

Kalman filter for position estimate. Due to the nonlinear nature of the pseudorange equations, an extended or unscented Kalman filter (EKF or UKF) must be used. It is obvious that in the three-dimensional space the Kalman filter state vector will consist of four elements (or more).

Use of Doppler-shift measurement

It has been already mentioned that the Doppler shift measurement is proportional to the pseudorange rate, i.e. the radial component of the relative velocity of the satellite and user. Thus, it is possible to extend the set of equations by the equations (2.63) for the pseudorange rate and estimate receiver velocity $\boldsymbol{v}_\mathrm{u}$ and clock bias $\dot{b}$ rate as well:

$$\dot{\rho}_\mathrm{i} = \mathbb{1}_{\mathrm{u,i}} \cdot (\boldsymbol{v}_\mathrm{u} - \boldsymbol{v}_\mathrm{i}) + \dot{b}. \tag{2.63}$$

Note that the satellite velocity $\boldsymbol{v}_\mathrm{i}$ can be evaluated for the time of transmission using the information included in the ephemeris. It is necessary to estimate the velocity jointly with the position, since the unitary direction vector cannot be evaluated without such information, see (2.61 on page 21).

The number of variables in the resulting equation set is doubled, however, the number of equations (i.e. measurements) is doubled as well. Therefore, at least four satellites are required for a successful PVT estimation. Should the Kalman filter be used for the solution, it has to operate with the state vector of length 8 or more – i.e. at least 3 position coordinates, 3 velocity coordinates, clock bias and drift.

Carrier smoothing

Since the phase measurements are one or two orders of magnitude less noisy than the pseudorange measurement, it is favorable to exploit them in the basic, stand-alone positioning as well. The carrier phase is inherently ambiguous, so it cannot be used for positioning in a direct way.

The carrier phase (φ, expressed in radians) can be easily recomputed to the range scale, i.e. it can be expressed in meters by multiplying by factor $\lambda_\mathrm{c}/(2\pi)$, where λ_c is the carrier wavelength. Let us denote this scaled value as accumulated delta-range. As long as the phase measurement is continuous, the ambiguity of the carrier phase measurement is constant. Thus, for the periods of continuous measurements it can be stated that the first derivative of the pseudorange is almost equal to the first derivative of the accumulated delta-range. For the pseudorange and carrier phase measurements taken in epochs i and $i-1$ it holds

$$\rho_\mathrm{i} - \rho_{\mathrm{i}-1} \approx \frac{\lambda_\mathrm{c}}{2\pi} \left(\varphi_\mathrm{i} - \varphi_{\mathrm{i}-1}\right), \tag{2.120}$$

provided the continuity criterion is satisfied.

It is possible to combine the unambiguous nature of pseudorange and smoothness of the pseudorange by a method called carrier smoothing [9]. The pseudorange in the epoch i can be computed on the basis of previous pseudorange measurement and the carrier phase difference:

$$\rho_\mathrm{i}^{(\varphi)} = \rho_{\mathrm{i}-1} + \frac{\lambda_\mathrm{c}}{2\pi} \left(\varphi_\mathrm{i} - \varphi_{\mathrm{i}-1}\right). \tag{2.121}$$

Similarly, epochs $i-2$, $i-3$, *et cetera* may be used as well. By means of combining multiple computed pseudorange measurements and the actual pseudorange measurement the smoothed value may be obtained. In [9] a recursive filter implementation is presented:

$$\rho_{\mathrm{i}}^{(\varphi)} = \frac{1}{K}\rho_{\mathrm{i}-1} + \frac{K-1}{K}\left(\rho_{\mathrm{i}-1}^{(\varphi)} + \frac{\lambda_{\mathrm{c}}}{2\pi}\left(\varphi_{\mathrm{i}} - \varphi_{\mathrm{i}-1}\right)\right). \tag{2.122}$$

The K coefficient sets the weight of the phase information, i.e. the length of the carrier-smoothing window; the algorithm is initialized with the raw pseudorange measurement.

It is necessary to remark that carrier smoothing does not suppress any systematic errors, it only reduces the noise and may partially suppress multipath errors. Moreover, for longer smoothing windows the pseudorange and carrier measurements may diverge, due to changing ionospheric conditions.[47] Such phenomena is often called code-carrier divergence.

Dilution of precision

Naturally, the accuracy of the final solutions depends on the shape of the visible satellite constellation with respect to the receiver. The receiver-satellite geometry is captured in the geometry matrix $\boldsymbol{G}$, which is the Jacobian matrix of the set of pseudorange equations (2.63). In the 3D-space, geometry matrix has 4 columns and number of rows equal to the number of pseudorange observations in the particular epoch. The first three columns are formed by the transposed unit vectors pointing from a satellite towards the receiver, the fourth column is all ones:

$$\boldsymbol{G} = \begin{bmatrix} \mathbb{1}_{\mathrm{u},1}^{\mathrm{T}} & 1 \\ \mathbb{1}_{\mathrm{u},2}^{\mathrm{T}} & 1 \\ \vdots & \vdots \\ \mathbb{1}_{\mathrm{u,M}}^{\mathrm{T}} & 1 \end{bmatrix} \tag{2.123}$$

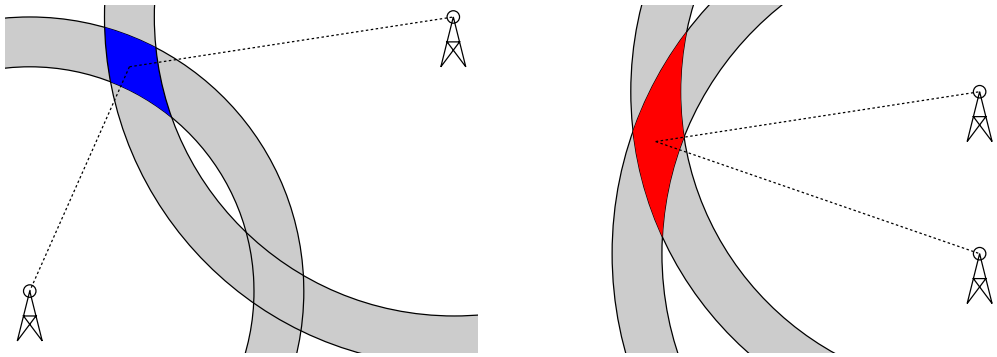
Then, the matrix $\boldsymbol{H}$ is defined as

$$\boldsymbol{H} = \left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\right)^{-1}, \tag{2.124}$$

which is a four-by-four, symmetric matrix – the DOP (dilution of precision) matrix. It can be shown that its diagonal elements describe how the pseudorange measurement variance affects the positioning accuracy in a particular direction, i.e. in the direction of three principal axes and in the time domain. Should the principal axes coincide with a local Cartesian frame, e.g. east-north-up, then the square roots of the diagonal elements are dilution of precision in the east, north, and up directions.

The principle can be described using a graphical representation of DOP in two dimensions, see Fig. 2.12. The circular bands depict the ranging measurements with non-ideal accuracy; the localized point lies certainly in the intersection of the two bands. In the (a) subfigure, the intersection is sharp, the DOP is low in all directions.

[47] The sign of the ionospheric delay is opposite for the pseudorange and carrier phase measurements. As long as it remains constant, it is canceled out. However, its changes accumulate in the opposite way in the two measurements.



(A) Low DOP (B) High DOP

FIGURE 2.12.: Graphical representation of DOP in two dimensions.

On the contrary, the intersection area is larger in the (b) subfigure – the DOP in the elongated direction as well as the PDOP is higher.[48]

Moreover, since the directions and time domain are orthogonal, we may obtain DOP in horizontal (or other) plane, in space by applying the Pythagorean theorem on the diagonal elements. The particular combinations to obtain the DOP coefficient in the local coordinate frame are summarized in Tab. 2.4.

The DOP coefficients should be interpreted as scaling factors of the ranging standard deviation. Thus, assuming the pseudorange standard deviation σ_ρ the total 3D position error would be

$$\sigma_{3\mathrm{D}} = \mathrm{PDOP} \cdot \sigma_\rho. \tag{2.125}$$

Lower values of DOP are desired; therefore, the satellites should be distributed evenly in the sky. By a rule of thumb the VDOP is 50 to 100 % higher than HDOP due to the nature of the constellations (there are no satellites below horizon). It is important to remark that introducing additional satellite never leads to higher DOP value.

II. Differential GNSS

A considerable amount of GNSS positioning error is related to the non-ideal propagation of the signals and ephemeris errors. However, receivers that are close to each other encounter rather similar errors, since the signal trajectories from satellite to both receivers are almost the same. The satellite-related errors are highly correlated, almost identical; the propagation delay errors decorrelate with the relative distance of the receivers – further, we will denote the vector between the receivers as baseline, and the distance as baseline length. Anyway, the ionospheric and tropospheric conditions

[48] An inverse, but principally similar, phenomena can be found in the following example: consider a ball attached firmly in the middle of a long string. The ball can be easily moved perpendicular to the string – the two halves of the string both stretch by a small fraction of the actual ball displacement magnitude. In contrast, moving the ball in the direction of the string is harder, since one of the halves of string has to stretch and the other has to shrink by the same amount – by the whole magnitude of displacement.

Table 2.4.: DOP coefficients in East-North-Up coordinate frame.

Label	Meaning	Formula
EDOP	East	$\sqrt{\boldsymbol{H}_{11}}$
NDOP	North	$\sqrt{\boldsymbol{H}_{22}}$
VDOP	Vertical	$\sqrt{\boldsymbol{H}_{33}}$
TDOP	Time	$\sqrt{\boldsymbol{H}_{44}}$
HDOP	Horizontal	$\sqrt{\boldsymbol{H}_{11} + \boldsymbol{H}_{22}}$
PDOP	Position (3D)	$\sqrt{\boldsymbol{H}_{11} + \boldsymbol{H}_{22} + \boldsymbol{H}_{33}}$
GDOP	General	$\sqrt{\boldsymbol{H}_{11} + \boldsymbol{H}_{22} + \boldsymbol{H}_{33} + \boldsymbol{H}_{44}}$

are usually changing in a smooth manner, so the applicability of assumption of identical errors may be valid over tens or a few hundreds of kilometers of baseline length, depending on the tropospheric and ionospheric conditions. Note that the multipath effects are related to the immediate surroundings of the receivers and consequently are uncorrelated.

Local area augmentation services (LAAS)

Consider a scenario with two receivers. The position of the first one is known, we will denote it as the reference station; the second one, the rover, needs to be localized. The two receivers are placed close enough to have the pseudorange errors similar. Clearly, the reference receiver should be able to determine the errors of the measured pseudoranges, since the true range to the visible satellites is available to it. In [9] two approaches of DGPS are described, nonetheless, the basic principles are valid for all GNSS.

In the first one, the DGNSS corrections are computed in the reference station by means of *differencing the pseudorange* measurements and the respective *true ranges*. The difference e_r at the reference station (denoted by the r-subscript) follows

$$\begin{aligned} e_\mathrm{r} &= \|\boldsymbol{r}_\mathrm{r} - \boldsymbol{r}_\mathrm{s}\| - \rho_\mathrm{r} \\ &= -b_\mathrm{r} - \epsilon_\mathrm{r}^{Sat} - \delta\rho_\mathrm{r}^{Iono} - \delta\rho_\mathrm{r}^{Tropo} - \epsilon_\mathrm{r}^{Rx}, \end{aligned} \tag{2.126}$$

where b_r is the station clock bias, $\delta\rho$ denotes the ionospheric and tropospheric delays, ϵ denotes the errors caused by the satellite (as observed by the station) and receiver (see 2.2.2–II for more information about the errors). A similar equation can be written for the rover, only the subscript would be different – further we will use u-subscript.

The correction is delivered to the rover (user) and the receiver corrects its pseudorange measurement by adding the reference stations correction. If the receivers are not too far[49] and the correction is not too old, we may assume

$$\epsilon_\mathrm{r}^{Sat} = \epsilon_\mathrm{u}^{Sat} \qquad \delta\rho_\mathrm{r}^{Iono} = \delta\rho_\mathrm{u}^{Iono} \qquad \delta\rho_\mathrm{r}^{Tropo} = \delta\rho_\mathrm{r}^{Tropo}, \tag{2.127}$$

[49] This statement is certainly quite vague, no precise limit of correction applicability exists. As a rule of thumb 100 km baseline length can be considered as a limit.

and consequently the corrected pseudorange is

$$\begin{aligned}\tilde{\rho}_{\mathrm{u}} &= \rho_{\mathrm{u}} + e_{\mathrm{r}} \\ &= \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{s}}\| + b_{\mathrm{u}} - b_{\mathrm{r}} + \epsilon_{\mathrm{u}}^{Rx} - \epsilon_{\mathrm{r}}^{Rx}.\end{aligned} \tag{2.128}$$

Naturally, the assumptions above become less valid with growing baseline length and age of corrections.

It is worth noting that in this approach, differential clock bias $(b_{\mathrm{u}} - b_{\mathrm{r}})$ is estimated alongside the rover position. However, the reference station may solve for its bias and compensate for it in the correction data. Also, the receiver-related errors differentiate; such behavior may be useful when identical receivers are used as reference station and rover, but may cause errors when different receivers are used. Regardless the choice of receivers, the random errors are inherently independent and therefore their variances sum up. It is also convenient to choose the reference station location so that the multipath effects are minimal.

The second approach is called *relative positioning*, and is based on differencing time-matched pseudorange measurements:

$$\rho_{\mathrm{ru}} = \rho_{\mathrm{r}} - \rho_{\mathrm{u}}. \tag{2.129}$$

Again, under the assumptions (2.127) it can be written that

$$\rho_{\mathrm{ru}} = \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{s}}\| - \|\boldsymbol{r}_{\mathrm{r}} - \boldsymbol{r}_{\mathrm{s}}\| + (b_{\mathrm{r}} - b_{\mathrm{u}})\, \epsilon_{\mathrm{r}}^{Rx} - \epsilon_{\mathrm{u}}^{Rx}. \tag{2.130}$$

The reference station position is known, and obviously, the rover position vector r_{u} and joint bias $(b_{\mathrm{r}} - b_{\mathrm{u}})$ have to be estimated.

The advantage of the relative positioning is that a modified version of carrier smoothing may be applied to the differenced pseudorange ρ_{ru}. In such case, the difference of phase measurement $\varphi_{\mathrm{ru}} = \varphi_{\mathrm{r}} - \varphi_{\mathrm{u}}$ is exploited, however, the smoothing filter remains the same as in stand-alone positioning, see (2.122). Favorably, due to the phase differencing, the possibility of code-carrier divergence is mitigated, since it would affect rover and reference station similarly. Naturally, both receivers have to observe the carrier phase continuously. Source [9] claims approximately meter-level accuracy for standard DGPS positioning and accuracy of a few decimeters for the carrier smoothing with the averaging length of 15 minutes or more.

Wide area augmentation services (WAAS)

For the broad areas, the approach described above is not feasible, since it requires quite dense network of reference stations.[50] Moreover, switching from one reference station to another is not possible without a loss of measurement continuity. In such case a whole area of coverage is monitored and models of the ephemeris and signal propagation errors for the whole area of coverage are provided to the users [9, 64].

Typically, the corrective information is provided via one or more geostationary satellites – therefore, the systems are often called SBAS (Satellite-based augmentation systems). There are several operational systems, for example:

[50] Such system was working in the U.S as Nationwide DGPS. However, it has been downsized in 2016 [62], and will be phased out between 2018 and 2020 [63].

WAAS(Wide-Area Augmentation System) operated by the U.S. Federal Aviation Administration (FAA), covering North-American continent, exploiting 3 geostationary satellites [65];

EGNOS(European Geostationary Navigation Overlay Service) operated by the European GNSS Agency (GSA), covering European continent, exploiting 3 geostationary satellites (only 2 operational) [66];

MSAS(Multi-functional Satellite Augmentation System) operated by the Japanese authorities, covering Japan islands and the surrounding ocean, exploiting 2 geostationary satellites [67];

GAGAN(GPS-Aided GEO Augmented Navigation) operated by the Airports Authority of India (AAI), covering Indian subcontinent, exploiting 3 geostationary satellites (only 2 operational) [68].

Nonetheless, other countries (e.g. Russia [69], China [64], South Korea [70]) are in development of their own systems.

According to experience obtained in the Integrated Satellite and Terrestrial Navigation Technologies Centre of the Faculty of Electrical engineering, the SBAS positioning performance is somehow comparable with DGNSS. In the higher latitude, the geostationary SBAS satellites appear on the south with rather low elevation angles. Therefore the augmentation service is often unavailable even for long periods of time, or the continuity of reception is compromised (especially for moving users).

III. Real-time kinematics

RTK is a differential GNSS positioning technique that fully exploits the *carrier-phase* measurement accuracy. Hence, it achieves position accuracy of a few centimeters for kinematic applications or even less for static measurements. It requires a data link between the reference station and rover, preferably with low latency. Since the full set of observables is transfered from the reference station to the rover, the required data throughput scales with the measurement rate, number of satellites and number of carrier frequencies received.

The implementation of RTK may be performed in several ways, but the most recognized one uses so-called double differences of the carrier-phase measurements, which are descibed by means of (2.131) and Figure 2.13.

$$\varphi_{\mathrm{ij,ur}} = \varphi_{\mathrm{i,ur}} - \varphi_{\mathrm{j,ur}} = (\varphi_{\mathrm{i,u}} - \varphi_{\mathrm{i,r}}) - (\varphi_{\mathrm{j,u}} - \varphi_{\mathrm{j,r}}) \tag{2.131}$$

The first difference is the same as in the relative positioning variant of DGNSS, i.e. between the rover and reference station. The second difference is performed between two satellites; such approach mitigates the internal biases of the receivers, since they are common to all the received satellite signals in the same frequency band.

The description of RTK is available in many GNSS-related books, e.g. [9, 45, 71]. Another source of information can be the manual of an open-source implementation of RTK positioning, *RTKlib* [72].

When the rover is not far from the reference station, only multipath and almost negligible residuals of the other errors affect the double-differenced phase measurement, moreover, the multipath impacts the carrier phase in far less severe manner than in the

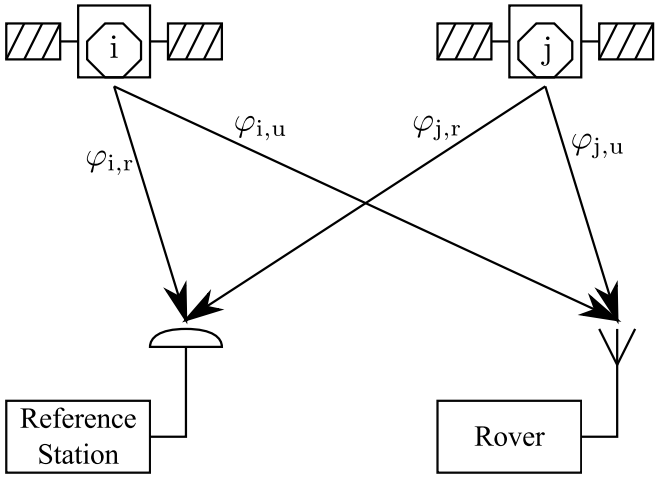


FIGURE 2.13.: Measurement schematic for RTK double differences.

case of pseudorange. Notwithstanding, the carrier phase ambiguity has to be resolved in order to perform the precise positioning. The integer nature of the carrier-phase ambiguity is preserved in the double differentiation process.[51] It is necessary to mention that the ambiguities cannot be resolved using the phase data measurements from a single measurement epoch. For each double-difference there is a unique ambiguity variable, and the baseline vector has to be estimated as well. Thus, the pseudorange measurements or data from multiple epochs have to be used.

In general, there is a grid of points in the space where each point corresponds to a certain combination of the ambiguities (and the fractional phase measurements, of course). It is important that it is a discrete space, since the ambiguities are discrete values, indeed. The positioning algorithm has to find the right point and get its position. Such search is certainly not trivial even for single-frequency case, since there are typically 4 or more satellites in view, and the wavelength is at the decimeter level.[52] Therefore, the *ambiguity resolution* (AR) is performed in two steps.

First, the ambiguities are assumed to be real numbers, i.e. continuous values. The set of double-differenced equations is solved for position and ambiguities in a least-squares manner, the initial guess may be obtained either from standalone or DGNSS solution. The solution with the real ambiguities is often denoted as the *float* solution, its accuracy is often similar to the DGPS one.

Second, the solution with integer ambiguities, the *fix* solution, has to be found. A rather straightforward way is to simply round the float ambiguities, however, it appears to be a non-optimal approach [74]. Another techniques of AR are for instance

[51] Sometimes, triple-differences technique is also used. The third difference is performed in time, i.e. between two consecutive epochs. This effectively cancels out the ambiguity terms (actually in the same manner as the carrier smoothing does). On the other hand, due to the time-differencing the integer constraint on the ambiguity values is relaxed, which results in the substantial loss of final solution accuracy.

[52] The number of grid points and volume of the search space is investigated in [73].

the ambiguity bootstrapping [75],[53] integer least-squares, *et cetera*; see [45]. Probably the most recognized method is LAMBDA (Least-squares ambiguity decorrelation adjustment) [76] which extends the integer least-squares, and searches the vicinity of the float solution for the best candidate point. In addition, the generally correlated ambiguities are transformed to another space, in order to decorrelate them. The solution is found in a more effective way when the values are less correlated, the two best candidate points are then transformed back to the "correlated space." The position estimate is obtained for all of these points and the final solution is verified by comparing the residuals of the two best candidate solutions.

The RTK solution is extremely precise. Typically, an error not larger than a few centimeters can be expected in kinematic modes, even less is achievable in the static modes. Nonetheless, the technique is inherently sensitive to the continuity of the phase measurements at both, rover and reference station receivers – the ambiguity remains constant between the discontinuities of the measurements.[54]

Since RTK is a differential technique, its accuracy and more importantly the ability to fix the ambiguities deteriorates with growing baseline length. The long-baseline RTK versions do exist (baselines longer than $\approx 100\,\mathrm{km}$), certain error modeling has to be applied in order to respect the different conditions at the ends of the baseline vector. Typically, the baseline accuracy is 1 to 10 ppm, i.e. 1 to 10 mm per each kilometer of baseline. The single baseline RTK is able to fix the ambiguities within a few tens of kilometers from the reference station. There are a few important measures of the final positioning quality. For instance, the ability to fix the ambiguity often expressed as the time to first ambiguity resolution, or the ambiguity-fix ratio. Also, the positioning accuracy is evaluated separately for the fix and float solutions, since they are typically different by at least one order of magnitude.

Of course, the same problem as in local version of DGNSS is faced here – a dense network of reference stations is required. For this purpose, network RTK concept has been introduced. There are several NRTK implementations; for instance, it has been proposed to provide corrections to a single reference station data, which are based on interpolated network measurements and approximate rover position [77]. Furthermore, VRS (virtual reference station) concept has been developed; based on the approximate rover location and reference station network observation the network software computes the observations as if they were measured at the approximate rover location [78]. The baseline (to the VRS) can be very short. Such approach does not require rover to perform the network computation itself. Not only the NRTK improves the available accuracy in a wider area, it also reduces the time required to fix the ambiguities [79]. The VRS-concept is not the only variant of the NRTK; the others are briefly described and compared in [80].

Still, the use of NRTK does not diminish the need for continuous carrier-phase observations. Regardless RTK or NRTK is used for survey, it is challenged by ionospheric

[53] The integer bootstrapping is a rounding that respects the covariances of the rounded variables.

[54] By such discontinuities we mean signal loss or receivers PLL cycle slip (i.e. the phase reading jumps across one or multiple phase cycles, however, the fractional part remains the same). Such discontinuities may not be limited to a single measurement epoch.

conditions – reduced ability to resolve the ambiguities and lower quality of measurements may be expected during periods of high ionospheric activity.

To sum up, RTK is able to provide excellent accuracy under favorable conditions. On the other hand, it is rather vulnerable to measurement discontinuities and its performance (time to ambiguity resolution, ambiguity-fix ratio, accuracy) deteriorates with the distance from the base station (unless NRTK is used) and instability of ionospheric conditions.

IV. Precise point positioning

Unlike the RTK, the precise point positioning (PPP) is not a differential technique, however, it aims on centimeter-to-decimeter level accuracy. Thereupon, all the errors, which were suppressed by differencing across receivers and satellites in RTK, have to be modeled and corrected by other means, or estimated during the observation. The errors have been summarized already in Section 2.2.2–II, Table 2.5 summarizes the methods of mitigation and residuals of the errors in PPP. The exact description of the various PPP techniques is beyond the scope of the thesis, however, we will provide a brief summary of its typical characteristics.

Table 2.5.: Error magnitudes in typical PPP processing (according to [81]).

Effect	**Typical magnitude**	**Mitigation method**	**Residual error magnitude**
Ionosphere	$< 100\,\mathrm{m}$	linear combination	mm
Troposphere	$< 10\,\mathrm{m}$	modeling / estimation	mm – dm
Relativistic	$\approx 10\,\mathrm{m}$	modeling	mm
Satellite antenna	$< 1\,\mathrm{m}$	modeling	mm
Code multipath, noise	$\approx 1\,\mathrm{m}$	filtering	dm – mm
Solid Earth tides	$\approx 20\,\mathrm{cm}$	modeling	mm
Phase wind-up	$\approx 10\,\mathrm{cm}$	modeling	mm
Ocean loading	$\approx 5\,\mathrm{cm}$	modeling	mm
Orbits and clocks	$\approx 20\,\mathrm{cm}$	modeling/filtering	cm – mm
Phase multipath	$\approx 1\,\mathrm{cm}$	filtering	cm – mm
Receiver antenna	$< 1\,\mathrm{cm}$	modeling	mm

The PPP is often performed in post-processing, since the precise ephemeris data, satellite clock states, ionospheric and tropospheric data is available a posteriori, see the update rates of the IGS products, for instance . On the other hand, real-time and near-real-time variants of PPP do exist, since rapid or predicted data is often available as well. Typically, the accuracy and integrity of such real-time data is lower in comparison with the final products. Moreover, some of the corrections, like satellite antenna offsets, code and phase biases, *et cetera* are temporally constant. There are several methods (e.g. PPP-AR and PPP-RTK) that do resolve and fix the carrier phase ambiguities of the undifferentiated measurements, which is a challenging task,

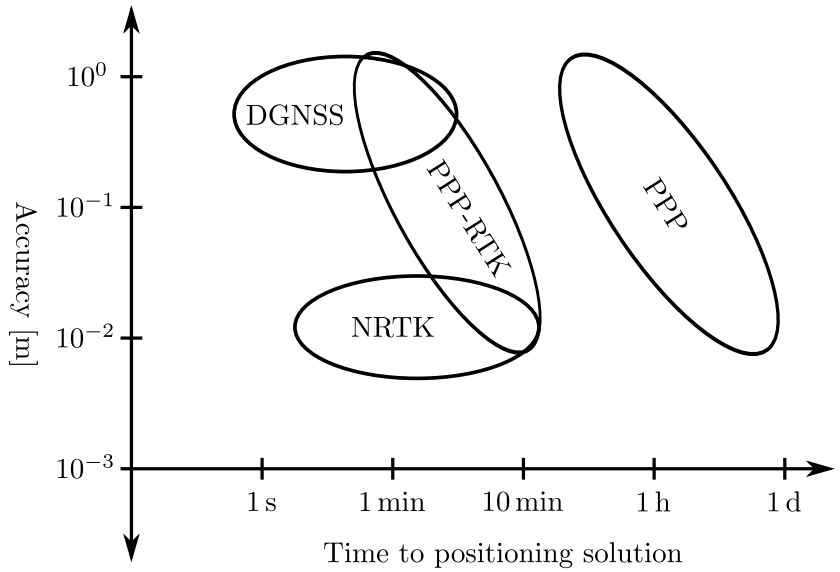


FIGURE 2.14.: Comparison of PPP and RTK and DGNSS (according to [82]).

indeed [82]. The majority of PPP methods requires dual-frequency receivers, since they are capable of elimination of the ionospheric delay. Nowadays, sub-decimeter accuracy is available even in real-time positioning with moving receiver applications [81].

The PPP estimate requires a certain duration of continuous observation (minutes or longer) to converge to the right solution. Since the phase observations are used and the ambiguities are often resolved, the loss of phase lock results in the need of re-initialization.

The amount of the information and the rate of its transfer is much lower than in the case of full observations (as is done in RTK). A comparison of the required data rates used in PPP-RTK and NRTK is provided within [80] a direct manner. It shows that the space-state representation of the GNSS errors (the one used in PPP-RTK) requires much less data bandwidth than the observation-space representation (used in NRTK).

Both RTK and PPP have their applications. The RTK is superior in areas, where high accuracy needs to be achieved very quickly or almost instantaneously. Nonetheless, it requires the reference station or a network, therefore its coverage is somehow limited. The PPP provides slightly worse performance, however is available worldwide. The accurate position estimate cannot be estimated immediately after measurement initiation. An impression of trade-offs between solution availability and accuracy is provided within Fig. 2.14.

2.2.3. Overview of the available systems

This section is dedicated to a short description of the GNSS that are currently in operation. The underlying principle is common to all of them, hence the text focuses on the unique features or specialties of the particular systems.

The summary of the available GNSS signals is provided at the end of this section by means of Fig. 2.15. A table summarizing the detailed technical specifications of the numerous signals can be found for instance at the end of manual [72].

I. GPS – Global Positioning System

The American GPS is the right system to start with, since it was the first ToA-based GNSS. Since the first satellite launch in 1978 several modernized generations of the satellites were deployed. Nonetheless, the original ranging signal, the C/A code on the L1 frequency is still transmitted by all the satellites. The full operational capability was declared in 1995. Naturally, the GPS performance is often considered as a benchmark for the other systems.

The space segment consists of 24 – 32 space vehicles in 6 orbital planes with 55° inclination and height of 20 180 km. It exploits CDMA technique, i.e. all the satellites broadcast at the same carrier frequency (set of frequencies), and their transmissions are distinguishable by means of correlation with pseudorandom ranging codes which are unique to each operational satellite.

The GPS time (GPST) is a continuous timescale that was set to zero at UTC midnight on 6^{th} January 1980; it is bound to the American approximation of UTC - UTC(USNO). Unlike UTC, GPST is not affected by leap seconds, and therefore it is now 18 s ahead of UTC (as of December 2018). The exact relation between GPST and UTC is given in the navigation message.

The modern satellites transmit on three frequencies – L1, L2 and L5.[55] Almost all the receivers are able to receive the C/A code, which is transmitted on L1 (in quadrature) only; the low-cost single-frequency receivers often support this signal solely. The C/A code is a BPSK-modulated Gold-code, with rate 1.023 Mchip/s; also it carries 50 bps of navigation data. In the oncoming generation of GPS satellites, i.e. "Block III", L1C (with time-multiplexed BOC modulation, described in detail in [83]) signal will be transmitted as well.

In the inphase component of L1, as well as in the L2 inphase component, an encrypted precision BPSK-modulated P(Y) code is transmitted. Although the encryption hinders full exploitation of this signal by non-authorized users, it is possible to process it in a semi-codeless manner [84].[56] In the modern satellites, the BOC-modulated military M-code is transmitted on the same carrier, however, the transmitted power is located at the edges of the L2 spectrum rather than around the center frequency, which is the case of P(Y).

[55] L1 1575.42 MHz, L2 1227.6 MHz and L5 1176.45 MHz.

[56] According to [85, p. 59], the semi-codeless access to the P(Y)-code will not be discontinued at least two years after 24 satellites broadcasting L5-signals are oribiting the earth. This is expected to happen in 2024, when the satellites will be transmitting a civil signal on L2 (L2C) as well. Note that before the introduction of L2C, the codeless or semi-codeless processing was the only option to obtain L2 measurements for the civil user.

The modernized GPS satellites transmit an open signal also on L2 (on older satellites only P(Y) code was available on L2). This code is called L2C and is formed by chip-by-chip multiplexing of two BPSK-modulated pseudorandom codes running at 511.5 kchip/s, so the resulting chip-rate is 1.023 Mchip/s, i.e. same as for C/A code. Nonetheless, both code-components of L2C are longer binary sequences, and one is even a pilot signal (no data message is modulated), therefore it promises better performance than the C/A code. Further information on signals on the L1 and L2 frequencies are available from [34], the documentation of L1C is separate [83].

The last generation of GPS satellites (IIF) transmits a civil signal on the L5 carrier. This carrier frequency is not shared with any military/restricted signals, the civil signal has both inphase and quadrature components (usually denoted L5-I and L5-Q). A 1 ms long pseudorandom code with rate 10.23 Mchip/s is BPSK-modulated in each component. Consequently, the resulting composite L5 signal appears to be QPSK-modulated. The 10-times higher chip rate (w.r.t. C/A code) results in 10-times wider bandwidth and should provide similar accuracy improvement. Further detailed information about the L5 interface is available in [86].

II. GLONASS – Globalnaya navigatsionnaya sputnikovaya sistema

The Russian GNSS has its roots in the former Soviet Union, its first satellite was launched in 1982 and the constellation of 24 satellites was completed in 1995.[57] Nonetheless, the number of active satellites dropped later; in 2002 only 7 satellites were available. By 2012, GLONASS was reconstructed and reached the nominal satellite count again, of course with modernized satellites.

The GLONASS timescale is deduced from the Russian UTC(SU) approximation, which is shifted by +3 hours (Moscow time) with respect to the "Greenwich time". It does respect the leap seconds. Consequently, there is a non-constant inter-system bias between the GPS and GLONASS.

The satellites orbit the Earth at 19 130 km altitude in three orbital planes. The inclination of the planes is 64.8°; consequently, the ground track of the satellites reaches to higher latitudes, which improves the coverage and system performance in the near-polar regions.

GLONASS is the only GNSS that uses the FDMA rather than CDMA to distinguish the transmissions of the particular satellites. Therefore, carrier frequencies in the L1 and L2 bands are slightly offset – one of the 14 channels is chosen.[58] The channels are numbered by indexes $k = \{-7, \ldots, +6\}$ nowadays, the L1 and L2 frequencies are

$$f_{\mathrm{L1}} = 1602.0000 + k \cdot 0.5625 \qquad [\mathrm{MHz}] \tag{2.132}$$

$$f_{\mathrm{L2}} = 1246.0000 + k \cdot 0.4375 \qquad [\mathrm{MHz}], \tag{2.133}$$

their ratio is 9:7, regardless which channel is assigned. The open service ranging code is a maximal length sequence of a 9-bit shift register (511 bits long) and is BPSK-modulated with a rate 0.511 Mchip/s on both L1 and L2 carriers (in the quadrature

[57] Although the system was formally declared operational in 1993.

[58] The same channels are assigned to the satellites so that they are never visible at the same time.

component).[59] The inphase component of the very same carriers is used by encrypted precision signal which has 10-times higher chip-rate.

The FDMA is quite challenging for RTK, since the hardware biases in the receivers are typically frequency-dependent and equipment-dependent. Therefore the receiver group delays for the individual satellites may be biased and would not cancel out completely by means of the second difference in the double-difference method.

In 2006 a new ICD [87] that introduces CDMA signals transmitted in L1, L2 and L3 bands for GLONASS has been released.[60] The properties of the signals are described in particular documents for each of the bands [88–90]. On the L1 and L2 carrier, both open and restricted signals are declared to be available in the future generation of the satellites along with the open signal in the L3 band (a few of the current satellites already provide this signal).

III. Galileo

The promising European GNSS; its first satellite was launched in 2011.[61] Although the initial services were declared at the end of 2016, the full operational capability is expected to be reached in 2020, when the full constellation of 30 satellites is to be completed.

The constellation is designed so that the satellites orbit in three planes which are separated by 120° in longitude. The inclination of the planes is rather similar to GPS – 56°, the orbital altitude is 23 222 km.

Galileo transmits in three bands denoted as E1, E5 and E6. The E1 carrier frequency is 1575.42 MHz, i.e. the same as the GPS–L1. The modulation, which is used for the open service in this band, is the composite BOC (CBOC) which resides the inphase component of the carrier. The pseudorandom code rate is 1.023 Mchip/s, the two BOC subcariers exhibit 1.023 MHz and 6.138 MHz, the latter carries 1/11 of the total signal power. Such non-trivial signal structure narrows the tip of the autocorrelation peak and therefore improves the ranging performance, as well as multipath imunity. The E1 CBOC signal consists of two components, E1-B and E1-C, from which the latter one lacks the navigation data message. Detailed information on CBOC signal properties and the Galielo E1 implementation can be found in [91] or [92], respectively.

In the E5 (1191.795 MHz) band, Galileo transmits AltBOC-modulated signal (Alternative BOC). The subcarrier frequency is 15.345 MHz and the chipping rate is 1.023 Mchip/s – the same of the GPS's P(Y) code. All the signal components are unencrypted, the inphase component carries a data message and the quadrature component serves as a pilot signal. The AltBOC signal exhibits two sidebands; the lower one is denoted E5a, the higher one E5b.[62] Both sidebands can be processed separately, or together by the receiver. In the former case, each sideband offers approximately

[59] It is not transmitted on L2 by the older GLONASS satellites.
[60] The carrier frequencies are 1600.995 MHz for L1, 1248.060 MHz for L2 and 1202.025 MHz for L3.
[61] Two experimental satellites Giove-A and Giove-B were launched in 2005 and 2008, respectively.
[62] E5a frequency is 1176.450 MHz; E5b frequency is 1207.140 MHz.

20 MHz bandwidth. When processed together, the bandwidth is more than 50 MHz. Receiving and processing such a wide bandwidth is a challenging task, certainly.[63]

The third Galileo carrier, E6 at 1278.75 MHz carries two BPSK-modulated signals in the inphase component. These signals have chip rate 5.115 Mchip/s and they share the transmission power evenly. The first component is capable of carrying 1 kbps data message, the second one is a pilot. The E6 signals are intended to be used for the Galileo Commercial Service (CA), or High Accuracy service (HA). Some sources claim that BOC-modulated signal of the public-regulated service is present in the quadrature component of E6 carrier.

IV. BeiDou

The Chinese GNSS was often referred as COMPASS. The system specifications did change in the previous years. For instance, the official English versions of the interface control documents for the BeiDou-2 (released in Nov. 2016) specifies the B1I and B2I signals[64] with QPSK modulation of the Gold-code-based ranging code. The full constellation consists of 5 geostationary, 3 IGSO satellites and 27 MEO satellites at 21 528 km altitude and 55° inclination w.r.t. the equatorial plane. The navigation data message differs for the geostationary ones, the IGSO and MEO share the data structure.

The Beidou-2 is however being superseded with BeiDou-3 which has a rather different signal structure. B1C signal is defined as a BOC modulated signal with 100 symbols-per-second data message, which is accompanied by a dataless quadrature-modulated BOC pilot. The carrier frequency is shifted and coincides with the GPS L1 and Galileo E1, i.e. is 1575.42 MHz. The B2a signal is transmitted in the GPS L5 band (1176.45 MHz). Again, the carrier is shared by the pilot and data component. The modulation is BPSK with 10.23 MHz rate for both components, the data rate is double w.r.t. the B1C signal. The B2a signal is not transmitted by the geostationary satellites.

V. Regional navigation satellite systems

The two systems below provide coverage of the specific areas in the Asian region only, however, they share the basic principles with GNSS. The satellites have to be persistently above the area of coverage, so their orbits have to be geosynchronous – the orbital period is exactly one day, which yields approximately 35 800 km orbit altitude.

IRNSS – Indian Regional Navigation Satellite System]

Sometimes, this system is denoted as NAVIC, and it covers the Indian subcontinent, obviously. The constellation consists of 7 satellites – 3 geostationary (34°, 83°, 131.5° east longitude) and 4 IGSO (equator crossings at 55° and 111.5° east longitude, 29° inclination, two alternating in each plane) [93]. IRNSS is a standalone system, and it should be able to provide 20 m (2-σ) accuracy over India and approx. 1500 km surround area.

[63] At least 100 Msamples/s (real) or 50 Msamples/s (complex) have to be taken in order to capture the entire AltBOC spectrum.

[64] The carrier frequencies are 1561.098 MHz and 1207.140 MHz, respectively.

Unlike all other systems, the IRNSS does not transmit any signals in the L1-band. It uses the L5-band (1191.795 MHz), and uniquely, it transmits navigation signals in S-band at 2491.75 MHz, which is at the edge of the ISM-band that is used by Wi-Fi, for example. On both carriers one publicly available BPSK-modulated signal (1.023 Mchip/s) is transmitted. It is accompanied with two regulated-service BOC(5,2)-modulated signals (data and pilot channel) that are placed in the spectral nulls of the open-service code [94].

QZSS – Quasi-Zenith Satellite System

This Japanese system is not stand-alone, it "enhances" the capabilities of GPS for the area of Japan islands and introduces new services. Both, overview information and detailed technical description (including ICD) are available through the official website [95]. The constellation consists of four satellites. One is placed in the geostationary orbit at 127° east longitude, and the other three are placed in the so-called evenly offset quasi-zenith orbits, which is unique among GNSS and gives the name to this system. The quasi-zenith (or tundra) orbit is an eccentric inclined geosynchronous orbit,[65] whose ground-track is an asymmetric eight-figure perpendicular to equator (crossing at 139° east). The advantage of this orbit is that the satellite stays for the significant amount of time in the smaller loop of the eight-figure. Since this loop is placed above Japan and because the orbits are offset, at least one of the satellites is observable near the zenith angle, which is desired especially in the obstructed sites, such as urban canyons.

The tight bond with GPS is obvious, the L1, L2 and L5 carrier frequencies of QZSS are shared with it, the satellites transmit the same signals (L1-C/A, L2C and L5). Moreover, the clocks are synchronized to the GPST. Additionally, QZSS satellites transmit signals in the E6/L6 band (1278.75 MHz), which carry information required by the CLAS – Centimeter Level Augmentation Service.[66] Also, Sub-meter Level Augmentation Service (SLAS) is provided by QZSS by means of a dedicated L1S signal.

[65] The nominal eccentricity is 0.075 and inclination is 41°.

[66] According to the available descriptions, the CLAS uses the state-space representation of the current GNSS errors (PPP-RTK data), provided to the user via satellites with a certain latency (10 to 20 s).

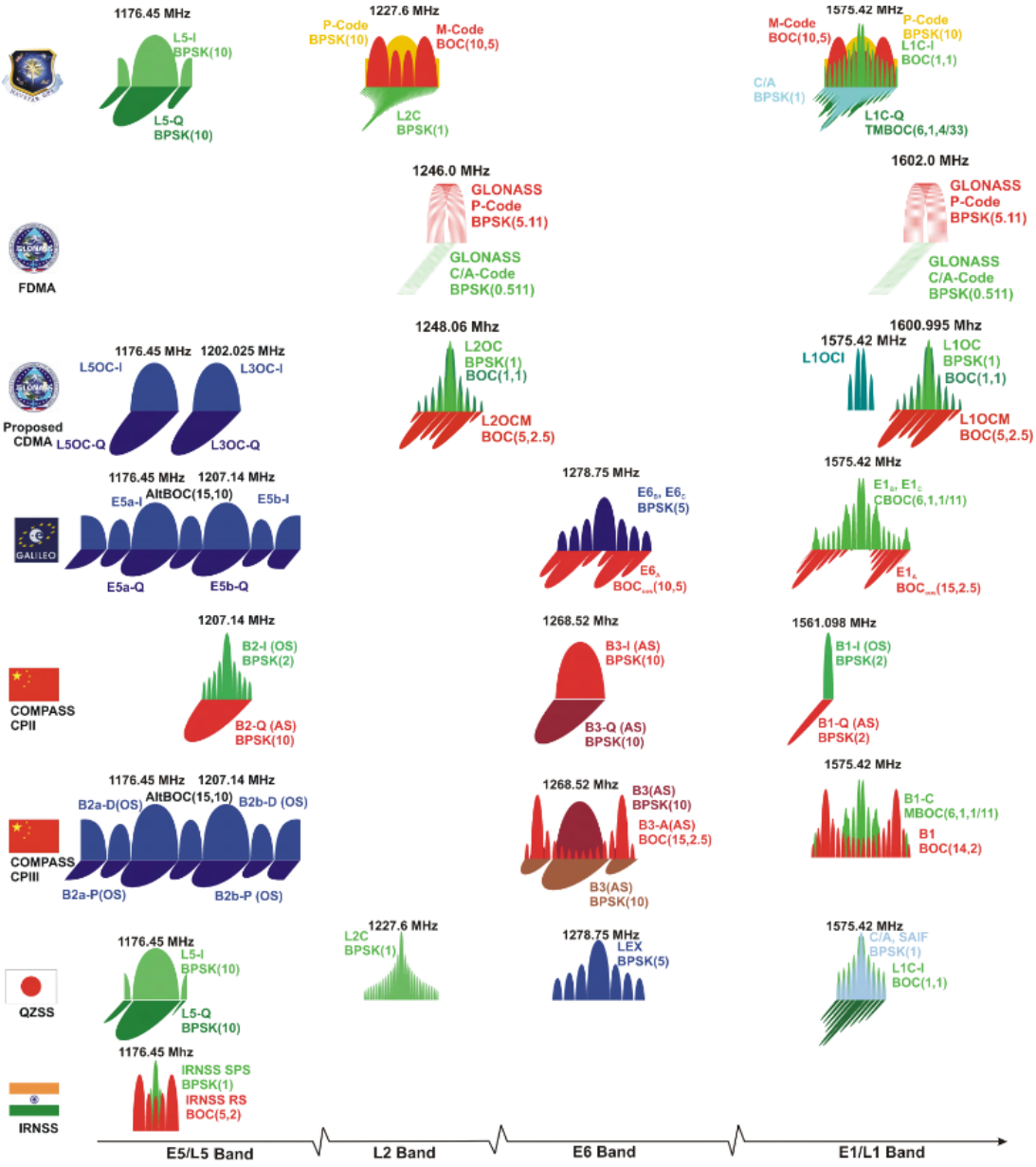


FIGURE 2.15.: Summary of available GNSS signals.
(Used with permission of Stefan Wallner; [96])

2.3. Ultra-Wide Band Positioning

Since the minimal possible standard deviation of signal time of arrival is inversely proportional to the signal effective bandwidth (see CRLB in Section 2.1.2), it is desirable to maximize the bandwidth. This is done for instance in the IEEE 812.15.4 standard, where the HRP UWB PHY (high rate pulse repetition frequency ultra-wide band physical layer) includes optional support for ranging [19]. This technology is also commonly denoted as IR-UWB (Impulse radio-UWB), since it uses a sequence of extremely short pulses to transfer the information. Although other UWB technologies do exist, we will focus solely on the impulse radio. The standard defines 15 UWB channels, most of them provide 499.2 MHz bandwidth.[67] For the sake of simplicity, we will neglect the channels with higher bandwidths. Channel 0 is in the sub-gigahertz band and it is not used for the localization purpose widely.[68] The other channels are located in 3.2 to 4.7 GHz (low-band) and 5.9 to 10.2 GHz (high-band) frequency bands.

The BPSK and BPM (burst position modulation) are used for data modulation in order to allow both coherent and non-coherent reception (only parity is carried by the BPSK modulation). Consequently, there is a single pulse-shape used for a particular channel bandwidth, only its polarity and position in time is varied. The shape of the pulse is defined in [19] by its cross correlation with a root-raised cosine reference pulse with defined duration and roll-off factor $\zeta = 0.5$ (the same as in Section 2.1.4–II). It can be shown that the CRLB of the signal delay or range estimator[69] is below the decimeter level for SNR better than 10 dB. Fig. 2.16 shows the CRLB for three different pulse durations; the $T_p = 2$ ns is the reference pulse for the 499.2 MHz channel bandwidth. It is worth noting that for the wider bands (and therefore shorter pulses) the CRLB drops even more.

Naturally, there are few disadvantages of the UWB positioning technology. The transmitted power, as well as transmitter duty cycle, are strictly regulated due to the very wide bandwidth. For instance, in the 3.4 to 4.8 GHz band the mean PSD is limited to -41.3 dBm/MHz with 1.5 % duty cycle (fixed indoor and mobile devices) and maximal continuous transmission time of 25 ms. Such limitations imply a rather low range of operation, and thus, the UWB localization is considered suitable for indoor applications, mostly. From the received signal power perspective, the UWB systems operate in the vicinity of the noise floor. The penetration of the signal through obstacles, such as walls or vegetation is rather poor for the 3 to 10 GHz signals; reflection or scattering are more likely to appear.

From the UWB characteristics mentioned above it arises that it is a relevant complement of the GNSS for indoor and similar areas, where the satellite signal is unavailable, or insufficient for positioning.

Obviously, the design of the RF equipment for such wide bandwidths is nontrivial. Wideband impedance matching of both power and low-noise amplifiers is rather challenging task, especially when power consumption is constrained. Moreover, the

[67] Channels 4, 7, 11, and 15 provide bandwidths over 1 GHz.

[68] The sub-GHz band has substantially higher relative bandwidth, which is hardware-challenging. Nonetheless, the lower frequency signals are able to reasonably penetrate under the surface of some materials. Applications such as avalanche victim search could benefit from this ability [97].

[69] The CRLBs of the estimators follow equations (2.16) and (2.17) from page 12.

distortion of the signal has to be kept within reasonable limits, thus, the gain and the group delay should not vary significantly throughout the band of interest. Such amplifiers are typically manufactured directly on a chip, where variations in the order of tens of picoseconds are achievable [98, 99]. The same applies for the antennae; not only the group delay should be constant w.r.t. frequency, it also should not vary with angle of signal arrival. Of course, the other design constraints on antennae (such as pattern, polarization, impedance matching, size, radiation efficiency, *et cetera*), are still to be respected.

The positioning is enabled by the introduction of precise counter that is able to timestamp the time of transmission and time of reception of the packets. The counter frequency is defined by IEEE 812.15.4 standard to be 63.8976 GHz and the counter should provide 32-bit unsigned integer value [19].[70] The timestamps can be used to estimate range or position using various methods. Further, we will focus on a description of UWB positioning aspects that were relevant for the author's work, however, detailed in-depth description is available e.g. in [2].

2.3.1. Available positioning methods

Generally the UWB localization system[71] needs to consist of two types of nodes: the *tags* and the *anchors*. The tags are the devices to be localized, while the anchors are the infrastructure of the network with known position. There are two methods that can be used in order to obtain position using an UWB network: the two-way ranging (TWR) and the time difference of arrival (TDoA). The basic principles of the methods and some of their variants have been already described within Sections 2.1.2–II and IV, respectively. The comparison of both positioning approaches is briefly summarized in Table 2.6, however, the rest of this section provides more detailed information.

The TWR always requires bi-directional communication. From the timestamps obtained during the TWR-negotiation, it is possible to compute the true radio range between two communicating network nodes (typically tag and anchor). However, it is also possible to determine the offset of the rangning counters (clocks) in the two nodes. It is usually considered a nuisance parameter, or it can be eliminated by a simple linear combination. The ultimate advantage of this method is that the true range is available as a raw measurement, it is not computed from the position difference as would be in the case of the other methods. This is advantageous especially when the proximity to an object or an area is to be evaluated, and the position of the tag is irrelevant. For example, a warning should be issued for a heavy machinery controller if anybody gets in the vicinity of the machine.

In order to obtain the position using the two way ranging a set of multilateration equations has to be solved. The equations are nonlinear, however, the least-squares method converge typically within a few (up to ten) iterations when the geometry is favorable. The problems typically arise when the measurement or anchor location uncertainty is comparable with the size of the constellation, or when the DOP is high

[70] The counter frequency is a 128-multiple of the HRP UWB PHY chipping rate (499.2 MHz). The 32-bit counter overflows every 67 ms, thus, some UWB modules provide extended counters [100].
[71] Sometimes, RTLS is used as an acronym for real-time localization systems.

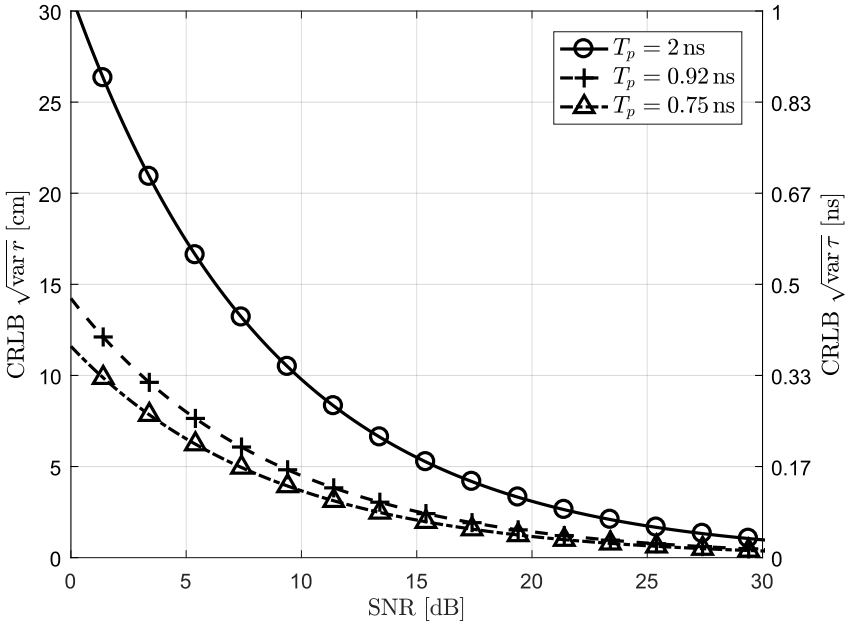


FIGURE 2.16.: Range (left axis) and delay (right axis) CRLB of a root-raised cosine pulses with different pulse duration (T_p) and roll-off factor 0.5.

TABLE 2.6.: Comparison of TWR and TDoA approach in UWB localization systems.

	TWR	**TDoA**
Communication	Bidirectional	Unidirectional
Infrastructure synchronization	Not required	Critical, sub-nanosecond accuracy
Tag/anchor power consumption	High/High	Low/High
Raw measurement useful	Yes (proximity detection)	No
Position fix rate	Tens per second	Thousands per second

(i.e. geometry is poor). Naturally, the number of the range measurements has to be higher or equal to the dimensionality of the localization space. Since each ranging negotiation takes at least a few milliseconds, the number of position fixes obtained in each second is rather limited (up to a couple of tens per second). The radio "traffic" becomes quite dense, and thus, some kind of media access policy (other than basic ALOHA)[72] has to be used. Anyway, either of the two negotiating nodes have to be ready to receive a message for an ineligible amount of time, which may drain the battery capacity of the mobile devices – the tags. Moreover, the number of position fixes may be further reduced due to the constraint on transmission duty-cycle imposed by the emission regulations.

It can be said that at least 3 messages have to be sent in total in order to obtain a range estimate, see scheme on Fig. 2.3. It is worth noting that some of the messages may be used as a part of a multiple ranging negotiations and thus, the overall number of messages per position fix may be reduced [101].

In contrast, when using the TDoA positioning principle, the tag may be located by exploiting a single message that is typically called *blink*. It is also possible to have a passive (receiving-only) tag, which receives messages from the anchors and computes its position.

The former variant is very power efficient at the tag side, especially if ALOHA media access policy is employed. Occasional transmission of the blink without the need to receive anything allows the transceiver to be turned off for a majority of the time. The anchors almost continuously receive the blinks from various tags, timestamp them and send them to a computer that matches the corresponding blink receptions, computes TDoA and estimates positions of the tags. The position is not known to th tag, unless delivered by other means. The capacity of such system is limited by the ability of anchors to process the blinks and by the maximal ALOHA air-time utilization. Typically, a few thousand blinks can be processed per second.

In the latter case, the position is not known at the anchor side of the system, the system works in a similar manner as GNSS with stationary satellites. The tag only has to receive at least three or four anchor positioning messages in order to be able to compute its position in horizontal plane or in 3D, respectively. Since the receiver does not know when the message will be sent, it has to be "armed" for reception for a while, typically. Therefore, this method may be more power-consuming at the tag side than the former one. On the contrary, there is no practical constraint on number of tags to be simultaneously localized. Without the loss of generality, we will focus on the case with the transmitting tag.

As mentioned earlier in Section 2.1.2–IV, the anchors have to share a common timescale. Since each nanosecond of timescale difference results in approximately 30 cm bias of the TDoA measurement, the performance of the time transfer should provide sub-nanosecond accuracy in order to keep the localization accuracy at a reasonable level for indoor applications. Let us consider the worst case scenario, when the reference clocks of two anchors are at the opposite limits of the ±20 ppm oscillator accuracy;

[72] The maximal air-time utilization when using ALOHA protocol is 18 %. Its advantage is that there is no need for media state assessment, nor need to transmit in a timeslot scheme.

it is straightforward to notice that the nanosecond offset value would be reached after 25 µs of operation.

The challenging task of clock synchronization may be solved in a wired or wireless manner. The dedicated high frequency cabling for synchronization is impractical and costly. The wireless synchronization exploits the timestamping ability of the transceiver and often the known propagation delay between two anchors. The wireless methods require unobstructed line of sight between the synchronized nodes. Provided that the propagation delay is known and compensated for, periodic transmission of one-way synchronization messages is sufficient to keep the track of the master timescale in the slave node. Should the propagation delay be unknown, TWR ranging can be used to estimate it [102]. Typically, there are free running clocks in both, master and slave nodes, and the slave node computes a correction that converts the slave timescale to the master timescale. An extensive comparison of the wireless synchronization methods is available in [103]; according to the paper, the best performance is achieved by Kalman filter that tracks the clock offset and the drift (1^{st} derivative of the offset). These methods require the master anchor to be clearly visible by each slave anchor.

Based on experience with synchronizing anchors, the author has incorporated drift-rate (the 2^{nd} derivative of the offset, oscillator frequency drift) tracking into the Kalman filter [5]. By means of such a small increase of computational complexity, the performance of the synchronization was boosted, especially during anchors startup (when the temperature changes rapidly). For detailed information see Appendix C.1. Moreover, the improved synchronization algorithm may be used to transfer the precise time information via a chain of relay anchors as well, not only from master to slave anchors directly. The description and error analysis of the chained synchronization algorithm is available in [5, 6] and is also briefly reviewed within Appendix C.1.

In order to compute position in plane (or in 3D space), at least two (three) TDoA measurements have to be available. Hence, the blink has to be received by three (four) anchors in the case when tag is transmitting. Only then the set of hyperbola or hyperboloid equations can be solved. Note that the dilution of precision due to anchor geometry (DOP, see Sec. 4.1.6) affects the quality of the position estimate, inherently. Each of the anchor's measurement may be interpreted as a pseudorange measurement, similarly as in GNSS. Consequently, it is possible to approach the positioning problem as a solution of a set of ToA equations.

However, according to the author's experience, the convergence of an iterative least-squares solver is more likely to fail on the ToA problem than TDoA problem. Since the offset of the tag's clock is a nuisance parameter in the ToA problem, the TDoA method is to be favored. An example is provided in Fig. 2.17 where the raw measurements from four anchors were approached in both ways. The geometry of the constellation and spatial similarity of the ToA and TDoA results is depicted in Fig. 2.17a. The iterative Levenberg-Maquardt algorithm (see Sec. 4.1.3) was used to solve the set of equation in a least-squares manner. The comparison of required iterations count in Fig. 2.17b clearly shows that the TDoA solution converges faster and more reliably than the ToA. The probable reason is that the time offset (bias) variable, which has

to be estimated in ToA, has often a few orders of magnitude broader range of possible values than the other variables. The success of the solver is rather dependent on the initial guess of the solution vector, which is easily set for the position components.[73]

2.3.2. Error sources

The accuracy of the final position estimate is affected by multiple factors. Most of them contribute to the overall error budget regardless of whether TDoA or TWR approach is used. The following characterization of major error sources is based on theoretical knowledge from the related literature, such as [2], and on author's hands-on experience with an experimental positioning system which is based on DecaWave DWM1000 ranging module [100].

Timestamping inaccuracy

Theoretically, the 64 GHz ranging counter provides resolution of approximately 15 ps (4.5 mm). However, the precision (standard deviation) of the time of arrival measurement is ten times worse even in high-SNR scenario; naturally, the precision degrades for low SNR values. Similar precision may be assigned to the transmission trigger, i.e. the transmission timestamp [103]. Note that this error component is random with zero or negligible mean value. Based on experimental work described in [5] and Appendix C.1, we state that the distribution of this error can be assumed Gaussian (see histograms of the synchronization Kalman filter measurement residuals C.1b on page 179).

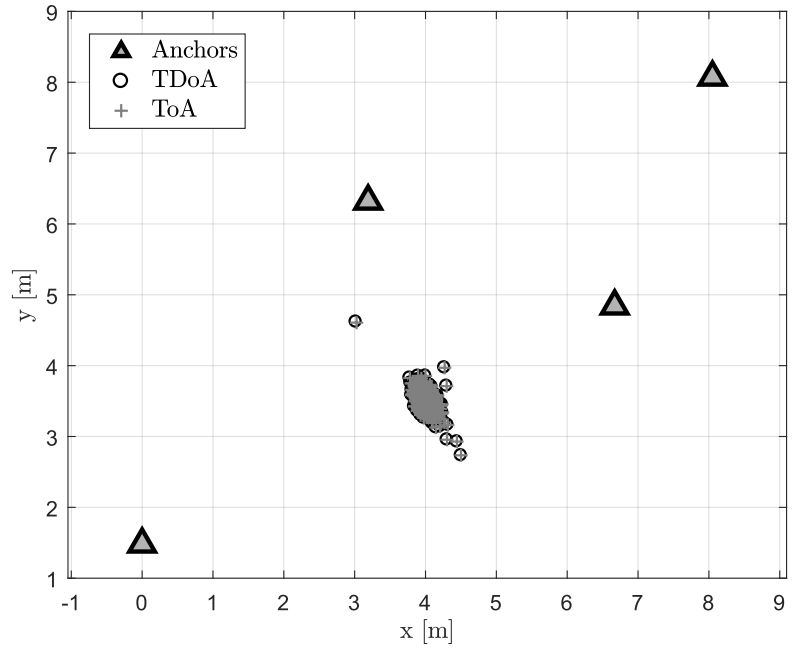
Timekeeping errors

The importance of timing accuracy is critical for both TDoA and TWR positioning approach. The clocks (oscillators) in the UWB transceivers are non-ideal, so in addition to the unknown offset they also suffer from drift, which is not necessarily constant. According to [19] the the clock frequency error should be within ±20 ppm. The problem is not that the actual time of flight is measured inaccurately – even the maximal ±20 ppm error w.r.t. the true value is negligible; the time of flight is highly unlikely to be more than a microsecond in the UWB system.

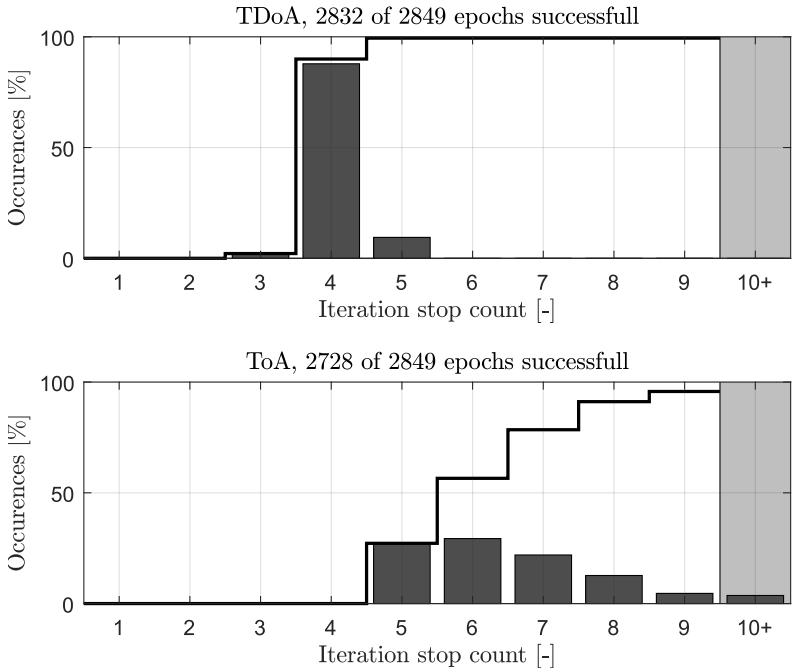
In the TDoA with *transmitting tag* case the timekeeping of the tag is relevant only to the transceiver internal circuitry, since it has to generate the blink signal of sufficient quality. Consider now that the anchors are synchronized using the KF-based algorithm from Appendix C.1. The clock offset, drift and drift rate are estimated on the basis of periodic synchronization messages from master to slave. The actual offset for any timestamp after the synchronization is predicted using the KF state obtained after the reception of synchronization message. Since the true values of the state vector (the clock error parameters) evolve temporally, the period of synchronization message transmission has to be kept short enough to encompass the development of the state vector (typically a few tenths of a second).[74] The TDoA values corrected for the

[73] The initial time-offset guess was set to the *minimum* of the ToA values for the purpose of generating Fig. 2.17, which resulted in 95.6 % positioning success rate. In contrast, when the time offset was initialized as the *mean* of the ToA values, the success rate was only 85.4 %, i.e. the count of failed attempts grew more than three times.

[74] In the case of DWM1000 modules, 400 ms period was sufficient to achieve sub-nanosecond accuracy, it also provided performance margin for occasional missing messages.



(A) Horizontal plot of the anchor constellation and the position solutions.



(B) Number of Levenberg-Marquardt algorithm iterations required to converge; histogram (bars) and CDF (line). The shaded area on the right shows failed positioning attempts.

FIGURE 2.17.: Comparison of ToA and TDoA problem convergence in a simple UWB localization test [104]. The positioning was successful when the RMS value of residuals fell below a threshold value in less than 10 iterations.

predicted clock offsets of the respective anchors are affected by the prediction errors. Those are kept below the nanosecond-level, however.

When the inverse variant of the TDoA is used, i.e. *tag is receiving*, the clock drift of the tag may become important as well. Only one message at the time can be received simultaneously; therefore the anchors should not transmit their messages concurrently, a small delay that would enable receiver to capture the consecutive messages from different anchors is necessary. Provided that the transmission timestamp (in the system master time) is included in each message, the tag can correct for the intentional transmission delay. When the delays to be corrected reach approximately a millisecond length, the difference between time interval measured at the tag and at the anchor becomes comparable with the ranging precision. Under such circumstances, the clock drift has to be taken into account and the simple TDoA measurement equation (2.65) from page 21 has to be modified. The TDoA, or the drift-affected range-difference between i and j anchors $\breve{d}_{\mathrm{i,j}}$ is to be expressed as

$$\begin{aligned} \breve{d}_{\mathrm{i,j}} &= \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{i}}\| - \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{j}}\| - c_0\,\delta\nu^{[\mathrm{T}]}\left(t^{[\mathrm{T}]}_{\mathrm{Rx,i}} - t^{[\mathrm{T}]}_{\mathrm{Rx,j}}\right) \\ c_0\left(t^{[\mathrm{T}]}_{\mathrm{Rx,i}} - t^{[\mathrm{M}]}_{\mathrm{Tx,i}} - t^{[\mathrm{T}]}_{\mathrm{Rx,j}} + t^{[\mathrm{M}]}_{\mathrm{Tx,j}}\right) &= \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{i}}\| - \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{j}}\| - c_0\,\delta\nu^{[\mathrm{T}]}\left(t^{[\mathrm{T}]}_{\mathrm{Rx,i}} - t^{[\mathrm{T}]}_{\mathrm{Rx,j}}\right), \end{aligned} \tag{2.134}$$

where $t^{[\mathrm{T}]}_{\mathrm{Rx,i}}$ and $t^{[\mathrm{M}]}_{\mathrm{Tx,i}}$ are the reception time of message from the i-anchor in the tags timescale, and the transmission time of this message in the master anchor timescale, respectively. The relative tag clock drift $\delta\nu^{[\mathrm{T}]}$ should be estimated as an additional nuisance parameter.[75]

In contrast, when using TWR, the time of flight (and therefore range as well) is computed from the measurements of the time intervals, which are a couple of orders of magnitude longer than the time of flight. Considering a few ppm clock offset, the nanosecond-level error is obtained when working with millisecond-long intervals. The actual TWR methods and their performance under assumption of drifting oscillators have been discussed in Section 2.1.2–II (p. 14). Generally, due to the limited accuracy of the oscillators, the single-sided TWR is inapplicable. Therefore the linear SDS-TWR or nonlinear ADS-TWR estimators are used. Some variants do exist, such as those presented in [105, 106], nonetheless, it can be shown that they are principally and performance-wise equivalent to the double-sided TWR approaches covered within this thesis [4].[76]

Hardware delays

The timestamping takes place in the digital part of the transmitter and receiver. Nonetheless, the time of flight should be measured between the antennae phase centers. There are substantial delays between antenna and the timestamping point, moreover, they are different for the transmission and reception path. The delays will be further denoted as τ_{Tx} and τ_{Rx}. The paths of the signal are depicted in Figure 2.18.

[75] Note that the clock drift is not likely to change swiftly, unlike the position coordinates. Such behavior may be exploited when estimating the position and drift by means of EKF or other nonlinear random process estimator.

[76] The performance-improving modifications of the DS-TWR are based on the reordering of the messages or minimizing the turnaround time.

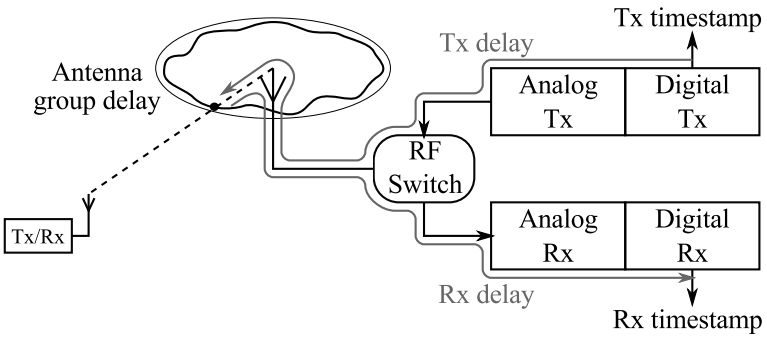


FIGURE 2.18.: Sources of HW delays in UWB transceiver.

Hardware delays for receiver and transmitter are often dependent on the choice of modulation settings and channel. Temperature variation affect the hardware delays as well.

The group delay of the antenna and its cable (or other transmission) is common to both, transmitting and receiving direction. The antenna group delay however varies also with the incident angle.[77] Unlike the other hardware delays, the angular group delay variation cannot be calibrated and accounted for effectively, since the angle of arrival and antenna orientation have to be known simultaneously. However, it is not impossible [107]. The magnitude of such variation is typically kept within $\pm 100\,\mathrm{ps}$, i.e. approximately to $\pm 3\,\mathrm{cm}$ of signal path in open space, see e.g. [108, 109].

The transmission and reception hardware delays cannot be measured separately, since there is typically no simultaneous access to the timestamping output and the antenna output. Only the aggregate delay

$$\tau_\Sigma = \tau_{\mathrm{Tx}} + \tau_{\mathrm{Rx}} \tag{2.135}$$

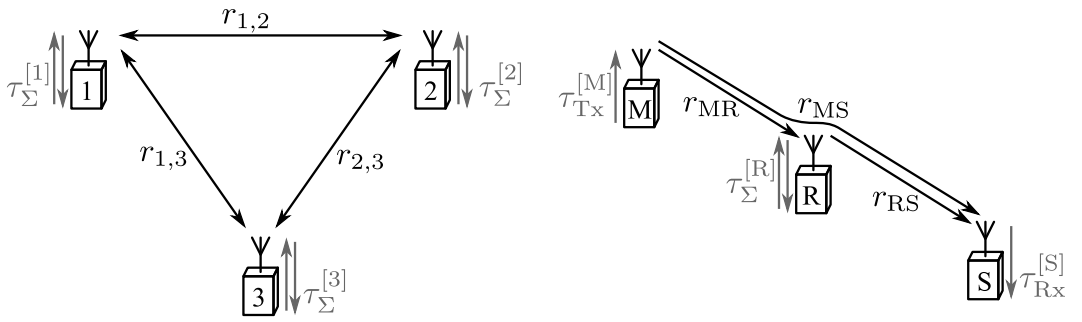
can be estimated using several methods [110, 111]; the ratio of τ_{Tx} and τ_{Rx} may be provided by the device manufacturer.

One of the calibration procedures uses [110] TWR measurements. Each of the range measurement $\hat{r}_{\mathrm{AB}}$ is affected by the aggregate delays of both transceivers involved and noise, which may be assumed white:

$$\hat{r}_{\mathrm{AB}} = r_{\mathrm{AB}} + \tau_\Sigma^{[\mathrm{A}]} + \tau_\Sigma^{[\mathrm{B}]} + \mathcal{N}(0, \sigma_r^2). \tag{2.136}$$

For the purpose of calibration, the true range r_{AB} is known. If a reference device with known τ_Σ is available, the τ_Σ of any other device can be computed using (2.136), assuming that the range measurements are averaged sufficiently in order to reduce the random noise impact below significant level. When no reference design is available, at least three devices are required for calibration. The matrix equation describing

[77] The antenna (and to some limited extent the transmission) are dispersive, i.e. their group delay varies with frequency as well. This may reduce UWB pulse fidelity, which may indirectly induce some additional ranging error.



(A) TWR-based method (B) Relay-based method

FIGURE 2.19.: Methods of UWB transceiver hardware delay calibration.

the situation for three transceivers and the general expression for arbitrary number of transceivers holds

$$\frac{1}{c_0}\begin{pmatrix}\bar{r}_{1,2}-r_{1,2}\\ \bar{r}_{1,3}-r_{1,3}\\ \bar{r}_{2,3}-r_{2,3}\end{pmatrix}=\begin{pmatrix}1&1&0\\1&0&1\\0&1&1\end{pmatrix}\begin{pmatrix}\tau_\Sigma^{[1]}\\ \tau_\Sigma^{[2]}\\ \tau_\Sigma^{[3]}\end{pmatrix}, \tag{2.137}$$

$$\frac{\bar{\boldsymbol{r}}-\boldsymbol{r}}{c_0}=\boldsymbol{A}\boldsymbol{\tau}_\Sigma. \tag{2.138}$$

Observe that $\bar{\boldsymbol{r}}$ is a vector of averaged $\hat{r}$ measurements, and $\boldsymbol{A}$ is a matrix of the measurement pairs, i.e. there are two ones in each row and all rows are unique. The number of rows is given by a binomial coefficient $\binom{2}{n}$, where n is number of transceivers used. Although the problem is solved by means of genetic algorithm in [110], it is possible to solve such linear problem using the ordinary least squares estimator

$$\hat{\boldsymbol{\tau}}_\Sigma=\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\frac{\bar{\boldsymbol{r}}-\boldsymbol{r}}{c_0}. \tag{2.139}$$

The second method of calibration uses two helper devices to calibrate the third transceiver. It may be implemented by exploiting the chained synchronization [6] or as described in [111]. We will denote the helpers as master (M) and slave (S) and the calibrated device as relay (R). The master sends a timestamped message that is captured by both, relay and slave; naturally, the reception time is known as well. Then, relay sends a timestamped message to slave. Under the assumption of driftless clocks, the clock offsets of the relay and slaves clock w.r.t. the master clock ($\delta t^{[\mathrm{MR}]}$, $\delta t^{[\mathrm{MS}]}$, respectively) remain constant. Reusing (2.22) we may write equations for transmission and reception times:

$$t_{\mathrm{RxM}}^{[\mathrm{S}]}=t_{\mathrm{Tx}}^{[\mathrm{M}]}+\delta t^{[\mathrm{MS}]}+\frac{r_{\mathrm{MS}}}{c_0}+\tau_{\mathrm{Tx}}^{[\mathrm{M}]}+\tau_{\mathrm{Rx}}^{[\mathrm{S}]} \tag{2.140}$$

$$t_{\mathrm{RxM}}^{[\mathrm{R}]}=t_{\mathrm{Tx}}^{[\mathrm{M}]}+\delta t^{[\mathrm{MR}]}+\frac{r_{\mathrm{MR}}}{c_0}+\tau_{\mathrm{Tx}}^{[\mathrm{M}]}+\tau_{\mathrm{Rx}}^{[\mathrm{R}]} \tag{2.141}$$

$$t_{\mathrm{RxR}}^{[\mathrm{S}]}=t_{\mathrm{Tx}}^{[\mathrm{R}]}+\left(\delta t^{[\mathrm{MS}]}-\delta t^{[\mathrm{MR}]}\right)+\frac{r_{\mathrm{RS}}}{c_0}+\tau_{\mathrm{Tx}}^{[\mathrm{R}]}+\tau_{\mathrm{Rx}}^{[\mathrm{S}]}. \tag{2.142}$$

By simple combination of the above equations we may obtain

$$\tau_{\Sigma}^{[\mathrm{R}]} = \left(t_{\mathrm{RxR}}^{[\mathrm{S}]} - t_{\mathrm{RxM}}^{[\mathrm{S}]}\right) - \left(t_{\mathrm{Tx}}^{[\mathrm{R}]} - t_{\mathrm{RxM}}^{[\mathrm{R}]}\right) - \frac{r_{\mathrm{MS}} - (r_{\mathrm{MR}} + r_{\mathrm{RS}})}{c_0}. \tag{2.143}$$

It is worth noting that the clock offset terms and delays at master and slave transceivers are eliminated, since they are common to both propagation paths – direct, and via relay. The true distances between the nodes r_{MR}, r_{RS} and r_{MS} are known, or the transceivers are placed in-line and thus condition

$$r_{\mathrm{MR}} + r_{\mathrm{RS}} = r_{\mathrm{MS}} \tag{2.144}$$

holds. In such case the propagation delays are inherently compensated. Naturally, a vast amount of measurements has to be performed in order to suppress the random noise. The time interval $(t_{\mathrm{Tx}}^{[\mathrm{R}]} - t_{\mathrm{RxM}}^{[\mathrm{R}]})$, and consequently $(t_{\mathrm{RxR}}^{[\mathrm{S}]} - t_{\mathrm{RxM}}^{[\mathrm{S}]})$ as well, should be kept as short as possible in order to mitigate the error induced by the clock drift.[78]

The same principle of mitigation master and slave delays may be exploited when using chained synchronization [5, 6]. Provided that the propagation ranges are compensated, the relays aggregate delay may be recognized as the difference of slaves clock offset observed from the messages sent by relay and the offset observed from the direct master messages. By the nature of the Kalman filter and its tuning, the noise is suppressed already.

Propagation errors

Arguably, the most noticeable errors originate from the multipath propagation. Generally, we are interested in the measurement of propagation delay on clear, direct line of sight; this is always the shortest and fastest path for the signal. Nonetheless, the signal propagates by reflections from walls, structures, *et cetera*, so there are other delayed "copies" of the signal received.

Based on whether the direct line of sight is available or not, the propagation channel is usually denoted as LOS or NLOS, respectively. In the LOS channel, the first received signal is typically the most dominant in terms of received energy. In contrast in case of the NLOS propagation, some of the reflected signal is the most powerful, i.e. the highest peak does not correspond to the correct delay to be measured. The direct component is often very weak or missing completely. Some algorithms of detecting the first path under the NLOS conditions are available in [2].

If the direct line of sight is obstructed by a dielectric material, the signal is not only partially reflected and attenuated, but is delayed as well. According to (2.13) and (2.14), we may say that the relative erroneous prolongation of the range is proportional to the averaged square-root of the dielectric constant (relative permitivity) ϵ_r along the signal path. Nevertheless, it is highly unlikely that a significant portion of the path will be filled by a medium with relative permitivity substantially higher than one, and the signal will not be attenuated below the sensitivity of the receiver at the same time.

[78] In [111] the clock-drift-related error is suppressed by extension of the negotiation.

2.3.3. Brief summary of UWB localization

The frequency bands for ranging and positioning are placed between 3 and 10 GHz and exhibit bandwiths in excess of 500 MHz. Although the bandwidth promises decimeter-to-centimeter level accuracy, the high-frequency signals are propagating through the open space almost exclusively. The signals tend to reflect from obstacles, rather than propagate through them.

Since the UWB system coexists with other radio systems operating in the 3 to 10 GHz band, its transmissions are strictly regulated in terms of maximal transmitted power and maximal air-time utilization. Usually, the systems have maximal range of not more than hundred meters. A typical application is in indoor localization systems, or as a complement of GNSS where it is not available.

2.4. LF and VLF systems

Low-frequency and very-low-frequency navigation (or positioning) systems are an opposite to the UWB systems. They typically cover vast areas, the transmitters are able to service ranges beyond hundreds or thousands of kilometers, depending on the frequency and power used. The long waves (the wavelength λ span for LF and VLF is 1 to 100 km) propagate through the various obstacles, since they are barely an ideal conductor and electrically reasonable fraction of wavelength in size.

At those frequencies the waves propagate between the ground and ionosphere [112]; unlike in the previous systems described within the thesis, the range is measured along the Earths surface.[79] Naturally, such systems are able to provide only a 2-D position fix, i.e. latitude and longitude coordinates.

Due to the size of the transmitting antennae and high radiated powers, it is almost impossible to jam these systems. It is a considerable advantage over the low-power, high-frequency solutions, such as GNSS.

2.4.1. Omega

Omega VLF navigation system can be considered to be the ground-based predecessor of the GNSS, since it provided global coverage. It was operating from 1982 to 1997. The global coverage was secured by 8 synchronized stations scattered around the globe – the locations were Norway, Liberia (earlier Trinidad), Hawaii, North Dakota, Reunion, Argentina, Australia and Japan. They transmitted concurrently in the 10 to 14 kHz frequency band; the radiated power was 10 kW per station. Due to the extreme wavelength, the transmitters typically used approximately 400 m tall grounded or isolated towers as the antennae.[80]

Detailed descriptions and principles of operation are available e.g. in [113, 114]. According to the sources, it was able to achieve 2 to 4 km accuracy via hyperbolic navigation (TDoA) approach. Indeed, the phase difference of the received signals

[79] Approximately the great circle distance, not the euclidean distance.

[80] The antennae in Norway and on Hawaii sites were valley-span cables that were several kilometers long.

was measured, rather than time of arrival difference, due to the wavelength used. Multiple frequencies were used in order to reduce the ambiguity of the phase difference measurements.

The VLF waves are considered to propagate in a spherical "waveguide" between the ground and ionosphere, the propagation is noticeably affected by the surface conductivity and ionosphere state [112]. Therefore, the night and day propagation differs and the waves are attenuated excessively in the snow-covered areas [113].

A Russian counterpart of the Omega system is called Alpha, or RSDN-20, both the frequency range used and principle of operation are very similar.

2.4.2. Loran family

The history of the hyperbolic navigation systems Loran (long range navigation) starts in the 2nd World War. Multiple generations of the system were developed, arguably the most widely used is Loran-C, which was developed in the late 50's. Although the first versions exploited HF signals, Loran-C uses approximately 300 µs long pulses modulated on a 100 kHz carrier wave.

There are two dominant modes of propagation of the 100 kHz waves – ground wave and sky wave [115]. The former propagates along the finitely conductive surface of the Earth and the latter is reflected from the ionosphere. The skywave arrives later, since its propagation path is longer; the delay is dependent on the effective height of ionosphere, which changes with daytime. However, the ground wave decays more quickly, therefore the less-stable skywave is more dominant at greater distances from the transmitter.

The Loran transmitters are organized in precisely timed chains, each consists of master and several secondaries. The TDoA of the ground waves[81] from at least two master-secondary pairs are required in order to obtain 2-D position fix. The system should provide 0.25 NM (460 m) absolute accuracy (95 %) and up to 90 m repeatable accuracy. The transmitters radiate 10 kW to 1.6 MW peak power, typically by means of approximately 200 m tall mast antennae.

Primarily, Loran operates over sea surface, which is more conductive than ground. For instance there is a chain of Tchayka, Russian analogy to Loran, with increased transmitter power that enables operation over the "dry" ground [115].

The position fix inaccuracy may be deteriorated by the DOP factor when the receiver to transmitter chain geometry is unfavorable, i.e. the measured hyperbolas of constant TDoA intersect in an obliquely. It is worth noting that the Omega system usually did not suffer from such phenomena, since the baselines between the transmitters were substantially longer.

2.5. Local terrestrial navigation systems

Within this section we are considering the local navigation systems that operate on the UHF frequencies and beyond, i.e. the electromagnetic waves propagate approximately via the line of sight. Such systems may often be understood as ground-based GNSS.

[81] Both delay and phase of arrival are measured in order to increase the accuracy.

Further we are discussing two commercially available systems – Locata and NextNav. The information provided is based on the published research articles and materials available via the respective websites.

2.5.1. Locata

The Locata system operates in the ISM band around the 2.45 GHz frequency. It heavily builds on the GNSS (or more precisely GPS) principles [116]. The system uses a passive (receiving-only) user equipment and a set of "LocataLites" that transmit the ranging signals. The LocataLites are precisely synchronized via line-of-sight links and they transmit 1023-chips long Gold code ranging sequences with a 10.23 MHz chipping rate. Since the transmitters are close to the ground, the difference of the received signal power for two LocataLites may be several tens of dB, which is beyond the code-division capabilities of the Gold codes.[82] Therefore, the TDMA is employed in addition to CDMA, i.e. timeslots are assigned to the LocataLites as well. No guard intervals are required between the slots, owing to the precise synchronization.

The Locata network operating range may be as wide as 50 km. The positioning methods are the same as in the case of GNSS, moreover, the position of the stationary LocataLites does not have to be computed from ephemeris and the ionosphere does not affect the propagation of the waves at all – only the local tropospheric conditions are relevant.

Locata receiver is able to provide both pseudorange and carrier phase measurements. Similarly to GNSS, meter-level accuracy is achievable when using the former, centimeter-level accuracy with the latter. The DOP is typically the key aspect of accuracy in case of ground-based constellation. For instance a very "flat" constellation cannot provide vertical accuracy as good as the vertical one.

2.5.2. NextNav

The NextNav system is a metropolitan beacon system (MBS) that is intended to complement GNSS in urban areas, mostly [117]. The beacons transmit ranging signals, whilst the users only receive. For the ranging purpose it exploits long Gold code pulses with the same chipping rate as GPS L1 C/A, the transmitters are ground based and precisely synchronized to the GPS time (the accuracy of synchronization is within a few nanoseconds). Only the main spectral lobe of the pseudorandom ranging code is transmitted in the licensed 919.75 to 927.25 MHz band with up to 30 W ERP.

Unlike Locata, the synchronization of the beacons rely on the GPS system. Moreover, the precise phase measurements are not of interest in case of NextNav. It targets on mass-market devices and related services, such as E911 localization (emergency localization).

The system is able to provide horizontal accuracy of a few tens of meters via the radio positioning [117]. Due to the unfavorable vertical dilution of precision (VDOP), which is caused by all transmitters located approximately in a single plane, the vertical accuracy of the radio localization is poor. Therefore the user should use barometric

[82] The cross-correlation levels of two different Gold codes are approximately 13 dB below the main peak of their the autocorrelation function.

measurements to determine the altitude. In order to enable floor-level accuracy the beacons make precise barometric and other meteorological measurements and provide this information to the users via the ranging signals. The receiver uses the data to eliminate the "meteorological biases" of its atmospheric pressure measurements, so that the resulting value corresponds solely to the altitude.

The precise timing and straightforward signal structure are favorable for exploitation of the NextNav for the purpose of passive radar. Especially, the ambiguity function of the pulses, i.e. correlation function with Doppler-shifted replicas, is unambiguous within the typical maximal range and provides reasonable resolution in both range and Doppler-shift domains [118, 119]. Due to the high level of prior knowledge about the signal structure[83], the signal can be demodulated with less effort then communication or broadcasting signals, for instance. When demodulation is performed, a clear noiseless replica of the transmitted signal can be reconstructed in the passive radar receiver, which brings performance improvement or enables the simplification of the radar receiver from coherent dual-channel architecture to single channel one [120]. In the cited sources the exploitation of NextNav signals by a passive radar for detecting low-altitude airborne targets is investigated.

2.6. Signals of Opportunity

This term, often abbreviated to SoO, is used to denote signals that are not primarily used for localization purposes, however their properties are suitable for their exploitation for navigation. The very basic requirements on the signal properties are outlined by the ranging CRLB – wide RMS bandwidth and signal power are favorable. If the ToA or TDoA estimation is targeted, some "features" of the signals have to be known *a priori* to the receiver.

The typical examples of SoO of the are signals of cell communication systems (e.g. LTE or CDMA2000), or broadcasting systems (e.g. DVB-T). It should be remarked that the transmitters of the SoO are uncooperative and the signals themselves are optimized for their original purpose, not for navigation. For instance, if the transmitters or base stations are synchronized, they are synchronized only to the extent required for delivering the data, performing handovers *et cetera*. Such requirements are rather weak in comparison with performance that would be ideal for localization purpose.

2.6.1. Basic principle of SoO processing

In this section we will focus on the ways of exploitation of SoO that are common regardless of which signals are used. The technique that leads to the ToA or TDoA estimation is of our interest. We will disregard methods of estimating the angle of arrival by means of antenna systems.

Most of the signals have some kind of pilot components or preambles that are known to the receiver *a priori*. We will denote such signal component $s_{\mathrm{p}}(t)$. The rest of the

[83] The ranging code is known *a priori*, similarly to GNSS. The variable component of the signal is the BPSK-modulated low-rate data message. The same holds for the GNSS signals themselves, however their low power levels are prohibitive with regard to the moving-target localization.

signal carries random data and will be denoted $s_{\mathrm{d}}(t)$. The transmitted signal is then formed as a superposition of such signals

$$s(t) = s_{\mathrm{p}}(t) + s_{\mathrm{d}}(t). \tag{2.145}$$

Assuming a simple linear time-invariant AWGN channel, the received signal $x(t)$ holds

$$s(t) = As(t-\tau) + w(t), \tag{2.146}$$

where A is a scaling factor expressing attenuation in the channel, τ is the propagation delay and $w(t)$ is the AWGN. The delay of the signal can then be estimated by means of finding the peak of correlation function of the received signal

$$R_{x,sp}(\tau) = \int x(t)s_{\mathrm{p}}^{*}(t-\tau)\mathrm{d}t. \tag{2.147}$$

It is worth noting that it is equivalent to peak search in the matched filter output since it holds

$$x_{\mathrm{MF}}(t) = x(t) \circledast s_{\mathrm{p}}^{*}(-t) \tag{2.148}$$

$$= \int x(\tau)s_{\mathrm{p}}^{*}(t-\tau)\mathrm{d}\tau = R_{x,sp}(t). \tag{2.149}$$

If a Doppler shift (or carrier frequency mismatch between receiver and transmitter) is present, the cross-ambiguity function (CAF)

$$\Psi_{x,sp}(\tau, f_{\mathrm{D}}) = \int x(t)e^{\mathrm{j}2\pi f_{\mathrm{D}}}s_{\mathrm{p}}^{*}(t-\tau)\mathrm{d}t. \tag{2.150}$$

has to be investigated. The correlation function or CAF may be computed sequentially over the batches of samples and searched for a peak that corresponds to the signal time of arrival.

Such methods of processing are used in the passive radar as well. Nonetheless, in the passive radar field it is usual to correlate the direct signal from the transmitter (or illuminator) of opportunity with the receivers' surveillance channel in order to search for much weaker reflections (delayer and frequency-shifted replicas) of the direct signal. In SoO processing a known signal component is compared with the whole received signal.

In some cases, when the SoO is continuously transmitted, the acquired signal can be tracked; the PLL or FLL loops may be used for tracking the carrier offsets, whilst DLL architecture may serve for tracking the temporal variations.[84] From the DLL the pseudorange corresponding to a particular SoO transmitter is obtained; of course, pseudoranges measured from various synchronized sources may be differentiated in order to acquire TDoA measurements. In case of the snapshot receiver, the TDoA can

[84]The DLL requires at least the early and late correlator outputs for its operation, prompt correlator is often available as well. Those outputs are indeed samples of the correlation function $R_{x,sp}$ before and after the maximum (early and late) or at the maximum (prompt). The rest of the loop exploits the correlator outputs in order to adjust the delay τ estimate and therefore keeps the evaluated correlation function samples in the vicinity of its maximum.

be sometimes obtained as a relative delay between two peaks in the evaluated cross-correlation function or CAF, provided that the signals transmitted by two transmitters are almost identical (as in case of broadcasting single-frequency network, for example).

It has already been mentioned several times that only a particular component of a SoO is known. In the case of snapshot approach to the processing, the algorithms need not to work in real time, the received signal portion can be saved and reprocessed. Such approach provides an advantage, since even the data component of the received signal can be demodulated and reconstructed. Decoding the actual data from the symbols is not necessary, only the modulated symbols are of interest. The whole signal can be then "re-correlated" with the received signal; clearly, the ranging performance should improve, since both SNR and RMS bandwidth become higher. Under assumption that the *a priori* known signal component and the unknown data component are orthogonal,[85] the improvement of the signal characteristics can be quantified based on the PSD of the respective components solely. The joint RMS bandwidth is not a sum of the individual bandwidths of the respective components, it follows

$$B_{\mathrm{RMS}}[s_{\mathrm{p}}+s_{\mathrm{p}}]=\sqrt{\frac{\int_{-\infty}^{+\infty}f^2\left|S_{\mathrm{p}}(f)+S_{\mathrm{d}}(f)\right|^2\mathrm{d}f}{\int_{-\infty}^{+\infty}\left|S_{\mathrm{p}}(f)+S_{\mathrm{d}}(f)\right|^2\mathrm{d}f}}=\sqrt{\frac{f^2\mathcal{S}_{\mathrm{p}}(f)+f^2\mathcal{S}_{\mathrm{d}}(f)\mathrm{d}f}{\int_{-\infty}^{+\infty}\mathcal{S}_{\mathrm{p}}(f)+\mathcal{S}_{\mathrm{d}}(f)\mathrm{d}f}}. \quad (2.151)$$

Obviously, the second equality is valid only because the orthogonality assumption. The joint SNR, however, is a sum of the individual SNRs, since the noise power remains unchanged P_{N}, i.e.

$$\mathrm{SNR}[s_{\mathrm{p}}+s_{\mathrm{d}}]=\frac{P_{\mathrm{p}}+P_{\mathrm{d}}}{P_{\mathrm{N}}}, \quad (2.152)$$

where P_{p} and P_{d} are power of the respective signal components.

The position of dedicated positioning transmitters is precisely surveyed and provided to the navigation equipment.[86] This is not always true for the navigation "transmitters of opportunity." Sometimes, the position of transmitter is publicly available (e.g. DVB-T transmitter positions are obtainable from regulators, typically), however the information may be inaccurate or incomplete. For reasonably low transmitter counts the transmitting sites may be at least approximately surveyed (e.g. [121]). Otherwise sensor fusion algorithms [122], or collaborative estimation techniques [123] may be used to estimate the position of the exploited transmitters.

The family of the potential SoO is rather broad and their structure is almost never as simple as was assumed in the preceding paragraphs. There are various aspects that hinder achieving the optimal performance or which impose some ambiguity. For instance, the CAF should ideally contain one single peak at zero delay and zero frequency shift; nonetheless, it often contains various other peaks that induce ambiguity in the signal (the pilot signals often possess periodic properties). From a different perspective, such

[85] A typical example can be seen in OFDM-modulated signals, where dedicated subcarriers are used as the pilot signals. All the subcarriers are orthogonal by definition.

[86] This assertion is rather simplified; it applies only to the navigation systems with passive user equipment, such as GNSS, LF and VLF systems or terrestrial beacon systems, obviously. The transmitter position information is either known *a priori* (e.g. Omega, Loran), included in the navigation message (e.g. Locata, NextNav), or computed using predicted orbit parameters included in the navigation message (GNSS satellites).

properties may be advantageous for signal acquisition. However, those aspects are typically bound to a specific type of the SoO (e.g. utilization of signal cyclostationary properties are described in [124]), and their detailed description is beyond the scope of the thesis; only a brief review of certain examples is provided in the following section.

2.6.2. Examples of SoO use

Digital video broadcasting uses powerful transmitters (e.g up to 0.1 MW in the Czech Republic), up to 7.6 MHz bandwidth is utilized in each channel. Transmissions of the DVB-T format are used often as a SoO. DVB-T uses OFDM modulation, which secures almost flat spectrum in the occupied bandwidth B_{OBW}. The RMS bandwidth is therefore

$$B_{\mathrm{RMS}}[\mathrm{OFDM}] = \frac{B_{\mathrm{OBW}}}{2\sqrt{3}}. \tag{2.153}$$

Note that full bandwidth of the signal may be exploited only when all data is known, i.e. only with a snapshot receiver; otherwise, only the pilot signals (both scattered and continuous) can be used for positioning. Examples of DVB-T signal snapshot processing are available in [121, 125], in [126, 127] the tracking of signal is described and evaluated.

Meter-level tracking accuracy has been reported in [127], however, the actual positioning performance is not mentioned. However, the positioning errors are notd determined by the accuracy only; for instance, DOP and propagation effects affect the position estimate accuracy. In [121] the observed TDoA positioning errors were between 3 and 115 m. The positioning algorithm used in [121, 125] was designed by the author and is described in Appendix B.1. It is worth noting that only the transmitters in a single-frequency network (SFN) have to be synchronized. Generally, the reference oscillators of the DVB-T transmitters are very accurate and stable, nonetheless, they need not to be synchronized to common timescale. However, most of them exploit GNSS-disciplined oscillators and therefore are synchronized to a common timescale.

In North America, DVB-T is not adopted; instead, the ATSC-standard signal is used. The possible ways of exploitation of ATSC synchronization signals for localization purpose are shown e.g. in [128, 129].

Another group of systems that provide useful SoO are cell communication systems. The comparison with the broadcasting systems is ambivalent. On one hand, the equivalent radiated powers of the base station transmitters are usually bellow 100 W; on the other hand, they are much more densely placed. Any *a priori* information about the transmitters is barely available, it is necessary to rely on the on-the-fly estimation of the transmitter parameters. The exploitation of the OFDM-modulated LTE signals is described in [122], where the tests on a road vehicle and an unmanned aerial vehicle resulted in less than 10 m RMS error w.r.t. the GPS ground truth. The results have been acquired with the widest possible LTE bandwidth – 20 MHz.

The cellular CDMA signals were already exploited as well [130]. Such choice of SoO is reasonable, since it uses spectrum spreading by means of pseudorandom sequences, i.e. similar technique to the GNSS ranging codes. The chipping rate is approx. 1.2 Mchip/s, however the peak of the ACF is not sharp. In [130] a mean error w.r.t. GPS ground truth was approx. 5.5 m.

3. Inertial Navigation

The radionavigation almost exclusively exploits communication between infrastructure and the user equipment. It deteriorates or fails completely when such link is unreliable or broken, for instance line of sight is lost or propagation effects reach beyond the envelope of correction models.

In contrast, inertial navigation relies solely on monitoring of the *motion* and orientation of the navigated object. The motion is sensed by means of a set of inertial sensors commonly denoted as the inertial measurement unit (IMU). Such sensors do not rely on any external information, principally. Within this thesis, we will consider *strapdown* IMU[87] with tri-axial accelerometer and tri-axial gyroscope as a de-facto standard equipment. However, for certain applications the IMUs with less sensors are used. Often a tri-axial magnetometer is included as well. Such sensors are nowadays available in almost every smartphone, similarly as GNSS chipsets.

An *accelerometer* measures the *specific force* (a) projected onto its sensitive axis. From the tri-axial accelerometer, i.e. three orthogonally-mounted accelerometers, a specific force vector is available. A *gyroscope* (commonly abbreviated as gyro) measures *angular rate* (ω) along a single axis. It is convenient (and usually respected) to align the axes of gyroscopes with the axes of the accelerometers, as is shown in Fig. 3.1. The measurement vectors are defined w.r.t. the body reference frame (denoted by the $^{\mathrm{b}}$ superscript);

$$\boldsymbol{a}^{\mathrm{b}} = \begin{bmatrix} a_x^{\mathrm{b}} \\ a_y^{\mathrm{b}} \\ a_z^{\mathrm{b}} \end{bmatrix}, \qquad \boldsymbol{\omega}^{\mathrm{b}} = \begin{bmatrix} \omega_x^{\mathrm{b}} \\ \omega_y^{\mathrm{b}} \\ \omega_z^{\mathrm{b}} \end{bmatrix}. \tag{3.1}$$

The quality of the measurements depends on the sensor quality grade, naturally. The sensors are affected by both systematic and random errors. The IMU grades reflect the parameters of the individual sensors, in [132] the typical biases of gyros and accelerometers for each grade are presented, see Tab. 3.1.

Within the thesis we focus on the low-cost IMUs, in particular. They typically use MEMS (micro electro-mechanical systems) devices as the sensors, often manufactured on a single chip. The technologies for higher grade gyros are reviewed e.g in [133]. In the following text, we will cover the basic approach of exploiting strapdown IMU measurements for navigation near the vicinity of Earth's surface.

[87] The sensors of a strapdown IMU are fixed on the navigated body, unlike gimballed IMUs. According to [131], almost all new IMU designs are strapdown type.

Table 3.1.: IMU performance grades according to [132]

IMU Grade	Accelerometer bias	Gyroscope bias
Consumer/Commercial	>1 deg /s	>50 mg
Tactical	∼1 deg /h	∼ 1 mg
Navigation	0.01 deg /h	25 µg
Strategic	∼0.001 deg /h	∼ 1 µg

3.1. Coordinate Frames and Attitude Parametrization

Several coordiante frames and their relations need to be defined in order to be able to process and use the measurements of inertial sensors. For instance, the strapdown IMU is mounted on the navigated body, whilst the navigation information has to be bound to a reference frame of the environment or with a frame of the continuously rotating Earth. The basic set of frames will be described within the second part of this section.

The attitude of a body within a coordinate system or relative rotation of two coordinate frames can be expressed by means of various parametrizations, each advantageous in a certain way. In order to be able to describe the relations between the coordinate frames, we need to describe the common parametrizations and transformations in the first part of the section.

3.1.1. Attitude parametrization and rotation formalisms

I. Euler angles

Probably the most intuitive way of describing an object attitude are the Euler angles. These define the amount of rotation around the axes, of a body (see Fig. 3.1 for the graphical reference): yaw ψ around the down-pointing z-axis; pitch θ around the right-pointing y-axis; roll ϕ around the forward-pointing x-axis.[88] It is obvious that the order of rotations does matter. The pitch is constrained to values between $\pm 90°$, in order to suppress ambiguous representation of a single attitude.

The representation can be used for any two frames, however it is convenient to introduce it as a transformation from navigation (n-frame, in the subscript) to the body frame (b-frame, in the superscript):[89] According ti [131], the vector of Euler angles follows

$$\boldsymbol{\psi}_{\mathrm{n}}^{\mathrm{b}} = \begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix}, \tag{3.2}$$

[88] This definition is coherent with Tait-Brian angles, however an imprecise denotation as Euler angles has became standard. The "original" Euler angles are defined with different axes of rotation, however, the information contained is equivalent.

[89] The frames will be defined below in Sec. 3.1.2.

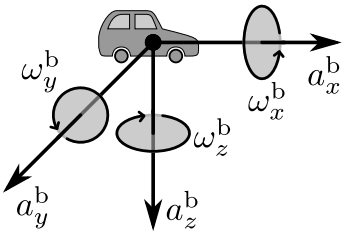


FIGURE 3.1.: IMU measurement axes.

where the order of rotations goes from bottom to top of the vector. The yaw is applied first, then the pitch is applied (around the body-frame axis) and the roll is applied as the last one. For the inverse rotation, the order of the rotations has to be reversed. It is necessary to remark that the euler angles always represent the transformation from the global frame to the local frame and not vice versa.

The problem of such attitude parametrization is the *gimbal lock* – situation when one degree of freedom is lost. This happens when the pitch angle is $\pm 90°$; then the yaw and roll axes become identical (vertical). Moreover, the stacking of rotations, i.e. combining multiple rotations cannot be done in a straightforward way, since for arbitrary frames a,b,c

$$\boldsymbol{\psi}_{\mathrm{a}}^{\mathrm{c}} \neq \boldsymbol{\psi}_{\mathrm{b}}^{\mathrm{c}} + \boldsymbol{\psi}_{\mathrm{a}}^{\mathrm{b}}. \tag{3.3}$$

II. Rotation matrix

In the 3-dimensional space the rotation matrix performs rotation from one coordinate frame to another frame by means of multiplication. The *from*-frame is again denoted in the subscript and the *to*-frame is indicated in the superscript, i.e. for transformation of vector $\boldsymbol{x}$ from the a-frame to the b-frame it holds

$$\boldsymbol{x}^{\mathrm{b}} = \boldsymbol{R}_{\mathrm{a}}^{\mathrm{b}} \boldsymbol{x}^{\mathrm{a}}. \tag{3.4}$$

Any rotation matrix should not change the length of the transformed vector, i.e. the determinant is always $\det \boldsymbol{R} = 1$. The inverse rotation can be performed by matrix division – multiplication by the inverse matrix. For arbitrary rotation matrix it holds

$$\boldsymbol{R}^{-1} = \boldsymbol{R}^{\mathrm{T}}, \tag{3.5}$$

therefore, it can be stated that the inverse rotation follows

$$\boldsymbol{R}_{\mathrm{b}}^{\mathrm{a}} = (\boldsymbol{R}_{\mathrm{a}}^{\mathrm{b}})^{\mathrm{T}}. \tag{3.6}$$

The transformation between multiple frames can be stacked due to the associativity of the matrix multiplication:

$$\boldsymbol{R}_{\mathrm{a}}^{\mathrm{c}} = \boldsymbol{R}_{\mathrm{b}}^{\mathrm{c}} \boldsymbol{R}_{\mathrm{a}}^{\mathrm{b}}. \tag{3.7}$$

However, the order of the matrices does matter, since the multiplication is non-commutative.

Relation to Euler angles

It is straightforward to derive the frame transformation matrix that corresponds to a specific set of Euler angles. Again, we will demonstrate the conversion on the transformation from the n-frame to the b-frame, provided that it can be generalized to two arbitrary frames. The three required rotations are stacked using (3.7), so that

$$\begin{aligned}\boldsymbol{R}_{\mathrm{n}}^{\mathrm{b}} &= \boldsymbol{R}\{\boldsymbol{\psi}_{\mathrm{n}}^{\mathrm{b}}\} = \boldsymbol{R}_{\phi}\boldsymbol{R}_{\theta}\boldsymbol{R}_{\psi} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \cos\theta\cos\psi & \cos\theta\sin\psi & -\sin\theta \\ \sin\phi\sin\theta\cos\psi - \cos\phi\sin\psi & \sin\phi\sin\theta\sin\psi + \cos\phi\cos\psi & \sin\phi\cos\theta \\ \cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi & \cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi & \cos\phi\cos\theta \end{bmatrix}.\end{aligned} \tag{3.8}$$

It is worth noting that $\boldsymbol{R}\{\boldsymbol{\psi}_{\mathbf{n}}^{\mathbf{b}}\}$ denotes an operator that converts a set of Euler angles to the rotation matrix. According to [131, 134], the inverse operator can be defined so that

$$\begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix} = \boldsymbol{\psi}\{\boldsymbol{R}_{\mathrm{n}}^{\mathrm{b}}\} = \begin{bmatrix} \tan^{-1}\dfrac{\boldsymbol{R}_{\mathrm{n}[2,3]}^{\mathrm{b}}}{\boldsymbol{R}_{\mathrm{n}[3,3]}^{\mathrm{b}}} \\ \tan^{-1}\dfrac{-\boldsymbol{R}_{\mathrm{n}[1,3]}^{\mathrm{b}}}{\sqrt{(\boldsymbol{R}_{\mathrm{n}[2,3]}^{\mathrm{b}})^2+(\boldsymbol{R}_{\mathrm{n}[3,3]}^{\mathrm{b}})^2}} \\ \tan^{-1}\dfrac{\boldsymbol{R}_{\mathrm{n}[1,2]}^{\mathrm{b}}}{\boldsymbol{R}_{\mathrm{n}[1,1]}^{\mathrm{b}}} \end{bmatrix} = \begin{bmatrix} \tan^{-1}\dfrac{\boldsymbol{R}_{\mathrm{n}[2,3]}^{\mathrm{b}}}{\boldsymbol{R}_{\mathrm{n}[3,3]}^{\mathrm{b}}} \\ -\sin^{-1}\boldsymbol{R}_{\mathrm{n}[1,3]}^{\mathrm{b}} \\ \tan^{-1}\dfrac{\boldsymbol{R}_{\mathrm{n}[1,2]}^{\mathrm{b}}}{\boldsymbol{R}_{\mathrm{n}[1,1]}^{\mathrm{b}}} \end{bmatrix}. \tag{3.9}$$

Near the singularity of gimbal lock the resolution of ϕ and ψ becomes practically impossible.[90]

III. Quaternions

The following text describes the basic properties of quaternions that are relevant for their usage for attitude parametrization, performing rotations and frame transformations. The description of quaternion algebra and the relation of quaternions to the other parametrization is based mostly on [134, 135]; the notation builds on [135], mainly. The advantage of the quaternion parametrization is that it does not suffer from the gimbal lock, unlike the Euler-angles parametrization. However, quaternion interpretation is not as straightforward.

[90] In [134] it is stated that in such case only the linear combinations of ϕ and ψ can be computed, the relevant equations are provided as well.

Quaternion algebra

A quaternion is commonly defined as a special column 4-vector:

$$\mathring{\boldsymbol{q}} \overset{def.}{\Leftrightarrow} \begin{bmatrix} q_w \\ \boldsymbol{q}_{xyz} \end{bmatrix} = \begin{bmatrix} q_w \\ \boldsymbol{q}_x \mathrm{i} \\ \boldsymbol{q}_y \mathrm{j} \\ \boldsymbol{q}_z \mathrm{k} \end{bmatrix} \tag{3.10}$$

$$\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = \mathrm{ijk} = -1. \tag{3.11}$$

The sum of two quaternions is performed as a simple sum of two vectors, therefore it is commutative and associative. The quaternion product, however, holds

$$\mathring{\boldsymbol{p}} \otimes \mathring{\boldsymbol{q}} = \begin{bmatrix} p_w q_w - \boldsymbol{p}_{xyz}^{\mathrm{T}} \boldsymbol{q}_{xyz} \\ p_w \boldsymbol{q}_{xyz} + q_w \boldsymbol{p}_{xyz} + \boldsymbol{p}_{xyz} \times \boldsymbol{q}_{xyz} \end{bmatrix} = \begin{bmatrix} p_w q_w - p_x q_x - p_y q_y - p_z q_z \\ p_w q_x + p_x q_w + p_y q_z - p_z q_y \\ p_w q_y - p_x q_z + p_y q_w - p_z q_x \\ p_w q_z + p_x q_y - p_y q_x - p_z q_w \end{bmatrix}, \tag{3.12}$$

where $\times$ denotes vector cross-product. Obviously, the quaternion product is non-commutative, though, it is associative and distributive over summation. The conjugate of the quaternion is defined as

$$\mathring{\boldsymbol{q}}^* = \begin{bmatrix} q_w \\ -\boldsymbol{q}_{xyz} \end{bmatrix}, \tag{3.13}$$

and it holds

$$\mathring{\boldsymbol{q}} \otimes \mathring{\boldsymbol{q}}^* = \mathring{\boldsymbol{q}}^* \otimes \mathring{\boldsymbol{q}} = q_w^2 + \boldsymbol{q}_{xyz} \cdot \boldsymbol{q}_{xyz} \tag{3.14}$$

$$(\mathring{\boldsymbol{p}} \otimes \mathring{\boldsymbol{q}})^* = \mathring{\boldsymbol{q}}^* \otimes \mathring{\boldsymbol{p}}^*. \tag{3.15}$$

The quaternion norm can be evaluated as follows

$$\|\mathring{\boldsymbol{q}}\| = \sqrt{\mathring{\boldsymbol{q}}^* \otimes \mathring{\boldsymbol{q}}} = \sqrt{q_w^2 + q_x^2 + q_y^2 + q_z^2}. \tag{3.16}$$

It can be shown that $\|\mathring{\boldsymbol{p}} \otimes \mathring{\boldsymbol{q}}\| = \|\mathring{\boldsymbol{p}}\| \, \|\mathring{\boldsymbol{q}}\|$. The norm is important for evaluation of the inverse quaternion, which satisfies the condition

$$\mathring{\boldsymbol{q}} \otimes \mathring{\boldsymbol{q}}^{-1} = \mathring{\boldsymbol{q}}^{-1} \otimes \mathring{\boldsymbol{q}} = \mathring{\mathbf{1}} \tag{3.17}$$

$$\mathring{\boldsymbol{q}}^{-1} = \frac{\mathring{\boldsymbol{q}}^*}{\|\mathring{\boldsymbol{q}}\|^2} \tag{3.18}$$

where $\mathring{\mathbf{1}} = [1, 0, 0, 0]^{\mathrm{T}}$ is the identity quaternion, for which $\mathring{\boldsymbol{q}} = \mathring{\mathbf{1}} \otimes \mathring{\boldsymbol{q}} = \mathring{\boldsymbol{q}} \otimes \mathring{\mathbf{1}}$. For the normalized quaternions the inverse and conjugate operations are equivalent, obviously, i.e.

$$\|\mathring{\boldsymbol{q}}\| = 1 \quad \Longrightarrow \quad \mathring{\boldsymbol{q}}^{-1} = \mathring{\boldsymbol{q}}^*. \tag{3.19}$$

This is rather important, because all the rotation quaternions are normalized.

The rotation of a vector is performed using

$$\boldsymbol{r}' = \mathring{\boldsymbol{q}} \otimes \boldsymbol{r} \otimes \mathring{\boldsymbol{q}}^* \tag{3.20}$$

where $\boldsymbol{x}$ is the column vector to be rotated.[91] Similarly, the transformation of the reference frame from a-frame to b-frame for a vector $\boldsymbol{r}$ is performed by means of

$$\boldsymbol{r}^{\mathrm{b}} = \mathring{\boldsymbol{q}}_{\mathrm{a}}^{\mathrm{b}} \otimes \boldsymbol{r}^{\mathrm{a}} \otimes (\mathring{\boldsymbol{q}}_{\mathrm{a}}^{\mathrm{b}})^{*}. \tag{3.21}$$

Using the (3.19) property, the inverse transformation follows

$$\boldsymbol{r}^{\mathrm{a}} = \mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{a}} \otimes \boldsymbol{r}^{\mathrm{b}} \otimes (\mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{a}})^{*} = (\mathring{\boldsymbol{q}}_{\mathrm{a}}^{\mathrm{b}})^{*} \otimes \boldsymbol{r}^{\mathrm{b}} \otimes \mathring{\boldsymbol{q}}_{\mathrm{a}}^{\mathrm{b}}. \tag{3.22}$$

The associativity of the quaternion product implies that the rotations, or frame transformations, are stackable, i.e.

$$\boldsymbol{r}^{\mathrm{c}} = \mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{c}} \otimes \mathring{\boldsymbol{q}}_{\mathrm{a}}^{\mathrm{b}} \otimes \boldsymbol{r}^{\mathrm{a}} \otimes (\mathring{\boldsymbol{q}}_{\mathrm{a}}^{\mathrm{b}})^{*} \otimes (\mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{c}})^{*}; \qquad \mathring{\boldsymbol{q}}_{\mathrm{a}}^{\mathrm{c}} = \mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{c}} \otimes \mathring{\boldsymbol{q}}_{\mathrm{a}}^{\mathrm{b}}. \tag{3.23}$$

A more detailed description of quaternion algebra, its connection to rotation and a review of a few alternative notations is available in [135].

Relation to rotation matrix

The representation of a quaternion in terms of rotation matrix follows [134]

$$\boldsymbol{R}\{\mathring{\boldsymbol{q}}\} = \begin{bmatrix} q_w^2 + q_x^2 - q_y^2 - q_z^2 & 2(q_x q_y - q_w q_z) & 2(q_x q_z + q_w q_y) \\ 2(q_x q_y + q_w q_z) & q_w^2 - q_x^2 + q_y^2 - q_z^2 & 2(q_y q_z - q_w q_x) \\ 2(q_x q_z - q_w q_y) & 2(q_y q_z + q_w q_x) & q_w^2 - q_x^2 - q_y^2 + q_z^2 \end{bmatrix} \tag{3.24}$$

The conversion of a rotation matrix to a quaternion form is more complicated. First, the values

$$\begin{aligned} P_1 &= 1 + \mathrm{tr}(\boldsymbol{R}) & P_2 &= 1 + 2R_{1,1} + \mathrm{tr}(\boldsymbol{R}) \\ P_3 &= 1 + 2R_{2,2} + \mathrm{tr}(\boldsymbol{R}) & P_4 &= 1 + 2R_{3,3} + \mathrm{tr}(\boldsymbol{R}) \end{aligned} \tag{3.25}$$

are computed. The $R_{i,j}$ are the elements of the rotation matrix and tr() denotes the trace of a matrix. The quaternion components are then computed depending on the maximum of the P_1 to P_4 values, as shown in Tab. 3.2 below.

Relation to Euler angles

According to [134], the conversion from Euler angles to quaternion follows

$$\mathring{\boldsymbol{q}}\{\boldsymbol{\psi}\} = \begin{bmatrix} \cos\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} + \sin\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} \\ \sin\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} - \cos\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} \\ \cos\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} + \sin\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} \\ \cos\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} - \sin\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} \end{bmatrix}. \tag{3.26}$$

[91] In a strict sense such notation used is incorrect, since we have not defined the product of vector and quaternion. In (3.20) the vector is inherently converted to quaternion and vice versa by attaching/detaching a zero-valued scalar component:

$$\mathring{\boldsymbol{x}} = \begin{bmatrix} 0 \\ \boldsymbol{x} \end{bmatrix}.$$

Although formally incorrect, such notation is very convenient, as the "vector-quaternions" are easily distinguished from the "rotation-quaternions."

TABLE 3.2.: Conversion of rotation matrix to quaternion

If:	**Then:**	$q_w =$	$q_x =$	$q_y =$	$q_z =$
$\max(P_i) = P_1$:		$\frac{\sqrt{P_1}}{2}$	$\frac{R_{3,2}-R_{2,3}}{2\sqrt{P_1}}$	$\frac{R_{1,3}-R_{3,1}}{2\sqrt{P_1}}$	$\frac{R_{2,1}-R_{1,2}}{2\sqrt{P_1}}$
$\max(P_i) = P_2$:		$\frac{R_{3,2}-R_{2,3}}{2\sqrt{P_2}}$	$\frac{\sqrt{P_2}}{2}$	$\frac{R_{2,1}+R_{1,2}}{2\sqrt{P_2}}$	$\frac{R_{1,3}+R_{3,1}}{2\sqrt{P_2}}$
$\max(P_i) = P_3$:		$\frac{R_{1,3}-R_{3,1}}{2\sqrt{P_3}}$	$\frac{R_{2,1}+R_{1,2}}{2\sqrt{P_3}}$	$\frac{\sqrt{P_3}}{2}$	$\frac{R_{3,2}+R_{2,3}}{2\sqrt{P_3}}$
$\max(P_i) = P_4$:		$\frac{R_{2,1}-R_{1,2}}{2\sqrt{P_4}}$	$\frac{R_{1,3}+R_{3,1}}{2\sqrt{P_4}}$	$\frac{R_{3,2}+R_{2,3}}{2\sqrt{P_4}}$	$\frac{\sqrt{P_4}}{2}$

The inverse conversion, from quaternion to Euler angle, belongs to the most important ones, since it allows us to visualize the quaternion-based attitude in a human readable form. The operation is described in [136] and holds

$$\boldsymbol{\psi}\{\mathring{\boldsymbol{q}}\} = \begin{bmatrix} \phi\{\mathring{\boldsymbol{q}}\} \\ \theta\{\mathring{\boldsymbol{q}}\} \\ \psi\{\mathring{\boldsymbol{q}}\} \end{bmatrix} = \begin{bmatrix} \tan^{-1} \frac{2(q_w q_x + q_y q_z)}{1-2\left(q_x^2+q_y^2\right)} \\ \sin^{-1}\left(2q_w q_y - 2q_x q_z\right) \\ \tan^{-1} \frac{2(q_w q_z + q_x q_y)}{1-2\left(q_y^2+q_z^2\right)} \end{bmatrix}. \tag{3.27}$$

Sometimes it is required to convert a covariance matrix of the Euler angles to quaternion covariance matrix (e.g. for initialization using *a priori* known attitude) or vice versa (e.g. for visualization). For reasonably small variances this can be performed by means of linearized operators (3.26) and (3.27), i.e. the Jacobian matrices of the transforming functions. The exact relations are provided as (3.28) and (3.29) on page 94, respectively.

IV. Rotation vector

The rotation vector $\boldsymbol{\nu}$ defines the axis of rotation by its direction (unitary vector $\boldsymbol{u}$) and the angle of rotation ϕ by its magnitude:

$$\boldsymbol{\nu} = \phi \boldsymbol{u}. \tag{3.30}$$

In order to perform transformation of vector $\boldsymbol{r}$ to $\boldsymbol{r}'$ using a rotation vector $\boldsymbol{\nu}$, it has to be decomposed first into parts parallel $\boldsymbol{r}_\parallel$, and perpendicular $\boldsymbol{r}_\perp$ to $\boldsymbol{\nu}$ (and $\boldsymbol{u}$, implicitly). The parallel component is obtained by means of a simple projection; the perpendicular one is then computed as well:

$$\boldsymbol{r}_\parallel = \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{r}, \qquad \boldsymbol{r}_\perp = \boldsymbol{r} - \boldsymbol{r}_\parallel. \tag{3.31}$$

According to [135], the vector rotation formula follows

$$\boldsymbol{r}' = \boldsymbol{r}_\parallel + \boldsymbol{r}_\perp \cos\phi + \boldsymbol{u} \times \boldsymbol{x} \sin\phi. \tag{3.32}$$

The rotation operation can be performed by means of converting to rotation matrix or quaternion as well.

Conversion of covariance $\boldsymbol{\psi} \to \mathring{\boldsymbol{q}}$:

$$\operatorname{cov} \mathring{\boldsymbol{q}}\{\boldsymbol{\psi}, \operatorname{cov} \boldsymbol{\psi}\} = \boldsymbol{J}_{\mathring{\boldsymbol{q}}\{\boldsymbol{\psi}\}} (\operatorname{cov} \boldsymbol{\psi}) \boldsymbol{J}^{\mathrm{T}}_{\mathring{\boldsymbol{q}}\{\boldsymbol{\psi}\}}, \tag{3.28}$$

where the Jacobian matrix follows

$$\boldsymbol{J}_{\mathring{\boldsymbol{q}}\{\boldsymbol{\psi}\}} = \frac{1}{2} \begin{bmatrix} -\sin\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} + \cos\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} & -\cos\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} + \sin\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} & -\cos\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} + \sin\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} \\ +\cos\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} + \sin\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} & -\sin\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} - \cos\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} & -\sin\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} - \cos\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} \\ -\sin\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} + \cos\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} & +\cos\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} - \sin\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} & -\cos\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} + \sin\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} \\ -\sin\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} - \cos\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} & -\cos\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} - \sin\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} & +\cos\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} + \sin\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} \end{bmatrix}.$$

Conversion of covariance $\mathring{\boldsymbol{q}} \to \boldsymbol{\psi}$:

$$\operatorname{cov} \boldsymbol{\psi}\{\mathring{\boldsymbol{q}}, \operatorname{cov} \mathring{\boldsymbol{q}}\} = \boldsymbol{J}_{\boldsymbol{\psi}\{\mathring{\boldsymbol{q}}\}} (\operatorname{cov} \mathring{\boldsymbol{q}}) \boldsymbol{J}^{\mathrm{T}}_{\boldsymbol{\psi}\{\mathring{\boldsymbol{q}}\}}, \tag{3.29}$$

where the Jacobian matrix follows

$$\boldsymbol{J}_{\boldsymbol{\psi}\{\mathring{\boldsymbol{q}}\}} = \begin{bmatrix} \frac{2q_x(1-2(q_x^2+q_y^2))}{(2q_x^2+2q_y^2-1)^2+(2q_wq_x+2q_yq_z)^2} & \frac{2q_w(1-2(q_x^2+q_y^2))+8q_x(q_wq_x+q_yq_z)}{(2q_x^2+2q_y^2-1)^2+(2q_wq_x+2q_yq_z)^2} & \frac{2q_z(1-2(q_x^2+q_y^2))+8q_y(q_wq_x+q_yq_z)}{(2q_x^2+2q_y^2-1)^2+(2q_wq_x+2q_yq_z)^2} & \frac{2q_y(1-2(q_x^2+q_y^2))}{D_\phi} \\ \frac{2y}{\sqrt{(1-4(q_wq_y-q_xq_z)^2)}} & \frac{-2z}{\sqrt{(1-4(q_wq_y-q_xq_z)^2)}} & \frac{2q}{\sqrt{(1-4(q_wq_y-q_xq_z)^2)}} & \frac{-2x}{\sqrt{(1-4(q_wq_y-q_xq_z)^2)}} \\ \frac{2q_z(1-2(q_y^2+q_z^2))}{(2q_y^2+2q_z^2-1)^2+(2q_wq_z+2q_xq_y)^2} & \frac{2q_y(1-2(q_y^2+q_z^2))}{(2q_y^2+2q_z^2-1)^2+(2q_wq_z+2q_xq_y)^2} & \frac{2q_x(1-2(q_y^2+q_z^2))+8q_y(q_wq_z+q_xq_y)}{(2q_y^2+2q_z^2-1)^2+(2q_wq_z+2q_xq_y)^2} & \frac{2q_w(1-2(q_y^2+q_z^2))+8q_z(q_wq_z+q_xq_y)}{(2q_y^2+2q_z^2-1)^2+(2q_wq_z+2q_xq_y)^2} \end{bmatrix}.$$

Relation to rotation matrix

According to [134, 135], the rotation matrix can be obtained as

$$\boldsymbol{R}\{\boldsymbol{\nu}\} = \mathbb{I} + \sin\phi\,[\boldsymbol{u}]_\times + (1-\cos\phi)\,[\boldsymbol{u}]_\times^2 = \mathbb{I} + \sin\phi\,[\boldsymbol{u}]_\times + (1-\cos\phi)\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} - \mathbb{I}, \quad (3.33)$$

where the skew operator $[]_\times$ is defined by

$$[\boldsymbol{a}]_\times = \begin{bmatrix} 0 & -a_z & +a_y \\ +a_z & 0 & -a_x \\ -a_y & +a_x & 0 \end{bmatrix}; \qquad \boldsymbol{a}\times\boldsymbol{b} = [\boldsymbol{a}]_\times\,\boldsymbol{b}. \quad (3.34)$$

The skew operator is often denoted as a cross operator, since it emulates vector cross product by means of matrix multiplication.

For small angles of rotation it can be assumed that $\sin\phi \approx \phi$ and $\cos\phi \approx 1$, thus the conversion can be further simplified to

$$\boldsymbol{R}\{\boldsymbol{\nu}\} \approx \mathbb{I} + \phi\,[\boldsymbol{u}]_\times = \mathbb{I} + [\boldsymbol{\nu}]_\times\,. \quad (3.35)$$

The rotation vector that corresponds to a given matrix can be found by means of eigendecomposition. The 3-dimensional rotation matrix has always 3 eigenvalues $\lambda_{\boldsymbol{R}} = \{1; e^{+\mathrm{j}\phi}; e^{-\mathrm{j}\phi}\}$. The axis of rotation $\boldsymbol{u}$ is the normalized eigenvector corresponding to the 1-eigenvalue; obtaining ϕ is straightforward.

Relation to quaternion

Rotation vector $\boldsymbol{\nu}$ from (3.30) can be converted to quaternion using [134, 135]

$$\mathring{\boldsymbol{q}}\{\boldsymbol{\nu}\} = \begin{bmatrix} \cos\frac{\|\boldsymbol{\nu}\|}{2} \\ \frac{\boldsymbol{\nu}}{\|\boldsymbol{\nu}\|}\sin\frac{\|\boldsymbol{\nu}\|}{2} \end{bmatrix} = \begin{bmatrix} \cos\frac{\|\phi\|}{2} \\ \boldsymbol{u}\sin\frac{\|\phi\|}{2} \end{bmatrix}. \quad (3.36)$$

The inverse operation holds

$$\phi = 2\tan^{-1}\frac{\|\boldsymbol{q}_{xyz}\|}{q_w}, \quad \boldsymbol{u} = \frac{\boldsymbol{q}_{xyz}}{\|\boldsymbol{q}_{xyz}\|} \quad \Longrightarrow \quad \boldsymbol{\nu}\{\mathring{\boldsymbol{q}}\} = 2\frac{\boldsymbol{q}_{xyz}}{\|\boldsymbol{q}_{xyz}\|}\tan^{-1}\frac{\|\boldsymbol{q}_{xyz}\|}{q_w}. \quad (3.37)$$

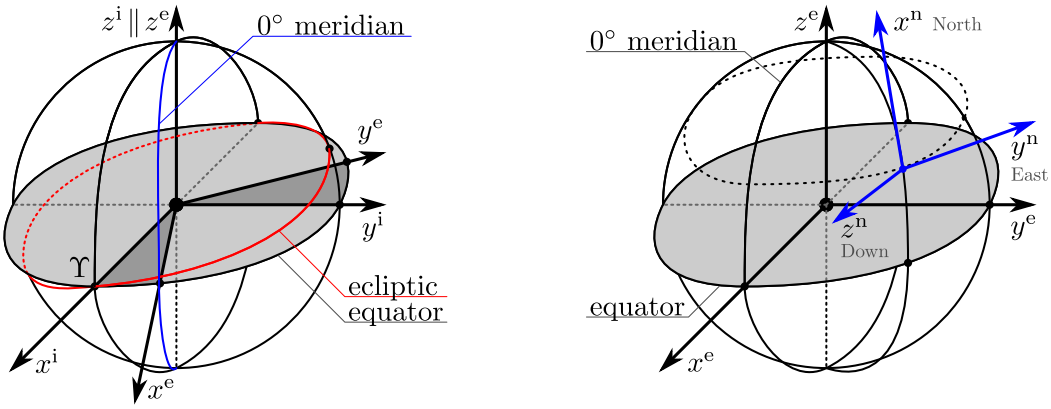
3.1.2. Coordinate frames

All the described frames are orthonormal, Cartesian coordinate frames. The order of the axes x, y and z follows the right-hand rule. In the description of the frames we will progress gradually towards the navigated body. The frame which the axis or vector relates to is denoted by means of superscript in the following text.

I. Earth-centered inertial (ECI) frame

The inertial measurements are always taken with respect to the ideal inertial frame. It is rather inconvenient to use a strictly inertial frame, therefore quasi-inertial ECI frame is defined to be the i-frame. The origin of the i-frame is aligned with the center of Earth, the z^{i}-axis coincides with the north semi-axis of the planet, as is shown in Fig. 3.2a. The ECI frame does not rotate with Earth, it is non-rotating w.r.t. distant galaxies. The x^{i}-axis is defined so that it passes through the vernal equinox[92] (labeled by Υ in the figure). The y^{i}-axis completes the right-handed system.

[92] Vernal equinox is the ascending node of the ecliptic and the equator, i.e. the point where the sun passes the equator from the Southern to the Northern hemisphere.



(A) ECI and ECEF frames.

(B) ECEF and NED frames.

FIGURE 3.2.: Relations of coordinate frames.

II. Earth-centered Earth-fixed (ECEF) frame

The origin and the z^e-axis of the ECEF, or e-frame coincide with the origin and z^i-axis of the i-frame. The Earth-fixed feature of the e-frame implies that it rotates with the planet's body – the x^e-axis passes through the intersection of the equator and Greenwich meridian, i.e. point with zero latitude and longitude; the y^i-axis completes the right-handed system.

The e-frame rotates with respect to the i-frame along the common z-axis with a rate ($\Omega_E = 15\,^\circ\,h^{-1}$). Obviously, the low-cost, consumer grade gyroscopes are often not able to reliably sense the Earth's rotation. In such cases it can be assumed that the inertial measurements are taken w.r.t. the ECEF frame, and the complexity of the computations may be reduced.

Geodetic representation

Near the vicinity of the Earth's surface the Cartesian coordinates cannot be considered human-understandable. It is far more useful to provide *latitude*, *longitude* and *height* which will be further denoted as φ, λ, and h, respectively. The geodetic system used by most GNSS is the WGS-84. The surface is approximated by the reference ellipsoid, which is symmetric w.r.t. the axis of Earth's rotation and w.r.t. the equatorial plane. The height is measured on the normal of the ellipsoid surface, as shown on the meridian-plane cut in Fig. 3.3.

It is curious that vernal equinox (or the x^i-axis) points towards the Pisces constellation, while the symbol ϒ is used for Aries. The reason is that the equinox has passed from Aries to Pisces in the the 1st century AD [137].

The conversion from the geodetic to the Cartesian ECEF frame is straightforward and unambiguous [22]:

$$\boldsymbol{r}^{\mathrm{n}} = \begin{bmatrix} r_x^{\mathrm{e}} \\ r_y^{\mathrm{e}} \\ r_z^{\mathrm{e}} \end{bmatrix} = \begin{bmatrix} \frac{a_{\mathrm{E}} \cos\lambda}{\sqrt{1+\left(1-e_{\mathrm{E}}^2\right)\tan^2\varphi}} + h\cos\lambda\cos\varphi \\ \frac{a_{\mathrm{E}} \sin\lambda}{\sqrt{1+\left(1-e_{\mathrm{E}}^2\right)\tan^2\varphi}} + h\sin\lambda\cos\varphi \\ \frac{a_{\mathrm{E}}\left(1-e_{\mathrm{E}}^2\right)\sin\varphi}{\sqrt{1-e_{\mathrm{E}}^2\sin^2\varphi}} + h\sin\varphi \end{bmatrix}. \tag{3.38}$$

The WGS-84 reference ellipsoid is defined by major semi-axis $a_{\mathrm{E}} = 6378.137\,\mathrm{km}$ and the squared eccentricity[93] $e_{\mathrm{E}}^2 = 0.006\,694\,379\,990\,14$. Note that the geodetic height (above the ellipsoid) is different from the height above geoid, or mean sea level; the difference between the two values is commonly referred as *geoid undulation*.

The conversion from cartesian system to the geodetic one is more complicated. The longitude may be determined simply by means of a four-quadrant arctangent function

$$\lambda = \tan^{-1}\frac{r_x^{\mathrm{e}}}{r_y^{\mathrm{e}}}, \tag{3.39}$$

however, the search for the ellipsoid normal that intersects the coordinate to be transformed is nontrivial. The iterative Bowring method [22] may be used for this purpose:

1. Initialize:

$$p = \sqrt{x^2+y^2} \qquad\qquad \tan u = \frac{z}{p}\cdot\frac{a}{b}$$

2. Iterate until $\tan u$ converges:

$$\cos^2 u = \frac{1}{1+\tan^2 u} \qquad\qquad \sin^2 u = 1-\cos^2 u$$
$$\tan\varphi = \frac{z+e'^2 b\sin^3 u}{p-e^2 a\cos^3 u} \qquad\qquad \tan u = \frac{b}{a_{\mathrm{E}}}\tan\varphi$$

3. Evaluate latitude and height above ellipsoid:

$$N = \frac{a_{\mathrm{E}}}{\sqrt{1-e_{\mathrm{E}}^2\sin^2\varphi}} \qquad h = \begin{cases} \frac{p}{\cos\varphi} - N & \text{for } \varphi \neq \pm\frac{\pi}{2} \\ \frac{z}{\sin\varphi} - N - e_{\mathrm{E}}^2 N & \text{for } \varphi \neq 0 \end{cases}$$

[93] According to [22], eccentricity e and the second eccentricity e' are determined by

$$e^2 = 1-\frac{b^2}{a^2} \qquad\qquad e'^2 = \frac{a^2}{b^2} - 1 = \left(\frac{a}{b}e\right).$$

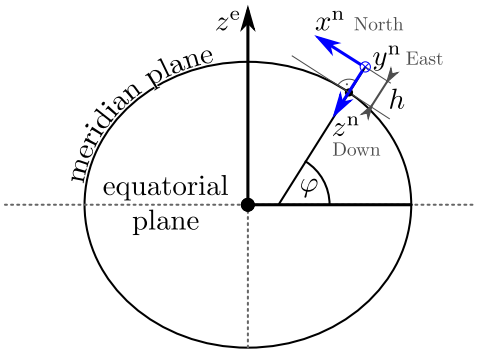


FIGURE 3.3.: Latitude in the meridian plane of the reference ellipsoid.

III. Local, navigational (NED) frame

The navigation frame (n-frame) is a local frame in the vicinity of navigated body. The x^{n}-axis points towards North, the y^{n}-axis points towards East, and the z^{n}-axis points down. An alternative system ENU (East-North-Up) is sometimes used, especially in the GNSS context; transformation between NED and ENU is straightforward.

The transformation of a point from the ECEF coordinates to the NED coordinates (and vice versa) requires translation and rotation:

$$\boldsymbol{r}^{\mathrm{n}} = \boldsymbol{R}^{\mathrm{n}}_{\mathrm{e}} \left(\boldsymbol{r}^{\mathrm{e}} - \boldsymbol{o}^{\mathrm{e}}_{\mathrm{n}} \right) \tag{3.40}$$

$$\boldsymbol{r}^{\mathrm{e}} = \boldsymbol{R}^{\mathrm{e}}_{\mathrm{n}} \boldsymbol{r}^{\mathrm{e}} + \boldsymbol{o}^{\mathrm{e}}_{\mathrm{n}} \tag{3.41}$$

$$\boldsymbol{R}^{\mathrm{n}}_{\mathrm{e}} = \left(\boldsymbol{R}^{\mathrm{e}}_{\mathrm{n}} \right)^{\mathrm{T}} = \begin{bmatrix} -\cos\lambda\sin\varphi & -\sin\lambda\sin\varphi & \cos\varphi \\ -\sin\lambda & \cos\lambda & 0 \\ -\cos\lambda\cos\varphi & -\sin\lambda\cos\varphi & -\sin\varphi \end{bmatrix}, \tag{3.42}$$

where $\boldsymbol{o}^{\mathrm{e}}_{\mathrm{n}}$ denotes the origin of the n-frame expressed in the e-frame; $\boldsymbol{R}^{\mathrm{e}}_{\mathrm{n}}$ is the rotation matrix[94] from the e-frame to the n-frame.

It is obvious that both geodetic and cartesian representations of the n-frame origin have to be known, since latitude and longitude values define the rotation. Remarkably, due to the "flatness" of the Earth (the reference ellipsoid) the vertical axis does not intersect the center of the Earth, which is the origin of the e-frame, unless latitude is 0° or ±90°.

Similarly as in case of the e-frame, the n-frame may be considered to be the inertial frame when the IMU is unable to sense the Earth's rotation.

Often, the origin of the n-frame is defined so that it coincides with the origin of the body frame (b-frame, see following section). In this thesis, however, the origins do not have to coincide perfectly, the body is moving in the n-frame that is stationary w.r.t.

[94] It is not necessary to use the rotation matrix notation; for example quaternion describing the same rotation as $\boldsymbol{R}^{\mathrm{e}}_{\mathrm{n}}$ holds

$$\mathring{\boldsymbol{q}}^{\mathrm{n}}_{\mathrm{e}} = \begin{bmatrix} \cos\left(-\frac{\pi}{4} - \frac{\varphi}{2}\right)\cos\frac{\lambda}{2} \\ \sin\left(-\frac{\pi}{4} - \frac{\varphi}{2}\right)\sin\frac{\lambda}{2} \\ -\sin\left(-\frac{\pi}{4} - \frac{\varphi}{2}\right)\cos\frac{\lambda}{2} \\ -\cos\left(-\frac{\pi}{4} - \frac{\varphi}{2}\right)\sin\frac{\lambda}{2} \end{bmatrix}.$$

the e-frame.[95] Such simplification may be understood as a flat-Earth approximation in the vicinity of the navigated body.[96]

IV. Body frame

For simplicity, we will consider the b-frame to be the common frame for the navigated body and the strapdown IMU – i.e. the axes and the reference point of the IMU coincide with the centre of mass and principal axes of the body, as shown in Fig. 3.1.

The transformation of vectors between b-frame and the n-frame follow:

$$\boldsymbol{r}^{\mathrm{b}} = \boldsymbol{R}_{\mathrm{n}}^{\mathrm{b}} \left(\boldsymbol{r}^{\mathrm{n}} - \boldsymbol{o}_{\mathrm{b}}^{\mathrm{n}} \right), \tag{3.43}$$

$$\boldsymbol{r}^{\mathrm{n}} = \boldsymbol{R}_{\mathrm{b}}^{\mathrm{n}} \boldsymbol{r}^{\mathrm{b}} + \boldsymbol{o}_{\mathrm{b}}^{\mathrm{n}}, \tag{3.44}$$

in the rotation matrix rotation, and

$$\boldsymbol{r}^{\mathrm{b}} = \mathring{\boldsymbol{q}}_{\mathrm{n}}^{\mathrm{b}} \otimes (\boldsymbol{r}^{\mathrm{n}} - \boldsymbol{o}_{\mathrm{b}}^{\mathrm{n}}) \otimes (\mathring{\boldsymbol{q}}_{\mathrm{n}}^{\mathrm{b}})^{*}, \tag{3.45}$$

$$\boldsymbol{r}^{\mathrm{n}} = \mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{n}} \otimes \boldsymbol{r}^{\mathrm{b}} \otimes (\mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{n}})^{*} + \boldsymbol{o}_{\mathrm{b}}^{\mathrm{n}} = (\mathring{\boldsymbol{q}}_{\mathrm{n}}^{\mathrm{b}})^{*} \otimes \boldsymbol{r}^{\mathrm{b}} \otimes \mathring{\boldsymbol{q}}_{\mathrm{n}}^{\mathrm{b}} + \boldsymbol{o}_{\mathrm{b}}^{\mathrm{n}}, \tag{3.46}$$

in the quaternion notation.

When the origins of the n-frame and b-frame coincide, i.e. $\boldsymbol{o}_{\mathrm{b}}^{\mathrm{n}} = \boldsymbol{0}$, further simplification of the equations above is possible.

The rotation matrices or quaternions can be obtained from the attitudes represented by the Euler angles using equations (3.8) and (3.26), respectively; the inverse conversion is performed by means of (3.9) and (3.27).

When other sensor is available at the navigated body, it is not placed in the reference point of the IMU, generally. As an example, we may assume a GNSS receiver, whose antenna (which is the localized element) is mounted on the top of the body, e.g. on a car roof. It is offset from the b-frame origin, however its position is stationary w.r.t. the b-frame – the vector describing the offset is commonly denoted as the *lever arm*. The measurements taken at such offset point have to be corrected for the lever arm prior to the fusion of the measurements. The GNSS measurement is referenced to the e-frame. Therefore the transformation between the b-frame and the e-frame (via the n-frame) has to be known in order to enable the correct transposition of the measurements to the b-frame origin. Such transposition is thoroughly described e.g. in [131].

3.2. Strapdown IMU mMechanization

The set of methods of estimating the position, velocity and attitude (the state of the body) on the basis of the inertial measurements is commonly reffered as the *mechanization*. The IMU senses rotation rates and accelerations in the body frame; the state of the body is to be obtained with respect to ECI, ECEF or NED frame, however. By nature of the inertial measurements it is not possible to determine the absolute position, velocity and attitude, only the change of those values can be observed. Let us

[95] In certain literature such frame is denoted as the tangential frame, see e.g. [131].

[96] Since the author's work only focuses on the inertial navigation with low-cost sensors in a limited area, the error induced by such approximation should not become significant.

now assume an ideally initialized system, where the initial state of the body is known. Moreover, we will neglect all the imperfections of the IMU measurements, for the sake of simplicity.

Regardless of the reference frame used, the most common way of mechanization comprises three steps that are performed in the following order: attitude update, velocity update and position update. We will consider that the IMU measurements are sampled at epochs referenced by index k; the sampling interval is denoted by τ_{sa}.

The mechanization described below is the forward mechanization, i.e. $\tau_{\mathrm{sa}} > 0$, which is applicable in real time. The described mechanization can be simply converted to backward mechanization by introducing negative τ_{sa} and using index $k-1$ instead of $k+1$.

3.2.1. Attitude update

The simplest version of the attitude update is performed w.r.t. the i-frame. In the quaternion and rotation matrix notation we may write for the attitude in the $k+1$ epoch that

$$\mathring{\boldsymbol{q}}^{\mathrm{i}}_{\mathrm{b}[k+1]} = \mathring{\boldsymbol{q}}^{\mathrm{i}}_{\mathrm{b}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]}; \qquad \boldsymbol{R}^{\mathrm{i}}_{\mathrm{b}[k+1]} = \boldsymbol{R}^{\mathrm{i}}_{\mathrm{b}[k]} \boldsymbol{R}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]}. \tag{3.47}$$

The terms $\mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]}$ and $\mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]}$ represent the rotation of the body from epoch k to $k+1$. Assuming that the change of the measured rotation vector between the epochs is small, we may obtain the incremental rotation vector by means of zeroth-order integration. Three kinds of the integrations may be used:

$$\text{Forward :} \quad \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]} = \mathring{\boldsymbol{q}}\{\boldsymbol{\omega}[k]\tau_{\mathrm{sa}}\} \qquad \boldsymbol{R}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]} = \boldsymbol{R}\{\boldsymbol{\omega}[k]\tau_{\mathrm{sa}}\} \tag{3.48}$$

$$\text{Backward :} \quad \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]} = \mathring{\boldsymbol{q}}\{\boldsymbol{\omega}[k+1]\tau_{\mathrm{sa}}\} \qquad \boldsymbol{R}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]} = \boldsymbol{R}\{\boldsymbol{\omega}[k+1]\tau_{\mathrm{sa}}\} \tag{3.49}$$

$$\text{Midward :} \quad \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]} = \mathring{\boldsymbol{q}}\left\{\tfrac{\boldsymbol{\omega}[k]+\boldsymbol{\omega}[k+1]}{2}\tau_{\mathrm{sa}}\right\} \qquad \boldsymbol{R}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]} = \boldsymbol{R}\left\{\tfrac{\boldsymbol{\omega}[k]+\boldsymbol{\omega}[k+1]}{2}\tau_{\mathrm{sa}}\right\} \tag{3.50}$$

The symbol $\boldsymbol{\omega}$ is the vector of the rotation rate measurements, which can be interpreted as the *rotation vector*; its subscript indicates the epoch of measurement. It is worth noting that the rotation rate in the middle of the sample interval is not necessarily the mean of the sampled values. Either way, $\boldsymbol{\omega}$ is assumed constant over the integration/sampling interval.[97] Due to the high sampling rates, the rotations between the sampling epochs are typically very small, i.e. the small angle approximation of $\boldsymbol{R}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]}$ is possible, see (3.35).

Should the e-frame be used, the Earth's rotation needs to be taken into account, since the e-frame rotates within the i-frame by a small angle between the epochs:

$$\mathring{\boldsymbol{q}}^{\mathrm{e}[k+1]}_{\mathrm{b}[k+1]} = \mathring{\boldsymbol{q}}^{\mathrm{e}[k+1]}_{\mathrm{e}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{e}[k]}_{\mathrm{b}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]}; \qquad \boldsymbol{R}^{\mathrm{e}[k+1]}_{\mathrm{b}[k+1]} = \boldsymbol{R}^{\mathrm{e}[k+1]}_{\mathrm{e}[k]} \boldsymbol{R}^{\mathrm{e}[k]}_{\mathrm{b}[k]} \boldsymbol{R}^{\mathrm{b}[k]}_{\mathrm{b}[k+1]}. \tag{3.51}$$

[97] It is possible to use first-order integration, as proposed in [135], which considers linear evolution of the $\boldsymbol{\omega}$ between the sampling epochs. Nonetheless, considering short sampling intervals and only a small change of the rotation-axis during the sampling interval, the advantage over the midward zeroth-order integration is usually negligible.

The small angle approximation is always applicable, therefore:

$$\mathring{\boldsymbol{q}}_{\mathrm{e}[k]}^{\mathrm{e}[k+1]} = \mathring{\boldsymbol{q}}\left\{\tau_{\mathrm{sa}}\begin{bmatrix}0\\0\\\Omega_{\mathrm{E}}\end{bmatrix}\right\}; \qquad \boldsymbol{R}_{\mathrm{e}[k]}^{\mathrm{e}[k+1]} = \boldsymbol{R}\left\{\tau_{\mathrm{sa}}\begin{bmatrix}0\\0\\\Omega_{\mathrm{E}}\end{bmatrix}\right\} \approx \mathbb{I} + \tau_{\mathrm{sa}}\begin{bmatrix}0\\0\\\Omega_{\mathrm{E}}\end{bmatrix}_{\times}. \tag{3.52}$$

In [131] the following approximation is suggested:

$$\boldsymbol{R}_{\mathrm{b}[k+1]}^{\mathrm{e}[k+1]} = \boldsymbol{R}_{\mathrm{b}[k]}^{\mathrm{e}[k]}(\mathbb{I} + [\boldsymbol{\omega}]_{\times}\,\tau_{\mathrm{sa}}) - \tau_{\mathrm{sa}}\begin{bmatrix}0\\0\\\Omega_{\mathrm{E}}\end{bmatrix}_{\times}\boldsymbol{R}_{\mathrm{b}[k]}^{\mathrm{e}[k]}. \tag{3.53}$$

When the n-frame is used as the reference frame, additional, but static rotation between the e-frame and n-frame $\mathring{\boldsymbol{q}}_{\mathrm{n}[k]}^{\mathrm{n}[k+1]}$ or $\boldsymbol{R}_{\mathrm{n}[k]}^{\mathrm{n}[k+1]}$ has to be taken into account

$$\mathring{\boldsymbol{q}}_{\mathrm{b}[k+1]}^{\mathrm{n}[k+1]} = \mathring{\boldsymbol{q}}_{\mathrm{n}[k]}^{\mathrm{n}[k+1]} \otimes \mathring{\boldsymbol{q}}_{\mathrm{b}[k]}^{\mathrm{n}[k]} \otimes \mathring{\boldsymbol{q}}_{\mathrm{b}[k+1]}^{\mathrm{b}[k]}; \qquad \boldsymbol{R}_{\mathrm{b}[k+1]}^{\mathrm{n}[k+1]} = \boldsymbol{R}_{\mathrm{n}[k]}^{\mathrm{n}[k+1]}\boldsymbol{R}_{\mathrm{b}[k]}^{\mathrm{n}[k]}\boldsymbol{R}_{\mathrm{b}[k+1]}^{\mathrm{b}[k]}. \tag{3.54}$$

The transformation from frame n$[k]$ to n$[k+1]$ is latitude-dependent and can be obtained as

$$\begin{aligned}
\mathring{\boldsymbol{q}}_{\mathrm{n}[k]}^{\mathrm{n}[k+1]} &= \mathring{\boldsymbol{q}}_{\mathrm{e}}^{\mathrm{n}} \otimes \mathring{\boldsymbol{q}}_{\mathrm{e}[k]}^{\mathrm{e}[k+1]} \otimes (\mathring{\boldsymbol{q}}_{\mathrm{e}}^{\mathrm{n}})^{*} = \tau_{\mathrm{sa}}\begin{bmatrix}\cos\frac{\Omega_{\mathrm{E}}}{2}\\ -\cos\varphi\sin\frac{\Omega_{\mathrm{E}}}{2}\\ 0\\ \sin\varphi\sin\frac{\Omega_{\mathrm{E}}}{2}\end{bmatrix};\\
\boldsymbol{R}_{\mathrm{n}[k]}^{\mathrm{n}[k+1]} &= \boldsymbol{R}_{\mathrm{e}}^{\mathrm{n}}\boldsymbol{R}_{\mathrm{e}[k]}^{\mathrm{e}[k+1]}(\boldsymbol{R}_{\mathrm{e}}^{\mathrm{n}})^{\mathrm{T}} \approx \tau_{\mathrm{sa}}\begin{bmatrix}1 & \Omega_{\mathrm{E}}\sin\varphi & 0\\ -\Omega_{\mathrm{E}}\sin\varphi & 1 & -\Omega_{\mathrm{E}}\cos\varphi\\ 0 & \Omega_{\mathrm{E}}\cos\varphi & 1\end{bmatrix},
\end{aligned} \tag{3.55}$$

where φ denotes the geodetic latitude. The small-angle approximation of the Earth's rotation has been used in the rotation-matrix notation.

The definition of the n-frame from Section 3.1.2 does not require coincidence of the n-frame and b-frame origins strictly, however, the distance between them should allow flat-Earth approximation. Therefore, the *transport rate*, i.e. the rotation of the n-frame with the curvature of the Earth as the coincident b-frame origin moves, is neglected. The formulas including the transport rate are available e.g. in [131, 133].

It is worth noting that for IMUs without the performance required for Earth's rotation sensing, both the n-frame and i-frame can be considered to be the inertial reference frame. In such case the attitude update equations may be simplified to

$$\mathring{\boldsymbol{q}}_{\mathrm{b}[k+1]}^{\mathrm{e}} \approx \mathring{\boldsymbol{q}}_{\mathrm{b}[k]}^{\mathrm{e}} \otimes \mathring{\boldsymbol{q}}_{\mathrm{b}[k+1]}^{\mathrm{b}[k]}; \qquad \boldsymbol{R}_{\mathrm{b}[k+1]}^{\mathrm{e}} \approx \boldsymbol{R}_{\mathrm{b}[k]}^{\mathrm{e}}\boldsymbol{R}_{\mathrm{b}[k+1]}^{\mathrm{b}[k]}, \tag{3.56}$$

$$\mathring{\boldsymbol{q}}_{\mathrm{b}[k+1]}^{\mathrm{n}} \approx \mathring{\boldsymbol{q}}_{\mathrm{b}[k]}^{\mathrm{n}} \otimes \mathring{\boldsymbol{q}}_{\mathrm{b}[k+1]}^{\mathrm{b}[k]}; \qquad \boldsymbol{R}_{\mathrm{b}[k+1]}^{\mathrm{n}} \approx \boldsymbol{R}_{\mathrm{b}[k]}^{\mathrm{n}}\boldsymbol{R}_{\mathrm{b}[k+1]}^{\mathrm{b}[k]}. \tag{3.57}$$

3.2.2. Velocity update

Prior to the velocity update, the specific force vector $\boldsymbol{a}^{\mathrm{b}}$ measured in the b-frame has to be transformed to the frame of interest. In the continuous time, the body attitude

and therefore the specific-force sensor rotates continuously. Usually, the situation is simplified so that the specific force and the attitude is considered constant during the sampling period. The attitude in halfway between the sampling epochs is chosen as the constant one, i.e. the measurement is interpreted as if $\boldsymbol{a}^{\mathrm{b}[k+\frac{1}{2}]}$ was measured.

Using the quaternion notation we may write

$$\boldsymbol{a}^{\mathrm{i}} = \mathring{\boldsymbol{q}}^{\mathrm{i}}_{\mathrm{b}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+\frac{1}{2}]} \otimes \boldsymbol{a}^{\mathrm{b}[k+\frac{1}{2}]} \otimes \left(\mathring{\boldsymbol{q}}^{\mathrm{i}}_{\mathrm{b}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+\frac{1}{2}]}\right)^* \tag{3.58}$$

$$\boldsymbol{a}^{\mathrm{e}[k+\frac{1}{2}]} = \mathring{\boldsymbol{q}}^{\mathrm{e}[k+\frac{1}{2}]}_{\mathrm{e}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{e}[k]}_{\mathrm{b}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+\frac{1}{2}]} \otimes \boldsymbol{a}^{\mathrm{b}[k+\frac{1}{2}]} \otimes \left(\mathring{\boldsymbol{q}}^{\mathrm{e}[k+\frac{1}{2}]}_{\mathrm{e}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{e}[k]}_{\mathrm{b}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+\frac{1}{2}]}\right)^* \tag{3.59}$$

$$\boldsymbol{a}^{\mathrm{n}[k+\frac{1}{2}]} = \mathring{\boldsymbol{q}}^{\mathrm{n}[k+\frac{1}{2}]}_{\mathrm{n}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{n}[k]}_{\mathrm{b}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+\frac{1}{2}]} \otimes \boldsymbol{a}^{\mathrm{b}[k+\frac{1}{2}]} \otimes \left(\mathring{\boldsymbol{q}}^{\mathrm{n}[k+\frac{1}{2}]}_{\mathrm{n}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{n}[k]}_{\mathrm{b}[k]} \otimes \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+\frac{1}{2}]}\right)^*, \tag{3.60}$$

where the forward integration of attitude half-update $\mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+\frac{1}{2}]}$ follows

$$\text{Forward :} \qquad \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+\frac{1}{2}]} = \mathring{\boldsymbol{q}}\left\{\frac{\boldsymbol{\omega}[k]\tau_{\mathrm{sa}}}{2}\right\} \tag{3.61}$$

$$\text{Backward :} \qquad \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+\frac{1}{2}]} = \mathring{\boldsymbol{q}}\left\{\frac{\boldsymbol{\omega}[k+1]\tau_{\mathrm{sa}}}{2}\right\} \tag{3.62}$$

$$\text{Midward :} \qquad \mathring{\boldsymbol{q}}^{\mathrm{b}[k]}_{\mathrm{b}[k+\frac{1}{2}]} = \mathring{\boldsymbol{q}}\left\{\frac{\boldsymbol{\omega}[k]+\boldsymbol{\omega}[k+1]}{4}\tau_{\mathrm{sa}}\right\} \tag{3.63}$$

The rotation of the Earth within the half of the sample period is given by $\mathring{\boldsymbol{q}}^{\mathrm{e}[k+\frac{1}{2}]}_{\mathrm{e}[k]}$ and $\mathring{\boldsymbol{q}}^{\mathrm{n}[k+\frac{1}{2}]}_{\mathrm{n}[k]}$, respectively. The actual value may be obtained using (3.52) where τ_{sa} is substitued for $\tau_{\mathrm{sa}}/2$. For the low-cost IMUs both $\mathring{\boldsymbol{q}}^{\mathrm{e}[k+\frac{1}{2}]}_{\mathrm{e}[k]}$ and $\mathring{\boldsymbol{q}}^{\mathrm{n}[k+\frac{1}{2}]}_{\mathrm{n}[k]}$ are negligible, i.e. they are assumed to be identity quaternions and may disappear from the equations.

In the rotation matrix notation the $\boldsymbol{R}^{\mathrm{e}[k+\frac{1}{2}]}_{\mathrm{b}[k+\frac{1}{2}]}$ can be approximated by the average of $\boldsymbol{R}^{\mathrm{e}[k+1]}_{\mathrm{b}[k+1]}$ and $\boldsymbol{R}^{\mathrm{e}[k]}_{\mathrm{b}[k]}$, which are both known from the attitude update step [131]; similar approach can be applied on $\boldsymbol{R}^{\mathrm{n}[k]}_{\mathrm{b}[k+\frac{1}{2}]}$. Consequently,

$$\boldsymbol{a}^{\mathrm{i}} = \frac{\boldsymbol{R}^{\mathrm{i}}_{\mathrm{b}[k]} + \boldsymbol{R}^{\mathrm{i}}_{\mathrm{b}[k+1]}}{2}\boldsymbol{a}^{\mathrm{b}[k+\frac{1}{2}]} \tag{3.64}$$

$$\boldsymbol{a}^{\mathrm{e}[k+\frac{1}{2}]} = \frac{\boldsymbol{R}^{\mathrm{e}[k]}_{\mathrm{b}[k]} + \boldsymbol{R}^{\mathrm{e}[k+1]}_{\mathrm{b}[k+1]}}{2}\boldsymbol{a}^{\mathrm{b}[k+\frac{1}{2}]} \tag{3.65}$$

$$\boldsymbol{a}^{\mathrm{n}[k+\frac{1}{2}]} = \frac{\boldsymbol{R}^{\mathrm{n}[k]}_{\mathrm{b}[k]} + \boldsymbol{R}^{\mathrm{n}[k+1]}_{\mathrm{b}[k+1]}}{2}\boldsymbol{a}^{\mathrm{b}[k+\frac{1}{2}]}. \tag{3.66}$$

A hybrid mechanism for the low-cost IMU is used for the specific-force vector transformation to the n-frame in Appendix D; the forward attitude integration is applied and Earth's rotation is neglected.

$$\boldsymbol{a}^{\mathrm{n}} = \boldsymbol{R}\left\{\mathring{\boldsymbol{q}}^{\mathrm{n}}_{\mathrm{b}[k]} \otimes \mathring{\boldsymbol{q}}\left\{\frac{\boldsymbol{\omega}[k]\,\tau_{\mathrm{sa}}}{2}\right\}\right\}\boldsymbol{a}^{\mathrm{b}[k+\frac{1}{2}]}. \tag{3.67}$$

The next step in the velocity update in all the frames is the correction for the gravity which is considered constant over the integration interval $\boldsymbol{g}(\boldsymbol{r})$. In the i-frame we can write

$$\boldsymbol{v}^{\mathrm{i}}[k+1] = \boldsymbol{v}^{\mathrm{i}}[k] + \tau_{\mathrm{sa}} \left(\boldsymbol{a}^{\mathrm{i}}[k+\tfrac{1}{2}] + \boldsymbol{g}(r^{\mathrm{i}}[k]) \right). \tag{3.68}$$

In the e-frame and n-frame the fictious forces have to be accounted for (if possibly sensed by the IMU). Neglecting the transport rate, we may simplify from [131] as following

$$\boldsymbol{v}^{\mathrm{e}}[k+1] = \boldsymbol{v}e_{[\mathrm{k}]} + \tau_{\mathrm{sa}} \left(\boldsymbol{a}^{\mathrm{e}}[k+\tfrac{1}{2}] - 2 \begin{bmatrix} 0 \\ 0 \\ \Omega_{\mathrm{E}} \end{bmatrix}_{\times} \boldsymbol{v}^{\mathrm{e}}[k] + \boldsymbol{g}(r^{\mathrm{i}}[k]) \right) \tag{3.69}$$

$$\boldsymbol{v}^{\mathrm{n}}[k+1] = \boldsymbol{v}^{\mathrm{n}}[k] + \tau_{\mathrm{sa}} \left(\boldsymbol{a}^{\mathrm{n}}[k+\tfrac{1}{2}] - 2 \begin{bmatrix} \cos\varphi\Omega_{\mathrm{E}} \\ 0 \\ -\sin\varphi\Omega_{\mathrm{E}} \end{bmatrix}_{\times} \boldsymbol{v}^{\mathrm{n}}[k] + \boldsymbol{g}(r^{\mathrm{i}}[k]) \right). \tag{3.70}$$

Under the assumption of navigation with a low-cost IMU the Coriolis term can be neglected and

$$\boldsymbol{v}^{\mathrm{e}}_{[\mathrm{k+1}]} \approx \boldsymbol{v}^{\mathrm{e}}_{[\mathrm{k}]} + \tau_{\mathrm{sa}} \left(\boldsymbol{a}^{\mathrm{e}[k+\frac{1}{2}]} + \boldsymbol{g}(r^{\mathrm{e}}_{[\mathrm{k}]}) \right); \tag{3.71}$$

$$\boldsymbol{v}^{\mathrm{n}}_{[\mathrm{k+1}]} \approx \boldsymbol{v}^{\mathrm{n}}_{[\mathrm{k}]} + \tau_{\mathrm{sa}} \left(\boldsymbol{a}^{\mathrm{n}[k+\frac{1}{2}]} + \boldsymbol{g}(r^{\mathrm{n}}_{[\mathrm{k}]}) \right). \tag{3.72}$$

The gravity compensation differs for the respective frames, moreover, the centrifugal component is often accounted for within the $\boldsymbol{g}$ vector. In [131, 133] a Somigliana model is mentioned, which provides the g_0 acceleration due to gravity in zero height as a function of latitude φ:

$$g_0(\varphi) = 9.780\,325\,335\,9 \frac{1 + 0.001\,931\,853 \sin^2\varphi}{\sqrt{1 - e_{\mathrm{E}}^2 \sin^2\varphi}} \qquad [\mathrm{m\,s^{-2}}]. \tag{3.73}$$

This value incorporates the centrifugal force, which is a few orders smaller than the gravitational component of the acceleration due to gravity. Therefore, for most applications it can be assumed that in the n-frame the acceleration due to gravity $\boldsymbol{g}^{\mathrm{n}}$ is aligned with the z-axis.

The gravitational acceleration decreases with height above the ellipsoid h, in contrast with the growing centrifugal force. Consequently, in order to extrapolate above the surface, the centrifugal force at $h = 0$ should be subtracted, then, the gravitational acceleration scaled, and as the last step the centrifugal force at the body height added. This is described exactly in [131]. Nonetheless, in the vicinity of Earth's surface and especially when using low-cost IMUs, the scaling can be applied to the acceleration due to gravity directly:

$$\boldsymbol{g}^{\mathrm{n}}(\varphi, h) \approxeq \left(1 - \frac{R_{\mathrm{S}}^2}{(|R_{\mathrm{S}} + h)^2} \right) \boldsymbol{g}^{\mathrm{n}}_0(\varphi) \approx \left(1 - \frac{2h}{R_{\mathrm{S}}} \right) \boldsymbol{g}^{\mathrm{n}}_0(\varphi), \tag{3.74}$$

where the symbol R_{S} is the Earth geocentric radius, i.e. distance from the ECEF/ECI origin to the particular surface point of the reference ellipsoid. It can be obtained as

a function of geodetic latitude

$$R_{\mathrm{S}} = a_{\mathrm{E}} \sqrt{\frac{\cos^2\varphi + (1-e_{\mathrm{E}}^2)^2 \sin^2\varphi}{1 - e_{\mathrm{E}}^2 \sin^2\varphi}}. \tag{3.75}$$

The vector of acceleration due to gravity in the e-frame $\boldsymbol{g}^{\mathrm{e}}_{(}\varphi, h)$ can be obtained via transformation (3.41).

In the non-rotating i-frame there is no centrifugal force; according to [131], the gravitation correction can be obtained directly from the position vector

$$\boldsymbol{g}^{\mathrm{i}}(\boldsymbol{r}^{\mathrm{i}}_{\mathrm{b}}) = \boldsymbol{\gamma}^{\mathrm{i}}(\boldsymbol{r}^{\mathrm{i}}_{\mathrm{b}}) = -\frac{\mu_{\mathrm{E}}}{\left\|\boldsymbol{r}^{\mathrm{i}}_{\mathrm{b}}\right\|^3} \left(\boldsymbol{r}^{\mathrm{i}}_{\mathrm{b}} + \frac{3 J_{\mathrm{E2}} a_{\mathrm{E}}^2}{2 \left\|\boldsymbol{r}^{\mathrm{i}}_{\mathrm{b}}\right\|^2} \begin{bmatrix} \left(1 - 5\left(\frac{r^{\mathrm{i}}_{\mathrm{b},z}}{\left\|\boldsymbol{r}^{\mathrm{i}}_{\mathrm{b}}\right\|}\right)^2\right) r^{\mathrm{i}}_{\mathrm{b},x} \\ \left(1 - 5\left(\frac{r^{\mathrm{i}}_{\mathrm{b},z}}{\left\|\boldsymbol{r}^{\mathrm{i}}_{\mathrm{b}}\right\|}\right)^2\right) r^{\mathrm{i}}_{\mathrm{b},y} \\ \left(1 - 5\left(\frac{r^{\mathrm{i}}_{\mathrm{b},z}}{\left\|\boldsymbol{r}^{\mathrm{i}}_{\mathrm{b}}\right\|}\right)^2\right) r^{\mathrm{i}}_{\mathrm{b},z} \end{bmatrix} \right), \tag{3.76}$$

where $\mu_{\mathrm{E}} = 3.986\,004\,418 \times 10^{14}\,\mathrm{m}^3/\mathrm{s}^2$ is the Earth's gravitational constant and $J_{\mathrm{E2}} = 1.082\,627 \times 10^{-3}$ is the dynamic form factor, both according to WGS-84.

3.2.3. Position update

The position update is thesis almost identical in all the frames described within this. Assuming linear evolution of the velocity vector throughout the sampling interval, the integration of velocity follows

$$\boldsymbol{r}^{\mathrm{i}}[k+1] = \boldsymbol{r}^{\mathrm{i}}[k] + \frac{\tau_{\mathrm{sa}}}{2}\left(\boldsymbol{v}^{\mathrm{i}}[k+1] + \boldsymbol{v}^{\mathrm{i}}[k]\right) \tag{3.77}$$

$$\boldsymbol{r}^{\mathrm{e}}[k+1] = \boldsymbol{r}^{\mathrm{e}}[k] + \frac{\tau_{\mathrm{sa}}}{2}\left(\boldsymbol{v}^{\mathrm{e}}[k+1] + \boldsymbol{v}^{\mathrm{e}}[k]\right) \tag{3.78}$$

$$\boldsymbol{r}^{\mathrm{n}}[k+1] = \boldsymbol{r}^{\mathrm{n}}[k] + \frac{\tau_{\mathrm{sa}}}{2}\left(\boldsymbol{v}^{\mathrm{n}}[k+1] + \boldsymbol{v}^{\mathrm{n}}[k]\right). \tag{3.79}$$

Note that these expressions are principally the midward integration as was used in the attitude update (3.50). Similarly, a forward and backward integration of velocity can be employed:

$$\text{Forward :} \qquad \boldsymbol{r}^{\mathrm{i}}[k+1] = \boldsymbol{r}^{\mathrm{i}}[k] + \tau_{\mathrm{sa}}\boldsymbol{v}^{\mathrm{i}}[k] \tag{3.80}$$

$$\text{Backward :} \qquad \boldsymbol{r}^{\mathrm{i}}[k+1] = \boldsymbol{r}^{\mathrm{i}}[k] + \tau_{\mathrm{sa}}\boldsymbol{v}^{\mathrm{i}}[k+1]., \tag{3.81}$$

and likewise for the other frames. The difference of the three integration approaches should not become significant provided that the sampling rate is high w.r.t. the system dynamics.

It is worth remarking that when using n-frame mechanization, the body coordinates are converted to the ECEF frame by means of (3.40) and then converted to the geodetic (latitude, longitude, height) format. In contrast, [131, 134] provide equations that operate within the geodetic coordinate system directly.

3.2.4. Initialization

Since the position, velocity and attitude are estimated by integration of their measured derivatives, it is necessary to provide boundary conditions, i.e. initial state, in order to make the integrated information meaningful. A solid analogy can be seen in simple indefinite integral of a one-dimensional function – the result is the *antiderivative* and *arbitrary constant*. The mechanization of the IMU was described by formulas providing transition from epoch k to $k+1$, thus initialization, i.e. definition of values for $k = 0$ is necessary.

The velocity and position have to be initialized by external means. The information may be passed from GNSS receiver or similar positioning system, for instance. Alternatively, the IMU is initialized as stationary (velocity in e-frame is zero vector) at *a priori* known location (position in e-frame is fixed).

The attitude can be partially resolved by means of the accelerometer measurements. The reaction to the gravity is sensed as a specific force vector pointing upwards. Consequently, the pitch θ and roll ϕ angles can be estimated from the accelerometer measurements by means of the *leveling* process [131] as

$$\theta = \tan^{-1} \frac{-a_x^{\mathrm{b}}}{\sqrt{(a_y^{\mathrm{b}})^2 + (a_z^{\mathrm{b}})^2}} \qquad \phi = \tan^{-1} \frac{-a_y^{\mathrm{b}}}{-a_z^{\mathrm{b}}}, \tag{3.82}$$

where four-quadrant arctangent should be used for the roll computation; note that pitch is constrained to $\pm 90°$ range. It is worth remarking that the body has to be stationary during leveling, otherwise the motion disrupts the specific force measurements and introduces an attitude error.

The yaw ψ cannot be obtained by similar approach. For high-performance grade IMUs the *gyrocompassing* can be used [131]. The gyrocopassing exploits sensing the Earth's rotation in the body frame by means of a gyroscope, which is beyond the capabilities of the lower-grade IMUs.

Consequently, the yaw information has to be obtained externally, e.g. from multi-antenna radio localization system or magnetometers (magnetic compass). In some applications the yaw can be estimated during motion, since the motion can be performed only in a certain direction in the b-frame. For example, the velocity vector of the car in Fig. 3.1 is always aligned with the x^{b}-axis.

If the lever-arm offset is present, the attitude has to be initialized prior to position and velocity in order to resolve the direction of the lever arm vector.

3.3. Effects of the IMU Errors

Both gyroscopes and accelerometers are imperfect devices, which suffer from both random and systematic errors. First, we will cover briefly the most common errors of the IMU measurements; then, the effect of biases and random walk on the navigation performance is presented.

3.3.1. IMU measurement errors

The errors of the IMU sensors have been analyzed and classified thoroughly in literature. A detailed error analysis of various sensor types can be found in books dedicated to inertial navigation, such as [133]. Below, only the very basic kinds of errors are covered.

The measurements suffer from random noise. Often the noise is expressed as the angular random walk (ARW) for the gyroscopes and velocity random walk (VRW) for the accelerometers; white noise is assumed. The ARW is commonly expressed in angle per square-root of time unit, e.g. $^\circ/\sqrt{\mathrm{s}}$; the VRW in velocity square-root of time unit, e.g. $\mathrm{ms}^{-1}/\sqrt{\mathrm{s}}$.

The IMU measurements are typically biased, i.e. the measurements are offset by a constant value. Due to the constant property of the value it is possible to estimate the biases, e.g. by means of Kalman filtering, as is shown in Appendix D. The bias has the same dimension as the measured value, naturally. In reality, the bias is not absolutely constant; it may change slightly in time. Therefore, it is characterized by two values: the bias *repeatability* (or turn-on bias) and bias *instability* (or in-run bias). The repeatability is the spread of the constant component of the bias that is present since the IMU is turned on. The bias instability expresses the change of bias in time during normal operation.

The biases may be estimated and compensated, at least to a certain extent. It is somehow problematic to distinguish the bias from the true measurement for the standalone IMU. One of the option is that a stationary condition is detected, i.e. the accelerometer and gyroscope measurements should be zero (of course up to the graviational acceleration and possibly Earth's rotation). Only noise and bias are then forming the actual measurement. Such approach of estimating biases is referred as zero-velocity update (ZUPT).

When some other positioning system si available (most commonly GNSS), the data from the system can be used to estimate the biases on a continuous basis. Reciprocally, the IMU measurements allow the system to overcome outages of the other positioning system. By means of such *sensor fusion* the weak points of the sensors may be suppressed and the performance of the system should outperform the individual sensors. Some of the methods of sensor fusion are theoretically described in Chapter 4, an example concerning author's work on fusion of IMU, GNSS and UWB is provided as Appendix D.

From the other possible errors, the scale factor error is common to the IMU sensors as well. The value is dimensionless, however very small, so it is usually expressed in terms of ppm.

The sensitive axes of the sensors are sometimes not perfectly orthogonal and some misalignment w.r.t. the true b-frame axes may be present. Consequently, a *cross-coupling* of the sensors may occur. For certain technologies the gyroscope scale factor is acceleration dependent, i.e. there is a coupling between the gyroscope and accelerometer readings. The model and correction for the most common errors is available e.g. in [131, 133]. In the author's work concerning IMUs, only random walks and biases were taken into account, see Appendix D or the corresponding article [138].

3.3.2. Bias and random walk propagation

Within this section we will show the effects of the IMU errors, namely biases and random walks, on the position velocity and attitude estimates. The focus will be put on the short time interval after initialization, since longer standalone operation of the consumer-grade IMUs is not viable.

For the short-term analysis of low-performance IMUs it is possible to neglect Earth's rotation, transport rate, *et cetera*. Moreover, the biases may be assumed constant. Therefore, we may model the biased and noisy measurements of gyroscopes and accelerometers in each epoch as

$$\breve{\boldsymbol{\omega}}[k] = \boldsymbol{\omega}[k] + \boldsymbol{\delta}_\omega + \mathcal{N}\left(0,\, \varsigma_\omega^2 \tau_{sa}\right) \tag{3.83}$$

$$\breve{\boldsymbol{a}}[k] = \boldsymbol{a}[k] + \boldsymbol{\delta}_a + \mathcal{N}\left(0,\, \varsigma_a^2 \tau_{sa}\right), \tag{3.84}$$

where $\boldsymbol{\delta}_\omega$ and $\boldsymbol{\delta}_a$ denote the biases and ς_ω^2 and ς_a^2 denote the AWGN random walk variance when integrated over a unit of time.

In the following paragraphs, the propagation of errors will be illustrated on an example considering parameters of a MEMS-based IMU. The simplified n-frame mechanization was used to investigate the error propagation in a stationary IMU. Constant biases and Gaussian random-walk sensor errors were considered, initial state (attitude, velocity and position) was known perfectly. For each run, the biases were chosen randomly from a pre-defined symmetric uniform distribution, the biases were considered uncorrelated. The parameters of the simulated IMU sensors are available in Tab. 3.3. The results of 10 simulation runs are provided in each of the figures presented further in this section.

Table 3.3.: Parameters of MEMS IMU error simulation.

Parameter		**Distribution**	**Value**	**Unit**
Gyroscope	Bias	uniform	$-0.3 < \boldsymbol{\delta}_\omega < +0.3$	°/s
	ARW	Gaussian	$\varsigma_\omega = 0.001$	$°/\sqrt{\text{s}}$
Accelerometer	Bias	uniform	$-6 < \boldsymbol{\delta}_\omega < +6$	mg
	VRW	Gaussian	$\varsigma_\omega = 0.004$	mg
Sample interval		constant	$\tau_{sa} = 10$	ms

During each attitude update step, both bias and AWGN affect the estimate. Unfortunately, the gyro measurements correspond to the Euler-angle rates only for very small angles – under such conditions, the attitude errors due to biases grow linearly in time. For larger errors the attitude errors grow non-linearly. In order to investigate the error propagation the attitude update mechanization has to be performed, and the estimate compared w.r.t. the ground truth. In a simple simulation, the rotation-matrix version of the simplified n-frame attitude update (3.57) has been used. The bias-induced errors are typically overpowering the ARW-related error. In Fig. 3.4a it is observable that initially the attitude errors grow linearly according to the biases, and later become nonlinear. Fig. 3.4b shows the attitude errors when the biases are neglected, i.e. only the ARW is considered. It is obvious that most of the errors of MEMS IMUs originate from biases, however, when coupled with some other system, the biases may be estimated and at least partially suppressed.

Considering a non-rotating body and perfectly known attitude, the velocity error would be simply integrated bias, i.e. linear drift in time with superposed VRW. Nonetheless, the attitude knowledge cannot be considered perfect and the error of attitude will always affect the velocity estimate. Not only the accelerometer readings will integrate in the wrong direction, but more importantly, the compensation of the specific force for the gravitational acceleration is affected, since the measurements are not correctly rotated from the b-frame to the mechanization frame. Therefore the velocity error may grow faster than linearly. As can be seen from Tab. 3.3, the accelerometer drift contribution to the velocity error may be a few orders of magnitude higher than the VRW contribution.

In Fig. 3.5 the result of the simulation is presented; the graphs were obtained using the rotation-matrix variant of the specific force vector transformation (3.66) and the simplified n-frame mechanization (3.72). In the top part of each figure the error along the principal axes is shown w.r.t. time. It is clearly observable that the error along the horizontal axes grows faster than along the vertical plane. The reason is that the "leakage" of the gravitational correction will be proportional to the sines of the tilt angles (pitch and roll), while the value on the vertical axis will be proportional to the cosines of the tilts (which are close to one for small errors). The bottom part of the figures depicts the magnitude of the velocity-error vector.

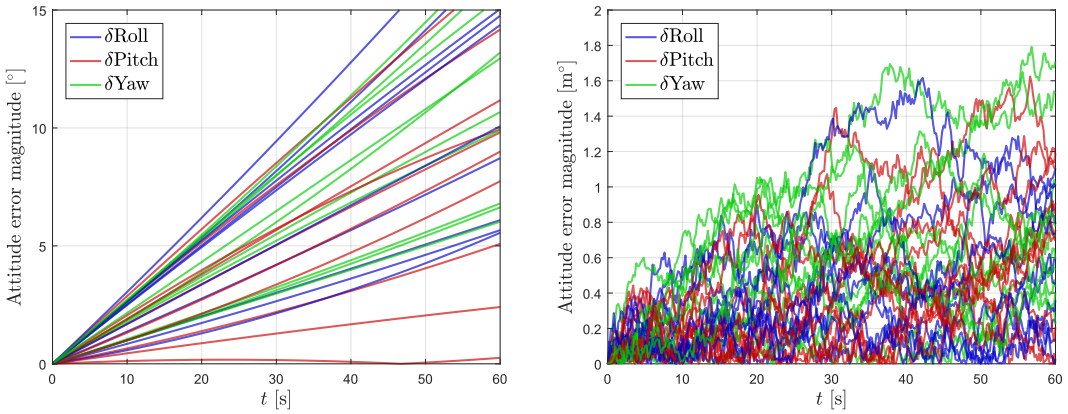
As well as in the attitude error demonstration, Fig. 3.5a takes the drifts and VRW into account, whilst in Fig. 3.5b only the VRW is considered. It is worth noticing that the gravitational acceleration leakage induces a significant lateral error component, when biases are neglected. In the vertical direction error is almost zero thanks to the small attitude error.

The position is obtained by means of velocity integration in the mechanization frame; in the simulation we have used the forward integration (3.80) in the n-frame. Naturally, all the errors affecting the velocity estimate accumulate. Under the assumption of non-rotating body and perfect attitude sensors the bias-induced error would grow quadratically. Taking into an account the attitude error and gravitational acceleration leakage, the error grows faster than quadratically, as can be seen in Fig. 3.6.

When the biases are present and not accounted for ideally (Fig. 3.6a), the errors may quickly grow to the unacceptably high values. In the upper plots within the figures

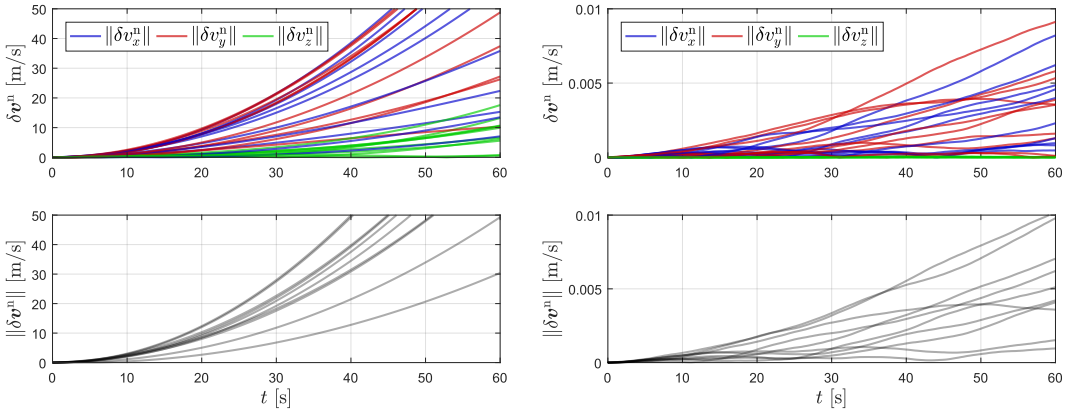
it can be observed that due to the gravitational acceleration leakage contribution, the position estimate is diverging from the correct value faster in the horizontal plane than it does in the vertical direction. In the bottom plot the position-error vector magnitude is depicted on a logarithmic vertical axis. The performance without biases is a few orders better, however, the error induced by the random walks becomes clearly noticeable (above decimeter level) after a few tens of seconds.

Even though this simulation was rather simple, it can be observed that the bias-related errors are the major component of the overall attitude, velocity and position errors for the MEMS-based IMUs. Therefore, suppression of biases is critical for the IMU performance. An example of the Kalman-filter-based sensor fusion algorithm that estimates the biases is provided in Appendix D.



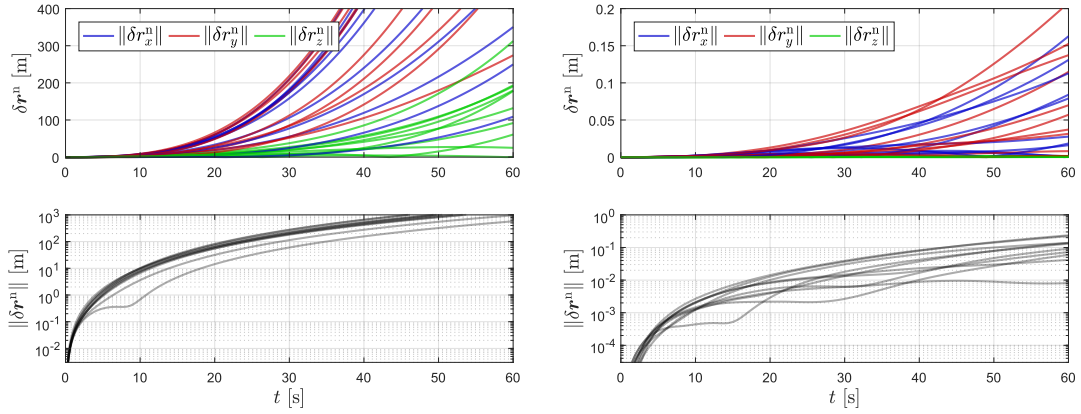
(A) Affected by biases and noise. (B) Affected only by noise.

FIGURE 3.4.: Propagation of IMU measurement errors to the attitude errors.



(A) Affected by biases and noise. (B) Affected only by noise.

FIGURE 3.5.: Propagation of IMU measurement errors to the velocity errors.



(A) Affected by biases and noise. (B) Affected only by noise.

FIGURE 3.6.: Propagation of IMU measurement errors to the position errors.

4. Positioning Solution and Data Fusion

Various approaches to radio positioning were presented within Chapter 2. Neither of them provides the position information directly; instead, certain observables are measured and position (and sometimes velocity) is obtained by means of solving set of equations that describe the relation of position and observables. Such equations are nonlinear, the closed-form solutions are nontrivial, and often impractical or unavailable at all. The first part of this chapter describes a few iterative methods that are suitable for solving the positioning equations in the least-squares sense. In addition to the general description of the methods, the actual application on the common problems will be presented.

It has been mentioned several times that data from multiple positioning systems can be combined in order to suppress the weaknesses of the individual sensors and improve the overall performance of the integrated system. Especially, in Chapter 3 the possibility to estimate the inertial sensor biases continuously based on other data was remarked. The second part of this chapter presents Kalman filtering techniques, which can be used to fuse the data and estimate states of the navigated body.

4.1. Iterative Solutions of Positioning Equations

The problems presented within the thesis, such as obtaining the position from either TWR, ToA, TDoA or AoA measurements are nonlinear. The closed-form solutions to the problems are rare and considerably less flexible than their iterative counterparts.[98] Moreover, the problems are often overdetermined, i.e. more measurements than required to solve the set of equations are available.

Here, we will focus on iterative solvers that converge to the solution with the least-square error. The aim is not to provide a rigorous derivation or describe the estimator theory, but to provide a guide how to use those algorithms and what are the strong or weak points of these algorithms. The theoretical background to these algorithms can be found in nonlinear optimization handbooks, such as [140].

Initially, the methods will be introduced in a general form. Section 4.1.5 then provides the exact formulas for the particular radionavigation problems.

[98] For example, the ToA problem, i.e. the satellite navigation positioning set of equations, can be solved by means of the Bancroft algorithm [139]. Such algorithm is designed for a single purpose, additional measurements and constraints cannot be applied.

It is common to all the described gradient algorithms that the parameter vector $\boldsymbol{x}$ is divided into the predicted part $\hat{\boldsymbol{x}}$ and its correction $\check{\boldsymbol{x}}$; the same is done for the vector of measurements:

$$\boldsymbol{x} = \tilde{\boldsymbol{x}} + \check{\boldsymbol{x}} \tag{4.1}$$

$$\boldsymbol{y} = \tilde{\boldsymbol{y}} + \check{\boldsymbol{y}}. \tag{4.2}$$

The relation of the measurement and parameters is given by a vector function $\boldsymbol{y} = \mathbf{f}(\boldsymbol{x})$. In the further text we will denote the covariance matrices of $\boldsymbol{x}$ and $\boldsymbol{y}$ by symbols $\boldsymbol{\Phi}_{\boldsymbol{x}}$ and $\boldsymbol{\Phi}_{\boldsymbol{y}}$. In the following sections we will provide three algorithms that are able to estimate the parameters $\boldsymbol{x}$ based on the set of measurements $\boldsymbol{y}$.

4.1.1. Newton-Raphson algorithm

First, we will show the Newton-Raphson (or Gauss-Newton) algorithm, which is the fastest iterative algorithm that uses only the first derivatives of the measurement function. It is presumed to exhibit quadratic convergence rate in the vicinity of the optimal solution. The estimate $\hat{\boldsymbol{x}}$ obtained by such method minimizes the least-squares condition [140]

$$\hat{\boldsymbol{x}} = \arg\min_{x} \left\| \boldsymbol{y}_i - \mathbf{f}(\boldsymbol{x}) \right\|^2. \tag{4.3}$$

First, linearization of the correction in the vicinity of the predicted value is necessary

$$\boldsymbol{y} \approx \mathbf{f}(\tilde{\boldsymbol{x}}) + \boldsymbol{J}_{\boldsymbol{f}}|_{\tilde{\boldsymbol{x}}}\, \check{\boldsymbol{x}}. \tag{4.4}$$

Note that $\boldsymbol{J}_{\boldsymbol{f}}$ is the Jacobian matrix of the $\mathbf{f}(\,)$ function, computed at point of prediction $\tilde{\boldsymbol{x}}$. The residual $\check{\boldsymbol{y}}$, i.e. the correcting part of the measurements, can be then approximated by means of

$$\check{\boldsymbol{y}} = \boldsymbol{y} - \tilde{\boldsymbol{y}} = \boldsymbol{y} - \mathbf{f}(\tilde{\boldsymbol{x}}) \approx \boldsymbol{J}_{\boldsymbol{f}} \check{\boldsymbol{x}}, \tag{4.5}$$

where the point of $\boldsymbol{J}_{\boldsymbol{f}}$ evaluation subscript has been omitted. Generally, the Jacobian matrix is not square and hence it cannot be simply inverted. Therefore, pseudoinverse (4.8) operation has to be performed. Obviously, $\boldsymbol{J}_{\boldsymbol{f}}$ has to be full-rank in order to satisfy that $\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{J}_{\boldsymbol{f}}$ is regular, i.e. invertible.

$$\boldsymbol{J}_{\boldsymbol{f}} \check{\boldsymbol{x}} = \check{\boldsymbol{y}} \tag{4.6}$$

$$\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{J}_{\boldsymbol{f}} \check{\boldsymbol{x}} = \boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \check{\boldsymbol{y}} \tag{4.7}$$

$$\check{\boldsymbol{x}} = \left(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{J}_{\boldsymbol{f}} \right)^{-1} \boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \check{\boldsymbol{y}} \tag{4.8}$$

The new estimate of $\boldsymbol{x}$ is then formed by means of (4.1).[99] In the next iteration, $\boldsymbol{x}$ is assumed to be the predicted part, $\tilde{\boldsymbol{x}}$, and the iteration loop consisting of evaluating (4.5), (4.8), (4.1) is repeated.

[99] Under certain conditions, the correction value may be excessively high, which may cause the algorithm to diverge, or reconverge rather slowly. If such condition is anticipated, logarithmic damping of the correction may be applied on each element of $\check{\boldsymbol{x}}$:

$$\check{x}' = \operatorname{sign}(\check{x}) \ln\left(1 + |\check{x}|\right).$$

Weighted Newton-Raphson algorithm

In case of weighted, or generalized, least squares (WLS, or GLS) the solution should satisfy condition [140, 141]

$$\hat{\boldsymbol{x}} = \arg\min_{x} \left(\boldsymbol{y}_i - \mathbf{f}(\boldsymbol{x})\right)^{\mathrm{T}} \boldsymbol{W} \left(\boldsymbol{y}_i - \mathbf{f}(\boldsymbol{x})\right), \tag{4.9}$$

where $\boldsymbol{W}$ is a positive-definite weight matrix. In the optimal case, $\boldsymbol{W}$ is the inverse of the measurement covariance matrix $\boldsymbol{\Phi}_{\boldsymbol{y}}$. Nonetheless, it is not always available, so the covariances are neglected and

$$\boldsymbol{W} = \boldsymbol{\Phi}_{\boldsymbol{y}}^{-1} \approx \begin{Bmatrix} 1/\sigma_{y_1}^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_{y_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & 1/\sigma_{y_n}^2 \end{Bmatrix}. \tag{4.10}$$

The diagonal elements represent the weight (or importance) of the corresponding measurement. In general, the values may be set arbitrarily in order to emphasize certain measurements among the others.

Only a slight modification of the non-weighted algorithm is required. In the weighted variant, the correction computation step becomes

$$\check{\boldsymbol{x}} = \left(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{J}_{\boldsymbol{f}}\right)^{-1} \boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{W} \check{\boldsymbol{y}} \tag{4.11}$$

instead of (4.8).[100] The iteration loop consists of residual computation (4.5), correction evaluation (4.11), and correction application (4.1).

4.1.2. Gradient descent algorithm

The gradient, or steepest descent method is substantially slower than the Newton-Raphson method, however, it is less likely to diverge. In each iteration a small parameter change is performed in the direction of the steepest descent of the objective function, i.e. $\|\boldsymbol{y}_i - \mathbf{f}(\boldsymbol{x})\|^2$. The direction of the steepest descent is opposite to the direction of the objective function gradient. It can be shown that the correction [141] can be computed from the residuals by means of

$$\check{\boldsymbol{x}} = \alpha \boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \check{\boldsymbol{y}}. \tag{4.12}$$

The residuals are computed in the same manner as in the Newton-Raphson method, i.e. using (4.5), the correction application follows (4.1). The α is a positive coefficient that determines the step size; smaller step sizes (low α value) slow down the convergence and further reduce risk of diverence.

[100] The non-weighted variant is equivalent to the weighted one, when $\boldsymbol{W}$ is identity matrix.

Weighted gradient descent algorithm

Naturally, a weighted variant of the gradient descent algorithm exists. The only difference from the non-weighted version is in computation of the parameter correction:

$$\check{\boldsymbol{x}} = \alpha \boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{W} \check{\boldsymbol{y}}. \tag{4.13}$$

The weight matrix should be chosen in the same way as in the case of weighted Newton-Raphson solver.

4.1.3. Levenberg-Marquardt algorithm

The Levenberg-Marquardt method combines the Newton-like and the gradient-like convergence. The purpose of such combination is to reduce the risk of divergence while maintaining fast convergence rate.

Only the correction update step is modified; additionally, two variants of the update step may be used. The variant originally proposed by Levenberg [142] can be written in the matrix form as

$$\check{\boldsymbol{x}} = \left(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{J}_{\boldsymbol{f}} + \lambda \mathbb{I} \right)^{-1} \boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \check{\boldsymbol{y}}. \tag{4.14}$$

The term $\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{J}_{\boldsymbol{f}}$ represents the Newton-like update step and the identity matrix designates the gradient-like step. The positive λ coefficient sets which convergence approach is the dominant one; higher values of λ enforce more gradient-like convergence and vice versa. In the extreme case of $\lambda = 0$, the equations is identical to (4.11), i.e. purely Newton-like update step is performed.

The behavior of the gradient-like term will change with the scale of the measurements, in contrast to the Newton-like, which are scale-invariant. Marquardt in [143] proposes pre-scaling of the measurements and accordingly whole problem (i.e. Jacobian matrix) to the units of their standard deviations to compensate for the scale inconsistency. However, more intuitive solution, where the identity matrix is substituted for the diagonal of the $\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{J}_{\boldsymbol{f}}$ matrix term, is widely used:

$$\check{\boldsymbol{x}} = \left(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{J}_{\boldsymbol{f}} + \lambda \operatorname{Diag}(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{J}_{\boldsymbol{f}}) \right)^{-1} \boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \check{\boldsymbol{y}}. \tag{4.15}$$

Such approach is not scale-dependent and the dimension of the measurements does not need to be normalized w.r.t. the standard deviation.

Weighted Levenberg-Marquardt algorithm

The rules for construction of the weight matrix are the same as in the previous cases, the correction update formulas follow

Original:
$$\check{\boldsymbol{x}} = \left(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{J}_{\boldsymbol{f}} + \lambda \mathbb{I} \right)^{-1} \boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{W} \check{\boldsymbol{y}} \tag{4.16}$$

Modified:
$$\check{\boldsymbol{x}} = \left(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{J}_{\boldsymbol{f}} + \lambda \operatorname{Diag}(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{J}_{\boldsymbol{f}}) \right)^{-1} \boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{W} \check{\boldsymbol{y}}. \tag{4.17}$$

Choice of λ parameter

The parameter λ for controlling the convergence damping of the Levenberg-Marquardt algorithm should be high enough to prevent divergence, and concurrently as low as possible in order to converge quickly. It is usually set in a heuristic manner, since the exact λ is usually computationally costly and may even require solving another least-squares problem within each iteration [140].

The common approach is to lower the λ value when the algorithm converges (the sum of square residuals is reduced), and increase λ when reduction of the residuals is not achieved:

$$\lambda_{k+1} = \begin{cases} \lambda_k / c_\lambda^- & \text{for} \quad \|\boldsymbol{y} - \mathbf{f}(\boldsymbol{x})\|^2 < \|\boldsymbol{y} - \mathbf{f}(\tilde{\boldsymbol{x}})\|^2 , \\ \lambda_k c_\lambda^+ & \text{for} \quad \|\boldsymbol{y} - \mathbf{f}(\boldsymbol{x})\|^2 \geq \|\boldsymbol{y} - \mathbf{f}(\tilde{\boldsymbol{x}})\|^2 . \end{cases} \tag{4.18}$$

If the weighted variant of the Levenberg-Marquard algorithms is used, weighted sum of residuals should be used, i.e. the metric follows (4.9) instead of (4.3):

$$\lambda_{k+1} = \begin{cases} \lambda_k / c_\lambda^- & \text{for } (\boldsymbol{y} - \mathbf{f}(\boldsymbol{x}))^{\mathrm{T}} \boldsymbol{W} (\boldsymbol{y} - \mathbf{f}(\boldsymbol{x})) < (\boldsymbol{y}_i - \mathbf{f}(\tilde{\boldsymbol{x}}))^{\mathrm{T}} \boldsymbol{W} (\boldsymbol{y} - \mathbf{f}(\tilde{\boldsymbol{x}})) , \\ \lambda_k c_\lambda^+ & \text{for } (\boldsymbol{y} - \mathbf{f}(\boldsymbol{x}))^{\mathrm{T}} \boldsymbol{W} (\boldsymbol{y} - \mathbf{f}(\boldsymbol{x})) \geq (\boldsymbol{y} - \mathbf{f}(\tilde{\boldsymbol{x}}))^{\mathrm{T}} \boldsymbol{W} (\boldsymbol{y} - \mathbf{f}(\tilde{\boldsymbol{x}})) . \end{cases} \tag{4.19}$$

Numerous variants of this heuristic approach do exist, for instance [143, 144] suggest that good results may be obtained when the λ-modification coefficients $c_\lambda^- = c_\lambda^+$, whilst [141] uses different values for increasing and decreasing λ.[101] It is rather convenient to provide bounds to the lambda value, which are spanning several orders of magnitude, e.g. $\lambda \in \langle 10^{-10}, 10^5 \rangle$.

Some implementations do not allow any update that would not lead to reduction in the minimized metric; for instance in [145, 146] a line search is performed in order to find the lowest possible λ. In each iteration it is checked whether the update leads to the reduction of the sum of squared residuals. If it does, the update is accepted and λ is reduced for the next step. Otherwise, the update is not applied and iteration is repeated with increased λ value until the update decreases the sum of squares.[102]

Authors of [146] and [141] suggest using "gain factor" metric[103] of estimate improvement for decision whether to change λ, [146] proposes a smooth mapping of the c_λ to the metric.

4.1.4. Termination conditions

The iterations of the algorithms are stopped when one or more conditions are met. First, the iterations should be stopped when the estimate fits to the measurements taken, i.e. when the residuals are sufficiently low. An apparent metric for the residual-based iterations is the (weighted) sum of residuals, which is the value to be minimized

[101] The actual values of the coefficients c_λ are only suggested in the publications. For instance in [143] $c_\lambda = 10$, in [144] $c_\lambda = 5$ and in [141] $c_\lambda^- = 9$, $c_\lambda^+ = 11$.

[102] In certain cases (e.g. when the metric function is not unimodal, and contains smaller false minima) it is not practical to disallow the temporary increase in the minimized metric.

[103] In simplicity, the actual change of the sums of squared residuals is normalized w.r.t. the anticipated improvement expected from the step (based on the Jacobian matrix). See [141] for the exact definition.

by the iterative algorithm. It is convenient, however not necessary, to use the mean of the residuals, since there is no need to change the threshold ϵ_{res} when the number of measurements (length of $\boldsymbol{y}$) is altered:

Non-weighted:
$$\frac{\|\boldsymbol{y} - \mathbf{f}(\boldsymbol{x})\|^2}{\dim \boldsymbol{y}} < \epsilon_{\text{res}} \tag{4.20}$$

Weighted:
$$\frac{(\boldsymbol{y} - \mathbf{f}(\boldsymbol{x}))^{\text{T}} \boldsymbol{W} (\boldsymbol{y} - \mathbf{f}(\boldsymbol{x}))}{\text{tr}\, \boldsymbol{W}} < \epsilon_{\text{res}}, \tag{4.21}$$

where $\text{tr}\, \boldsymbol{W}$ denotes the trace of the weight matrix and $\dim \boldsymbol{y}$ denotes the dimension of $\boldsymbol{y}$, i.e. the number of measurements.

Another terminating occasion is when the magnitude of the correction falls below a threshold, i.e. $\|\check{\boldsymbol{x}}\| < \epsilon_{\text{cor}}$. In certain sources the maximum of the $\check{\boldsymbol{x}}$-elements is compared to the threshold. Moreover, in [141], the elements of the correction are normalized w.r.t. the predicted value $\check{x}_{\text{i}}/\tilde{x}_{\text{i}}$. Furthermore, the relative improvement of the sum of squared residuals is a possible metric that can be used to terminate the iterations [140].

It is also possible to stop the algorithm after a fixed number of iterations. This may be considered as a stopping condition for problems that are known to converge within a few iterations, or it can be used to break the infinite loop when the algorithm does not converge, i.e. when no other stop condition is met.

Certainly, other options of terminating the iterations are possible, either exact or ad-hoc.

4.1.5. Application on the navigation problems

I. Time of arrival (ToA)

For simplicity, we will consider that the pseudorange ρ and the bias of the user clock w.r.t. the system clock is expressed in the units of length (it is multiplied via the propagation velocity). Recapitulating from Chapter 2, pseudorange can be expressed as

$$\rho_{\text{i}} = \|\boldsymbol{r}_{\text{u}} - \boldsymbol{r}_{\text{i}}\| + ., \tag{2.59}$$

Borrowing the GNSS terminology, $\boldsymbol{r}_{\text{u}}$ is the unknown user receiver position vector and $\boldsymbol{r}_{\text{i}}$ is the known position of the satellite, to which the pseudorange ρ_{i} is measured. Hence, it is straightforward to map these variables to the generic vector of parameters $\boldsymbol{x}$ and vector of measurements $\boldsymbol{y}$

$$\boldsymbol{x} \leftrightarrow \begin{bmatrix} \boldsymbol{r}_{\text{u}} \\ b \end{bmatrix} = \begin{bmatrix} r_{\text{u};x} \\ r_{\text{u};y} \\ r_{\text{u};z} \\ b \end{bmatrix} \qquad \boldsymbol{x} \leftrightarrow \boldsymbol{\rho} = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_{\text{i}} \\ \vdots \\ \rho_N \end{bmatrix}. \tag{4.22}$$

In order to be able to perform the iterative solution, Jacobian matrix $\boldsymbol{J}$ is needed (the $\boldsymbol{f}$ subscript is omitted). Each row of the matrix (further denoted as $\boldsymbol{J}_{\text{i}}$) contains

partial derivatives w.r.t. the user position $\boldsymbol{r}_\mathrm{u}$ and clock bias b:

$$\begin{aligned}\boldsymbol{J}_\mathrm{i} &= \begin{bmatrix}\frac{\partial\rho_\mathrm{i}}{\boldsymbol{r}_\mathrm{u}} & \frac{\partial\rho_\mathrm{i}}{b}\end{bmatrix} = \begin{bmatrix}\frac{\partial\rho_\mathrm{i}}{r_{\mathrm{u};x}} & \frac{\partial\rho_\mathrm{i}}{r_{\mathrm{u};y}} & \frac{\partial\rho_\mathrm{i}}{r_{\mathrm{u};z}} & \frac{\partial\rho_\mathrm{i}}{b}\end{bmatrix} \\ &= \begin{bmatrix}\frac{(\boldsymbol{r}_\mathrm{u}-\boldsymbol{r}_\mathrm{i})^\mathrm{T}}{\|\boldsymbol{r}_\mathrm{u}-\boldsymbol{r}_\mathrm{i}\|} & 1\end{bmatrix} = \begin{bmatrix}\mathbb{1}_{\mathrm{u,i}} & 1\end{bmatrix} = \begin{bmatrix}\frac{r_{\mathrm{u}x}-r_{\mathrm{i};x}}{\|\boldsymbol{r}_\mathrm{u}-\boldsymbol{r}_\mathrm{i}\|} & \frac{r_{\mathrm{u}y}-r_{\mathrm{i};y}}{\|\boldsymbol{r}_\mathrm{u}-\boldsymbol{r}_\mathrm{i}\|} & \frac{r_{\mathrm{u}z}-r_{\mathrm{i};z}}{\|\boldsymbol{r}_\mathrm{u}-\boldsymbol{r}_\mathrm{i}\|} & 1\end{bmatrix}. \end{aligned} \tag{4.23}$$

From the various notations of the same relation it can be seen that each row contains the direction vector with unit length pointing from the satellite to the user, i.e. from infrastructure node to the localized node.[104] The general $N \times 4$ Jacobian matrix is then computed in each iteration of the chosen least-squares algorithm. The residuals are computed as

$$\breve{\rho}_\mathrm{i} = \rho_\mathrm{i} - \left(\|\tilde{\boldsymbol{r}}_\mathrm{u} - \boldsymbol{r}_\mathrm{i}\| + \tilde{b}\right). \tag{4.24}$$

The individual pseudorange measurements can be considered independent. Consequently, the covariance matrix is diagonal and easy to invert. The weight matrix can be formulated as (4.25) where the approximation is valid when the pseudorange variances are approximately equal.

$$\boldsymbol{W} = \mathrm{Diag}\begin{bmatrix}\sigma_{\rho 1}^{-2} & \cdots & \sigma_{\rho \mathrm{i}}^{-2} & \cdots & \sigma_{\rho N}^{-2}\end{bmatrix} \approx \sigma_\rho^{-2}\mathbb{I} \tag{4.25}$$

II. Two-way ranging (TWR)

In this case, the observables are the estimates of the "radio" ranges between the localized equipment and the localizing infrastructure. Such problem can be described by a set of the of the equations

$$r_{\mathrm{u}i} = \|\boldsymbol{r}_\mathrm{u} - \boldsymbol{r}_\mathrm{i}\|. \tag{2.11}$$

The equation is not valid for the TWR exclusively; theoretically, it can be employed in positioning by means of signal strength or another system, which uses bi-directional message transfer in order to obtain range, such as DME in aviation. The mapping of the parameter and measurement vectors is straightforward:

$$\boldsymbol{x} \leftrightarrow \boldsymbol{r}_\mathrm{u} = \begin{bmatrix}r_{\mathrm{u};x}\\ r_{\mathrm{u};y}\\ r_{\mathrm{u};z}\end{bmatrix} \qquad \boldsymbol{x} \leftrightarrow \boldsymbol{r}_{\mathrm{u}i} = \begin{bmatrix}r_{\mathrm{u}1}\\ \vdots\\ r_{\mathrm{u}i}\\ \vdots\\ r_{\mathrm{u}N}\end{bmatrix}. \tag{4.26}$$

The rows of the Jacobian (or geometry) matrix are the same as in (4.23) up to the last element, which is missing in the case of range measurements:

$$\begin{aligned}\boldsymbol{J}_\mathrm{i} &= \frac{\partial r_{\mathrm{u}i}}{\boldsymbol{r}_\mathrm{u}} = \begin{bmatrix}\frac{\partial r_{\mathrm{u}i}}{r_{\mathrm{u};x}} & \frac{\partial r_{\mathrm{u}i}}{r_{\mathrm{u};y}} & \frac{\partial r_{\mathrm{u}i}}{r_{\mathrm{u};z}}\end{bmatrix} \\ &= \frac{(\boldsymbol{r}_\mathrm{u} - \boldsymbol{r}_\mathrm{i})^\mathrm{T}}{\|\boldsymbol{r}_\mathrm{u} - \boldsymbol{r}_\mathrm{i}\|} = \mathbb{1}_{\mathrm{u,i}} = \begin{bmatrix}\frac{r_{\mathrm{u}x}-r_{\mathrm{i};x}}{\|\boldsymbol{r}_\mathrm{u}-\boldsymbol{r}_\mathrm{i}\|} & \frac{r_{\mathrm{u}y}-r_{\mathrm{i};y}}{\|\boldsymbol{r}_\mathrm{u}-\boldsymbol{r}_\mathrm{i}\|} & \frac{r_{\mathrm{u}z}-r_{\mathrm{i};z}}{\|\boldsymbol{r}_\mathrm{u}-\boldsymbol{r}_\mathrm{i}\|}\end{bmatrix}. \end{aligned} \tag{4.27}$$

[104] In the GNSS-related literature this particular Jacobian matrix is called the geometry matrix and denoted by $\boldsymbol{G}$. Such notation has been used in the GNSS-related section, see (2.123).

The residual computation follows

$$\check{r}_{\mathrm{ui}} = r_{\mathrm{ui}} - (\|\tilde{\boldsymbol{r}}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{i}}\|)\,. \tag{4.28}$$

Under the assumption of stochastically independent range measurement is the weight matrix $\boldsymbol{W}$ constructed in the same manner as in (4.25).

III. Time difference of Arrival (TDoA)

According to Chapter 2, the TDoA problem is described by

$$d_{\mathrm{i,j}} = \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{i}}\| - \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{j}}\|\,. \tag{2.64}$$

It is possible to derive the corresponding row of the Jacobian matrix as

$$\begin{aligned}\boldsymbol{J}_{d\mathrm{i,j}} &= \frac{\partial d_{\mathrm{i,j}}}{\boldsymbol{r}_{\mathrm{u}}} = \frac{(\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{i}})^{\mathrm{T}}}{\|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{i}}\|} - \frac{(\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{j}})^{\mathrm{T}}}{\|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{j}}\|} = \mathbb{1}_{\mathrm{u,i}} - \mathbb{1}_{\mathrm{u,j}} \\ &= \left[\frac{r_{\mathrm{u}x} - r_{\mathrm{i};x}}{\|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{i}}\|} - \frac{r_{\mathrm{u}x} - r_{\mathrm{j};x}}{\|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{j}}\|} \quad \frac{r_{\mathrm{u}y} - r_{\mathrm{i};y}}{\|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{i}}\|} - \frac{r_{\mathrm{u}y} - r_{\mathrm{j};y}}{\|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{j}}\|} \quad \frac{r_{\mathrm{u}z} - r_{\mathrm{i};z}}{\|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{i}}\|} - \frac{r_{\mathrm{u}z} - r_{\mathrm{j};z}}{\|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{j}}\|}\right]. \end{aligned} \tag{4.29}$$

It is necessary to avoid the situation when the $\boldsymbol{J}$-matrix is rank deficient. This is equivalent to saying that the lines of $\boldsymbol{J}$ are linearly-dependent; for instance, such condition may occur when three differences are produced from three measurements, e.g. $d_{\mathrm{i,j}}$, $d_{\mathrm{j,k}}$ and $d_{\mathrm{i,k}}$.

In the Section 2.1.2 it was mentioned that TDoA can be constructed as a difference of two pseudoranges and that the variance of the TDoA measurement is the sum of the two pseudorange variances. Consider now three measurements $d_{\mathrm{i,j}}$, $d_{\mathrm{i,k}}$ and $d_{\mathrm{l,i}}$. Unless the geometry of the nodes is linear, the corresponding lines of the Jacobian matrix are independent. Nonetheless, the measurements are not stochastically independent, since they share the pseudorange measurement to the "i." Therefore, the weight matrix $\boldsymbol{W}$ should be constructed using a non-diagonal covariance matrix:

$$\boldsymbol{\Phi}_D = \begin{bmatrix} \sigma_{\rho\mathrm{i}}^2 + \sigma_{\rho\mathrm{j}}^2 & +\sigma_{\rho\mathrm{i}}^2 & -\sigma_{\rho\mathrm{i}}^2 \\ +\sigma_{\rho\mathrm{i}}^2 & \sigma_{\rho\mathrm{i}}^2 + \sigma_{\rho\mathrm{k}}^2 & 0 \\ -\sigma_{\rho\mathrm{i}}^2 & 0 & \sigma_{\rho\mathrm{i}}^2 + \sigma_{\rho\mathrm{k}}^2 \end{bmatrix}. \tag{4.30}$$

Under the assumption of equal pseudorange standard deviations σ_ρ we may define the TDoA standard deviation as $\sigma_{\mathrm{d}} = 2\sigma_\rho$. Further, the covariance matrix follows

$$\boldsymbol{\Phi}_d = \sigma_\rho \begin{bmatrix} 2 & 1 & & -1 & & \\ 1 & 2 & 0 & -1 & 0 & 2 \end{bmatrix} = \sigma_{\mathrm{d}} \begin{bmatrix} 1 & 0.5 & -0.5 \\ 0.5 & 1 & 0 \\ -0.5 & 0 & 1 \end{bmatrix}. \tag{4.31}$$

IV. Angle of arrival (AoA)

For the angle of arrival observable, we will investigate only the two-dimensional case; however, three-dimensional description is available e.g. in [147]. In the following example, the x-axis is pointing North, the y-axis is pointing East. The measured bearing angle β_i is the azimuth from the user to the node with position vector $\boldsymbol{r}_\mathrm{i}$. Thus,

$$\beta_\mathrm{i} = \tan^{-1} \frac{r_{\mathrm{i};y} - r_{\mathrm{u};y}}{r_{\mathrm{i};x} - r_{\mathrm{u};x}} \tag{4.32}$$

$$\boldsymbol{J}_\mathrm{i} = \frac{\partial \beta_\mathrm{i}}{\partial \boldsymbol{r}_\mathrm{u}} = \left[-\frac{r_{\mathrm{i};y} - r_{\mathrm{u};y}}{\|\boldsymbol{r}_\mathrm{i} - \boldsymbol{r}_\mathrm{u}\|^2} \quad \frac{r_{\mathrm{i};x} - r_{\mathrm{u};x}}{\|\boldsymbol{r}_\mathrm{i} - \boldsymbol{r}_\mathrm{u}\|^2} \right]. \tag{4.33}$$

If the bearing measurement is not taken w.r.t. the North (x-axis), but with respect to an unknown azimuth ψ, it holds

$$\beta_\mathrm{i} = \tan^{-1} \frac{r_{\mathrm{i};y} - r_{\mathrm{u};y}}{r_{\mathrm{i};x} - r_{\mathrm{u};x}} - \psi, \tag{4.34}$$

$$\boldsymbol{J}_\mathrm{i} = \left[\frac{\partial \beta_\mathrm{i}}{\partial \boldsymbol{r}_\mathrm{u}} \quad \frac{\partial \beta_\mathrm{i}}{\partial \psi} \right] = \left[-\frac{r_{\mathrm{i};y} - r_{\mathrm{u};y}}{\|\boldsymbol{r}_\mathrm{i} - \boldsymbol{r}_\mathrm{u}\|^2} \quad \frac{r_{\mathrm{i};x} - r_{\mathrm{u};x}}{\|\boldsymbol{r}_\mathrm{i} - \boldsymbol{r}_\mathrm{u}\|^2} \quad -1 \right]. \tag{4.35}$$

In such case, ψ is part of the parameter vector, and has to be estimated. The need to estimate the azimuth ψ can be eliminated by differentiation of the bearing measurements:

$$B_{\mathrm{i,j}} = \beta_\mathrm{i} - \beta_\mathrm{j} = \tan^{-1} \frac{r_{\mathrm{i};y} - r_{\mathrm{u};y}}{r_{\mathrm{j};x} - r_{\mathrm{u};x}} - \tan^{-1} \frac{r_{\mathrm{j};y} - r_{\mathrm{u};y}}{r_{\mathrm{j};x} - r_{\mathrm{u};x}}, \tag{4.36}$$

$$\boldsymbol{J}_{B\mathrm{i,j}} = \frac{\partial B_\mathrm{i}}{\partial \boldsymbol{r}_\mathrm{u}} = \left[-\frac{r_{\mathrm{i};y} - r_{\mathrm{u};y}}{\|\boldsymbol{r}_\mathrm{i} - \boldsymbol{r}_\mathrm{u}\|^2} + \frac{r_{\mathrm{j};y} - r_{\mathrm{u};y}}{\|\boldsymbol{r}_\mathrm{j} - \boldsymbol{r}_\mathrm{u}\|^2} \quad \frac{r_{\mathrm{i};x} - r_{\mathrm{u};x}}{\|\boldsymbol{r}_\mathrm{i} - \boldsymbol{r}_\mathrm{u}\|^2} - \frac{r_{\mathrm{j};x} - r_{\mathrm{u};x}}{\|\boldsymbol{r}_\mathrm{j} - \boldsymbol{r}_\mathrm{u}\|^2} \right]. \tag{4.37}$$

It is worth noting that the tree described variants can be seen as an angular analogy to the TWR, ToA and TDoA problems, respectively. The weight matrices of the angular measurements are constructed using the same principles as were described in the previous sections. It should be remarked that the possible covariances should be taken into account when using the differentiated measurements in order to avoid estimation of ψ.

V. Combined estimation

It is worth remarking that it is possible to combine the ToA, TWR and TDoA equations in a common set of equations, if the coordinate frames are identical. All the equations feature the same user-position vector. The TWR and TDoA measurements are insensitive to the clock bias value, therefore their partial derivative w.r.t. b is zero. For example, the Jacobian matrix for estimating position from $\mathbf{y} = [\rho_1,\, \rho_2,\, r_{\mathrm{u}3},\, r_{\mathrm{u}4},\, d_{5,6}]^\mathrm{T}$ and solving for $\boldsymbol{x} = [\boldsymbol{r}_\mathrm{u},\, b]^\mathrm{T}$ would follow

$$\boldsymbol{J} = \begin{bmatrix} \mathbb{1}_{\mathrm{u},1} & 1 \\ \mathbb{1}_{\mathrm{u},2} & 1 \\ \mathbb{1}_{\mathrm{u},3} & 0 \\ \mathbb{1}_{\mathrm{u},4} & 0 \\ \mathbb{1}_{\mathrm{u},5} - \mathbb{1}_{\mathrm{u},6} & 0 \end{bmatrix}.$$

4.1.6. Estimate covariance and dilution of precision

In this section we consider the weighted scenario, where the weighted matrix is the inverse of the covariance matrix, or its approximation (when only the diagonal elements are assumed). In such case the covariance matrix of the parameter vector $\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}$ is

$$\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}} = \left(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{J}_{\boldsymbol{f}}\right)^{-1}, \tag{4.38}$$

regardless which one of the algorithms described above is used. The vector containing the standard deviations of the estimated parameters $\boldsymbol{\sigma}_{\hat{\boldsymbol{x}}}$ can by obtained by means of

$$\boldsymbol{\sigma}_{\hat{\boldsymbol{x}}} = \sqrt{\operatorname{diag} \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}} = \sqrt{\operatorname{diag} \left(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{J}_{\boldsymbol{f}}\right)^{-1}}, \tag{4.39}$$

where the "diag" operator produces the vector of the diagonal elements of a matrix.

Error ellipse

When assuming multivariate Gaussian distribution, it is possible to construct an error ellipse which is based on the covariance matrix. Such shape depicts is the area, where the true value lies with a certain probability.

For simplicity, we will now consider only the 1σ-error ellipse in an xy-plane. When there is no covariance between the random variables x and y, the axes of the error ellipse are aligned with the x and y-axes, the respective semi-axes have length of the corresponding standard deviations σ_x and σ_y.

When a covariance between x and y is present, the error ellipse is scaled and rotated. It can be shown that the length of the semi-axes is given by the square root of *eigenvalues* of the covariance matrix, whilst the direction of the semi-axes is determined by the respective *eigenvectors*. The situation is depicted in Figure 4.1; the rotation of the ellipse θ is the arctangent of the elements of the eigenvector corresponding to the maximal eigenvalue.

Regardless of whether a covariance is present, the center of the ellipse coincides with the means of the random variables, or the actual estimates. The 1σ-ellipse may be scaled by parameter η in order to reflect a particular probability of the true value lying within the error ellipse. Under assumption of multivariate Gaussian distribution, it holds

$$\Pr\left(|x - \hat{x}| < \eta\sigma\right) = \operatorname{erf}\left(\frac{\eta}{\sqrt{2}}\right), \tag{4.40}$$

where $\operatorname{erf}(\cdot)$ denotes the error function.[105] Table 4.1 lists some of the useful relations between η coefficient and the probability.

It is possible to use a similar approach to construct error ellipsoid for visualization in three dimensions; the three eigenvalues and eigenvectors still define the ellipsoid. The ellipses are more common than ellipsoids, however.

Relation to DOP

Estimate covariance is inherently related to the dilution of precision, DOP, which was mentioned in the GNSS context within Section 2.2.2–I. Under the assumption

[105] The error function is bound to the distribution function of the normal (Gaussian) distribution.

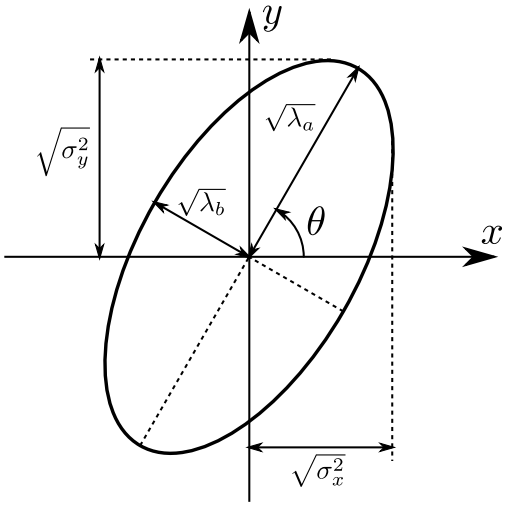


FIGURE 4.1.: 1σ-error ellipse and its parameters.

TABLE 4.1.: Probability of true value laying within the error ellipse.

η	$\Pr(\lvert x-\hat{x}\rvert < \eta\sigma)$	η	$\Pr(\lvert x-\hat{x}\rvert < \eta\sigma)$
1	0.682 689	0.674 490	50.000 %
1.5	0.866 386	1.644 853	90.000 %
2	0.954 500	1.959 963	95.000 %
3	0.997 300	2.575 829	99.000 %
4	0.999 936	3.290 526	99.900 %
5	0.999 999	3.890 591	99.990 %

that all the measurements have equal standard deviation σ_y and are independent, i.e. covariances are zero, the standard deviation of the estimates $\boldsymbol{\sigma}_{\hat{\boldsymbol{x}}}$ follow:

$$\boldsymbol{\sigma}_{\hat{\boldsymbol{x}}} = \sqrt{\operatorname{diag}\left(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}}(\sigma_y \mathbb{I})^{-2}\boldsymbol{J}_{\boldsymbol{f}}\right)^{-1}} = \sigma_y \sqrt{\operatorname{diag}\left(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}}\boldsymbol{J}_{\boldsymbol{f}}\right)^{-1}}. \tag{4.41}$$

The matrix is often abbreviated to $\boldsymbol{H} = \left(\boldsymbol{J}_{\boldsymbol{f}}^{\mathrm{T}}\boldsymbol{J}_{\boldsymbol{f}}\right)^{-1}$ (sometimes written as $\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\right)^{-1}$); consequently,

$$\sigma_{\hat{\boldsymbol{x}};1} = \sigma_y\sqrt{H_{11}} \qquad \sigma_{\hat{\boldsymbol{x}};1,2} = \sqrt{\sigma_{\hat{\boldsymbol{x}};1}^2 + \sigma_{\hat{\boldsymbol{x}};2}^2} = \sigma_y\sqrt{H_{11}+H_{22}}, \tag{4.42}$$

et cetera. The square roots of the diagonal elements of $\boldsymbol{H}$ and their combinations are referred as the DOP factors. They can be understood as the measures of the user-infrastructure geometry quality, since they reflect how the imprecision of the measurement affects the position estimate with a particular geometry.

When the DOP factors are computed in the ENU frame, they can be interpreted as was suggested in Table 2.4 on page 55. The DOP factors are not limited to the

GNSS field, they can be computed for any of the ToA, TWR, TDoA, or AoA problems. Nonetheless, extra care has to be taken when computing DOP in the TDoA case, since the measurements may be significantly correlated.[106]

4.1.7. Pseudomeasurement as a soft equality constraint

Consider an example: An UWB tag is mounted on a helmet, which is worn by a person. It is practical to aid the position estimate process by such information in a form of a constraint. It is not possible to define a hard constraint on the vertical position, since we cannot say that the tag is e.g. 177.5 cm above the reference plane. However, it is likely that the helmet is located *around* 175 cm – the actual height of the person with helmet is uncertain, and the vertical position of head is not constant when walking.

Such information may be input to the least-squares iterative solution as a pseudomeasurement. We will denote the vector of pseudomeasurements by $\boldsymbol{y}'$ and their covariance matrix $(\boldsymbol{W}')^{-1}$. For the sake simplicity we will assume Gaussian distribution of the pseudomeasurement; the model follows

$$\boldsymbol{y}' = \mathrm{g}(\boldsymbol{x}) + \mathcal{N}\left(0,\, (\boldsymbol{W}')^{-1}\right), \tag{4.43}$$

where g is the constraint function. In the example mentioned above, the function returns a scalar value – the height above ground.

Here, the extension of the Newton-Raphson algorithm will be shown, however, the modification of the gradient and Levenberg-Marquardt method is analogical. Provided that there is no correlation between the measurements and the psudomeasurements, the modified weighted least-squares criterion is

$$\hat{\boldsymbol{x}} = \arg\min_{x} \begin{bmatrix} \boldsymbol{y} - \mathbf{f}(\boldsymbol{x}) \\ \boldsymbol{y}' - \mathbf{g}(\boldsymbol{x}) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{W} & \mathbb{O} \\ \mathbb{O} & \boldsymbol{W}' \end{bmatrix} \begin{bmatrix} \boldsymbol{y} - \mathbf{f}(\boldsymbol{x}) \\ \boldsymbol{y}' - \mathbf{g}(\boldsymbol{x}) \end{bmatrix}, \tag{4.44}$$

where $\mathbb{O}$ is a properly sized matrix of zeros.

The first modification of the Newton-Raphson algorithm is in the extension of residual computation:

$$\begin{bmatrix} \check{\boldsymbol{y}} \\ \check{\boldsymbol{y}}' \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} - \tilde{\boldsymbol{y}} \\ \boldsymbol{y}' - \tilde{\boldsymbol{y}}' \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} - \mathbf{f}(\tilde{\boldsymbol{x}}) \\ \boldsymbol{y}' - \mathbf{g}(\tilde{\boldsymbol{x}}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}' \end{bmatrix} - \begin{bmatrix} \mathbf{f}(\tilde{\boldsymbol{x}}) \\ \mathbf{g}(\tilde{\boldsymbol{x}}) \end{bmatrix}. \tag{4.45}$$

The second modification is in the correction update computation. From comparison of (4.45) with (4.5) we may deduce that first order approximation of the residuals follows

$$\begin{bmatrix} \check{\boldsymbol{y}} \\ \check{\boldsymbol{y}}' \end{bmatrix} \approx \begin{bmatrix} \boldsymbol{J_f}|_{\tilde{\boldsymbol{x}}} \\ \boldsymbol{J_g}|_{\tilde{\boldsymbol{x}}} \end{bmatrix} \check{\boldsymbol{x}}. \tag{4.46}$$

Consequently, for the constrained weighted Newton-Raphson algorithm the correction update equation becomes

$$\check{\boldsymbol{x}} = \left(\begin{bmatrix} \boldsymbol{J_f} \\ \boldsymbol{J_g} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{W} & \mathbb{O} \\ \mathbb{O} & \boldsymbol{W}' \end{bmatrix} \begin{bmatrix} \boldsymbol{J_f} \\ \boldsymbol{J_g} \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{J_f} \\ \boldsymbol{J_g} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{W} & \mathbb{O} \\ \mathbb{O} & \boldsymbol{W}' \end{bmatrix} \begin{bmatrix} \check{\boldsymbol{y}} \\ \check{\boldsymbol{y}}' \end{bmatrix}. \tag{4.47}$$

[106] The correlation (covariance normalized by variance) of the TDoA measurements is $\pm 1/2$.

The application of the parameter vector correction is identical to the unconstrained algorithm, i.e. follows (4.1). The pseudomeasurement approach is applied in Appendix B, where the TDoA-based estimation in an UWB system using Levenberg-Marquardt algorithm with pseudomeasurement is presented.

Clearly, the pseudomeasurements are not the only possible approach to constraining the least-squares solution. For instance in [148] the solution using Lagrange multipliers is shown. For the hard-constrained cases, the modifications of Levenberg-Marquardt algorithm is available in [149].

4.2. Kalman Filtering and Data Fusion

The Kalman filtering is a method that is used to estimate a system state (in navigation often position and velocity) on the basis of a related series of measurements of the system, observed over time. According to [150], it is the minimum-variance state estimator for linear dynamic systems, regardless whether the system noise is Gaussian or not.

In the following text we will show the basic equations for the linear Kalman filter (KF) and its nonlinear variants extended and unscented Kalman filters (EKF and UKF). Examples of use by author are available in the last two appendices of the thesis. Within Appendix C, the linear KF is used for estimating the clock state of a UWB transceiver, and for smoothing the position estimates as well. In Appendix D, EKF is used as a complementary filter in the integrated GNSS-UWB-IMU system.

4.2.1. Linear Kalman filter

Within the thesis we will provide only the basic description of the filters, the rigorous derivation is beyond the scope of the thesis; it is available e.g. in [151].

For the linear system holds that the evolution of the state vector $\boldsymbol{x}$ from epoch k and $k+1$ can be described by

$$\boldsymbol{x}[k+1] = \boldsymbol{F}\boldsymbol{x}[k] + \mathcal{N}\{0, \boldsymbol{Q}\}, \tag{4.48}$$

where $\boldsymbol{F}$ is the *transition matrix* and $\boldsymbol{Q}$ is the covariance matrix of the system noise. The transition matrix reflects the dynamic model of the system. The measurements, or observations, $\boldsymbol{y}$ of the system follow

$$\boldsymbol{y}[k] = \boldsymbol{H}\boldsymbol{x}[k] + \mathcal{N}\{0, \boldsymbol{R}\}, \tag{4.49}$$

where $\boldsymbol{H}$ is the measurement model matrix and $\boldsymbol{R}$ is the measurement covariance matrix. The goal of the Kalman filter is to estimate the state vector $\boldsymbol{x}$ and its covariance matrix $\boldsymbol{P}$.

The algorithm consists of two steps – the prediction and the measurement update. First, we will describe the prediction of state and its covariance from epoch $k-1$ to epoch k. The filter state $\boldsymbol{x}^+[k-1]$ is considered known, the $+$ superscript denotes the *a posteriori* value. The covariance matrix is denoted by $\boldsymbol{P}^+[k-1]$. For the linear system it holds that the *a priori* (denoted by the $-$ superscript) state for epoch k and its covariance can be predicted by means of

$$\boldsymbol{x}^-[k] = \boldsymbol{F}\boldsymbol{x}^+[k-1] \tag{4.50}$$

$$\boldsymbol{P}^-[k] = \boldsymbol{F}\boldsymbol{P}^+[k-1]\boldsymbol{F}^{\mathrm{T}} + \boldsymbol{Q}. \tag{4.51}$$

In this step, the state estimate becomes more uncertain, since the variances (i.e. the covariance matrix) always grow.

During the measurement update, the prediction is confronted with the measurements, and the predictions are corrected accordingly. First, innovations $\boldsymbol{z}[k]$ and their

covariance matrix $\boldsymbol{S}[k]$ are computed using

$$\boldsymbol{z}[k] = \boldsymbol{y}[k] - \boldsymbol{H}\boldsymbol{x}^{-}[k] \tag{4.52}$$

$$\boldsymbol{S}[k] = \boldsymbol{H}\boldsymbol{P}^{-}[k]\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}. \tag{4.53}$$

It is worth noting that the innovations may be understood as the measurement residuals. Next, the Kalman gain is obtained as

$$\boldsymbol{K}[k] = \boldsymbol{P}^{-}[k]\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{S}[k])^{-1}. \tag{4.54}$$

The Kalman gain is a matrix that decides how much the innovations will be applied to the state vector. The *a posteriori* estimate of the state vector and its covariance is found by means of

$$\boldsymbol{x}^{+}[k] = \boldsymbol{x}^{-}[k] + \boldsymbol{K}[k]\boldsymbol{z}[k] \tag{4.55}$$

$$\boldsymbol{P}^{+}[k] = (\mathbb{I} - \boldsymbol{K}[k]\boldsymbol{H}[k])\,\boldsymbol{P}^{-}[k]\,(\mathbb{I} - \boldsymbol{K}[k]\boldsymbol{H}[k])^{\mathrm{T}} + \boldsymbol{K}[k]\boldsymbol{R}\boldsymbol{K}^{\mathrm{T}}[k]. \tag{4.56}$$

Remarkably, this is not the only possible way of propagating the state covariance, other variants are available e.g. in [151].

The measurement model and the transition matrices $\boldsymbol{H}$ and $\boldsymbol{F}$ have to be known, naturally. In the ideal case, the system noise and measurement covariance matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ are known as well. Nonetheless, the full covariance matrices are often not available, so they have to be guessed and tuned by the KF designer ad hoc, in order to provide the desired results. Of course, such guess should reflect the nature and expected variance of the measurements. Practical considerations for tuning the filter are provided at the end of this section.

It is worth noting that all of the matrices describing the system and measurements may be different in each epoch.

Variable epoch interval

Sometimes, the time interval between epochs is not constant (e.g. in Appendix C). Inherently, the temporally-dependent terms of the transition matrix $\boldsymbol{F}$ need to be adjusted. However, the process noise covariance $\boldsymbol{Q}$ has to be scaled as well. Assuming the process noise to be multivariate zero-mean Gaussian random walk, the $\boldsymbol{Q}$ for k-th epoch (i.e. transition from k to $k+1$) is

$$\boldsymbol{Q}[k] = \tau[k]\boldsymbol{Q}_0, \tag{4.57}$$

where $\tau[k]$ is the time interval between the respective epochs and $\boldsymbol{Q}_0$ is the covariance matrix normalized w.r.t. the unit time. The measurement model and covariance are usually independent of the epoch interval $\tau[k]$, thus, $\boldsymbol{H}$ and $\boldsymbol{R}$ are not affected.

Practical considerations

The real systems are often not described perfectly by the $\boldsymbol{F}$, $\boldsymbol{Q}$, $\boldsymbol{H}$ and $\boldsymbol{R}$ matrices; also, the noise is not always Gaussian and zero-mean. Therefore, the theoretical values have

to be adapted to the real scenario – typically by means of simulations, or experimentally (either directly or by post-processing of captured measurements).

The system noise covariance matrix $\boldsymbol{Q}$ is commonly chosen as a diagonal matrix. The variances (diagonal elements of $\boldsymbol{Q}$) then imply how quickly the state is allowed to change in time. Rapidly changing state-variables exhibit higher variances than the slowly changing ones. Too low values in $\boldsymbol{Q}$ may lead to divergence of the Kalman filter; it is often practical to overestimate the variance.[107] Overestimating $\boldsymbol{Q}$ may be considered as trading the reliable operation for the accuracy of estimation. Similarly, it may be beneficial to reduce the variances (or covariances) of measurements ($\boldsymbol{R}$-matrix) in order to emphasize the measurements over the predictions. Alternatively, fading-memory filter may be used [151].

It is necessary to consider arithmetic precision limitations as well. Most often, the covariance matrices and operations with them cause problems due to their quadratic nature. Solution to this may be convenient scaling of the state variables or increasing arithmetic precision. In software such as MATLAB the problems with arithmetic are uncommon, since double-precision floating-point representation is used. However, in case of poor variable scaling, the algorithm may fail. It is rather dangerous when the orders of magnitude of the state variables differ substantially. In such case the rounding errors of the large values may be greater than the small values. Moreover, the inversion of $\boldsymbol{S}$ within (4.54) becomes problematic.

The proper scaling is vital when fixed-point arithmetic is used which is typical for embedded processors.[108] The "dynamic range" of the state is defined by the number of bits used for the variable.

The initialization of the KF may become problematic as well. The initial state $\boldsymbol{x}[0]$ should not be too far from the expected estimate. Similarly, the covariance matrix $\boldsymbol{P}[0]$ should allow for quick convergence. Therefore, the variances of the states should be set large in order to reflect the initial uncertainty of the state value.

4.2.2. Exended Kalman filter

The EKF is the modification of the Kalman filter that allows operation with nonlinear measurement model and nonlinear state model. It is worth remarking that the notation used below is not the only one possible. For the nonlinear system the process model is described by means of a function $\mathbf{f}$, whilst the measurement model is described by function $\mathbf{h}$:

$$\boldsymbol{x}[k+1] = \mathbf{f}\left(\boldsymbol{x}[k]\right) + \mathcal{N}\{0, \boldsymbol{Q}\} \tag{4.58}$$

$$\boldsymbol{y}[k] = \mathbf{h}\left(\boldsymbol{x}[k]\right) + \mathcal{N}\{0, \boldsymbol{R}\}. \tag{4.59}$$

It is apparent that the prediction of the state from epoch $k-1$ to k of the EKF is performed via the function $\mathbf{f}$ as

$$\boldsymbol{x}^{-}[k] = \mathbf{f}\left(\boldsymbol{x}^{+}[k-1]\right). \tag{4.60}$$

[107] Even for the states that should be constant it is recommended introduce an artificial variance [151].

[108] Embedded processors often contain only the fixed-point arithmetic unit, the floating-point operations are therefore emulated and remarkably slower than the fixed-point operations.

In order to propagate the covariance matrix $\boldsymbol{P}$, the function $\mathbf{f}$ is linearized, i.e. its Jacobian matrix is used.

$$\boldsymbol{F}[k-1] = \left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}^+[k-1]} = \begin{bmatrix} \left.\frac{\partial \mathrm{f}_1}{\partial x_1}\right|_{x_1^+[k-1]} & \left.\frac{\partial \mathrm{f}_1}{\partial x_2}\right|_{x_2^+[k-1]} & \cdots \\ \left.\frac{\partial \mathrm{f}_2}{\partial x_1}\right|_{x_1^+[k-1]} & \left.\frac{\partial \mathrm{f}_2}{\partial x_2}\right|_{x_2^+[k-1]} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{4.61}$$

$$\boldsymbol{P}^-[k] = \boldsymbol{F}[k-1]\boldsymbol{P}^+[k-1]\boldsymbol{F}^{\mathrm{T}}[k-1] + \boldsymbol{Q} \tag{4.62}$$

Similarly, the innovations are computed using the measurement model function, and the innovation covariance by means of linearization of the function at the point of *a priori* state (the state prediction):

$$\boldsymbol{z}[k] = \boldsymbol{y}[k] - \mathbf{h}\left(\boldsymbol{x}^-[k]\right) \tag{4.63}$$

$$\boldsymbol{H}[k] = \left.\frac{\partial \mathbf{h}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}^-[k]} \tag{4.64}$$

$$\boldsymbol{S}[k] = \boldsymbol{H}[k]\boldsymbol{P}^-[k]\boldsymbol{H}^{\mathrm{T}}[k] + \boldsymbol{R}. \tag{4.65}$$

Next, the Kalman gain is computed in the same way as for the linear KF. However, obtaining *a posteriori* estimate of the state vector and its covariance matrix is slightly different:

$$\boldsymbol{K}[k] = \boldsymbol{P}^-[k]\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{S}[k])^{-1} \tag{4.66}$$

$$\boldsymbol{x}^+[k] = \boldsymbol{x}^-[k] + \boldsymbol{K}[k]\boldsymbol{z}[k] \tag{4.67}$$

$$\boldsymbol{P}^+[k] = \left(\mathbb{I} - \boldsymbol{K}[k]\boldsymbol{H}[k]\right)\boldsymbol{P}^-[k]. \tag{4.68}$$

It is possible to observe that (4.68) does not necessarily produce a symmetric matrix. Nonetheless, symmetry is the basic property of any covariance matrix; thus, it is recommended to perform a simple diagonalization of $\boldsymbol{P}^+[k]$ by means of

$$\boldsymbol{P} = \frac{\boldsymbol{P} + \boldsymbol{P}^{\mathrm{T}}}{2}. \tag{4.69}$$

The example of use of the EKF by the author is available in Appendix D.

Practical considerations

Since the propagation of the covariance matrix is only approximate, it is desirable to overestimate the process noise ($\boldsymbol{Q}$) in order to reduce the confidence in the prediction and provide margin for the linearization error. Due to the linearization, there is a risk of underestimating the predicted covariance matrix, which may lead to filter divergence.

The linearization error in the covariance matrix prediction may be reduced by means of breaking the prediction step into two or more smaller ones.[109] The linear approximation is then broken into a piecewise-linear approximation and the resulting error is

[109] This approach is an analogy to reducing the error of a simple Euler integration by means of reducing the step size.

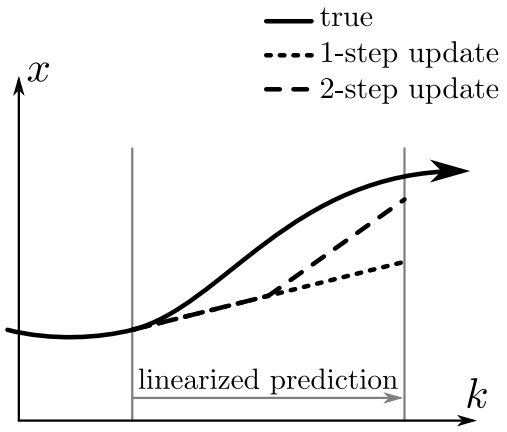


FIGURE 4.2.: Reducing linearization error by partitioning of the update step.

likely to be smaller, see Figure 4.2. If the nonlinearity-related problems arise from the measurement model, the iterated EKF may provide improvement [151].

Of course, the considerations mentioned in regard to the linear KF are mostly applicable to the EKF as well.

4.2.3. Unscented Kalman filter

In order to show that EKF is not the only possibility to deal with nonlinear systems, the unscented Kalman filter (UKF) is described here, although not used directly in the author's work presented within the appendices. Unlike EKF, it does not use linearization to propagate the state and its covariance matrix though the nonlinear models, but rather uses the *unscented transformation* (UT). The description of the UKF in this thesis follows [151], however, other variants of the unscented transformation do exist, e.g. [152, 153].

When a random vector is transformed nonlinearly, it is not secured that the transformation of the mean is the mean value of the transformed vector. The distribution of the transformed variable is not necessarily Gaussian. The UT uses several samples of the original random variable distribution – those are denoted *sigma-points*. The points are then propagated via the nonlinear function, and transformed mean and variance are computed based on them.

In order to show the principle of the UT, we will propagate the n-variate random variable $\boldsymbol{x}$ to m-variate random variable $\boldsymbol{y}$ via vector function $\boldsymbol{y} = \mathbf{f}(\boldsymbol{x})$. First, $2n$ sigma points are chosen so that

$$\begin{aligned} \boldsymbol{x}_i^\sigma &= \boldsymbol{x} + \sqrt{n}\left(\sqrt{\boldsymbol{\Phi}_{\boldsymbol{x}}}\right)_i^{\mathrm{T}} \qquad \text{for} \quad 1 \le i \le n \\ \boldsymbol{x}_i^\sigma &= \boldsymbol{x} - \sqrt{n}\left(\sqrt{\boldsymbol{\Phi}_{\boldsymbol{x}}}\right)_i^{\mathrm{T}} \qquad \text{for} \quad n+1 \le i \le 2n, \end{aligned} \tag{4.70}$$

where $\left(\sqrt{\boldsymbol{\Phi}_{\boldsymbol{x}}}\right)_i$ is the i-th row of the square root of $\boldsymbol{\Phi}_{\boldsymbol{x}}$ (the covariance matrix of x). Then, the sigma points are propagated through the nonlinear function $\boldsymbol{y}_i^\sigma = \mathbf{f}(\boldsymbol{x}_i^\sigma)$ and

the transformed mean $\bar{\boldsymbol{y}}$ and covariance matrix $\boldsymbol{\Phi}_{\boldsymbol{y}}$ are computed using

$$\bar{\boldsymbol{y}} = \frac{1}{2n}\sum_{i=1}^{2n} \boldsymbol{y}_i^{\sigma} \tag{4.71}$$

$$\boldsymbol{\Phi}_{\boldsymbol{y}} = \frac{1}{2n}\sum_{i=1}^{2n} (\boldsymbol{y}_i^{\sigma} - \bar{\boldsymbol{y}})(\boldsymbol{y}_i^{\sigma} - \bar{\boldsymbol{y}})^{\mathrm{T}}. \tag{4.72}$$

As an example we will provide a transformation of a 2-dimensional Cartesian coordinates to the distance and angle representation, i.e.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \sqrt{x_1^2 + x_2^2} \\ \tan^{-1}\frac{x_2}{x_1} \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mathcal{N}(3,\,1) \\ \mathcal{N}(4,\,1) \end{bmatrix}. \tag{4.73}$$

In Figure 4.3 the transformation of the mean and variance is presented by means of the error ellipses. The "true" values of the mean and covariance were obtained by propagating 10^3 random samples, which are depicted as dots in the figures. The EKF-like linearization approach was used in Figure 4.3a; clearly, the propagated mean, i.e. $\mathbf{f}(\bar{\boldsymbol{x}})$, does not coincide with the true mean of $\boldsymbol{y}$. The UT was performed in Figure 4.3b; it is obvious that the original, symmetric constellation of the sigma points has been distorted by the nonlinear function. The true and propagated means almost coincide, the covariance propagation is not substantially wrong.

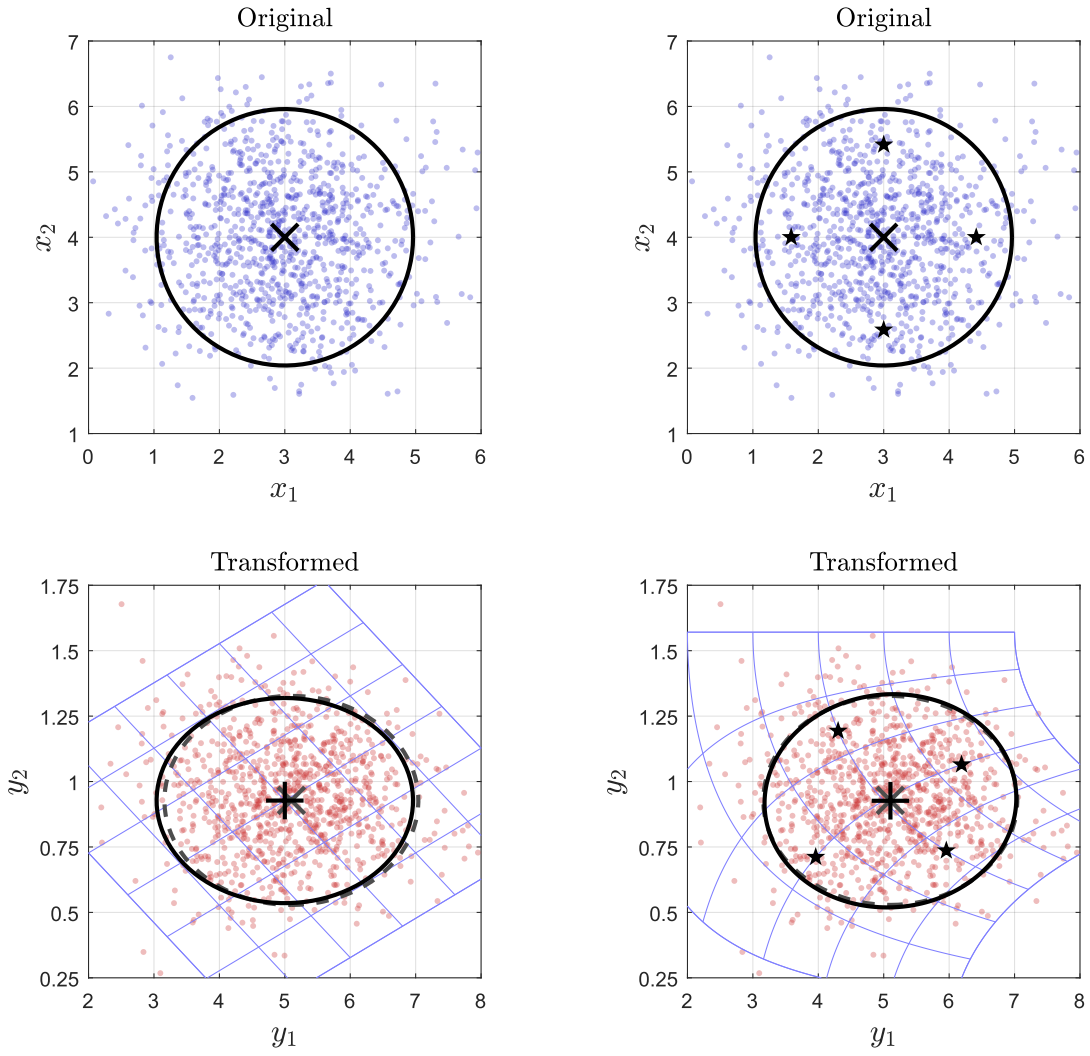
The difference of the linearized and the nonlinear transformation is visualized by means of the transformation of the background grids of the upper (original) plots in Fig. 4.3a to the lower (transformed) plot. It is apparent that the nonlinearity causes deformation of the grid (Fig. 4.3b), which is ignored in the linearized case (Fig. 4.3a). Naturally, the sigma-point constellation is deformed in the same way as the grid is.

The actual UKF algorithm uses the UT in several places. First, rather straightforward application is in the prediction step; the sigma points $\boldsymbol{x}_i^{\sigma+}[k-1]$ are generated according to (4.70) using the *a posteriori* state $\boldsymbol{x}^{+}[k-1]$ and its covariance matrix $\boldsymbol{P}^{+}[k-1]$. The sigma-points are then propagated through the nonlinear system process model function $\boldsymbol{x}_i^{\sigma-}[k] = \mathbf{f}(\boldsymbol{x}_i^{\sigma+}[k-1])$, and the *a priori* state $\boldsymbol{x}^{-}[k]$ is obtained using (4.71). In case of the *a priori* covariance matrix $\boldsymbol{P}^{-}[k]$, the computation is similar to (4.72), nonetheless, the process noise $\boldsymbol{Q}$ has to be taken into account. Thus,

$$\boldsymbol{P}^{-}[k] = \frac{1}{2n}\sum_{i=1}^{2n} (\boldsymbol{x}_i^{\sigma-}[k] - \boldsymbol{x}^{-}[k])(\boldsymbol{x}_i^{\sigma-}[k] - \boldsymbol{x}^{-}[k])^{\mathrm{T}} + \boldsymbol{Q}. \tag{4.74}$$

Second, more complicated application is in the measurement update step; the sigma-points $\boldsymbol{x}_i^{\sigma-}[k]$ may be re-generated according to (4.70). Otherwise, the ones from the prediction step are used, in order to avoid computation of the matrix square root. The sigma-points are propagated through the nonlinear measurement model $\boldsymbol{y}_i^{\sigma}[k] = \mathbf{h}(\boldsymbol{x}_i^{\sigma-}[k])$ and their mean $\bar{\boldsymbol{y}}[k]$ is obtained using (4.71). Then, the covariance of the predicted measurements is computed using

$$\boldsymbol{\Phi}_{\boldsymbol{y}}[k] = \frac{1}{2n}\sum_{i=1}^{2n} (\boldsymbol{y}_i^{\sigma}[k] - \bar{\boldsymbol{y}}[k])(\boldsymbol{y}_i^{\sigma}[k] - \bar{\boldsymbol{y}}[k])^{\mathrm{T}} + \boldsymbol{R}, \tag{4.75}$$



(A) Linearized (EKF-like)

(B) Unscented transformation

FIGURE 4.3.: Mean and covariance propagation via nonlinear function. The gray dashed 95 % error ellipse with "×" center is based on mean and covariance computed from 10^5 random samples (only 10^3 are plotted); the solid black error ellipse with "+" center depicts the transformed mean and covariance. The stars depict the sigma-points of the UT.
The blue mesh is the grid from the upper plot, transformed in the linearized (A) and the nonlinear (B) manner.

where $\boldsymbol{R}$ is the measurement covariance matrix. The cross-covariance between the prediction and the corresponding measurement is obtained by means of

$$\boldsymbol{\Phi}_{\boldsymbol{xy}}[k] = \frac{1}{2n}\sum_{i=1}^{2n}(\boldsymbol{x}_i^{\sigma-}[k] - \boldsymbol{x}^{-}[k])(\boldsymbol{y}_i^{\sigma}[k] - \bar{\boldsymbol{y}}[k])^{\mathrm{T}}. \tag{4.76}$$

The Kalman gain evaluation and the actual measurement update of the state follows

$$\boldsymbol{K}[k] = \boldsymbol{\Phi}_{\boldsymbol{xy}}[k]\boldsymbol{\Phi}_{\boldsymbol{y}}^{-1}[k] \tag{4.77}$$

$$\boldsymbol{x}^{+}[k] = \boldsymbol{x}^{-}[k] + \boldsymbol{K}[k]\,(\boldsymbol{y}[k] - \bar{\boldsymbol{y}}[k]) \tag{4.78}$$

$$\boldsymbol{P}^{+}[k] = \boldsymbol{P}^{-}[k] - \boldsymbol{K}[k]\boldsymbol{\Phi}_{\boldsymbol{y}}[k]\boldsymbol{K}^{\mathrm{T}}[k]. \tag{4.79}$$

By obtaining the *a posteriori* state in the epoch k, the filter evaluation loop is closed. It is worth noting that when the system process model or the measurement model is linear, there is no need to use the UT for the prediction or the measurement update, respectively. In such cases a hybrid (partially linear, partially unscented) Kalman filter can be used.

According to [153], the UKF provides 3$^{\mathrm{rd}}$ order accuracy of the mean and covariance transformation, whilst the linearization in EKF implies 1$^{\mathrm{st}}$ order accuracy only. It is necessary to consider whether the performance improvement brought by UKF in comparison with EKF is worth the increased computational efforts – especially the sigma-point generation, which contains the matrix square root operation.

The number of nonlinear function evaluations can be a performance factor as well. Consider that m and n are the dimensions of the measurement ($\boldsymbol{y}$) and state ($\boldsymbol{x}$) vectors, respectively. In each unscented transformation, as was described here, $2nm$ evaluations of a vector-to-scalar nonlinear function are performed; a set of m functions is evaluated n-times. In contrast, each linearization within the EKF requires m evaluations of a vector-to-scalar nonlinear function and nm Jacobian-matrix element evaluations; this yields evaluation of up to $n(m+1)$ different nonlinear functions.[110] For certain applications it may be advantageous that the unscented transformation, unlike the linearization, does not require knowledge of the partial derivatives of the transforming function at all.

4.2.4. Constraining the Kalman filter

The constraints carry an additional information about the system of interest that is not available to the filter otherwise – it is not included in the measurement nor in the system model.

There are various ways of constraining the numerous types of the Kalman filters; an exhaustive review of the available methods is provided in [150]. In this work we are particularly interested in soft equality constraints. The ease of implementation and versatility of the constraints is considered to be an important attribute. Therefore, a *pseudomeasurement* approach is presented.

[110] The number of evaluated functions in EKF is usually lower, since a number of partial derivatives is typically zero. In case of both filters, not all of the vector-to-scalar functions have to be nonlinear.

I. Pseudomeasurement approach

The pseudomeasurement approach is applicable to all the presented filters, i.e. linear KF, EKF and UKF. We may model the constraints as a generally nonlinear vector function

$$\boldsymbol{y'} = \mathbf{g}(\boldsymbol{x}) + \mathcal{N}\left(0,\, \boldsymbol{R'}\right) \tag{4.80}$$

where $\boldsymbol{y'}$ is the pseudomeasurement. Note that the constraints are commonly defined with the vector of zeros on the right hand side of the equation, however, for the pseudomeasurement approach this definition is more convenient.

The prediction step is performed in the same way as it would be performed in the unconstrained case. In the measurement update step, the *a priori* state is compared with both the measurements and the pseudomeasurements. For the sake of clarity we will drop the epoch identifier (k-index). The concatenated measurement vector and its covariance matrix are defined as

$$\boldsymbol{y}^{\circ} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y'} \end{bmatrix}, \qquad \boldsymbol{R}^{\circ} = \begin{bmatrix} \boldsymbol{R} & \mathrm{O} \\ \mathrm{O} & \boldsymbol{R'} \end{bmatrix}. \tag{4.81}$$

Intuitively, the measurement model function is concatenated similarly as the measurement vector. For the use in the EKF the concatenated Jacobian matrix of $\mathbf{h}^{\circ}(\boldsymbol{x}^{-})$ is obtained as a vertical concatenation of the Jacobian matrix of the measurement model $\boldsymbol{H}$ and the Jacobian matrix of the constraint model $\boldsymbol{G}$:

$$\mathbf{h}^{\circ}(\boldsymbol{x}^{-}) = \begin{bmatrix} \mathbf{h}(\boldsymbol{x}^{-}) \\ \mathbf{g}(\boldsymbol{x}^{-}) \end{bmatrix}, \qquad \boldsymbol{H}^{\circ} = \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{G} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{h}(\boldsymbol{x}^{-})}{\partial \boldsymbol{x}^{-}} \\ \frac{\partial \mathbf{g}(\boldsymbol{x}^{-})}{\partial \boldsymbol{x}^{-}} \end{bmatrix}. \tag{4.82}$$

The measurement update is then performed with the concatenated vectors, functions or matrices. For the linear KF the equations (4.52) to (4.56) are used, EKF update follows (4.63) to (4.68) and the UKF update is directed by (4.75) to (4.79).

II. Projection on the constraint subspace

In certain cases it is convenient to define constraint values, i.e. the pseudomeasurement vector $\boldsymbol{y'}$ as a function of the state.

As an example we may take the car from Figure 3.1 (page 89). The x^{b}-axis of the cars body frame points forward. Unless the car is skidding, the velocity should only be present in the direction of x^{b}. The Kalman filter state, i.e. the position, velocity and attitude, are expressed w.r.t. the n-frame; the dynamic model runs in the n-frame as well and does not reflect the b-frame related constraint. It is straightforward to see that the lateral and vertical velocity components v_y^{b} and v_z^{b} should be both zero. The constrained n-frame velocity vector $\boldsymbol{v'}^{\mathrm{n}}$ is then the velocity vector projected onto the x^{b}-axis. Provided that the quaternion $\mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{n}}$ describes the transformation from b-frame to n-frame, such projection can be written as

$$\boldsymbol{v'}^{\mathrm{n}} = \mathbb{1}_{x^{\mathrm{b}}}^{\mathrm{n}} (\mathbb{1}_{x^{\mathrm{b}}}^{\mathrm{n}})^{\mathrm{T}} \boldsymbol{v}^{\mathbf{n}} = \left(\mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{n}} \otimes \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \otimes (\mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{n}})^{*} \right) \left(\mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{n}} \otimes \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \otimes (\mathring{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{n}})^{*} \right)^{\mathrm{T}} \boldsymbol{v}^{\mathbf{n}}, \tag{4.83}$$

where $\boldsymbol{v}^{\mathbf{n}}$ denotes the unconstrained velocity vector, and $\mathbb{1}^{\mathrm{n}}_{x^{\mathrm{b}}}$ is the x^{b}-axis direction vector expressed in the n-frame. It is worth reminding that the vectors are preceded by a zero-element for the purpose of quaternion multiplication.

The general expression for projection of vector $\boldsymbol{x}$ into the direction of vector $\boldsymbol{c}$ is

$$\boldsymbol{x}' = \frac{\boldsymbol{c}\boldsymbol{c}^{\mathrm{T}}}{\|\boldsymbol{c}\|^2}\boldsymbol{x}, \qquad \boldsymbol{x}' = \mathbb{1}_{\boldsymbol{c}}\mathbb{1}_{\boldsymbol{c}}^{\mathrm{T}}\boldsymbol{x}, \tag{4.84}$$

where $\mathbb{1}_{\boldsymbol{c}}$ is the unit-length vector in the direction of $\boldsymbol{c}$. The projection of $\boldsymbol{x}$ onto a plane normal to vector $\boldsymbol{c}$ (sometimes denoted as vector rejection) can be performed as

$$\boldsymbol{x}' = \boldsymbol{x} - \frac{\boldsymbol{c}\boldsymbol{c}^{\mathrm{T}}}{\|\boldsymbol{c}\|^2}\boldsymbol{x} = \left(\mathbb{I} - \frac{\boldsymbol{c}\boldsymbol{c}^{\mathrm{T}}}{\|\boldsymbol{c}\|^2}\right)\boldsymbol{x}, \qquad \boldsymbol{x}' = \boldsymbol{x} - \mathbb{1}_{\boldsymbol{c}}\mathbb{1}_{\boldsymbol{c}}^{\mathrm{T}}\boldsymbol{x} = \left(\mathbb{I} - \mathbb{1}_{\boldsymbol{c}}\mathbb{1}_{\boldsymbol{c}}^{\mathrm{T}}\right)\boldsymbol{x}. \tag{4.85}$$

Note that the normal vector of a plane can be obtained as the cross-product any two linearly independent vectors lying within the plane to be projected onto.

It is acknowledged that projection onto a nonlinear surface was not considered within the thesis. The pseudomeasurement approach with linearly-projected constraints has been used by the author in the complementary EKF which is described within Appendix D.

4.2.5. Kalman smoother

In the post-processing applications it is possible to run the Kalman filters backwards in time. Of course, the models have to be modified to reflect the direction of processing. In each epoch, the forward and backward *a posteriori* states can be fused to a single state estimate by means of a multidimensional weighted average [134]. Omitting the epoch identifier we may write

$$\boldsymbol{P}_{\leftrightarrow} = \left(\boldsymbol{P}_{\rightarrow}^{-1} + \boldsymbol{P}_{\leftarrow}^{-1}\right)^{-1} \tag{4.86}$$

$$\boldsymbol{x}_{\leftrightarrow} = \boldsymbol{P}_{\leftrightarrow}\left(\boldsymbol{P}_{\rightarrow}^{-1}\boldsymbol{x}_{\rightarrow} + \boldsymbol{P}_{\leftarrow}^{-1}\boldsymbol{x}_{\leftarrow}\right), \tag{4.87}$$

where the $\rightarrow$, $\leftarrow$ and $\leftrightarrow$ subscripts denote the forward, backward and smoothed solution, respectively.

The computational burden of combining the forward and backward solution may be reduced by setting certain parts of the covariance matrices to zero, which may render the combination suboptimal. It is convenient to make the $\boldsymbol{P}$-matrices block-diagonal in order to trade the inversions of large matrices for more inversions of smaller matrices, or even scalars. Taken even further, if all the state covariances are neglected, all $\boldsymbol{P}$-matrices become diagonal. In such case the weighted average may be performed element-wise, all the inversions can be interpreted as scalar divisions.

A specific situation is in the error-state KF in Appendix D, where the blockwise diagonalization of $\boldsymbol{P}$-matrices has been used for additional reason. The error-state KF estimates the error of the position, velocity and attitude estimate. Whilst the position and velocity estimated values and the respective error values may be combined directly

using (4.86) and (4.87), the attitude estimate and attitude error estimate parimetrizations have to be altered. Therefore, the attitude-related part of the KF state and the covariance matrix is processed separately; consequently, covariance between attitude and other state-vector elements is neglected.

4.3. Alternative methods of data fusion

The use of Kalman filters is not the only possible way of sensor fusion. For instance, the particle filters or factor graphs may be considered as an alternative.

The particle filters (PF), sometimes called sequential Monte-Carlo (SMC) methods, share the prediction – measurement update architecture with the Kalman filters. However, they propagate a number of randomly distributed state estimates through the generally nonlinear models of the measurements and the system. The background idea is that a sufficient number of samples provides an approximation of the true probability density functions (PDFs) of the state, its prediction, innovation or the *a posteriori* estimate. Consequently, unlike KF, the PF is not bound by the assumption of Gaussian distribution nor linearity of the models. The disadvantage of the particle filter is high computational load and memory requirements. The detailed description of the particle filters is available e.g. in [154]. Examples of data-fusion-related PFs are available e.g. in [155]; in [136] the unscented particle filter (a hybrid between UKF and PF) is introduced. The particle filter can be used in the standalone GNSS positioning as well, see [156, 157].

The factor graphs can be considered as a more general approach, since they do not build on the prediction – measurement update architecture, nor they are constraining the underlying models, nor the state or noise PDFs. The factor graphs provide a graphical method of factorization of a complicated global function into a product of several local functions; an overview of the principles can be found e.g. in [158]. Although most commonly used in the information theory field, especially in digital communication and coding, the factor graphs can be used in the navigation field as well. For instance, [159] proposes a factor-graph-based approach for fusing data from IMU, GNSS and visual sensors. It is worth remarking that the Kalman filter may be visualized in a form of factor graph as well, see e.g. [160].

5. Conclusion

The main text of the thesis provides a theoretical context of the author's work in the field of radionavigation and integrated navigation; the appendices of the thesis present utilization of the particular principles in specific applications within author's work. The interconnection of the more general topics through the applications was already shown in the mindmap (Figure 1.2), which appeared in the introductory chapter. The following paragraphs will summarize the novel approaches found in this work.

Since the thesis is not deeply focused on any particular area but rather provides a link of several topics via applications of more general methods, it is worth to highlight a few important points this thesis has brought up:

Evaluation of bias and variance of the asymmetric double-sided TWR

The two-way ranging approach is very popular ranging method in the UWB systems; the variants of the method are described in Section 2.1.2–II. In particular, the asymmetric double-sided TWR was reported to reduce the sensitivity to clock-drift without the need of symmetric reply times as is required by the symmetric DS-TWR. However, the nonlinearity of the ADS-TWR estimator was overlooked and its properties have not been investigated, even though it has been patented [11] and is the reference TWR implementation by the UWB-chip manufacturers Application note [161].

The thorough analysis provided in Appendix A confirmed that the estimator is safe to use and that the increased immunity to drifting clocks was not traded for accuracy, nor significant bias of the estimator. The variance of the ADS-TWR estimator is proven to be similar to the variance of SDS-TWR, indeed.

Chained synchronization for UWB-network

The synchronization of UWB anchors (the fixed infrastructure nodes of the UWB localization network) is necessary to enable the TDoA positioning. The wired synchronization is rather impractical and expensive, thus, a wireless solution is favored. A sub-nanosecond timing accuracy is required in order to be able to achieve decimeter-level positioning. The Kalman-filter based algorithm described in Appendix C and published in [5, 6] provides sufficiently accurate time-transfer not only between a pair of master and slave nodes with line-of-sight link available. Relaying of the timing information over several nodes is possible with only a minor modification of the algorithm. Consequently, it is possible to synchronize two nodes without the direct line-of-sight via multiple line-of-sight segments.

Under most circumstances, the synchronization algorithm achieves sub-nanosecond accuracy (typically 0.3 to 0.4 ns RMS). Details of the algorithm principle and imple-

mentation are available in Appendix C.1; a number of experimental results is provided as well.

The exploitation of the results is shown in the Appendix B.2; the precise synchronization of the UWB network made possible to collect the TDoA measurements.

It is worth remarking that the author's chained synchronization algorithm is implemented in an UWB-based localization system developed jointly by Czech Technical University in Prague and RCD Radiokomunikace.

Soft-constrained error-state Kalman filter

The Kalman filters, which are suitable for performing sensor fusion, have been described in Section 4.2. It provides grounds for the development of the error-state Kalman filter (ESKF, see Appendix D) which is able to combine the data from a loosely coupled system consisting of a GNSS receiver, UWB transceiver and MEMS IMU. This sensor-fusion filter is an important part of the work [138], where accurate navigation in the GNSS-challenged environment is targeted.

Since the motion of the localized platform is deterministic to a considerable extent, dynamic constraints were considered in the ESKF. The constraints were applied in a "soft" manner, because there is an uncertainty of the constraint validity – the platform was man-powered on a rough surface. Additionally, the classification of the motion model, which was performed by a neural network, did not provide perfect results.

A pseudomeasurement implementation (general description in Section 4.2.4) of the soft constraints was used; such an approach allows simple adjustment of the constraints to match the chosen motion model. Attitude, velocity and partial position constraints were available for application, see Appendix D.3.

The filter architecture is modular, i.e. measurements and pseudomeasurements can be incorporated in a straightforward manner. The presented solution is able to overcome short radionavigation outages. The improvement in performance brought by the soft constraints is indisputable; naturally, the improvement (quantified by means of reduction of position RMS error) is greater when the motion is more constrained. For instance, the improvement was 35 % on the straight portion trajectory and 16 % in a turn.

Signal-power-based AoA in Rician channel

For the signal-power based measurement of range (Sec. 2.1.1) or angle of arrival (AoA, Sec. 2.1.4–I) two different channel models were assumed. First, a fairly common log-normal fading channel model was assumed. It is rather convenient for computations, since in dB-scale, the power distribution is normal (Gaussian); however, it neglects the presence of the thermal noise floor. Thus, the model with Rician channel, i.e. non-central χ^2 signal power distribution was investigated for both cases. It is worth remarking that the numerical integration of the distribution required in the CRLB evaluation was nontrivial and required adaptive setting of integration bounds in order to provide reliable results.

Iterative solutions of navigation problems

A number of iterative least-squares algorithms were presented in Section 4.1. Their application on two different TDoA problems is described in Appendix B – the DVB-T signal of opportunity positioning and UWB positioning. It needs to be remarked that we have not focused on the estimated position vector solely, the position covariance matrix, its importance and interpretation was highlighted as well within the thesis.

Appendices

A. Bias and Variance of the Asymmetric Double-sided TWR

In this appendix, the performance of three range estimators that were already mentioned in section 2.1.2 is investigated. The estimators (their equations are repeated below for the sake of clarity)

$$r_{\mathrm{AB},2} = \frac{c_0}{2}\,\frac{\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]}}{\tau_{\mathrm{A1}}^{[\mathrm{A}]} + \tau_{\mathrm{A2}}^{[\mathrm{A}]}} \tag{2.48}$$

$$r_{\mathrm{AB},1} = \frac{c_0}{2}\,\frac{\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]}}{\tau_{\mathrm{B2}}^{[\mathrm{B}]} + \tau_{\mathrm{B1}}^{[\mathrm{B}]}} \tag{2.49}$$

$$r_{\mathrm{AB},3} = c_0\,\frac{\tau_{\mathrm{A1}}^{[\mathrm{A}]}\tau_{\mathrm{B2}}^{[\mathrm{B}]} - \tau_{\mathrm{A2}}^{[\mathrm{A}]}\tau_{\mathrm{B1}}^{[\mathrm{B}]}}{\tau_{\mathrm{A1}}^{[\mathrm{A}]} + \tau_{\mathrm{A2}}^{[\mathrm{A}]} + \tau_{\mathrm{B2}}^{[\mathrm{B}]} + \tau_{\mathrm{B1}}^{[\mathrm{B}]}} \tag{2.50}$$

are claimed to reduce the negative impact of clock frequency bias in the range estimate, while relaxing the requirement on equality of the reply delays in the two negotiating nodes [3, 11]. The advantage of the estimators is indisputable, however the performance of the estimators, especially their variance has not been assessed yet. It is worth noting that the equation (2.50) is considered as a reference TWR implementation in an application note [161] authored by a UWB-ranging module manufacturer.

Since in all the equations of range estimate (2.48) to (2.50) are formed by a fraction, it is not possible to find the exact value of variance of such estimator. Moreover, due to the non-linearity of the equations, the range estimator may be biased. In the following sections, an approach for obtaining approximation of bias and variance of the estimators will be presented. It will be shown that under certain assumptions, the respective formulas can be simplified to reasonably long expressions. Graphical representation of the results will be provided, moreover, it will be compared to the results obtained by Monte–Carlo method.

A.1. Approximation of the Estimator Performance Parameters

In order to obtain an approximation of bias and variance, the multivariate Taylor expansion of the corresponding equations will be used. A similar approach has been already used for estimating the bias and variance of a simple ratio estimator in [162], nonetheless, the application in our case is a bit more complicated due to two reasons:

Number of parameters – It is known that the number $N_{\mathrm{m,k}}$ of m-th order terms of the Taylor series of a function with k parameters is given by

$$N_{\mathrm{m,k}} = \left(\!\!\binom{k}{m}\!\!\right) = \binom{m+k-1}{m}, \tag{A.1}$$

where $\left(\!\!\binom{a}{b}\!\!\right)$ denotes a multiset number[111] and $\binom{a}{b}$ the binomial coefficient.[112] With the higher order and more parameters the complexity of the multivariate Taylor expansion formula grows and the chance of achieving reasonable analytic result is reduced, naturally.

Dependency of numerator and denominator – In the case of a simple ratio estimator, stochastic independence of the numerator and denominator is usually assumed. In our case, however, the numerator and denominator are inherently dependent, by virtue of at least two parameters appearing both above and below the fraction bar.

Prior to presentation of the simplified formulas for the bias and variance of the particular estimators a generic approach to the approximation is presented. As the first step, it is necessary to find the Taylor expansion of the particular range estimator around the means of the involved delay measurements, see (A.2). For the sake of simplicity, we will drop the superfixes from those equations and denote the vector of means by $\boldsymbol{\mu}_\tau = [\mu_{\tau\mathrm{A1}}, \mu_{\tau\mathrm{B2}}, \mu_{\tau\mathrm{A2}}, \mu_{\tau\mathrm{B1}}]^{\mathrm{T}}$. The second-order Taylor expansion of a generic ADS-TWR estimator yields

$$\begin{aligned}
r \approx\ & r|_{\boldsymbol{\mu}_\tau} + (\tau_{\mathrm{A1}} - \mu_{\tau\mathrm{A1}})\left.\frac{\partial r}{\partial \tau_{\mathrm{A1}}}\right|_{\boldsymbol{\mu}_\tau} + (\tau_{\mathrm{B2}} - \mu_{\tau\mathrm{B2}})\left.\frac{\partial r}{\partial \tau_{\mathrm{B2}}}\right|_{\boldsymbol{\mu}_\tau} \\
& + (\tau_{\mathrm{A2}} - \mu_{\tau\mathrm{A2}})\left.\frac{\partial r}{\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} + (\tau_{\mathrm{B1}} - \mu_{\tau\mathrm{B1}})\left.\frac{\partial r}{\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} \\
& + \frac{(\tau_{\mathrm{A1}} - \mu_{\tau\mathrm{A1}})^2}{2}\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A1}}^2}\right|_{\boldsymbol{\mu}_\tau} + \frac{(\tau_{\mathrm{B2}} - \mu_{\tau\mathrm{B2}})^2}{2}\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{B2}}^2}\right|_{\boldsymbol{\mu}_\tau} \\
& + \frac{(\tau_{\mathrm{A2}} - \mu_{\tau\mathrm{A2}})^2}{2}\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A2}}^2}\right|_{\boldsymbol{\mu}_\tau} + \frac{(\tau_{\mathrm{B1}} - \mu_{\tau\mathrm{B1}})^2}{2}\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{B1}}^2}\right|_{\boldsymbol{\mu}_\tau} \\
& + (\tau_{\mathrm{A1}} - \mu_{\tau\mathrm{A1}})(\tau_{\mathrm{B2}} - \mu_{\tau\mathrm{B2}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A1}}\partial \tau_{\mathrm{B2}}}\right|_{\boldsymbol{\mu}_\tau} + (\tau_{\mathrm{A1}} - \mu_{\tau\mathrm{A1}})(\tau_{\mathrm{A2}} - \mu_{\tau\mathrm{A2}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A1}}\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} \\
& + (\tau_{\mathrm{A1}} - \mu_{\tau\mathrm{A1}})(\tau_{\mathrm{B1}} - \mu_{\tau\mathrm{B1}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A1}}\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} + (\tau_{\mathrm{B2}} - \mu_{\tau\mathrm{B2}})(\tau_{\mathrm{A2}} - \mu_{\tau\mathrm{A2}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{B2}}\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} \\
& + (\tau_{\mathrm{B2}} - \mu_{\tau\mathrm{B2}})(\tau_{\mathrm{B1}} - \mu_{\tau\mathrm{B1}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{B2}}\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} + (\tau_{\mathrm{A2}} - \mu_{\tau\mathrm{A2}})(\tau_{\mathrm{B1}} - \mu_{\tau\mathrm{B1}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A2}}\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau},
\end{aligned} \tag{A.2}$$

[111] Multiset number in the notation mentioned above is a number of b-tuples chosen from a types of elements. There may be multiple elements of the same type in the b-tuple.

[112] Binomial coefficient in the notation mentioned above is a number of b-tuples chosen from a set of a elements. Each element can be present only once in the b-tuple, therefore $a \geq b$ must hold.

where $|_{\boldsymbol{\mu}_\tau}$ denotes that the value is obtained at the expansion point, i.e. when the parameters are equal to their respective means. In such notation, $r|_{\boldsymbol{\mu}_\tau}$ is the estimator output value when the parameter means are input to the function.

In order to find the mean value of the range estimate, at least the second-order expansion must be used. Provided that random values of the estimator function parameters are unbiased, all the first-order terms are eliminated, since $\mathrm{E}[\tau - \mu_\tau] = 0$. When the expectation operator is applied on the second-order terms of the Taylor expansion, the defining expressions of variances and covariances of the estimator parameters can be recognized. Therefore, the generic formula for the estimator variance is

$$\begin{aligned}
\mathrm{E}[r] \approx\ & r|_{\boldsymbol{\mu}_\tau} + \frac{\operatorname{var}\tau_{\mathrm{A1}}}{2}\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A1}}^2}\right|_{\boldsymbol{\mu}_\tau} + \frac{\operatorname{var}\tau_{\mathrm{B2}}}{2}\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{B2}}^2}\right|_{\boldsymbol{\mu}_\tau} \\
& + \frac{\operatorname{var}\tau_{\mathrm{A2}}}{2}\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A2}}^2}\right|_{\boldsymbol{\mu}_\tau} + \frac{\operatorname{var}\tau_{\mathrm{B1}}}{2}\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{B1}}^2}\right|_{\boldsymbol{\mu}_\tau} \\
& + \operatorname{cov}(\tau_{\mathrm{A1}}, \tau_{\mathrm{B2}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A1}}\partial \tau_{\mathrm{B2}}}\right|_{\boldsymbol{\mu}_\tau} + \operatorname{cov}(\tau_{\mathrm{A1}}, \tau_{\mathrm{A2}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A1}}\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} \\
& + \operatorname{cov}(\tau_{\mathrm{A1}}, \tau_{\mathrm{B1}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A1}}\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} + \operatorname{cov}(\tau_{\mathrm{B2}}, \tau_{\mathrm{A2}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{B2}}\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} \\
& + \operatorname{cov}(\tau_{\mathrm{B2}}, \tau_{\mathrm{B1}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{B2}}\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} + \operatorname{cov}(\tau_{\mathrm{A2}}, \tau_{\mathrm{B1}})\left.\frac{\partial^2 r}{\partial \tau_{\mathrm{A2}}\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau}.
\end{aligned} \tag{A.3}$$

It is straightforward to see that the bias of the estimator can be obtained by omitting the term $r|_{\boldsymbol{\mu}_\tau}$ from the equation above

$$\operatorname{Bias}[r] = \mathrm{E}[r] - r|_{\boldsymbol{\mu}_\tau}. \tag{A.4}$$

The second task is obtaining an approximation of the estimator variance. For the sake of reasonable complexity, only the first order expansion will be used; likewise, the bias of the estimator will be neglected. Since we assume $\mathrm{E}[r] \approxeq r|_{\boldsymbol{\mu}_\tau}$, this value is eliminated from the following equation. The generic formula for the ADS-TWR range estimator variance is

$$\begin{aligned}
\mathrm{E}[r - \mathrm{E}[r]]^2 \approx \mathrm{E}\Bigg[& \left((\tau_{\mathrm{A1}} - \mu_{\tau\mathrm{A1}})\left.\frac{\partial r}{\partial \tau_{\mathrm{A1}}}\right|_{\boldsymbol{\mu}_\tau}\right)^2 + \left((\tau_{\mathrm{B2}} - \mu_{\tau\mathrm{B2}})\left.\frac{\partial r}{\partial \tau_{\mathrm{B2}}}\right|_{\boldsymbol{\mu}_\tau}\right)^2 \\
& + \left((\tau_{\mathrm{A2}} - \mu_{\tau\mathrm{A2}})\left.\frac{\partial r}{\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau}\right)^2 + \left((\tau_{\mathrm{B1}} - \mu_{\tau\mathrm{B1}})\left.\frac{\partial r}{\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau}\right)^2 \\
& + 2(\tau_{\mathrm{A1}} - \mu_{\tau\mathrm{A1}})(\tau_{\mathrm{B2}} - \mu_{\tau\mathrm{B2}})\left.\frac{\partial r}{\partial \tau_{\mathrm{A1}}}\right|_{\boldsymbol{\mu}_\tau}\left.\frac{\partial r}{\partial \tau_{\mathrm{B2}}}\right|_{\boldsymbol{\mu}_\tau} \\
& + 2(\tau_{\mathrm{A1}} - \mu_{\tau\mathrm{A1}})(\tau_{\mathrm{A2}} - \mu_{\tau\mathrm{A2}})\left.\frac{\partial r}{\partial \tau_{\mathrm{A1}}}\right|_{\boldsymbol{\mu}_\tau}\left.\frac{\partial r}{\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} \\
& + 2(\tau_{\mathrm{A1}} - \mu_{\tau\mathrm{A1}})(\tau_{\mathrm{B1}} - \mu_{\tau\mathrm{B1}})\left.\frac{\partial r}{\partial \tau_{\mathrm{A1}}}\right|_{\boldsymbol{\mu}_\tau}\left.\frac{\partial r}{\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau}
\end{aligned}$$

$$
\begin{aligned}
&+2\,(\tau_{\mathrm{B2}}-\mu_{\tau\mathrm{B2}})\,(\tau_{\mathrm{A2}}-\mu_{\tau\mathrm{A2}})\,\frac{\partial r}{\partial \tau_{\mathrm{B2}}}\bigg|_{\boldsymbol{\mu}_\tau}\frac{\partial r}{\partial \tau_{\mathrm{A2}}}\bigg|_{\boldsymbol{\mu}_\tau}\\
&+2\,(\tau_{\mathrm{B2}}-\mu_{\tau\mathrm{B2}})\,(\tau_{\mathrm{B1}}-\mu_{\tau\mathrm{B1}})\,\frac{\partial r}{\partial \tau_{\mathrm{B2}}}\bigg|_{\boldsymbol{\mu}_\tau}\frac{\partial r}{\partial \tau_{\mathrm{B1}}}\bigg|_{\boldsymbol{\mu}_\tau}\\
&+2\,(\tau_{\mathrm{A2}}-\mu_{\tau\mathrm{A2}})\,(\tau_{\mathrm{B1}}-\mu_{\tau\mathrm{B1}})\,\frac{\partial r}{\partial \tau_{\mathrm{A2}}}\bigg|_{\boldsymbol{\mu}_\tau}\frac{\partial r}{\partial \tau_{\mathrm{B1}}}\bigg|_{\boldsymbol{\mu}_\tau}\Bigg]\\
=\;&\operatorname{var}\tau_{\mathrm{A1}}\left(\frac{\partial r}{\partial \tau_{\mathrm{A1}}}\bigg|_{\boldsymbol{\mu}_\tau}\right)^2 + \operatorname{var}\tau_{\mathrm{B2}}\left(\frac{\partial r}{\partial \tau_{\mathrm{B2}}}\bigg|_{\boldsymbol{\mu}_\tau}\right)^2\\
&+\operatorname{var}\tau_{\mathrm{A2}}\left(\frac{\partial r}{\partial \tau_{\mathrm{A2}}}\bigg|_{\boldsymbol{\mu}_\tau}\right)^2 + \operatorname{var}\tau_{\mathrm{B1}}\left(\frac{\partial r}{\partial \tau_{\mathrm{B1}}}\bigg|_{\boldsymbol{\mu}_\tau}\right)^2\\
&+2\operatorname{cov}(\tau_{\mathrm{A1}},\,\tau_{\mathrm{B2}})\,\frac{\partial r}{\partial \tau_{\mathrm{A1}}}\bigg|_{\boldsymbol{\mu}_\tau}\frac{\partial r}{\partial \tau_{\mathrm{B2}}}\bigg|_{\boldsymbol{\mu}_\tau} + 2\operatorname{cov}(\tau_{\mathrm{A1}},\,\tau_{\mathrm{A2}})\,\frac{\partial r}{\partial \tau_{\mathrm{A1}}}\bigg|_{\boldsymbol{\mu}_\tau}\frac{\partial r}{\partial \tau_{\mathrm{A2}}}\bigg|_{\boldsymbol{\mu}_\tau}\\
&+2\operatorname{cov}(\tau_{\mathrm{A1}},\,\tau_{\mathrm{B1}})\,\frac{\partial r}{\partial \tau_{\mathrm{A1}}}\bigg|_{\boldsymbol{\mu}_\tau}\frac{\partial r}{\partial \tau_{\mathrm{B1}}}\bigg|_{\boldsymbol{\mu}_\tau} + 2\operatorname{cov}(\tau_{\mathrm{B2}},\,\tau_{\mathrm{A2}})\,\frac{\partial r}{\partial \tau_{\mathrm{B2}}}\bigg|_{\boldsymbol{\mu}_\tau}\frac{\partial r}{\partial \tau_{\mathrm{A2}}}\bigg|_{\boldsymbol{\mu}_\tau}\\
&+2\operatorname{cov}(\tau_{\mathrm{B2}},\,\tau_{\mathrm{B1}})\,\frac{\partial r}{\partial \tau_{\mathrm{B2}}}\bigg|_{\boldsymbol{\mu}_\tau}\frac{\partial r}{\partial \tau_{\mathrm{B1}}}\bigg|_{\boldsymbol{\mu}_\tau} + 2\operatorname{cov}(\tau_{\mathrm{A2}},\,\tau_{\mathrm{B1}})\,\frac{\partial r}{\partial \tau_{\mathrm{A2}}}\bigg|_{\boldsymbol{\mu}_\tau}\frac{\partial r}{\partial \tau_{\mathrm{B1}}}\bigg|_{\boldsymbol{\mu}_\tau}.
\end{aligned}
\tag{A.5}
$$

The format of the result is rather convenient, since only second central moments, i.e. variances and covariances, are present in the expression.

A.2. Evaluation for the Particular Estimators

In order to estimate the variance and the mean (bias) of the particular estimator, it is necessary to find its Jacobian and Hessian matrices, i.e. the matrices of first and second partial derivatives for the particular ranging equation. The full Jacobian Hessian matrices, i.e. the elements of the matrices, of the estimators (2.48) to (2.50) are available at the end of this appendix in Section A.4.

In exchange for reasonably complicated equations for means and variances of the estimators, assumptions on the measurement covariance matrices have to be made. In the numbered subsections below, the behavior of the estimators under three different measurement covariance matrix assumptions are investigated. The first two (unnumbered) subsections should serve as a guide how to read the graphical results that are provided with the formulas.

Evaluation for a representative set of values

The formulas in the following sections describe mean and variance of the estimators, however, a plot of the results for representative intervals of parameter values may provide more straightforward information. Clearly, it is not possible to visualize a real function of four parameters, only a real function with two real parameters can be visualized by means of contour or surface plot. It is rather inconvenient to use two of the TWR delays as variables and fix the other two delays. Therefore we have chosen

the parameters to be the true range r and the normalized asymmetry of the reply delays according to

$$\mathcal{A}_{\mathrm{n}} = (\mu_{\tau\mathrm{A2}} - \mu_{\tau\mathrm{B1}})/(\mu_{\tau\mathrm{A2}} + \mu_{\tau\mathrm{B1}}). \tag{A.6}$$

Note that this dimensionless value is plotted on interval from -0.5 to $+0.5$, which corresponds to the interval of the ratio $\mu_{\tau\mathrm{A2}} : \mu_{\tau\mathrm{B1}}$ from $1:3$ to $3:1$. The interval of the range values spans from 10^0 to 10^6 in order to show the behavior of the mean and variance of the estimator, when the propagation delays become comparable to the response delays. In order to be able to solve for the values of the range estimator parameters, we had to provide a fixed value; it was found convenient to fix the duration of the negotiation to 10 milliseconds (measured on one node), which is a typical value for UWB ranging systems. For the sake of simplicity, we will use substitution

$$\Sigma_\mu = 2(\mu_{\tau\mathrm{A1}} + \mu_{\tau\mathrm{A2}}) = 2(\mu_{\tau\mathrm{B1}} + \mu_{\tau\mathrm{B2}}) = \mu_{\tau\mathrm{A1}} + \mu_{\tau\mathrm{A2}} + \mu_{\tau\mathrm{B1}} + \mu_{\tau\mathrm{B2}}. \tag{A.7}$$

Therefore we can write for the assumption on duration of the negotiation

$$\frac{\Sigma_\mu}{2} = 10\,\mathrm{ms}. \tag{A.8}$$

Also, it is far more convenient to plot estimator bias instead of the estimate mean value. Expression σ^2_{Rx} will be used to denote the variance of the reception timestamp. It will become obvious in the following sections that the bias of the estimator can be normalized by the variance σ^2_{Rx}. Since the variance of the timestamp is expressed in squared seconds and the bias in meters, it is convenient to include the speed of light in the normalization.

$$\mathrm{Bias}_{\mathrm{n}}\,[r] = \frac{\mathrm{Bias}\,[r]}{\sigma^2_{\mathrm{Rx}} c_0} \tag{A.9}$$

Consequently, the dimension of the plotted normalized bias value is s^{-1}. In order to obtain the actual bias value in meters, it is necessary to multiply the reception timestamp variance (in s^2) by speed of light (in ms^{-1}) and the plot contour line value for particular true range and normalized asymmetry value.

In the other sets of plots the normalized standard deviation of the estimate is plotted, instead of the variance. Similarly to the case of estimator bias, it will become apparent that the standard deviation of the estimator can be normalized by the standard deviation of the reception timestamp (further denoted by σ_{Rx}). Therefore the plotted value is

$$\mathrm{std}_{\mathrm{n}}\,[r] = \frac{\mathrm{std}\,[r]}{\sigma_{\mathrm{Rx}} c_0} = \sqrt{\frac{\mathrm{var}\,[r]}{\sigma^2_{\mathrm{Rx}} c_0^2}}, \tag{A.10}$$

and has dimension $\mathrm{m}^{-1}\mathrm{s}$. The same holds for the variance plot that is based on Monte–Carlo Evaluation.

Monte–Carlo evaluation of estimator variance

An experimental evaluation of the estimator performance was performed. The approach was rather similar to the parametric evaluation of the analytical expressions. The steps to obtain the Monte–Carlo results follow:

1. Choose a point from the range vs. normalized asymmetry plane.
2. Solve for vector of delay means $\boldsymbol{\mu}_\tau$, assume fixed duration.
3. Perform N_{MC} trials:
   a) Add AWGN according to the corresponding covariance matrix (scaled by standard deviation σ_{Rx}) to the vector $\boldsymbol{\mu}_\tau$.
   b) Apply one of the range estimators (2.48), (2.49) or (2.50) and store the result.
4. Evaluate bias and variance (standard deviation) of the estimate, normalize the values according to (A.9) and (A.10).
5. Choose the next point from the range vs. normalized asymmetry plane and continue with step 2. Stop if all point have been already evaluated.

The parameters of the evaluation are summarized in Table A.1. The bias and variance (standard deviation) of the respective estimator was the outcome of each Monte–Carlo algorithm run. However, the bias was not observable due to its very low amplitude and therefore these results are not presented in the following sections.

In order to be able to produce the contour plots of the estimator standard deviations, the results were smoothed by a two-dimensional moving average filter.[113] The 201×250 matrix of standard deviation data was convolved with a 3×3 kernel, whose all entries were $1/9$. It is worth noting that the data matrix was padded in a nearest neighbor manner prior to the convolution, in spite of eliminating edge effects that would occur if the matrix would be padded with zeros. The contours to be plotted were based on the central part of the convolution result with a size 201×250.

The contour plots in the following sections should be read as a map. The horizontal axis (that would represent longitude on a map) represents the true range; the vertical axis (that would represent latitude on a map) represents the normalized asymmetry. The contours (lines with constant elevation in a map) are lines with constant bias or standard deviation here. The bias contours in all figures share the same color gradient scheme; smoothly from blue color for the positive values through pure black on the zero contour and further to the red for the negative values. In contrast, the normalized standard deviation plots use the blue-green-yellow-red gradient scaled on the interval from zero to one in a linear way.

I. Three-message TWR with perfect time of transmission

Let us assume for this case that all transmission times are known precisely (i.e. their variance is zero) and all the reception times have the same variance σ^2_{Rx}. Therefore, all the delay measurements have the same variance σ^2_{Rx}, too.

A case when the negotiation between the two nodes follows the three-message TWR scheme is investigated. The diagram of the negotiation is provided in Figure 2.3 on page 16. For that particular case, τ_{A1} and τ_{A2} are inherently correlated since they share

[113] Without smoothing the contours would be inhomogeneous and uneven, so the readability of the plot would be reduced.

TABLE A.1.: Parameters of the Monte–Carlo simulation.

	Parameter	**Unit**	**Value / Interval**	**Points per interval**
$\mathcal{A}_\mathrm{n}$	Normalized asymmetry	[-]	$\langle -0.5;\, +0.5 \rangle$	201 *linear*
r	True range	[m]	$\langle 10^0;\, 10^6 \rangle$	250 *logarithmic*
$\frac{1}{2}\Sigma_\mu$	Duration	[s]	$10 \cdot 10^{-6}$	*fixed value*
σ_Rx	Std. of reception time	[s]	$0.250 \cdot 10^{-9}$	*fixed value*
N_MC	Number of trials	[-]	10^6	*per point*

the receive time information. On the contrary, τ_B1 and τ_B2 are mutually independent, because they share the non-random transmission time while the random reception times are kept separate. Obviously, the covariance matrix for such case is

$$\boldsymbol{\Phi}_{\boldsymbol{\tau}1} = \sigma_\mathrm{Rx}^2 \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{A.11}$$

The approximate mean values of the range estimates can be obtained by substitution of covariance matrix to (A.3). The corresponding Hessian matrix elements for the three range estimators are provided as equations (A.33), (A.35) and (A.37).

$$\mathrm{E}[r_1] \approx r\big|_{\boldsymbol{\mu}_\tau} + 0 \tag{A.12}$$

$$\begin{aligned} \mathrm{E}[r_2] &\approx r\big|_{\boldsymbol{\mu}_\tau} + c_0\sigma_\mathrm{Rx}^2 \left(\frac{8(\mu_{\tau\mathrm{A}1}\mu_{\tau\mathrm{B}2} - \mu_{\tau\mathrm{A}2}\mu_{\tau\mathrm{B}1})}{(\Sigma_\mu)^3} + \frac{2(\mu_{\tau\mathrm{A}2} - \mu_{\tau\mathrm{A}1})}{(\Sigma_\mu)^2} \right) \\ &= r\big|_{\boldsymbol{\mu}_\tau} + \sigma_\mathrm{Rx}^2 \left(\frac{8\, r\big|_{\boldsymbol{\mu}_\tau} + 2c_0(\mu_{\tau\mathrm{A}2} - \mu_{\tau\mathrm{A}1})}{(\Sigma_\mu)^2} \right) \end{aligned} \tag{A.13}$$

$$\begin{aligned} \mathrm{E}[r_3] &\approx r\big|_{\boldsymbol{\mu}_\tau} + c_0\sigma_\mathrm{Rx}^2 \left(\frac{2(\mu_{\tau\mathrm{A}1}\mu_{\tau\mathrm{B}2} - \mu_{\tau\mathrm{A}2}\mu_{\tau\mathrm{B}1})}{(\Sigma_\mu)^3} + \frac{\mu_{\tau\mathrm{A}2} - \mu_{\tau\mathrm{A}1}}{(\Sigma_\mu)^2} \right) \\ &= r\big|_{\boldsymbol{\mu}_\tau} + \sigma_\mathrm{Rx}^2 \left(\frac{2\, r\big|_{\boldsymbol{\mu}_\tau} + c_0(\mu_{\tau\mathrm{A}2} - \mu_{\tau\mathrm{A}1})}{(\Sigma_\mu)^2} \right) \end{aligned} \tag{A.14}$$

It is obvious that only the r_1 estimator is unbiased. Nonetheless, the bias of the other two estimators is likely to be negligible when typical values for UWB ranging systems are assumed.

Figures A.1 to A.3 present the normalized bias with respect to normalized asymmetry and true range. In the biased cases, the bias is almost range-independent until the propagation delay r/c_0 is comparable with the mean response delay times $\mu_{\tau\mathrm{B}1}$ and $\mu_{\tau\mathrm{A}2}$. For the extremely long ranges (that are not achievable by the UWB systems) the zero-bias contour of the estimator r_2 turns towards the negative values of the normalized asymmetry. The zero-bias contour of the estimator r_3 remains on the zero value of the normalized asymmetry, additionally, the bias magnitude is reduced. It is worth

noting that contour $10\,\mathrm{s}^{-1}$ in the figure corresponds to a few nanometers of ranging bias under typical conditions of UWB ranging. This renders the zero-bias assumption required for variance approximation valid.

The approximate variances of the range estimates can be obtained by substitution of covariance matrix to (A.5). The corresponding Jacobian matrix elements for the three range estimators are provided as equations (A.32), (A.34) and (A.36).

$$\mathrm{E}\left[r_1 - \mathrm{E}\left[r\right]\right]^2 \approx \sigma_{\mathrm{Rx}}^2 \left(\frac{c_0^2(\mu_{\tau\mathrm{B2}}{}^2 + \mu_{\tau\mathrm{A1}}{}^2 + \mu_{\tau\mathrm{B1}}{}^2 + \mu_{\tau\mathrm{A2}}{}^2 + 2\mu_{\tau\mathrm{B1}}\mu_{\tau\mathrm{B2}})}{(\Sigma_\mu)^2} \right) \tag{A.15}$$

$$\mathrm{E}\left[r_2 - \mathrm{E}\left[r\right]\right]^2 \approx \sigma_{\mathrm{Rx}}^2 \left(\frac{c_0^2(\mu_{\tau\mathrm{B2}}{}^2 + \mu_{\tau\mathrm{A1}}{}^2 + \mu_{\tau\mathrm{B1}}{}^2 + \mu_{\tau\mathrm{A2}}{}^2 + 2\mu_{\tau\mathrm{B1}}\mu_{\tau\mathrm{B2}})}{(\Sigma_\mu)^2} + \frac{4c_0 r(\mu_{\tau\mathrm{A2}} - \mu_{\tau\mathrm{A1}}) + 8r^2}{(\Sigma_\mu)^2} \right) \tag{A.16}$$

$$\mathrm{E}\left[r_3 - \mathrm{E}\left[r\right]\right]^2 \approx \sigma_{\mathrm{Rx}}^2 \left(\frac{c_0^2(\mu_{\tau\mathrm{B2}}{}^2 + \mu_{\tau\mathrm{A1}}{}^2 + \mu_{\tau\mathrm{B1}}{}^2 + \mu_{\tau\mathrm{A2}}{}^2 + 2\mu_{\tau\mathrm{B1}}\mu_{\tau\mathrm{B2}})}{(\Sigma_\mu)^2} + \frac{2c_0 r(\mu_{\tau\mathrm{A2}} - \mu_{\tau\mathrm{A1}}) + 2r^2}{(\Sigma_\mu)^2} \right) \tag{A.17}$$

The contour graphs of the normalized standard deviations of the estimators, i.e. square roots of the above expressions, are provided in Figures A.4 to A.6. In the (A) subfigures, the data is based on the three approximate equations, in contrast, the (B) subfigures present the results of the Monte–Carlo simulation.

As well as in the case of biases, the variances of all the estimators can be considered range-independent for the corresponding propagation delays that are significantly shorter than the reply delays. However, for the extremely long ranges the area with lower variance shrinks and "bends" towards the positive normalized asymmetry values for r_1, towards the negative normalized asymmetry values for r_2. For the estimator r_3 it remains in the center and enlarges with the growing range. This mutual symmetry of r_1 and r_2 is somehow expectable due to the nature of the estimator equations. Nonetheless, there is no explicit mention of true range in the variance approximation (A.15) for r_1; the range dependency is inherently included in the common term of all three expressions.

The contours of the Monte–Carlo results follow both the shape and the magnitude of the analytic approximation; there is an almost perfect match. The dependence of the standard deviation on the normalized asymmetry is quite weak. We may conclude that the standard deviation of all three estimators is approximately the same in the region of realistic ranges. As a rule of thumb the standard deviation of the range estimate is less than 2/3 of the reception time standard deviation scaled by the speed of light c_0, i.e. a nanosecond standard deviation of reception time corresponds to 20 centimeter range error.

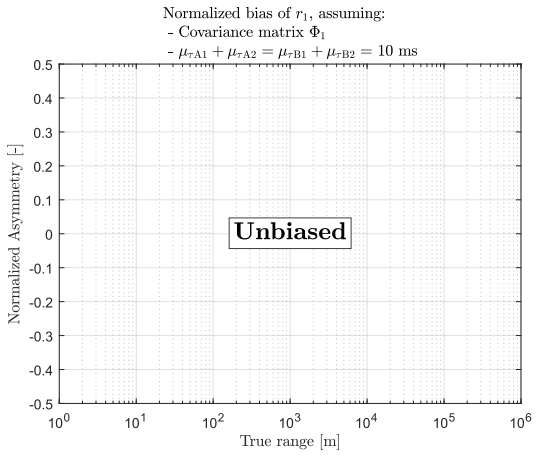


FIGURE A.1.: Normalized bias of r_1, assuming $\boldsymbol{\Phi_{\tau 1}}$.

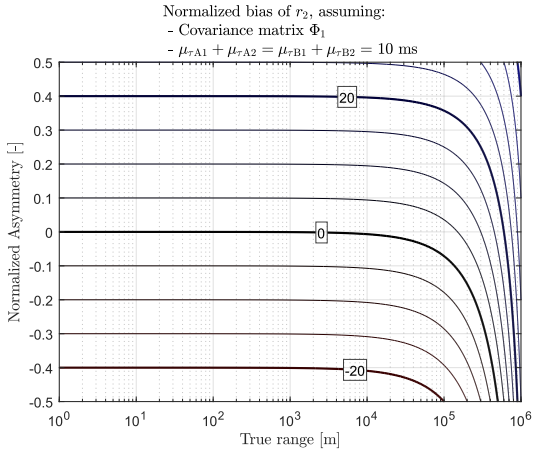


FIGURE A.2.: Normalized bias of r_2, assuming $\boldsymbol{\Phi_{\tau 1}}$.

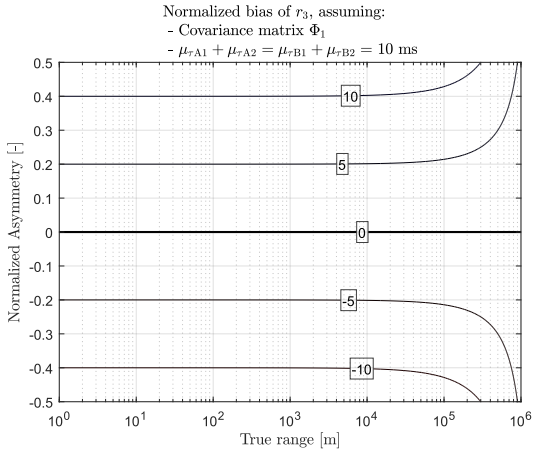
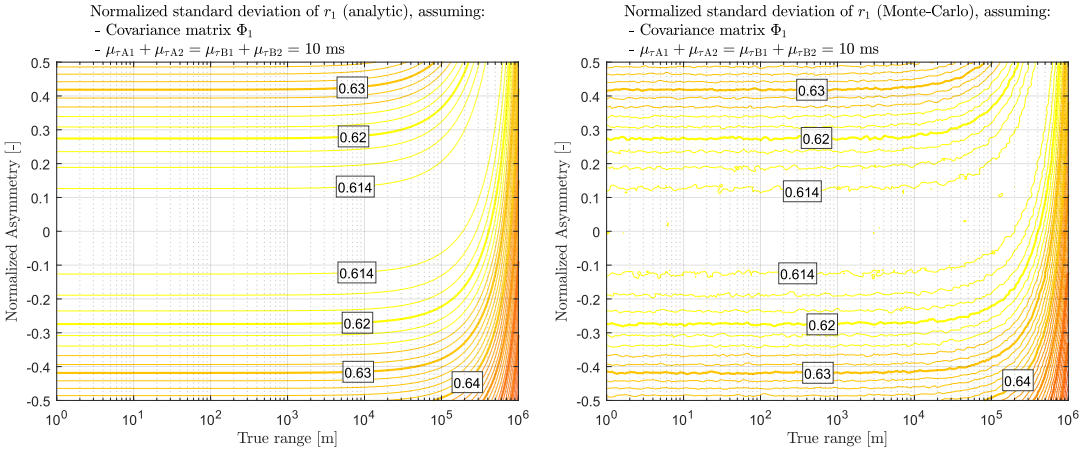
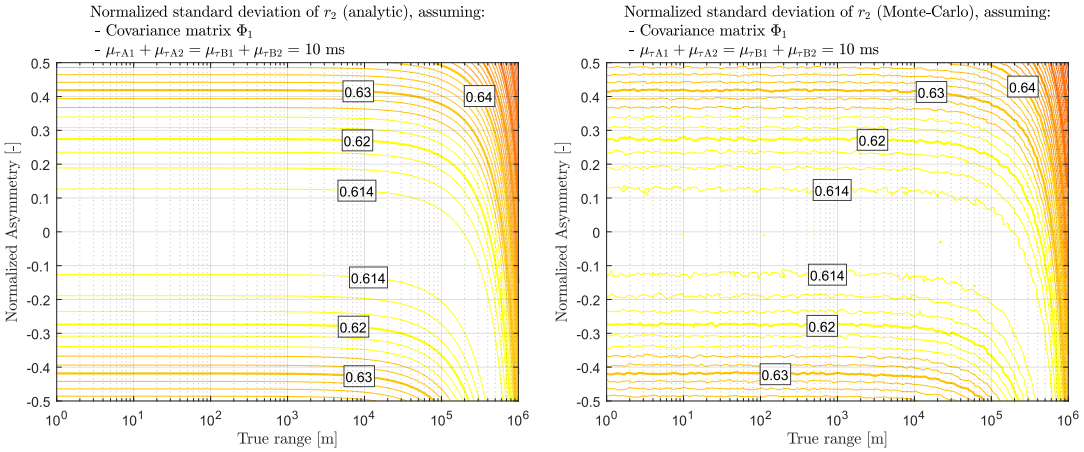


FIGURE A.3.: Normalized bias of r_3, assuming $\boldsymbol{\Phi_{\tau 1}}$.



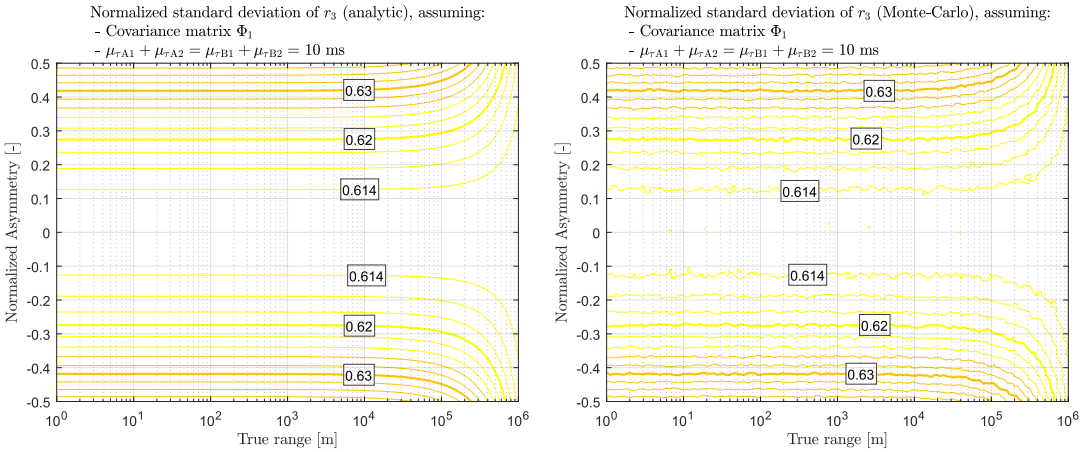
(A) Approximation (B) Monte–Carlo

FIGURE A.4.: Normalized standard deviation of r_1, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}1}$.



(A) Approximation (B) Monte–Carlo

FIGURE A.5.: Normalized standard deviation of r_2, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}1}$.



(A) Approximation (B) Monte–Carlo

FIGURE A.6.: Normalized standard deviation of r_3, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}1}$.

II. Four-message TWR with perfect time of transmission

A scenario where the delays measured by the A node are not correlated is considered within this subsection. Such decorrelation can be accomplished by using 4-message negotiation as is described in Figure A.7. In that case, two complete, independent single-sided TWR negotiations are performed; of course, each node initiates one of the negotiations. Consequently, the same measurements are available as in the 3-message case without the correlation between the τ_{A1} and τ_{A2}. Nonetheless, the cost of the extra message has to be paid by virtue of decorrelation.

It is obvious that the corresponding covariance matrix is purely diagonal. Assuming that all the reception timestamps have the same variance σ^2_{Rx} it is

$$\boldsymbol{\Phi}_{\boldsymbol{\tau}2} = \sigma^2_{\mathrm{Rx}} \mathbb{I}_{4\times 4}, \tag{A.18}$$

where $\mathbb{I}_{4\times 4}$ is an identity matrix of size 4 by 4.

The mean of output values for the particular range estimators is approximated by the known procedure of substituting the covariance matrix elements and corresponding Hessian matrix elements into (A.3).

$$\begin{aligned}
\mathrm{E}[r_1] &\approx r|_{\boldsymbol{\mu}_\tau} + c_0\sigma^2_{\mathrm{Rx}}\left(\frac{8(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{2(\mu_{\tau\mathrm{B1}} - \mu_{\tau\mathrm{B2}})}{(\Sigma_\mu)^2}\right) \\
&= r|_{\boldsymbol{\mu}_\tau} + \sigma^2_{\mathrm{Rx}}\left(\frac{8\, r|_{\boldsymbol{\mu}_\tau} + 2c_0(\mu_{\tau\mathrm{B1}} - \mu_{\tau\mathrm{B2}})}{(\Sigma_\mu)^2}\right)
\end{aligned} \tag{A.19}$$

$$\begin{aligned}
\mathrm{E}[r_2] &\approx r|_{\boldsymbol{\mu}_\tau} + c_0\sigma^2_{\mathrm{Rx}}\left(\frac{8(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{2(\mu_{\tau\mathrm{A2}} - \mu_{\tau\mathrm{A1}})}{(\Sigma_\mu)^2}\right) \\
&= r|_{\boldsymbol{\mu}_\tau} + \sigma^2_{\mathrm{Rx}}\left(\frac{8\, r|_{\boldsymbol{\mu}_\tau} + 2c_0(\mu_{\tau\mathrm{A2}} - \mu_{\tau\mathrm{A1}})}{(\Sigma_\mu)^2}\right)
\end{aligned} \tag{A.20}$$

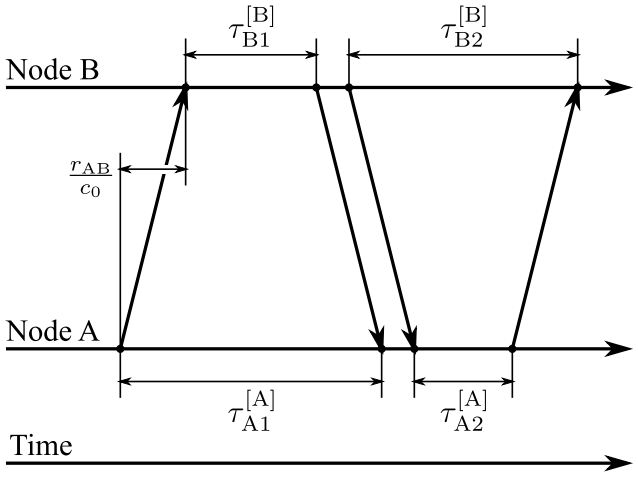


FIGURE A.7.: Double-sided two-way ranging timing diagram with 4 messages for measurement decorrelation.

$$\begin{aligned}\mathrm{E}[r_3] &\approx r|_{\boldsymbol{\mu}_\tau} + c_0\sigma_{\mathrm{Rx}}^2\left(\frac{4(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{\mu_{\tau\mathrm{A2}} - \mu_{\tau\mathrm{A1}} + \mu_{\tau\mathrm{B1}} - \mu_{\tau\mathrm{B2}}}{(\Sigma_\mu)^2}\right) \\ &= r|_{\boldsymbol{\mu}_\tau} + \sigma_{\mathrm{Rx}}^2\left(\frac{4\,r|_{\boldsymbol{\mu}_\tau} - 4\,r|_{\boldsymbol{\mu}_\tau}}{(\Sigma_\mu)^2}\right) = r|_{\boldsymbol{\mu}_\tau} + 0 \end{aligned} \tag{A.21}$$

The graphical representation of the expressions above is provided in Figures A.8 to A.10. Under assumption of no delay covariance, the only unbiased estimator is the r_3. The bias of r_1 and r_2 exhibits mutual symmetry, see Figures A.8 and A.9. Such behavior is understandable due to the symmetric nature of the estimator equations (2.48), (2.49). In all three plots it is noticeable that the range dependency of the normalized bias is observable for extremely long ranges only. Likewise the 3-message TWR case, the magnitude of the bias is negligible for reasonable parameters.

The formulas for the range estimator variances are obtained by means of substitution covariance matrix elements and corresponding Jacobian matrix elements into (A.5).

$$\mathrm{E}\left[r_1 - \mathrm{E}\left[r\right]\right]^2 \approx \sigma_{\mathrm{Rx}}^2\left(\frac{c_0^2(\mu_{\tau\mathrm{B2}}{}^2 + \mu_{\tau\mathrm{A1}}{}^2 + \mu_{\tau\mathrm{B1}}{}^2 + \mu_{\tau\mathrm{A2}}{}^2) + 4c_0 r(\mu_{\tau\mathrm{B1}} - \mu_{\tau\mathrm{B2}}) + 8r^2}{(\Sigma_\mu)^2}\right) \tag{A.22}$$

$$\mathrm{E}\left[r_2 - \mathrm{E}\left[r\right]\right]^2 \approx \sigma_{\mathrm{Rx}}^2\left(\frac{c_0^2(\mu_{\tau\mathrm{B2}}{}^2 + \mu_{\tau\mathrm{A1}}{}^2 + \mu_{\tau\mathrm{B1}}{}^2 + \mu_{\tau\mathrm{A2}}{}^2) + 4c_0 r(\mu_{\tau\mathrm{A2}} - \mu_{\tau\mathrm{A1}}) + 8r^2}{(\Sigma_\mu)^2}\right) \tag{A.23}$$

$$\mathrm{E}\left[r_3 - \mathrm{E}\left[r\right]\right]^2 \approx \sigma_{\mathrm{Rx}}^2\left(\frac{c_0^2(\mu_{\tau\mathrm{B2}}{}^2 + \mu_{\tau\mathrm{A1}}{}^2 + \mu_{\tau\mathrm{B1}}{}^2 + \mu_{\tau\mathrm{A2}}{}^2) - 4r^2}{(\Sigma_\mu)^2}\right) \tag{A.24}$$

The resulting graphs of normalized standard deviations, along with the Monte–Carlo results, in Figures A.11 to A.13 show that the variance is the lowest when the negotiation is symmetric. However, the impact of the asymmetry is not critical. The range dependency is observable only for extreme ranges. We may say that the standard deviation of the estimate is 50 to 55 % of the reception timestamp standard deviation scaled by speed of light. This is a bit lower than in the case of 3-message TWR; the reason for the slightly improved performance is the mutual independence of all the delay measurements.

The simulation results match the approximation very well for all three range estimators.

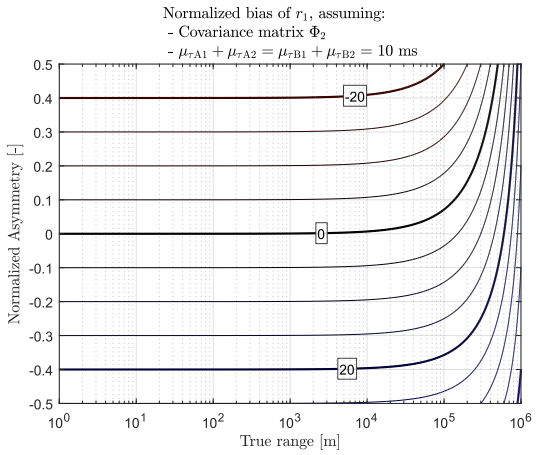


FIGURE A.8.: Normalized bias of r_1, assuming $\boldsymbol{\Phi}_{\boldsymbol{\tau}2}$.

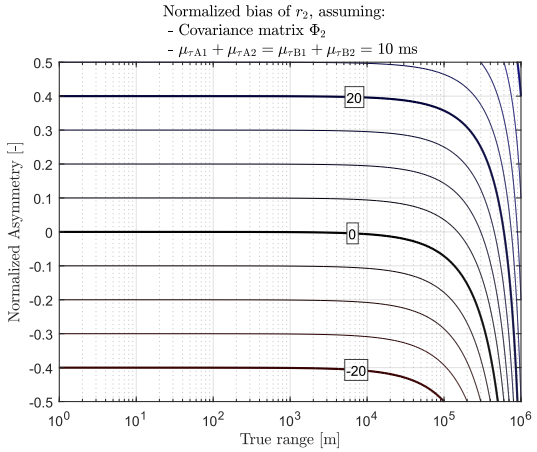


FIGURE A.9.: Normalized bias of r_2, assuming $\boldsymbol{\Phi}_{\boldsymbol{\tau}2}$.

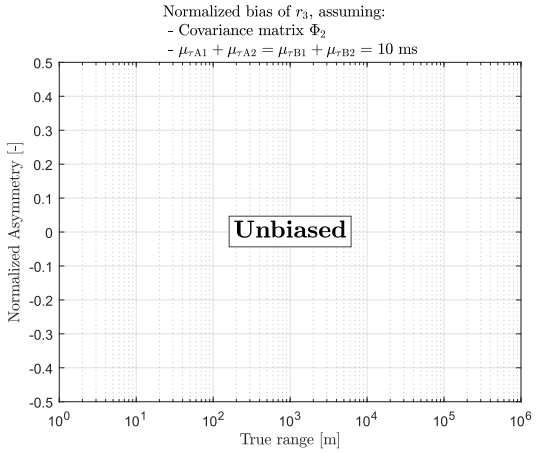
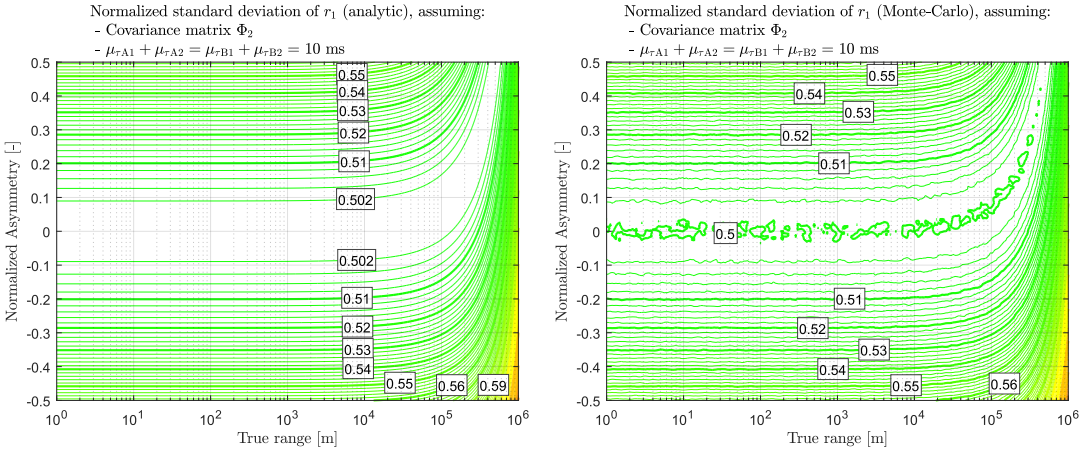
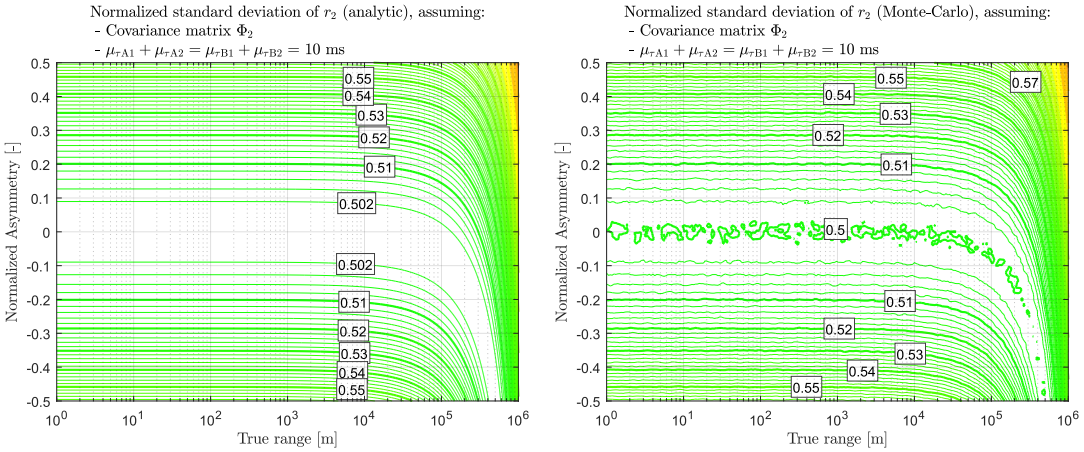


FIGURE A.10.: Normalized bias of r_3, assuming $\boldsymbol{\Phi}_{\boldsymbol{\tau}2}$.



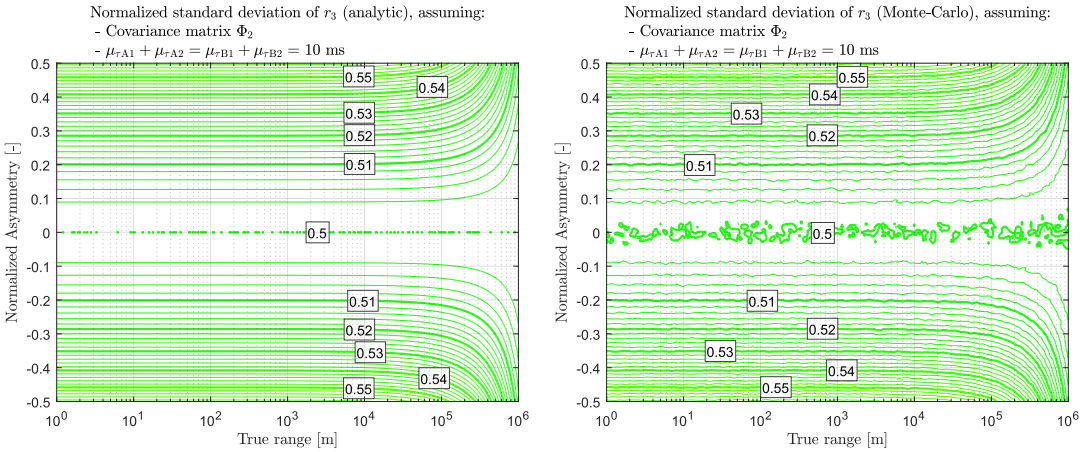
(A) Approximation (B) Monte–Carlo

FIGURE A.11.: Normalized standard deviation of r_1, assuming $\mathbf{\Phi_\tau 2}$.



(A) Approximation (B) Monte–Carlo

FIGURE A.12.: Normalized standard deviation of r_2, assuming $\mathbf{\Phi_\tau 2}$.



(A) Approximation (B) Monte–Carlo

FIGURE A.13.: Normalized standard deviation of r_3, assuming $\mathbf{\Phi_\tau 2}$.

III. Three-message TWR with imperfect time of transmission

For the sake of completeness, we may consider a 3-message TWR negotiation where both transmission and reception timestamps have the same variance, i.e. $\sigma^2_{\mathrm{Tx}} = \sigma^2_{\mathrm{Rx}}$. Consequently, all the delays have the variance equal to the double of the time of reception variance. In this case, there is a correlation between the measurements taken on both nodes. The variance-covariance matrix yields

$$\boldsymbol{\Phi}_{\boldsymbol{\tau}3} = \sigma^2_{\mathrm{Rx}} \begin{bmatrix} 2 & 0 & -1 & 0 \\ 0 & 2 & 0 & -1 \\ -1 & 0 & 2 & 0 \\ 0 & -1 & 0 & 2 \end{bmatrix}, \tag{A.25}$$

because one of the timestamps that define the delay is shared with the other delay measurement performed on the same node.

The mean values of the estimates can be approximated by:

$$\begin{aligned} \mathrm{E}[r_1] &\approx r|_{\boldsymbol{\mu}_\tau} + c_0\sigma^2_{\mathrm{Rx}}\left(\frac{8(\mu_{\tau\mathrm{A}1}\mu_{\tau\mathrm{B}2} - \mu_{\tau\mathrm{A}2}\mu_{\tau\mathrm{B}1})}{(\Sigma_\mu)^3} + \frac{2(\mu_{\tau\mathrm{B}1} - \mu_{\tau\mathrm{B}2})}{(\Sigma_\mu)^2}\right) \\ &= r|_{\boldsymbol{\mu}_\tau} + \sigma^2_{\mathrm{Rx}}\left(\frac{8\, r|_{\boldsymbol{\mu}_\tau} + 2c_0(\mu_{\tau\mathrm{B}1} - \mu_{\tau\mathrm{B}2})}{(\Sigma_\mu)^2}\right) \end{aligned} \tag{A.26}$$

$$\begin{aligned} \mathrm{E}[r_2] &\approx r|_{\boldsymbol{\mu}_\tau} + c_0\sigma^2_{\mathrm{Rx}}\left(\frac{8(\mu_{\tau\mathrm{A}1}\mu_{\tau\mathrm{B}2} - \mu_{\tau\mathrm{A}2}\mu_{\tau\mathrm{B}1})}{(\Sigma_\mu)^3} + \frac{2(\mu_{\tau\mathrm{A}2} - \mu_{\tau\mathrm{A}1})}{(\Sigma_\mu)^2}\right) \\ &= r|_{\boldsymbol{\mu}_\tau} + \sigma^2_{\mathrm{Rx}}\left(\frac{8\, r|_{\boldsymbol{\mu}_\tau} + 2c_0(\mu_{\tau\mathrm{A}2} - \mu_{\tau\mathrm{A}1})}{(\Sigma_\mu)^2}\right) \end{aligned} \tag{A.27}$$

$$\begin{aligned} \mathrm{E}[r_3] &\approx r|_{\boldsymbol{\mu}_\tau} + c_0\sigma^2_{\mathrm{Rx}}\left(\frac{4(\mu_{\tau\mathrm{A}1}\mu_{\tau\mathrm{B}2} - \mu_{\tau\mathrm{A}2}\mu_{\tau\mathrm{B}1})}{(\Sigma_\mu)^3} + \frac{\mu_{\tau\mathrm{A}2} - \mu_{\tau\mathrm{A}1} + \mu_{\tau\mathrm{B}1} - \mu_{\tau\mathrm{B}2}}{(\Sigma_\mu)^2}\right) \\ &= r|_{\boldsymbol{\mu}_\tau} + \sigma^2_{\mathrm{Rx}}\left(\frac{4\, r|_{\boldsymbol{\mu}_\tau} - 4\, r|_{\boldsymbol{\mu}_\tau}}{(\Sigma_\mu)^2}\right) = r|_{\boldsymbol{\mu}_\tau} + 0. \end{aligned} \tag{A.28}$$

The contour plots of the expressions above are provided in Figures A.14 to A.16. As in the 4-message decorrelated case, only the r_3 is unbiased. Naturally, due to the double variances and possibly due to the covariance terms, the bias magnitude is significantly higher than in the uncorrelated case, i.e. when covariance matrix $\boldsymbol{\Phi}_{\boldsymbol{\tau}2}$ was assumed. Nevertheless, the bias is still negligible component of the ranging error.

The approximate variances of the three estimators under assumption of covariance matrix $\boldsymbol{\Phi}_{\boldsymbol{\tau}3}$ follow:

$$\mathrm{E}\left[r_1 - \mathrm{E}\left[r\right]\right]^2 \approx \sigma_{\mathrm{Rx}}^2 \left(\frac{2c_0^2(\mu_{\tau\mathrm{B}2}{}^2 + \mu_{\tau\mathrm{A}1}{}^2 + \mu_{\tau\mathrm{B}1}{}^2 + \mu_{\tau\mathrm{A}2}{}^2 + \mu_{\tau\mathrm{A}1}\mu_{\tau\mathrm{A}2} + \mu_{\tau\mathrm{B}1}\mu_{\tau\mathrm{B}2})}{(\Sigma_\mu)^2} + \frac{4c_0 r(\mu_{\tau\mathrm{B}1} - \mu_{\tau\mathrm{B}2}) + 8r^2}{(\Sigma_\mu)^2} \right) \tag{A.29}$$

$$\mathrm{E}\left[r_2 - \mathrm{E}\left[r\right]\right]^2 \approx \sigma_{\mathrm{Rx}}^2 \left(\frac{2c_0^2(\mu_{\tau\mathrm{B}2}{}^2 + \mu_{\tau\mathrm{A}1}{}^2 + \mu_{\tau\mathrm{B}1}{}^2 + \mu_{\tau\mathrm{A}2}{}^2 + \mu_{\tau\mathrm{A}1}\mu_{\tau\mathrm{A}2} + \mu_{\tau\mathrm{B}1}\mu_{\tau\mathrm{B}2})}{(\Sigma_\mu)^2} + \frac{4c_0 r(\mu_{\tau\mathrm{A}2} - \mu_{\tau\mathrm{A}1}) + 8r^2}{(\Sigma_\mu)^2} \right) \tag{A.30}$$

$$\mathrm{E}\left[r_3 - \mathrm{E}\left[r\right]\right]^2 \approx \sigma_{\mathrm{Rx}}^2 \left(\frac{2c_0^2(\mu_{\tau\mathrm{B}2}{}^2 + \mu_{\tau\mathrm{A}1}{}^2 + \mu_{\tau\mathrm{B}1}{}^2 + \mu_{\tau\mathrm{A}2}{}^2 + \mu_{\tau\mathrm{A}1}\mu_{\tau\mathrm{A}2} + \mu_{\tau\mathrm{B}1}\mu_{\tau\mathrm{B}2})}{(\Sigma_\mu)^2} + \frac{4r^2}{(\Sigma_\mu)^2} \right). \tag{A.31}$$

The contour plots of the standard deviations based on the equations above and the respective Monte–Carlo results are provided in Figures A.14 to A.16. The shape of the contours is rather similar to the 4-message decorrelated case, the magnitude is higher, however. If the covariances were neglected, the two cases would differ only by scaling factor of $\sqrt{2}$, since the variances are doubled in $\boldsymbol{\Phi}_{\boldsymbol{\tau}3}$ with respect to $\boldsymbol{\Phi}_{\boldsymbol{\tau}2}$. Due to the covariance, the multiplying factor is even higher.

As well as in the previous sections we can assert that there is an almost perfect match of the approximation and Monte–Carlo simulation results.

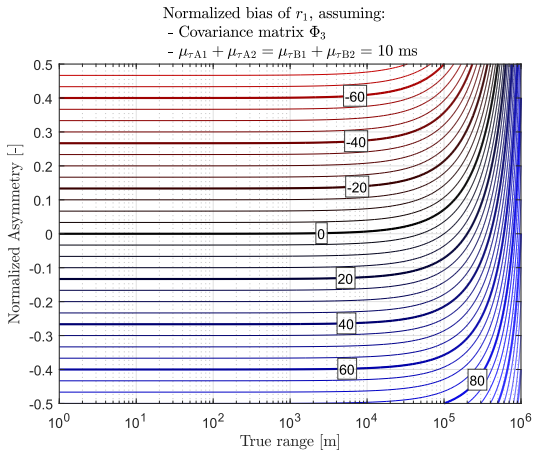


FIGURE A.14.: Normalized bias of r_1, assuming $\boldsymbol{\Phi_{\tau}}_3$.

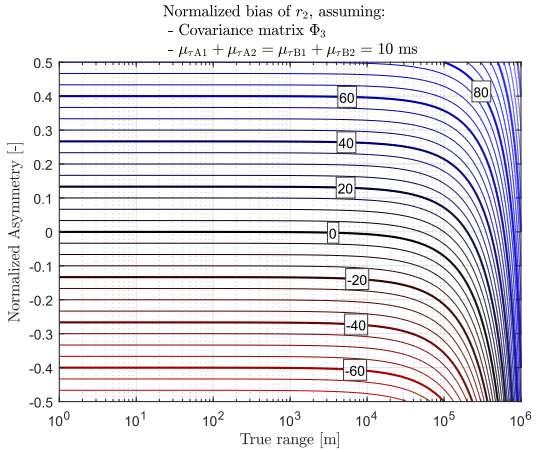


FIGURE A.15.: Normalized bias of r_2, assuming $\boldsymbol{\Phi_{\tau}}_3$.

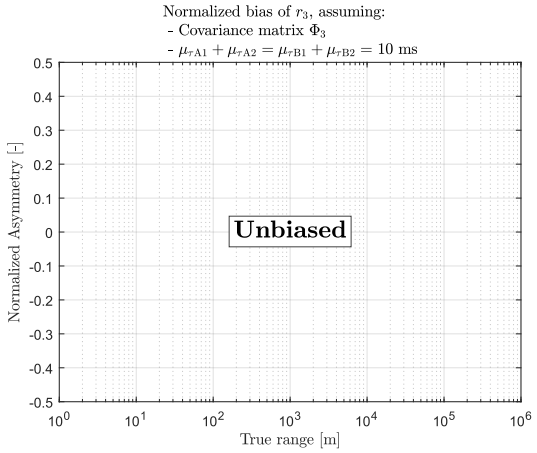
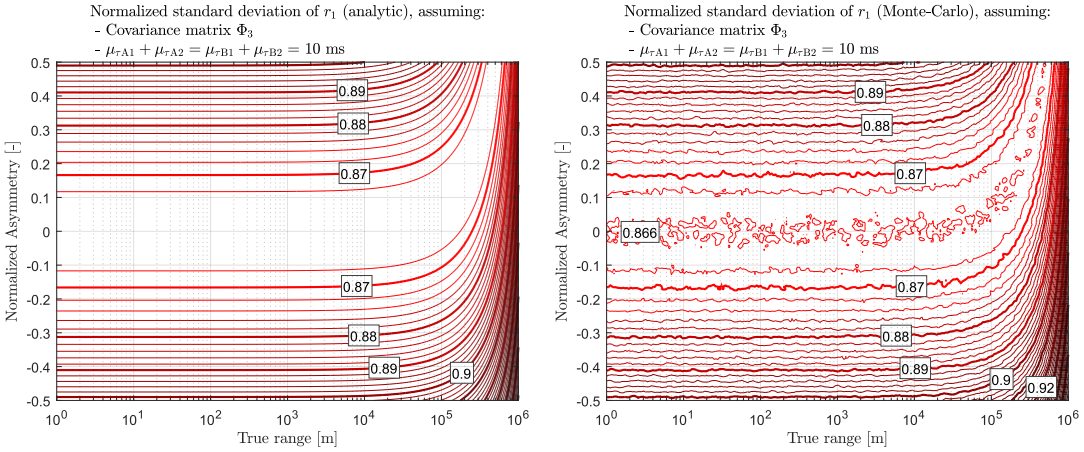
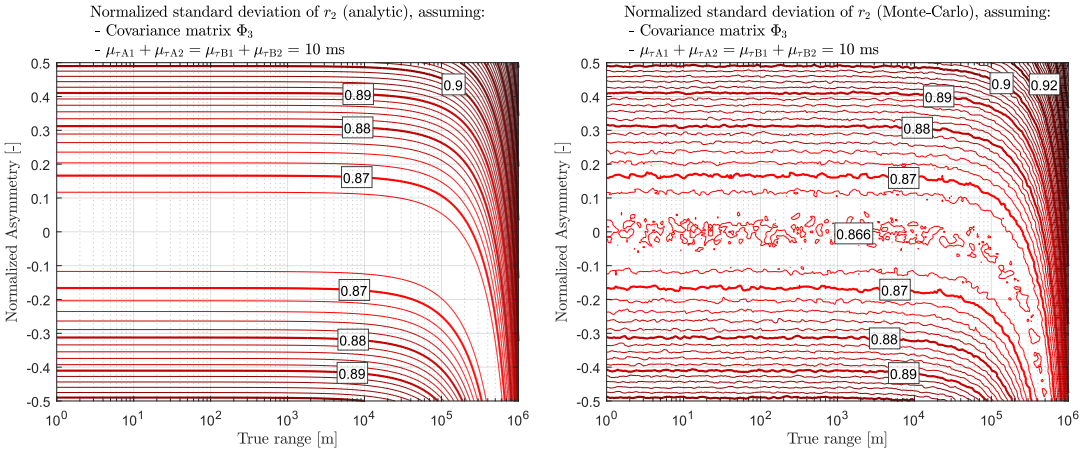


FIGURE A.16.: Normalized bias of r_3, assuming $\boldsymbol{\Phi_{\tau}}_3$.



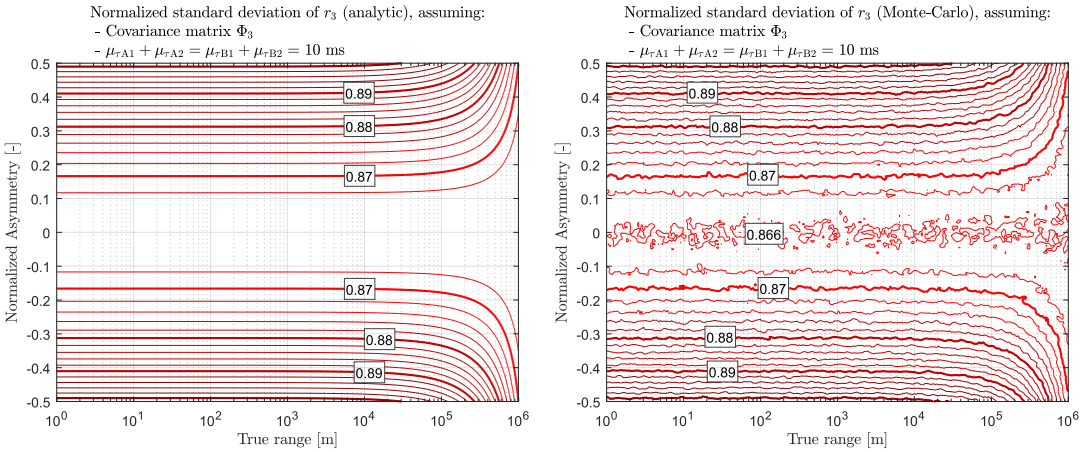
(A) Approximation (B) Monte–Carlo

FIGURE A.17.: Normalized standard deviation of r_1, assuming $\boldsymbol{\Phi_\tau 3}$.



(A) Approximation (B) Monte–Carlo

FIGURE A.18.: Normalized standard deviation of r_2, assuming $\boldsymbol{\Phi_\tau 3}$.



(A) Approximation (B) Monte–Carlo

FIGURE A.19.: Normalized standard deviation of r_3, assuming $\boldsymbol{\Phi_\tau 3}$.

A.3. Remarks

The biases and variances of the ranging estimators for ADS-TWR were approximated by means of Taylor expansion within this appendix. The results were verified by a Monte–Carlo simulation, almost perfect match of the simulation results and analytical approximation was observed. The estimators were evaluated under three different delay covariance matrix assumptions. When the correlation of the measurements taken on both nodes were symmetric, only the r_3 estimator (2.50) was unbiased. Under the assumption of perfect transmission timestamps and 3-message negotiation, r_1 estimator (2.48) was unbiased. Nonetheless, the magnitude of the bias was negligible (nanometer-scale) in all the investigated cases.

From the standpoint of variance (standard deviation) it was observed that for response delays significantly longer than time-of-flight, the range dependency of the variance is negligible. A weak dependency on the asymmetry of the response delays was observed. In all the cases, the estimated range standard deviation is lower than the standard deviation of the delay measurement scaled by speed of light. Results suggest that decorrelation of the delay measurements, e.g. by introduction of the fourth message in the TWR negotiation (see Fig. A.7), should provide slight improvement of the estimator performance in terms of its variance.

We may conclude that the performance of the nonlinear TWR estimators described by (2.48) to (2.50) is similar to the performance of the linear estimator defined by (2.33). The bias that is induced by some of the nonlinear estimators is negligible under common operating conditions.

A.4. The Complete Jacobian and Hessian Matrices

Note that for the sake of reasonable length substitution

$$\Sigma_\mu = 2(\mu_{\tau\mathrm{A1}} + \mu_{\tau\mathrm{A2}}) = 2(\mu_{\tau\mathrm{B1}} + \mu_{\tau\mathrm{B2}}) = \mu_{\tau\mathrm{A1}} + \mu_{\tau\mathrm{A2}} + \mu_{\tau\mathrm{B1}} + \mu_{\tau\mathrm{B2}} \tag{A.7}$$

is used in the following equations.

Jacobian matrix elements for range estimator (2.48):

$$\begin{aligned}
\left.\frac{\partial r_1}{\partial \tau_{\mathrm{A1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{\mu_{\tau\mathrm{B2}}}{\Sigma_\mu} - \frac{2(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial r_1}{\partial \tau_{\mathrm{B2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{\mu_{\tau\mathrm{A1}}}{\Sigma_\mu} \right) \\
\left.\frac{\partial r_1}{\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{-\mu_{\tau\mathrm{B1}}}{\Sigma_\mu} - \frac{2(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial r_1}{\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{-\mu_{\tau\mathrm{A2}}}{\Sigma_\mu} \right)
\end{aligned} \tag{A.32}$$

Hessian matrix elements for range estimator (2.48):

$$\begin{aligned}
\left.\frac{\partial^2 r_1}{\partial \tau_{\mathrm{A1}}^2}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{8(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^3} - \frac{4\mu_{\tau\mathrm{B2}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_1}{\partial \tau_{\mathrm{B2}}^2}\right|_{\boldsymbol{\mu}_\tau} &= 0 \\
\left.\frac{\partial^2 r_1}{\partial \tau_{\mathrm{A2}}^2}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{8(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{4\mu_{\tau\mathrm{B1}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_1}{\partial \tau_{\mathrm{B1}}^2}\right|_{\boldsymbol{\mu}_\tau} &= 0 \\
\left.\frac{\partial^2 r_1}{\partial \tau_{\mathrm{A1}} \partial \tau_{\mathrm{B2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{-2\mu_{\tau\mathrm{A1}}}{(\Sigma_\mu)^2} + \frac{1}{\Sigma_\mu} \right) \\
\left.\frac{\partial^2 r_1}{\partial \tau_{\mathrm{A1}} \partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{8(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{2(\mu_{\tau\mathrm{B1}} - \mu_{\tau\mathrm{B2}})}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_1}{\partial \tau_{\mathrm{A1}} \partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2\mu_{\tau\mathrm{A2}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_1}{\partial \tau_{\mathrm{B2}} \partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{-2\mu_{\tau\mathrm{A1}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_1}{\partial \tau_{\mathrm{B2}} \partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= 0 \\
\left.\frac{\partial^2 r_1}{\partial \tau_{\mathrm{A2}} \partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2\mu_{\tau\mathrm{A2}}}{(\Sigma_\mu)^2} - \frac{1}{\Sigma_\mu} \right)
\end{aligned} \tag{A.33}$$

Jacobian matrix elements for range estimator (2.49):

$$\begin{aligned}
\left.\frac{\partial r_2}{\partial \tau_{\mathrm{A1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{\mu_{\tau\mathrm{B2}}}{\Sigma_\mu}\right) \\
\left.\frac{\partial r_2}{\partial \tau_{\mathrm{B2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{\mu_{\tau\mathrm{A1}}}{\Sigma_\mu} - \frac{2(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^2}\right) \\
\left.\frac{\partial r_2}{\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{-\mu_{\tau\mathrm{B1}}}{\Sigma_\mu}\right) \\
\left.\frac{\partial r_2}{\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{-\mu_{\tau\mathrm{A2}}}{\Sigma_\mu} - \frac{2(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^2}\right)
\end{aligned} \tag{A.34}$$

Hessian matrix elements for range estimator (2.49):

$$\begin{aligned}
\left.\frac{\partial^2 r_2}{\partial \tau_{\mathrm{A1}}^2}\right|_{\boldsymbol{\mu}_\tau} &= 0 \\
\left.\frac{\partial^2 r_2}{\partial \tau_{\mathrm{B2}}^2}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{8(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^3} - \frac{4\mu_{\tau\mathrm{A1}}}{(\Sigma_\mu)^2}\right) \\
\left.\frac{\partial^2 r_2}{\partial \tau_{\mathrm{A2}}^2}\right|_{\boldsymbol{\mu}_\tau} &= 0 \\
\left.\frac{\partial^2 r_2}{\partial \tau_{\mathrm{B1}}^2}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{8(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{4\mu_{\tau\mathrm{A2}}}{(\Sigma_\mu)^2}\right) \\
\left.\frac{\partial^2 r_2}{\partial \tau_{\mathrm{A1}}\partial \tau_{\mathrm{B2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{-2\mu_{\tau\mathrm{B2}}}{(\Sigma_\mu)^2} + \frac{1}{\Sigma_\mu}\right) \\
\left.\frac{\partial^2 r_2}{\partial \tau_{\mathrm{A1}}\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} &= 0 \\
\left.\frac{\partial^2 r_2}{\partial \tau_{\mathrm{A1}}\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{-2\mu_{\tau\mathrm{B2}}}{(\Sigma_\mu)^2}\right) \\
\left.\frac{\partial^2 r_2}{\partial \tau_{\mathrm{B2}}\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2\mu_{\tau\mathrm{B1}}}{(\Sigma_\mu)^2}\right) \\
\left.\frac{\partial^2 r_2}{\partial \tau_{\mathrm{B2}}\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{8(\mu_{\tau\mathrm{A1}}\mu_{\tau\mathrm{B2}} - \mu_{\tau\mathrm{A2}}\mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{2(\mu_{\tau\mathrm{A2}} - \mu_{\tau\mathrm{A1}})}{(\Sigma_\mu)^2}\right) \\
\left.\frac{\partial^2 r_2}{\partial \tau_{\mathrm{A2}}\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2(\mu_{\tau\mathrm{A2}} + \mu_{\tau\mathrm{B1}})}{(\Sigma_\mu)^2} - \frac{1}{\Sigma_\mu}\right)
\end{aligned} \tag{A.35}$$

Jacobian matrix elements for range estimator (2.50):

$$\begin{aligned}
\left.\frac{\partial r_2}{\partial \tau_{\mathrm{A1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{\mu_{\tau \mathrm{B2}}}{\Sigma_\mu} - \frac{\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial r_2}{\partial \tau_{\mathrm{B2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{\mu_{\tau \mathrm{A1}}}{\Sigma_\mu} - \frac{\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial r_2}{\partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{-\mu_{\tau \mathrm{B1}}}{\Sigma_\mu} - \frac{\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial r_2}{\partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{-\mu_{\tau \mathrm{A2}}}{\Sigma_\mu} - \frac{\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}}}{(\Sigma_\mu)^2} \right)
\end{aligned} \tag{A.36}$$

Hessian matrix elements for range estimator (2.50):

$$\begin{aligned}
\left.\frac{\partial^2 r_3}{\partial \tau_{\mathrm{A1}}^2}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2(\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}})}{(\Sigma_\mu)^3} - \frac{2\mu_{\tau \mathrm{B2}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_3}{\partial \tau_{\mathrm{B2}}^2}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2(\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}})}{(\Sigma_\mu)^3} - \frac{2\mu_{\tau \mathrm{A1}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_3}{\partial \tau_{\mathrm{A2}}^2}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2(\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{2\mu_{\tau \mathrm{B1}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_3}{\partial \tau_{\mathrm{B1}}^2}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2(\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{2\mu_{\tau \mathrm{A2}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_3}{\partial \tau_{\mathrm{A1}} \partial \tau_{\mathrm{B2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2(\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}})}{(\Sigma_\mu)^3} - \frac{\mu_{\tau \mathrm{B2}} + \mu_{\tau \mathrm{A1}}}{(\Sigma_\mu)^2} + \frac{1}{\Sigma_\mu} \right) \\
\left.\frac{\partial^2 r_3}{\partial \tau_{\mathrm{A1}} \partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2(\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{\mu_{\tau \mathrm{B1}} - \mu_{\tau \mathrm{B2}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_3}{\partial \tau_{\mathrm{A1}} \partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2(\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{\mu_{\tau \mathrm{A2}} - \mu_{\tau \mathrm{B2}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_3}{\partial \tau_{\mathrm{B2}} \partial \tau_{\mathrm{A2}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2(\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{\mu_{\tau \mathrm{B1}} - \mu_{\tau \mathrm{A1}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_3}{\partial \tau_{\mathrm{B2}} \partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2(\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{\mu_{\tau \mathrm{A2}} - \mu_{\tau \mathrm{A1}}}{(\Sigma_\mu)^2} \right) \\
\left.\frac{\partial^2 r_3}{\partial \tau_{\mathrm{A2}} \partial \tau_{\mathrm{B1}}}\right|_{\boldsymbol{\mu}_\tau} &= c_0 \left(\frac{2(\mu_{\tau \mathrm{A1}}\mu_{\tau \mathrm{B2}} - \mu_{\tau \mathrm{A2}}\mu_{\tau \mathrm{B1}})}{(\Sigma_\mu)^3} + \frac{\mu_{\tau \mathrm{A2}} + \mu_{\tau \mathrm{B1}}}{(\Sigma_\mu)^2} - \frac{1}{\Sigma_\mu} \right)
\end{aligned} \tag{A.37}$$

B. Examples of TDoA Positioning

This Appendix shows practical examples of solving TDoA sets of equations by means of the iterative least-squares solvers described in Section 4.1 (page 111).

B.1. DVB-T Opportunistic Localization

In the first example, a two-dimensional localization is performed in a metropolitan area by means of measuring TDoA of signals from three synchronized transmitters of a DVB-T (Terrestrial digital video broadcasting) single-frequency network (SFN). The related signal processing and method of assigning measurements to the corresponding transmitters is described thoroughly in [125, 163] and further developed in [121]. Herein, only a brief recapitulation is provided.

B.1.1. Obtaining the TDoA measurements

The TDoA measurements were obtained by means of batch processing of a captured signal. The DVB-T signal was demodulated and its noiseless replica (or at least the replica of the pilot signals) is correlated with the received signal. It is possible to evaluate the cross-ambiguity function (CAF), according to its definition in (2.150), and find the temporal difference between the peaks that correspond to the time delays of signals received from multiple transmitters [125]. An exemplary evaluation of CAF, which is based on a signal captured in Prague, Czech republic, is depicted in Figure B.1. Three peaks can be easily found.

In reality, the full CAF is barely computed, since the frequency of all transmitters in the SFN is almost identical and Doppler shift is typically negligible. The offset of the receiver center frequency is estimated and eliminated; then the correlation of signal and its replica is performed by means of matched filtration.[114] In other words a single fixed-frequency cut of the CAF is obtained. The peaks corresponding to the respective signal delays can be found e.g. by means of a CA-CFAR detection[115] algorithm, as suggested in [121].

B.1.2. Applying the Newton-Raphson algorithm

Anyway, three peak delays w.r.t. an unknown start of the recording are available, and therefore 2 linearly independent TDoA measurements are obtainable. Consequently,

[114] Mismatched filtration is suggested in [121] in order to suppress time-domain sidelobes of the peaks, which may be misinterpreted as additional transmitter on the expense of minor loss of accuracy and peak resolution.

[115] Cell-averaging constant false alarm rate detection.

only two-dimensional problem can be solved. The receiver is assumed to be in the vicinity of Earth's surface, which is further approximated by a plane. An ENU coordinate frame is constructed at the centroid of the transmitter positions, and its height is set so that the error caused by the Earth's surface curvature for certain area is minimized (detailed analysis is available in [164]). The transmitter coordinates are transformed to this ENU frame (including height) and the receiver position coordinates are estimated in the East-North plane only. Thus, we define the unknown user position 2-vector $\boldsymbol{r}_\mathrm{u}$ and the *a priori* known i-th transmitter position 3-vector $\boldsymbol{r}_\mathrm{i}$ as

$$\boldsymbol{r}_\mathrm{u} = \begin{bmatrix} r_{\mathrm{u},E} \\ r_{\mathrm{u},N} \end{bmatrix}; \qquad \boldsymbol{r}_\mathrm{i} = \begin{bmatrix} r_{\mathrm{i},E} \\ r_{\mathrm{i},N} \\ r_{\mathrm{i},U} \end{bmatrix} \boldsymbol{r}_\mathrm{i}. \tag{B.1}$$

The TDoA measurement equation (2.64) from page 21 has to be adapted to the different dimensions of the user and transmitter position vectors. The missing user height $r_{\mathrm{u},U} = 0$, and therefore

$$\begin{aligned} d_{\mathrm{i,j}} =& \left\| \begin{bmatrix} \boldsymbol{r}_\mathrm{u} \\ 0 \end{bmatrix} - \boldsymbol{r}_\mathrm{i} \right\| - \left\| \begin{bmatrix} \boldsymbol{r}_\mathrm{u} \\ 0 \end{bmatrix} - \boldsymbol{r}_\mathrm{j} \right\| \\ =& \sqrt{(r_{\mathrm{u},E} - r_{\mathrm{i},E})^2 + (r_{\mathrm{u},N} - r_{\mathrm{i},N})^2 + r_{\mathrm{i},U}^2} \\ & - \sqrt{(r_{\mathrm{u},E} - r_{\mathrm{j},E})^2 + (r_{\mathrm{u},N} - r_{\mathrm{j},N})^2 + r_{\mathrm{j},U}^2}. \end{aligned} \tag{B.2}$$

The Jacobian matrix corresponding to such expression of the TDoA measurement can be obtained straightforwardly. Note that the height component of the transmitters can be neglected in the evaluation, since the direction of the gradient in the horizontal plane remains the same, and the magnitude difference is negligible (unless very close to a transmitter).

$$\begin{aligned} \boldsymbol{J}_{d\mathrm{i,j}} =& \begin{bmatrix} \frac{\partial d_{\mathrm{i,j}}}{\partial r_{\mathrm{u},E}} \\ \frac{\partial d_{\mathrm{i,j}}}{\partial r_{\mathrm{u},N}} \end{bmatrix} \\ =& \begin{bmatrix} \frac{r_{\mathrm{u},E} - r_{\mathrm{i},E}}{\sqrt{\left(r_{\mathrm{u},E} - r_{\mathrm{i},E}\right)^2 + \left(r_{\mathrm{u},N} - r_{\mathrm{i},N}\right)^2 + r_{\mathrm{i},U}^2}} - \frac{r_{\mathrm{u},E} - r_{\mathrm{j},E}}{\sqrt{\left(r_{\mathrm{u},E} - r_{\mathrm{j},E}\right)^2 + \left(r_{\mathrm{u},N} - r_{\mathrm{j},N}\right)^2 + r_{\mathrm{j},U}^2}} \\ \frac{r_{\mathrm{u},N} - r_{\mathrm{i},N}}{\sqrt{\left(r_{\mathrm{u},E} - r_{\mathrm{i},E}\right)^2 + \left(r_{\mathrm{u},N} - r_{\mathrm{i},N}\right)^2 + r_{\mathrm{i},U}^2}} - \frac{r_{\mathrm{u},N} - r_{\mathrm{j},N}}{\sqrt{\left(r_{\mathrm{u},E} - r_{\mathrm{j},E}\right)^2 + \left(r_{\mathrm{u},N} - r_{\mathrm{j},N}\right)^2 + r_{\mathrm{j},U}^2}} \end{bmatrix} \\ \approx& \begin{bmatrix} \frac{r_{\mathrm{u},E} - r_{\mathrm{i},E}}{\sqrt{\left(r_{\mathrm{u},E} - r_{\mathrm{i},E}\right)^2 + \left(r_{\mathrm{u},N} - r_{\mathrm{i},N}\right)^2}} - \frac{r_{\mathrm{u},E} - r_{\mathrm{j},E}}{\sqrt{\left(r_{\mathrm{u},E} - r_{\mathrm{j},E}\right)^2 + \left(r_{\mathrm{u},N} - r_{\mathrm{j},N}\right)^2}} \\ \frac{r_{\mathrm{u},N} - r_{\mathrm{i},N}}{\sqrt{\left(r_{\mathrm{u},E} - r_{\mathrm{i},E}\right)^2 + \left(r_{\mathrm{u},N} - r_{\mathrm{i},N}\right)^2}} - \frac{r_{\mathrm{u},N} - r_{\mathrm{j},N}}{\sqrt{\left(r_{\mathrm{u},E} - r_{\mathrm{j},E}\right)^2 + \left(r_{\mathrm{u},N} - r_{\mathrm{j},N}\right)^2}} \end{bmatrix} \end{aligned} \tag{B.3}$$

In order to solve a set of equations of type (B.2), we will apply the non-weighted Newton-Raphson algorithm, which was described in Section 4.1.1. In our scenario there are 3 transmitters, the initial guess (prediction) of the user position is denoted by $\tilde{\boldsymbol{r}}$. First, the residuals are computed for all TDoA measurement pairs, i.e.

$$\check{d}_{1,2} = d_{1,2} - \left(\left\| \begin{bmatrix} \tilde{\boldsymbol{r}}_\mathrm{u} \\ 0 \end{bmatrix} - \boldsymbol{r}_\mathrm{i} \right\| - \left\| \begin{bmatrix} \tilde{\boldsymbol{r}}_\mathrm{u} \\ 0 \end{bmatrix} - \boldsymbol{r}_\mathrm{j} \right\| \right), \tag{B.4}$$

and similarly for $\check{d}_{1,3}$ and $\check{d}_{2,3}$. The concatenated Jacobian matrix for all measurements is constructed according to

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_{d1,2} \\ \boldsymbol{J}_{d1,3} \\ \boldsymbol{J}_{d2,3} \end{bmatrix}. \tag{B.5}$$

The correction of the user position prediction is obtained using the pseudoinverse (4.8):

$$\check{\boldsymbol{r}}_{\mathrm{u}} = \left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}\right)^{-1}\boldsymbol{J}^{\mathrm{T}} \begin{bmatrix} \check{d}_{1,2} \\ \check{d}_{1,3} \\ \check{d}_{2,3} \end{bmatrix}. \tag{B.6}$$

The correction is applied $\boldsymbol{r} = \tilde{\boldsymbol{r}} + \check{\boldsymbol{r}}$ and the corrected value is taken as the next, more accurate prediction (guess) $\tilde{\boldsymbol{r}} \leftarrow \boldsymbol{r}$. The iteration of this loop is stopped successfully when the correction magnitude drops below a certain threshold and unsuccessfully when maximal permitted number of iterations is reached.

It is obvious that the TDoA measurements $\check{d}_{1,2}$, $\check{d}_{1,3}$ and $\check{d}_{2,3}$ are linearly dependent. It does not cause problems in (B.6), since the 2×2 matrix $\left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}\right)$ is inverted. However, such choice of the measurement vector does not allow to use the weighted variant of the Newton-Raphson algorithm where the weighting matrix is the inverse of the TDoA measurement covariance matrix. The optimal covariance matrix that takes into account the covariances of the measurements is inherently rank-deficient for the set of TDoA measurements $\check{d}_{1,2}$, $\check{d}_{1,3}$ and $\check{d}_{2,3}$. An apparent solution is to omit one of the linearly-dependent measurements, which renders the corresponding covariance matrix full-rank. Here, the non-weighted method is used, thus a sub-optimal solution that neglects covariances and presumes equal variances of the measurements is obtained.

B.1.3. Experimental results

The DVB-T positioning was tested on 6 signal captures obtained at three different sites (Rx_1 to Rx_3) in Prague, Czech Repubic. A signal from three transmitters (Tx_1 to Tx_3) of a SFN at DVB-T channel 42 were received at each site; the TDoA values were measured according to [121].

The constellation of the transmitters and receiver sites is provided in Figure B.2. It is rather unfortunate that the transmitters are located almost on a single axis; the situation is nearly axially symmetric.[116] The horizontal dilution of precision (HDOP) is evaluated for each point in the area of interest in order to show where the performance is expected to be poor due to the unfavorable transmitter geometry. The HDOP is obtained as

$$HDOP = \sqrt{\mathrm{tr}\left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}\right)^{-1}} \tag{B.7}$$

evaluated at each point. The HDOP values are depicted in color in the plot; higher values (warmer colors) correspond to the areas with poor expected positioning performance.[117]

[116] The symmetry of the problem may cause the solution to be ambiguous, since there would be no information whether the user is located east or west from the axis of symmetry.

[117] It is acknowledged that the background map was plotted using the Map–Matlab utility [165], which uses tiles of the OpenStreetMap.

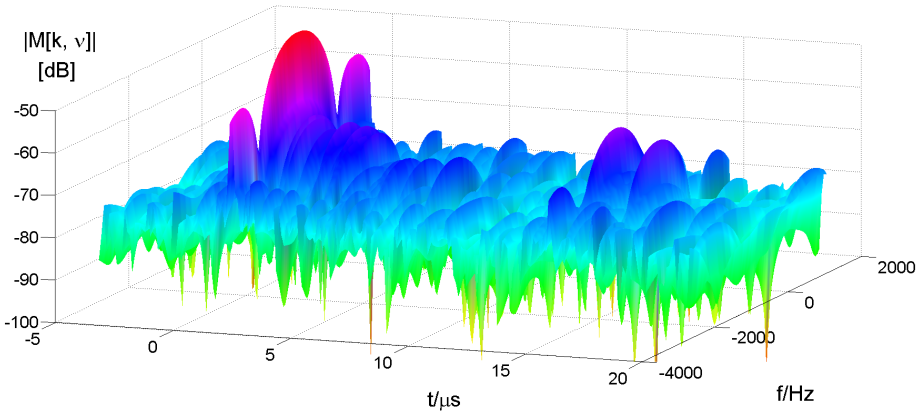


FIGURE B.1.: CAF of a DVB-T signal [125].

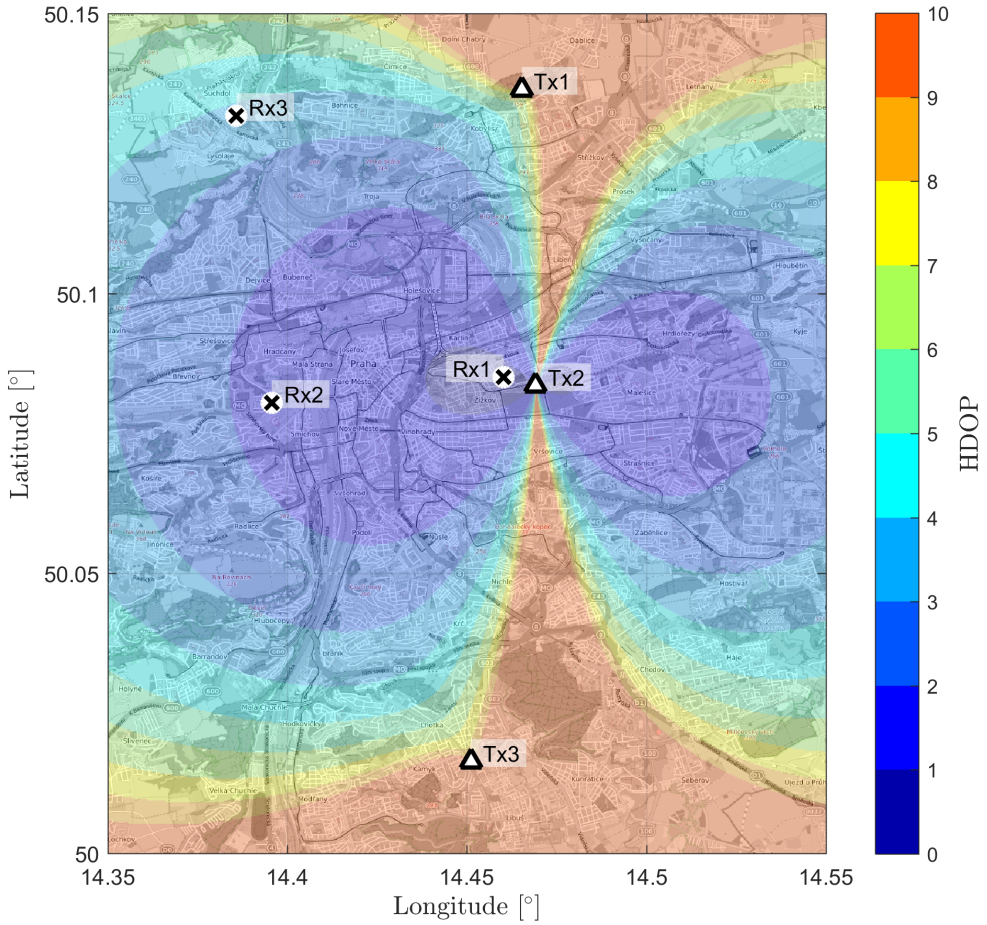


FIGURE B.2.: Constellation of DVB-T transmitters, receiver testing sites, HDOP in color; area map in background.
(Map source: `OpenStreetMap.org`)

TABLE B.1.: Experimental results of DVB-T – TDoA positioning in Prague [121].

Receiver	**Latitude** [deg]	**Longitude** [deg]	**HDOP**	**Error** [m]
Rx_1 I.	50.085 07	14.460 34	0.8	9
Rx_1 II.	50.085 10	14.460 18	0.8	3
Rx_2 I.	50.080 34	14.396 04	1.7	30
Rx_2 II.	50.080 31	14.395 87	1.7	22
Rx_3 I.	50.131 75	14.385 94	3.8	9
Rx_3 II.	50.131 23	14.387 21	3.7	115

The quantitative results of the 6 runs of Newon-Raphson-based positioning algorithm are presented in Table B.1. For low values of HDOP the horizontal positioning error is not higher than a few tens of meters w.r.t. true position (obtained by a GNSS receiver). In case of the test site Rx_3, the HDOP is close to 4, the error was higher than 100 m. It is probable that the error propagated from the TDoA measurement itself and was not caused by the positioning algorithm.

In all cases the solution was found within 8 or less iterations of the Newton-Raphson algorithm, even with the unfavorable geometry of the constellation.

B.2. Ultra-Wide Band System

More general approach was used in case of UWB-TDoA positioning. An UWB tag (the localized equipment) sends periodically a blink (positioning request) that is captured by multiple anchors. Each anchor timestamps the reception in its local timescale. Further, the timestamp is expressed in a master, system-wide, timescale; note that the exact mechanism of obtaining the master timescale information in the slave anchors is described in Appendix C.1.

The times of arrival in a common timescale are *de facto* pseudoranges, thus ToA positioning approach may be applied. However, we are not interested in estimating clock offset between the tag and master clocks; consequently, it is possible to eliminate it by means of differencing the measurements, i.e. obtaining TDoA measurements.[118] Moreover, from several experiments it can be observed that the convergence of TDoA problem is faster and more reliable than of the ToA problem (see e.g. comparison in Fig. 2.17 on page 74).

B.2.1. General definition of the TDoA measurements

Consider now that in a single epoch we have n UWB-pseudorange measurements available, let us denote their vector $\boldsymbol{\rho}$. We may obtain the vector of TDoA (denoted by $\boldsymbol{d}$) by means of linear transformation of $\boldsymbol{\rho}$

$$\boldsymbol{d} = \boldsymbol{D}\boldsymbol{\rho}, \tag{B.8}$$

[118] In case of GNSS, the estimation of the clock offset (bias) is necessary in order to be able to determine the satellite positions, *et cetera*.

where $\boldsymbol{D}$ is a matrix with n columns and $n-1$ rows which performs the differencing. Matrix $\boldsymbol{D}$ should have following properties:

- In each row must be one $+1$ and one -1 element, other elements are zero.
- In each column must be at least one nonzero element.
- Consequently, it is full-rank.

One of the possible forms of $\boldsymbol{D}$ follows

$$\boldsymbol{D} = \begin{bmatrix} +1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & +1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & +1 & -1 \end{bmatrix}. \tag{B.9}$$

The covariance matrix of the TDoA measurements can be obtained from the covariance matrix of the pseudoranges by means of

$$\boldsymbol{\Phi_d} = \boldsymbol{D}\boldsymbol{\Phi_\rho}\boldsymbol{D}^{\mathrm{T}}. \tag{B.10}$$

B.2.2. Applying the (un)constrained weighted solver

It is obvious that the position of $+1$ corresponds to the pseudorange measured at the i-th anchor and position of -1 corresponds to the pseudorange measured at the j-th anchor when related to the definition of TDoA measurement in Section 2.1.2–IV, eq. (2.64). Consequently, the residual computation for k-th element of $\boldsymbol{d}$ is

$$\check{d}_{\mathrm{k}} = \tilde{d}_{\mathrm{k}} - \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{i}}\| - \|\boldsymbol{r}_{\mathrm{u}} - \boldsymbol{r}_{\mathrm{j}}\|, \tag{B.11}$$

where i and j denote the position of $+1$ and -1 in the k-th line of $\boldsymbol{D}$, respectively. The symbol $\boldsymbol{r}_{\mathrm{u}}$ denotes the 3-dimensional user position vector and $\boldsymbol{r}_{\mathrm{i}}$, $\boldsymbol{r}_{\mathrm{j}}$ are the 3-dimensional position vectors of the respective anchors. For convenience, the vectors are expressed in a local, ENU frame.

Similarly, the k-th line of the Jacobian matrix that is required for obtaining the correction of the estimated position vector follows (4.29) and thus

$$\boldsymbol{J}_{\mathrm{k}} = \mathbb{1}_{\mathrm{u,i}} - \mathbb{1}_{\mathrm{u,j}}. \tag{B.12}$$

By means of vertical concatenation of $n-1$ lines, it is possible to obtain the whole $\boldsymbol{J}$ matrix.

Since the tag in our scenario is expected to be located in approximately constant height, it is convenient to apply the according constraint. This kind of constraint is rather useful in the typical UWB-based localization scenario due to geometrical reasons, which will be further commented in the following section. Consistently with the pseudomeasurement approach from Section 4.1.7 and denoting the assumed tag height h and its variance σ_h^2, we may write the extension of the equation of the constraint as

$$h = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \boldsymbol{r}_{\mathrm{u}}. \tag{B.13}$$

Such linear constraint yields following expressions for the residual evaluation and Jacobian matrix extension:

$$\check{h} = h - \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \tilde{\boldsymbol{r}}_{\mathrm{u}} \tag{B.14}$$

$$\boldsymbol{J}_h = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}. \tag{B.15}$$

It is worth remarking that if non-ENU frame was used, the $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$ expression would be replaced by the transpose of the horizontal-plane normal vector in that particular frame.

Now, we have all the information available to write the optimally weighted Levenberg-Marquardt correction update equations, according to (4.17). One of the following equations is used, depending on whether the constraint is applied or not.

Unconstrained:
$$\check{\boldsymbol{r}}_{\mathrm{u}} = \left(\boldsymbol{J}^{\mathrm{T}} \boldsymbol{\Phi}_{\boldsymbol{d}}^{-1} \boldsymbol{J} + \lambda \operatorname{Diag}(\boldsymbol{J}^{\mathrm{T}} \boldsymbol{\Phi}_{\boldsymbol{d}}^{-1} \boldsymbol{J})\right)^{-1} \boldsymbol{J}^{\mathrm{T}} \boldsymbol{\Phi}_{\boldsymbol{d}}^{-1} \check{\boldsymbol{d}} \tag{B.16}$$

Constrained:
$$\check{\boldsymbol{r}}_{\mathrm{u}} = \left(\begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{J}_h \end{bmatrix}^{\mathrm{T}} \boldsymbol{W}' \begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{J}_h \end{bmatrix} + \lambda \operatorname{Diag} \left(\begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{J}_h \end{bmatrix}^{\mathrm{T}} \boldsymbol{W}' \begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{J}_h \end{bmatrix} \right) \right)^{-1} \begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{J}_h \end{bmatrix}^{\mathrm{T}} \boldsymbol{W}' \begin{bmatrix} \check{\boldsymbol{d}} \\ \check{h} \end{bmatrix} \tag{B.17}$$

$$\boldsymbol{W}' = \begin{bmatrix} \boldsymbol{\Phi}_{\boldsymbol{d}}^{-1} & \boldsymbol{0} \\ \boldsymbol{0}^{\mathrm{T}} & \sigma_h^{-2} \end{bmatrix} \tag{B.18}$$

The update of the user position prediction $\boldsymbol{r}_{\mathrm{u}} = \tilde{\boldsymbol{r}}_{\mathrm{u}} + \check{\boldsymbol{r}}_{\mathrm{u}}$ is the same for both, constrained and unconstrained version. It should be remarked that λ was set according to the heuristic method described in Section 4.1.3; it was initialized at value $\lambda = 5$, and coefficients $c_\lambda^- = c_\lambda^+ = 5$. Typically, the algorithm converges within a few iterations, until a stop condition is reached.

From (4.38) it is apparent that the covariance matrix of the final estimate is

Unconstrained:
$$\boldsymbol{\Phi}_{\boldsymbol{r}\mathrm{u}} = \left(\boldsymbol{J}^{\mathrm{T}} \boldsymbol{\Phi}_{\boldsymbol{d}}^{-1} \boldsymbol{J}\right)^{-1} \tag{B.19}$$

Constrained:
$$\boldsymbol{\Phi}_{\boldsymbol{r}\mathrm{u}} = \left(\begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{J}_h \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{\Phi}_{\boldsymbol{d}}^{-1} & \boldsymbol{0} \\ \boldsymbol{0}^{\mathrm{T}} & \sigma_h^{-2} \end{bmatrix} \begin{bmatrix} \boldsymbol{J} \\ \boldsymbol{J}_h \end{bmatrix} \right)^{-1}, \tag{B.20}$$

where the Jacobian matrix is obtained at the point of the final estimate.

B.2.3. Experimental results

As an example we will present results of TDoA positioning that were partially published in [5]. The raw measurements were obtained with a PeLoc device at the UWB measurement polygon of the company RCD Radiokomunikace. Four anchors were available for the measurement, their geometry is captured in Figures B.3 and B.4; they were all mounted in the height 280 cm above ground. Note, that the z-axis is vertical and zero-value corresponds to the ground level. The other two axes were aligned with the walls of the room, which are depicted in the figures by the heavy gray lines. The positions of anchors are marked by the triangles. The test run was 85 seconds

long, the algorithm described in Appendix C.1 was used to synchronize the constellation of the anchors. During the test the tag was moved diagonally across the room, then along two walls. The tag was stationary in the last 25 seconds.

By means of the following figures, we will provide comparison of the unconstrained and constrained solution. The Figure B.3 presents the estimated positions in the horizontal plane, obtained by means of the unconstrained Levenberg-Marquardt algorithm, respectively. In Figure B.3 the identical data was processed by means of the constrained algorithm – the height (z-position) of the tag was softly constrained to 120 cm above ground, the standard deviation of the constraint was set to 30 cm; in other words, the tag is located between 60 and 180 cm with 95 % probability.

The time series of the position estimates in all three dimensions are available in Figure B.5 and Figure B.6 for the unconstrained and constrained solution, respectively. Note that the last 15 seconds were omitted for the sake of readability. The error bars in the plot depict 1-σ confidence intervals of the position estimate.

Such constraint solves a couple of problems. By its nature, it improves the accuracy of the solution in the vertical direction, which is poor otherwise due to unfavorably high VDOP. Inherently, it deals with the planar symmetry of the whole problem. In the unconstrained case there are two equivalent solutions above and below the plane of anchors; in Fig. B.5 it is noticeable that in several epochs the solution was found above the plane. It is worth remarking that the symmetry may be broken by introduction of another anchor that is placed substantially out of the plane.

By augmenting the vertical position, the accuracy of the horizontal position estimates is improved as well. The most apparent improvement may be found in the "corners" of the constellation, where the DOP is poor in general.

Moreover, the pseudomeasurement improved the solution success rate. It is apparent from comparison of the solutions that in several epochs the unconstrained algorithm failed to find a valid solution.[119]

In the test it has been shown that a TDoA-UWB system is capable of achieving decimeter level accuracy. It is necessary to remark that the positioning accuracy depends on the quality of time of reception measurements, on the constellation-tag geometry (the DOP coefficients) and the quality of synchronization of the anchor network. Exploitation of constraints can lead to significant improvement of the positioning performance in various aspects. The pseudomeasurement implementation of the soft constraints allows the assumption of constraint uncertainty.

[119] The estimate was discarded when considered obviously wrong (the absolute value of any position vector component was more than 10^4 m), or the iterations resulted in not-a-number value.

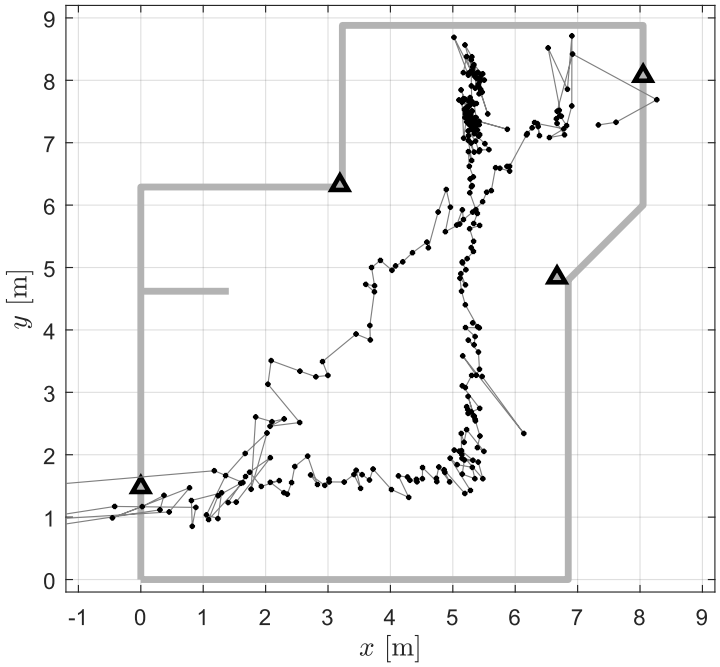


FIGURE B.3.: Unconstrained TDoA-UWB position estimates in horizontal plane.

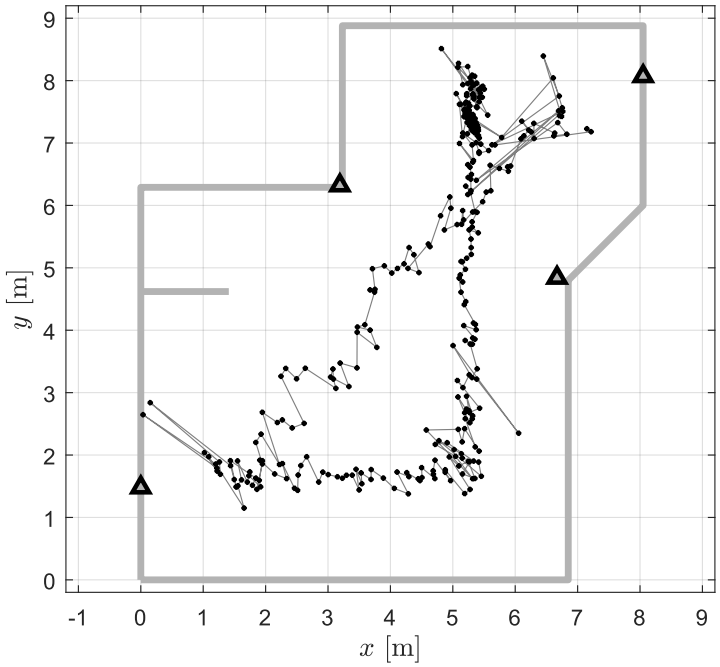


FIGURE B.4.: Constrained TDoA-UWB position estimates in horizontal plane.

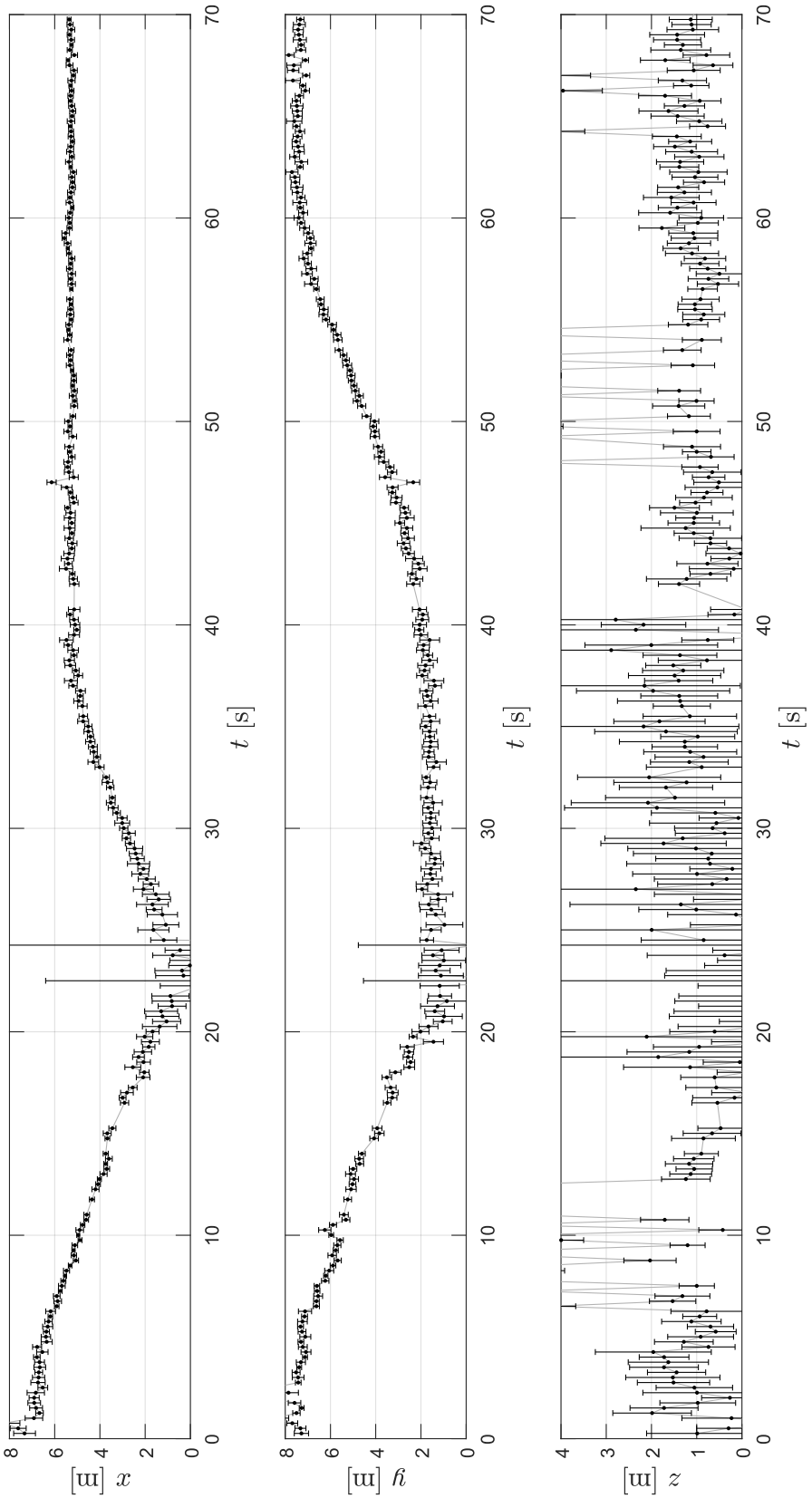


FIGURE B.5.: Time series of unconstrained TDoA-UWB position estimate with 1-σ error bars. Only first 70 seconds are shown.

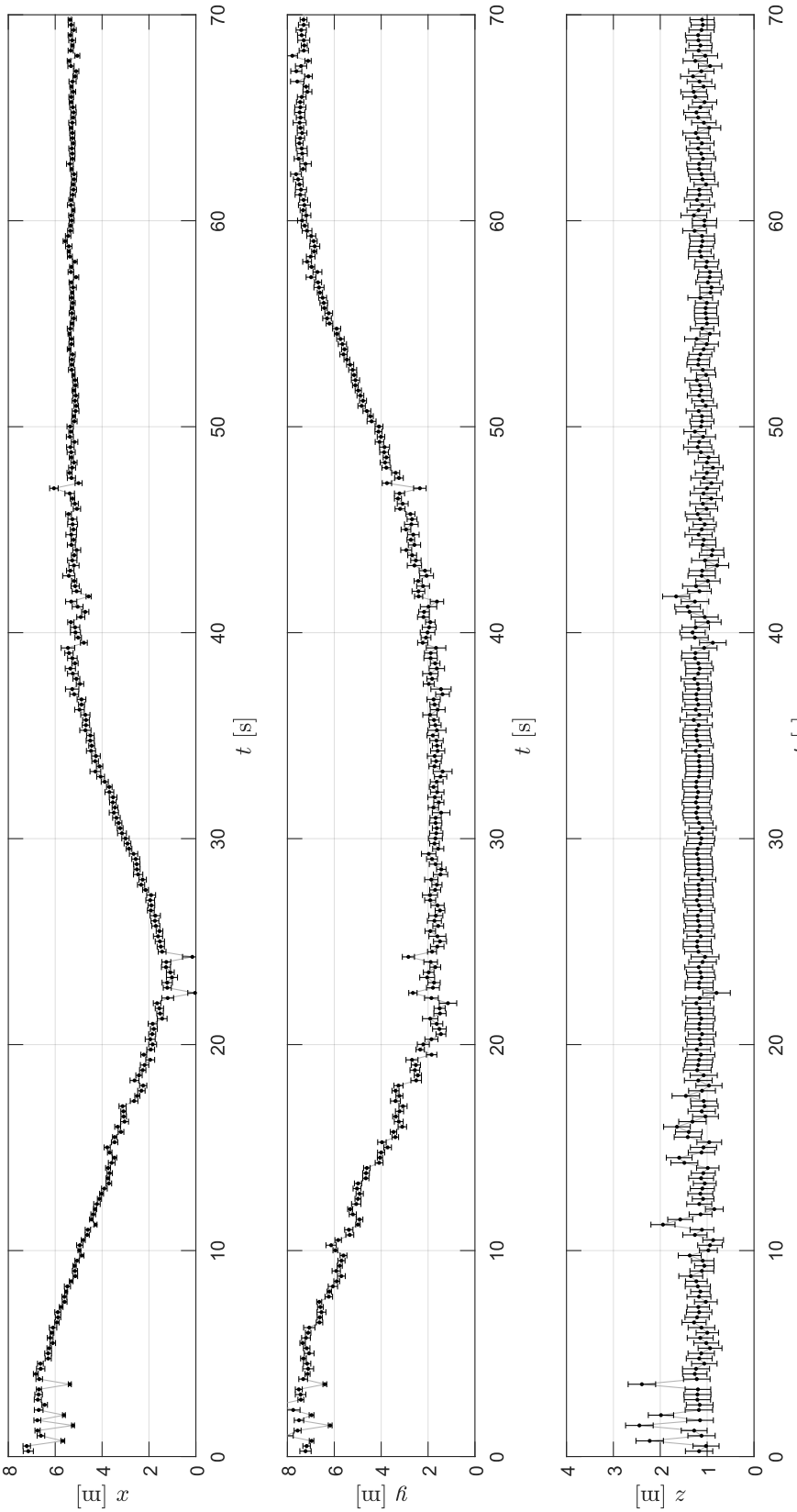


FIGURE B.6.: Time series of constrained TDoA-UWB position estimate with 1-σ error bars. Only first 70 seconds are shown.

C. Dual Application of Linear KF in UWB systems

This Appendix documents the exploitation of relatively simple linear Kalman filters by an UWB-TDoA positioning system. The system consists of PeLoc devices, which were developed in cooperation with company RCD Radiokomunikace. First use of Kalman filter is in the wireless synchronization algorithm, that has to achieve sub-nanosecond accuracy; this task can be understood as a pre-requisity to the TDoA position estimation, which is described in Appendix B.2. The second use is in filtering of the positioning estimates that were obtained by the algorithm described in Appendix B.2.

C.1. Synchronization

It has been mentioned several times that the infrastructure nodes of an UWB positioning network (the anchors) have to share a common timescale in order to make the ToA or TDoA measurements meaningful. It is rather inconvenient to provide a time reference signal to all the anchors via a wired connection. Most probably, a dedicated coaxial cable would be required in order to achieve the desired accuracy. Fortunately, it is possible to use the precise timestamping capability of the UWB transceivers to perform the synchronization in a wireless manner over a line-of-sight link.

Several wireless synchronization algorithms for UWB systems were compared in [103]; the Kalman-filter solution appears to be superior to the others. However, during experimental evaluation it became apparent that the algorithm described in [103] suffers from excessive error when the oscillator frequency is drifting. Therefore, an improved algorithm for estimating the clock state was developed by the author and presented in [5]. The advantage of this algorithm is that its synchronization segments can be chained in order to disseminate the precise timing information to the anchors that do not have direct line-of-sight access to the anchor with the master clock (this feature was experimentally verified in [6]).

C.1.1. Two-node synchronization

We will begin with a simple synchronization of two nodes; the one-way synchronization was chosen in order to minimize the time required for synchronization of the network. This can be done since the anchors are stationary and their position is pre-surveyed; thus the propagation delay is known and can be compensated. The equipment delays of the transceivers are known as well. In each anchor, there is a free running reference clock with approximately 15 ps resolution.

One of the nodes is considered to be the *master* – its clock defines the master (system) timescale; the other node is referenced as the *slave*. The values related to the

master and slave timescale (or the node itself) are denoted by the [M] or [S] superscript. The synchronization is performed by means of sending a message from master to slave. The reception time stamp at the slaves receiver $t_{Rx}^{[\mathrm{S}]}$ can be described by

$$t_{Rx}^{[\mathrm{S}]} = t_{Tx}^{[\mathrm{M}]} + \tau_{Tx}^{[\mathrm{M}]} + \tau_{Rx}^{[\mathrm{S}]} + \Delta t_{\mathrm{MS}} + \frac{r_{\mathrm{MS}}}{c_0}, \tag{C.1}$$

where $t_{Tx}^{[\mathrm{M}]}$ denotes transmission time stamp in the master timescale – this information is included in the message and therefore available in the slave node. Symbol r_{MS} denotes the geometric distance between master and slave, $\tau_{Tx}^{[\mathrm{M}]}$ and $\tau_{Rx}^{[\mathrm{S}]}$ are the respective equipment delays at master transmitter and slave receiver.

The offset between master and slave timescale is denoted by Δt_{MS}. For ideal clocks this value would be constant, however, in reality the clock frequency is different from the nominal, therefore the clock offset changes in time. According to [19], the clock frequency error should not be worse than $\pm 20\,\mathrm{ppm}$.

Therefore, the synchronization messages have to be sent periodically and the temporal development of Δt_{MS} has to be estimated. Consider now that the state of a linear Kalman filter vector $\boldsymbol{x} = [x_0\ x_1\ x_2]$ corresponds to the clock offset at the particular epoch Δt_{MS}, the clock drift $\Delta \dot{t}^{\mathrm{MS}}$ and the clock drift rate $\Delta \ddot{t}^{\mathrm{MS}}$, respectively. It is possible to adopt a model which is based on a second-order approximation of the clock offset

$$\boldsymbol{x}[k] \approx \begin{bmatrix} x_0[k-1] + \int_{t[k-1]}^{t[k]} x_1 + \int x_2\,\mathrm{d}t\,\mathrm{d}t + \int_{t[k-1]}^{t[k]} \mathcal{N}(0, \sigma_{x0}^2)\,\mathrm{d}t \\ \int_{t[k-1]}^{t[k]} x_2\,\mathrm{d}t + \int_{t[k-1]}^{t[k]} \mathcal{N}(0, \sigma_{x1}^2)\,\mathrm{d}t \\ x_2 + \int_{t[k-1]}^{t[k]} \mathcal{N}(0, \sigma_{x2}^2)\,\mathrm{d}t \end{bmatrix}, \tag{C.2}$$

where $t[k-1]$ and $t[k]$ are the times of the respective epochs. In the linear KF notation (4.50), (4.51) the prediction step holds

$$\boldsymbol{x}^-[k] = \boldsymbol{F}\boldsymbol{x}^+[k-1] \tag{4.50}$$

$$\boldsymbol{P}^-[k] = \boldsymbol{F}\boldsymbol{P}^+[k-1]\boldsymbol{F}^{\mathrm{T}} + \boldsymbol{Q} \tag{4.51}$$

$$\boldsymbol{F} = \begin{bmatrix} 1 & T[k-1] & \frac{1}{2}T[k-1]^2 \\ 0 & 1 & T[k-1] \\ 0 & 0 & 1 \end{bmatrix} \tag{C.3}$$

$$\boldsymbol{Q} = \begin{bmatrix} T[k-1]\,\sigma_{x0}^2 & 0 & 0 \\ 0 & T[k-1]\,\sigma_{x1}^2 & 0 \\ 0 & 0 & T[k-1]\,\sigma_{x2}^2 \end{bmatrix}, \tag{C.4}$$

where $T[k-1] = t[k] - t[k-1]$ and σ_{x0}^2 to σ_{x2}^2 are the state random walk variances normalized w.r.t. a unit of time.[120]

The measurement model is rather simple, since the first element of the state vector is the actual measurement; thus

$$H = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}. \tag{C.5}$$

[120] From a practical point of view, it does not matter whether T is computed as a difference of the reception or transmission timestamps.

The measurement itself can be derived from (C.1) as

$$\boldsymbol{y}[k] = \Delta t_{\mathrm{MS}} = t_{Rx}^{[\mathrm{S}]} - t_{Tx}^{[\mathrm{M}]} - \tau_{Tx}^{[\mathrm{M}]} - \tau_{Rx}^{[\mathrm{S}]} - \frac{r_{\mathrm{MS}}}{c_0}. \tag{C.6}$$

Considering that the equipment delays $\tau_{Tx}^{[\mathrm{M}]}$, $\tau_{Rx}^{[\mathrm{S}]}$ and the geometric range r_{MS} are known perfectly, the covariance of the measurement is a simple sum of the transmission and reception timestamp variances.

$$\boldsymbol{R} = \sigma_{\mathrm{Tx}}^2 + \sigma_{\mathrm{Rx}}^2 \tag{C.7}$$

The measurement update is described by equations (4.52) to (4.56).

The approach described above is able to provide us an estimate of the clock offset Δt_{MS} at any epoch as the first state value $x_0[k]$. Its variance is the first element of the $\boldsymbol{P}[k]$ matrix. Nonetheless, we are interested in the Δt_{MS} value at the time of reception of a bling from a tag, in order to be able to express it in the master timescale.

It is possible to adopt the first line of the model (C.2), i.e. the first line of the $\boldsymbol{F}$-matrix. Denoting the blink time t_b and using $t[k]$ for the time of last available epoch with measurement update already performed, we may obtain the

$$\Delta t_{\mathrm{MS}}(t_b) \approx x_0[k] + \int_{t[k]}^{t_b} x_1 + \int x_2 \,\mathrm{d}t \,\mathrm{d}t \tag{C.8}$$

$$\Delta t_{\mathrm{MS}}(t_b) \approx \begin{bmatrix} 1 & t_b - t[k] & \frac{1}{2}(t_b - t[k])^2 \end{bmatrix} \boldsymbol{x}[k]. \tag{C.9}$$

This is rather similar to the prediction step of the Kalman filter; we are indeed predicting the first element of the state vector into an arbitrary moment t_b . The variance of this prediction is obtained by means of

$$\begin{aligned} \operatorname{var} \Delta t_{\mathrm{MS}}(t_b) &\approx \boldsymbol{F}_1 \boldsymbol{P}[k] \boldsymbol{F}_1^{\mathrm{T}} + (t_b - t[k])\sigma_{x0}^2 \\ \boldsymbol{F}_1 &= \begin{bmatrix} 1 & t_b - t[k] & \frac{1}{2}(t_b - t[k])^2 \end{bmatrix}. \end{aligned} \tag{C.10}$$

Note that $\boldsymbol{F}_1$ is the equivalent of the first line of the state-transition matrix $\boldsymbol{F}$; the time interval between epoch is of course substitute by the interval from the last available epoch (i.e. the time of last synchronization message reception) to the time of blink reception.

The blink reception time is converted from the slave timescale to the master timescale by means of the following equation. Its variance is modified accordingly.

$$t_b^{[M]} = t_b^{[S]} - \Delta t_{\mathrm{MS}}(t_b) \tag{C.11}$$

$$\operatorname{var} t_b^{[M]} = \operatorname{var} t_b^{[S]} + \operatorname{var} \Delta t_{\mathrm{MS}}(t_b) \tag{C.12}$$

The exploitation of Kalman filter is advantageous over the linear interpolation or feedback control loop (PI, PII, PID) approaches described in [103], since it provides the variance of the clock offset inherently – an information about the quality of the synchronization. In contrast to the poor-performing linear interpolation it is able to predict the clock state, the synchronization message received after the bling is not

Table C.1.: Parameters of the proposed three-state synchronization KF.

	Parameter	Unit	Value	St. Dev.
σ^2_{Tx}	Transmission timestamp variance	[s]	$1.5 \cdot 10^{-20}$	$\sim 120\,\mathrm{ps}$
σ^2_{Rx}	Reception timestamp variance	[s]	$1.5 \cdot 10^{-20}$	$\sim 120\,\mathrm{ps}$
σ^2_{x0}	Clk. offset proc. noise norm. var.	[s^2/s]	$5 \cdot 10^{-20}$	$\sim 0.2\,\mathrm{ps}/\sqrt{\mathrm{s}}$
σ^2_{x1}	Clk. drift proc. noise norm. var.	[1/s]	$5 \cdot 10^{-18}$	$\sim 2\,\mathrm{ppb}/\sqrt{\mathrm{s}}$
σ^2_{x2}	Clk. drift rate proc. noise norm. var.	[1/s^3]	$5 \cdot 10^{-18}$	$\sim 2\,(\mathrm{ppb/s})/\sqrt{\mathrm{s}}$
T	Synchronization message period	[s]	~ 0.4	

necessary; consequently, there is no significant lag in obtaining the blink reception time in the master timescale.

Additionally, it is possible to further disseminate the synchronization messages with the transmission timestamps expressed in the master timescale, which is advantageous for areas with multiple line-of-sight obstructions. The synchronization message forwarding will be described in Section C.1.2 below.

Last but not least, it is possible to tune this filter in a fairly straightforward manner.

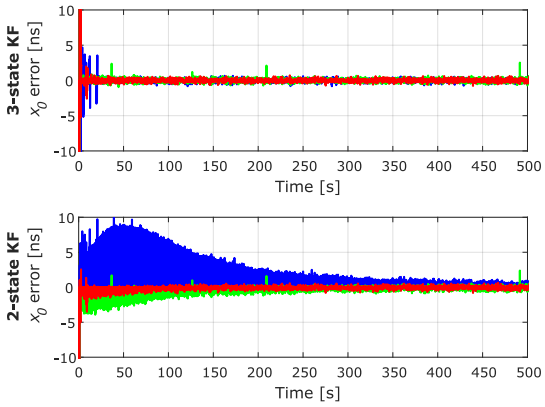
Experimental results

For the purpose of demonstration of the synchronization algorithm capabilities, we will use a dataset already presented in [125]. Nonetheless, numerous laboratory test runs and in-field tests were already successfully performed, see e.g. [104].

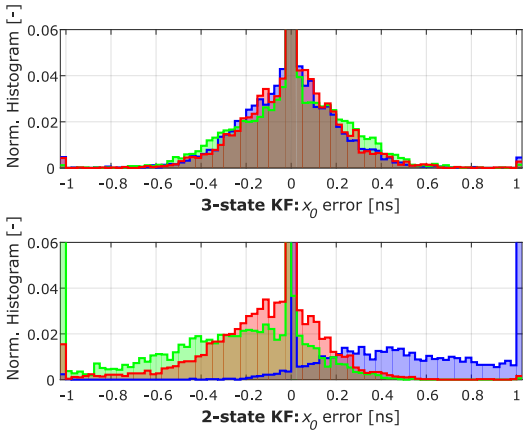
The test run is approximately 500 seconds long; three slave anchors are synchronized to a single master. Even though the anchor hardware is identical, their clock state evolution is various.[121] In the following text, the anchors and the respective clocks are denoted by the colors that are used to plot the particular data in the figures with graphical results. The synchronization message was sent by the master every 100 milliseconds, however, the filter measurement update was performed on every fourth message. The complete set of messages was used to evaluate the errors, since there was no "ground truth" available.

In the following paragraphs, we will compare the performance of the proposed three-state Kalman filter (tracking clock, offset, drift, and drift rate) to the two-state Kalman filter (tracking only clock offset and drift) described in [103]. The filter tuning was performed on the basis of available information about the clock and timestamping performance and was slightly adjusted during several trials. The parameters are summarized in Tab. C.1. The variances from the table can be interpreted with ease through the related standard deviations, which are included in the last column. Note that the synchronization message period T is only approximate, the exact value is always computed from the difference of the respective transmission timestamps. Mostly, it oscillates around the nominal value, however, it can be multiplied when some synchro-

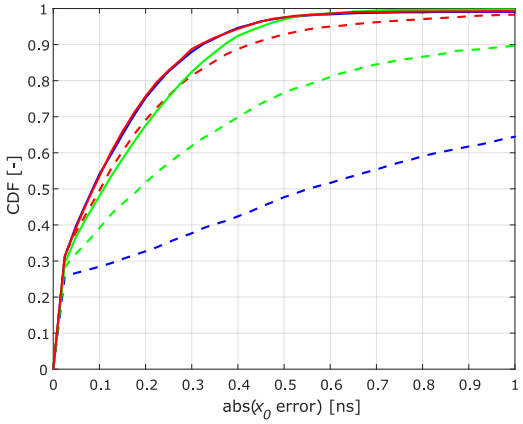
[121] Obviously, this was the reason for choosing the particular dataset for demonstration.



(A) Error in time



(B) Histograms



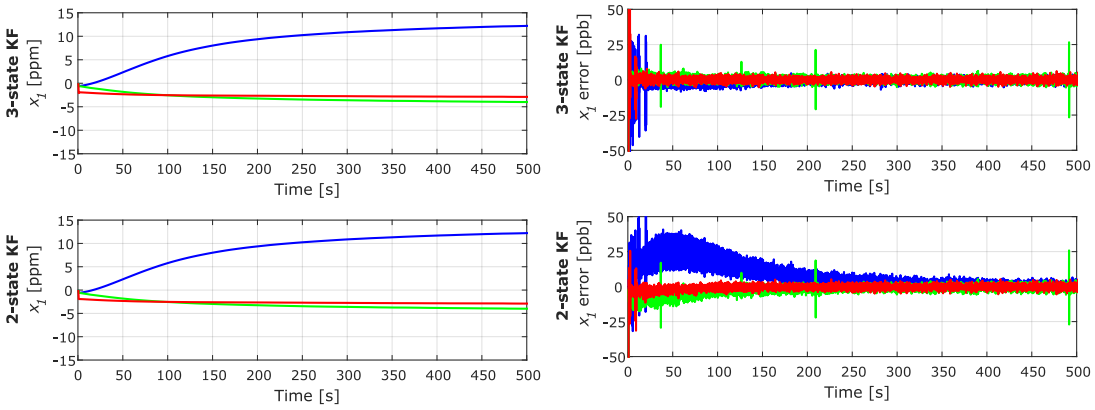
(C) CDF (three-state KF solid; two-state KF dashed).

FIGURE C.1.: Comparison of clock offset (x_0) estimate errors.

nization messages are lost. The state vector was initialized with zero states, the state covariance matrix was set to identity matrix.

The time development of the clock offset estimate error (error of the state x_0) provided by the three-state and two-state KFs is depicted in Figure C.1a. It is apparent that in case of the red anchor, the KF converged to a very accurate estimate almost instantly. The two-state Kalman filter struggles with the estimate of the green, and especially the blue anchor clock offset. It is because it inherently assumes the clock drift rate to be zero, whilst the three-state KF estimates the drift rate. The clock offset error magnitude is more than several nanoseconds for a couple of minutes. It needs to be remarked that such error would cause the pseudoranges (and consequently the TDoA measurements) to be biased by several meters, which is likely to corrupt the indoor localization completely.

The clock drift estimates (state x_1) from the two compared KFs occur rather similar in Figure C.2a. It is obvious that the clock drift of the red anchor does not change, while the green drift changes very slowly and the clock drift of the blue anchor grows



(A) Estimates (B) Errors

FIGURE C.2.: Comparison of clock drift (x_1) estimates and respective errors.

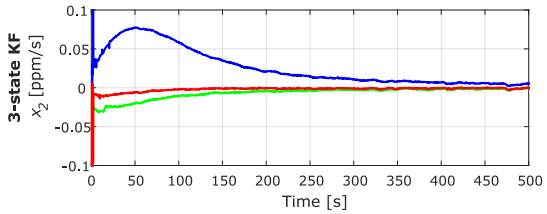


FIGURE C.3.: Clock drift rate (x_2) estimated by three-state KF.

rather rapidly. When the estimate errors are examined (see Figure C.2b), it shows up that the two-state KF estimate exhibits more than 20 ppb error. Although the error seems to be very small at a first glance, it causes the unacceptable clock offset estimate. It is worth remarking that the algorithm needs to keep the synchronization accuracy below sub-nanosecond level with the 400 ms measurement-update period; the required accuracy is at the part-per-billion level.

The clock drift rate estimated by the proposed three-state KF is depicted in Figure C.3. The shape estimated curve is very similar to the shape of the clock offset error envelope of the two-state KF in the bottom part of Figure C.1a.

The substantial difference in the quality of the clock offset estimates provided by the two filters can be observed on the histograms of the estimate errors (Fig. C.1b); the two-state estimates of the green and blue anchors clock offsets are severely biased. The outstanding performance of the three-state KF is documented by the cumulative distribution function (CDF) of the clock offset error (the solid lines in Fig. C.1c). Even with rapidly drifting clock it is able to achieve error below 500 ps with 95 % probability. The performance of the two-state approach is noticeably worse even for the red anchor with the stable clock.

The quantitative representation of the improvement brought by estimating the clock drift rate is available in table C.2. Therein, the RMS value of the clock offset error is employed as a metric. Note that the corresponding range errors are provided as

Table C.2.: RMS values of the clock offset error and their comparison.

Anchor	Two-state KF		Three-state KF		Improvement		
	[ps]	[cm]	[ps]	[cm]	[ps]	[cm]	[%]
Blue	2302	69.0	358	10.7	1944	58.3	84.4
Green	687	20.6	221	6.6	466	14.0	67.8
Red	306	9.2	329	9.9	−23	−0.7	−7.5
Total (RMS)	1398	41.9	308	9.2	1090	32.7	77.9

well. The greatest improvement was achieved at the blue anchor, where the error was suppressed by more than 84 %. The overall improvement (RMS computed over the three anchors) is almost 78 %.

Clearly, the proposed three-state KF approach promises reliable and accurate estimate of the clock offset. Thanks to the estimation of the clock drift rate, the algorithm is able to operate even when the clock frequency is unstable, e.g. during the warm-up period after anchor power-on. It also provides the information about the current accuracy and does not need to wait for reception of the next synchronization message.

C.1.2. Chained synchronization

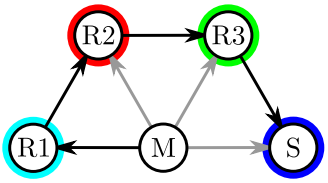
The node-to-node synchronization requires a line-of-sight link in order to be able to compensate for the propagation delay accurately and reliably; the reliability of the wireless link is required as well. Therefore, only the anchors with the master node "in sight" have the estimate of the master timescale available.

For such anchor, however, it is possible to update and forward the synchronization message. Such relayed synchronization message includes the local transmission timestamp corrected to the master timescale, the accuracy of the time information, and indication that it is a relayed information. The next slave node can then process the relayed messages in the same way as it would process the direct ones from the master node. For convenience, we will denote the forwarding anchor as the *relay*.

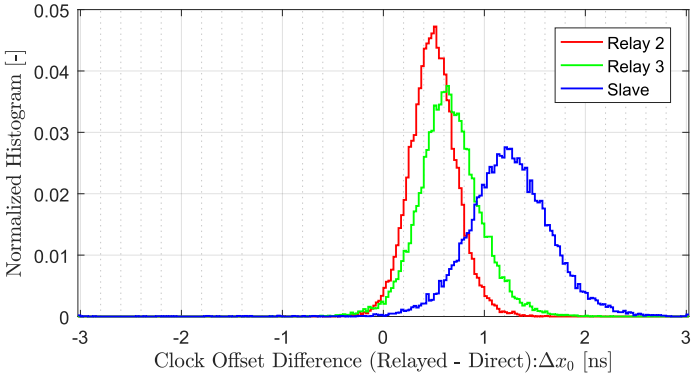
Such approach allows master time dissemination to the nodes that do not have direct line-of-sight to the master node available. The chain may consist of multiple hops (node-to-node segments) organized in a couple of branches. The branches are allowed to split and merge, however, they should not form closed loops, i.e. no timing information feedback from the end of the chain to its beginning should occur.

Experimental results

An experimental evaluation of the chained synchronization in UWB network was experimentally tested and published in [6]. Unsurprisingly, the accuracy of the time information deteriorates with the quality of received messages and number of hops. Yet, simulations [5] and experiments [6] have shown that the degradation may be reduced by providing multiple paths of the timing information – the slave node receives messages from two or more relays and performs the KF measurement update for each of them. The incoming information can be weighted by means of setting the $\boldsymbol{R}$ matrix.

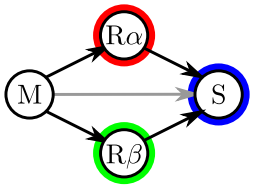


(A) Constellation

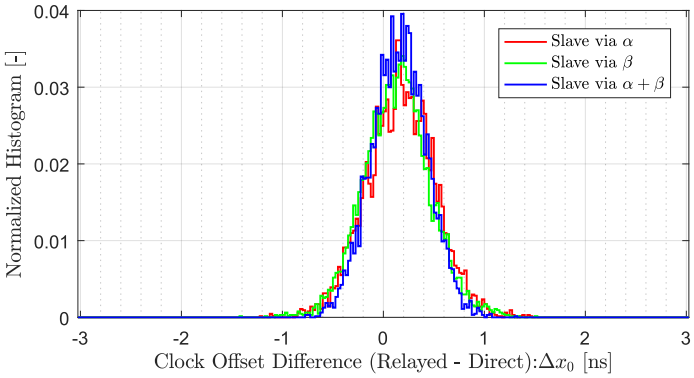


(B) Histograms of clock offset differences

FIGURE C.4.: Serial multi-hop test.



(A) Constellation



(B) Histograms of clock offset differences

FIGURE C.5.: Parallel branches test.

Two different tests are presented herein; in both the direct line-of-sight link to the master anchor is available and serves as the reference. Hence, two KF synchronization algorithms are running in each node. The first one processes the relayed messages and the other one uses only the direct messages. The performance of the chained synchronization is assessed by means of comparison of the outputs of these two filters.

In the first test, the relays and slave node were located around the receiver, as depicted in Figure C.4a. The anchors were equally spaced, all node-to-node distances were 30 cm long. The direct, reference, synchronization links are illustrated by the gray arrows, whilst the black arrows show the relayed-message path. The histograms of the clock offset error (w.r.t. the direct sync. path) are available in Figure C.4b; the colors of the lines correspond to the colors used for anchors in Fig. C.4a. It is apparent that the bias and variance of the clock offset estimate grow with increasing number of hops – the histogram peaks are offset from the center and wider.

The more detailed quantitative results of the test are available in Table C.3. The standard deviation of the clock offset is below half-nanosecond level even after four hops. The sub-nanosecond biases from each hop have accumulated into a 1.2 ns bias at

the slave node. It is believed, however, that this value can be further reduced by means of more careful survey and compensation of the propagation and equipment delays, or by the on-the-fly survey of the propagation delay through a TWR measurement. The errors in the estimation of the clock drift and its rate are negligible.

In the second test it was shown that it is possible to use multiple paths for the synchronization information. The messages from master were relayed by anchors α and β, the constellation is presented in Figure C.4a. Four KFs were running in the slave node:

- The reference KF, using only the direct messages from master;
- KF using only messages from relay α;
- KF using only messages from relay β;
- KF using messages from both relays.

The histograms of the clock offset error (Figure C.5b) are very similar for all three relayed KFs. Nevertheless, the quantitative results from Table C.4 prove that the KF that uses both available paths features lower standard deviation than both single-path filters. Moreover, the bias is approximately averaged when using two independent paths. As well as in the case of the first test, the clock drift and its rate are estimated almost perfectly by means of the chained Kalman filters.

C.1.3. Remarks

The synchronization algorithm that was developed by the author is able to provide reliable and accurate timing information for the anchors of an UWB network. It is required for all the synchronized anchors to have link to the master anchor via one or more line-of-sight segments. Under most circumstances, the sub-nanosecond accuracy is achieved with a considerable margin. Thanks to the synchronization, it is possible to provide inputs for the TDoA positioning algorithm, which was described in Appendix B.2.

There are a few aspects that were not covered in this text, however, may be important for the actual implementation. For the purpose of synchronization, the equipment delays were assumed to be known exactly for the transmission and reception operation. Nonetheless, for production devices it is only possible to measure an aggregate delay of an anchor, i.e. the sum of the transmission and reception equipment delays. Typically, the division of the aggregate delay between the two operations is based on the information from the device manufacturer.

The analysis of the systematic errors, which was presented in author's work [6], shows that only the aggregate delay does matter when measuring the TDoA. The transmission and reception delays affect the blink receptions in the very same way as they affect the synchronization messages. Therefore, they either subtract (e.g. the reception delay in the slave node) or they sum up to the aggregate delays (e.g. blink reception in a master node, or message relaying). The detailed analysis is available in [6].

TABLE C.3.: Comparison of direct and relayed synchronization in serial multi-hop setup.

Anchor	**KF residuals (RMS)**				**Clock offset Δx_0**						**Freq. offset Δx_1**		**Freq. drift Δx_2**	
	direct		relayed		RMS		mean		std		RMS	mean	RMS	mean
	[ps]	[cm]	[ps]	[cm]	[ps]	[cm]	[ps]	[cm]	[ps]	[cm]	[ppb]	[ppb]	[ppb/s]	[ppb/s]
Relay 2	158	4.7	213	6.4	549	16.5	497	14.9	232	7.0	0.581	0.003	0.416	0.002
Relay 3	163	4.9	292	8.8	724	21.7	654	19.6	310	9.3	0.744	0.003	0.531	0.003
Slave	286	8.6	366	11.0	1318	39.5	1249	37.4	423	12.7	1.053	0.000	0.764	0.006

TABLE C.4.: Comparison of direct and relayed synchronization of slave node in parallel setup.

Chain	**KF residuals (RMS)**				**Clock offset Δx_0**						**Freq. offset Δx_1**		**Freq. drift Δx_2**	
Branch	direct		relayed		RMS		mean		std		RMS	mean	RMS	mean
Used	[ps]	[cm]	[ps]	[cm]	[ps]	[cm]	[ps]	[cm]	[ps]	[cm]	[ppb]	[ppb]	[ppb/s]	[ppb/s]
Only α	190	5.7	311	9.3	379	11.4	172	5.2	337	10.1	0.835	0.002	0.583	0.001
Only β	190	5.7	278	8.3	356	10.7	134	4.0	329	9.9	0.783	0.000	0.538	−0.001
Both	190	5.7	243	7.3	316	9.5	158	4.7	273	8.2	0.838	0.001	0.634	0.000

The results presented herein were obtained by means of processing the captured raw time of transmission and reception data in the Matlab environment. Thus, double precision floating point arithmetic was used. For the purpose of implementation into the anchor processor, the algorithm was converted into a fixed point arithmetic. 64-bit long variables were required, appropriate scaling of the state-vector values and the covariances was necessary. Due to advantageous properties of the synchronization KF matrices, it was possible to perform number of the KF computations in a smarter way than by strictly adhering to the general definitions, which were described in Section 4.2.1. For instance, there is no need to evaluate all multiplications by zero in every KF iteration. The symmetry of the covariance matrices is to be exploited as well; a 3-by-3 symmetric matrix has only 6 unique elements. From a numerical point of view it is convenient to adopt the Horner's method when evaluating the elements of the state prediction and its covariance – the expression can be interpreted as a polynomial with indeterminate $T[k-1]$ and coefficients determined by the state vector. The description of the optimized fixed-point KF and comparison with floating-point implementation for a particular embedded processor is available in [166].

C.2. Position Filtering

The position estimates obtained on epoch-by-epoch basis are often considered as noisy. A representative example can be found in Appendix B.2. A Kalman filter, which represents the dynamic model of the localized body, can be used to make the measurements smoother.

The data from Appendix B.2 will be processed by two different linear KFs. One would assume stationary tag, whilst the second would assume constant-velocity linear motion. Obviously, the tags trajectory is not fully consistent with either of the models; the process noise setting should reflect the fact.

Assumption: Stationary tag

In this very simple case, the Kalman filter state is the filtered position vector itself. The position is not expected to change in time, thus, the system model matrix is the identity matrix. In the equations it yields

$$\boldsymbol{x} = \hat{\boldsymbol{r}}_{\mathrm{u}} \qquad \boldsymbol{F} = \mathbb{I}_{3\times 3} \tag{C.13}$$

$$\boldsymbol{y} = \boldsymbol{r}_{\mathrm{u}} \qquad \boldsymbol{H} = \mathbb{I}_{3\times 3}. \tag{C.14}$$

The process noise covariance matrix is dependent on the time period between epochs, which is denoted by T; note that this value may not be constant due to missing measurements. Rather low values of the normalized process noise variances were set, since the tag movement is slow

$$\boldsymbol{Q} = T\,\mathrm{Diag}\begin{bmatrix}\sigma^2_{\mathrm{Q}rx}\\ \sigma^2_{\mathrm{Q}ry}\\ \sigma^2_{\mathrm{Q}rz}\end{bmatrix}; \qquad \begin{bmatrix}\sigma^2_{\mathrm{Q}rx}\\ \sigma^2_{\mathrm{Q}ry}\\ \sigma^2_{\mathrm{Q}rz}\end{bmatrix} = \begin{bmatrix}(0.2\,\mathrm{m}/\sqrt{\mathrm{s}})^2\\ (0.2\,\mathrm{m}/\sqrt{\mathrm{s}})^2\\ (0.1\,\mathrm{m}/\sqrt{\mathrm{s}})^2\end{bmatrix}. \tag{C.15}$$

The measurement covariance $\boldsymbol{R}$ is the covariance matrix of the position estimate $\boldsymbol{\Phi}_{\boldsymbol{r}\mathrm{u}}$, which can be obtained as an output of the least-squares positioning algorithm (see Appendix B.2).

The Linear Kalman filter was implemented according to Section 4.2.1 and initialized with $\boldsymbol{x}^{+}[0] = \boldsymbol{r}_{\mathrm{u}}[0]$ and $\boldsymbol{P}^{+}[0] = \boldsymbol{\Phi}_{\boldsymbol{r}\mathrm{u}}[0]$. The 0^{th} epoch measurements were not used in the first measurement update.

The filtered trajectory is depicted by the red line in Figure C.6. Obviously, the trajectory is smoother than the raw position estimates (black). In the time series visualization (Figure C.8) it is observable that the variance of the filtered position estimates is lower due to the additional knowledge provided by the system model. Note that the *a posteriori* KF state variance is depicted by the blue area and the original, raw, variance is visualized by the T-shaped bars, both are related to the right axis.

Due to the assumption of a stationary tag, the filtered trajectory estimate (red line, left axis) tends to lag behind the actual position of the tag – the filtered estimate is "slowed down" by the system model. It is worth noting that increasing $\boldsymbol{Q}$ or decreasing $\boldsymbol{R}$ would reduce this effect, and vice versa.

Assumption: Linear motion

In the second presented case, the state vector contains position and velocity estimates. The system model matrix $\boldsymbol{F}$ reflects that constant velocity is assumed (bottom-right $\mathbb{I}$); the position prediction is a sum of the current position (top-left $\mathbb{I}$) and time-integrated constant velocity ($\mathbb{I}_{3\times3}T$). It is worth reminding that T denotes the time interval between the epochs. Still, only the position vector is available as the input to the KF, no direct link of the velocity-related part of the state vector to the measurement vector is available. This is confirmed by a matrix of zeros ($\mathbb{O}$) in the right part of measurement model $\boldsymbol{H}$.

$$\boldsymbol{x} = \begin{bmatrix} \hat{\boldsymbol{r}} \\ \hat{\boldsymbol{v}} \end{bmatrix} \qquad \boldsymbol{F} = \begin{bmatrix} \mathbb{I}_{3\times3} & \mathbb{I}_{3\times3}T \\ \mathbb{O}_{3\times3} & \mathbb{I}_{3\times3} \end{bmatrix} \tag{C.16}$$

$$\boldsymbol{y} = \boldsymbol{r} \qquad \boldsymbol{H} = \begin{bmatrix} \mathbb{I}_{3\times3} & \mathbb{O}_{3\times3} \end{bmatrix} \tag{C.17}$$

The system process noise matrix was set in a similar way as in the stationary case:

$$\boldsymbol{Q} = T\,\mathrm{Diag}\begin{bmatrix} \sigma^2_{\mathrm{Q}rx} \\ \sigma^2_{\mathrm{Q}ry} \\ \sigma^2_{\mathrm{Q}rz} \\ \sigma^2_{\mathrm{Q}vx} \\ \sigma^2_{\mathrm{Q}vy} \\ \sigma^2_{\mathrm{Q}vz} \end{bmatrix}; \qquad \begin{bmatrix} \sigma^2_{\mathrm{Q}rx} \\ \sigma^2_{\mathrm{Q}ry} \\ \sigma^2_{\mathrm{Q}rz} \\ \sigma^2_{\mathrm{Q}vx} \\ \sigma^2_{\mathrm{Q}vy} \\ \sigma^2_{\mathrm{Q}vz} \end{bmatrix} = \begin{bmatrix} \left(0.05\,\mathrm{m}/\sqrt{\mathrm{s}}\right)^2 \\ \left(0.05\,\mathrm{m}/\sqrt{\mathrm{s}}\right)^2 \\ \left(0.025\,\mathrm{m}/\sqrt{\mathrm{s}}\right)^2 \\ \left(0.1\,\mathrm{ms}^{-1}/\sqrt{\mathrm{s}}\right)^2 \\ \left(0.1\,\mathrm{ms}^{-1}/\sqrt{\mathrm{s}}\right)^2 \\ \left(0.05\,\mathrm{ms}^{-1}/\sqrt{\mathrm{s}}\right)^2 \end{bmatrix}. \tag{C.18}$$

Non-zero variance is provided to both, position and velocity states. The latter indicates that the velocity vector may change, the former suggests that the true position may deviate from the linear-motion model. The implementation and initialization of the KF is almost identical to the stationary tag case; the velocity is initialized as stationary value with velocity variance $5\,\mathrm{m}^2\mathrm{s}^{-2}$.

Figure C.7 presents the filtered results in the horizontal plane, the smoothness of the trajectory is similar to the one in the previous case. Since the velocity is observed indirectly, through the state covariances, the filter tends to "overshoot" the corners, or the true value after initialization, especially when the system noise is underestimated. From the time-series representation of the results (Figure C.9) it is apparent, that the KF converges slower than in the previous case. Nonetheless, it does not suffer from the lag on the straight portions of the trajectory.

It is evident that neither of the two models fits the actual motion perfectly. However, it was shown that even rather simple KFs may be used to smoothen the position estimates. With careful tuning the artifacts appearing from the motion model unsuitability suppressed to a reasonable level.

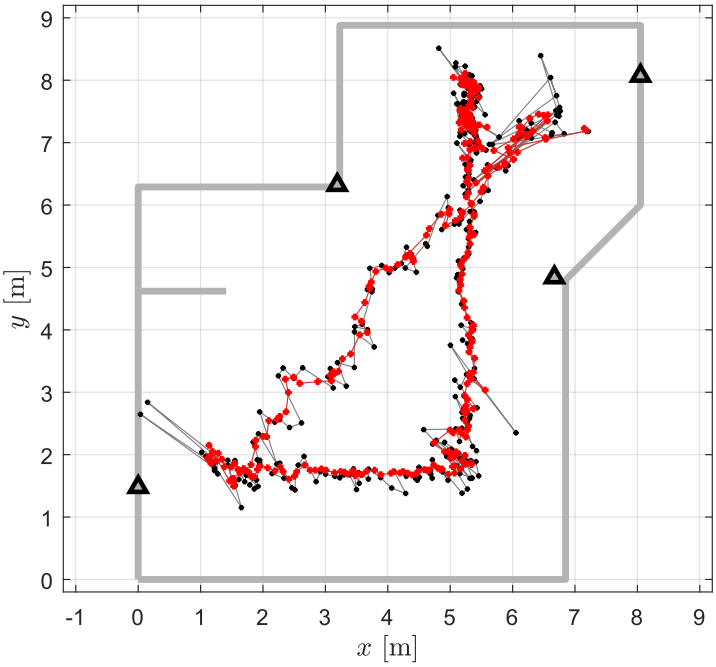


FIGURE C.6.: Filtered TDoA-UWB positions in horizontal plane – stationary model.

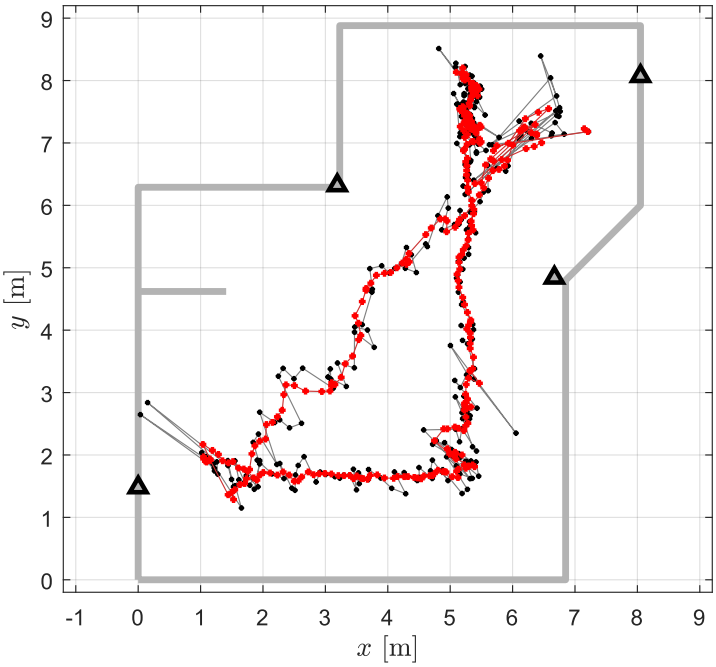


FIGURE C.7.: Filtered TDoA-UWB positions in horizontal plane – linear motion model.

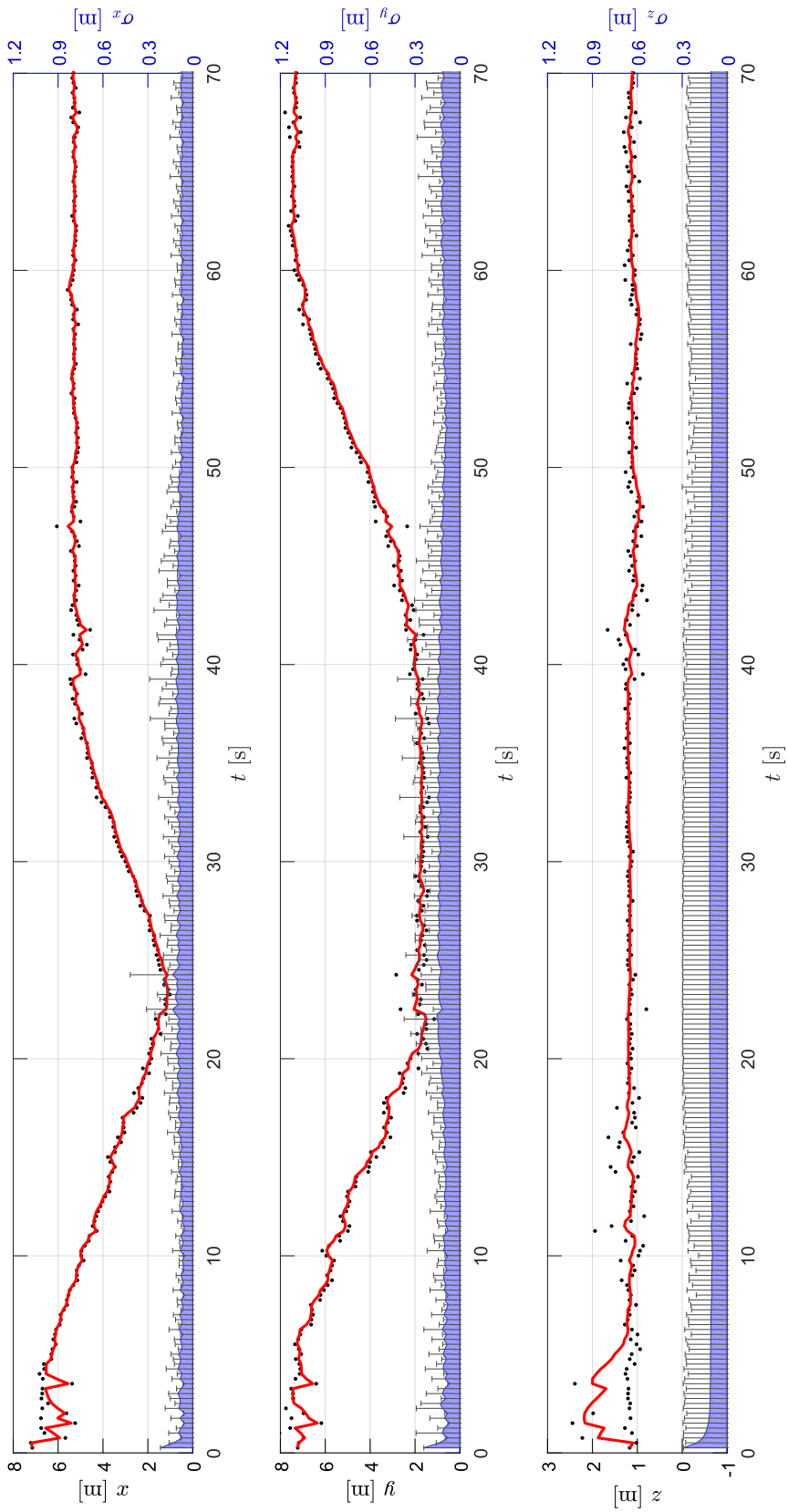


FIGURE C.8.: Time series of filtered TDoA-UWB position estimate (left axis; red line – filtered, dots – raw) and standard deviation (right axis; blue area – filtered *a posteriori*, T-bars – raw). Stationary tag assumed. Only first 70 seconds are shown.

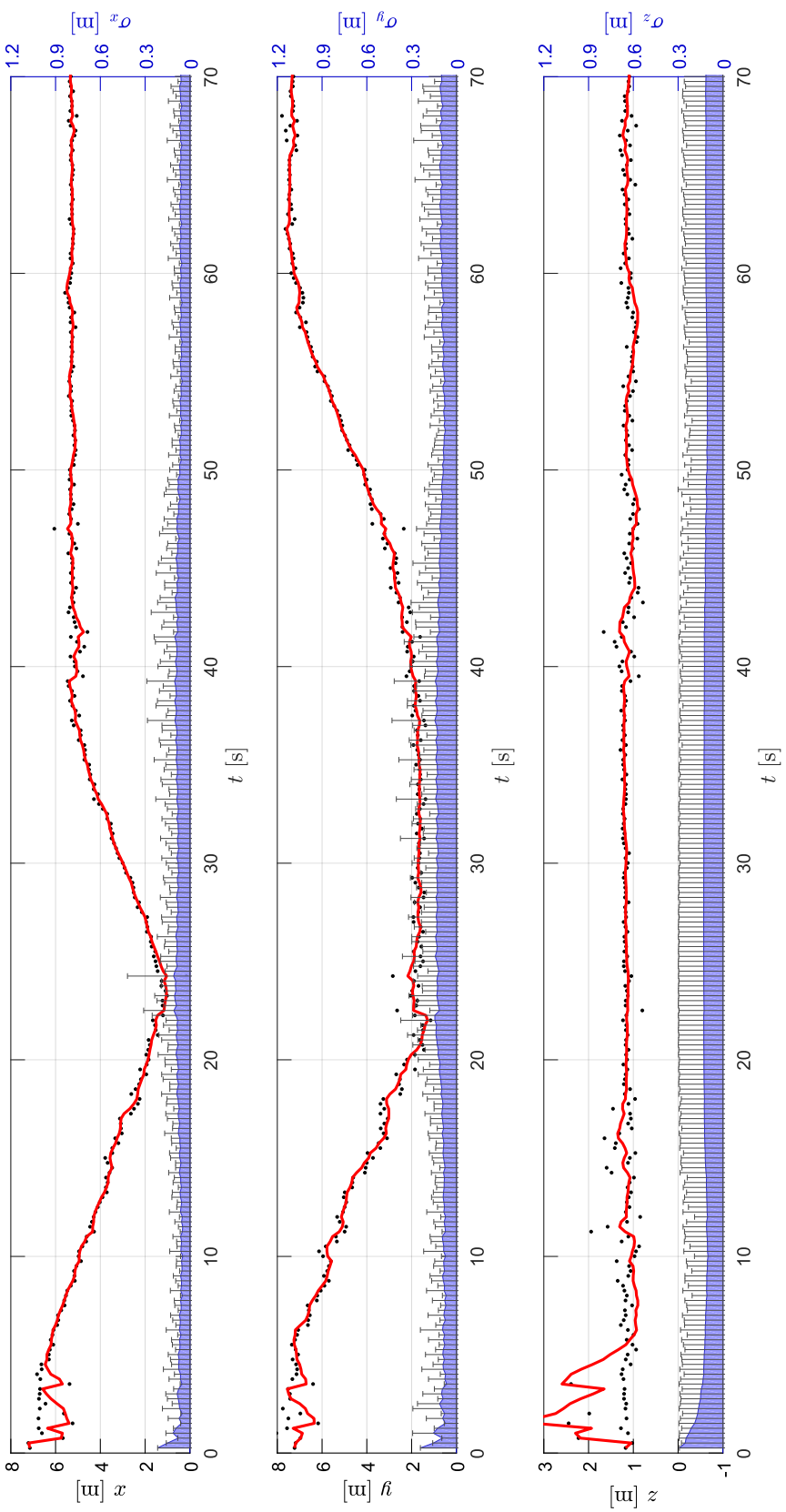


FIGURE C.9.: Time series of filtered TDoA-UWB position estimate (left axis; red line – filtered, dots – raw) and standard deviation (right axis; blue area – filtered *a posteriori*, T-bars – raw). Linear motion assumed. Only first 70 seconds are shown.

D. Application of Constrained EKF for UXO Mapping

The use of Extended Kalman filter, which is constrained by means of pseudomeasurement, is illustrated on a positioning algorithm for a platform for UXO (Unexploded ordnance) mapping. Publications [138, 167, 168] focus on various aspects of the system described here. For the localization purpose, three sensors were available – the MEMS-IMU, GNSS receiver, and UWB transceiver. The mapping platform is a cart and during mapping only a few distinctive motion types are performed, thus, application of constraints is considered beneficial. Nonetheless, the constraints are not known exactly *a priori*, they are deduced and adjusted during the mapping run itself. Thus, soft constrains have to be used in order to allow some amount of uncertainty of the constraints.

The IMU strapdown mechanization and unconstrained version of the filter will be described first. Then the constraints will be defined and their inclusion to the filter will be presented.

The IMU is a MEMS device (Microstrain 3DM-GX3-35) provides raw measurements from triaxial gyroscope and triaxial accelerometer at 100 Hz rate, nonetheless, the data was downsampled by factor 2 prior to mechanization and fusion. The IMU measuremenst were synchronized to the GPST via its own GNSS receiver. Although the IMU is able to integrate the GNSS and inertial position estimates, only the raw data was used and the GNSS was used only for the synchronization purpose.

RTK-quality position estimates with 5 Hz rate were produced by the GNSS receiver; under open-sky conditions the standard deviation was not worse than a few centimeters. In certain tests the GNSS solution was not available due to vegetation foliage.

The UWB system outputs position estimate with approximately decimeter-level accuracy. The UWB position estimate was based on TWR measurements, the period of the positioning fixes was not constant. Usually, the time separation of the UWB-epochs was 0.25 to 0.45 s. Prior to the fusion of the measurements by means of an EKF, the data are converted to the navigation frame as was defined in 3.1.2. The n-frame was referenced to the initialization point of the test.

D.1. The IMU Mechanization

It is assumed that the IMU measurements are degraded by biases and random walks. The vectors of accelerometer (acceleration) and gyroscope (angular rate) measurements from epoch k are denoted by $\boldsymbol{a}[k]$ and $\boldsymbol{\omega}[k]$, and they are referenced to the b-frame.[122] Symbols $\boldsymbol{\delta}_a[k]$, $\boldsymbol{\delta}_\omega[k]$ correspond to the respective biases; note that the biases are estimated by the error-state KF, which is described in the following section.

The mechanization of the IMU measurement is performed in the n-frame using quaternion representation. The attitude is expressed by means of $\mathring{\boldsymbol{q}}$, which describes the transformation from the b-frame to the n-frame, i.e. in the notation of Chapter 3 it would be denoted by $\mathring{\boldsymbol{q}}^{\mathrm{n}}_{\mathrm{b}}$. The attitude update is performed using the forward integration as described in (3.48) on page 100 and application of (3.54). Note that the transport rate can be neglected for the MEMS IMU involved. Prior to the integration, the bias is subtracted from the measurement. In a single line it can be written as

$$\mathring{\boldsymbol{q}}[k+1] = \mathring{\boldsymbol{q}}[k] \otimes \mathring{\boldsymbol{q}}\left\{\left(\boldsymbol{\omega}[k] - \boldsymbol{\delta}_\omega[k]\right)\tau_{sa}\right\}, \tag{D.1}$$

where τ_{sa} is the time between the successive IMU measurement samples (in this particular case 20 ms).

The attitude in the halftime between epochs is obtained using forward integration according to (3.61). Since MEMS IMU is used, the simplified n-frame update (3.72) can be used. Note that both, accelerometer and gyroscope biases have to be subtracted from the measurement, thus

$$\boldsymbol{v}[k+1] = \boldsymbol{v}[k] + \left(\boldsymbol{R}\left\{\mathring{\boldsymbol{q}}[k] \otimes \mathring{\boldsymbol{q}}\left\{\tfrac{\tau_{sa}(\boldsymbol{\omega}[k]-\boldsymbol{\delta}_\omega[k])}{2}\right\}\right\}\left(\boldsymbol{a}[k] - \boldsymbol{\delta}_a[k]\right) + \boldsymbol{g}\right)\tau_{sa}. \tag{D.2}$$

The gravity is compensated by means of the $\boldsymbol{g}$ vector, which is a downward-pointing vector. The position update in the n-frame is performed using backward integration:

$$\boldsymbol{r}[k+1] = \boldsymbol{p}[k] + \boldsymbol{v}[k+1]\,\tau_{sa}. \tag{D.3}$$

Since the data is post-processed, it is possible to run backward mechanization; in such case, the attitude, velocity and position updates follow

$$\mathring{\boldsymbol{q}}[k-1] = \mathring{\boldsymbol{q}}[k] \otimes \mathring{\boldsymbol{q}}\left\{-\left(\boldsymbol{\omega}[k] - \boldsymbol{\delta}_\omega[k]\right)\tau_{sa}\right\} \tag{D.4}$$

$$\boldsymbol{v}[k-1] = \boldsymbol{v}[k] - \left(\boldsymbol{R}\left\{\mathring{\boldsymbol{q}}[k] \otimes \mathring{\boldsymbol{q}}\left\{\tfrac{-\tau_{sa}(\boldsymbol{\omega}[k]-\boldsymbol{\delta}_\omega[k])}{2}\right\}\right\}\left(\boldsymbol{a}[k] - \boldsymbol{\delta}_a[k]\right) + \boldsymbol{g}\right)\tau_{sa} \tag{D.5}$$

$$\boldsymbol{r}[k-1] = \boldsymbol{p}[k] - \boldsymbol{v}[k-1]\,\tau_{sa}. \tag{D.6}$$

It is necessary to remark that the the IMU was initialized to the origin of local NED coordinate system, the platform was always stationary during initialization. The attitude was initialized partially from the accelerometers, i.e. the tilts were estimated from the direction of gravity vector, see (3.82) on page 105. The heading, or yaw angle, was coarsely initialized by means of a magnetometer measurements; local magnetic declination was taken into account.

[122] The b-frame axes are aligned with the platform's forward, right and down directions. The b-superscript denoting the frame (as in Chapter 3) was dropped for the sake of simplicity.

D.2. The Error-state Kalman Filter

Each of the subsystems to be integrated produces position estimates, in addition, the IMU mechanization outputs attitude and velocity. The system that integrates the subsystem outputs at the position- and velocity-estimate level is sometimes referred as a loosely coupled system.

There are two goals of the measurement fusion algorithm: first, it should provide the best possible position estimate; second, it should estimate the IMU biases. According to [135], the error-state Kalman filter (ESKF) is not estimating the position, velocity and attitude (PVA) directly, but the state vector is considered to be the error of the *nominal* PVA solution (the output of the strapdown mechanization) w.r.t the *true* PVA. The true PVA is unknown, only its position component is observed via imperfect GNSS and UWB measurements. The input to the ESKF is therefore the difference of the position vector obtained from GNSS (or UWB) and from the nominal position. It is necessary to remark that the underlying model is nonlinear, since the attitude error is incorporated in the ESKF as well. Since linearization is required, error-state EKF is exploited.

The state vector $\boldsymbol{x}$ is defined as

$$\boldsymbol{x} = \begin{bmatrix} \check{\boldsymbol{r}}^{\mathrm{T}} & \check{\boldsymbol{v}}^{\mathrm{T}} & \check{\boldsymbol{\nu}}^{\mathrm{T}} & \boldsymbol{\delta}_a^{\mathrm{T}} & \boldsymbol{\delta}_\omega^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \tag{D.7}$$

where $\check{\boldsymbol{r}}$, $\check{\boldsymbol{r}}$ denote the position and velocity error, $\check{\boldsymbol{\nu}}$ is the attitude error expressed as a rotation vector. The six elements of the state vector are the accelerometer and gyroscope biases, respectively. By modifying expressions from [135] we may obtain the state transition matrix as follows

$$\boldsymbol{F} = \begin{bmatrix} \mathbb{I} & \mathbb{I}\,\tau_{sa} & \mathbb{0} & \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{I} & -\boldsymbol{R}\{\mathring{\boldsymbol{q}}\}\,[\boldsymbol{a}-\boldsymbol{\delta}_a]_\times\,\tau_{sa} & -\boldsymbol{R}\{\mathring{\boldsymbol{q}}\}\,\tau_{sa} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & -\boldsymbol{R}\{\boldsymbol{\omega}-\boldsymbol{\delta}_\omega\}^{\mathrm{T}} & \mathbb{0} & -\mathbb{I}\,\tau_{sa} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{I} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{0} & \mathbb{I} \end{bmatrix}. \tag{D.8}$$

The small-angle linear approximation is used in the state transition matrix, so the application of the function process model (4.60) can be performed by means of left matrix multiplication by $\boldsymbol{F}$. The *a priori* state vector and its covariance can be obtained by means of

$$\boldsymbol{x}^-[k] = \mathbf{F}\boldsymbol{x}^+[k-1] \tag{D.9}$$

$$\boldsymbol{P}^-[k] = \boldsymbol{F}[k-1]\boldsymbol{P}^+[k-1]\boldsymbol{F}^{\mathrm{T}}[k-1] + \boldsymbol{Q}. \tag{4.62}$$

It is worth noting that the state covariance propagation remains the same as in case of a regular EKF.

The measurement vectors are defined by

$$\boldsymbol{y}_{\mathrm{S}} = \boldsymbol{r}_{\mathrm{S}} - \boldsymbol{r}\boldsymbol{y}_{\mathrm{U}} = \boldsymbol{r}_{\mathrm{U}} - \boldsymbol{r}. \tag{D.10}$$

The measurements should be equivalent to the first three elements of the state vector (the position error vector) and thus the measurement model is linear. It can be described by means of

$$\boldsymbol{H}_{\mathrm{S}} = \boldsymbol{H}_{\mathrm{U}} = \begin{bmatrix} \mathbb{I}_{3\times3} & \mathbb{0}_{3\times12} \end{bmatrix}. \tag{D.11}$$

In the measurement update, there is a minor modification of the innovation evaluation, since the measurement model is linear. The *a posteriori* update is therefore obtained using

$$\boldsymbol{z}[k] = \boldsymbol{y}[k] - \boldsymbol{H}\boldsymbol{x}^{-}[k] \tag{D.12}$$
$$\boldsymbol{S}[k] = \boldsymbol{H}[k]\boldsymbol{P}^{-}[k]\boldsymbol{H}^{\mathrm{T}}[k] + \boldsymbol{R} \tag{4.65}$$
$$\boldsymbol{K}[k] = \boldsymbol{P}^{-}[k]\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{S}[k])^{-1} \tag{4.66}$$
$$\boldsymbol{x}^{+}[k] = \boldsymbol{x}^{-}[k] + \boldsymbol{K}[k]\boldsymbol{z}[k] \tag{4.67}$$
$$\boldsymbol{P}^{+}[k] = (\mathbb{I} - \boldsymbol{K}[k]\boldsymbol{H}[k])\,\boldsymbol{P}^{-}[k]. \tag{4.68}$$

The matrix $\boldsymbol{H}$ is formed according to the availability of the measurements.

Only GNSS: $\boldsymbol{H} = \boldsymbol{H}_{\mathrm{S}}$ $\boldsymbol{y} = \boldsymbol{y}_{\mathrm{S}}$ $\boldsymbol{R} = \boldsymbol{R}_{\mathrm{S}}$ (D.13)

Only UWB: $\boldsymbol{H} = \boldsymbol{H}_{\mathrm{U}}$ $\boldsymbol{y} = \boldsymbol{y}_{\mathrm{U}}$ $\boldsymbol{R} = \boldsymbol{R}_{\mathrm{U}}$ (D.14)

Both: $\boldsymbol{H} = \begin{bmatrix}\boldsymbol{H}_{\mathrm{S}}\\ \boldsymbol{H}_{\mathrm{U}}\end{bmatrix}$ $\boldsymbol{y} = \begin{bmatrix}\boldsymbol{y}_{\mathrm{S}}\\ \boldsymbol{y}_{\mathrm{U}}\end{bmatrix}$ $\boldsymbol{R} = \begin{bmatrix}\boldsymbol{R}_{\mathrm{S}} & \mathbb{0}\\ \mathbb{0} & \boldsymbol{R}_{\mathrm{U}}\end{bmatrix}$ (D.15)

Naturally, the measurement update is omitted when neither of the radio position estimates is available. It is important to remark that such architecture of the ESKF allows us to disregard one of the measurement sources; the system can be considered modular.

The *true* PVA estimates $\hat{\boldsymbol{r}}$, $\hat{\boldsymbol{v}}$, $\hat{\mathring{\boldsymbol{q}}}$ are obtained by means of addition of the *a posteriori* estimated errors to the *nominal* PVA estimates. Note that in case of the quaternion attitude, the erroneous rotation has to be applied by means of quaternion multiplication. Omitting the KF epoch identifier, which is the same for all variables involved, we may write

$$\hat{\boldsymbol{r}} = \boldsymbol{r} + \check{\boldsymbol{r}}^{+} \tag{D.16}$$
$$\hat{\boldsymbol{v}} = \boldsymbol{v} + \check{\boldsymbol{v}}^{+} \tag{D.17}$$
$$\hat{\mathring{\boldsymbol{q}}} = \mathring{\boldsymbol{q}} \otimes \mathring{\boldsymbol{q}}\left\{\check{\boldsymbol{\nu}}^{+}\right\}. \tag{D.18}$$

It is necessary to correct the nominal PVA estimate occasionally, in order to avoid excessive growth of the error, since the linearization (the small-angle approximation) used in construction of the $\boldsymbol{F}$-matrix would become inaccurate. When correcting, the *true* PVA is assigned to the *nominal* PVA, the corresponding elements (the first nine) of the state vector $\boldsymbol{x}$ are set to zeros, however, the covariances remain unchanged. The bias estimates and their covariances are not affected as well.

It is clear that the biases and the attitude and velocity errors are observed via the state propagation. Therefore, the *nominal* PVA correction cannot be performed after each measurement update, because the filter state would be partially reset before the prediction step. In order to make all state vector elements observable a few full iterations of the ESKF are necessary.

On one hand, the *nominal* PVA should be corrected as often as possible, in order to keep the error low; on the other hand, the intervals between the *nominal* PVA

cor rections should be long in order to make the state vector more observable. For this particular application, a suitable value of 4 corrections per second was found experimentally.

It is possible to modify the ESKF to run backwards in time in a strightforward manner. In such case, the prediction step runs from epoch $k+1$ to k, and the value of τ_{sa} is negative.

D.3. Motion Constraints

As anticipated in the beginning of the appendix, the motion of the platform can be described by means of constraints rather accurately. The platform is moved over an area in parallel strips, as shown in Figure D.2. Before each strip, the platform is stationary at least for a few seconds; the velocity of motion along the strip is constant. There are three kinds of movements:

Stationary – Zero-velocity update (ZUPT) is performed. No attitude change is possible, since the platform (cart) cannot be rotated when stationary; the height (n-frame z-axis position) is constant. The complete nominal velocity estimate is the error, since the true value is zero. It is necessary to remark that these constraints are rotationally invariant, thus they can be applied directly in the n-frame. The pseudomeasurement related to this motion type will be denoted by $\bullet$ subscript.

Straight motion – In this case the attitude and height are constrained in the same way as for the stationary condition. The platform moves along its b-frame x-axis with a constant velocity $v'_{x\uparrow}$, the lateral and vertical velocity components are zero. The velocity constraint originates in the b-frame, therefore it has to be transformed to the n-frame in order to become a pseudomeasurement for the ESKF. The pseudomeasurement related to this motion type will be denoted by $\uparrow$ subscript.

Turn – During the turn, only the height and velocity direction are constrained. The velocity pseudomeasurement is obtained as a projection of the nominal velocity estimate into the direction of b-frame x-axis. The attitude is unconstrained completely. The pseudomeasurement related to this motion type will be denoted by $\circlearrowleft$ subscript.

The motion constraints for each of the kinds of movements are illustrated in Figure D.3.

It is necessary to define the pseudomeasurements and the respective pseudomeasurement matrices. Since we are constraining the state values, the pseudomeasurement matrices may be interpreted as a "mapping" to the relevant components of the ESKF state vector.

The constant-height constraint is rather straightforward; the vertical component nominal position should be the same as the initialization height, i.e. 0. Therefore the pseudomeasurement follows

$$\boldsymbol{y}'_{\mathrm{h}\bullet} = \boldsymbol{y}'_{\mathrm{h}\uparrow} = \boldsymbol{y}'_{\mathrm{h}\circlearrowleft} = -r_z^- \tag{D.19}$$

$$\boldsymbol{H}'_{\mathrm{h}} = \begin{bmatrix} \mathbb{O}_{1\times 2} & 1 & \mathbb{O}_{1\times 12} \end{bmatrix}. \tag{D.20}$$

It is worth noting that this constraint should not be applied when the surface is not flat and level.

The velocity constraint is different for each of the motion types. In the stationary case, ZUPT, any nominal velocity is considered to be an error, see (D.21). The velocity pseudomeasurement of the straight movement (D.21) is a difference of constant velocity vector in the direction of b-frame x-axis transformed into n-frame and the nominal velocity estimate. The rotation is performed using the last known nominal attitude. During turns, the velocity pseudomeasurement (D.24) is the difference of the nominal velocity *projected*[123] into the direction of the b-frame x-axis expressed in n-frame, and the nominal velocity estimate.

$$\boldsymbol{y}'_{\mathrm{v}\bullet} = -\boldsymbol{v}[k] \tag{D.21}$$

$$\boldsymbol{y}'_{\mathrm{v}\uparrow} = \mathring{\boldsymbol{q}}[k] \otimes \begin{bmatrix} v'_{x\uparrow} & 0 & 0 \end{bmatrix}^{\mathrm{T}} \otimes \mathring{\boldsymbol{q}}[k] - \boldsymbol{v}[k] \tag{D.22}$$

$$\boldsymbol{y}'_{\mathrm{v}\circlearrowleft} = \left(\mathring{\boldsymbol{q}}[k] \otimes \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \otimes \mathring{\boldsymbol{q}}[k] \right) \left(\mathring{\boldsymbol{q}}[k] \otimes \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \otimes \mathring{\boldsymbol{q}}[k] \right)^{\mathrm{T}} \boldsymbol{v}[k] - \boldsymbol{v}[k] \tag{D.23}$$

$$\boldsymbol{H}'_{\mathrm{v}} = \begin{bmatrix} \mathbb{O}_{3\times3} & \mathbb{I}_{3\times3} & \mathbb{O}_{3\times9} \end{bmatrix} \tag{D.24}$$

The attitude constraint is applied only when the platform is not turning. The attitude-error pseudomeasurement is defined as the rotation between the current and previous epoch in both, stationary and straight movement, cases. It follows

$$\boldsymbol{y}'_{\theta\bullet} = \boldsymbol{y}'_{\theta\uparrow} = \boldsymbol{\nu}\left\{\mathring{\boldsymbol{q}}^{*}[k] \otimes \mathring{\boldsymbol{q}}[k-1]\right\} \tag{D.25}$$

$$\boldsymbol{H}'_{\theta} = \begin{bmatrix} \mathbb{O}_{3\times6} & \mathbb{I}_{3\times3} & \mathbb{O}_{3\times6} \end{bmatrix}, \tag{D.26}$$

where $\boldsymbol{\nu}\{\}$ denotes the attitude parametrization conversion from quaternion to rotation vector. The uncertainty, or "softness," of the constraints can be set by setting the respective pseudomeasurement covariance matrices $\boldsymbol{R}'$. In our implementation the covariances were neglected, thus all the $\boldsymbol{R}'$ matrices were diagonal.

The relevant constraints, i.e. pseudomeasurement vectors, matrices and covariance matrices, are appended to the regular measurements and the respective matrices in the same way, as was described in Section 4.2.4. In order to avoid performing the computationally demanding ESKF measurement update in every epoch, the constraint update was performed in every fourth epoch only.

The choice of the constraint set was performed by means of a neural network (NN). The detailed description of the NN does not fall in the scope of the thesis, however, is available in [138, 167]. Two different NN approaches are compared in [138] – the feed-forward NN and the convolutional NN; only the results of the latter are presented here. The inputs of the convolutional neural network were the raw accelerometer and gyroscope measurements.

It is worth remarking that the soft constraints deal with two nuisance characteristics of the platform motion. First, the platform was pushed by men power over an uneven surface, thus, the velocity is not perfectly constant and the straight segment may not be perfectly straight. Second, after the turn, the platform was raised in order to get

[123] The projection is performed using (4.84), from page 133.

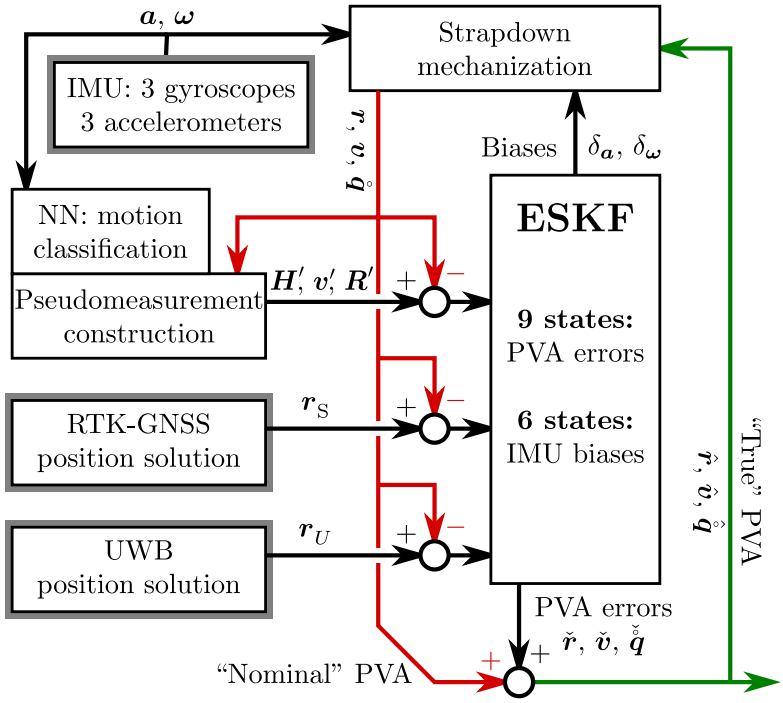


FIGURE D.1.: Block diagram of the integrated system.

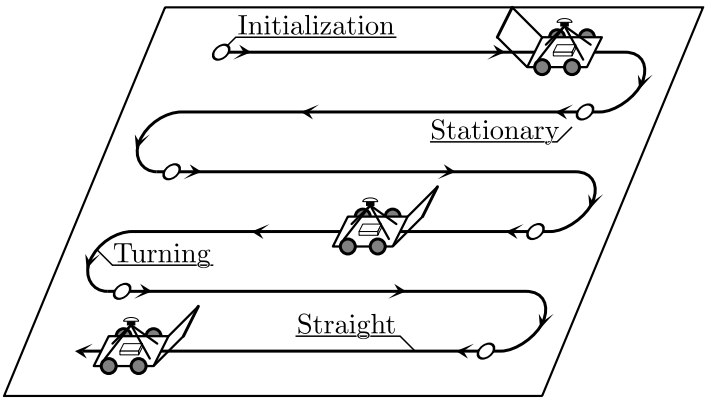
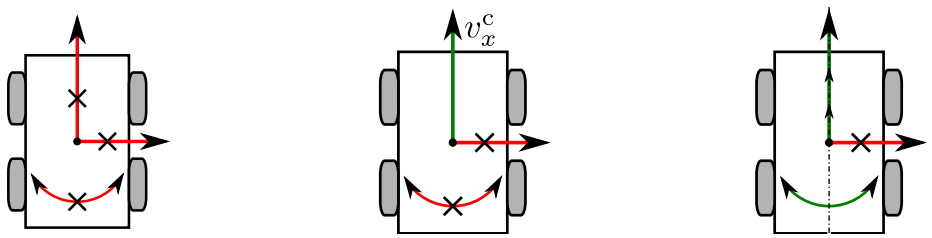


FIGURE D.2.: Motion pattern of the platform.



(A) Stationary (B) Straight (C) Turning

FIGURE D.3.: Constraints for different motion types.

aligned with the following mapping strip. Such motion violates the constraints, the platform moves sideways and is rotated. Should the hard constraints be used, such motion singularities may lead to filter instability.

D.4. Forward-backward Solution Combination

In the postprocessing, it is possible to run the IMU strapdown mechanization and ESKF both forward and backward in time. When *nominal* PVA estimates, PVA error estimates (states of the ESKF) are stored for each epoch in both directions, the forward and backward estimates of the *true* PVA can be obtained for each epoch by means of (D.16), (D.17), and (D.18). The ESKF *a posteriori* state covariance is assumed to be the variance of the *true* PVA estimate.

The forward and backward solution can be combined. The optimal combination is a covariance-weighted average of the two solutions, as was declared in (4.86), (4.87) in Section 4.2.5. Due to the attitude representations, this approach cannot be used directly. For the purpose of solution combination, the ESKF state covariance matrix will be considered block-diagonal

$$\tilde{\boldsymbol{P}} = \begin{bmatrix} \tilde{\boldsymbol{P}}^{rv} & \mathbb{O}_{6\times3} & \mathbb{O}_{6\times6} \\ \mathbb{O}_{3\times6} & \tilde{\boldsymbol{P}}^{\theta} & \mathbb{O}_{3\times6} \\ \mathbb{O}_{6\times6} & \mathbb{O}_{6\times3} & \tilde{\boldsymbol{P}}^{\delta} \end{bmatrix}, \tag{D.27}$$

where $\tilde{\boldsymbol{P}}^{rv}$, $\tilde{\boldsymbol{P}}^{\theta}$, $\tilde{\boldsymbol{P}}^{\delta}$ denote the covariance submatrices related to the position and velocity, the attitude and the biases, respectively. Certainly, such assumption brings certain suboptimality to the forward-backward combination, since a number of covariances is neglected.

Denoting the forward, backward and combined (smoothed) solutions by means of $\rightarrow$, $\leftarrow$ and $\leftrightarrow$ subscripts, the combination of the estimates of the *true* position, velocity and biases can be performed by a straightforward application of (4.86), (4.87):

$$\tilde{\boldsymbol{P}}^{rv}_{\leftrightarrow} = \left((\tilde{\boldsymbol{P}}^{rv}_{\rightarrow})^{-1} + (\tilde{\boldsymbol{P}}^{rv}_{\leftarrow})^{-1}\right)^{-1} \tag{D.28}$$

$$\begin{bmatrix} \hat{\boldsymbol{r}}_{\leftrightarrow} \\ \hat{\boldsymbol{v}}_{\leftrightarrow} \end{bmatrix} = \tilde{\boldsymbol{P}}^{rv}_{\leftrightarrow} \left((\tilde{\boldsymbol{P}}^{rv}_{\rightarrow})^{-1} \begin{bmatrix} \hat{\boldsymbol{r}}_{\rightarrow} \\ \hat{\boldsymbol{v}}_{\rightarrow} \end{bmatrix} + (\tilde{\boldsymbol{P}}^{rv}_{\leftarrow})^{-1} \begin{bmatrix} \hat{\boldsymbol{r}}_{\leftarrow} \\ \hat{\boldsymbol{v}}_{\leftarrow} \end{bmatrix} \right) \tag{D.29}$$

$$\tilde{\boldsymbol{P}}^{\delta}_{\leftrightarrow} = \left((\tilde{\boldsymbol{P}}^{\delta}_{\rightarrow})^{-1} + (\tilde{\boldsymbol{P}}^{\delta}_{\leftarrow})^{-1}\right)^{-1} \tag{D.30}$$

$$\begin{bmatrix} \hat{\boldsymbol{\delta}}_{a\leftrightarrow} \\ \hat{\boldsymbol{\delta}}_{\omega\leftrightarrow} \end{bmatrix} = \tilde{\boldsymbol{P}}^{\delta}_{\leftrightarrow} \left((\tilde{\boldsymbol{P}}^{\delta}_{\rightarrow})^{-1} \begin{bmatrix} \hat{\boldsymbol{\delta}}_{a\rightarrow} \\ \hat{\boldsymbol{\delta}}_{\omega\rightarrow} \end{bmatrix} + (\tilde{\boldsymbol{P}}^{\delta}_{\leftarrow})^{-1} \begin{bmatrix} \hat{\boldsymbol{\delta}}_{a\leftarrow} \\ \hat{\boldsymbol{\delta}}_{\omega\leftarrow} \end{bmatrix} \right). \tag{D.31}$$

Apparently, two quaternions representing the forward and backward attitude estimate cannot be combined in such straightforward manner; moreover, the covariance submatrix $\tilde{\boldsymbol{P}}^{\theta}$ corresponds to the rotation vector attitude parametrization instead of quaternions. The combined attitude is obtained by means of partial rotation from the forward to the backward attitude solution. The amount of the rotation along each axis

is given by the ratio of the diagonal elements of $\tilde{\boldsymbol{P}}^{\theta}_{\rightarrow}$ and $\tilde{\boldsymbol{P}}^{\theta}_{\leftarrow}$ submatrices:

$$\mathring{\boldsymbol{q}}_{\leftrightarrow} = \mathring{\boldsymbol{q}}_{\rightarrow} \otimes \mathring{\boldsymbol{q}} \left\{ \frac{\mathrm{Diag}\left(\tilde{\boldsymbol{P}}^{\theta}_{\rightarrow}\right)}{\mathrm{Diag}\left(\tilde{\boldsymbol{P}}^{\theta}_{\rightarrow} + \tilde{\boldsymbol{P}}^{\theta}_{\leftarrow}\right)} \boldsymbol{\nu} \left\{ \mathring{\boldsymbol{q}}^{*}_{\rightarrow} \otimes \mathring{\boldsymbol{q}}_{\leftarrow} \right\} \right\}. \tag{D.32}$$

It is necessary to remark that the fraction of the two diagonal matrices with the same dimensions can be performed as an element-wise division of the corresponding diagonal elements.

D.5. Experimental Results

Two sets of results are presented here, in order to demonstrate the performance advantage brought by applying the constraints. In the first, open-sky area scenario, accurate and frequent GNSS position estimates are available. Therefore, UWB/IMU integrated solution is compared to a GNSS/IMU reference solution. During the second, canopied area scenario, the GNSS is unavailable; consequently a reference solution is missing. In order to quantitatively asses the impact of the constraints, artificial outages of the UWB positioning were simulated.

All the presented results were obtained as a combination of the forward and backward solutions.

Open-sky area

In the open-sky test, the parallel strips were oriented in the North-South direction. The length of the strips was approximately 80 m. The reference GNSS/IMU solution and the raw GNSS data is depicted in Figure D.4a as a black line and red ×-marks, respectively. In the left part of the figure, the Northern turning point is zoomed in; the whole trajectory is in the right part, the zoomed part is enclosed by a gray rectangle.

The *unconstrained* UWB/IMU solution is presented in Figure D.4b. The trajectory starts at the point with zero coordinates, the mapping strips were stacked from East towards West. The UWB position estimates are depicted by the blue circles; clearly, they are less frequent and less accurate than the GNSS measurements. The red ×-marks are the GNSS position estimates, which were not used. It is apparent that the UWB measurements are sparse when approaching the turn; the operator pushing the cart was obstructing the line of sight for the UWB anchors, which were located on the opposite end of the strip. The length of the strip is rather close to the maximal range of the UWB ranging system, thus, there was not sufficient signal-power margin. The UWB/IMU solution becomes rather inaccurate when the UWB position update is not available, the estimated trajectory (solid black line) is uneven and diverging. In certain parts of the trajectory, the UWB and GNSS position estimates are noticeably different.

Results which were obtained with the *constrained* ESKF are provided in Figure D.4c. The trajectory is considerably smoother than in the unconstrained case, however, still offset from the GNSS position estimates. Since the GNSS position data is unavailable to the filter, the offset cannot be eliminated by the constraints; the solution correctly

follows the UWB position estimates. In the right part of the figure, the NN-based motion model classification is shown. The blue +-marks depict points with stationary cart, red and green colors indicate straight and turning motion. The NN was trained and verified on a different part of the open-sky area test.

Regardless of the positioning method used, it can be observed that before some strips, there is a move that does not fit to any of the three motion models considered; it manifests as a "dent" in the trajectory right after the turn.

The quantitative analysis of the two UWB/IMU trajectory estimates is provided in Table D.1. The RMSE column shows the RMS error of the UWB/IMU solution w.r.t. the GNSS/IMU solution. The Δ_a and Δ_r columns present the absolute and relative improvement of the constrained solution w.r.t. the unconstrained one. Apparently, the constraints bring approximately 10 % accuracy advantage, nonetheless, the RMS error is still close to 30 cm. Arguably, a dominant part of the error is caused by the occasional offset between UWB and GNSS position estimates.

Canopied area

The sky view was obstructed by a tree canopy in the second test. The strips were approximately 30 m long and oriented from Northwest to Southeast direction, the mapping was performed from Southwest toward Northeast.

Since GNSS was not available, the unconstrained UWB/IMU solution was taken as a reference. The trajectory estimates of Northwestern turns are presented in Figure D.5a. Thanks to the shorter mapping strips, the loss of UWB position estimates (blue circles) due to line of sight shadowing is not a significant problem in the canopied area test. An overview of the whole trajectory is available as Figure D.6.

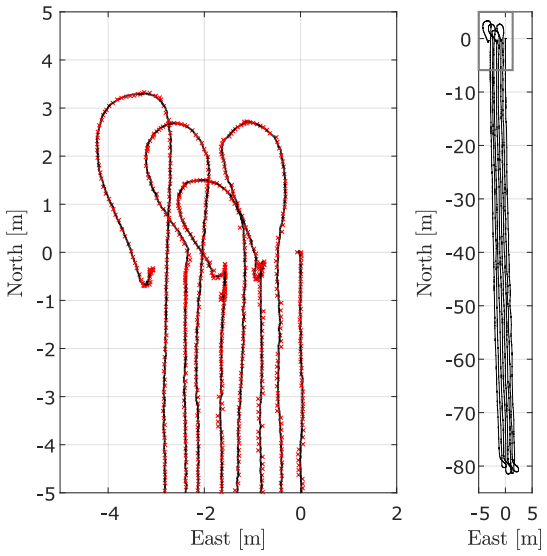
In order to demonstrate the difference of the unconstrained and constrained trajectory estimates, two 5 to 10 s long UWB outages were simulated. During that periods, only the trajectory was estimated using the IMU measurements only. The outage A occurs in a turn, whilst outage B happens during straight motion. The unconstrained trajectory solution (Figure D.5c) appears uneven, whilst the constrained solution (Figure D.5d) is rather smooth. Moreover, the constrained trajectory seems to be more similar in shape to the reference solution (without outages).

The results are quantitatively compared in Table D.2. The RMS errors of the unconstrained solution are rather close to the 0.5 m level during the outages. The introduction of constraints reduces the horizontal RMS error by 16 % in the turn and by 35 % when moving straight. Such result is consistent with the fact that tighter constraints are applied when moving straight (no attitude change is allowed and constant velocity is forced).

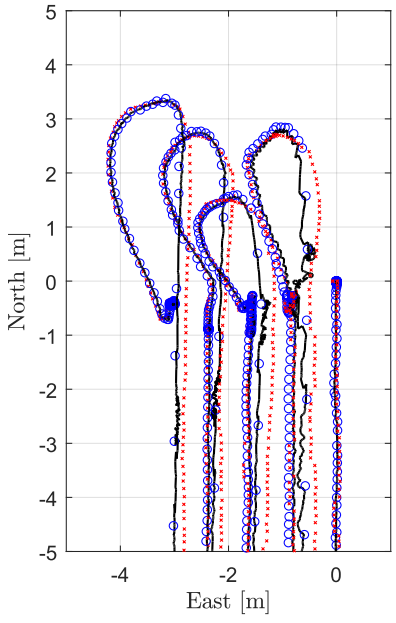
The motion type classification was performed by the same NN as in the open-sky area; visualization is provided in Figure D.5b. Apparently, a few segments were misclassified: the blue ×-marks depicting the stationary platform are appearing on a straight line (should be red); nearby, a straight segment is incorrectly classified as a turn (green). Since the constraints are implemented in a soft manner, the effect of the misclassification is not completely destructive.

D.6. Remarks

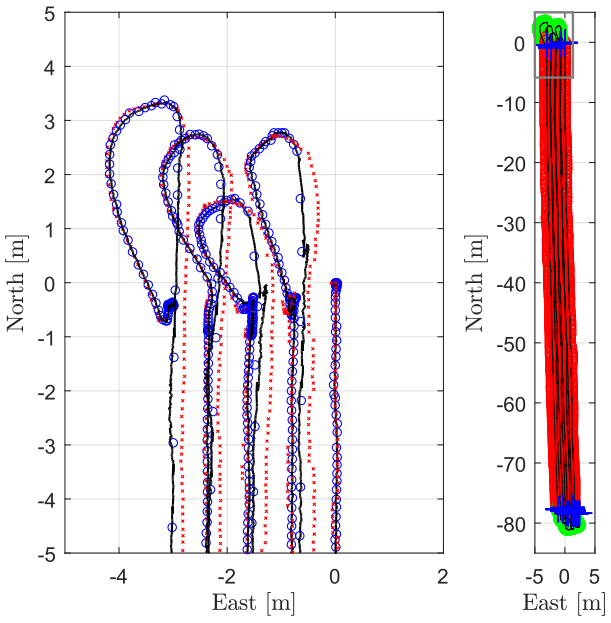
The application of the error-state Kalman filter (ESKF) in a loosely coupled system exploiting GNSS receiver, UWB transceiver and MEMS IMU was presented in this appendix. Moreover, it has been shown that introduction of soft constraints that are implemented as pseudomeasurements (see Section 4.2.4) brings noteworthy performance improvement. The presented architecture is easy to modify; it would be possible to incorporate additional systems or constraints. More detailed information about the system can be found in [138, 167, 168].



(A) GNSS/IMU – reference solution

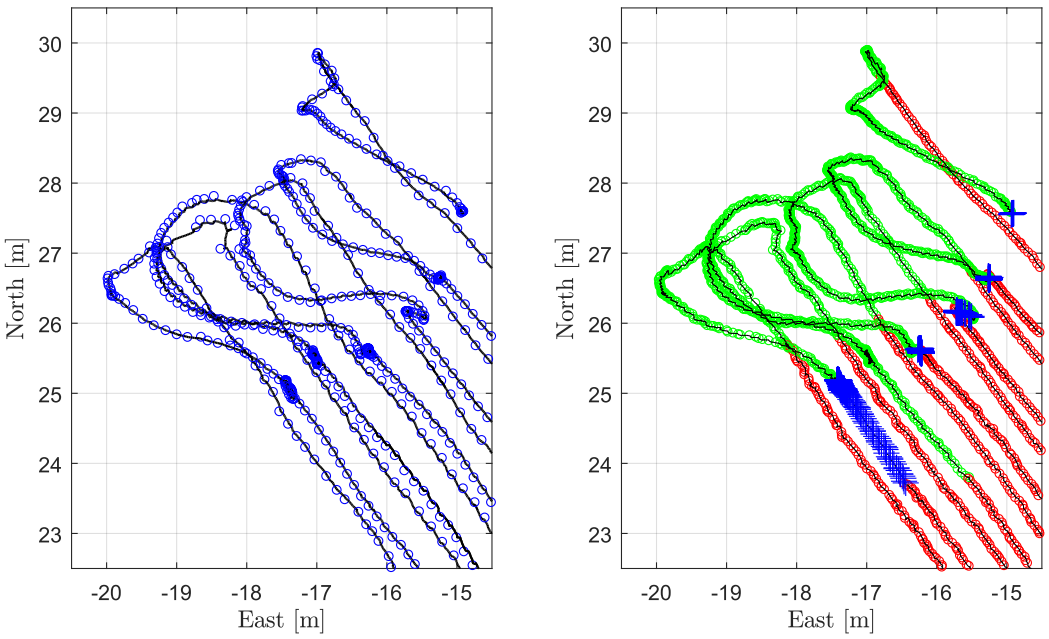


(B) UWB/IMU unconstrained sol.



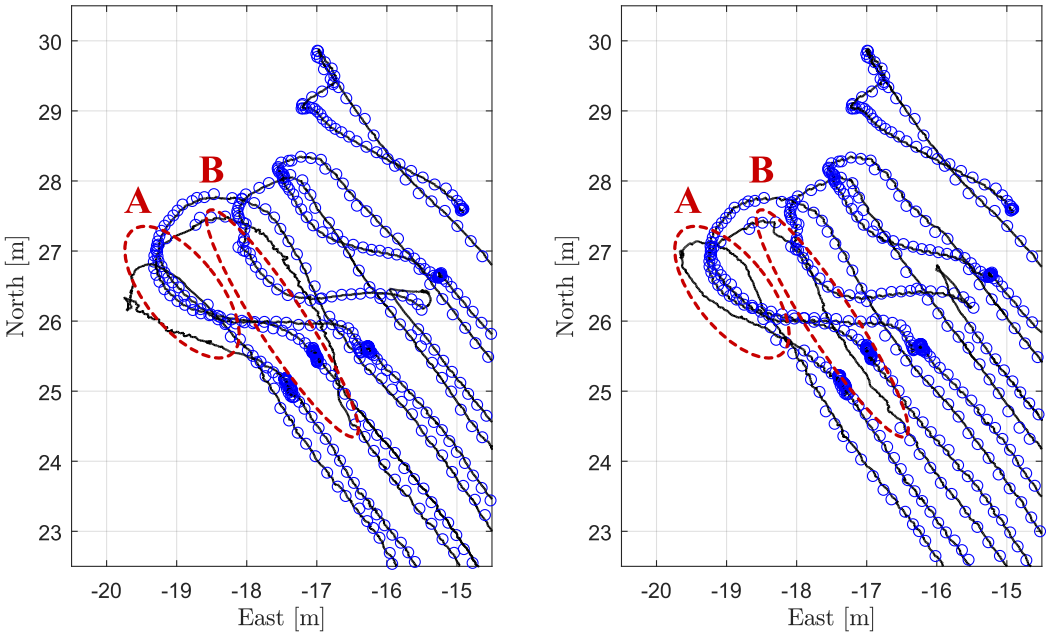
(C) UWB/IMU constrained solution

FIGURE D.4.: Open-sky area trajectory estimate results.



(A) Unconstrained reference solution

(B) Motion type according to NN



(C) Simulated outages, unconstrained sol.

(D) Simulated outages, constrained sol.

FIGURE D.5.: Canopied area UWB/IMU trajectory estimate results.

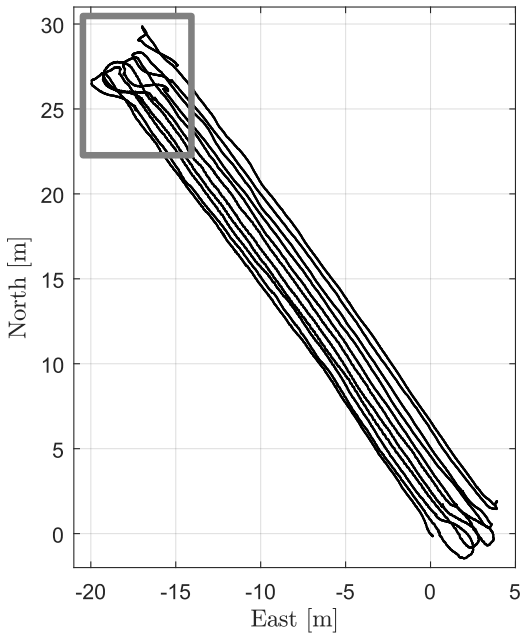


FIGURE D.6.: Canopied area trajectory overview.

TABLE D.1.: Comparison of UWB/IMU trajectory estimate errors w.r.t. GNSS/IMU reference in the open-sky area scenario.

		Error w.r.t. GNSS/IMU		
	UWB/IMU	RMSE [cm]	Δ_a [cm]	Δ_r [%]
2D	unconstrained	32.1		
	constrained	28.9	3.2	10.0
3D	unconstrained	32.9		
	constrained	29.5	3.4	10.3

TABLE D.2.: Comparison of UWB/IMU trajectory estimate errors in the simulated outages in the canopied area scenario.

		A – Turn			**B – Straight**		
	UWB/IMU	RMSE [cm]	Δ_a [cm]	Δ_r [%]	RMSE [cm]	Δ_a [cm]	Δ_r [%]
2D	unconstrained	49.1			46.0		
	constrained	41.1	8.0	16.3	29.7	16.3	35.4
3D	unconstrained	51.4			47.2		
	constrained	42.2	9.2	17.9	31.8	15.4	32.6

Bibliography

1. S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice Hall, 1993, ISBN: 0-13-345711-7.
2. Z. Sahinoglu, S. Gezici, and I. Gvenc, *Ultra-wideband Positioning Systems: Theoretical Limits, Ranging Algorithms, and Protocols*. New York, NY, USA: Cambridge University Press, 2011, ISBN: 978-0-521-18783-1.
3. D. Neirynck, E. Luk, and M. McLaughlin, "An alternative double-sided two-way ranging method", in *2016 13th Workshop on Positioning, Navigation and Communications (WPNC)*, 2016-10, pp. 1–4. DOI: 10.1109/WPNC.2016.7822844.
4. V. Navrátil and F. Vejražka, "Bias and variance of the asymmetric double-sided two-way ranging", *Navigation*, 2019, Accepted on 2019-04-19.
5. V. Navrátil, J. Krška, F. Vejražka, and V. Koreček, "Chained wireless synchronization algorithm for UWB-TDOA positioning", in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, IEEE, 2018-04. DOI: 10.1109/plans.2018.8373376.
6. V. Navrátil, J. Krška, and F. Vejražka, "Experimental evaluation of chained synchronization for UWB systems", in *Proceedings of the 16th IAIN World Congress 2018 (IAIN2018)*, (2018-11-26), Chiba, Japan, 2018, pp. 137–145.
7. A. F. Molisch, K. Balakrishnan, C.-C. Chong, *et al.*, "IEEE 802.15.4a channel model – final report", Tech. Rep., 2004. URL: http://grouper.ieee.org/groups/802/15/archive/802-15-sg4alist/pdf00025.pdf.
8. L. Mailaender, "Geolocation bounds for received signal strength (RSS) on Rician faded channels", LGS Innovations LLC, Tech. Rep., 2011. DOI: 10.13140/RG.2.1.2584.5922.
9. P. Misra and P. Enge, *Global Positioning System: Signals, Measurements, and Performance*, G.-J. Press, Ed. Lincoln, Mass.: Ganga-Jamuna Press, 2001, ISBN: 0-9709544-0-9.
10. J. W. Betz, *Engineering Satellite-Based Navigation and Timing: Global Navigation Satellite Systems, Signals, and Receivers*. John Wiley & Sons, 2015, ISBN: 9781118615935. URL: https://books.google.cz/books?id=%5C_htcCwAAQBAJ.
11. M. McLaughlin and B. Verso, "Asymmetric double-sided two-way ranging in an ultra-wideband communication system", EP 3 158 359 B1, 2016.
12. S. J. Orfanidis, *Electromagnetic waves and antennas*. Rutgers University New Brunswick, NJ, 2002, [rev. 2016-08]. URL: http://www.ece.rutgers.edu/~orfanidi/ewa/.
13. S. Silver, Ed., *Microwave Antenna Theory and Design*, First, ser. Radiation Laboratory series. McGraw-Hill Book Company, 1949, vol. 12.
14. G. Ramos-Llordén, "Cramér-Rao analysis of Inversion Recovery sequence", iMinds-Vision Lab, Department of Physics, University of Antwerp, 2014-04. URL: https://visielab.uantwerpen.be/sites/default/files/profile/Gabriel%20Ramos%20Llorden/crlb_ir.pdf.

15. R. J. Muirhead, "The multivariate normal and related distributions", in *Aspects of Multivariate Statistical Theory*, ser. Wiley Series in Probability and Statistics. Wiley-Blackwell, 2008, ch. 1, pp. 1–49, ISBN: 9780471094425. DOI: 10.1002/9780470316559.ch1.

16. A. Mallat, J. Louveaux, and L. Vandendorpe, "UWB based positioning: Cramer Rao bound for angle of arrival and comparison with time of arrival", in *2006 Symposium on Communications and Vehicular Technology*, 2006-11, pp. 65–68. DOI: 10.1109/SCVT.2006.334374.

17. A. Mallat, J. Louveaux, and L. Vandendorpe, "UWB based positioning in multipath channels: CRBs for AOA and for hybrid TOA-AOA based methods", in *2007 IEEE International Conference on Communications*, 2007-06, pp. 5775–5780. DOI: 10.1109/ICC.2007.957.

18. J. Proakis, *Digital Communications*, ser. Communications and signal processing. McGraw-Hill, 1995, ISBN: 978-0070517264.

19. "IEEE standard for low-rate wireless networks", *IEEE Std 802.15.4-2015*, 2016, (Revision of IEEE Std 802.15.4-2011). DOI: 10.1109/IEEESTD.2016.7460875.

20. "IEEE standard letter designations for radar-frequency bands", *IEEE Std. 521-2002*, 2003, (Revision of IEEE Std 521-1984). DOI: 10.1109/IEEESTD.2003.94224.

21. B. W. Parkinson and J. J. Spilker Jr., *Global Positioning System: Theory and Applications*. Washington: American Institute of Aeronautics and Astronautics, 1996, vol. 1, ISBN: 9781563471063.

22. E. D. Kaplan and C. J. Hegarty, *Understanding GPS: principles and applications*, 2nd ed. Artech House, 2006, ISBN: 1-58053-894-0.

23. International GNSS Service (IGS) RINEX Working Group and Radio Technical Commission for Maritime Services Special Committee 104 (RTCM – SC104). (2015-07). RINEX – The Receiver Independent Exchange Format, Version 3.03, URL: ftp://igs.org/pub/data/format/rinex303.pdf (vid. 2018-09-02).

24. S. Schaer, W. Gurtner, and J. Feltens, "IONEX: The IONosphere Map EXchange Format version 1", IGS, 1998. URL: ftp://igs.org/pub/data/format/ionex1.pdf (vid. 2018-09-02).

25. "SINEX - Solution (software/technique) INdependent EXchange Format", Earth Rotation and Reference Systems Service, 2006. URL: https://www.iers.org/SharedDocs/Publikationen/EN/IERS/Documents/ac/sinex/sinex_v202_pdf.pdf (vid. 2018-09-02).

26. S. Hilla, "The extended standard product 3 orbit format (SP3–c)", NGS, 2010. URL: ftp://igs.org/pub/data/format/sp3c.txt (vid. 2018-09-02).

27. M. Rothacher and R. Schmid, "ANTEX: The antenna exchange format, version 1.4", TU München, 2010. URL: ftp://igs.org/pub/station/general/antex14.txt (vid. 2018-09-02).

28. (2017). BINEX: Binary exchange format, URL: http://binex.unavco.org/binex.htm (vid. 2018-09-02).

29. Radio Technical Commission for Maritime Services. (2018). RTCM - general overview, URL: http://www.rtcm.org/about.html (vid. 2018-08-13).

30. Radio Technical Commission for Maritime Services. (2016-10). RTCM 10403.3, differential GNSS services - version 3, URL: http://www.rtcm.org/differential-global-navigation-satellite--dgnss--standards.html (vid. 2018-08-13).

31. P. Waller, F. Gonzalez, S. Binda, *et al.*, "The in-orbit performances of GIOVE clocks", *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 57, no. 3, pp. 738–745, 2010-03. DOI: 10.1109/tuffc.2010.1472.

32. A. Cernigliaro, S. Valloreia, G. Fantino, L. Galleani, and P. Tavella, "Analysis on GNSS space clocks performances", in *2013 Joint European Frequency and Time Forum & International Frequency Control Symposium (EFTF/IFC)*, IEEE, 2013-07. DOI: 10.1109/eftf-ifc.2013.6702289.

33. H. Gong, W. Yang, H. Sha, X. Zhu, and F. Wang, "Frequency stability estimation of compass on-board clock based on smoothed broadcast ephemeris", in *2013 Joint European Frequency and Time Forum & International Frequency Control Symposium (EFTF/IFC)*, IEEE, 2013-07. DOI: 10.1109/eftf-ifc.2013.6702233.

34. *IS-GPS-200J, navstar GPS space segment/navigation user interfaces*, 2018-04-25. URL: https://www.gps.gov/technical/icwg/IS-GPS-200J.pdf (vid. 2018-11-16).

35. (2018). IGS analysis GPS broadcast, URL: http://www.igs.org/analysis/gps-broadcast (vid. 2018-08-30).

36. N. Wang, Y. Yuan, Z. Li, O. Montenbruck, and B. Tan, "Determination of differential code biases with multi-GNSS observations", *Journal of Geodesy*, vol. 90, no. 3, pp. 209–228, 2015-11. DOI: 10.1007/s00190-015-0867-4.

37. O. Montenbruck, A. Hauschild, and P. Steigenberger, "Differential code bias estimation using multi-GNSS observations and global ionosphere maps", *Navigation*, vol. 61, no. 3, pp. 191–201, 2014-09. DOI: 10.1002/navi.64.

38. (2018). IGS analysis GPS final, URL: http://www.igs.org/analysis/gps-final (vid. 2018-10-16).

39. (2018). NGA – FTP product repository, NGA GNSS Division, URL: ftp://ftp.nga.mil/pub2/ (vid. 2018-10-17).

40. European GNSS Service Centre. (2019). Galileo satellite metadata, URL: https://www.gsc-europa.eu/support-to-developers/galileo-satellite-metadata (vid. 2019-04-11).

41. F. Dilssner, T. Springer, C. Flohrer, and J. Dow, "Estimation of phase center corrections for GLONASS–M satellite antennas", *Journal of Geodesy*, vol. 84, no. 8, pp. 467–480, 2010-04. DOI: 10.1007/s00190-010-0381-7.

42. J. T. Wu, S. C. Wu, G. A. Hajj, W. I. Bertiger, and S. M. Lichten, "Effects of antenna orientation on GPS carrier phase", in *Astrodynamics 1991; Proceedings of the AAS/AIAA Astrodynamics Conference*, 1992.

43. "Recommendation ITU–R P.676–11: Attenuation by atmospheric gases", International Telecommunication Union (ITU), 2016. URL: https://www.itu.int/rec/R-REC-P.676-11-201609-I/en.

44. "Recommendation ITU–R P.618–13: Propagation data and prediction methods required for the design of Earth – space telecommunication systems", International Telecommunication Union (ITU), 2017. URL: https://www.itu.int/rec/R-REC-P.618-13-201712-I/en.

45. A. Leick, L. Rapoport, and D. Tatarnikov, *GPS Satellite Surveying*. John Wiley & Sons, Inc, 2015-03. DOI: 10.1002/9781119018612.

46. "Recommendation ITU–R P.531–13: Ionospheric propagation data and prediction methods required for the design of satellite services and systems", International Telecommunication Union (ITU), 2016. URL: https://www.itu.int/rec/R-REC-P.531-13-201609-I/en.

47. M. Hernández-Pajares, J. M. Juan, J. Sanz, *et al.*, "The ionosphere: Effects, GPS modeling and the benefits for space geodetic techniques", *Journal of Geodesy*, vol. 85, no. 12, pp. 887–907, 2011-09. DOI: 10.1007/s00190-011-0508-5.

48. J. Boehm, M. Hernández-Pajares, U. Hugentobler, *et al.*, "Models for atmospheric propagation delays", in *IERS Technical Note 36*, G. Petit and B. Luzum, Eds. Verlag des Bundesamts für Kartographie und Geodäsie, 2010, ch. 9, pp. 132–150. URL: http://www.iers.org/TN36/ (vid. 2018-09-04).

49. L. Urquhart, "An analysis of multi-frequency carrier phase linear combinations for GNSS", Department of Geodesy and Geomatics Engineering, University of New Brunswick, Tech. Rep. 263, 2009. URL: http://www2.unb.ca/gge/Pubs/TR263.pdf (vid. 2018-09-26).

50. M. Cocard, S. Bourgon, O. Kamali, and P. Collins, "A systematic investigation of optimal carrier-phase combinations for modernized triple-frequency GPS", *Journal of Geodesy*, vol. 82, no. 9, pp. 555–564, 2008-01. DOI: 10.1007/s00190-007-0201-x.

51. J. P. Collins, "An overview of inter-frequency carrier phase combinations", UNB, 1999. URL: http://gauss.gge.unb.ca/papers.pdf/L1L2combinations.collins.pdf (vid. 2018-09-26).

52. G. A. McGraw, "High resolution correlator technique for spread spectrum ranging system code and carrier multipath mitigation", US 6,687,316 B1, 2004-02-03.

53. R. van Nee, J. Siereveld, P. Fenton, and B. Townsend, "The multipath estimating delay lock loop: Approaching theoretical accuracy limits", in *Proceedings of 1994 IEEE Position, Location and Navigation Symposium - PLANS*, IEEE, 1994. DOI: 10.1109/plans.1994.303320.

54. R. van Nee, "Method of estimating a line of sight signal propagation time using a reduced-multipath correlation function", US5615232A, 1997-05-25.

55. A. J. Fernandez, J. A. Pulido, M. Wis, *et al.*, "ARTEMISA: New GNSS receiver processing techniques for positioning and multipath mitigation", in *2012 6th ESA Workshop on Satellite Navigation Technologies (Navitec 2012) & European Workshop on GNSS Signals and Signal Processing*, IEEE, 2012-12. DOI: 10.1109/navitec.2012.6423092.

56. B.-K. Choi, J.-U. Park, K. M. Roh, and S.-J. Lee, "Comparison of GPS receiver DCB estimation methods using a GPS network", *Earth, Planets and Space*, vol. 65, no. 7, pp. 707–711, 2013-07. DOI: 10.5047/eps.2012.10.003.

57. N. I. Ziedan, *GNSS Receivers for Weak Signals*. Artech House, 2006, ISBN: 1-59693-052-7.

58. D. Doberstein, *Fundamentals of GPS Receivers*. Springer New York, 2011-10-21, 348 pp., ISBN: 1461404088. URL: https://www.ebook.de/de/product/14877917/dan_doberstein_fundamentals_of_gps_receivers.html.

59. (2015). OEM638 product sheet. version 8, Novatel Inc., URL: https://www.novatel.com/assets/Documents/Papers/OEM638-PS-D17916.pdf (vid. 2018-10-17).

60. (2018-01-12). DELTA–3 datasheet. version 1.5, Javad GNSS, URL: https://www.javad.com/downloads/javadgnss/sheets/Delta-3_Datasheet.pdf (vid. 2018-10-17).

61. C4ADS, "Above us only stars, Exposing GPS spoofing in Russia and Syria", 2019-03-26. URL: https://c4ads.org/s/Above-Us-Only-Stars.pdf (vid. 2019-05-17).

62. "Nationwide differential global positioning system (NDGPS), A notice by the coast guard, the transportation department, and the engineers corps", in *Federal Register*, 2016, pp. 43 613–43 615. URL: https://www.federalregister.gov/documents/2016/07/05/2016-15886/nationwide-differential-global-positioning-system-ndgps.

63. "Discontinuance of the Nationwide Differential Global Positioning System (NDGPS), A Notice by the Coast Guard", in *Federal Register*, 2018-03-21, pp. 12 402–12 403. URL: https://www.federalregister.gov/documents/2018/03/21/2018-05684/discontinuance-of-the-nationwide-differential-global-positioning-system-ndgps.

64. Novatel Inc. (2019). An introduction to GNSS – satellite based augmentation systems, URL: https://www.novatel.com/an-introduction-to-gnss/chapter-5-resolving-errors/satellite-based-augmentation-systems/.

65. U.S. Dept. of Transportation and FAA, *Global positioning system wide area augmentation system (WAAS) performance standard*, 2008. URL: https://www.gps.gov/technical/ps/2008-WAAS-performance-standard.pdf (vid. 2018-11-05).

66. *EGNOS open service (OS) service definition document*, version 2.3, 2017. URL: https://egnos-user-support.essp-sas.eu/new_egnos_ops/sites/default/files/documents/egnos_os_sdd_in_force.pdf (vid. 2018-11-05).

67. Japan Civil Aviation Bureau, *Overview of MSAS*, 2009. URL: http://www.unoosa.org/documents/pdf/icg/activities/2007/icg2/presentations/04_01.pdf (vid. 2018-11-05).

68. IRSO, Dept. of Space. (2017). Satellite navigation programme, URL: https://www.isro.gov.in/applications/satellite-navigation-programme (vid. 2018-11-05).

69. Russian Space Systems, *System of differential correction and monitoring - interface control document, Radiosignals and digital data structure of glonass wide area augmentation system, system of differential correction and monitoring*, 2012. URL: http://www.sdcm.ru/smglo/ICD_SDCM_1dot0_Eng.pdf (vid. 2018-11-05).

70. GPS World Staff. (2018-02-05). Korea will launch its own satellite positioning system, URL: http://gpsworld.com/korea-will-launch-its-own-satellite-positioning-system/.

71. B. W. Parkinson and J. J. Spilker Jr., *Global Positioning System: Theory and Applications*. Washington: American Institute of Aeronautics and Astronautics, 1996, vol. 2, ISBN: 9781563471070.

72. T. Takasu, *RTKLIB ver.2.4.2 manual*, 2013-04. URL: http://www.rtklib.com/prog/manual_2.4.2.pdf.

73. P. J. G. Teunissen, P. J. de Jonge, and C. C. J. M. Tiberius, "The volume of the GPS ambiguity search space and its relevance for integer ambiguity resolution", in *Proceedings of the 9th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1996)*, Kansas City, MO, 1996-09-17, pp. 889–898. URL: https://www.ion.org/publications/abstract.cfm?articleID=2621 (vid. 2018-11-12).

74. P. J. Teunissen, "Mixed integer estimation and validation for next generation GNSS", in *Handbook of Geomathematics*, W. Freeden, M. Z. Nashed, and T. Sonar, Eds. Springer Berlin Heidelberg, 2010, pp. 1101–1127, ISBN: 978-3-642-01546-5. DOI: 10.1007/978-3-642-01546-5_37.

75. P. J. G. Teunissen, "Theory of carrier phase ambiguity resolution", *Wuhan University Journal of Natural Sciences*, vol. 8, no. 2, pp. 471–484, 2003-06. DOI: 10.1007/bf02899809.

76. P. de Jonge and C. Tiberius, "The LAMBDA method for integer ambiguity estimation: Implementation aspects", Delft University of Technology, Tech. Rep. 12. LGR-Series, 1996. URL: https://www.tudelft.nl/citg/over-faculteit/afdelingen/geoscience-remote-sensing/research/research-themes/positioning-navigation-and-timing-pnt/gps/lambda-method/ (vid. 2015-03-25).

77. G. Wübbena, A. Bagge, G. Seeber, V. Böder, and P. Hankemeier, "Reducing distance dependent errors for real-time precise DGPS applications by establishing reference station networks", in *Proceedings of the 9th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1996)*, Kansas City, MO, 1996-09-17, pp. 1845–1852. URL: http://www.geopp.de/pdf/kansas96.pdf (vid. 2018-11-06).

78. U. Vollath, A. Buecherl, H. Landau, C. Pagels, and B. Wagner, "Multi-base rtk positioning using virtual reference stations", in *Proceedings of the 13th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2000)*, Salt Lake City, UT, 2000-09-19, pp. 123–131. URL: http://geosoft.ee.dev.mikare.ee/uploads/userfiles/file/ION2000-Paper-MultiBase.pdf (vid. 2018-11-06).

79. U. Vollath, A. Buecherl, H. Landau, C. Pagels, and B. Wagner, "Long-range RTK positioning using virtual reference stations", in *Proceedings of the 13th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2000)*, Salt Lake City, UT, 2000, pp. 1143–1147. URL: https://www.ucalgary.ca/engo_webdocs/SpecialPublications/KIS%2001/PDF/0802.PDF (vid. 2018-11-06).

80. G. Wübbena, M. Schmitz, and A. Bagge, "PPP-RTK: Precise point positioning using state-space representation in RTK networks", in *Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2005)*, Long Beach, CA, 2005-09-16, pp. 2584–2594. URL: https://www.ion.org/publications/abstract.cfm?articleID=6467 (vid. 2018-11-01).

81. S. Gunawardena, S. Bisnath, and S. Kennedy. (2014). Precise positioning techniques: GNSS error sources & mitigation. L. Dearman, Ed. Webinar, Inside GNSS, URL: http://insidegnss.com/wp-content/uploads/2014/12/Inside_GNSS_Precise_Positioning_webinar_121514_final_secured.pdf (vid. 2018-11-10).

82. S. Choy. (2018). GNSS precise point positioning (PPP), URL: http://www.unoosa.org/documents/pdf/icg/2018/ait-gnss/16_PPP.pdf (vid. 2018-11-08).

83. *IS-GPS-800E navstar GPS space segment/user segment L1C interfaces*, 2018-04-25. URL: https://www.gps.gov/technical/icwg/IS-GPS-800E.pdf (vid. 2018-11-16).

84. K. T. Woo, "Optimum semicodeless carrier-phase tracking of L2", *Navigation*, vol. 47, no. 2, pp. 82–99, 2000-06. DOI: 10.1002/j.2161-4296.2000.tb00204.x.

85. "2017 federal radionavigation plan", U.S. Department of Defense, Department of Homeland Security, and Department of Transportation, 2017. URL: https://www.navcen.uscg.gov/pdf/FederalRadionavigationPlan2017.pdf (vid. 2018-11-13).

86. *IS-GPS-705E navstar GPS space segment/user segment L5 interfaces*, 2018-04-25. URL: https://www.gps.gov/technical/icwg/IS-GPS-705E.pdf (vid. 2018-11-16).

87. Russian Space Systems, JSC, *GLONASS ICD – general description of code division multiple access signal system*, version Edition 1.0, Moscow, 2016. URL: http://russianspacesystems.ru/wp-content/uploads/2016/08/ICD-GLONASS-CDMA-General.-Edition-1.0-2016.pdf (vid. 2018-11-06).

88. Russian Space Systems, JSC, *GLONASS ICD – code division multiple access open service navigation signal in L1 frequency band*, version Edition 1.0, Moscow, 2016. URL: http://russianspacesystems.ru/wp-content/uploads/2016/08/ICD-GLONASS-CDMA-L1.-Edition-1.0-2016.pdf (vid. 2018-11-06).

89. Russian Space Systems, JSC, *GLONASS ICD – code division multiple access open service navigation signal in L2 frequency band*, version Edition 1.0, Moscow, 2016. URL: http://russianspacesystems.ru/wp-content/uploads/2016/08/ICD-GLONASS-CDMA-L2.-Edition-1.0-2016.pdf (vid. 2018-11-06).

90. Russian Space Systems, JSC, *GLONASS ICD – code division multiple access open service navigation signal in L3 frequency band*, version Edition 1.0, Moscow, 2016. URL: http://russianspacesystems.ru/wp-content/uploads/2016/08/ICD-GLONASS-CDMA-L3.-Edition-1.0-2016.pdf (vid. 2018-11-06).

91. J.-A. Avila-Rodriguez, S. Wallner, G. Hein, *et al.*, "CBOC : An implementation of MBOC", in *CNES-ESA, 1st Workshop on GALILEO Signals and Signal Processing*, Toulouse, France, 2006-10-27. URL: https://hal-enac.archives-ouvertes.fr/hal-01021795 (vid. 2018-11-18).

92. E. Union, *European GNSS (Galileo) open service signal-in-space interface control document*, 2016. URL: https://www.gsc-europa.eu/system/files/galileo_documents/Galileo-OS-SIS-ICD.pdf (vid. 2018-11-18).

93. P. Majithiya, K. Khatri, S. C. Bera, S. Sarkar, and K. S. Parikh, "Future space service of NavIC (IRNSS) constellation", *InsideGNSS*, July/August 2017. URL: http://insidegnss.com/auto/julyaug17-IRNSS.pdf (vid. 2018-11-12).

94. IRSO Satellite Centre, *IRNSS SIS ICD for standard positioning service*, L. Mruthyunjaya, Ed., version 1.1, 2017. URL: https://www.isro.gov.in/sites/default/files/irnss_sps_icd_version1.1-2017.pdf (vid. 2018-11-12).

95. Cabinet Office, Government of Japan. (2018). Quasi-zenith satellite system (QZSS), URL: http://qzss.go.jp/en/ (vid. 2018-11-12).

96. J. S. Subirana, J. M. J. Zornoza, and M. Hernandez-Pajares. (2011). GNSS signal (Navipedia), URL: https://gssc.esa.int/navipedia/index.php/GNSS_signal (vid. 2018-11-11).

97. B. Denis, J. Keignart, and N. Daniele, "UWB measurements and propagation models for snowy environments", in *2005 IEEE International Conference on Ultra-Wideband*, Zurich, Switzerland: IEEE, 2005-09-05. DOI: 10.1109/icu.2005.1569969.

98. K. Yousef, H. Jia, R. Pokharel, *et al.*, "A 0.18 um CMOS current reuse ultra-wideband low noise amplifier (UWB-LNA) with minimized group delay variations", in *2014 44th European Microwave Conference*, IEEE, 2014-10. DOI: 10.1109/eumc.2014.6986705.

99. H. Mosalam, A. Allam, H. Jia, A. B. Abdelrahman, and R. K. Pokharel, "High efficiency and small group delay variations 0.18 um CMOS UWB power amplifier", *IEEE Transactions on Circuits and Systems II: Express Briefs*, pp. 1–1, 2018. DOI: 10.1109/tcsii.2018.2870165.

100. Decawave Ltd. (2017). DWM1000 datasheet, URL: https://www.decawave.com/sites/default/files/dwm1000-datasheet-v1.6.pdf (vid. 2018-08-25).

101. V. Djaja-Josko and J. Kolakowski, "A new transmission scheme for wireless synchronization and clock errors reduction in UWB positioning system", in *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2016-10, pp. 1–6. DOI: 10.1109/IPIN.2016.7743635.

102. V. Djaja-Josko and J. Kolakowski, "A new method for wireless synchronization and TDOA error reduction in UWB positioning system", in *2016 21st International Conference on Microwave, Radar and Wireless Communications (MIKON)*, 2016-05, pp. 1–4. DOI: 10.1109/MIKON.2016.7492077.

103. C. McElroy, D. Neirynck, and M. McLaughlin, "Comparison of wireless clock synchronization algorithms for indoor location systems", in *2014 IEEE International Conference on Communications Workshops (ICC)*, 2014-06, pp. 157–162. DOI: 10.1109/ICCW.2014.6881189.

104. V. Navrátil and J. Krška, "Určování polohy UWB-TDoA: Popis metody, výsledky a doporučení", Czech, Czech Technical University, Internal report no. TE01020186/CTU/-2017/8, 2017.

105. Y. Jiang and V. C. Leung, "An asymmetric double sided two-way ranging for crystal offset", in *2007 International Symposium on Signals, Systems and Electronics*, IEEE, 2007-07. DOI: 10.1109/issse.2007.4294528.

106. M. Kwak and J. Chong, "A new double two-way ranging algorithm for ranging system", in *2010 2nd IEEE InternationalConference on Network Infrastructure and Digital Content*, IEEE, 2010-09. DOI: 10.1109/icnidc.2010.5657814.

107. B. Merkl, A. Fathy, and M. Mahfouz, "Base station orientation calibration in 3-D indoor UWB positioning", in *2008 IEEE International Conference on Ultra-Wideband*, IEEE, 2008-09. DOI: 10.1109/icuwb.2008.4653292.

108. Taoglas Ltd., *UWC.20 specification*, 2018. URL: https://cdn.taoglas.com/datasheets/UWC.20.pdf (vid. 2019-02-09).

109. DecaWave Ltd., "WB001 hardware build instruction", Application note, 2014-11-07. URL: https://www.decawave.com/uwb-antennae-design-files/ (vid. 2019-02-11).

110. DecaWave Ltd., "APS014: Antenna delay calibration of DW1000-based products and systems.", Application note, 2014.

111. K. A. Horváth, G. Ill, and Á. Milánkovich, "Calibration method of antenna delays for UWB-based localization systems", in *2017 IEEE 17th International Conference on Ubiquitous Wireless Broadband (ICUWB)*, 2017-09, pp. 1–5. DOI: 10.1109/ICUWB.2017.8250969.

112. B. Burgess and T. Jones, "The propagation of l.f. and v.l.f. radio waves with reference to some systems applications", *Radio and Electronic Engineer*, vol. 45, no. 1-2, p. 47, 1975. DOI: 10.1049/ree.1975.0007.

113. J. Kasper and C. Hutchinson, "The omega navigation system–an overview", *IEEE Communications Society Magazine*, vol. 16, no. 3, pp. 23–35, 1978-05. DOI: 10.1109/mcom.1978.1089729.

114. E. Swanson, "Omega", *Proceedings of the IEEE*, vol. 71, no. 10, pp. 1140–1155, 1983. DOI: 10.1109/proc.1983.12743.

115. W. J. Pelgrum, "New potential of low-frequency radionavigation in the 21st century", PhD thesis, Technische Universiteit Delft, 2006-11-28, ISBN: 978-90-811198-1-8. URL: http://resolver.tudelft.nl/uuid:90450409-f146-4c45-839c-a4b484f723ff (vid. 2019-02-13).

116. Locata Corporation Pty Ltd, *Locata signal interface control document*, version 100E, 2014-01-29. URL: http://www.locata.com/icd/ (vid. 2019-02-13).

117. S. Meiyappan, A. Raghupathy, and G. Pattabiraman, "Positioning in GPS challenged locations - the NextNav terrestrial positioning constellation", in *Proceedings of the 26th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2013)*, 2013, pp. 426–431.

118. V. Navrátil, J. L. Garry, A. O'Brien, and G. E. Smith, "Utilization of terrestrial navigation signals for passive radar", in *2017 IEEE Radar Conference (RadarConf)*, 2017-05, pp. 825–829. DOI: 10.1109/RADAR.2017.7944317.

119. V. Navrátil, A. O'Brien, J. L. Garry, and G. E. Smith, "Demonstration of Space-Time Adaptive Processing for DSI Suppression in a Passive Radar", in *The 18th International Radar Symposium IRS 2017*, Praha: DGON, 2017-06.

120. V. Navrátil, J. L. Garry, A. J. O'Brien, and G. E. Smith, "Exploiting terrestrial positioning signals to enable low-cost passive radar", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 5, pp. 2246–2256, 2018-10. DOI: 10.1109/taes.2018.2812458.

121. R. Karásek and F. Vejražka, "The DVB-T-based positioning system and single frequency network offset estimation", *Radioengineering*, vol. 27, no. 4, pp. 1155–1165, 2018-09. DOI: 10.13164/re.2018.1155.

122. K. Shamaei, J. Khalife, and Z. M. Kassas, "Exploiting LTE signals for navigation: Theory to implementation", *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2173–2189, 2018-04. DOI: 10.1109/twc.2018.2789882.

123. J. J. Morales and Z. M. Kassas, "Optimal collaborative mapping of terrestrial transmitters: Receiver placement and performance characterization", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 2, pp. 992–1007, 2018-04. DOI: 10.1109/taes.2017.2773238.

124. P. Figueiredo E Silva, "Signals of opportunity for positioning purposes", PhD thesis, Tampere University of Technology, 2018-12-04, ISBN: 978-952-15-4280-0. URL: http://urn.fi/URN:ISBN:978-952-15-4280-0 (vid. 2019-02-25).

125. V. Navratil, R. Karasek, and F. Vejrazka, "Position estimate using radio signals from terrestrial sources", in *2016 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 2016-04, pp. 799–806. DOI: 10.1109/PLANS.2016.7479775.

126. D. Serant, P. Thevenon, M.-L. Boucheret, *et al.*, "Development and validation of an OFDM/DVB-T sensor for positioning", in *IEEE/ION Position, Location and Navigation Symposium*, IEEE, 2010-05. DOI: 10.1109/plans.2010.5507273.

127. L. Chen, O. Julien, P. Thevenon, *et al.*, "TOA estimation for positioning with DVB-T signals in outdoor static tests", *IEEE Transactions on Broadcasting*, vol. 61, no. 4, pp. 625–638, 2015-12. DOI: 10.1109/tbc.2015.2465155.

128. M. Rabinowitz and J. J. Spilker, "A new positioning system using television synchronization signals", *IEEE Transactions on Broadcasting*, vol. 51, no. 1, pp. 51–61, 2005-03. DOI: 10.1109/tbc.2004.837876.

129. C. Yang, T. Nguyen, D. Qiu, *et al.*, "Mobile positioning with DTV signals (ATSC 8VSB and M/H standards)", in *Proceedings of the 25th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2012)*, Nashville, TN, 2012-09-17, pp. 1208–1216.

130. J. Khalife, K. Shamaei, and Z. M. Kassas, "A software-defined receiver architecture for cellular CDMA-based navigation", in *2016 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, IEEE, 2016-04. DOI: 10.1109/plans.2016.7479777.

131. P. D. Groves, *Principles of GNSS, Inertial, and Multi-Sensor Integrated Navigation Systems*, 1st ed., ser. GNSS technology and applications series. Artech House, 2008, ISBN: 978-1-58053-255-6.

132. N. M. Barbour, "Inertial navigation sensors", in *Low-Cost Navigation Sensors and Integration Technology*, RTO-EN-SET-116(2011), NATO – Research and Technology Organisation, 2011, ch. 2, ISBN: 978-92-837-0137-8. URL: https://www.sto.nato.int/publications/STO%20Educational%20Notes/RTO-EN-SET-116-2011/EN-SET-116(2011)-02.pdf (vid. 2019-02-27).

133. D. Titterton and J. Weston, *Strapdown Inertial Navigation Technology*. IET, 2004-10-29, 576 pp., ISBN: 0863413587. DOI: 10.1049/PBRA017E.

134. E.-H. Shin, "Estimation techniques for low-cost inertial navigation", PhD thesis, The University of Calgary, 2005. URL: https://www.ucalgary.ca/engo_webdocs/NES/05.20219.EHShin.pdf (vid. 2019-03-04).

135. J. Solà, "Quaternion kinematics for the error-state Kalman filter", 2017-11-03.

136. J. Zhou, "Low-cost MEMS-INS/GPS integration using nonlinear filtering approaches", PhD thesis, Universität Siegen, 2013-04-18. URL: http://www.zess.uni-siegen.de/cms/papers/Zhou2013LowCost.pdf (vid. 2019-03-06).

137. J. Meeus, *Mathematical Astronomy Morsels*. Willmann-Bell, 1997, ISBN: 0-943396-51-4.

138. Z. Koppanyi, V. Navrátil, H. Xu, C. K. Toth, and D. Grejner-Brzezinska, "Using adaptive motion constraints to support UWB/IMU based navigation", *Navigation*, vol. 65, no. 2, pp. 247–261, 2018-04. DOI: 10.1002/navi.227.

139. S. Bancroft, "An algebraic solution of the GPS equations", *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-21, no. 1, pp. 56–59, 1985-01. DOI: 10.1109/taes.1985.310538.

140. G. A. F. Seber and C. J. Wild, *Nonlinear Regression*. John Wiley & Sons, Inc., 1989-02. DOI: 10.1002/0471725315.

141. H. P. Gavin. (2019-01-10). The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems, Duke University, URL: http://people.duke.edu/~hpgavin/ce281/lm.pdf (vid. 2019-03-27).

142. K. Levenberg, "A method for the solution of certain non-linear problems in least squares", *Quarterly of Applied Mathematics*, vol. 2, no. 2, pp. 164–168, 1944-07. DOI: 10.1090/qam/10666.

143. D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters", *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963-06. DOI: 10.1137/0111030.

144. J. Dobeš and V. Žalud, *Moderní radiotechnika*, Czech. Praha: BEN – technická literatura, 2006, ISBN: 80-7300-132-2.

145. H. B. Nielsen, "Damping parameter in Marquardt's method", DTU, Dept. of Mathematical Modelling, Tech. Rep., 1999. URL: http://www2.imm.dtu.dk/documents/ftp/tr99/tr05_99.pdf (vid. 2019-04-03).

146. M. I. A. Lourakis, "A brief description of the Levenberg-Marquardt algorithm implemened by levmar", Foundation for Research and Technology- Hellas (FORTH), 2005-02-11. URL: http://users.ics.forth.gr/~lourakis/levmar/levmar.pdf (vid. 2019-04-02).

147. P. D. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*, 2nd ed. Artech House, 2013-03-01, ISBN: 1608070050.

148. L. Nielsen, "Least-squares estimation using Lagrange multipliers", *Metrologia*, vol. 35, no. 2, pp. 115–118, 1998-04. DOI: 10.1088/0026-1394/35/2/6.

149. C. Kanzow, N. Yamashita, and M. Fukushima, "Levenberg–Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints", *Journal of Computational and Applied Mathematics*, vol. 172, no. 2, pp. 375–397, 2004-12. DOI: 10.1016/j.cam.2004.02.013.

150. D. Simon, "Kalman filtering with state constraints: A survey of linear and nonlinear algorithms", *IET Control Theory & Applications*, vol. 4, no. 8, pp. 1303–1318, 2010-08. DOI: 10.1049/iet-cta.2009.0032.

151. D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. John Wiley & Sons, 2006-06-12, 554 pp., ISBN: 0471708585. URL: https://www.ebook.de/de/product/5388550/simon_optimal_state_estimation.html.

152. S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems", in *Signal Processing, Sensor Fusion, and Target Recognition VI*, I. Kadar, Ed., SPIE, 1997-07. DOI: 10.1117/12.280797.

153. E. A. Wan and R. van der Merwe, "The unscented Kalman filter for nonlinear estimation", in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, IEEE, 2002-08-06. DOI: 10.1109/asspcc.2000.882463.

154. A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer New York, 2013-03-09. URL: https://www.ebook.de/de/product/25438357/sequential_monte_carlo_methods_in_practice.html.

155. F. Gustafsson, F. Gunnarsson, N. Bergman, *et al.*, "Particle filters for positioning, navigation, and tracking", *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 425–437, 2002. DOI: 10.1109/78.978396.

156. J. A. F. P. Closas C. Fernández-Prades, "Optimizing the likelihood with sequential Monte-Carlo methods", in *proc. XXI Simpsium Nacional de la Unin Cientfica Internacional de Radio*, 2006. URL: http://www.cttc.es/wp-content/uploads/2013/03/071116-18-closas-ursi0906-19163.pdf (vid. 2019-04-14).

157. P. Closas, C. Fernandez-Prades, and J. A. Fernandez-Rubio, "Maximum likelihood estimation of position in GNSS", *IEEE Signal Processing Letters*, vol. 14, no. 5, pp. 359–362, 2007-05. DOI: 10.1109/lsp.2006.888360.

158. F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm", *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001. DOI: 10.1109/18.910572.

159. V. Indelman, S. Williams, M. Kaess, and F. Dellaert, "Information fusion in navigation systems via factor graph based incremental smoothing", *Robotics and Autonomous Systems*, vol. 61, no. 8, pp. 721–738, 2013-08. DOI: 10.1016/j.robot.2013.05.001.

160. P. O. Vontobel, "Kalman filters, factor graphs, and electrical networks", ETH Zurich, Post-Diploma Project, 2002. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.58.4006&rep=rep1&type=pdf (vid. 2019-04-14).

161. Decawave, "APS013: The implementation of two-way ranging with the DW1000", Application note, version 2.3, 2018. URL: https://www.decawave.com/wp-content/uploads/2018/10/APS013_The-Implementation-of-Two-Way-Ranging-with-the-DW1000_v2.3.pdf (vid. 2019-04-15).

162. G. van Kempen and L. van Vliet, "Mean and variance of ratio estimators used in fluorescence ratio imaging", *Cytometry*, vol. 39, no. 4, pp. 300–305, 2000. DOI: 10.1002/(SICI)1097-0320(20000401)39:4<300::AID-CYTO8>3.0.CO;2-O.

163. V. Navrátil and R. Karásek, "Position estimate using DVB-T signal", in *Proceedings of the 20th International Scientific Student Conferenece POSTER 2016*, awarded as 3rd best poster in Communication session, Prague: Czech Technical University in Prague, 2016-05, ISBN: 978-80-01-05950-0.

164. R. Karásek and V. Navrátil, "Odhad polohy využitím DVB-T signálů", Czech, Czech Technical University, Internal report no. TE01020186/CTU/2015/??, 2018.

165. B. Bechtold, *Map-matlab*, comp. software, 2017. URL: https://github.com/bastibe/Map-Matlab (vid. 2019-04-18).

166. J. Krška and V. Navrátil, "Implementace Kalmanova filtru v aritmetice s pevnou řádovou čárkou", Czech, Czech Technical University, Internal report no. TE01020186/-CTU/2018/10, 2018.

167. Z. Koppanyi, V. Navrátil, H. Xu, C. K. Toth, and D. Grejner-Brzezinska, "UWB/IMU integration with adaptive motion constraints to support UXO mapping", in *Proceedings of the ION 2017 Pacific PNT Meeting*, Honolulu, Hawaii: ION, 2017-05, pp. 429–438.

168. C. K. Toth, Z. Koppanyi, V. Navrátil, and D. Grejner-Brzezinska, "Georeferencing in GNSS-challenged environment: Integrating UWB and IMU technologies", in *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-1/W1, 2017, pp. 175–180. DOI: 10.5194/isprs-archives-XLII-1-W1-175-2017.

Lists

List of Acronyms

ACF	Auto-Correllation Function.	*p. 13, 85*
ADF	Automatic Direction Finder.	*p. 1*
AoA	Angle of Arrival.	*p. 2, 3, 26–36, 111, 119, 122, 136*
AR	Ambiguity Resolution.	*p. 58*
ARW	Angular Random Walk.	*p. 106, 108*
AWGN	Additive White Gaussian Noise.	*p. 12, 83, 107, 108, 146, 223*
BOC	Binary-Offset Carrier.	*p. 14, 62, 64–66*
BPM	Burst Position Modulation.	*p. 68*
BPSK	Binary Phase-Shift Keying.	*p. 38, 39, 62, 63, 65, 66, 68, 82*
CAF	Cross-Ambiguity Function.	*p. 83, 84, 163, 223*
CDF	Cumulative Density Function.	*p. 74, 180*
CDMA	Code-Division Multiple-Acess.	*p. 39, 62–64, 81, 85*
CRLB	Cramer-Rao Lower Bound.	*p. 2, 3, 8–10, 12, 13, 15, 17, 20, 22, 28–36, 68, 70, 82, 136*
DCB	Differential Code Bias.	*p. 44, 49*
DGNSS	Differential GNSS.	*p. 55, 57–59*
DGPS	Differential GPS.	*p. 41, 55, 56, 58*
DLL	Delay-Locked Loop.	*p. 39, 40, 49, 50, 83*
DOP	Dilution of Precision.	*p. 53, 54, 69, 72, 80, 81, 85, 120–122, 170*
DVB-T	Digital Video Broadcasting — Terrestrial.	*p. 4, 5, 22, 82, 84, 85, 137, 163, 165*
ECEF	Earth-Centered, Earth-Fixed; global coordinate frame.	*p. 96–99, 103, 104*
ECI	Earth-Centered inertial; global coordinate frame.	*p. 95, 96, 99, 103*
EKF	Extended Kalman Filter.	*p. 4, 5, 52, 75, 124, 126–129, 131–133, 191, 193*
ENU	East-North-Up; local coordinate frame.	*p. 98, 121, 164, 168, 169*
ERP	Equivalent Radiated Power (w.r.t. half-wave dipole).	*p. 81*
ESKF	Error-State Kalman Filter.	*p. 5, 136, 193–199, 201*
FDMA	Frequency-Division Multiple-Acess.	*p. 39, 63, 64*
FDoA	Frequency Difference of Arrival.	*p. 23*
FLL	Frequency-Locked Loop.	*p. 39, 83*
FSL	Free-Space Loss.	*p. 7*
GLONASS	GLObalnaya NAvigatsionnaya Sputnikovaya Sistema.	*p. 38, 41, 43, 63, 64*
GNSS	Global Satellite Navigation System(s).	*p. 3, 5, 14, 20, 37–67, 71, 72, 79–82, 85, 87, 96, 98, 99, 105, 106, 116, 120, 122, 124, 134, 136, 167, 191, 193, 194, 199–201, 204*

GPS	Global Positioning System.	*p. 38–41, 43, 44, 47, 49, 50, 62–64, 66, 81*
GPST	GPS Time.	*p. 43, 62, 66, 191*
GST	Galileo System Time.	*p. 43*
HDOP	Horizontal Dilution of Precision.	*p. 54, 165, 167*
HF	High Frequency ($f = 3 - 30\,\mathrm{MHz}$; $\lambda = 10 - 100\,\mathrm{m}$).	*p. 80*
ICD	Interface Control Document.	*p. 39, 43, 47, 64, 66*
IGS	International GNSS Service.	*p. 41, 43, 44, 60*
IGSO	Inclined Geosynchronous Orbit.	*p. 65*
IMU	Inertial Measurement Unit.	*p. 87, 98–103, 105–109, 124, 134, 136, 191–193, 198–201, 204*
IPP	Ionospheric Pierce Point.	*p. 47*
IRNSS	Indian Regional Navigation Satellite System.	*p. 65, 66*
KF	Kalman Filter (linear).	*p. 4, 5, 73, 124–128, 132–134, 176, 179–181, 183, 185–187, 194, 223, 224*
LAMBDA	Last-squares AMBiguity Decorrelation Adjustment.	*p. 59*
LF	Low Frequency ($f = 30 - 300\,\mathrm{kHz}$; $\lambda = 1 - 10\,\mathrm{km}$).	*p. 79*
LFM	Linear Frequency Modulation.	*p. 14*
LLI	Loss-of-Lock Indicator.	*p. 41*
M-PSK	M-ary Phase-Shift Keying.	*p. 38*
MEDLL	Multipath-Estimating Delay-Locked Loop.	*p. 49*
MEMS	Micro-Electro-Mechanical Systems.	*p. 87, 107–109, 136, 191, 192, 201*
MEO	Medium Earth Orbit.	*p. 37, 43, 51, 65*
NDB	Non-Directional Beacon.	*p. 1*
NED	North-East-Down; local coordinate frame.	*p. 96, 98, 99, 192*
NGA	National Geospatial-Intelligence Agency.	*p. 44*
NN	Neural Network.	*p. 196, 200*
OFDM	Orthogonal Frequency-Division Multiplexing.	*p. 84, 85*
PDF	Probability Density Function.	*p. 9, 30–32, 134*
PF	Particle Filter.	*p. 134*
PLL	Phase-Locked Loop.	*p. 39–41, 47, 50, 59, 83*
ppb	Parts per Billion.	*p. 178, 184*
ppm	Parts per Million.	*p. 19, 59, 71, 73, 75, 106, 176*
PPP	Precise Point Positioning.	*p. 60, 61, 66*
PSD	Power Spectral Density.	*p. 12–14, 68, 84, 223*
PVA	Position, Velocity and Attitude.	*p. 193, 194, 198*

PVT	Position, Velocity and Time.	*p. 37, 41, 49, 50, 52*
QPSK	Quadrature Phase-Shift Keying.	*p. 63, 65*
QZSS	Quasi-Zenith Satellite System.	*p. 66*
RF	Radio Frequency.	*p. 30, 68*
RINEX	The Receiver Independent Exchange Format.	*p. 41*
RMS	Root Mean Square; quadratic average.	*p. 12, 36, 40, 44, 47, 50, 74, 82, 84, 85, 180, 181, 200*
RTCM	Radio Technical Commission for Maritime Services.	*p. 41, 42*
RTK	Real-Time Kinematics.	*p. 41, 57, 59–61, 64, 191*
SBAS	Satellite-based Augmentation System.	*p. 47, 56, 57*
SLAM	Simultaneous Localisation And Mapping.	*p. 11*
SNR	Signal-to-Noise Ratio.	*p. 12–14, 30, 31, 33, 41, 42, 68, 73, 84*
SoO	Signals of Opportunity.	*p. 4, 5, 82–85*
STEC	Slant Total Electron Content.	*p. 46*
TDMA	Time-Division Multiple-Acess.	*p. 39, 81*
TDoA	Time Difference of Arrival, hyperbolic navigation.	*p. 3–5, 21–23, 26, 69–73, 75, 79, 80, 82, 83, 85, 111, 118–119, 122, 123, 135, 137, 163–170, 175, 183, 222*
TEC	Total Electron Content.	*p. 46, 47*
ToA	Time of Arrival.	*p. 3, 19–22, 26, 37, 62, 72, 73, 82, 111, 116–117, 119, 122, 167, 175*
TWR	Two-Way Ranging.	*p. 3, 5, 14–19, 69, 70, 73, 75, 76, 111, 117–119, 122, 135, 141–159, 183, 191*
UHF	Ultra High Frequency ($f = 300\,\mathrm{MHz} - 3\,\mathrm{GHz}$; $\lambda = 10\,\mathrm{cm} - 1\,\mathrm{m}$).	*p. 37, 80*
UKF	Unscented Kalman Filter.	*p. 4, 5, 52, 124, 128–132, 134*
UT	Unscented Transformation.	*p. 128–131*
UTC	Temps Universel Coordonné; Universal coordinated time.	*p. 62, 63*
UWB	Ultra-Wide Band.	*p. 3–5, 8, 34, 36, 68–79, 106, 122–124, 135, 136, 141, 145, 147, 148, 167–170, 175, 181, 183, 191, 193, 194, 199–201*
UXO	Unexploded ordnance.	*p. 4, 191*
VDOP	Vertical Dilution of Precision.	*p. 54, 81, 170*
VLF	Very Low Frequency ($f = 3 - 30\,\mathrm{kHz}$; $\lambda = 10 - 100\,\mathrm{km}$).	*p. 79, 80*
VRW	Velocity Random Walk.	*p. 106, 108*
VTEC	Vertical Total Electron Content.	*p. 46*
ZUPT	Zero-velocity update.	*p. 106, 195, 196*

List of Symbols and Operators

$\boldsymbol{a}^{\mathrm{b}}$	Specific force vector (in b-frame).
$\mathrm{Av}\{\cdot\}$	Average value, often w.r.t. time.
b	Receiver clock bias (in ToA, GNSS context) .
B	Bandwidth; B_{RMS} – RMS bandwidth.
β	Bearing angle.
c_0	Speed of light in vacuum.
$\mathrm{cov}\,(A, B)$	Covariance of random variables; $\mathrm{cov}\,(A, B) = \mathrm{E}\left[(A - \mathrm{E}[A])(B - \mathrm{E}[B])\right]$.
$d_{\mathrm{i,j}}$	Time difference of Arrival (TDoA) from/to i-th and j-th node.
$\mathrm{diag}\,\cdot$	Vector of diagonal elements of a matrix.
$\mathrm{Diag}\,\cdot$	Diagonal matrix, may be constructed from a vector or a matrix.
$\mathrm{E}[\cdot]$	Expected value, expectation; probability-weighted average.
$\mathcal{E}$	Energy.
f	Frequency.
$\boldsymbol{F}$	KF state transition matrix.
G	Gain (e.g. antenna gain).
$\boldsymbol{H}$	KF measurement model matrix.
I_m	Modified Bessel function of the first kind and order m.
$\mathbb{I}$	Identity matrix, eye matrix.
$\boldsymbol{\mathcal{I}}$	Fisher information matrix.
j	Imaginary unit; $\mathrm{j} = \sqrt{-1}$.
$\boldsymbol{K}$	Kalman gain.
$\mathcal{L}(\theta\|x)$	Likelihood function of the parameter θ given that the outcome is x.
$\Lambda(\theta\|x)$	Log-Likelihood function, $\ln \mathcal{L}(\theta\|x)$.
λ	Wavelength; Coefficient of the Levenberg-Marquardt iterative method.
$\ln(\cdot)$	Natural logartithm.
$\log_b(\cdot)$	Logartithm of base b, base-10 logarithm if no base is specified.

n	Path loss exponent (one-slope model); Refractive index; Sample index.
ν	Clock drift (relative).
$\boldsymbol{\nu}$	Rotation vector.
N_0	Constant level of the twosided PSD of AWGN.
$\mathcal{N}(\mu\,\sigma^2)$	Normally distributed random variable with mean μ and variance σ^2.
$\boldsymbol{\omega}^{\mathrm{b}}$	Angular rate vector (in b-frame).
P	Power (e.g. signal power).
$\boldsymbol{P}$	KF state covariance matrix.
$\boldsymbol{\Phi}_{\boldsymbol{x}}$	Variance-covariance matrix of vector $\boldsymbol{x}$.
ϕ	Accumulated delta range (Carrier phase in meters); Roll angle (attitude).
φ	Carrier phase in radians.
ψ	Yaw angle (attitude).
$\Psi(\tau, f_{\mathrm{D}})$	Cross-ambiguity function (CAF); function of delay and frequency shift.
$\boldsymbol{Q}$	KF state process noise covariance matrix.
$\mathring{\boldsymbol{q}}_{\mathrm{a}}^{\mathrm{b}}$	Quaternion for transformation from a-frame to b-frame.
$\mathring{\boldsymbol{q}}\{\cdot\}$	Quaternion equivalent to a different rotation representation.
r_{ab}	Range (between a and b).
$\boldsymbol{r}_{\mathrm{a}}$	Position vector (of a).
$r_{\mathrm{a};x}$	x-axis component of $\boldsymbol{r}_{\mathrm{a}}$.
$R_{a,b}$	Cross-correlation function of a and b; R_a denotes Auto-corr. function of a.
$\boldsymbol{R}$	KF measurement noise covariance matrix.
$\boldsymbol{R}_{\mathrm{a}}^{\mathrm{b}}$	Rotation matrix for transformation from a-frame to b-frame.
$\boldsymbol{R}\{\cdot\}$	Rotation matrix equivalent to a different rotation representation.
ρ	Pseudorange.
s	Signal.
S	Signal spectrum.
$\mathcal{S}$	Power spectral density.

θ	Pitch angle (attitude).
t	Time (continuous).
τ	Delay.
$\mathrm{tr}\,(\cdot)$	Trace of a matrix.
$\boldsymbol{v}$	Velocity vector.
var	Variance; $\mathrm{var}A = \mathrm{E}\left[(A - \mathrm{E}[A])^2\}\right]$.
$\boldsymbol{x}$	KF state vector.
$\boldsymbol{y}$	KF measurement vector.
$\mathbb{1}_{\mathrm{B,A}}$	Unitary direction vector from A to B.
$\mathbf{0}$	Column vector of zeros.
$\mathbb{O}$	Matrix of zeros.
(a, b)	Open interval; for all $x \in (a, b)$ it holds $a < x < b$.
$\langle a, b\rangle$	Closed interval; for all $x \in \langle a, b\rangle$ it holds $a \leq x \leq b$.
$\boldsymbol{a} \cdot \boldsymbol{b}$	Dot (scalar) product of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$; Occasionally used as multiplication sign between two scalar numbers.
$\boldsymbol{a} \times \boldsymbol{b}$	Cross product of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.
$\boldsymbol{p} \otimes \boldsymbol{q}$	Quaternion product product of two quaternions $\boldsymbol{p}$ and $\boldsymbol{q}$.
$a \circledast b$	Convolution of two functions a and b.
$a \propto b$	Variable a is proportional to variable b.

List of Figures

1.1. Aircraft navigated towards a beacon. 2
1.2. Mindmap of the thesis contents. 5

2.1. CRLB of range est. based on power measurement 11
2.2. Single-sided two-way ranging timing diagram. 15
2.3. Double-sided two-way ranging timing diagram. 16
2.4. Geometry of Doppler effect in two dimensions. 24
2.5. Parametric plots of Doppler values. 25
2.6. Visualization of the constant-Doppler cones. 26
2.7. Approximation of antenna patterns 27
2.8. CRLB of signal-power-based AoA. 33
2.9. Schematic description of the principal assumption on a linear array. 35
2.10. CRLB of delay-based AoA for raised-cosine pulse. 36
2.11. Schematic description of the principal assumption on a linear array. 42
2.12. Graphical representation of DOP in two dimensions. 54
2.13. Measurement schematic for RTK double differences. 58
2.14. Comparison of PPP and RTK and DGNSS (according to [82]). 61
2.15. Summary of available GNSS signals 67
2.16. Range and delay CRLB of a root-raised cosine pulse. 70
2.17. Comparison of ToA and TDoA problem convergence in UWB. 74
2.18. Sources of HW delays in UWB transceiver. 76
2.19. Methods of UWB transceiver hardware delay calibration. 77

3.1. IMU measurement axes. 89
3.2. Relations of coordinate frames. 96
3.3. Latitude in the meridian plane of the reference ellipsoid. 98
3.4. Propagation of IMU measurement errors to the attitude errors. 110
3.5. Propagation of IMU measurement errors to the velocity errors. 110
3.6. Propagation of IMU measurement errors to the position errors. 110

4.1. 1σ-error ellipse and its parameters. 121
4.2. Reducing linearization error by partitioning of the update step. 128
4.3. Mean and covariance propagation via nonlinear function 130

A.1. Normalized bias of r_1, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}1}$. 149
A.2. Normalized bias of r_2, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}1}$. 149
A.3. Normalized bias of r_3, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}1}$. 149
A.4. Normalized standard deviation of r_1, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}1}$. 150
A.5. Normalized standard deviation of r_2, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}1}$. 150
A.6. Normalized standard deviation of r_3, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}1}$. 150
A.7. Double-sided two-way ranging timing diagram with 4 messages. 151
A.8. Normalized bias of r_1, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}2}$. 153
A.9. Normalized bias of r_2, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}2}$. 153
A.10. Normalized bias of r_3, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}2}$. 153
A.11. Normalized standard deviation of r_1, assuming $\mathbf{\Phi}_{\boldsymbol{\tau}2}$. 154

A.12. Normalized standard deviation of r_2, assuming $\mathbf{\Phi}_{\tau 2}$. 154
A.13. Normalized standard deviation of r_3, assuming $\mathbf{\Phi}_{\tau 2}$. 154
A.14. Normalized bias of r_1, assuming $\mathbf{\Phi}_{\tau 3}$. 157
A.15. Normalized bias of r_2, assuming $\mathbf{\Phi}_{\tau 3}$. 157
A.16. Normalized bias of r_3, assuming $\mathbf{\Phi}_{\tau 3}$. 157
A.17. Normalized standard deviation of r_1, assuming $\mathbf{\Phi}_{\tau 3}$. 158
A.18. Normalized standard deviation of r_2, assuming $\mathbf{\Phi}_{\tau 3}$. 158
A.19. Normalized standard deviation of r_3, assuming $\mathbf{\Phi}_{\tau 3}$. 158

B.1. CAF of a DVB-T signal [125]. 166
B.2. Constellation of DVB-T transmitters, HDOP 166
B.3. Unconstrained TDoA-UWB position estimates in horizontal plane. 171
B.4. Constrained TDoA-UWB position estimates in horizontal plane. 171
B.5. Time series of unconstrained TDoA-UWB position estimate 172
B.6. Time series of constrained TDoA-UWB position estimate 173

C.1. Comparison of clock offset (x_0) estimate errors. 179
C.2. Comparison of clock drift (x_1) estimates and respective errors. 180
C.3. Clock drift rate (x_2) estimated by three-state KF. 180
C.4. Serial multi-hop test. 182
C.5. Parallel branches test. 182
C.6. Filtered TDoA-UWB positions in horizontal plane – stationary model. . . 188
C.7. Filtered TDoA-UWB positions in horizontal plane – linear motion model. 188
C.8. Time series of filtered TDoA-UWB positions (stationary) 189
C.9. Time series of filtered TDoA-UWB positions (linear motion) 190

D.1. Block diagram of the integrated system. 197
D.2. Motion pattern of the platform. 197
D.3. Constraints for different motion types. 197
D.4. Open-sky area trajectory estimate results. 202
D.5. Canopied area UWB/IMU trajectory estimate results. 203
D.6. Canopied area trajectory overview. 204

List of Tables

2.1. Common values of channel model coefficients [2, 7] 9
2.2. Approx. beamwidth and directivity for $\cos^n$ main lobe shape 28
2.3. Directivity pattern shape functions and their derivatives 31
2.4. DOP coefficients in East-North-Up coordinate frame. 55
2.5. Error magnitudes in typical PPP processing (according to [81]). 60
2.6. Comparison of TWR and TDoA approach in UWB localization systems. . . 70

3.1. IMU performance grades according to [132] 88
3.2. Conversion of rotation matrix to quaternion 93
3.3. Parameters of MEMS IMU error simulation. 107

4.1. Probability of true value laying within the error ellipse. 121

A.1. Parameters of the Monte–Carlo simulation. 147

B.1. Experimental results of DVB-T – TDoA positioning in Prague [121]. 167

C.1. Parameters of the proposed three-state synchronization KF. 178
C.2. RMS values of the clock offset error and their comparison. 181
C.3. Comparison of synchronization in serial multi-hop setup. 184
C.4. Comparison of synchronization of slave node in parallel setup. 184

D.1. UWB/IMU trajectory estimate errors in the open-sky area scenario. 204
D.2. UWB/IMU trajectory estimate errors in the canopied area scenario. 204

List of Author's Publications and Citations

Unless explicitly noted, the authorship is divided equally among the listed authors.

Publications Directly Related to Dissertation Thesis

Articles in Journals (with IF) excerpted in WoS

A1. Z. Koppanyi, V. Navrátil, H. Xu, C. K. Toth, and D. Grejner-Brzezinska, "Using adaptive motion constraints to support UWB/IMU based navigation", *Navigation*, vol. 65, no. 2, pp. 247–261, 2018-04. DOI: 10.1002/navi.227.

A2. V. Navrátil and F. Vejražka, "Bias and variance of the asymmetric double-sided two-way ranging", *Navigation*, 2019, Accepted on 2019-04-19.

Other publications excerpted in WoS

B1. V. Navrátil, R. Karásek, and F. Vejražka, "Position estimate using radio signals from terrestrial sources", in *2016 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 2016-04, pp. 799–806. DOI: 10.1109/PLANS.2016.7479775.

B2. V. Navrátil, J. Krška, F. Vejražka, and V. Koreček, "Chained wireless synchronization algorithm for UWB-TDOA positioning", in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, IEEE, 2018-04. DOI: 10.1109/plans.2018.8373376, [35 % authorship].

Other publications not excerpted in WoS

C1. V. Navrátil, J. Krška, and F. Vejražka, "Experimental evaluation of chained synchronization for UWB systems", in *Proceedings of the 16th IAIN World Congress 2018 (IAIN2018)*, (2018-11-26), Chiba, Japan, 2018, pp. 137–145.

C2. Z. Koppanyi, V. Navrátil, H. Xu, C. K. Toth, and D. Grejner-Brzezinska, "UWB/IMU integration with adaptive motion constraints to support UXO mapping", in *Proceedings of the ION 2017 Pacific PNT Meeting*, Honolulu, Hawaii: ION, 2017-05, pp. 429–438.

C3. C. K. Toth, Z. Koppanyi, V. Navrátil, and D. Grejner-Brzezinska, "Georeferencing in GNSS-challenged environment: Integrating UWB and IMU technologies", in *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-1/W1, 2017, pp. 175–180. DOI: 10.5194/isprs-archives-XLII-1-W1-175-2017.

C4. V. Navrátil and R. Karásek, "Position estimate using DVB-T signal", in *Proceedings of the 20th International Scientific Student Conferenece POSTER 2016*, awarded as 3rd best poster in Communication session, Prague: Czech Technical University in Prague, 2016-05, ISBN: 978-80-01-05950-0.

Other Publications within Radio Engineering Field

Articles in Journals (with IF) excerpted in WoS

D1. V. Navrátil, J. L. Garry, A. J. O'Brien, and G. E. Smith, "Exploiting terrestrial positioning signals to enable low-cost passive radar", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 5, pp. 2246–2256, 2018-10. DOI: 10.1109/taes.2018.2812458.

Other publications excerpted in WoS

E1. V. Navrátil, J. L. Garry, A. O'Brien, and G. E. Smith, "Utilization of terrestrial navigation signals for passive radar", in *2017 IEEE Radar Conference (RadarConf)*, 2017-05, pp. 825–829. DOI: 10.1109/RADAR.2017.7944317.

E2. V. Navrátil, A. O'Brien, J. L. Garry, and G. E. Smith, "Demonstration of Space-Time Adaptive Processing for DSI Suppression in a Passive Radar", in *The 18th International Radar Symposium IRS 2017*, Praha: DGON, 2017-06.

E3. J. Dobeš, D. Černý, F. Vejražka, and V. Navrátil, "Comparing the steady-state procedures based on epsilon-algorithm and sensitivity analysis", in *2015 IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*, 2015-12, pp. 601–604. DOI: 10.1109/ICECS.2015.7440388, [15 % authorship].

E4. D. Černý, J. Dobeš, and V. Navrátil, "Functional chaining mechanism allowing definable models of electronic devices", in *ECCTD - European Conference on Circuit Theory and Design*, 2017, [15 % authorship].

Citations

[B1] was cited in:
J. Sadowski and J. Stefanski, "Asynchronous phase-location system", *Journal of Marine Engineering & Technology*, vol. 16, no. 4, pp. 400–408, 2017-02. DOI: 10.1080/20464177.2017.1376372

[E2] was cited in:
K. Strom, O. Lie-Svendsen, I. Norheim-Nass, *et al.*, "DSI suppression with adaptive beamforming in DVB-T passive radar measurements", in *2018 19th International Radar Symposium (IRS)*, IEEE, 2018-06. DOI: 10.23919/irs.2018.8448023

[E2] was also cited in:
T. Petö and R. Seller, "Space-time adaptive cancellation in passive radar systems", *International Journal of Antennas and Propagation*, vol. 2018, pp. 1–16, 2018. DOI: 10.1155/2018/2467673

This thesis was typeset by the author
in Latin Modern font family
using LaTeX.