MAT-45806 Mathematics for Positioning
TKT-2546 Methods for Positioning

Simo Ali-Löytty
Jussi Collin
Helena Leppäkoski
Hanna Sairo
Niilo Sirola

2008

Preface

Positioning techniques and algorithms have been studied for some years at the Tampere University of Technology within several research groups. This course hand-out brings together the expertise of researchers of the TUT Department of Computer Systems and the TUT Department of Mathematics. The course is divided into two parts: *MAT-45806 Mathematics for Positioning*, that covers Sections 1 and 2, and *TKT-2546 Methods for Positioning*, Sections 3–5.

Our objective has been to collect together most important algorithms and mathematical tools used in positioning including examples and starting from the basics. We do not go into details of specialized techniques and equipments, but after this course student should be able to solve application dependent problems without having to "re-invent the wheel" again and again.

Prerequisites are first-year engineering mathematics and basics of probability. Additionally, the course *TKT-2536 Introduction to Satellite Positioning* is a useful but not compulsory prerequisite. There is no official course text book in addition to this hand-out, mostly because the authors have not managed to find a single book to cover all the material on the level of abstraction we need. The battery of positioning computation methods is collected from different areas of mathematics and engineering sciences, and there are often school and interpretation differences between them, so we have tried to use common notations and represent connections between different ways of thinking as best as we could.

The authors would like to thank Sami Tiainen for the initial translation of the manuscript, and professor Robert Piché for suggesting various corrections.

<div align="right">

Tampere, January 29, 2008

the authors

</div>

Contents

Chapter 1

Preliminaries

SIMO ALI-LÖYTTY

In this chapter, we review the mathematical concepts and tools necessary for digesting the rest of the material. The reader is assumed to be familiar with most of the material, and is encouraged to look up some of the cited references if this is not the case. Mathematically speaking, positioning is finding the "best"[*] estimator for state $x \in \mathbb{R}^n$ using the equation

$$y = f(x) + \varepsilon. \tag{1.1}$$

Here y is a vector that contains measurements and ε is called the error (unknown). An important special case of equation (1.1) is when function f is linear; this special case is handled in Section 1.2. Often the error term ε and possibly also the state x are modeled as random variables. Because of this we review some probability theory in section 1.3. At the end of the chapter we handle coordinate systems in 1.4 and moving coordinate systems in 1.5, which are a natural part of positioning. First of all we give short introduction to linear algebra in 1.1.

1.1 Linear Algebra

Here we give short introduction to linear algebra, which is covered in detail in, for example, [31, 32]. The following lists some properties of a matrix $A \in \mathbb{R}^{m \times n}$.

- The null space of matrix A is $\mathcal{N}(A) = \{x \in \mathbb{R}^n | Ax = 0\}$.

- The column space of matrix A is $\mathcal{R}(A) = \{y \in \mathbb{R}^m | y = Ax; x \in \mathbb{R}^n\}$.

- $\dim(\mathcal{R}(A^\mathrm{T})) = \dim(\mathcal{R}(A)) = \mathrm{rank}(A)$.

---

[*]It is not at all obvious what the word "best" means. One of the tasks of the mathematical modeller is to define a criterion to compare estimators. Often this criterion is a so-called cost function, for instance in the form $\|y - f(\hat{x})\|^2$ or $\mathrm{E}(\|x - \hat{x}\|^2)$.

- $\dim(\mathcal{N}(A)) + \operatorname{rank}(A) = n$ (The Rank Theorem).

- A is orthogonal if $A^T A = I$.

- $A \in \mathbb{R}^{n \times n}$ is symmetric if $A^T = A$.

- $A \in \mathbb{R}^{n \times n}$ is idempotent if $AA = A$.

- $A \in \mathbb{R}^{n \times n}$ and $Ax = \lambda x$, where $x \neq 0$. Now λ is an eigenvalue of the matrix A and the corresponding eigenvector is x.

Example 1. *Let* A *be idempotent and* λ *is arbitrary eigenvalue of* A *corresponding eigenvector x. Now*

$$\lambda x = Ax = AAx = \lambda^2 x \Longrightarrow \lambda = 1 \text{ or } \lambda = 0,$$

so eigenvalues of an idempotent matrix A *are all either zero or one.*

Let matrix $A \in \mathbb{R}^{n \times n}$ be symmetric. Then

- A is positive definite, denoted $A > 0$, if $x^T A x > 0$ for all $x \neq 0$.

- A is positive semi-definite, denoting $A \geq 0$, if $x^T A x \geq 0$ for all x.

- $A > B$ is interpreted as $A - B > 0$, and $A \geq B$ as $A - B \geq 0$.

Example 2. *If* $A \geq 0$ *and* λ *is an arbitrary eigenvalue of matrix* A *corresponding to eigenvector x, then*

$$\lambda x = Ax \Longrightarrow \lambda \|x\|^2 = x^T A x \geq 0 \Longrightarrow \lambda \geq 0,$$

so all eigenvalues of a positive semi-definite matrix are non-negative.

Theorem 1 (Schur decomposition)**.** *Let* $A \in \mathbb{R}^{n \times n}$ *be symmetric. Then there is an orthogonal square matrix* Q *and diagonal matrix* Λ *so that*

$$A = Q\Lambda Q^T. \tag{1.2}$$

Because Q is orthogonal, the inverse matrix of Q is Q^T. From Eq (1.2) follows that $AQ = Q\Lambda$, so diagonal elements of diagonal matrix Λ are eigenvalues of matrix A and columns of matrix Q are normalized eigenvectors corresponding diagonal elements of Λ.

Definition 1 (Square root of matrix $A \geq 0$)**.** *Using Schur decomposition we get*

$$A = Q\lceil \lambda_1, \ldots \lambda_n \rfloor Q^T.$$

where $\lambda_i \geq 0$ *for all* $i \in \{1, \ldots, n\}$ *(Example 2). We define the square root* * *of matrix* A *as*

$$A^{\frac{1}{2}} = Q\lceil \sqrt{\lambda_1}, \ldots \sqrt{\lambda_n} \rfloor Q^T.$$

---

*Usually matrix B is called a square root of matrix A if $A = BB$. In some cases also matrix C is called a square root of matrix A if $A = CC^T$. Neither one of the above matrices (B or C) is unique. Notice that our definition of the matrix $A^{\frac{1}{2}}$ fulfills both definitions.

1.2 Overdetermined linear system of equations

Consider the linear system of equations

$$Ax = y, \tag{1.3}$$

where $A \in \mathbb{R}^{m \times n}$ and $y \in \mathbb{R}^m$. Assume that the columns of the matrix A are linearly independent i.e.

$$Ax = 0 \text{ only if } x = 0.$$

If $m > n$, the system is called overdetermined, which can happen when for instance there are more measurements than unknowns. Generally the overdetermined equation has no exact solution, in which case one approach is to search for the least squares solution

$$\hat{x} = \operatorname{argmin}_x \|y - Ax\|^2. \tag{1.4}$$

Because (see Exercise 1.3)

$$\|y - Ax\|^2 = \|A((A^T A)^{-1} A^T y - x)\|^2 + \|y - A(A^T A)^{-1} A^T y\|^2, \tag{1.5}$$

the solution of Equation (1.4) is

$$\hat{x} = (A^T A)^{-1} A^T y. \tag{1.6}$$

If the system $Ax = y$ is to be solved with Matlab software, it is recommended to use backslash-command `A\y` instead of computing the inverses explicitly.

Example 3. *Let y be measurement and initial guess is x_0. Compute estimate $\hat{x}$ that minimizes $\|y - Hx\|^2 + \|x_0 - x\|^2$.*

$$\hat{x} = \operatorname{argmin}_x (\|y - Hx\|^2 + \|x_0 - x\|^2)$$

$$= \operatorname{argmin}_x \left(\left\| \begin{bmatrix} y \\ x_0 \end{bmatrix} - \begin{bmatrix} H \\ I \end{bmatrix} x \right\|^2 \right)$$

$$\overset{(1.6)}{=} (H^T H + I)^{-1} (H^T y + x_0).$$

Example 4. *We get following two-dimensional $\begin{bmatrix} x_i \\ y_i \end{bmatrix}$ measurements*

$$\left\{ \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \right\}$$

Estimate parameters of equation $y = ax + b$ so that error $\sum_{i=1}^{5} \|y_i - (ax_i + b)\|^2$ is as small as possible.

Now

$$\sum_{i=1}^{5}\|y_i - (ax_i + b)\|^2 = \left\| \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \right\|^2 = \|y - \mathrm{A}z\|^2$$

It follows from Eq (1.6) that solution is

$$\hat{z} = (\mathrm{A}^T\mathrm{A})^{-1}\mathrm{A}^T y = \begin{bmatrix} 6 & 2 \\ 2 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 4 \end{bmatrix} = \frac{1}{26} \begin{bmatrix} 5 & -2 \\ -2 & 6 \end{bmatrix} \begin{bmatrix} 5 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{17}{26} \\ \frac{14}{26} \end{bmatrix}.$$

So the approximated equation is $y = \frac{17}{26}x + \frac{7}{13}$.

1.3 Probability theory

In this section, we briefly review the basics of probability theory focus being on normal distribution and conditional probability. For a proper definition of a random variable, see [44, 14]. In this section, a random variable is denoted in boldface, e.g. $\mathbf{x}$, and can be either vector or scalar. (In later sections, it should be clear from the context whether boldface is used to refer to a vector or to a random variable.) The density function of random variable $\mathbf{x}$ is denoted as $f_{\mathbf{x}}(x)$ or $p_{\mathbf{x}}(x)$ and cumulative function as $F_{\mathbf{x}}(x)$. The independence of random variables is a central concept: it is needed for instance in filtering in Chapter 3.

Definition 2 (Independence). *Random variables $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are independent if*

$$F_{\mathbf{x}_1, \ldots, \mathbf{x}_k}(x_1, \ldots, x_k) = \prod_{i=1}^{k} F_{\mathbf{x}_i}(x_i), \quad \forall x_1, \ldots, x_n \in \mathbb{R}^n.$$

Theorem 2. *Random variables $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are independent if and only if*

$$f_{\mathbf{x}_1, \ldots, \mathbf{x}_k}(x_1, \ldots, x_k) = \prod_{i=1}^{k} f_{\mathbf{x}_i}(x_i), \quad \forall x_1, \ldots, x_n \in \mathbb{R}^n.$$

Example 5. *Let the density function of a two-dimensional random variable $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2]$ be*

$$f_{\mathbf{x}_1, \mathbf{x}_2}(x_1, x_2) = \begin{cases} \frac{1}{\pi}, & \text{when} \quad x_1^2 + x_2^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

Are the random variables $\mathbf{x}_1$ and $\mathbf{x}_2$ independent?

Now the density functions of the marginal distributions are

$$f_{\mathbf{x}_1}(x_1) = \int_{-\infty}^{\infty} f_{\mathbf{x}_1, \mathbf{x}_2}(x_1, x_2) dx_2 = \begin{cases} \frac{2}{\pi}\sqrt{1 - x_1^2}, & \text{when} \quad |x_1| \leq 1 \\ 0, & \text{otherwise} \end{cases} \text{ and}$$

$$f_{\mathbf{x}_2}(x_2) = \int_{-\infty}^{\infty} f_{\mathbf{x}_1, \mathbf{x}_2}(x_1, x_2) dx_1 = \begin{cases} \frac{2}{\pi}\sqrt{1 - x_2^2}, & \text{when} \quad |x_2| \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

Because

$$f_{\mathbf{x}_1}(0)f_{\mathbf{x}_2}(0) = \frac{4}{\pi^2} \neq \frac{1}{\pi} = f_{\mathbf{x}_1,\mathbf{x}_2}(0,0),$$

it follows from Theorem 2 that the random variables $\mathbf{x}_1$ and $\mathbf{x}_2$ are not independent.

Definition 3 (Expectation). *Assume that $\mathbf{x}$ is a continuous random variable and that the integral*

$$\int_{-\infty}^{\infty} |g(u)| f_{\mathbf{x}}(u) du$$

converges. Then the expectation of the random variable matrix $g(\mathbf{x})$ is

$$E(g(\mathbf{x})) = \int_{-\infty}^{\infty} g(u) f_{\mathbf{x}}(u) du.$$

The next theorem follows from the linearity of integration.

Theorem 3. *If $A \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^p$ are constant, then*

$$E(A\mathbf{x} + b) = A E(\mathbf{x}) + b.$$

The mean $\mu_{\mathbf{x}} = E(\mathbf{x})$ and the covariance matrix $\Sigma_{\mathbf{x}} = V(\mathbf{x}) = E((\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{x} - \mu_{\mathbf{x}})^T)$, of random variable $\mathbf{x}$ are important expectation values.

The correlation matrix of random variable $\mathbf{x}$ is

$$R_{\mathbf{x}} = D\Sigma_{\mathbf{x}}D, \tag{1.7}$$

where D is a diagonal matrix whose diagonal contains the square roots of the reciprocals of the diagonal elements of the covariance matrix $\Sigma_{\mathbf{x}}$.

Example 6. *Let $\mathbf{x}_i$ be independent identically distributed continous random variables, so that $E(\mathbf{x}_i) = \mu$ and $V(\mathbf{x}_i) = \Sigma$. Define random variable $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$. Now*

$$E(\bar{\mathbf{x}}) = \int \left(\frac{1}{n}\sum_{i=1}^{n} x_i \right) \prod_{k=1}^{n} f_{\mathbf{x}_k}(x_k) dx_1 \cdots dx_n = \frac{1}{n}\sum_{i=1}^{n} \int x_i f_{\mathbf{x}_i}(x_i) dx_i = \mu \quad \text{and}$$

$$V(\bar{\mathbf{x}}) = \int \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu) \right) \left(\frac{1}{n}\sum_{j=1}^{n}(x_j - \mu) \right)^T \prod_{k=1}^{n} f_{\mathbf{x}_k}(x_k) dx_1 \cdots dx_n$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \int (x_i - \mu)(x_j - \mu)^T \prod_{k=1}^{n} f_{\mathbf{x}_k}(x_k) dx_1 \cdots dx_n = \frac{1}{n}\Sigma.$$

The random variable $\bar{\mathbf{x}}$ is an unbiased estimator of the parameter μ because $E(\bar{\mathbf{x}} - \mu) = 0$. Furthermore, we see that $V(\bar{\mathbf{x}}) = \frac{1}{n}\Sigma \to 0$, when $n \to \infty$. Thus, for all $\varepsilon > 0$, the probability $P(\|\bar{\mathbf{x}} - \mu\| > \varepsilon) \to 0$ when $n \to \infty$. Therefore, the estimator $\bar{\mathbf{x}}$ is also consistent.

1.3.1 Normal distribution

We assume that the reader is familiar with the one-dimensional normal distribution. We define n-dimensional normal distribution as follows.

Definition 4 (Normal distribution). *An n-dimensional random variable $\mathbf{x}$ is normal distributed with parameters $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$, $\Sigma \geq 0$, denoted $\mathbf{x} \sim N(\mu, \Sigma)$ or $\mathbf{x} \sim N_n(\mu, \Sigma)$, if the random variable $a^T \mathbf{x}$ is an one-dimensional normal distribution or constant for all vectors $a \in \mathbb{R}^n$.*

Let $\mathbf{x} \sim N(\mu, \Sigma)$. Then the parameters are the mean $\mu = E(\mathbf{x})$ and the covariance $\Sigma = V(\mathbf{x})$ of the random variable $\mathbf{x}$. If Σ is positive definite, then the density function of random variable $\mathbf{x}$ is

$$f_{\mathbf{x}}(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) \right). \tag{1.8}$$

If Σ positive semi-definite but singular, then we call the distribution a singular normal distribution. For instance, in Chapter 3 the state model error $\mathbf{w}_{k-1}$ (3.1) can quite possibly follow singular normal distribution. This happens for instance when we know that the user is stationary, then the position state model is simply $\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{w}_{k-1}$, where $\mathbf{w}_{k-1} \sim N(0,0)$.

Example 7 (Visualizing the normal distribution). *Figure 1.1 shows the density function of the random variable*

$$\mathbf{x} \sim N(0, \Sigma), \tag{1.9}$$

where $\Sigma = \begin{bmatrix} 16 & 24 \\ 24 & 52 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 8^2 & 0 \\ 0 & 2^2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}.$
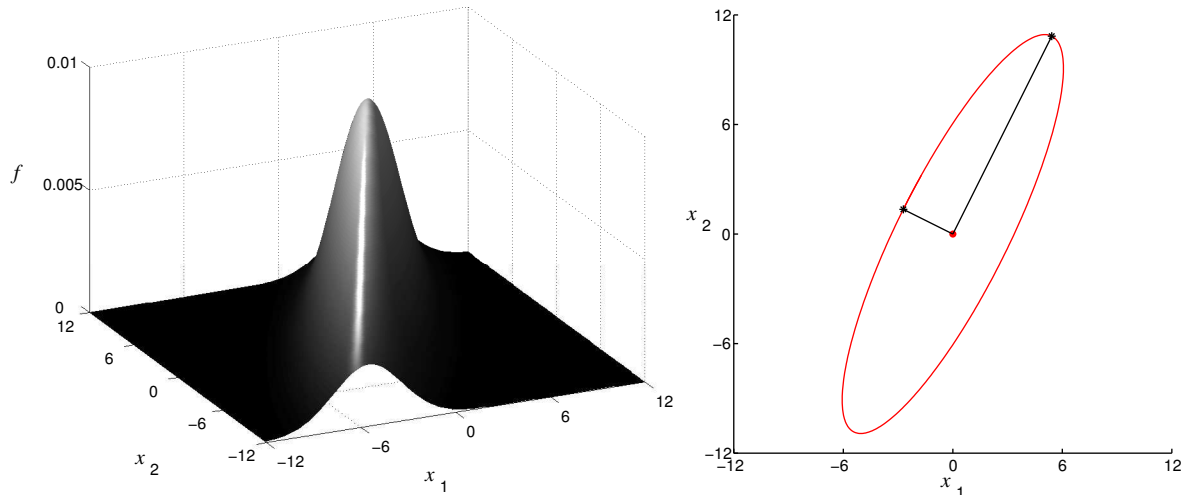
Figure 1.1: Two visualizations of the density function of normal distribution (1.9) in the area $|x_1| \leq 12$, $|x_2| \leq 12$

The picture on the left shows density function (1.8) values. The picture on the right shows the level curve of the density function that contains 68% of the probability mass, see Exercise 1.10. The picture on the right also contains line segments which are directed along the eigenvectors of Σ and whose lengths are proportional to the square roots of corresponding eigenvalues.

Theorem 4. *If $\mathbf{x} \sim N(\mu, \Sigma)$ and $\mathbf{y} = A\mathbf{x} + b$ then $\mathbf{y} \sim N(A\mu + b, A\Sigma A^T)$.*

Example 8 (Generation of normal distributed random varibales)**.** *If $\mathbf{u} \sim N_n(0, I)$ then random variable $\mathbf{x} = \Sigma^{\frac{1}{2}}\mathbf{u} + \mu$ has normal distribution with parameters μ ja Σ (Theorem 4) that is*

$$\mathbf{x} \sim N_n(\mu, \Sigma).$$

If we can generate random variables from distribution $N_n(0, I)$ then we can generate random variables from arbitrary normal distribution $N_n(\mu, \Sigma)$. For example, in Matlab software we can use the command:

```
x=sqrtm(Sigma)*randn(n,1)+mu
```

Figure 1.2 shows one hundred samples (with some samples outside the figure) from normal distribution (1.9). Compare Figure 1.2 and Figure 1.1.
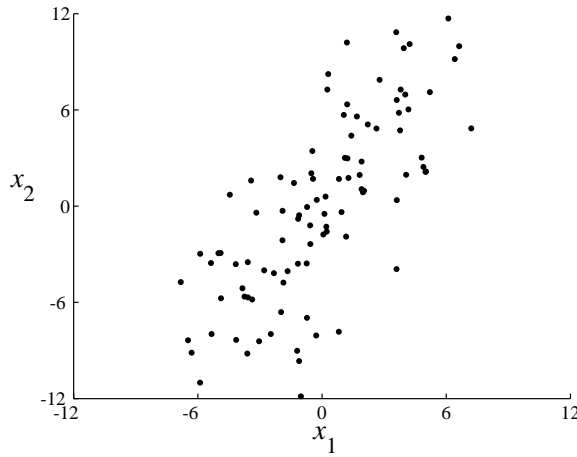
Figure 1.2: 100 samples from normal distribution (1.9).

Theorem 5. *Let $\mathbf{x} \sim N(\mu, \Sigma)$. Then $A\mathbf{x}$ and $B\mathbf{x}$ are independent if and only if $A\Sigma B^T = 0$.*

Example 9. *Let*

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N\left(\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right) \quad \text{ja} \quad \mathbf{z} = \mathbf{x} - \Sigma_{xy}\Sigma_{yy}^{-1}\mathbf{y}.$$

Are random variables $\mathbf{z}$ and $\mathbf{y}$ independent?

Note that

$$\begin{bmatrix} I & -\Sigma_{xy}\Sigma_{yy}^{-1} \end{bmatrix} \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \begin{bmatrix} 0 \\ I \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} & \Sigma_{xy} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yy} \end{bmatrix} \begin{bmatrix} 0 \\ I \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ I \end{bmatrix} = 0.$$

Thus, the random variables $\mathbf{z}$ and $\mathbf{y} = \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$ are independent (Theorem 5).
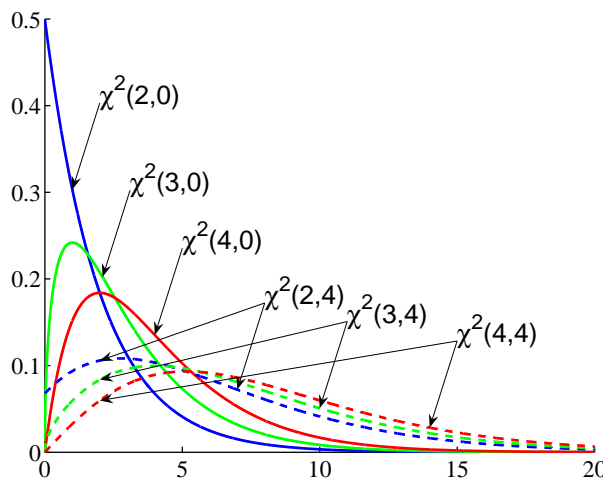
1.3.2 χ^2 distribution

Definition 5. *Random variable* $\mathbf{z}$ *is* χ^2 *distributed with degrees of freedom n, denoted* $\mathbf{z} \sim \chi^2(n)$, *if it has the same distribution as random variable* $\mathbf{x}^T\mathbf{x}$, *where* $\mathbf{x} \sim N_n(0, I)$.

Example 10. *Let* $\mathbf{x} \sim N_n(\mu, \Sigma)$ *be a non-singular random variable. Now* $\Sigma^{-\frac{1}{2}}(\mathbf{x} - \mu) \sim N(0, I)$, *and so*

$$(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu) \sim \chi^2(n).$$

Definition 6. *Random variable* $\mathbf{z}$ *is noncentral* χ^2 *distributed with n degrees of freedom and noncentral parameter* λ, *denoted* $\mathbf{z} \sim \chi^2(n, \lambda)$, *if it has the same distribution as the random variable* $\mathbf{x}^T\mathbf{x}$, *where* $\mathbf{x} \sim N_n(\mu, I)$ *ja* $\lambda = \mu^T\mu$.

Figure 1.3: Density functions of noncentral χ^2 distributions.

Theorem 6. *Let* $\mathbf{x} \sim N_n(\mu, \Sigma)$ *be non-singular random variable, let matrix* A *be symmetric and let matrix* $A\Sigma$ *be idempotent. Now*

$$\mathbf{x}^T A \mathbf{x} \sim \chi^2\left(\text{rank}(A), \mu^T A \mu\right).$$

Proof. Let $B = \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$. Matrix B is symmetric because matrix A and $\Sigma^{\frac{1}{2}}$ are symmetric and idempotent because

$$BB = \Sigma^{\frac{1}{2}} A \Sigma A \Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}} A \Sigma A \Sigma \Sigma^{-\frac{1}{2}} \stackrel{idemp.}{=} \Sigma^{\frac{1}{2}} A \Sigma \Sigma^{-\frac{1}{2}} = B.$$

Using Schur decomposition (Theorem 1) and Example 1 we get

$$B = Q \begin{bmatrix} I_{\text{rank}(B)} & 0 \\ 0 & 0 \end{bmatrix} Q^T,$$

where Q orthogonal. In Exercise 1.2, we show that rank(B) = rank(A). Define rank(A)-dimensional random variable $\mathbf{y} = [I_{\text{rank}(A)}, 0] Q^T \Sigma^{-\frac{1}{2}} \mathbf{x}$. Now using Theorem 4 we get

$$\mathbf{y} \sim N([I_{\text{rank}(A)}, 0] Q^T \Sigma^{-\frac{1}{2}} \mu, I_{\text{rank}(A)}).$$

11

and so, by Definition 6,

$$\mathbf{x}^T A \mathbf{x} = \mathbf{x}^T \Sigma^{-\frac{1}{2}} Q \begin{bmatrix} I_{\text{rank}(A)} & 0 \\ 0 & 0 \end{bmatrix} Q^T \Sigma^{-\frac{1}{2}} \mathbf{x} = \mathbf{y}^T \mathbf{y} \sim \chi^2(\text{rank}(A), \mu^T A \mu).$$

$\square$

1.3.3   Conditional density function

The conditional density function plays a central role in Bayesian probability theory.

Definition 7 (Conditional density function). *The conditional density function of a random variable* $\mathbf{x}$ *with condition* $\mathbf{y} = y$ *is defined with equation*

$$f_{\mathbf{x}|\mathbf{y}}(x|y) = \frac{f_{\mathbf{x},\mathbf{y}}(x,y)}{f_{\mathbf{y}}(y)},$$

at points where the denominator is positive.

According to Theorem 2 we see that random variables $\mathbf{x}$ and $\mathbf{y}$ are independent if and only if $f_{\mathbf{x}|\mathbf{y}}(x|y) = f_{\mathbf{x}}(x)$. Conditional expectation is defined correspondingly:

Definition 8 (Conditional expectation). *The conditional expectation of a random variable* $\mathbf{x}$ *with condition* $\mathbf{y} = y$ *is defined with equation*

$$E(\mathbf{x}|\mathbf{y} = y) = \int_{-\infty}^{\infty} x f_{\mathbf{x}|\mathbf{y}}(x|y) dx.$$

Notice that conditional expectation $E(\mathbf{x}|\mathbf{y} = y)$ depends on the value y of the random variable $\mathbf{y}$ and it is therefore also a random variable, defined in the same probability space $\mathbf{y}$.

Example 11 (Conditional density function). *Assume that the user's state* $\mathbf{x}$ *is normally distributed according to* $\mathbf{x} \sim N(0, 16)$, *and we get measurement*

$$\mathbf{y} = \frac{3}{2}\mathbf{x} + \mathbf{v},$$

where the error $\mathbf{v} \sim N(0, 16)$ *is independent of the state (of the random variable* $\mathbf{x}$). *Then density function of joint distribution* $\begin{bmatrix} \mathbf{x} \\ \mathbf{v} \end{bmatrix}$ *is the product of the density functions of random variables* $\mathbf{x}$ *and* $\mathbf{v}$. *Easy calculation shows that*

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{v} \end{bmatrix} \sim N(0, 16 \cdot I), \tag{1.10}$$

then according to Theorem 4

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N\left(0, \begin{bmatrix} 16 & 24 \\ 24 & 52 \end{bmatrix}\right). \tag{1.11}$$

It can then be shown (Exercise 1.14) that

$$\mathbf{x}|\mathbf{y} \sim N\left(\frac{6}{13}y, \frac{64}{13}\right). \tag{1.12}$$

For instance, if we get an observation $\mathbf{y} = -12$, then the conditional distribution of random variable $\mathbf{x}$ with condition $\mathbf{y} = -12$ is $N(-5\frac{7}{13}, 4\frac{12}{13})$. Graphically, the conditional density function is a normalized slice of the density function in the plane $\mathbf{y} = y$. Compare the function in plane $\mathbf{y} = -12$ on the left side of Figure 1.1 with the calculated density function of normal distribution (Figure 1.4).
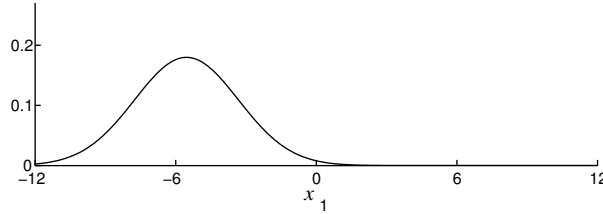
Figure 1.4: The density function of normal distribution $N(-5\frac{7}{13}, 4\frac{12}{13})$ within interval $[-12, 12]$, which can be compared to the left hand side of Figure 1.1.

1.4 Coordinate systems

JUSSI COLLIN

The precise definition of the coordinate systems used is a fundamental part of navigation. A linear coordinate system consists of an origin and a basis, f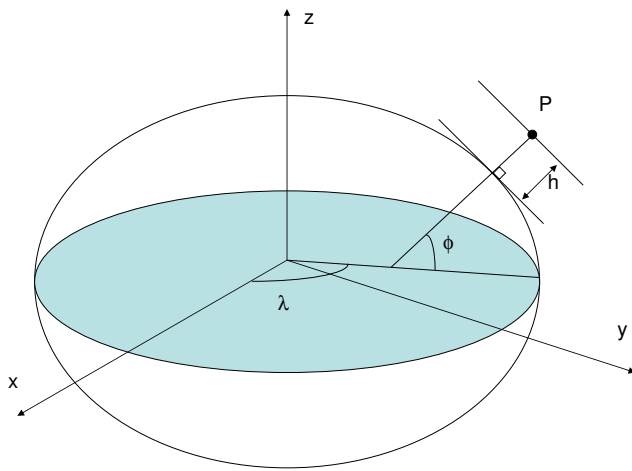or instance $(O, \mathbf{u}, \mathbf{v}, \mathbf{w})$, where vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are linearly independent. Now any position can be given in component representation $\mathbf{r} = x\mathbf{u} + y\mathbf{v} + z\mathbf{w}$, and coordinates x, y and z are uniquely defined.

When positioning with respect to Earth, curvilinear coordinate systems are often necessary. For example, in a spherical coordinate system, position is given as coordinates (r, θ, φ). Here r is the point's distance from origin, angle θ is the colatitude and φ is the azimuth angle. The commonly used geodetic coordinate system is slightly different from this. The geodetic coordinate system is generated by rotating an ellipse around its minor axis. This ellipsoid is usually characterized by two parameters, the major axis length a and flattening ratio $f = \frac{a-b}{a}$ (b is minor axis length). Geodetic coordinates in this coordinate system are defined as shown in Figure 1.5, i.e.

Figure 1.5: Geodetic coordinates and x-, y- and z-axis of the corresponding rectangular coordinate system.

- ϕ, geodetic latitude. Notice that the angle in question is the angle between the equatorial plane and the normal vector of the ellipsoid surface.

- λ, geodetic longitude. The angle between reference meridian and meridian of the location P.

- h, geodetic height. Signed distance from the ellipsoid.

In addition to mathematical definition, a coordinate system has to be fixed to some physical entity. The problem here is the dynamic behavior of Earth – a planet is not a rigid body. An *realization* of a coordinate is therefore fixed to some time instant, i.e. epoch. Typically the origin is fixed to be the Earth's mass center, the positive z-axis goes through the North Pole, the reference meridian goes through the Greenwich observatory and major axis length as well as flattening ratio are found by fitting an ellipsoid to the mean sea level of Earth. GPS uses World Geodetic System 84 (WGS-84) coordinate system [43]. The Finnish coordinate system realization (EUREF-FIN) has been defined in the Finnish Government Recommendation 153 [13].

1.4.1 Coordinate transformation

Navigation algorithms often require transformations between different coordinate systems. For instance, GPS computations start from inertial coordinates (satellites' orbits), switch to geodetic coordinates (coordinate system rotating with Earth) and finally to local coordinates for showing the position on a 2D map. In inertial positioning computations the coordinates have to be transformed perhaps even more often; sensors form their own coordinate system, sensor measurements are with respect to inertial coordinates, and gravitation computation requires position in geodetic coordinates.

Let A-coordinates be given as vector $\mathbf{m}^A = [\alpha_1 \ \alpha_2 \ \alpha_3]^T$ and B-coordinates as vector $\mathbf{m}^B = [\beta_1 \ \beta_2 \ \beta_3]^T$. Here α_i are coordinates of position P in coordinate system A and β_i are the same position's coordinates in coordinate system B. Coordinate transformation is then the function

$$\mathbf{f}_A^B(\mathbf{m}^A) = \mathbf{m}^B. \tag{1.13}$$

When the transformation is between linear coordinate systems this function is easy to find. Assume that both coordinate systems have the same origin. The position vector $\mathbf{q}$ can be written using the basis vectors and coordinates:

$$\begin{aligned} \mathbf{q} &= X_A \mathbf{m}^A \\ \mathbf{q} &= X_B \mathbf{m}^B \end{aligned},$$

where columns of X are the basis vectors of the coordinate system in question. A matrix formed from the basis vectors is non-singular, and thus the transformation is just

$$\mathbf{m}^B = X_B^{-1} X_A \mathbf{m}^A.$$

In the following equations, the basis vectors are assumed orthonormal, and therefore $X^{-1} = X^T$. Additionally, we set a natural base $X_A = I = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$ for A. Then the transformation is simplified into form

$$\mathbf{m}^B = X_B^T \mathbf{m}^A.$$

Now we can define a matrix often used in positioning, called the *direction cosine matrix* C_A^B:

$$\mathbf{m}^B = C_A^B \mathbf{m}^A, \tag{1.14}$$

where $\mathbf{m}^A$ is a vector given in coordinate system A and $\mathbf{m}^B$ is the same vector in coordinate system B. It can be shown (Exercise 1.15) that

$$C_A^B = \begin{bmatrix} \cos(\mathbf{u}, \mathbf{e}_1) & \cos(\mathbf{u}, \mathbf{e}_2) & \cos(\mathbf{u}, \mathbf{e}_3) \\ \cos(\mathbf{v}, \mathbf{e}_1) & \cos(\mathbf{v}, \mathbf{e}_2) & \cos(\mathbf{v}, \mathbf{e}_3) \\ \cos(\mathbf{w}, \mathbf{e}_1) & \cos(\mathbf{w}, \mathbf{e}_2) & \cos(\mathbf{w}, \mathbf{e}_3) \end{bmatrix}, \tag{1.15}$$

where $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ is the orthonormal basis of B. Coordinate transformation back to A is easy, because $C_B^A = (C_A^B)^{-1} = (C_A^B)^T$. The following chain rule is needed especially in inertial navigation algorithms:

$$C_A^D = C_B^D C_A^B. \tag{1.16}$$

Equation (1.15) is in quite nice form for handling coordinate frame rotations.

Example 12. *Assume that coordinate frames A1 and A2 are initially identical. Rotate the frame A2 around the z-axis by angle* θ. *An observer using frame A1 sees an interesting object in location* $\mathbf{a}^{A1} = [1\ 1\ 1]^T$. *Where does an observer using frame A2 find this object?*

Answer: Define rotations as right handed, i.e. a positive rotation fo the vector **v** *around the vector* **w** *takes the tip of* **v** *towards the direction* $\mathbf{w} \times \mathbf{v}$. *Note that every base vector is rotated in coordinate transformation. Now it is easy to see that the z-axis is not transformed. The other angles in equation (1.15) are also easy to find:*

$$C_{A1}^{A2} = \begin{bmatrix} \cos(\theta) & \cos(\frac{\pi}{2} - \theta) & \cos(\frac{\pi}{2}) \\ \cos(\frac{\pi}{2} + \theta) & \cos(\theta) & \cos(\frac{\pi}{2}) \\ \cos(\frac{\pi}{2}) & \cos(\frac{\pi}{2}) & \cos(0) \end{bmatrix} \tag{1.17}$$

The solution is therefore $\mathbf{a}^{A2} = C_{A1}^{A2}\mathbf{a}^{A1} = \begin{bmatrix} \cos(\theta) + \sin(\theta) \\ \cos(\theta) - \sin(\theta) \\ 1 \end{bmatrix}$.

In order to deal with more general rotations, it is good to introduce some mathematical tools that make things easier. Denote with $(\mathbf{a}\times)$ the matrix form of the cross product:

$$\mathbf{a} \times \mathbf{b} = (\mathbf{a}\times)\mathbf{b} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \mathbf{b} \tag{1.18}$$

Direction cosine matrix is not the only way to describe orientations of the coordinates. It can be shown (see for instance [36]), that direction cosine matrix can be represented in the form

$$C_{A2}^{A1} = I + \frac{\sin(p)}{p}(\mathbf{p}\times) + \frac{1 - \cos(p)}{p^2}(\mathbf{p}\times)(\mathbf{p}\times), \tag{1.19}$$

where **p** is the rotation vector and p its length. When coordinate frame A1 is rotated around axis **p** by angle p, we get a new coordinate frame A2, and Equation (1.19) shows the connection between the direction cosine matrix and the rotation vector. As an exercise, you will verify that it does not matter whether the rotation vector is given in coordinates A1 or A2. In Example 12, the rotation vector is $\mathbf{p} = [0\ 0\ \theta]^T$, and it is left as an exercise to show that when this is substituted into formula (1.19) we get the transpose of the matrix (1.17), as expected.

1.5   Rotating coordinate frames

Transformations between coordinate frames that are moving linearly relative to each other are easily handled by adding velocities and translating the origin. In this section we rather focus on rotating coordinate frames. We start with simple rotation, i.e. keep the angular velocity and direction of rotation vector constant. Assume again that coordinate frames A1 and A2 are initially ($t = 0$) equivalent. Then the rotation vector can be given in the form

$$\mathbf{p} = \omega t \mathbf{u}, \tag{1.20}$$

where $\omega \, (= \dot{p})$ is angular speed (rad/s), t is time and $\mathbf{u}$ is unit vector. The derivative of Equation (1.20) with respect to time is thus the angular velocity

$$\mathbf{w} = \dot{\mathbf{p}} = \omega\mathbf{u}. \tag{1.21}$$

Now the derivative of Equation (1.19) with respect to time is

$$\begin{aligned}
\dot{\mathbf{C}}_{A2}^{A1} &= \omega\cos(\omega t)(\mathbf{u}\times) + \omega\sin(\omega t)(\mathbf{u}\times)(\mathbf{u}\times) \\
&= [\cos(\omega t)\mathbf{I} + \sin(\omega t)(\mathbf{u}\times)](\mathbf{w}\times)
\end{aligned}. \tag{1.22}$$

Here you should notice that $\mathbf{C}_{A2}^{A1}(\mathbf{w}\times) = [\mathbf{I}\cos(\omega t) + \sin(\omega t)(\mathbf{u}\times)](\mathbf{w}\times)$ [*]. This form is more compact, for instance the Coriolis equation can now be easily derived:

$$\begin{aligned}
\mathbf{n}^A &= \mathbf{C}_B^A \mathbf{n}^B & \Rightarrow \\
\frac{d\mathbf{n}^A}{dt} &= \frac{d\mathbf{C}_B^A \mathbf{n}^B}{dt} & \Rightarrow \\
\frac{d\mathbf{n}^A}{dt} &= \frac{d\mathbf{C}_B^A}{dt}\mathbf{n}^B + \mathbf{C}_B^A\frac{d\mathbf{n}^B}{dt} & \Rightarrow \\
\frac{d\mathbf{n}^A}{dt} &= \mathbf{C}_B^A(\mathbf{w}\times)\mathbf{n}^B + \mathbf{C}_B^A\frac{d\mathbf{n}^B}{dt} & \Rightarrow \\
\frac{d\mathbf{n}^A}{dt} &= \mathbf{C}_B^A[\frac{d\mathbf{n}^B}{dt} + \mathbf{w}\times\mathbf{n}^B]
\end{aligned} \tag{1.23}$$

Using the chain rule, we get rid of the initial assumption; if at first $(t=0)$ the connection between coordinate systems is $\mathbf{C}_{A2(0)}^{A1}$, then

$$\mathbf{C}_{A2}^{A1} = \mathbf{C}_{A2(0)}^{A1}\mathbf{C}_{A2}^{A2(0)}$$

and matrix $\mathbf{C}_{A2(0)}^{A1}$ is constant with respect to time, i.e.

$$\dot{\mathbf{C}}_{A2}^{A1} = \mathbf{C}_{A2(0)}^{A1}\dot{\mathbf{C}}_{A2}^{A2(0)}.$$

After this, we need to be more careful with the angular velocity. INS literature often uses the notation $\mathbf{w}_{IB}^B$, which reads as the angular velocity in B-coordinate frame with respect to I-coordinate frame ($\mathbf{w}_{IB}$), and the vector is expressed in B-frame coordinates ($\mathbf{w}^B$). The subscript conventions thus differ from those in the direction cosine matrix. I-coordinate frame means inertial coordinate frame (it does not rotate nor accelerate in space), and B-coordinate frame is a coordinate system fixed to the measurement unit. The vector $\mathbf{w}_{IB}^B$ is in fact the output of an ideal gyro triad.

The following coordinate frames are used in inertial navigation:

- *Inertial frame (I)*, non-rotating, non-accelerating coordinate system. Newton's laws of motion apply in this coordinate system. In some cases (relativity theory) we additionally need to assume that there is no gravitation in I-frame, but in this course the computations in I-frame can be approximated with the ECI (Earth Centered Inertial) coordinate system. The origin of ECI is the Earth's mass center, and the axes keep their direction with respect to stars.

---

[*]This can be verified by using the formula $(\mathbf{a}\times)(\mathbf{a}\times) = -\|\mathbf{a}\|^2\mathbf{I} + \mathbf{a}\mathbf{a}^T$, for example

- *Earth frame (E)*, coordinate axes are fixed to Earth, z-axis has the same direction as Earth rotation axis, x- and y-axes are in equatorial plane. In literature one often sees abbreviation ECEF (Earth Centered Earth Fixed). The Earth-fixed frame rotates with respect to the inertial fram with $\mathbf{w}_{IE}^E \approx [0\ 0\ 7.29 \times 10^{-5}]^T$ (rad/s).

- *Local frame (L)*, axes define directions "up–down", "north–south" and "west–east". It is common to use ENU (East, North, Up) ordering, i.e. x-axis points east, y-axis north and z-axis points up. When the origin is fixed to user's position, C_L^E is a function of time and $\mathbf{w}_{IL}^L = \mathbf{w}_{IE}^L + \mathbf{w}_{EL}^L$, where the last term describes user's movement with respect to Earth.

- *Body frame (B)*, frame fixed to the vehicle in which the navigation system is used, or the mobile unit in the case of personal navigation. This is the main coordinate system in inertial measurements. Gyro triad measurement is $\mathbf{w}_{IB}^B$ and the acceleration triad outputs $\mathbf{a}^B - \mathbf{g}^B$, where $\mathbf{a}$ is acceleration in ECI and $\mathbf{g}$ is local gravitation acceleration.

Because the laws of motion apply only in I-frame, and sensors thus measure motion with respect to this coordinate frame, it is important to have a good understanding of dynamics of coordinate frames B-L-E-I. We start by finding the time derivative of the equation

$$C_B^L = C_I^L C_B^I. \tag{1.24}$$

First, note that

$$\dot{C}_L^I = C_L^I (\mathbf{w}_{IL}^L \times)$$

and that its transpose is

$$\dot{C}_I^L = -(\mathbf{w}_{IL}^L \times) C_I^L.$$

Differentiating equation (1.24) with respect to time from both sides gives:

$$\dot{C}_B^L = \dot{C}_I^L C_B^I + C_I^L \dot{C}_B^I$$

When we also remember that $\dot{C}_B^I = C_B^I (\mathbf{w}_{IB}^B \times)$, we end up with equation

$$\dot{C}_B^L = C_B^L (\mathbf{w}_{IB}^B \times) - (\mathbf{w}_{IL}^L \times) C_B^L. \tag{1.25}$$

Numerical integration of equation (1.25) is the beef of inertial navigation algorithms.

Example 13. *A satellite is in a circular orbit with radius R and inclination i. An observer at the equator sees the satellite at zenith, going towards the northern hemisphere. What is the satellite's velocity and acceleration in E-coordinate system?*

Answer: Choose an Earth-centered satellite-fixed coordinate frame S, where the satellite is in location $\mathbf{r}^S = [R\ 0\ 0]^T$. Let the x-axis of ECI- coordinate frame point to the satellite at instant

$t = 0$, and let E- and I-coordinate coincide at the time of observation. Now we can see that $\mathbf{w}_{IS}^I = \omega_s[0 \ -\sin(i) \ \cos(i)]^T$, where $\omega_s = \frac{2\pi}{T}$ and T is the orbital period, and

$$
C_S^E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(i) & -\sin(i) \\ 0 & \sin(i) & \cos(i) \end{bmatrix}.
$$

Using Earth's angular speed ω_e, we get $\mathbf{w}_{IE}^E = [0 \ 0 \ \omega_e]^T$. We use Equation (1.25) in these coordinate systems, i.e.

$$
\dot{C}_S^E = C_S^E(\mathbf{w}_{IS}^S \times) - (\mathbf{w}_{IE}^E \times)C_S^E
$$

and we get for instant $t = 0$

$$
\dot{C}_S^E = \begin{bmatrix} 0 & -\omega_s + \omega_e\cos(i) & -\omega_e\sin(i) \\ \omega_s\cos(i) - \omega_e & 0 & 0 \\ \omega_s\sin(i) & 0 & 0 \end{bmatrix} \tag{1.26}
$$

and

$$
\ddot{C}_S^E = \begin{bmatrix} -\omega_s^2 + 2\omega_s\omega_e\cos(i) - \omega_e^2 & 0 & 0 \\ 0 & -\omega_s^2\cos(i) + 2\omega_s\omega_e - \omega_e^2\cos(i) & \omega_e^2\sin(i) \\ 0 & -\omega_s^2\sin(i) & 0 \end{bmatrix}. \tag{1.27}
$$

Notice! "Velocity in E-coordinate system" now means: change of E-fixed position vector given in E-frame with respect to time. It is not I-velocity vector given in E-coordinates (we will get back to this in the INS part). That means $\dot{\mathbf{r}}^E$ needs to be thought as $\frac{d}{dt}(\mathbf{r}^E)$. The position of the satellite is constant in S-coordinate system, i.e. $\dot{\mathbf{r}}^S = \mathbf{0}$ and $\ddot{\mathbf{r}}^S = \mathbf{0}$. So the answer is

$$
\dot{\mathbf{r}}^E = \dot{C}_S^E \mathbf{r}^S = [0 \ R(\omega_s\cos(i) - \omega_e) \ R(\omega_s\sin(i))]^T \tag{1.28}
$$

and

$$
\ddot{\mathbf{r}}^E = \ddot{C}_S^E \mathbf{r}^S = [R(-\omega_s^2 + 2\omega_s\omega_e\cos(i) - \omega_e^2) \ 0 \ 0]^T. \tag{1.29}
$$

In the case of a GPS satellite, applicable numerical values would be e.g. $i = 0.967$ rad, $R = 26600$ km, $\omega_s = 151x10^{-6}$ rad/s, and

$$
\dot{\mathbf{r}}^E = [0 \ 341 \ 3310]^T m/s
$$

$$
\ddot{\mathbf{r}}^E = [-0.414 \ 0 \ 0]^T m/s^2
$$

The transformations between coordinate systems presented above were based on matrix algebra. Another approach way is to use quaternion algebra ([8, 35]), where for instance the matrix (1.19) corresponds to quaternion

$$
q_{A2}^{A1} = \begin{bmatrix} \cos(\frac{1}{2}p) \\ \sin(\frac{1}{2}p)\mathbf{p}/p \end{bmatrix}. \tag{1.30}
$$

Exercises

1.1. Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Show that

$$A > 0 \iff \text{the eigenvalues of A are positive}$$
$$A > 0 \Rightarrow \det(A) > 0$$
$$A > 0 \iff A^{-1} > 0$$

1.2. Suppose that matrixes $B > 0$ and A have same size. Show that

$$\text{rank}(BAB) = \text{rank}(A).$$

1.3. (a) Let columns of matrix $A \in \mathbb{R}^{m \times n}$ with $m > n$ be linearly independent ($\text{rank}(A) = n$). Show that matrix $A^T A$ is invertible.

   (b) Show that equation (1.5) is correct.

1.4. Let matrices $A > 0$ and $B > 0$. Show that matrix $A + B$ is invertible.

1.5. With $A = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix}$, compute $A^{\frac{1}{2}}$.

1.6. Find all matrices $B \in \mathbb{R}^{2 \times 2}$ such that $BB = I$. Which ones of the obtained matrices are symmetric and positive definite?

1.7. User is inside an equilateral triangle building. The length of each side of the triangle is $6l$. User has distance measurements from each side, these measurements are l, $2l$ and $3l$. What is the least squares estimate of the user's position?

1.8. Let $A \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^p$ be constants and $\mathbf{x}$ a random variable such that $V(\mathbf{x}) = \Sigma$. Compute
$$V(A\mathbf{x} + b)$$

1.9. Consider the system $\mathbf{y} = Hx + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 I)$. Compute the distribution of the estimator
$$\hat{\mathbf{x}} = (H^T H)^{-1} H^T \mathbf{y}.$$

Is the estimator unbiased ($E(x - \hat{\mathbf{x}}) = 0$)?

1.10. Give the equation of the ellipse shown on the right of Figure 1.1. (help. Matlab: `chi2inv(0.68,2)` ≈ 2.279)

1.11. Let

$$\mathbf{x} \sim N\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix}\right) \quad \text{ja} \quad a = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Compute the probability $P\left(a^T\mathbf{x} \le 0\right)$.

1.12. Let $\mathbf{x} \sim \chi^2(n, \lambda)$. Compute $E(\mathbf{x})$.

1.13. Let $\mathbf{x} \sim N(\bar{x}, \Sigma_x)$ and $\mathbf{y} \sim N(\bar{y}, \Sigma_y)$ be independent random variable, for simplicity we assume that $\Sigma_x > 0$ and $\Sigma_y > 0$. Show that

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N\left(\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \begin{bmatrix} \Sigma_x & 0 \\ 0 & \Sigma_y \end{bmatrix}\right).$$

tip: If A and C are square matrixes, then $\det\left(\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}\right) = \det(A)\det(C)$.

1.14. Prove Equation (1.12) starting from Definition 7.

1.15. Derive the direction cosine matrix (1.15).

1.16. Let $\mathbf{p}$ be rotation vector between frames $A1$ and $A2$. Show that $\mathbf{p}^{A1} = \mathbf{p}^{A2}$. If $C_{A2}^{A1} \ne I$, can you find $\mathbf{q} \ne \alpha\mathbf{p}$, $\alpha \in \mathbb{R}$ which satisfies $\mathbf{q}^{A1} = \mathbf{q}^{A2}$?

1.17. Try formula (1.19) with vectors $\mathbf{p}_1 = \begin{bmatrix} 0 & 0 & \theta \end{bmatrix}^T$ ja $\mathbf{p}_2 = \begin{bmatrix} 0 & 0 & -\theta \end{bmatrix}^T$

1.18. A hiker walks 5 km south, 5 km east and 5 km north, and ends up to the initial location. What was the origin of his journey? Express the movement using direction cosine matrices, and try to find those locations where $C_2^1 C_3^2 C_4^3 = I$.

1.19. A maglev train runs from Helsinki to Rovaniemi at 300 km/h. How large lateral force is needed to keep the train in north directed course?
What is the angle between Earth's rotation axis and the direction of the train with respect to time?
Assume a spherical Earth with radius R, train mass M and simplify the coordinates as:
Helsinki: $60^o 10'00''$ north latitude $25^o 0'00''$ east longitude
Rovaniemi: $66^o 30'00''$ north latitude $25^o 0'00''$ east longitude

1.20. Let a mobile phone coordinate system (B) be defined as follows: x-axis points to direction given by numbers 9-6-3, y-axis points to direction given by numbers 9-8-7, and z-direction is the cross product of x-axis and y-axis (from the screen to viewer of the screen). Rotate the phone -90 degrees around phone's y-axis. Next, rotate phone 90 degrees around x-axis. Finally, rotate phone 90 degrees around y-axis. Do the same mathematically, i.e. write direction cosine matrices. What happens if you change the order of the rotations? What if you do the rotations with respect to a non-rotating (e.g. fixed to your desktop) coordinate system?

1.21. Show that

$$\dot{\mathbf{C}}_{A2}^{A1} = \mathbf{C}_{A2}^{A1}(\mathbf{w}\times).$$

Hint: (1.19), (1.22) and $\mathbf{u}^T(\mathbf{u}\times) = \dots$.

Chapter 2

Static positioning

Niilo Sirola

Positioning (or localization) means estimation of a receiver's coordinates and possibly other interesting quantities (velocity, orientation, clock bias, etc). Many kinds of measurements coming from different sources can be used, and to solve the problem we first form a *mathematical model* to describe the measurements, and then apply suitable mathematical machinery to the model. No model describes reality perfectly, and often even if we knew the phenomenon quite accurately we may still decide to use some simpler model that fits better to the mathematical method we want to (or know how to) use for the problem.

Static positioning means computing a single location estimate from several simultaneously taken measurements, independent of previous or future measurements or estimates. The more general time series solution will be tackled later in Chapter 3, Depending on the choice of measurement model, the static positioning problem can be formulated as solving a nonlinear system of equations either in a sense of least squares (Section 2.3), in closed form (Section 2.4), or with respect to likelihoods (Section 2.5) or probability (Section 2.6).

Most of the methods needed for static positioning come from estimation theory, see e.g. [37, 41].

2.1 Measurement equations

Usually, not all details of the measurement process are known exactly, or it is not sensible to include them all into the model. Simple models often are sufficiently accurate without making computations too complicated. For instance, measurement error result from several different factors that can have complex mutual dependencies, but still modelling the error with one normally distributed term often works surprisingly well.

The measurement model can be written as a probability density function or in equation form. We start from the equation form, which is simpler but has some limitations (e.g. all measurements

have to be numerical and continuous). Measurements are thought to consist of a "true value" and an error:

$$\text{measurement} = \text{theoretical value at location } \mathbf{x} + \text{error}.$$

All the measured values are collected into a vector $\mathbf{y}$ and the system of equations is written in the form

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) + \mathbf{v} \tag{2.1}$$

where $\mathbf{h}$ is a known vector valued vector function and $\mathbf{v}$ is the measurement error vector, whose probability distribution is assumed known.[*]

When measurements have been made and the measurement equations known, we search for a location/state estimate $\hat{\mathbf{x}}$ that fits the measurements best. There are different solutions and solution methods depending on what we mean by the words "fit" and "best". The solution is not always unique, that is, more than one location candidate may fit the measurements equally well.

Example 14 (Measurement equations). *Range measurement to a station located at* $\mathbf{s}$ *can be written as*

$$r = \|\mathbf{s} - \mathbf{x}\| + v,$$

where r is the measured range, $\mathbf{x}$ *is location and* $\|\mathbf{s} - \mathbf{x}\|$ *(* $= h(\mathbf{x})$ *) is the true range.*

As another example, GPS pseudorange measurement can be written as

$$\rho = \|\mathbf{s} - \mathbf{x}\| + b + v,$$

where b is the additional error in meters caused by receiver's clock bias. Because the clock bias b is also an unknown to be solved in the same way as location, we define a 4-dimensional state vector such that the first three state components, denoted by $\mathbf{x}_{1:3}$*, contain location and the remaining component* x_4 *is the clock bias. If there are measurements to several different satellites/stations, then the measurement equations (2.1) can be written in vector form*
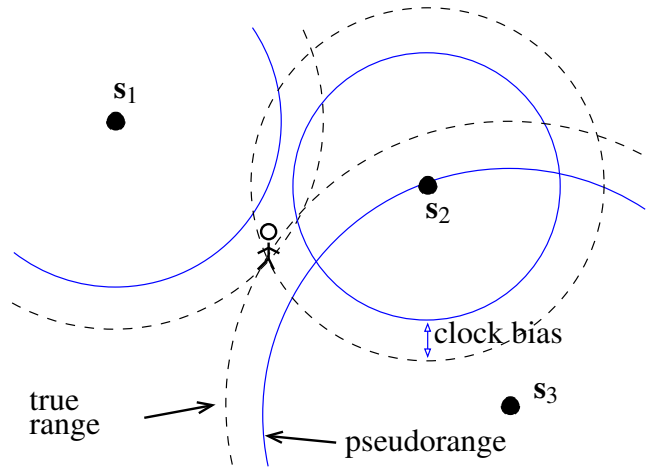
Figure 2.1: *Positioning with pseudoranges. Clock bias shows up as an error with same magnitude in all pseudoranges.*

$$\begin{bmatrix} \rho_1 \\ \vdots \\ \rho_n \end{bmatrix} = h(\mathbf{x}) + \mathbf{v} = \begin{bmatrix} \|\mathbf{s}_1 - \mathbf{x}_{1:3}\| + x_4 \\ \vdots \\ \|\mathbf{s}_n - \mathbf{x}_{1:3}\| + x_4 \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

[*] We do not get into the estimation of real-life measurement error distributions based on the properties of the transmission path, transmitter and receiver or based on empirical measurement data. Just assume for now that the measurement error distribution or at least its variance has been given (or guessed) beforehand.

2.2 Error analysis and sensitivity

In addition to location or state estimates, we are interested about the accuracy and reliability of the estimate. Measurements always contain errors, so after computing the location estimate an essential question is how close to the true location the estimate can be promised to be. Before actual error analysis, it is useful to classify different kinds of errors and to specify which of them we even try to consider.

- **modelled errors: caused by measurement errors and measurement geometry**
- modeling errors: wrong assumptions/guesses about errors' distributions or their dependencies, measurement resolution or detection threshold, incorrect values of natural constants, etc.
- numerical errors: caused by limited accuracy in computations (and/or careless programming)
- approximation, truncation and sampling errors: often result from intentional "optimization" of the algorithm
- others: environmental factors, equipment misuse, coding bugs, etc.

In this section, we only consider modelled errors, i.e. the errors that our chosen measurement model says will be there. The obtained error bound and estimates strictly speaking hold only if there are no other kinds of errors, but all the models are correct, random errors follow their assumed distributions, and computations are done without significant numerical errors. Error bound computed this way have to be carefully interpreted along the lines of: "If the models were correct, then...". Other error types are covered in the discussion of robustness and reliability in Chapter 5.

The error of the estimation process can be characterized in different ways:

- Error propagation: Given the distribution (or at least the covariance) of measurement errors (=input), compute the distribution (covariance) of the solution (=output)

- Confidence intervals: Compute the probability that the true location is inside a given area, or determine an area that includes the true position with e.g. 95% probability

- Sensitivity analysis: How much do small changes in measurements change the solution. If for measurements $\mathbf{y}$ we get estimate $\hat{\mathbf{x}}$, how much does the estimate change if the measurement is $\widetilde{\mathbf{y}} = \mathbf{y} + \Delta\mathbf{y}$, where the change $\Delta\mathbf{y}$ is assumed relatively small?

2.3 Iterative least squares

Write the system of measurement equations (2.1) again with the help of *residual*:

$$\mathbf{p}(\mathbf{x}) = \mathbf{h}(\mathbf{x}) - \mathbf{y}.$$

The residual describes the incompatibility between measurements and an assumed location $\mathbf{x}$. If measurements are error-free, then the residual goes to zero when $\mathbf{x}$ is the true location (but possibly also in other points). The idea of the least squares method is to determine a location estimate $\hat{\mathbf{x}}$ that minimizes the expression $\mathbf{p}(\mathbf{x})^T \mathbf{p}(\mathbf{x})$.

If there are no measurement errors, then the residual goes to zero in true location $\bar{\mathbf{x}}$:

$$\mathbf{p}(\bar{\mathbf{x}}) = \mathbf{0},$$

in which case the solution(s) could be found also by directly solving the system of measurement equations (see Section 2.4). Usually, the measurements contain some error, then the residual does not necessarily have roots at all, and the location estimate is obtained by solving the minimization problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \mathbf{p}(\mathbf{x})^T \mathbf{p}(\mathbf{x}). \tag{2.2}$$

Nonlinear optimization is a demanding problem and there is no general way to solve it analytically.

If the function to be minimized is well-behaved and there is an initial guess "close enough" to the minimum, then we can use an iterative optimization method. The idea is to start from an initial guess $\mathbf{x}_0$ and then compute refined estimates $\mathbf{x}_1, \mathbf{x}_2$, etc., until the solution does not change anymore. In Gauss-Newton method [15, section 4.2.1], [20, p. 93], [33, section 9.6] (which in positioning context is usually called iterative least squares method) the residual *is linearized* in the neighborhood of $\mathbf{x}_k$ using the first order Taylor expansion $\mathbf{p}(\mathbf{x}_k + \Delta\mathbf{x}_k) \approx \mathbf{p}(\mathbf{x}_k) + \mathbf{p}'(\mathbf{x}_k)\Delta\mathbf{x}_k$ and a step $\Delta\mathbf{x}_k$ is sought such that the linearized residual goes to zero:

$$\mathbf{p}(\mathbf{x}_{k+1}) = \mathbf{p}(\mathbf{x}_k + \Delta\mathbf{x}_k) \approx \mathbf{p}(\mathbf{x}_k) + \mathbf{p}'(\mathbf{x}_k)\Delta\mathbf{x}_k = \mathbf{0}.$$

Denoting the residual's *Jacobian matrix* as $J_k = \mathbf{p}'(\mathbf{x}_k) = \mathbf{h}'(\mathbf{x}_k)$, the equation can be written in the form

$$J_k \Delta\mathbf{x}_k = -\mathbf{p}(\mathbf{x}_k).$$

As long as there are enough independent measurements, this is an overdetermined system of linear equations, whose least squares solution (1.6) is

$$\Delta\mathbf{x}_k = -(J_k^T J_k)^{-1} J_k^T \mathbf{p}(\mathbf{x}_k).$$

Thus, a better estimate for the minimizer is obtained with the iteration

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (J_k^T J_k)^{-1} J_k^T \mathbf{p}(\mathbf{x}_k). \tag{2.3}$$

For algorithm implementation, the initial guess $\mathbf{x}_0$ is needed, and starting from that, the iteration is repeated until the solution has converged to some point $\hat{\mathbf{x}}$ (or the maximum number of iterations has been reached). It can be shown that the iteration converges if the initial point is "close enough" to the minimum and the second derivative of the residual is "small enough" [20].

Algorithm 1 Gauss-Newton method (Iterative Least Squares)
___

1. Choose initial guess $\mathbf{x}_0$ and stopping tolerance δ. Set $k = 0$.

2. Compute $\mathrm{J}_k = \mathbf{h}'(\mathbf{x}_k)$.

3. Step: $\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta\mathbf{x}_k$, where $\Delta\mathbf{x}_k = -\mathrm{J}_k \backslash (\mathbf{h}(\mathbf{x}_k) - \mathbf{y})$

4. If stopping condition $\|\Delta\mathbf{x}_k\| < \delta$ is not satisfied, then increase k and repeat from Step 2.
___

When using just satellite measurements, center of the Earth works as an initial guess when no better information is available, because the measurement equations are nearly linear.

If some measurements are more accurate than others, the algorithm can further be improved by using a weight matrix that forces good-quality measurement equations to be realized more accurately than poor-quality equations. When the inverse of the measurement covariance matrix is used as a weight matrix the problem to be solved is thus

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \mathbf{p}(\mathbf{x})^T \Sigma^{-1} \mathbf{p}(\mathbf{x}). \tag{2.4}$$

This produces the minimum variance estimate (proved in [26, chapter 6.A.1]), and in case of normally distributed errors, also the maximum likelihood estimate (see Example 17 on page 31).

Algorithm 2 Weighted Gauss-Newton method (Iterative Weighted Least Squares)
___

1. Choose initial guess $\mathbf{x}_0$ and stopping tolerance δ. Additionally, measurement covariance matrix $\Sigma = \mathrm{cov}(\mathbf{v})$ is required. Set $k = 0$.

2. Compute $\mathrm{J}_k = \mathbf{h}'(\mathbf{x}_k)$

3. Step: $\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta\mathbf{x}_k$, where $\Delta\mathbf{x}_k = -(\Sigma^{-\frac{1}{2}}\mathrm{J}_k) \backslash \left(\Sigma^{-\frac{1}{2}}(\mathbf{h}(\mathbf{x}_k) - \mathbf{y})\right)$

4. If stopping condition $\|\Delta\mathbf{x}_k\| < \delta$ is not satisfied, increase k and repeat from item 2.
___

Remarks:

1. The Weighted Least Squares method gives optimal results (in variance sense) as long as the measurement covariance is known (that is, if measurement errors follow the presumed distribution). The least squares criterion is sensitive to large measurement errors, outliers, which should be detected and removed.

2. If the measurement equation is strongly nonlinear, then the Gauss-Newton method can behave poorly and some other optimization method such as line search or Levenberg-Marquardt can be better. If the measurement equation has discontinuities, then it usually

has to be either approximated with some continuous function or minimized with some Monte Carlo type algorithm. [33]

3. In addition to measurement equations, inequality constraints (e.g. mobile phone network cell sector boundary) can also be taken into the measurement model. Then the problem becomes a constrained minimization problem, for which there are special technoques.

Example 15 (Jacobian matrix computation). *The system of measurement equations for range measurements to three stations in locations* $\mathbf{s}_1$, $\mathbf{s}_2$ *and* $\mathbf{s}_3$ *is*

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \|\mathbf{s}_1 - \mathbf{x}\| \\ \|\mathbf{s}_2 - \mathbf{x}\| \\ \|\mathbf{s}_3 - \mathbf{x}\| \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

The Jacobian matrix required in Newton algorithm is

$$\mathbf{J}(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} \mathbf{h}(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} \begin{bmatrix} \|\mathbf{s}_1 - \mathbf{x}\| \\ \|\mathbf{s}_2 - \mathbf{x}\| \\ \|\mathbf{s}_3 - \mathbf{x}\| \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}} \|\mathbf{s}_1 - \mathbf{x}\| \\ \frac{\partial}{\partial \mathbf{x}} \|\mathbf{s}_2 - \mathbf{x}\| \\ \frac{\partial}{\partial \mathbf{x}} \|\mathbf{s}_3 - \mathbf{x}\| \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{(\mathbf{s}_1 - \mathbf{x})^T}{\|\mathbf{s}_1 - \mathbf{x}\|} \\ -\frac{(\mathbf{s}_2 - \mathbf{x})^T}{\|\mathbf{s}_2 - \mathbf{x}\|} \\ -\frac{(\mathbf{s}_3 - \mathbf{x})^T}{\|\mathbf{s}_3 - \mathbf{x}\|} \end{bmatrix}.$$

In this case, each row of the Jacobian matrix is the transpose of a unit vector pointing from the presumed position $\mathbf{x}$ *to a station.*

2.3.1 Least squares error analysis

Let $\bar{\mathbf{x}}$ be the true location, $\bar{\mathbf{y}}$ the error-free measurements and $\Delta \mathbf{y}$ the realized measurement error. If the least squares method converges to solution $\hat{\mathbf{x}} = \bar{\mathbf{x}} + \Delta \mathbf{x}$, then the iteration step (2.3) length at this point is zero, i.e.:

$$(\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \left(\mathbf{h}(\bar{\mathbf{x}} + \Delta \mathbf{x}) - (\bar{\mathbf{y}} + \Delta \mathbf{y}) \right) = \mathbf{0}.$$

The Jacobian matrix J is a function of $\mathbf{x}$, but if $\Delta \mathbf{x}$ is assumed small, J is almost constant in the neighborhood and measurement equations can be linearized $\mathbf{h}(\hat{\mathbf{x}}) \approx \mathbf{h}(\bar{\mathbf{x}}) + \mathbf{J}\Delta \mathbf{x}$. We get

$$(\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \left(\underbrace{\mathbf{h}(\bar{\mathbf{x}})}_{=\bar{\mathbf{y}}} + \mathbf{J}\Delta \mathbf{x} - (\bar{\mathbf{y}} + \Delta \mathbf{y}) \right) = \mathbf{0}$$

$$(\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \mathbf{J}\Delta \mathbf{x} - (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \Delta \mathbf{y} = \mathbf{0}$$

$$\Delta \mathbf{x} = (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \Delta \mathbf{y}.$$

If the measurement errors are normally distributed, $\Delta \mathbf{y} \sim N(\mathbf{0}, \Sigma)$, then it follows from Theorem 4 (p. 10) that

$$\Delta \mathbf{x} \sim N \left(\mathbf{0}, (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \Sigma \mathbf{J} (\mathbf{J}^T \mathbf{J})^{-1} \right). \tag{2.5}$$

In the special case where measurement errors are distributed independently and identically, i.e. $\Sigma = \sigma^2 I$, the location error simplifies to $\Delta \mathbf{x} \sim \mathrm{N}\left(\mathbf{0}, \sigma^2 (\mathbf{J}^T \mathbf{J})^{-1}\right)$. Then the covariance of estimation error derived from measurement errors is simply the product of the variance σ^2 of measurement errors and matrix $(\mathbf{J}^T \mathbf{J})^{-1}$ which depends on measurement geometry. In the context of GPS, this matrix is called DOP matrix (Dilution of Precision) [26, sections 6.A.1–2], from which different DOP-numbers can be computed as follows:

$$\text{Global/Geometric Dilution of Precision: } GDOP = \sqrt{\mathrm{tr}(\mathbf{J}^T \mathbf{J})^{-1}}$$

$$\text{Position Dilution of Precision: } PDOP = \sqrt{\mathrm{tr}[(\mathbf{J}^T \mathbf{J})^{-1}]_{(1:3,1:3)}}$$

$$\text{Horizontal Dilution of Precision: } HDOP = \sqrt{\mathrm{tr}[(\mathbf{J}^T \mathbf{J})^{-1}]_{(1:2,1:2)}}$$

$$\text{Vertical Dilution of Precision: } VDOP = \sqrt{[(\mathbf{J}^T \mathbf{J})^{-1}]_{(3,3)}}$$

$$\text{Time Dilution of Precision: } TDOP = \sqrt{[(\mathbf{J}^T \mathbf{J})^{-1}]_{(4,4)}}$$

DOP figures are useful because they compress the quality of the satellite geometry into one number. For example, the position error standard deviation can be estimated simply by multiplying measurement error deviation with the PDOP figure. If measurements have different deviations, however, then the error formula (2.5) does not simplify into DOP form, see Computer Exercise 2.5.

Let us emphasize that in the derivation of formula (2.5), the error $\Delta \mathbf{x}$ was assumed small so that the measurement function $\mathbf{h}(\mathbf{x})$ could be linearized in the neighborhood of the estimate. The more curved the measurement equations are, the poorer the estimate for location error obtained this way is. (see Exercise 2.17)

2.4   Closed form solutions

Closed form solution, or direct solution, is a general name for various non-iterative algorithms, where an initial guess is not needed to start the iteration. The advantage is that termination conditions are not needed for iteration, and there is no fear of converging into a wrong solution. If the problem has several solutions, the closed form solution gives them all. There are no general closed form solutions capable of digesting any kind of measurement, but it has to be derived separately for each measurement model, e.g. [4, 9, 24, 29].

In geometric solutions, every measurement is thought to define a surface in the parameter space on which the measurement equation is fulfilled. For instance, range measurement defines a spherical surface in three-dimensional location-space, range difference measurement defines a hyperboloid, etc. The problem's solutions are all the points where all measurement equations are satisfied, i.e. where all measurement surfaces intersect. Another approach, of which the Bancroft method is shown below as an example, is algebraic where different kinds of calculation tricks are used for trying to find a special solution for the measurement equations.

Closed form solutions are useful as initial guesses for iterative methods, but otherwise mostly in theoretical research and problem visualization, because measurement errors are difficult to model with them.

Example 16 (Bancroft method [4]). *If all the measurements are pseudorange measurements (for instance if only GPS-measurements are used), then a least squares solution*[*] *can be computed in closed form as follows. Start from the system of measurement equations*

$$\|\mathbf{s}_1 - \mathbf{x}\| + b = y_1$$
$$\vdots$$
$$\|\mathbf{s}_n - \mathbf{x}\| + b = y_n.$$

Trick 1: Move the b's to the right hand side of the equations and square both sides of the equations. We then get n equations of the form:

$$\|\mathbf{s}_i - \mathbf{x}\|^2 = (y_i - b)^2$$
$$\Leftrightarrow \|\mathbf{s}_i\|^2 - 2\mathbf{s}_i^T\mathbf{x} + \|\mathbf{x}\|^2 = y_i^2 - 2y_ib + b^2.$$

Remark. Squaring both sides of an equation can cause the solution to have several branches that have to be treated separately. In this case this does not happen because the norm is non-negative, and thus the left-hand side has to be as well.

Trick 2: Collect the squared terms into a new variable $\lambda = \|\mathbf{x}\|^2 - b^2$, then we get a linear equation whose variables are $\mathbf{x}$, b and λ:

$$2\mathbf{s}_i^T\mathbf{x} - 2y_ib = \lambda + \|\mathbf{s}_i\|^2 - y_i^2.$$

Collect all the equations into a linear system and solve for $\mathbf{x}$ and b in least squares sense:

$$\begin{bmatrix} 2\mathbf{s}_1^T & -2y_1 \\ & \vdots & \\ 2\mathbf{s}_n^T & -2y_n \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}\lambda + \begin{bmatrix} \|\mathbf{s}_1\|^2 - y_1^2 \\ \vdots \\ \|\mathbf{s}_n\|^2 - y_n^2 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{x} \\ b \end{bmatrix} = \underbrace{\begin{bmatrix} 2\mathbf{s}_1^T & -2y_1 \\ & \vdots & \\ 2\mathbf{s}_n^T & -2y_n \end{bmatrix}^\dagger \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}}_{=\mathbf{p}}\lambda + \underbrace{\begin{bmatrix} 2\mathbf{s}_1^T & -2y_1 \\ & \vdots & \\ 2\mathbf{s}_n^T & -2y_n \end{bmatrix}^\dagger \begin{bmatrix} \|\mathbf{s}_1\|^2 - y_1^2 \\ \vdots \\ \|\mathbf{s}_n\|^2 - y_n^2 \end{bmatrix}}_{=\mathbf{q}}$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{x} \\ b \end{bmatrix} = \mathbf{p}\lambda + \mathbf{q} = \begin{bmatrix} \mathbf{d} \\ f \end{bmatrix}\lambda + \begin{bmatrix} \mathbf{e} \\ g \end{bmatrix}$$

where $A^\dagger = (A^TA)^{-1}A^T$ and vectors $\mathbf{p}$ and $\mathbf{q}$ can be computed from known quantities. The original system of equations is thus fulfilled (in a least squares sense) if and only if

$$\mathbf{x} = \mathbf{d}\lambda + \mathbf{e}$$
$$b = f\lambda + g. \tag{2.6}$$

---

[*]For over-determined systems, the solution is not the same as obtained with iterative methods, because the methods minimize slightly different functions.

Substitute these back into the definition of λ:

$$\lambda = \|\mathbf{x}\|^2 - b^2 = \|\mathbf{d}\lambda + \mathbf{e}\|^2 - (f\lambda + g)^2,$$

expand square terms and rearrange into

$$(\|\mathbf{d}\|^2 - f^2)\lambda^2 + (2\mathbf{d}^T\mathbf{e} - 2fg - 1)\lambda + \|\mathbf{e}\|^2 - g^2 = 0$$

which is a second degree polynomial with respect to λ *and all other terms are known. Solve roots and compute the corresponding* $\mathbf{x}$ *and b by substituting the roots back into formula (2.6). Although there are two solution candidates, GPS satellite geometry is such that the other solution is in outer space and can be neglected.*[*]

2.5   Maximum likelihood

A more general form of measurement model The conditional probability density function is a more general form of measurement model. It is denoted with $p(\mathbf{y} \mid \mathbf{x})$, which can be read "the probability that $\mathbf{y}$ is measured in location $\mathbf{x}$" (see Section 1.3.3). When $\mathbf{x}$ is considered as the variable and measurements $\mathbf{y}$ as a constant in the expression, the function is then called the *likelihood function* (or simply the likelihood) and it is interpreted such that the larger the value of likelihood with measurements $\mathbf{y}$ and in location $\mathbf{x}$, the better the location in question "fits" the measurements. The likelihood function is sometimes denoted $L(\mathbf{x}|\mathbf{y})$ to emphasize that $\mathbf{x}$ is considered a variable.

If the measurement model is given in equation form $\mathbf{y} = \mathbf{h}(\mathbf{x}) + \mathbf{v}$, then it is easily written in likelihood form because $\mathbf{v} = \mathbf{y} - \mathbf{h}(\mathbf{x})$ and the distribution of $\mathbf{v}$ is known, so

$$p(\mathbf{y} \mid \mathbf{x}) = p_{\mathbf{v}}(\mathbf{y} - \mathbf{h}(\mathbf{x})),$$

where $p_{\mathbf{v}}$ is the known error distribution. Converting a likelihood function into a measurement equation is however possible only in special cases. Maximum likelihood as a method is therefore more generally usable than the ones previously presented.

When the measurement model is in likelihood form and we obtain a measurement, the solution $\hat{\mathbf{x}}$ is sought such that its likelihood in light of the measurements is as large as possible: $\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})$. To accomplish this, we can use nonlinear optimization methods, for instance the Newton method.

Example 17. *Consider a system of n measurements*

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) + \mathbf{v}$$

where the error $\mathbf{v}$ *is normally distributed,* $\mathbf{v} \sim \mathrm{N}(0, \Sigma)$. *Remember that the density function of a multidimensional normal distribution is*

$$p_{\mathbf{v}}(\mathbf{z}) = C\exp\left(-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}\right)$$

---

[*]Unless the application is e.g. space rocket positioning…

(the constant $C = (2\pi)^{-n/2}\det(\Sigma)^{-1/2}$ is not relevant here). Now the likelihood function is

$$p(\mathbf{y} \mid \mathbf{x}) = p_{\mathbf{v}}(\mathbf{y} - \mathbf{h}(\mathbf{x})) = C\exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{h}(\mathbf{x}))^T\Sigma^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}))\right).$$

Because the covariance matrix Σ is positive definite, the exponential function's argument is always non-positive. Therefore the solution with largest likelihood is found by minimizing the expression $(\mathbf{y} - \mathbf{h}(\mathbf{x}))^T\Sigma^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}))$, which is exactly the same as weighted least squares cost function with weight matrix Σ^{-1}.

Thus, for normally distributed errors, the maximum likelihood and weighted least squares method are identical if the inverse of the measurement covariance matrix is used as the weight matrix.

2.6 Bayesian solution

The likelihood function $p(\mathbf{y} \mid \mathbf{x})$ does not have a concrete interpretation as such. A more interesting quantity would be the probability of the state (i.e. location) conditioned on obtained measurements $p(\mathbf{x} \mid \mathbf{y})$. This quantity can be computed using Bayes' formula:

$$p(\mathbf{x} \mid \mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y} \mid \mathbf{x})}{\int p(\mathbf{x})p(\mathbf{y} \mid \mathbf{x})d\mathbf{x}} \tag{2.7}$$

where $p(\mathbf{y} \mid \mathbf{x})$ is the likelihood function of the measurement model and $p(\mathbf{x})$ is *prior distribution*, independent of measurements, that gives an "educated guess" of possible values of $\mathbf{x}$ in the form of a density function[*]. When the obtained measurement $\mathbf{y}$ is substituted into Bayes' formula, the result is the *posterior distribution* i.e. the conditional probability distribution for state $\mathbf{x}$.

Bayes' formula gives a probability distribution $p(\mathbf{x} \mid \mathbf{y})$, and this is more informative than just the state estimate $\hat{\mathbf{x}}$. The distribution contains all information we have on the state, based on prior information and measurement. From the posterior distribution we can, when needed, compute expectation or maximum likelihood estimate, covariance, confidence intervals, etc. and it can be used as prior information for the solution of next time instant .

If densities are any more complicated, then the posterior density cannot usually be found analytically, and we resort to Monte Carlo -integration which works for arbitrary functions. More about this in Section 3.5.

Example 18. *Consider the one-dimensional example where we measure the location of an object in a 10 meter long pipe. Let the measurement error distribution be $\mathrm{N}(0, 3^2)$.*

The likelihood function is now the normal density function (the constant coefficient is not relevant)

$$p(y \mid x) = C\exp\left(-\frac{(x-y)^2}{2\cdot 3^2}\right)$$

---

[*]If we know nothing beforehand about $\mathbf{x}$, the prior can be chosen to have uniform density.

and a uniform distribution inside the pipe is used as prior:

$$p(x) = \begin{cases} \frac{1}{10} & \textit{if } 0 \le x \le 10 \\ 0 & \textit{otherwise} \end{cases}$$

If measurement is, say $y = 8$ meters, then using Bayes formula (2.7) we get

$$p(x \mid y) = \frac{p(x)p(8 \mid x)}{\int p(x)p(8 \mid x)dx} = \frac{p(x)C\exp\left(-\frac{(x-8)^2}{2\cdot 3^2}\right)}{\int p(x)C\exp\left(-\frac{(x-8)^2}{2\cdot 3^2}\right)dx}$$

$$= \begin{cases} \dfrac{\exp\left(-\frac{(x-8)^2}{18}\right)}{\int_0^{10}\exp\left(-\frac{(x-8)^2}{18}\right)dx} & \textit{if } 0 \le x \le 10 \\ 0 & \textit{otherwise} \end{cases}$$

$$= \begin{cases} 0.1788\exp\left(-\frac{(x-8)^2}{18}\right) & \textit{if } 0 \le x \le 10 \\ 0 & \textit{otherwise} \end{cases}.$$

Prior information was used here to restrict the estimate to be inside the allowed area. If the rough location of the object in the pipe had been known more precisely, this information could also have been used as prior.

To the best of our knowledge, the obtained posterior distribution $p(x|y)$ now describes the object's location probability in different parts of the pipe. From the posterior we can compute any estimates required, such as expectation $\mu = \int xp(x|y)dx \approx 6.76$ and variance $\sigma^2 = \int (x-\mu)^2 p(x|y)dx \approx 2.12^2$. Compared to measurement $y = 8$, the use of prior information thus moved the

estimate a little more to the center and reduced estimate's standard deviation from 3 meters to 2.12 meters. If we were to get further independent measurements (assuming the object does is not moving), then the currently solved posterior can be used as new prior, thus using all the information from all the measurements optimally.
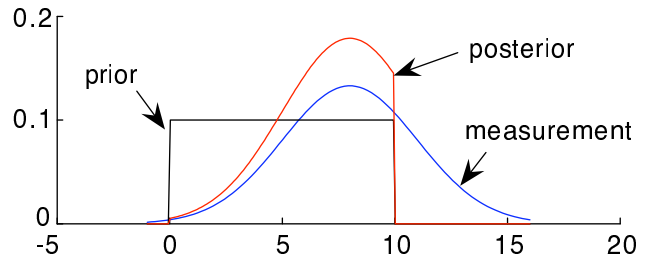
2.6.1 Bayesian error analysis

It was previously mentioned that a Bayesian location solution produces a probability density function of the location, from which we can compute both location estimate and its mean square error estimate

$$MSE = \int (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T p(\mathbf{x} \mid \mathbf{y})d\mathbf{x}$$

where the integral is usually computed numerically [28]. If the estimate $\hat{\mathbf{x}}$ is the expectation $E(\mathbf{x} \mid \mathbf{y})$, then the mean square error is by definition the posterior covariance.

Example 19. *Here we show that for normally distributed measurements, the error covariance is (nearly) the same as obtained in least squares method error analysis. Let measurement errors' distribution be* $\mathbf{v} \sim N(\mathbf{0}, \sigma^2 I)$, *and prior* $p(\mathbf{x})$ *be uniform distributed*[*]. *Now the likelihood function is*

$$p(\mathbf{y} \mid \mathbf{x}) = C \exp\left(-\frac{(\mathbf{h}(\mathbf{x}) - \mathbf{y})^T (\mathbf{h}(\mathbf{x}) - \mathbf{y})}{2\sigma^2} \right)$$

and thus

$$
\begin{aligned}
p(\mathbf{x} \mid \mathbf{y}) &= \frac{p(\mathbf{x}) p(\mathbf{y} \mid \mathbf{x})}{\int p(\mathbf{x}) p(\mathbf{y} \mid \mathbf{x}) d\mathbf{x}} = \frac{p(\mathbf{y} \mid \mathbf{x})}{\int p(\mathbf{y} \mid \mathbf{x}) d\mathbf{x}} \\
&= \frac{C \exp\left(-\frac{(\mathbf{h}(\mathbf{x}) - \mathbf{y})^T (\mathbf{h}(\mathbf{x}) - \mathbf{y})}{2\sigma^2} \right)}{\int C \exp\left(-\frac{(\mathbf{h}(\mathbf{x}) - \mathbf{y})^T (\mathbf{h}(\mathbf{x}) - \mathbf{y})}{2\sigma^2} \right) d\mathbf{x}} \\
&= C_2 \exp\left(-\frac{(\mathbf{h}(\mathbf{x}) - \mathbf{y})^T (\mathbf{h}(\mathbf{x}) - \mathbf{y})}{2\sigma^2} \right).
\end{aligned}
$$

If we use the same linearization in the neighborhood of true location as in least squares analysis, $\mathbf{h}(\mathbf{x}) \approx \mathbf{h}(\bar{\mathbf{x}}) + J(\bar{\mathbf{x}} - \mathbf{x}) = \bar{\mathbf{y}} + J\Delta\mathbf{x} = \mathbf{y} - \Delta\mathbf{y} + J\Delta\mathbf{x}$, *then we get*

$$p(\mathbf{x} \mid \mathbf{y}) \approx C_2 \exp\left(-\frac{(J\Delta\mathbf{x} - \Delta\mathbf{y})^T (J\Delta\mathbf{x} - \Delta\mathbf{y})}{2\sigma^2} \right),$$

which, unlike the corresponding least squares variance, depends on the realized measurement error $\Delta\mathbf{y}$. *For illustrative purposes, see what happens when* $\Delta\mathbf{y} = \mathbf{0}$ *(reasonable because the measurement errors are zero centered):*

$$
\begin{aligned}
p(\mathbf{x} \mid \mathbf{y})|_{\Delta\mathbf{y}=\mathbf{0}} &\approx C_2 \exp\left(-\frac{1}{2} \frac{\Delta\mathbf{x}^T J^T J \Delta\mathbf{x}}{\sigma^2} \right) \\
&= C_2 \exp\left(-\frac{1}{2} \Delta\mathbf{x}^T \left[\sigma^2 (J^T J)^{-1} \right]^{-1} \Delta\mathbf{x} \right)
\end{aligned}
$$

which is density function of distribution $\Delta\mathbf{x} \sim N(\mathbf{0}, \sigma^2 (J^T J)^{-1})$ *i.e. the same result which was obtained in previous section for least squares method. The Bayesian error estimates in this case thus fluctuate around the one given by least squares, depending on the realized measurements.*

---

[*]The prior is often chosen to be a uniform density in whole space, which is a contradiction because a probability density function's integral should equal to 1. In practice, the uniformly distributed prior cancels out from the computations, and it can be thought that we used prior that is uniformly distributed in an area "large enough" and outside of it remains so little probability mass that it does not have any significant effect on the results. Another approach is to use a Gaussian prior with covariance λI and let $\lambda \to \infty$

2.6.2 Confidence intervals

In one-dimensional case, confidence interval means an interval inside of which the estimate is with e.g. 99% probability.* Confidence interval is usually centered around the estimate, but nothing prevents setting it asymmetrically if needed. In multidimensional case, we usually use confidence spheres or ellipsoids centered around the estimate, but again nothing stops us from using area of any shape, as long as it contains the desired fraction of the distribution.

If the estimate's distribution is known, then the probability of being inside a certain interval is computed by integrating the distribution over the desired interval:

$$P(a \leq x \leq b \mid y) = \int_a^b p(x \mid y)dx.$$

In the case of Gaussian distribution, we can use the precomputed values of the cumulative distribution function and its inverse function, and get the familiar rules of thumb: inside $\pm 2\sigma$ is 95% of the probability, inside $\pm 3\sigma$ is 99.7% etc.

Even when the distribution of the estimate is not known, its variance can usually be approximated. Confidence intervals derived from the normal distribution do not hold in general, but we can apply Chebyshev's inequality, which gives a lower bound for how much probability mass lies outside a certain interval when the distribution variance is σ^2:

$$P(|x - \hat{x}| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}. \tag{2.8}$$

For example, according to this formula, the probability mass lying over three sigmas away from the average can be at most 1/9, which means that whatever the distribution is, there is always at least $8/9 \approx 89\%$ of the probability inside the interval $\hat{x} \pm 3\sigma$.

The corresponding n-dimensional Chebyshev inequality is

$$P\left(\|\mathbf{x} - \hat{\mathbf{x}}\| \geq \lambda\sqrt{\text{trace}(\Sigma)}\right) \leq \frac{1}{\lambda^2}$$

where Σ is the covariance matrix of the distribution. If variables have large correlations or they have variances with different orders of magnitude, then the less simplified formula

$$P\left(\sqrt{(\mathbf{x} - \hat{\mathbf{x}})^T \Sigma^{-1} (\mathbf{x} - \hat{\mathbf{x}})} \geq \lambda\sqrt{n}\right) \leq \frac{1}{\lambda^2} \tag{2.9}$$

gives a significantly tighter limit.

---

*This is the Bayesian interpretation which we prefer over the traditional one in this context. Basically the question is how randomness should be modeled mathematically. According to the frequentistic interpretation, probability means the proportion of successes if a test is repeated "many" times. In the Bayesian interpretation probability is a subjective measure that tells us how much we know (or think we know) about something.

The difference in the views is seen e.g. in the interpretation of the 99% confidence interval. In the frequentistic sense the true location is some fixed value and in 99% of the cases we manage to compute the confidence interval such that it contains the true location. The Bayesian interpretation is slightly more intuitive: we compute the confidence interval just once, and the true location is considered random and falls inside our computed confidence interval with 99% probability.

Formula (2.9) can also be used for defining the equation for confidence ellipsoid, for instance the 95% confidence ellipsoid results from choosing on the right hand side of the equation $1/\lambda^2 = 0.05$, meaning that at most 5% of the probability is allowed to be outside the ellipsoid. Thus, $\lambda = \sqrt{1/0.05}$, and the ellipsoid

$$E_{95\%} = \left\{ \mathbf{x} \mid (\mathbf{x} - \hat{\mathbf{x}})^T \Sigma^{-1} (\mathbf{x} - \hat{\mathbf{x}}) < \frac{n}{0.05} \right\}$$

contains at least 95% of the probability mass (compare with Exercise 1.10 on page 20).

2.7 Conclusion and observations

- The static positioning problem can be formulated as either solving a nonlinear system of equations or seeking maximum likelihood or the maximum of the Bayesian posterior. Problems are solved either with standard optimization methods or with numerical integration [28], which we return to in Section 3.5.

- If measurement equations are nearly linear such that $J(\mathbf{x}) = \mathbf{h}'(\mathbf{x})$ is nearly constant near the solution (then it can be denoted only by J) and measurement errors are normally distributed or can be modeled as such, then the least squares estimate and Bayesian mean estimate are practically identical.

- Closed form solutions are harder to derive, and usually more sensitive to measurement errors. They are however sometimes useful for searching initial guess or all the possible solutions for the problem.

Exercises

2.1. The unknown to be solved is the two-dimensional position $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$. Write the measurement modelm when the measurement y is

   (a) a noisy measurement of the coordinate x_2,

   (b) a noisy range from a fixed station at $\mathbf{s} = \begin{bmatrix} s_1 & s_2 \end{bmatrix}^T$,

   (c) a noisy angle to the station $\mathbf{s}$ so that zero angle points north, and the angle grows clockwise,

   (d) the difference of noisy ranges to stations $\mathbf{s}^1$ and $\mathbf{s}^2$.

2.2. An ideal INS unit measures the velocity vector in B-coordinates and the rotation vector $\mathbf{p}$ defining the attitude of the B-frame with respect to the local L-frame. Write the measurement equation assuming $\mathbf{p}$ is error-free.

2.3. The unknown to be solved is the 2D position $\mathbf{x}$. The measurement model is

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) + \mathbf{v} = \begin{bmatrix} \|[100\ 0]^T - \mathbf{x}\| \\ \|[0\ 50]^T - \mathbf{x}\| \end{bmatrix} + \mathbf{v}, \quad \text{where } \mathbf{v} \sim N(0, 10^2 I)$$

(a) Give the formulas for the residual $\mathbf{p}(\mathbf{x})$ and its derivative matrix $J = \mathbf{p}'(\mathbf{x})$ kaavat.

(b) Let the measurement be $\mathbf{y} = \begin{bmatrix} 108 & 46 \end{bmatrix}^T$. What is the first step of the Gauss-Newton method, if the starting point is $\mathbf{x} = [0\ 0]^T$?

2.4. Derive the iteration step of the Weighted Gauss-Newton method (Algorithm 2), that is, show that the step minimizes the cost $\mathbf{p}(\mathbf{x})^T \Sigma^{-1} \mathbf{p}(\mathbf{x})$.

––––––––

2.5. Derive the distribution of the error of the Weighted Gauss-Newton method (Algorithm 2) similarly to Section 2.3.1. The weighting matrix is W and the measurement covariance Σ.

2.6. A radar measures the range to a target and the rate of the change of the range, i.e. the radial velocity. The antenna is installed on a rotating rod, so that it's coordinates at time t are $\mathbf{s}(t) = [\cos t\ \sin t]^T$.
Write the measurement equation and its Jacobian matrix at time t with respect to the state vector $[r_1\ r_2\ r_1'\ r_2']^T$.
(Note: be careful to distinguish between differentation with respect to time and differentation with respect to state.)

2.7. Given a linear measurement model $\mathbf{y} = H\mathbf{x} + \mathbf{v}$ and a vector $\mathbf{x}_0$, derive an estimator that minimizes the formula $\|\mathbf{y} - H\mathbf{x}\|^2 + \|\mathbf{x}_0 - \mathbf{x}\|^2$.

2.8. Derive a closed-form solution (similar to Example 16 on page 30) for the measurement system

$$\|\mathbf{s}_1 - \mathbf{x}\| = y_1$$
$$\vdots$$
$$\|\mathbf{s}_n - \mathbf{x}\| = y_n.$$

Then apply your algorithm to the situation of Exercise 2.3.

2.9. With WLAN signals, for example, it is not often possible or reasonable to convert the measured signal strength into a range measurement. Instead, in *fingerprint methods*, the positioning system is first calibrated by choosing a set of locations $\mathbf{p}_i$, $i = 1\dots n$, and measuring the signal strength of all stations $\mathbf{a}_i \in \mathbb{R}^m$ in each calibration point. After this, positioning is performed by comparing received measurements with the calibration database.
Write the measurement model assuming that measurement noise is i.i.d. $N(0, \sigma^2)$ and that the calibration database contains the accurate expected values of the measurements at each calibration point. Consider also, what kind of algorithm would be needed to solve position from your measurement model.

2.10. Write the likelihood functions $p(y \mid x)$ corresponding to the following measurement equations:

(a) $y = x^2 + v, \quad v \sim \text{Uniform}[-1, 1]$

(b) $y = h(x) + v, \quad v \sim \text{Uniform}[-1, 1]$

(c) $y = 1$ if $\|\mathbf{s} - \mathbf{x}\| \leq 100$, otherwise $y = 0$

(d) $y = 1$ if $\|\mathbf{s} - \mathbf{x}\| + v \leq 100$, $v \sim \text{N}(0, 1)$, otherwise $y = 0$

(e) $y = \|\mathbf{s} - \mathbf{x}\|^2, \quad \mathbf{s} \sim \text{N}(\hat{\mathbf{s}}, \mathbf{I})$

Note that in (c) and (d), the measurement y is binary, so that its likelihood function consists of just the two values $p(0 \mid x)$ and $p(1 \mid x)$.

2.11. The measurement model is $y = \lfloor x + v \rfloor$, where $v \sim \text{N}(0, (1/2)^2)$ and $\lfloor \cdot \rfloor$ denotes the floor function.

(a) What is the likelihood function of the measurement? Sketch graphs of the likelihood as a function of y (with some fixed value of x), and as a function of x (with some fixed y).

(b) What is the likelihood of n independent measurements?

(c) (optional Matlab exercise) The values of five independent measurements are $0, -2, -1, 0$ ja -1. Ignoring the floor function, the estimate of x would be the mean of the measurements, i.e. -0.8.

Plot the joint likelihood over some suitable interval, for example $[-2, 1]$, and estimate the maximum likelihood point from the graph.

2.12. Derive the Bayes formula (2.7) from the definition of conditional density (page 12).

2.13. When both the prior and likelihood are one-dimensional normal distributions, also the posterior is a normal distribution. Derive its parameters.

2.14. Let the density of the prior be $p(x) = \begin{cases} 1/2 & \text{when } -1 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$

What is the mean and variance of this distribution?

The measurement model is $p(y \mid x) = \begin{cases} (y+1)/2 & \text{when } x \geq 0 \text{ and } -1 \leq y \leq 1 \\ (1-y)/2 & \text{when } x \leq 0 \text{ and } -1 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$

Work out what this means, e.g. how the value of x affects the distribution of measured y. Compute the posterior distribution, and based on that the Bayesian state estimate and its variance.

2.15. Consider the Chebyshevin inequality (2.8).

    (a) If x is distributed according to a *two-sided exponential distribution*, $p(x) = \frac{1}{2}e^{-|x|}$, its expectation value is 0 and standard deviation $\sqrt{2}$. Compute the 67% confidence interval $[-a, a]$ both accurately and using the Chebyshevin inequality.

    (b) Give an example of a distribution that satisfies the equality $P(|x - \hat{x}| \geq \sigma) = 1$?

Computer exercises

2.16. Continues Exercise 2.3:

    (a) Program the Newton method (Algorithm 1 on page 27), and run it using initial guess $\mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Use the stopping tolerance $\delta = 10^{-3}$. What solution did you end up and how many iteration steps were needed?

    (b) Run algorithm again using initial guess $\mathbf{x}_0 = \begin{bmatrix} 100 \\ 100 \end{bmatrix}$. What solution did you end up and how many iteration steps were needed?

    (c) Draw the residual norm in area $x \in [-50, 150]$, $y \in [-50, 150]$ with Matlab's `surf` command. Estimate how close to the correct solution initial guess should be so that iteration finds correct solution.

2.17. Continues previous exercise.
Because all the measurements have the same variance σ^2, location error covariance can be estimated in the form $\sigma^2 (J^T J)^{-1}$.
Let true location be $\mathbf{x}_{\text{true}} = [0\ 0]^T$. Location estimate distribution can be approximated numerically with simulated measurements. Now $\mathbf{y} \sim N([100\ 50], \sigma^2 I)$.

    (a) Compute location estimate a thousand times with different measurement realizations, and compute covariance of the obtained estimates and its error compared to computed covariance estimate $\sigma^2 (J^T J)^{-1}$ when $\sigma = 1$ and when $\sigma = 10$.

    (b) Draw DOP values in area $x \in [-50, 150]$, $y \in [-50, 150]$ with Matlab's `surf` command. It may be necessary to cut off too large values from the figure, e.g. to draw `min(DOP,10)`.

Chapter 3

Filtering

SIMO ALI-LÖYTTY

A filter uses all the earlier measurements in addition to the current ones for computing the current state estimate. A basic prerequisite for the use of earlier measurements in positioning is a dynamic *state model*

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1}, \tag{3.1}$$

which describes how current state depends on the previous ones. We use the board term *state* instead of location, so that we can include for example velocities, clock biases, or other interesting quantities as unknowns. In equation (3.1) we use the following notations: $\mathbf{x}_k$ is the state, $\mathbf{f}_k$ is the state transfer function and $\mathbf{w}_k$ is the state model error. The subscript k refers to time instant t_k. In this chapter, the state $\mathbf{x}$ of an observable is interpreted as a stochastic process [3, 12, 16].

Definition 9 (Stochastic process). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and T a parameter set. Stochastic process is a mapping $\mathbf{x} : \Omega \times T \to \mathbb{R}^n$, such that for every fixed $t \in T$, $\mathbf{x}(\cdot, t)$ is a random variable, which is usually denoted $\mathbf{x}_t$ or $\mathbf{x}(t)$.*

Because state is a stochastic process, computation of the state estimate is done in two phases. At each time instant, we find the state distribution conditioned on measurements:

$$p_{\mathbf{x}_k|\mathbf{y}_{1:k}}(x_k|y_{1:k}).$$

Here the notation $y_{1:k}$ is used for all measurements up to current time instant. Measurements are also interpreted as realizations of random variables. Measurement equations are the same as in the static case (2.1):

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \tag{3.2}$$

where $\mathbf{y}_k$ are the measurements, $\mathbf{h}_k$ is measurement function and $\mathbf{v}_k$ measurement error. In addition to the state model and the measurement equations, we also need an initial state $\mathbf{x}_0$ for the computation of the conditional distribution. Once we have the conditional state distribution,

we can compute an estimate that is optimal with respect to some chosen criterion. These optimal estimates are handled in Section 3.4. From the state distribution, we can compute also other interesting quantities such as the covariance matrix of the estimation error.

In order to do estimation within reasonable time, we must be able to compute the state's conditional distribution recursively instead of reusing the whole measurement history at each time instant. This can be done if we make some independence assumptions. The first assumption is that errors $\mathbf{w}_k$ and $\mathbf{v}_k$ are mutually independent and *white noise*, which means that the errors independent of past errors. For simplicity we also assume that the errors are zero-mean; if necessary we can always change the state transfer- and measurement functions so that the errors become zero-mean. Additionally we assume that errors $\mathbf{w}_k$ and $\mathbf{v}_k$ are independent of the initial state $\mathbf{x}_0$. With these assumptions the state's conditional distribution can be computed recursively. This is covered in Sections 3.2 and 3.4.

Computing the conditional state distribution in a time series, from which we obtain the desired estimate, is called *filtering*. Filtering has many advantages compared to a static location solution. First of all, we can use all the measurements nearly optimally independent of how many measurements we get at each time instant. This enhances positioning accuracy considerably compared to static positioning. Additionally, a location solution can also be computed even if we have not got any measurement at the concerned time instant, because with the help of the state model we can compute the distribution of the coming state without new measurements.

Filtering makes it easy to use possible other information in positioning such as map information and sector/maximum range information of the base stations. Also, different kinds of error models that take into account multipath effects or other faulty measurements can be used to make positioning more fault tolerant. On the other hand, filtering has disadvantages also, first of all in several cases computation of conditional distribution is not possible analytically and although the computation can be done numerically it typically takes many times more computation time than static positioning. Additionally, filtering needs an initial state, the state dynamics model, and the distributions of the errors, any of which we do not necessarily need in static positioning. Often these are not well known, and thus the model does not necessarily correspond to reality as well as in static positioning.

As previously mentioned, we often cannot solve the conditional state distribution analytically. However, if both the state model and the measurement equations are linear and the errors and the initial state are normally distributed, then the state conditional distribution can be computed recursively. The algorithm doing this is called the Kalman filter, named after Rudolf E. Kalman [17]. The Kalman filter also solves a wider class of problems, as it remains the best linear unbiased estimator (BLUE, Best Linear Unbiased Estimator) even when the distributions are non-Gaussian. The Kalman filter is presented in Section 3.2 from the BLUE point of view, and in Exercise 3.5 you are asked to confirm that it also computes the state's conditional distribution. There also exists numerous approximations of the Kalman filter in cases where the state model and/or the measurement equations are not linear; these non-linear Kalman filters are presented in Section 3.3.

The general Bayesian filtering problem is presented in Section 3.4. Numerical methods for solving the general Bayesian filtering problem are presented in Section 3.5.

3.1 Constant velocity model

The state dynamics model (3.1), like the measurement model, is dependent on the observed system. Clearly, for a stationary target, a suitable state model is different than for a cyclist weaving through a crowd. The problem is that we rarely know the correct state model in advance. This is why, for instance, the interactive multiple model method (IMM) uses a set of several measurement/state models from which the filter tries to choose the best during runtime [34]. Another approach to this problem is adaptive filter that try to estimate and update the state model parameters along with the state itself. It is good to remember that often we have to settle for broad generalizations in the selection of the models, because the purpose is to get models that work as good as possible and are simple enough. In this section, we introduce a simple but commonly used model called the *constant velocity model*.

The constant velocity model has 6 state variables, the three-dimensional location and three-dimensional velocity. The clock bias and its derivative required in satellite positioning can be included in the same manner, as is done for example in the books [6, 10, 19, 27].

We model the state as a stochastic differential equation

$$d\mathbf{x} = \mathbf{F}\mathbf{x}dt + \mathbf{G}d\beta, \tag{3.3}$$

where

$$\mathbf{F} = \begin{bmatrix} 0 & \mathbf{I} \\ 0 & 0 \end{bmatrix} \quad \text{ja} \quad \mathbf{G} = \begin{bmatrix} 0 \\ \mathbf{I} \end{bmatrix},$$

Here the velocity error is modeled as Brownian motion, denoted with β. The diffusion matrix for the Brown motion β is chosen to be a simple diagonal matrix $\mathbf{Q}_c = \sigma_c^2 \mathbf{I}$, where σ_c^2 describes velocity error in the axis directions. To be more precise, σ_c^2 describes how much variance needs to be added to the velocity error within one second in north-south, east-west and up-down directions.

The solution of the stochastic differential equation (3.3) can be discretized into a state model (see e.g.[25, s.171] or [12, s.200])

$$\mathbf{x}_k = \Phi_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1},$$

where

$$\Phi_{k-1} = e^{(t_k - t_{k-1})\mathbf{F}} = \begin{bmatrix} \mathbf{I} & \Delta t \mathbf{I} \\ 0 & \mathbf{I} \end{bmatrix}.$$

The term $\Delta t = t_k - t_{k-1}$ is the length of time step. State model error $\mathbf{w}_k$ is white zero-mean normally distributed noise, which is assumed independent of initial condition $\mathbf{x}_0$. The covariance matrix $\mathbf{Q}_k$ of the state model error is

$$\mathbf{Q}_k = \mathbf{V}(\mathbf{w}_{k-1}) = \int_{t_{k-1}}^{t_k} \Phi(t_k, t)\mathbf{G}\mathbf{Q}_c\mathbf{G}^T\Phi(t_k, t)^T dt = \begin{bmatrix} \frac{1}{3}\Delta t^3 \mathbf{I} & \frac{1}{2}\Delta t^2 \mathbf{I} \\ \frac{1}{2}\Delta t^2 \mathbf{I} & \Delta t \mathbf{I} \end{bmatrix} \sigma_c^2.$$

42

3.2   The Kalman filter

The Kalman filter solves the filtering problem where the state model and the measurement model (measurement equations) are linear. Then the state model is

$$\mathbf{x}_k = \Phi_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \tag{3.4}$$

where $\mathbf{x}_k$ is the process state, Φ_k is the state transfer matrix and $\mathbf{w}_k$ is the state model error, which is assumed zero-mean white noise and independent of the errors in the measurements and the initial state. The covariance matrix of the state model error is denoted with Q_k. The measurement model is

$$\mathbf{y}_k = H_k\mathbf{x}_k + \mathbf{v}_k, \tag{3.5}$$

where $\mathbf{y}_k$ are the measurements, H_k is the measurement matrix (or system matrix) and $\mathbf{v}_k$ is the measurement model error, which is assumed zero-centered white noise and independent of the errors in the measurements and the initial state. The covariance matrix of the measurement model error is denoted with R_k, which is assumed positive definite. In addition to the state model and the measurement model, the Kalman filter needs initial state $\mathbf{x}_0$. We denote the the expectation of the initial state with vector $\hat{\mathbf{x}}_0$ and the covariance matrix with constant matrix P_0, which is assumed positive definite. The hat notation refers to an estimator, which is a function of the measurements. In the case of $\hat{\mathbf{x}}_0$, the estimator is a constant random variable, because at time instant t_0 we do not yet have any measurements. Covariance matrix P_0 can also be interpreted as a mean square error (MSE) matrix $E\left(\mathbf{e}_0\mathbf{e}_0^T\right)$, where $\mathbf{e}_0 = \mathbf{x}_0 - \hat{\mathbf{x}}_0$.

The purpose of the Kalman filter is to solve a best linear unbiased estimator for state $\mathbf{x}_k$. Because of the independence assumptions, this can be done recursively and then the estimator can be written as

$$\hat{\mathbf{x}}_k = \begin{bmatrix} J_{k-1} & K_k \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_{k-1} \\ \mathbf{y}_k \end{bmatrix}, \tag{3.6}$$

where $\hat{\mathbf{x}}_{k-1}$ is the best linear unbiased estimator for state $\mathbf{x}_{k-1}$ at time instant t_{k-1}. The corresponding mean square error matrix is denoted P_{k-1}. The measurements at time instant t_k are denoted with random variable $\mathbf{y}_k$.

In the remainder of this section, we derive such matrices J_{k-1} and K_k that (3.6) becomes the best linear unbiased estimator, i.e. the Kalman filter. The best here means that the mean square error of the estimator $E(\mathbf{e}_k^T\mathbf{e}_k) = \mathrm{tr}(E(\mathbf{e}_k\mathbf{e}_k^T))$ would be as small as possible. This estimator is called the least squares estimator (MSE-estimator). Because the estimator is required to be unbiased, we get

$$E(\mathbf{x}_k) = E(\hat{\mathbf{x}}_k) = \begin{bmatrix} J_{k-1} & K_k \end{bmatrix} \begin{bmatrix} E(\hat{\mathbf{x}}_{k-1}) \\ E(\mathbf{y}_k) \end{bmatrix}.$$

Using the unbiasedness of the estimator $\hat{\mathbf{x}}_{k-1}$, the state model (3.4) and the measurement model (3.5) we get

$$\Phi_{k-1}E(\mathbf{x}_{k-1}) = J_{k-1}E(\mathbf{x}_{k-1}) + K_kH_k\Phi_{k-1}E(\mathbf{x}_{k-1}).$$

This equation needs to hold for any expectation $E(\mathbf{x}_{k-1})$, thus

$$J_{k-1} = \Phi_{k-1} - K_k H_k \Phi_{k-1}. \tag{3.7}$$

The best linear unbiased estimator of state $\mathbf{x}_k$ which does not use measurements at the current time instant t_k is called the *prior estimator*. In other words, the prior estimator is the same form as the estimator in formula (3.6), but with matrix K_k set to zero. Then the prior estimator, denoted by a minus superscript, is according to formula (3.7)

$$\hat{\mathbf{x}}_k^- = \Phi_{k-1}\hat{\mathbf{x}}_{k-1}. \tag{3.8}$$

The mean square error matrix of the prior estimator is thus

$$\begin{aligned}
P_k^- &= E\left((\mathbf{x}_k - \hat{\mathbf{x}}_k^-)(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)^T\right) \\
&\overset{(3.4)}{=} E\left((\Phi_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1})(\Phi_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1})^T\right) \\
&\overset{\text{indep.}}{=} \Phi_{k-1}E\left(\mathbf{e}_{k-1}\mathbf{e}_{k-1}^T\right)\Phi_{k-1}^T + E\left(\mathbf{w}_{k-1}\mathbf{w}_{k-1}^T\right) \\
&= \Phi_{k-1}P_{k-1}\Phi_{k-1}^T + Q_{k-1}.
\end{aligned}$$

The estimator in formula (3.6) that also uses measurements at time instant t_k is called the *posterior estimator*. The error of the posterior estimator is

$$\begin{aligned}
\mathbf{e}_k &= \mathbf{x}_k - \hat{\mathbf{x}}_k \\
&= \mathbf{x}_k - (\Phi_{k-1} - K_k H_k \Phi_{k-1})\hat{\mathbf{x}}_{k-1} - K_k \mathbf{y}_k \\
&\overset{(3.5)}{=} (I - K_k H_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k^-) - K_k \mathbf{v}_k
\end{aligned} \tag{3.9}$$

According to the assumption, the error of the prior estimator $\mathbf{e}_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^-$ is independent of the measurement error $\mathbf{v}_k$, so that the mean square error matrix of the posterior estimator is

$$\begin{aligned}
P_k &= E\left(\mathbf{e}_k \mathbf{e}_k^T\right) \\
&\overset{(3.9)}{=} (I - K_k H_k)P_k^-(I - K_k H_k)^T + K_k R_k K_k^T \\
&\overset{\text{unbiased,ex.3.2}}{=} P_k^- - P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} H_k P_k^- \\
&\qquad + (K_k B - P_k^- H_k^T B^{-1})(K_k B - P_k^- H_k^T B^{-1})^T,
\end{aligned} \tag{3.10}$$

where $B = (H_k P_k^- H_k^T + R_k)^{\frac{1}{2}}$. The trace of P_k is at minimum when the trace of the matrix

$$(K_k B - P_k^- H_k^T B^{-1})(K_k B - P_k^- H_k^T B^{-1})^T \tag{3.11}$$

is minimized, because the other terms in the sum do not depend on matrix K_k. The trace of matrix (3.11) is smallest possible when $K_k B - P_k^- H_k^T B^{-1} = 0$ (Exercise 3.3), and the solution to this equation is

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1}. \tag{3.12}$$

This matrix is called the Kalman gain. Combining the results, we get the Kalman filter estimator

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + K_k(\mathbf{y}_k - H_k \hat{\mathbf{x}}_k^-), \tag{3.13}$$

where the prior estimator has been given in formula (3.8) and the Kalman gain in formula (3.12). The mean square error matrix of this estimator (3.10) can be written as

$$P_k = (I - K_k H_k) P_k^-.$$

The Kalman filter has been given in compact form in Algorithm 3.

In formula (3.13), the difference between the measurement $\mathbf{y}_k$ and the predicted measurement $\hat{\mathbf{y}}_k = H_k \hat{\mathbf{x}}_k^-$, i.e. $\mathbf{y}_k - H_k \hat{\mathbf{x}}_k^-$ is called the *innovation*. The mean square error matrix of the measurement prediction (the covariance matrix of the innovation) is $H_k P_k^- H_k^T + R_k$ (Exercise. 3.1). The use of the innovation is almost the only way of monitoring the quality of the filter solution during runtime. Because the covariance matrix of the innovation is known, we can do statistical tests on whether the realized measurement is probable or not. Many *robust filters* do these kinds of tests and if the result is that the measurement is unlikely, it is either not used at all or it is given less weight than the "likely" measurements. Additionally, many *adaptive filters* use innovations to define or update filter parameters during runtime, for instance the covariances of the errors [5, Chapter 11].

Algorithm 3 The Kalman filter

- The state model: $\quad\quad\quad\quad \mathbf{x}_k = \Phi_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \quad V(\mathbf{w}_{k-1}) = Q_{k-1}$
- The measurement model: $\quad \mathbf{y}_k = H_k\mathbf{x}_k + \mathbf{v}_k, \quad\quad\quad\quad V(\mathbf{v}_k) = R_k$
- The measurements: $\quad\quad\quad y_{1:m} = \{y_1, y_2, \ldots, y_m\}$
- The initial state estimate and its MSE: $\hat{x}_0$ and P_0

1. Set $k = 1$.

2.
   - The prior estimate: $\quad\quad\quad \hat{x}_k^- = \Phi_{k-1}\hat{x}_{k-1}$
   - The prior MSE: $\quad\quad\quad\quad P_k^- = \Phi_{k-1}P_{k-1}\Phi_{k-1}^T + Q_{k-1}$
   - The Kalman gain: $\quad\quad\quad K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1}$
   - The posterior estimate: $\quad \hat{x}_k = \hat{x}_k^- + K_k(y_k - H_k\hat{x}_k^-)$
   - The posterior MSE: $\quad\quad P_k = (I - K_k H_k)P_k^-$

3. Stop if $k = m$, otherwise set $k = k + 1$ and get back to step 2.

Example 20 (A Kalman filter). *Figure 3.1 shows an example of a situation where a two-dimensional constant velocity model has been used, see Section 3.1. Constant velocity model satisfies assumptions of Kalman filter. The parameters used in the example are as following:*

$$\mathbf{x}_0 = \begin{bmatrix} 0 & 0 & 5 & 0 \end{bmatrix}^T, \; P_0 = \begin{bmatrix} 10^2 I & 0 \\ 0 & 3^2 I \end{bmatrix},$$

$$\Phi = \begin{bmatrix} I & I \\ 0 & I \end{bmatrix}, \quad\quad Q = \begin{bmatrix} \frac{1}{3}I & \frac{1}{2}I \\ \frac{1}{2}I & I \end{bmatrix}, \quad\quad\quad (3.14)$$

$$H = \begin{bmatrix} I & 0 \end{bmatrix} \quad and \quad R = \begin{bmatrix} 30^2 & 0 \\ 0 & 50^2 \end{bmatrix}.$$

The true route over one minute period has been drawn with a black dashed line. The positioning system gives location measurements according to the measurement model at one second intervals (red dots linked with narrow lines). The Kalman filter uses the locations as measurements, and the resulting state estimate is drawn with a continuous blue line.
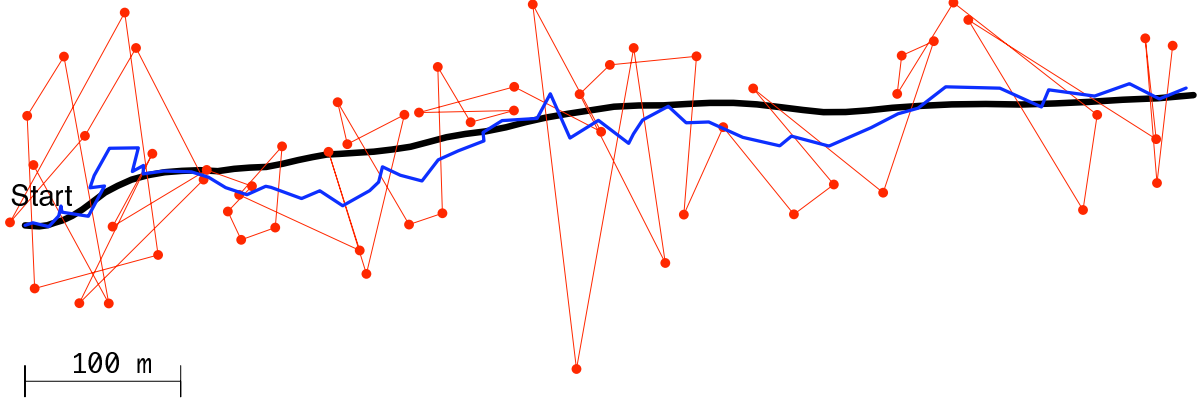
Figure 3.1: Example of a Kalman filter applied to two-dimensional location measurements.

3.3 Nonlinear Kalman filters

The Kalman filter is very popular for linear systems, because it is relatively easy to compute and requires few assumptions. For this reason, there are several methods of applying something similar to the Kalman filter also to nonlinear systems (3.1), (3.2). In this section, we introduce the basic ideas on which most Kalman filter extensions are based. First, we give a slightly generalized version of the BLU-estimator. Let

$$\mathrm{E}\left(\left[\begin{array}{c} \mathbf{x}_k \\ \mathbf{y}_k \end{array}\right]\right) = \left[\begin{array}{c} \bar{x}_k \\ \bar{y}_k \end{array}\right], \text{ and } \mathrm{V}\left(\left[\begin{array}{c} \mathbf{x}_k \\ \mathbf{y}_k \end{array}\right]\right) = \left[\begin{array}{cc} \mathrm{P}_{xx_k} & \mathrm{P}_{xy_k} \\ \mathrm{P}_{yx_k} & \mathrm{P}_{yy_k} \end{array}\right].$$

Now the BLU-estimator of state $\mathbf{x}_k$ is [5]

$$\begin{aligned} \hat{\mathbf{x}}_k &= \bar{x}_k + \mathrm{P}_{xy_k}\mathrm{P}_{yy_k}^{-1}(\mathbf{y}_k - \bar{y}_k), \\ \mathrm{E}((\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^T) &= \mathrm{P}_{xx_k} - \mathrm{P}_{xy_k}\mathrm{P}_{yy_k}^{-1}\mathrm{P}_{yx_k}. \end{aligned} \tag{3.15}$$

Nonlinear extensions of the Kalman filter try to solve the unknown quantities of equation (3.15), $\bar{x}_k, \bar{y}_k, \mathrm{P}_{xx_k}, \mathrm{P}_{xy_k} = \mathrm{P}_{yx_k}^T, \mathrm{P}_{yy_k}$ by using some approximation. For these kinds of algorithms we use a general name *nonlinear Kalman filter*, whose algorithm has been given on page 47. Next, we treat two commonly used nonlinear Kalman filters in more detail, the so-called extended Kalman filter (EKF) and the unscented Kalman filter (UKF) [34].

When using nonlinear Kalman filters, it is good to remember that they do not actually even approximate the optimal solution which is handled in Section 3.4. For this reason, EKF for instance may give completely false results like in the example on page 55. Nevertheless, in many

practical application nonlinear Kalman filters work fairly well (see Exercise 3.10) and they often require significantly less computation than the numerical approximations of the general Bayesian filter, which are handled in Section 3.5. Because of these reasons, nonlinear Kalman filters are popular in engineering.

Algorithm 4 Nonlinear Kalman filter (EKF and UKF)

- The state model: $\qquad \mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1}, \quad V(\mathbf{w}_{k-1}) = Q_{k-1}$
- The measurement model: $\quad \mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \qquad\qquad V(\mathbf{v}_k) = R_k$
- The initial state estimate, its MSE and the measurements: $\quad \hat{x}_0, \quad P_0$ and $y_{1:m} = \{y_1, y_2, \ldots, y_m\}$

1. Set $k = 1$.

2.
   - The prior estimate: $\qquad \hat{x}_k^- = \bar{x}_k$
   - The prior MSE: $\qquad\quad P_k^- = P_{xx_k}$
   - The posterior estimate: $\quad \hat{x}_k = \hat{x}_k^- + P_{xy_k}P_{yy_k}^{-1}(y_k - \bar{y}_k)$
   - The posterior MSE: $\qquad P_k = P_{xx_k} - P_{xy_k}P_{yy_k}^{-1}P_{yx_k},$

3. Stop, if $k = m$, otherwise set $k = k + 1$ and go back to step 2.

3.3.1 Extended Kalman filter

The extended Kalman filter is a nonlinear Kalman filter that is based on the first order Taylor approximations of the state model and the measurement model.

- The state model approximation:
  $\mathbf{x}_k \approx \mathbf{f}_{k-1}(\hat{x}_{k-1}) + \Phi_{k-1}(\mathbf{x}_{k-1} - \hat{x}_{k-1}) + \mathbf{w}_{k-1}$ where $\Phi_{k-1} = \mathbf{f}'_{k-1}(\hat{x}_{k-1})$ and

- The measurement model approximation:
  $\mathbf{y}_k \approx \mathbf{h}_k(\hat{x}_k^-) + H_k(\mathbf{x}_k - \hat{x}_k^-) + \mathbf{v}_k$, where $H_k = \mathbf{h}'_k(\hat{x}_k^-)$.

Then the unknown quantities in Algorithm 4 are easily computed:
$$\begin{bmatrix} \bar{x}_k \\ \bar{y}_k \end{bmatrix} = \begin{bmatrix} \mathbf{f}_{k-1}(\hat{x}_{k-1}) \\ \mathbf{h}_k(\hat{x}_k^-) \end{bmatrix}, \text{ and}$$
$$\begin{bmatrix} P_{xx_k} & P_{xy_k} \\ P_{yx_k} & P_{yy_k} \end{bmatrix} = \begin{bmatrix} P_k^- & P_k^- H_k^T \\ H_k P_k^- & H_k P_k^- H_k^T + R_k \end{bmatrix},$$

where $P_k^- = \Phi_{k-1}P_{k-1}\Phi_{k-1}^T + Q_{k-1}$. It is also possible to use a higher order Taylor approximation of the quantities listed above. On the other hand, higher order approximations often need more restrictive assumptions about the prior and posterior distributions, for instance normality, so that the quantities in formula (3.15) can be computed. Additionally, the analytical computation of even a first order Taylor approximation is difficult and in some cases impossible. This the reason for developing nonlinear Kalman filters that do not require the derivative of the state transfer function or the measurement function, such as the UKF in the next section.

3.3.2 Unscented Kalman filter

Unscented Kalman filter (UKF) is a derivative-free nonlinear Kalman filter based on a numerical integration method called the unscented transformation (UT). The UT approximates the expectation of function $\mathbf{f}(\mathbf{x})$ using the so-called σ-points $\{\chi_0, \ldots, \chi_{N-1}\}$ when the distribution of random variable $\mathbf{x}$ is known. This numerical integration method can be written as

$$E(\mathbf{f}(\mathbf{x})) = \int \mathbf{f}(x)p(x)dx \approx \sum_{i=0}^{N-1} \omega_i \mathbf{f}(\chi_i), \tag{3.16}$$

where weights ω_i and sigma-points χ_i are selected such that the approximation is accurate for certain degree polynomials $\mathbf{f}(x)$ when random variable $\mathbf{x}$ is normally distributed.

Let n be the state dimension, $\hat{x}$ be the expectation of random variable $\mathbf{x}$ and the covariance matrix be P. Then the generally used σ-points selection contains $2n+1$ points which are

$$
\begin{array}{ccc}
\text{Index } (i) & \text{Weight } (\omega_i) & \text{σ-point } (\chi_i) \\
\hline
0 & \frac{\kappa}{\kappa+n} & \hat{x} \\
1,\ldots,n & \frac{1}{2(\kappa+n)} & \hat{x}+\sqrt{\kappa+n}\sqrt{P}e_i \\
n+1,\ldots,2n & \frac{1}{2(\kappa+n)} & \hat{x}-\sqrt{\kappa+n}\sqrt{P}e_{i-n},
\end{array}
\tag{3.17}
$$

where $\sqrt{P}$ means such a matrix which has the property that $\sqrt{P}\sqrt{P}^T = P$. Parameter $\kappa > -n$ is a freely selectable constant. With these σ-points and weights, the approximation (3.16) is exact if the random variable $\mathbf{x}$ is normally distributed and function $\mathbf{f}(x)$ is a third degree polynomial. In addition, the choice $\kappa = 3 - n$ minimizes the error of the integral approximation of a fourth degree polynomial and it is therefore commonly used, see Exercise 3.6.

Let now $\{\chi_{0_{k-1}}, \ldots, \chi_{(N-1)_{k-1}}\}$ be the σ-points generated based on the posterior distribution at time instant t_k. Then the unknown quantities of Algorithm 4 can be computed:

$$\bar{x}_k = \sum_{i_{k-1}=0}^{N-1} \omega_{i_{k-1}} \mathbf{f}_{k-1}(\chi_{i_{k-1}}), \quad \bar{y}_k = \sum_{i_{k-1}=0}^{N-1} \omega_{i_{k-1}} \mathbf{h}_k(\chi_{i_{k|k-1}}),$$

$$P_{xx_k} = \sum_{i_{k-1}=0}^{N-1} \omega_{i_{k-1}}(\mathbf{f}_{k-1}(\chi_{i_{k-1}}) - \bar{x}_k)(\mathbf{f}_{k-1}(\chi_{i_{k-1}}) - \bar{x}_k)^T + Q_{k-1},$$

$$P_{xy_k} = \sum_{i_{k-1}=0}^{N-1} \omega_{i_{k-1}}(\chi_{i_{k|k-1}} - \bar{x}_k)(\mathbf{h}_k(\chi_{i_{k|k-1}}) - \bar{y}_k)^T \text{ and}$$

$$P_{yy_k} = \sum_{i_{k-1}=0}^{N-1} \omega_{i_{k-1}}(\mathbf{h}_k(\chi_{i_{k|k-1}}) - \bar{y}_k)(\mathbf{h}_k(\chi_{i_{k|k-1}}) - \bar{y}_k)^T + R_k,$$

where $\chi_{i_{k|k-1}} = \mathbf{f}_{k-1}(\chi_{i_{k-1}})$. Another way is to generate these σ-points of the prior distribution from a prior distribution whose expectation is $\bar{x}_k$ and whose covariance matrix is P_{xx_k}, then the results are a little different from the previous implementation.

3.4 Bayesian filter

In this section, we use the state model, the measurement model, and the initial state as well as the usual independence assumptions introduced at the beginning of the chapter on page 40. The aim in Bayesian filtering is to determine the state's conditional density function conditioned on the measurements,

$$p_{\mathbf{x}_k|\mathbf{y}_{1:k}}(x_k|y_{1:k}) \stackrel{\triangle}{=} p(x_k|y_{1:k}).$$

The abbreviated notation is used where it is easy to see from the context what the notation means. Assume that the state's conditional density function at the previous time instant is $p(x_{k-1}|y_{1:k-1})$. The conditional density function for the current instant conditioned on previous measurements, $p(x_k|y_{1:k-1})$, can be determined with the help of the Chapman-Kolmogorov equation

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1}.$$

This distribution is called the prior distribution. The conditional density function $p(x_k|x_{k-1})$ can be derived from the state model

$$p(x_k|x_{k-1}) \stackrel{(3.1)}{=} p_{\mathbf{w}_{k-1}}(x_k - \mathbf{f}_{k-1}(x_{k-1})).$$

Now, by Bayes' rule and the independence assumptions of the errors, we can derive a formula for the current state's conditional density function conditioned on current and previous measurements:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{\int p(y_k|x_k)p(x_k|y_{1:k-1})dx_k}.$$

This distribution is called the posterior distribution. The first term of the numerator

$$p(y_k|x_k) = p_{\mathbf{y}_k=y_k|\mathbf{x}_k}(y_k|x_k) \stackrel{(3.2)}{=} p_{\mathbf{v}_k}(y_k - \mathbf{h}_k(x_k))$$

is called the likelihood. Note that this is not a probability density function. The density function of the posterior distribution is proportional to the product of the prior distribution and the likelihood. The denominator is the normalization constant $p(y_k|y_{1:k-1})$, which is the conditional density function of the new measurement conditioned on previous measurements. From this the filtering can be continued recursively.

As already mentioned, in linear-Gaussian cases the Kalman filter is the exact solution for this filtering problem, see Exercise 3.5. Nevertheless, in the general case the posterior distributions cannot be analytically solved. Because of this, several numerical methods for the approximation of the posterior distribution have been developed. These methods are described in Section 3.5.

Optimal estimator

In practical applications, we usually are interested in finding an optimal state estimator and an estimate of the estimation error, rather than the full Bayesian posterior distribution. By "estimator" we mean a function whose arguments are the measurements and the initial state, and by "optimal" we mean that the estimator minimizes the expectation of some scalar cost function $L(\mathbf{x}_k - \hat{\mathbf{x}}_k)$, in short

$$\hat{\mathbf{x}}_k^{opt} = \text{argmin}_{\hat{\mathbf{x}}_k} \, \text{E}(L(\mathbf{x}_k - \hat{\mathbf{x}}_k)).$$

The cost function can be chosen with many different ways and thus we get different kinds of optimal estimators. It is usually assumed, as we shall, that the cost function is non-negative and that it has the value zero in the origin. Because

$$\text{E}(L(\mathbf{x}_k - \hat{\mathbf{x}}_k)) = \text{E}(\text{E}(L(\mathbf{x}_k - \hat{\mathbf{x}}_k)|\mathbf{y}_{1:k})), \tag{3.18}$$

it suffices to find an estimator that minimizes the conditional expectation

$$\hat{\mathbf{x}}_k^{opt} = \text{argmin}_{\hat{\mathbf{x}}_k} \, \text{E}(L(\mathbf{x}_k - \hat{\mathbf{x}}_k)) = \text{argmin}_{\hat{\mathbf{x}}_k} \, \text{E}(L(\mathbf{x}_k - \hat{\mathbf{x}}_k)|\mathbf{y}_{1:k} = y_{1:k}), \tag{3.19}$$

for all values of the measurements $y_{1:k}$. A common estimator is the *least squares estimator*, that we get by minimizing the expectation value of the square of the error norm, that is,

$$L(\mathbf{x}_k - \hat{\mathbf{x}}_k) = \|\mathbf{x}_k - \hat{\mathbf{x}}_k\|^2. \tag{3.20}$$

It can be shown that the estimator corresponding to the cost function (3.20) is the same as the expectation of the posterior

$$\hat{\mathbf{x}}_k^{\text{MSE}} = \text{E}(\mathbf{x}_k|\mathbf{y}_{1:k} = y_{1:k}),$$

(see Exercise 3.7). In fact, this same cost function is used for deriving the Kalman filter (Section 3.2). It is however good to remember that the Kalman filter minimizes the expectation (3.18) among linear estimators only, whereas the optimal estimators handled in this section minimize the conditional expectation (3.19) from the set of all estimators, not just linear ones.

Another common estimator is the *maximum a posteriori*, i.e. MAP estimator, then accordingly the estimator is

$$\hat{\mathbf{x}}_k^{\text{MAP}} = \text{argmax}_{x_k} \, p(x_k|y_{1:k}).$$

The MAPestimator minimizes the so-called *hit or miss* cost function $L = \lim_{\delta \downarrow 0} L_\delta$, where

$$L_\delta(\mathbf{x}_k - \hat{\mathbf{x}}_k) = \begin{cases} 1, & \|\mathbf{x}_k - \hat{\mathbf{x}}_k\| \geq \delta \\ 0, & \|\mathbf{x}_k - \hat{\mathbf{x}}_k\| < \delta \end{cases},$$

3.5 Numerical methods for filtering

This section presents some numerical methods for solving filtering problems. A hallmark for a good numerical approximation method is that there is a way to control the trade-off between accuracy and computational load. for example parameter N that can be increased in order to make the solution approach the correct one. In numerical Bayesian filtering, this parameter often is the number of the pairs $\{p_k^i(x), \omega_k^i\}$ ($i = 1, \dots, N$), with which we approximate the density function of the posterior distribution

$$p(x_k|y_{1:k}) \approx \sum_{i=1}^{N} \omega_k^i p_k^i(x_k).$$

For instance, for the particle filter (Section 3.5.2) and for the point mass filter

$$p_k^i(x_k) = \delta(x_k - x_k^i),$$

where δ is the delta function (Dirac's delta measure). For the grid filter, the function $p_k^i(x_k)$ is the characteristic function of set A_i and for the Gaussian mixture filter $p_k^i(x_k)$ is the density function of a normal distribution.

The biggest difference between the particle filter and the point mass filter is that in the particle filter particles $\{x_k^i\}$ are chosen randomly whereas in the point mass filter they are deterministically set. The particle filter is based on Monte Carlo integration, described next.

3.5.1 Monte Carlo integration

Monte Carlo integration is based on the law of large numbers.

Theorem 7 (Strong law of large numbers)**.** *Let* $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$ *be independent identically distributed random variables. Then*[*]

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i = \mu\right) = 1$$

if and only if the distribution has an expectation value and $\mathrm{E}(\mathbf{z}) = \mu$.

Write the integral to be computed in the form

$$I = \int g(z)dz = \int \frac{g(z)}{f(z)} f(z)dz \triangleq \int h(z)f(z)dz, \tag{3.21}$$

where $f(z) > 0$ is the density function of a random variable $\mathbf{z}$. The random variable $\mathbf{z}$ can be chosen almost freely, so it is good to pick one from which it is convenient to draw random samples. Then according to the Strong law of large numbers (above), the sample mean

$$\bar{h}_n = \frac{1}{n} \sum_{i=1}^{n} h(\mathbf{z}_i) \tag{3.22}$$

---

[*]Then it is said that the sample mean *converges almost surely*. From almost certain convergence follows *convergence in the probability sense*, from which furthermore follows *convergence in the distribution sense*.

converges almost surely towards integral I (3.21). If $\mathrm{E}(h(z)^2)$ exists, then the variance of random variable $\bar{h}_n$ can be approximated

$$\mathrm{V}(\bar{h}_n) = \frac{1}{n} \int [h(z) - \mathrm{E}(h(z))]^2 f(z) dz \approx \frac{1}{n^2} \sum_{i=1}^{n} \left(h(\mathbf{z}_i) - \bar{h}_n \right)^2 \triangleq \sigma_{\bar{h}_n}^2. \qquad (3.23)$$

Additionally, it can be shown that the quotient (compare to the central limit theorem)

$$\frac{\bar{h}_n - \mathrm{E}(h(z))}{\sqrt{\sigma_{\bar{h}_n}^2}}$$

approaches in a distribution sense the distribution of the random variable $\mathrm{N}(0,1)$.

Example 21 (Monte Carlo integration). *Let a random variable be* $\mathbf{z} \sim \mathrm{N}(0,1)$ *and*

$$g(z) = \begin{cases} \frac{3}{4}(1 - z^2), & |z| \le 1 \\ 0, & |z| > 1 \end{cases}$$

When the number of the simulated random variables is $n = 10^4$, *one realization of Monte Carlo approximation of integral* $\int g(z) dz = 1$ *is about* 1.005, *equation (3.22) and the approximation of the variance of this approximation is about* $3.045 \cdot 10^{-5}$ *(3.23). The density function of the normal distribution corresponding to these values is visualized on the left in Figure 3.2. In the background of the density function there has been drawn a histogram of a thousand such Monte Carlo simulations. The histogram resembles a normal distribution whose expectation is one, just like in theory. In Figure 3.2 on the right is a visualization of the error of the Monte Carlo integration compared to the number n of random draws used. The line in the plot reflects the theoretic convergence rate* $O(n^{-\frac{1}{2}})$.
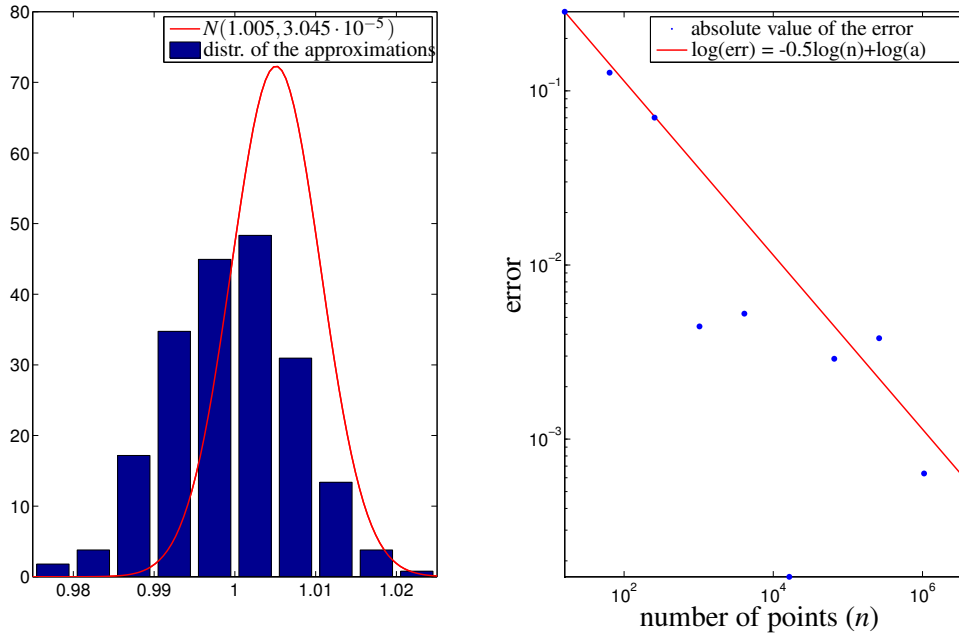
Figure 3.2: Results of the Monte Carlo simulation.

3.5.2 Particle filter

In this section, we describe different variants of particle filters [34]. A particle filter is based on the Monte Carlo method and it approximates prior and posterior distributions using samples, called particle clouds, of the distributions in question. One particle filter is given as the Algorithm 5. In fact, the algorithm in question can be considered as an algorithm for a family of particle filters, because by choosing different model parameters we get different kinds of particle filters. These choices are discussed in the next section. Additionally, it is good to remember that this is not the most general formulation for particle filter but there exists even more general or in some way differing particle filters.

Algorithm 5 Particle filter

- The state model: $\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1},\ p_{\mathbf{w}_{k-1}}(w)$
- The measurement model: $\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k,\ p_{\mathbf{v}_k}(v)$
- The initial state and the measurements: $p_{\mathbf{x}_0}(x)$ and $y_{1:m} = \{y_1, y_2, \ldots, y_m\}$
- The number of particles and the proposal distributions: N and $g(x_k|x_{k-1}^i, y_k)$

In this algorithm $i = 1, \ldots, N$.

1. - Simulate samples x_0^i from distribution $p_{\mathbf{x}_0}(x)$
   - Set $\omega_0^i = \frac{1}{N}$ and $k = 1$

2. - Simulate particles x_k^i from distributions $g(x_k|x_{k-1}^i, y_k)$
   - Set $\omega_k^i = \omega_{k-1}^i \dfrac{p_{\mathbf{v}_k}(y_k - h_k(x_k^i)) p_{\mathbf{w}_{k-1}}(x_k^i - f_{k-1}(x_{k-1}^i))}{g(x_k^i|x_{k-1}^i, y_k)}$

3. Normalization: $\omega_k^i = \dfrac{\omega_k^i}{\sum_{i=1}^N \omega_k^i}$

4. Compute point estimates, e.g.:
   - The posterior expectation: $\hat{x}_k \approx \sum_{i=1}^N \omega_k^i x_k^i$
   - The covariance matrix of the posterior: $P_k \approx \sum_{i=1}^N \omega_k^i \left(x_k^i - \hat{x}_k\right)\left(x_k^i - \hat{x}_k\right)^T$

5. Resampl when needed:
   - Simulate particles x_k^i from distribution $\sum_i^k \omega_k^i \delta(x_k - x_k^i)$
   - Set $\omega_k^i = \frac{1}{N}$

6. Stop, if $k = m$, otherwise set $k = k+1$ and go back to Step 2.

Sequential importance sampling (SIS) is a simple particle filter and it is obtained from Algorithm 5 when the resampling (Step 5) is left out. However, this eventually causes all the weight to accumulate in just a few particles. Then all the other weights are practically zero, and a lot of computation capacity has to be spent without having any effect on the approximation of the posterior distribution. Because of this, in the SIR filter (sampling importance

resampling) the resampling is done at every step. Additionally, prior distributions $p(x_k|x_{k-1}^i)$ are used as proposal distributions $g(x_k|x_{k-1}^i, y_k)$, so that the formula of the weights in step 2 of the algorithm simplifies into $\omega_k^i = p_{\mathbf{v}_k}(y_k - h_k(x_k^i))$.

In the SIS filter, resampling is not done at all, whereas in the SIR filter it is done at every time instant. Generally speaking, neither of these is the optimal way to operate. Different heuristics have been developed to decide whether it is worth resampling or not. One of these ways is to compute an approximation for the number of effective samples

$$N_{\text{eff}} \approx \frac{1}{\sum_{i=1}^{N}(\omega_k^i)^2},$$

and resample if it is smaller than a certain threshold value.

The resampling can be done in many ways. Algorithm 6 introduces *systematic resampling*. A computationally heavier resampling algorithm (so-called multinomial resampling) is obtained when in the systematic resampling the algorithm comparison points $z_i \sim \text{Uniform}(0,1]$ are simulated each time. Systematic resampling takes one sample from each interval $\left(\frac{i-1}{N}, \frac{i}{N}\right]$. This guarantees that if the weight of a particle is $\omega_k^j \geq \frac{l}{N}$, where l is a natural number, then the corresponding particle appears at least l times in the resampled particle set. This also holds for the so-called stratified resampling where one comparison point is simulated from each interval $\left(\frac{i-1}{N}, \frac{i}{N}\right]$ and otherwise it functions the same as Algorithm 6.

Algorithm 6 Systematic resampling

- Particles and weights $\{x_k^i, \omega_k^i\}$, where $i = 1, \ldots, N$.

1. Simulate the starting point: $z_1 \sim \text{Uniform}(0, \frac{1}{N}]$ and set $i = 1$.

2. Compute current comparison point $z_i = z_1 + (i-1)\frac{1}{N}$

3. Set $\omega_k^i = \frac{1}{N}$ and $x_k^i = x_k^j$, where j is set in such a way that $\sum_{l=1}^{j-1} \omega_k^l < z_i \leq \sum_{l=1}^{j} \omega_k^l$.

4. Stop, if $i = N$, otherwise set $i = i+1$ and go back to Step 2

The proposal distribution $g(x_k|x_{k-1}^i, y_k)$ was mentioned earlier in connection with the SIR filter, which uses prior distributions $p(x_k|x_{k-1}^i)$ as proposal distributions. Prior distribution is a popular choice for the proposal distribution, although they are not the optimal choice when an effective sample size is considered. The optimal choice would be to use posterior distributions $p(x_k|x_{k-1}^i, y_k)$ as proposal distributions. However, usually they are not known[*], and then we often end up using either the prior distributions or the approximations of the posterior distributions which can be computed for example with the nonlinear Kalman filter (Section 3.3).

Example 22 (PF and EKF). *Figure 3.3 shows a positioning example using particle filter (PF) and extended Kalman filter (EKF). The system state is four-dimensional, two location coordinates and two velocity coordinates, and state model parameters Φ and Q are the same as in*

---

[*]Because they are the ones we are seeking.

the Kalman filter example (3.14). A measurement is taken at one second intervals as a range measurement from two base stations, which have also been drawn in the figure. Then the measurement model is

$$\mathbf{y}_k = \begin{bmatrix} \|\mathbf{x}_{bs_1} - \mathbf{x}_k\| \\ \|\mathbf{x}_{bs_2} - \mathbf{x}_k\| \end{bmatrix} + \mathbf{v}_k,$$

where $\mathbf{x}_{bs_1} = \begin{bmatrix} -250 \\ 0 \end{bmatrix}$, $\mathbf{x}_{bs_2} = \begin{bmatrix} 250 \\ 0 \end{bmatrix}$ and $\mathbf{v}_k \sim N(0, 10^2 I)$. The initial state of the system is

$$\mathbf{x}_0 \sim N\left(\begin{bmatrix} -150 \\ -30 \\ 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 20^2 & 0 & 0 & 0 \\ 0 & 20^2 & 0 & 0 \\ 0 & 0 & 5^2 & 0 \\ 0 & 0 & 0 & 5^2 \end{bmatrix} \right).$$

The location estimates given by the filters have been drawn in the figure, which are in this case approximations of the expectations of the posterior distributions. At time instants $\{0, 10, 20, 30, 40\}$ the ellipses corresponding to the covariance matrices given by the EKF have been drawn. The ellipses are defined in such a way that if the posterior distributions were normal with parameters given by the EKF, then the ellipse would represent the level curve of the density function of the posterior distribution such that with about 68% probability the state is inside the ellipse.

The particle filter used is the SIR filter (p. 54) with a million particles. The particle clouds have been visualized at time instants $\{0, 10, 20, 30, 40\}$ with a hundred particles, which have been got from the original set of one million particles by using systematic resampling (algorithm 6). Because the number of particles is relatively large, it can be thought that the particle filter represents some kind of reference filter in the example.

From the figure we see that at the last time instant ($t = 40$), the posterior distribution clearly has two separate peaks. The EKF has started following one of these peaks, which is very characteristic for the EKF. In this case the EKF has "chosen" the wrong peak and then we say that the EKF has gone astray.

Figure 3.3: *Figure of the situation of the example, more detailed explanation is found in the text.*

Although the EKF gives completely false results, the covariance matrix shows that the EKF trusts very much the estimate in question. After this, the EKF does not easily "find" the correct location even when measurements from additional sources become available and the bimodal (two-peaked) posterior becomes unimodal. As a rule of thumb, we can say that it is worth applying the EKF to the system if the posterior distributions resemble normal distribution i.e. they are unimodal and the tails of the distribution rapidly decay towards zero, but in this kind of two-peaked case applying EKF starts to be quite risky. This bimodal situation would completely

change if there were range measurements from one more base station or if the route did not go near the line connecting the base stations.

Exercises

3.1. Show that the mean square error matrix of prediction $\hat{\mathbf{y}}_k = H_k\hat{\mathbf{x}}_k^-$ of measurement $\mathbf{y}_k = H_k\mathbf{x}_k + \mathbf{v}_k$ is $H_kP_k^-H_k^T + R_k$.

3.2. Assume that matrix R is symmetric positive definite and matrix P is symmetric positive semi-definite. Show that

$$(I - KH)P(I - KH)^T + KRK^T = P - AA^T + (KB - A)(KB - A)^T,$$

where $B = (HPH^T + R)^{\frac{1}{2}}$ and $A = PH^TB^{-1}$.

3.3. Let A be an arbitrary real matrix. Show that the zero matrix is the unique solution for the problem $\text{argmin}_A \text{tr}(AA^T)$.

3.4. For a partitioned matrix it holds that (check it)

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix},$$

if the required inverse matrices exist. On the other hand, we know that the inverse matrix of a symmetric matrix is symmetric. Applying this information to matrix

$$\begin{bmatrix} P^{-1} & H^T \\ H & -R \end{bmatrix}$$

may be of help in the next task. Additionally, let A and B be symmetric positive definite matrices. Now, if $A \leq B$ then $B^{-1} \leq A^{-1}$ [14, Corollary 3.3.4.].

(a) Show that $(P^{-1} + H^TR^{-1}H)^{-1} = P - PH^T(HPH^T + R)^{-1}HP$, when $P > 0$ and $R > 0$.

(b) Show that in linear Gaussian case posterior covariance matrix is "increasing". That is, show that if $0 < P_1 \leq P_2$ then $0 < P_1^+ \leq P_2^+$, where $P_i^+ = (I - KH)P_i$ and $K = PH^T(HPH^T + R)^{-1}$.

3.5. (a) Let $\mathbf{x}_{k-1} \sim N(\hat{x}_{k-1}, P_{k-1})$, $\mathbf{w}_{k-1} \sim N(0, Q_{k-1})$ and $\mathbf{v}_k \sim N(0, R_k)$ be independent random variables. What is the distribution of the random variable

$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} \Phi_{k-1} & I & 0 \\ H_k\Phi_{k-1} & H_k & I \end{bmatrix} \begin{bmatrix} \mathbf{x}_{k-1} \\ \mathbf{w}_{k-1} \\ \mathbf{v}_k \end{bmatrix} ?$$

(b) Let

$$
\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathrm{N} \left(\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right),
$$

show that $\mathbf{x}|\mathbf{y} = y \sim \mathrm{N}(\bar{x} + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \bar{y}), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})$.

(c) Apply the result of part (b) to the distribution from part (a), and compare the obtained result with the BLU-estimator introduced in Section 3.2.

3.6. Show that in one-dimensional case, the unscented transformation using the points given in formula (3.17) make the approximation (3.16) exact for a third degree polynomial. Naturally, here $p(x)$ is the density function of distribution $\mathrm{N}(\mu, \sigma^2)$. On what grounds is the choice $\kappa = 2$ reasonable? (Hint: $\int x^4 p(x)dx = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$)

3.7. Show that the posterior expectation minimizes the expectation $\mathrm{E}(\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|^2 \mid \|\mathbf{y}_{1:k} = y_{1:k})$.

Computer exercises

3.8. Implement the Kalman filter (KF), whose specification is

```
%   [X,P] = kalman(x0,P0,F,Q,Y,H,R)
%
%   x_{k+1} = Fx_{k}+w
%     y_{k} = Hx_{k}+v
%
% IN: x0 = initial state at time 0 (R^n)
%     P0 = initial covariance matrix
%     F  = state transformation matrix
%     Q  = covariance matrix of process noise (w)
%     Y  = measurements Y=[y_1,...,y_k](R^(m \times k),
%     H  = ''measurement model''
%     R  = covariance matrix of measurements errors (v)
%
%OUT:   X  = [x_0,x_1,...,x_k], where x_i is estimate at time i.
% P(:,:,i)  = MSE matrix at time i

function [X,P] = kalman(x0,P0,F,Q,Y,H,R)
```

There are data in web page http://www.students.tut.fi/~aliloytt/menetelmat/harkat.html that you can test your filter. Visualize these tests.

3.9. Compute by using Monte Carlo integration the expression

$$
I = \frac{\int_A \frac{x}{2\pi} \exp\left(-\frac{1}{2}x^T x\right) dx}{\int_A \frac{1}{2\pi} \exp\left(-\frac{1}{2}x^T x\right) dx},
$$

where $A = [0,2] \times [0,2]$. Additionally, study the convergence of Monte Carlo simulations towards the exact value

$$I = \frac{1 - \exp(-2)}{\sqrt{2\pi}(\Phi(0) - \Phi(-2))} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

where Φ is the cumulative function of distribution $N(0,1)$. The integral in question represents the expectation of a posterior distribution whose prior distribution is $N(0,I)$ and the measurement is the information that the state is inside the area A.

3.10. Implement two filters of this group {EKF, UKF, PF}, whose specifications is

```
%   [X,P,...] = filter(x0,P0,F,Q,Y,S,R)
%
%   x_{k+1} = Fx_{k}+w
%     y_{k} = h(x_{k})+v,
% where (h(x_{k}))_i is range measurement from base station s_i
%
% IN: x0 = initial state at time 0 (R^n)
%     P0 = initial covariance matrix
%     F  = state transformation matrix
%     Q  = covariance matrix of process noise (w)
%     Y  = measurements Y=[y_1,...,y_k](R^(m \times k),
%     S  = coordinates of base stations  [s_1,...,s_l]
%     R  = covariance matrix of measurements errors (v)
%
%OUT:   X  = [x_0,x_1,...,x_k], where x_i is estimate at time i.
% P(:,:,i)  = MSE matrix at time i

function [X,P,...] = filter(x0,P0,F,Q,Y,S,R)
```

There are data in web page http://www.students.tut.fi/~aliloytt/menetelmat/harkat.html that you can test your filter. Visualize these tests.

Chapter 4

Positioning techniques

JUSSI COLLIN

4.1 Inertial navigation

In principle, inertial navigation algorithms are based on the coordinate transformations studied in Chapter 1. Before going through the basic formulas, we have to understand what we are measuring. The measurement of an accelerometer triad, $\mathbf{a}^B - \mathbf{g}^B$, was already mentioned. This is perhaps the most difficult measurement to comprehend, because vector $\mathbf{a}$ is the acceleration in inertial coordinate frame. The measurement device however outputs this vector in B-frame. Another problem is that the use of Newton's laws of motion also requires measurement of gravitational forces, and that cannot be done directly with the accelerometers! For instance, stationary accelerometer triad at the North Pole is not in accelerating motion with respect to the ECI-coordinates, but the accelerometer output is still $-\mathbf{g}$ (i.e., upward acceleration). The Newtonian approach in this situation is to draw the gravitation force vector downwards and the normal force upwards. These forces cancel each other, and then the acceleration is $\mathbf{0}$. From this it follows that if location changes are computed in ECI with the help of accelerometers, the local gravity vector needs to be separately computed and added to the measurement. Note that $\mathbf{g}$ is a function of location, which causes some feedback connections (see Exercise 4.7).

As mentioned previously, the acceleration $\mathbf{a}^B$ is given in the B-coordinate system. It is not reasonable to integrate acceleration measurements taken in this coordinate system. Instead, we first have to move to I- or E-coordinate system. In practice, we need to compute and update the rotation matrices C_B^I, C_B^E and C_B^L during navigation, and for this we need the gyro triad measurements. As said, a gyro triad* outputs vector $\mathbf{w}_{IB}^B$. Next we briefly go through the basic equations and measurement errors of inertial sensors.

---

*The simplest gyro and accelerometer assembly consists of three sensors whose measurement axes are perpendicular to each other.

4.1.1 Accelerometers and angular velocity sensors

The term *inertia* refers to a property of matter by which it remains in uniform motion unless acted upon by some external force. Accelerometers and some gyros are indeed based on this principle. To build an accelerometer we need a mass, a spring, a case, and something to indicate the location of the mass with respect to the case (Figure 4.1).* The relation between force and acceleration is known, and the compression or stretching of the string shows the magnitude of this force. Yet, in Figure 4.1 the visible spring compression can be either the combined effect of the normal force and gravitation while the sensor is stationary, or it may be caused by acceleration in the inertial frame (without gravitation). The sensor does not know which case it is.
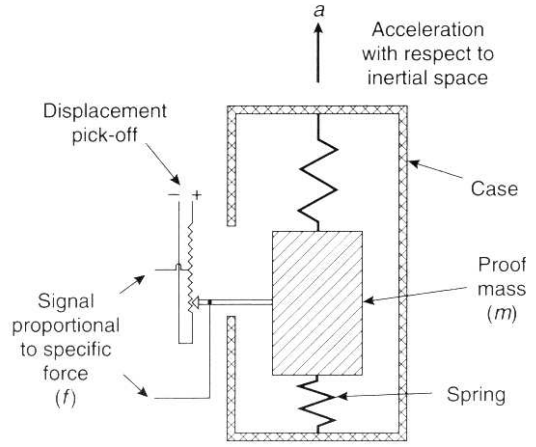
Figure 4.1: Operational principle of an accelerometer. Source: [40]
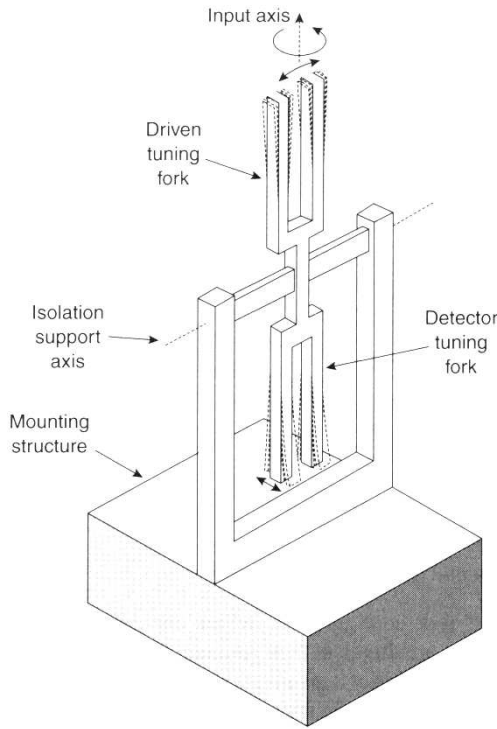
Figure 4.2: Example of a gyro based on the Coriolis force. Source: [40]

There are several different choices to measure angular velocity. Mechanical gyros are also based on inertia, namely the moment of inertia of a rotating body. One way to measure angular velocity is to install a rotating mass (M) into gimballed case (B), so that $\mathbf{w}_{BM}^B = [0\ 0\ \omega_M]^T$. The rotating mass is supported in such a way that it and its bearings are able to rotate in direction $\mathbf{u}_{out}^B = [0\ \pm 1\ 0]^T$, but not in direction $\mathbf{u}_{in} = [\pm 1\ 0\ 0]^T$. Then the motion $\mathbf{w}_{IB}^B = [\omega_{in}\ 0\ 0]^T$ causes this mass to rotate in direction $\mathbf{w}_{BM}^B \times \mathbf{w}_{IB}^B$, i.e. the output axis $\mathbf{u}_{out}$. When we measure rotation with respect to this axis, we get information about the angular velocity with respect to $\mathbf{u}_{in}$ axis.

A rotating mass is not always practical if there are design limitations with respect to size or power consumption. Another type of mechanical gyro is a vibrating gyro, where motion occurs back and forth. Figure 4.2 presents the operational principle of a tuning fork gyro. The upper fork is electrically driven to resonate. If the sensor rotates around the input axis, then the Coriolis force causes an oscillating motion perpendicular to the forced resonance direction and to the input axis direction (the arrows in the lower fork). This motion is

---

*Additionally, we can have a feedback element that tries to keep the mass stationary with respect to the case, resulting in a closed loop accelerometer.

measured (usually capacitively), and the result is an angular velocity modulated signal, which can be used to derive the angular velocity. Micro-mechanical (MEMS) gyros use this principle.

Moving parts are not necessarily needed for measuring rotation, because mechanical rotation has an electro-magnetic counterpart, the Sagnac phenomenon. To explain the phenomenon, relativity theory is needed, which is beyond the scope of this course. The operation principle of a Ring Laser Gyro (Figure 4.3) is as follows. We send a laser beam, and with the help of mirrors we direct the beam back to its initial location, following a clockwise path. We add another laser beam which follows a counter-clockwise path. Two standing waves are formed, whose frequency difference is measured. When the case rotates with respect to I-coordinates (around the normal of the plane defined by the laser route), this frequency difference changes[*]. The use of optic measurements has several benefits:

- The gyro is extremely accurate.

- The input bandwidth is unlimited.

- There are no moving parts, jitter is not a problem and reliability is better than with mechanical gyros.

- Linear acceleration does not affect the measurement.

On the other hand, its price, size and power consumption prevent its use in personal applications.
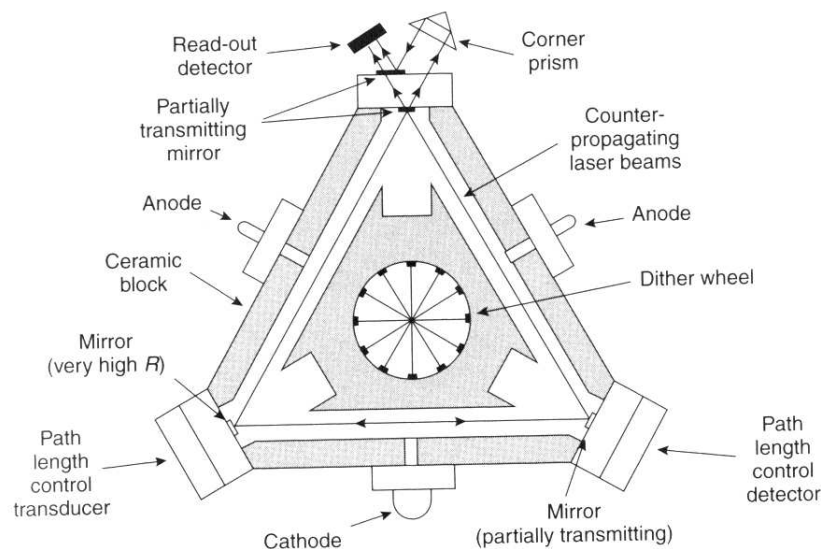
Figure 4.3: A laser gyro. Source: [40]

---

[*]The beam going to the rotation direction takes a longer trip viewed in the I-coordinates ...

4.1.2 INS equations

This section follows the notation of [35]. The origin of the location vector is the origin of the E-frame, and $\mathbf{g}^L$ is computed separately as a function of location (e.g. from a gravitation model). The time derivative of the B-coordinates and L-frame was already presented (1.25):

$$\dot{C}_B^L = C_L^B(\mathbf{w}_{IB}^B\times) - (\mathbf{w}_{IL}^L\times)C_B^L. \tag{4.1}$$

Initial condition C_B^L is obtained with a separate initialization algorithm (Exercise 4.8). The angular velocity term $\mathbf{w}_{IB}^B$ is obtained from a gyrotriad, and the term $\mathbf{w}_{IL}^L$ is the sum of the rotation of Earth and the rotation of local coordinates with respect to Earth due to ground speed:

$$\mathbf{w}_{IL}^L = \mathbf{w}_{IE}^L + \mathbf{w}_{EL}^L \tag{4.2}$$

Local frame (L) keeps z-axis perpendicular to a sphere (or ellipsoid) surface, and while an object is moving on the surface of the Earth this coordinate system rotates with angular velocity:

$$\mathbf{w}_{EL}^L = F_c(\mathbf{u}_{ZL}^L \times \mathbf{v}^L) + \rho_{ZL}\mathbf{u}_{ZL}^L \tag{4.3}$$

The 3×3 matrix F_c depends on whether an ellipsoid or a spherical model is used. Additionally, the term ρ_{ZL} controls L's north-axis motion (Exercise 4.1). The unit vector $\mathbf{u}_{ZL}^L$ points to the "up" direction in the L-frame. With our conventions this means $\mathbf{u}_{ZL}^L = [0\ 0\ 1]^T$.

The transformation of the measured acceleration vector (denote $\mathbf{a}_{SF}^B = \mathbf{a}^B - \mathbf{g}^B$) to L-frame is

$$\mathbf{a}_{SF}^L = C_B^L\mathbf{a}_{SF}^B. \tag{4.4}$$

Finally, we need the gravitational force ($\mathbf{g}_P$=gravitation plus the rotation effect of Earth)

$$\mathbf{g}_P^L = \mathbf{g}^L - (\mathbf{w}_{IE}^L\times)(\mathbf{w}_{IE}^L\times)\mathbf{R}^L, \tag{4.5}$$

the rate of change of the velocity vector (with respect to Earth, E-frame)

$$\dot{\mathbf{v}}^L = \mathbf{a}_{SF}^L + \mathbf{g}_P^L - (\mathbf{w}_{EL}^L + 2\mathbf{w}_{IE}^L) \times \mathbf{v}^L, \tag{4.6}$$

and the rate of change of the horizontal location

$$\dot{C}_L^E = C_L^E(\mathbf{w}_{EL}^L\times) \tag{4.7}$$

and the rate of change of the height

$$\dot{h} = \mathbf{u}_{ZL}^L \bullet \mathbf{v}^L. \tag{4.8}$$

That's all. With a suitable initial condition, position and velocity can be computed from the ordinary differential equations (4.6)–(4.8). For clarity, group the most important terms.

- The information from the sensors: $\mathbf{a}_{SF}^B$ and $\mathbf{w}_{IB}^B$

- The terms related to Earth properties: $\mathbf{w}_{IE}$, $\mathbf{g}$, F_c

- The terms that we originally wanted to compute: $\mathbf{R}^E$ (location vector in ECEF, C_L^E and h, height from reference geoid, provides the same information, Exercise 4.2), $\mathbf{v}^L$ (velocity with respect to Earth), and C_B^L (INS device attitude with respect to L-coordinates.)

4.1.3 Measurement errors

Gyro errors usually define the quality of the whole INS system. This results from the nature of INS mechanization, and on the other hand from the fact that angular velocity measurement is more difficult than acceleration measurement. Accordingly, with the notations of Chapter 2, let the measured gyro output be $\mathbf{y}$. Because of the measurement errors this is not in practice $\mathbf{w}_{IB}^B$ but for instance in accordance with model

$$\mathbf{y} = \mathbf{M}\mathbf{w}_{IB}^B + \mathbf{b} + \mathbf{n}. \tag{4.9}$$

Here matrix M contains scaling and alignment errors (scale factor, misalignment). The noise terms are divided into bias-like ($\mathbf{b}$) and uncorrelated noise ($\mathbf{n}$). Notice that the error term is now

$$\mathbf{v} = (\mathbf{M} - \mathbf{I})\mathbf{w}_{IB}^B + \mathbf{b} + \mathbf{n},$$

i.e. the scaling and alignment errors cause the error to be dependent on the signal itself. This causes problems in the statistical analysis of the measurement errors.

A basic assembly contains three sensors whose measurement axes are perpendicular to each other. Then the diagonal elements of the 3×3 matrix M are the scaling errors of the corresponding sensors. An alignment error results from the fact that in real system the sensors are not exactly perpendicular to each other. Thus, data from one axis leaks to the other sensors. In high-quality INS devices, the alignment error is in order of one thousandth of a degree, and the scale factor errors are a few parts per million (ppm) of the signal magnitude. Notice that it is not necessary to align the sensors perpendicular to each other as long as we know the angles between the axes. There can be more sensors than the required three to ensure continued functionality in case of a defect. In that case, a perpendicular installation is not optimal.

Correlated noise $\mathbf{b}$ is in practice the most challenging to deal with. The term refers to the errors that stay the same or almost the same for a long time. The separation of noise into $\mathbf{b}$ and $\mathbf{n}$ is not at all obvious. The limiting situation, where $\mathbf{b}$ would be always constant and $\mathbf{n}$ would be completely uncorrelated between samples, is not found in the real world. If it was, then the bias term could be completely solved already in a factory calibration and remaining errors would all be white. In practice, the correlation time of $\mathbf{b}$ is of the same order as the time between the device startups (from hours to months), and the correlation time of $\mathbf{n}$ is significantly smaller. Characterization of these error terms can be done using Allan Variance plots [2].

Previous chapters already contain a lot of information about the handling of the error term $\mathbf{n}$. You should notice that INS mechanization includes a lot of integrations, so the growth of the bias-like errors with respect to time has a considerable effect on the positioning accuracy. If a short sample from data looks noisy, one should not make too radical conclusions. Trying to compare gyros' quality using the standard deviation estimates computed from samples taken over just a few minutes can be very misleading.

Similarly to the gyro measurements, the model (4.9) works also for acceleration sensors, as long as term $\mathbf{w}_{IB}$ is replaced by term $\mathbf{a}_{SF}^B$. Table (4.1.3) lists the requirements for the sensors if the INS solution error is desired to be within a few kilometers after one hour navigation. The requirements are stringent, especially for the gyros. Depending on the intended use, there

might also be requirements for the dynamic range, for example that a rotation of 500 degrees per second has to be measured. An INS device of this level costs about 100 000–200 000 US dollars, but the prices are steadily falling. In addition to price, the export limitations cause problems, because accurate INS devices are classified as weapons technology in several countries. MEMS sensors do not satisfy these requirements yet, they are not even close, but they are easily available and naturally less expensive. And although INS by itself would not suffice for positioning because of measurement errors, inertial measurements give valuable additional information to positioning algorithms [7].

Table 4.1: Accuracy requirements of sensors, positioning errors 0,1 or 1 knots [36]

	0.1 nmph	1 nmph
Acceleration sensor, bias	$5\,\mu g$	$40\,\mu g$
Acceleration sensor, scaling error	$40\,\mathrm{ppm}$	$200\,\mathrm{ppm}$
Acceleration sensor, alignment error	$1/3600^\circ$	$7/3600^\circ$
Gyro, bias	$0.0007^\circ/\mathrm{h}$	$0.007^\circ/\mathrm{h}$
Gyro, scaling error	$1\,\mathrm{ppm}$	$5\,\mathrm{ppm}$
Gyro, alignment error	$0.7/3600^\circ$	$3/3600^\circ$

4.2 Satellite positioning based on carrier measurements

The basic measurements of a GPS receiver are often considered to be the pseudorange measurement and the delta-pseudorange measurement. These are actually formed using (replica) code phase- and (replica) carrier phase measurement. The replica code phase can be converted to transmission time from which the pseudorange can then be computed. The replica carrier phase can be converted to a delta-pseudorange or to so-called integrated Doppler measurement.

Carrier measurements were not designed for utilization when GPS was implemented, instead GPS was designed assuming that only code measurements would be used. Use of carrier measurements in GPS-positioning was developed from research in which the relative locations of two astronomical antennas were measured.

With the help of carrier measurements, a **change** of location can be measured very accurately. If some reference point (where a range is measured) is accurately known, the *absolute accuracy* is also transfered into accuracy of the measured location (Figure 4.4). The most important application area of relative positioning is land surveying, i.e. geodesy. In personal positioning the carrier signal has been used for smoothing code measurements (*carrier smoothing*). Relative positioning is however also coming to personal positioning [1].

4.2.1 Doppler effect

Satellites are in continuous motion with respect to the user. GPS satellites orbit the Earth in elliptical, nearly circular orbits about two times per day. Thus their orbit velocity is about 4
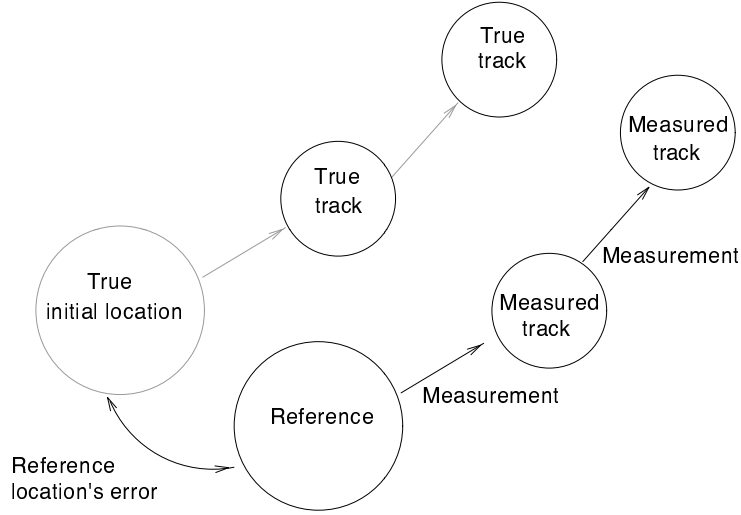
Figure 4.4: Relative positioning. The accuracy of the initial location translates into accuracy of the whole route if the route can be followed without errors.

km/s. Of course the user's own motion also affects the relative velocity between satellites and the user.

The relative motion between a satellite and a receiver receiving the satellite's transmitted signal affects the frequency of the received signal. This effect of relative motion between transmitter and receiver on received frequency is called the Doppler effect. Correspondingly, the difference between transmitted signal frequency and received signal frequency is called the Doppler shift. GPS receiver can measure the Doppler shift very accurately. By continuously following Doppler shift information, the motion between the satellite and the user is obtained, and when the satellite's velocity is known, the user's velocity can be found.

The Doppler effect is modeled as follows:

$$ f_R = f_T \left(1 - \frac{\mathbf{v}_r \cdot \mathbf{a}}{c} \right) \tag{4.10} $$

where f_R is the received frequency, f_T is the transmitted frequency, $\mathbf{v}_R$ is the relative velocity between the satellite and the user, $\mathbf{a}$ is a unit vector from receiver's location towards the satellite and c is the speed of light.

The Doppler shift is thus

$$ f_D = f_R - f_T. \tag{4.11} $$

By integrating the Doppler shift f_D over a measurement period (typically 1 s) the change of the carrier measurement is obtained. In other words, f_D is the rate of change of the carrier and thus integration over the measurement period gives the *carrier phase measurement*.

4.2.2 Connection between Doppler frequency and change of range measurement

Example 14 (page 24) presented a measurement model for pseudorange measurement. Using the same notations, we now express the pseudorange derivative i.e. *deltarange*, assuming the measurement error is zero:

$$\dot{\rho} = \frac{d}{dt}(\|\mathbf{s} - \mathbf{x}\| + b) = (\mathbf{v}_{\text{sat}} - \mathbf{v}_{\text{u}})^T \frac{\mathbf{s} - \mathbf{x}}{\|\mathbf{s} - \mathbf{x}\|} + \dot{b} \tag{4.12}$$

where $\mathbf{v}_{\text{sat}}$ is satellite's velocity vector, $\mathbf{v}_{\text{u}}$ is receiver's velocity vector and $\dot{b}$ the receiver's *clock drift* [s/s].

Now we see that $\mathbf{a} = \frac{\mathbf{s} - \mathbf{x}}{\|\mathbf{s} - \mathbf{x}\|}$ and that the relative velocity $\mathbf{v}_{\text{r}} = \mathbf{v}_{sat} - \mathbf{v}_u$, and therefore (4.12) becomes

$$\dot{\rho} = \mathbf{v}_{\text{r}} \cdot \mathbf{a} + \dot{b}. \tag{4.13}$$

When the receiver's clock drift effect is ignored and referring to Equations (4.10) and (4.11), we get the result

$$f_D = f_R - f_T = \frac{f_T}{c} \mathbf{v}_r \cdot \mathbf{a} = \frac{f_T}{c} \dot{\rho}. \tag{4.14}$$

4.2.3 Connection between Doppler frequency and phase measurement

The carrier measurement is also called "integrated Doppler". A new carrier measurement is obtained by adding to the old measurement the integral of the Doppler over the measurement epoch:

$$\phi_n = \phi_{n-1} + \int_{t_{n-1}}^{t_n} f_D(\tau) d\tau \tag{4.15}$$

where ϕ_n is the carrier measurement at epoch n and f_D is the Doppler frequency as a function of time.

Let's look at a simplified example. A general alarm signal is tested by transmitting a sinusoidal (with changing frequency) sound signal at time 12.00.00. Assume that the sender and the person who hears the signal have perfectly synchronized clocks. The receiver knows the transmission frequency and the transmission time. The receiver notices that the sound frequency is at its highest at t_R seconds after noon. What can the receiver conclude about the range to the transmitter?

The sound wave travels at speed v, i.e. the frequency of the transmission time instant is received s/v seconds later, where s is range. The transmission frequency had been modulated by a sinusoid, i.e.

$$f_T(t) = A \sin(\omega t + \phi_0) + f_0.$$

The constant terms A, ω, f_0 ja ϕ_0 are known by the receiver. On the receiver side it holds that

$$f_R(t) = f_T(t - s/v) = A\sin(\omega(t - s/v) + \phi_0) + f_0,$$

i.e. phase-difference (R-T) is $-\omega s/v$. From this we can solve s but what remains is the so-called integer ambiguity k, because the same measurement would also be obtained at range $s_1 + \frac{2k\pi v}{\omega}$. These are multiples of the wavelength, $\frac{2\pi v}{\omega} = \lambda$, and the receiver has a multitude of range solutions that differ by multiples of the wavelength. The wavelength is thus significant when phase measurements are used for positioning. The wavelength of GPS L1-frequency is 19 cm, from which follows bad news and good news:

- There are several solutions

- If the correct k is somehow found, and the phase can be measured with an accuracy of few degrees, we have a position solution to within a few millimeters.

Unfortunately, the phase difference can not be directly studied, in the example case there was no noise corrupting the phase measurement. Because of noise, the GPS receiver needs to measure the phase indirectly with the help of Doppler measurements and equation (4.15).

4.2.4 Solving the integer ambiguity

At this point it should be clear that a carrier measurement describes motion occurred between two consecutive measurement epochs, and that a carrier measurement can be done very accurately. The number of whole waves is however unknown when initializing the positioning (for every signal this number is thus different). That is like initializing from a blind situation, where an accurately known wave measurement is added to "some number" in each measurement epoch, and relative positioning is accurate.

Resolving the numbers of whole waves in the initial case makes accurate and absolute positioning possible. *Carrier cycle integer ambiguity resolution* is an active research area of positioning. The most famous methods are the LAMBDA method developed by Teunissen [39] and Hatch's least squares method [11].

4.2.5 Differential fixes

Combining the code and the carrier measurements of two near-by receivers eliminates the range errors that are common to both receivers, such as satellite ephemeris errors and unmodelled ionospheric erros. *Single difference* is the difference taken over signals that are obtained from the same satellite with two different receivers, as shown in Figure 4.5. Double difference is the difference taken over the single differences associated to two separate satellites, as shown in Figure 4.6. When double differences can be used for code and carrier measurements, then most of the common errors can be removed [19].
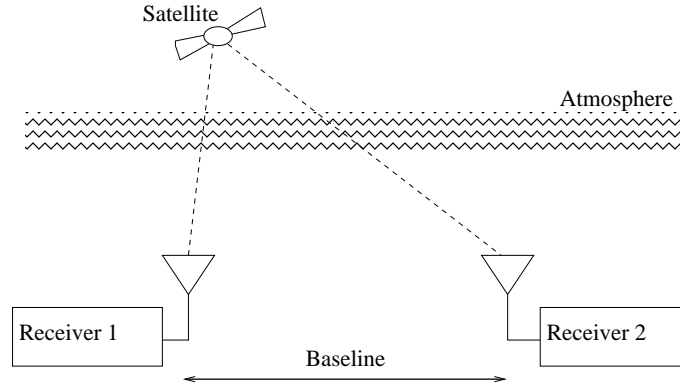
Figure 4.5: Single difference
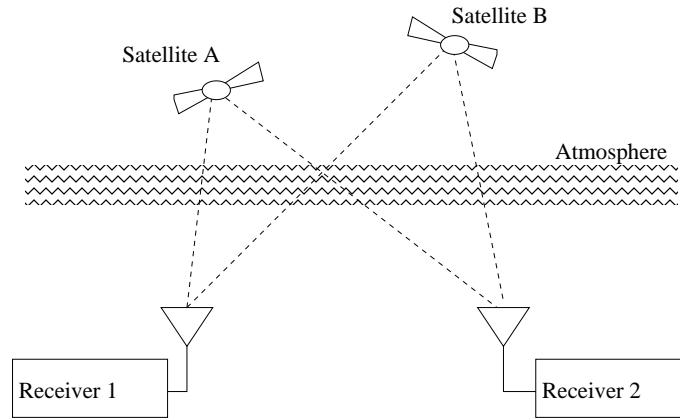
Figure 4.6: Double difference

Exercises

4.1. Consider formula

$$\mathbf{w}_{EL}^L = \mathrm{F}_c(\mathbf{u}_{ZL}^L \times \mathbf{v}^L) + \rho_{ZL}\mathbf{u}_{ZL}^L.$$

What kind of matrix F_c do we get, if we use a sphere model, with sphere radius R? What does the term ρ_{ZL} do? Hint: think about an airplane that flies very near the north pole. If the L-frame's y-axis is always kept pointing north, what happens to vector $\mathbf{w}_{EL}^L$?

4.2. Is C_L^E together with h sufficient information for unique location solution?

4.3. Three gyros are enough in INS applications, but to guard against fault situations there can be more. Assume that three gyros are perpendicular with each other, and add one gyro such that its measurement axis is not parallel with any of the other gyros. Can we now detect whether some sensor gives completely faulty measurements? If yes, can we identify the sensor in question and thus isolate false measurements? What if we have 5 sensors such that none of the three axes are in the same plane?

68

4.4. Show that the error in computation of $\mathbf{g}$ (gravitation model error) causes positive feedback connection to the height error of INS.

4.5. Assume that in future MEMS-gyro with analog output can satisfy the 1 nmph INS requirements, see Table (4.1.3). In order to digitize the signal, an A/D converter is needed. Estimate the required dynamic range, i.e. how many bits are needed for quantization.

4.6. How many integrators does the term $\mathbf{w}_{IB}^{B}$ (i.e. gyrodata) go through in INS equations? What happens to (nearly) uncorrelated noise after that many integrations? What about bias-type noise?

4.7. *Computer exercise:* Feedback connections in INS mechanization: see course web page.

4.8. *Computer exercise:* C_B^L initialization with the help of earth rotation vector and normal force (see course web page).

4.9. Derive Equation (4.12) or show that it is false, i.e.

$$\frac{d}{dt}(\|\mathbf{s} - \mathbf{x}\| + b) = (\dot{\mathbf{s}} - \dot{\mathbf{x}})^T \frac{\mathbf{s} - \mathbf{x}}{\|\mathbf{s} - \mathbf{x}\|} + \dot{b}$$

Chapter 5

Reliability

HELENA LEPPÄKOSKI

From a user perspective, positioning is not reliable if positioning cannot be done because of system or device error, or the result is clearly wrong although positioning seemingly succeeds. In this section we take a closer look at methods that may help a receiver detect large errors in the positioning solution and to remove the measurements causing the errors.

Reliability relates to several evaluation criteria describing the performance of positioning, and these are presented in Section 5.1. Then, principles of statistical inference and hypothesis testing are reviewed. Finally, the last section presents error detection methods and removal methods applied to a static location solution.

5.1 Performance measures of positioning

In the following we concentrate on satellite positioning, though some of the given indicators can be used on other positioning methods as well. This subject is covered e.g. in [38].

5.1.1 Accuracy

Accuracy of a positioning technology means how close the computed position is to the true position. Often accuracy is evaluated using the root mean square error (RMSE) of the positioning results. Another common way to describe error is to give its statistical distribution. This can be done either by using a probability density function or cumulative function. One simplification of these is CERP (Circular Error Probable), where a circle radius is defined such that the circle contains a certain percentage of the positioning results when the center of the circle is the true location of the object. For instance 95% CERP 56 m means that at least 95% of the positioning results are at most 56 meters away from the true location. CERP is used for describing

a horizontal 2-dimensional location error. Correspondingly, the magnitude of a 3-dimensional positioning error can be given using SERP (Spherical Error Probable), where instead of the circle radius, the sphere radius is given*.

Positioning accuracy depends not only on accuracy of measurement signals but also on how measurement errors affect the error of the location solution. The latter depends on the measurement geometry, and its goodness is evaluated with DOP (see Section 2.3.1).

5.1.2   Integrity

Integrity of a positioning system means the ability of the system to warn a user in real time in case the system should not be used for positioning. Integrity information is supposed to warn about a situation where the system seemingly works and produces a positioning signal, but the signal is for some reason so faulty that the computed location based on it is worthless.

GPS satellites transmit information to a user about their integrity in navigation messages, but that information always comes delayed. The Earth based control segment must first have realized the error in the satellite transmitted signals, and after that transmitted information about detected integrity problem to the satellites, which is then transmitted to users in navigation messages. The use of the positioning result during this delay can have fatal consequences e.g. in airplane or ambulance navigation.

Because of the delay in the offered integrity information of the GPS system, methods have been developed where the receiver itself checks the integrity of the signals. These methods are called RAIM (Receiver Autonomous Integrity Monitoring). The methods are based on consistency checks of the positioning signals. The methods can be used if there is redundancy in the measurements, i.e. there are enough signals in use to make the location estimation problem over-determined. Lack of consistency in the signals can be caused by either an error in the signal itself or an error in the measurement model.

Originally, the development of RAIM methods were driven by civil aviation security requirements. The aim has been to eliminate large position errors that arise for instance as a result of a satellite clock malfunction, and to eliminate errors in the measurement model that are caused by faulty information about a satellite's location (ephemeris error). Thus, RAIM aims to detect a possible error in the positioning system itself, where the error is caused by either the control segment or the space segment. These kinds of errors are very rare, their frequency is in the order of one error every 18–24 months. Because of the rareness of these errors, it is usually assumed that errors affect only one of the measurement equations at a time.

A positioning signal can be faulty also for other reasons than those caused by the control or space segment. Signal reflection and multipath can cause considerable errors compared to the normal noise in the positioning measurements. High sensitivity receivers (HSGPS, High Sensitivity GPS) can receive strongly attenuated signals, and the poor signal to noise ratio of the measurement causes large amplification of noise in the measurement itself. Additionally, it is possible to lock on to a wrong signal when the signal/noise ratio is poor. When an error-free

---

*In some sources CERP and SERP are abbreviated with CEP and SEP

satellite location is combined with a wrong measurement, the final result is as bad as in the large ephemeris-error case, when seen from the measurement equation perspective.

The errors mentioned above are typical in personal positioning. Because their effect appears the same way as the errors caused by control or space segment, the same methods as in traditional RAIM can be used to detect and remove them. Designing RAIM algorithm for handling errors of personal positioning has a few additional challenges: these errors are more frequent than the errors caused by control and space segment, and also the probability of errors occurring simultaneously in several different measurement equations is large.

In addition to a receiver independently performing integrity checking, there are monitoring station networks external to GPS that monitor satellites and transmit information about GPS integrity to users. Like the integrity information in the navigation message, also these are delayed, and cannot detect problems caused by the immediate surroundings of the user.

5.1.3 Reliability

The term *reliability* is used in many different meanings:

- The applicability of a device or system to its intended use, as a function of time
- The capability of a device or system to accomplish a required operation in defined conditions over a defined time period
- The probability that the operational unit accomplishes the required operation in defined conditions for a defined amount of time
- Fulfillment of performance design requirements by a device or system
- Fault tolerance of a device or system

Reliability of satellite positioning can mean two things, either a system reliability or statistical reliability of the results. In the first case, reliability is linked to reliability of devices and components, i.e. the probability of a correct operation. The analysis tools are then for instance probability densities of components to fail as function of time. Statistical reliability on the other hand means consistency of the measurement series. In satellite positioning this means that simultaneously taken measurements from different sources are available, whose mutual consistency can be studied if the positioning problem is over-determinend, i.e. there are more measurements available than necessary for positioning. In these notes positioning reliability is looked at from the point of view of statistical reliability of the positioning results.

In the analysis of geodetic measurement networks, reliability means the ability of the estimation to detect large errors [21], [22]. Positioning reliability defined this way is as a concept similar to integrity studied by RAIM algorithms. This kind of reliability definition can also be used in the performance analysis of personal positioning [23].

The geometry of the positioning problem affects not only the positioning accuracy but also how well the reliability or integrity of the positioning problem can be examined with consistency tests. The usual DOP used in the estimation of the accuracy of the positioning result does not always describe sufficiently a problem's applicability to consistency checks.

5.1.4 Availability

Positioning availability is affected by several factors. In the case of satellite positioning, the first prerequisite for positioning availability is that a receiver is able to receive a sufficient number of satellite signals. This depends not only on the satellite constellation, but also on the positioning environment — signals do not go through thick concrete structures. Although positioning availability may be excellent on the roofs of the highest buildings in the positioning environment, it would not necessarily be good at street level, inside buildings, in tunnels or in underground spaces.

When evaluating positioning availability, requirements can also be set for positioning quality (accuracy, reliability), then the number describing availability (probability or percentage of time available) usually gets smaller. For instance, positioning availability can be studied when the requirement is that the satellites used in positioning have elevation angle at least $5°$ and their PDOP (position dilution of precision) can be at most 6.

Availability of the integrity or reliability information can also be examined separately. To produce such information more visible satellites are required than just for positioning, and additional requirements for the geometry are imposed. If the positioning result is rejected because of the integrity or reliability check and the error can not be isolated, or the reliability check is not available because of geometry or number of satellites, then positioning availability gets worse. On the other hand, if the above checks are not done, but all measurements are accepted then, especially in personal positioning, the average positioning accuracy gets worse.

5.2   Statistical inference and hypothesis testing[*]

Many methods developed for integrity or reliability checks of positioning problems are based on statistical inference and hypothesis testing. The methods represented in this chapter are covered in more detail in references [18, 21, 22, 23, 30, 42]. The purpose of statistical inference is to determine the truth of some statement in the light of collected observations. A typical question is, for instance, if an industrial process really produces different quality product than before, or if the changed test results are just coincidental. When examining the measurement set, a common question is whether the measurement differing from the others is faulty or if the deviation of the detected magnitude could be a coincidence, in other words, is the deviation probable when the usual variation of the measurement in question is taken into account.

For inference, the question to be solved is rephrased into a hypothesis i.e. an assumption of the state. These assumptions, which can be either true or false, are called statistical hypotheses. Usually a hypothesis concerns one parameter of the probability distribution of the examined phenomenon.

Hypothesis testing is not meant to prove the hypothesis true or false. Instead, the aim is to show that the hypothesis is not believable, because it leads to very small probability, that is, the

---

[*]This section is based on the frequentist approach to statistics, as opposed to the bayesian approach.

collected observations deviate considerably from those which would have been probable if the hypothesis was true.

Example 23. *We want to find out whether the coin used in coin tossing is fair, i.e. whether the probability of tossing a head the same as tossing a tail. To study this, we form the hypothesis "the coin is fair": $p = 0.5$, where p is the probability to get a head in coin tossing. As an experiment, we study the fairness of the coin by tossing it 20 times. Say 16 of the tosses gives a head. Then we are quite ready to reject the hypothesis "the coin is fair", because when the hypothesis holds, the obtained result "16 heads" has a very small probability, that is, the test result deviates considerably from the one predicted by the hypothesis.*

The hypothesis to be contradicted is chosen such that if it holds, nothing changes or has not changed, or such that it presents a neutral situation that does not require any actions. For this reason it is often called the null hypothesis H_0. When studying the state of a production process, the null hypothesis would be "there are no changes in product quality" (no need to start looking and fixing the defect in the process that has caused a change in the quality). The null hypothesis in a measurement data analysis is "measurements are normal" (no need to start thinking what is faulty and needs to be removed from the sample set), and in the case of coin tossing "the coin is fair" (no need to change the coin, random hits can be generated with it).

Hypothesis testing can be described as a five-phased process:

1. Formulate the practical problem into a hypothesis.

   Start from the alternative hypothesis H_1. This should cover those cases that are to be diagnosed, where a positive test result means that it is time to take actions.

   The null hypothesis is chosen in such a way that it is simple and represents the status quo, and together with the alternative hypothesis it covers all possible situations. Examples:

   - H_1: "The coin is not fair", $p \neq 0.5$ (the coin has to be changed). H_0: "The coin is fair", $p = 0.5$. In these p is probability that the result in coin tossing is head.

   - H_1: "The new process B improves product quality compared to the original process A", $\mu_B > \mu_A$. H_0: "The new process B does not improve product quality compared to the original process A", $\mu_B \leq \mu_A$. In these μ_A and μ_B are expectations of a quantity describing product quality, where higher value describes better quality.

2. Choose a test statistics T, which is a function of the observations (measurements).

   Good test statistics have the following properties:

   (a) It behaves differently when H_0 is true compared to when H_1 is true.

   (b) Its probability distribution can be computed when H_0 is assumed to be true.

3. Choose a critical area.

   If the value of the test statistics T is in the critical area, H_0 is rejected in favor of H_1. If the value of T is not in the critical area we conclude that there is not enough evidence to support hypothesis H_1, and hypothesis H_0 is not rejected. No decision is made about whether or not to accept H_0; if there is no evidence for H_1, we fail to reject H_0.

The problem is to decide what kind of test statistics values strongly imply that H_1 is true. Critical areas are for instance the following:

Right-sided $H_1 : T > T_{cr}$. If the test statistics is larger than the (right-sided) critical value, H_0 is rejected.

Left-sided $H_1 : T < T_{cr}$. If the test statistics is smaller than the (left-sided) critical value, H_0 is rejected.

Two-sided $H_1 : T < T_{cr_1}$ or $T > T_{cr_2}$. If the test statistics is either larger than the right-sided critical value or smaller than the left-sided critical value, H_0 is rejected.

4. Decide the size of the critical area.

   Statistical inference is based on the idea that the decision made about rejecting or not rejecting H_0 can be correct or incorrect. Inference can lead to a situation where H_0 is rejected, although it is correct in reality – this is called a *type I error*. H_0 can also not get rejected, although it is incorrect in reality and should get rejected – this is called a *type II error*.

   When defining the critical area, we decide how big a risk of making a wrong decision are we ready to take. The probability that a type I error occurs is called the *significance level* of the test or the risk level and it is denoted with α. Often used significance levels are 5%, 1% and 0.1%. Our confidence in the decision to not reject the null hypothesis is described with confidence level $1 - \alpha$, which is the probability of the correct decision "H_0 is not rejected when H_0 is true".

   The probability of a type II error is denoted β. The *test efficiency* $1 - \beta$ is the probability that the real change from the status quo situation represented by H_0 is detected from the variation of the test statistics. Table 5.1 shows the four different possibilities occurring in hypothesis testing to end up to the correct or wrong decision .

5. Determine whether and where computed T is located in the critical area.

   If T is located inside but near the edge of the critical area, we can deduce that we have some evidence that H_0 should be rejected. On the other hand, if T is deep in the critical area, far away from the edge, we have significant evidence to support the rejection of H_0.

Table 5.1: Null hypothesis testing

		Decision	
		H_0 is not rejected	H_0 is rejected
Reality	H_0 true	correct decision probability $1 - \alpha$ = confidence level	type I error probability α = significance level
	H_0 false	type II error probability β	correct decision probability $1 - \beta$ = test power

Example 24 (α, β and the magnitude of the difference observed with a test, [42]). *We illustrate hypothesis testing with an example where the problem geometry is simpler than in a typical positioning problem. We observe the weights of male students at a certain university. The null hypothesis is that the male students weigh on average 68 kg, so we set the hypotheses as $H_0 : \mu = 68$ and $H_1 : \mu \neq 68$. The alternative hypothesis covers the cases $\mu < 68$ or $\mu > 68$.*

The test statistics is chosen to be the sample mean $\bar{x}_0$. If it occurs near the assumed value 68, we have little or no evidence to reject H_0. If the sample mean is considerably larger or smaller than 68, the data is not consistent with H_0 but it supports alternative hypothesis H_1. The critical area is chosen (rather arbitrarily) to be in intervals $\bar{x} < 67$ and $\bar{x} > 69$; The non-rejecting area, where H_0 is not rejected, is $67 \leq \bar{x} \leq 69$. These have been visualized in Figure 5.1.
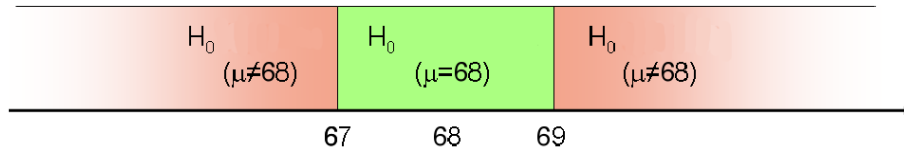
Figure 5.1: Critical area (on the edges).

Compute the probabilities for type I and II errors when the inference criterion is in accordance with Figure 5.1. Assume that in the population the standard deviation of weight is $\sigma = 3.6$. The sample mean $\bar{X}$, which is used as a test statistics, is computed from a randomly selected sample set whose size is $n = 3.6$. Based on the central limit theorem it is known that the distribution of $\bar{X}$ approaches a normal distribution whose standard deviation is $\sigma_{\bar{X}} = \sigma / \sqrt{n} = 3.6/6 = 0.6$.
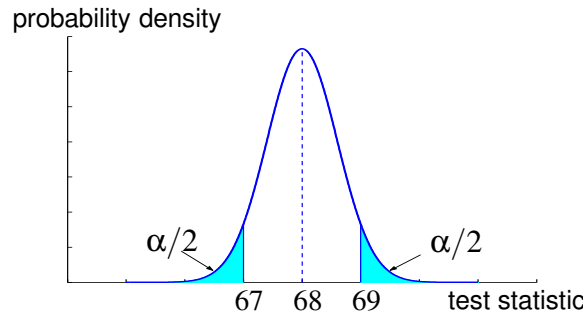
Figure 5.2: Critical area when testing hypothesis $\mu = 68$ against hypothesis $\mu \neq 68$.

The significance level of the test α, i.e. the probability of a type I error (where H_0 is rejected, though H_0 is true), is equal to the sum of the shaded areas in Figure 5.2. The probability for the case where $\bar{X} \sim N\left(68, 0.6^2\right)$ and $\bar{X}$ is in the critical area is

$$\alpha = P\left(\bar{X} < 67\right) + P\left(69 < \bar{X}\right) = 2P\left(\bar{X} < 67\right) = 0.0956.$$

According to the result, about 9.6% of all 36-sample sets would cause rejection of hypothesis $\mu = 68$ although the hypothesis is true. To lower the probability α, we should either increase the

sample set size (which will reduce the sample standard deviation) or increase the non-rejection area. Assume that the sample set size is increased to n = 64. Then $\sigma_{\bar{X}} = 3.6/8 = 0.45$. Now

$$\alpha = 2\,\mathrm{P}\,(\bar{X} < 67) = 0.0263.$$

A low probability of a type I error i.e. a small value of α, does not guarantee a good test. We also want that the probability β of a type II error is sufficiently small. To study the probability β, we compute values of β with fixed values of μ that satisfy the alternative hypothesis $\mu \neq 68$. For instance, we want to make sure that H_0 is rejected with high probability, if the distribution's true expectation is $\mu \leq 66$ or $\mu \geq 70$. Then β can be computed for alternative hypothesis $\mu = 66$ or $\mu = 70$. Because of symmetry it suffices to find the probability that the null hypothesis $\mu = 68$ is not rejected when the alternative hypothesis $H_1 : \mu = 70$ is true. Then a type II error occurs if the sample mean occurs in the interval $[67, 69]$ while H_1 is true. From Figure 5.3 we get

$$\beta = \mathrm{P}\,(67 \leq \bar{X} \leq 69 \text{ when } \mu = 70)\,.$$
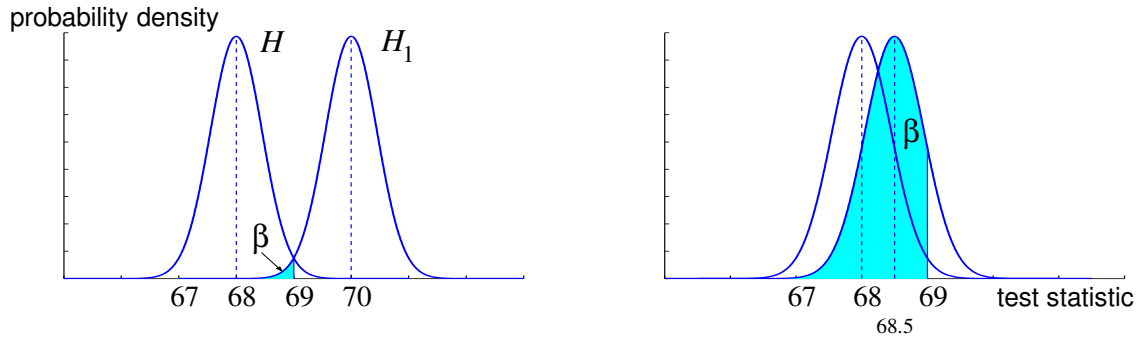
Figure 5.3: Probability of type II error when testing $\mu = 68$ against $\mu = 70$ (left) and against $mu = 68.5$ (right).

Assume further that n = 64. Then, according to the alternative hypothesis $\bar{X} \sim \mathrm{N}\,(70, 0.45^2)$, and

$$\beta = \mathrm{P}\,(67 \leq \bar{X} \leq 69) = \mathrm{P}\,(\bar{X} \leq 69) - \mathrm{P}\,(\bar{X} \leq 67) = 0.0131 - 0.0000 = 0.0131.$$

If the true μ is $\mu = 66$, then β is the same 0.0131. For other possible values $\mu < 66$ or $\mu > 70$, then the probability β is even smaller, so the probability of not rejecting H_0 is small if H_0 is false. The probability of type II error increases if true μ is not equal but approaches the assumed value of the null hypothesis. On the left side of Figure 5.3 we have an example where the true $\mu = 68.5$. Then $\bar{X} \sim \mathrm{N}\,(68.5, 0.45^2)$, and

$$\beta = \mathrm{P}\,(67 \leq \bar{X} \leq 69) = 0.8667 - 0.0004 = 0.8663.$$

To conclude it can be stated that the probabilities of type I and II errors, α and β, depend on each other, and β depends on distance between the true value and the value assumed in the null hypothesis. A more theoretical view on this matter is handled for instance in [44] or [14].

5.3 Residuals

In this section, we review the properties of residuals, and introduce the notations used in the remainder of this chapter, which in some ways differ from the notations used earlier in this hand-out. In the analysis it is assumed that the relation between measurements and parameters to be estimated can be presented using the linear measurement model

$$\mathbf{y} = H\mathbf{x} + \boldsymbol{\varepsilon}, \tag{5.1}$$

where the $n \times 1$-vector $\mathbf{y}$ contains measurements (observations), $\boldsymbol{\varepsilon}$ contains measurement errors, $m \times 1$-vector $\mathbf{x}$ contains the states (parameters to be estimated), and the measurement matrix H defines a linear relation between measurements and states. In positioning applications, the matrix H is also called the geometry matrix. In practical positioning applications, the computations are based on a nonlinear model, and H is obtained by linearizing the system of nonlinear measurement equations.

Assume that a location has been solved using the weighted least squares method. If weights are chosen to be the inverse of the measurement variance matrix, the weighted least squares method gives the BLU estimate (Best Linear Unbiased Estimate).

The covariance of the measurement errors is assumed known: $V(\mathbf{y}) = V(\boldsymbol{\varepsilon}) = \Sigma$. Then the weighted least squares solution (WLS-solution) is

$$\hat{\mathbf{x}} = \left(H^T \Sigma^{-1} H \right)^{-1} H^T \Sigma^{-1} \mathbf{y} = K\mathbf{y}. \tag{5.2}$$

In the following it is assumed that the covariance matrix is diagonal, i.e. measurement errors do not correlate with each other. Residuals can be written as a function of the measurements:

$$\mathbf{v} = H\hat{\mathbf{x}} - \mathbf{y} = HK\mathbf{y} - \mathbf{y} = (HK - I)\mathbf{y} = -R\mathbf{y}, \tag{5.3}$$

where matrix R is called the redundancy matrix. It is left as an exercise to show that matrix R is idempotent. The covariance of residuals can be represented using the geometry matrix and the covariance of measurement errors (exercise):

$$C_v = V(\mathbf{v}) = (HK - I)V(\mathbf{y})(HK - I)^T = \Sigma - H\left(H^T \Sigma^{-1} H \right)^{-1} H^T. \tag{5.4}$$

The redundancy matrix can be represented using covariances of residuals and measurements (exercise):

$$R = C_v \Sigma^{-1}. \tag{5.5}$$

5.3.1 Quadratic form of residuals

Quadratic form of residuals $\mathbf{v}^T \Sigma^{-1} \mathbf{v}$ is a common test parameter in reliability analysis of positioning. To accomplish our study, we assume that the measurement errors in model (5.1) are normally distributed: $\boldsymbol{\varepsilon} \sim N(\boldsymbol{\mu}, \Sigma)$.

The quadratic form can be written using the measurements:

$$\mathbf{v}^T \Sigma^{-1} \mathbf{v} = (-R\mathbf{y})^T \Sigma^{-1} (-R\mathbf{y}) = \boldsymbol{\varepsilon}^T R^T \Sigma^{-1} R \boldsymbol{\varepsilon}.$$

Denote $A = R^T \Sigma^{-1} R$. In order to use the result of Theorem 6 (on page 11), we determine whether matrix $A\Sigma$ is idempotent. Using Equation (5.5), we can write

$$A = R^T \Sigma^{-1} R = \left(C_v \Sigma^{-1}\right)^T \Sigma^{-1} C_v \Sigma^{-1} = \Sigma^{-1} C_v \Sigma^{-1} C_v \Sigma^{-1} = \Sigma^{-1} R R = \Sigma^{-1} R, \qquad (5.6)$$

because Σ^{-1} and C_v are symmetric and R is idempotent. From this we get $A\Sigma$ and $A\Sigma A\Sigma$:

$$A\Sigma = \Sigma^{-1} R \Sigma$$
$$A\Sigma A\Sigma = \Sigma^{-1} R \Sigma \Sigma^{-1} R \Sigma = \Sigma^{-1} R R \Sigma = \Sigma^{-1} R \Sigma = A\Sigma,$$

so $R^T \Sigma^{-1} R \Sigma$ is idempotent and $\mathbf{v}^T \Sigma^{-1} \mathbf{v} \sim \chi^2 (n-m, \lambda)$, where $\text{rank}(R^T \Sigma^{-1} R) = n - m$ (exercise) is the number of degrees of freedom, where n is the number of measurement equations and m is the number of states to be estimated. If measurement errors are assumed zero mean, i.e. $\boldsymbol{\mu} = \mathbf{0}$, then the non-centrality parameter λ is 0, and then the quadratic form follows a central χ^2 distribution. If not all measurement errors are assumed zero mean, i.e. $\boldsymbol{\mu} \neq \mathbf{0}$, then the non-centrality parameter $\lambda = \boldsymbol{\mu}^T A \boldsymbol{\mu} = \boldsymbol{\mu}^T R^T \Sigma^{-1} R \boldsymbol{\mu}$ and the quadratic form's distribution is non-central χ^2 distribution.

Consider the case where $\boldsymbol{\varepsilon} \sim N(\boldsymbol{\mu}, \Sigma)$, where measurement y_i contains bias error, i.e. the expectation of that error is non-zero, while other measurement errors are zero-mean: $E(\varepsilon_i) = \mu_i \neq 0$, $E(\varepsilon_j) = \mu_j = 0 \; \forall j \neq i$, where $j = 1 \ldots n$. Then the value of the non-centrality parameter can be computed using results (5.6) and (5.5) and diagonality of Σ:

$$\begin{aligned} \lambda = \boldsymbol{\mu}^T A \boldsymbol{\mu} &= \mu_i^2 [A]_{(i,i)} = \mu_i^2 \left[\Sigma^{-1} R\right]_{(i,i)} \\ &= \mu_i^2 \left[\Sigma^{-1} C_v \Sigma^{-1}\right]_{(i,i)} = \mu_i^2 / \sigma_i^4 [C_v]_{(i,i)} . \end{aligned} \qquad (5.7)$$

5.3.2   Standardized residuals

A standardized residual w_i is obtained by dividing the residual by the square of its variance:

$$w_i = \frac{v_i}{\sqrt{[C_v]_{(i,i)}}}, \quad i = 1 \ldots n. \qquad (5.8)$$

Consider again the case where $\boldsymbol{\varepsilon} \sim N(\boldsymbol{\mu}, \Sigma)$, where measurement y_i contains bias error, i.e. the expectation of this error is non-zero, while other measurement errors are zero-mean: $\mu_i \neq 0$, $\mu_j = 0 \; \forall j \neq i$, where $j = 1 \ldots n$. Denote diagonal items of Σ with $[\Sigma]_{(i,i)} = \sigma_i^2$.

For the expectation of residual vector items we get based on (5.3)

$$E(v_j) = E\left([-R\mathbf{y}]_j\right) = -[R]_{(j,i)} \mu_i.$$

For the expectation of standardized residuals we get

$$E\left(w_j\right) = E\left(v_j / \sqrt{[C_v]_{(j,j)}}\right) = -\mu_i \left[R\right]_{(j,i)} / \sqrt{[C_v]_{(j,j)}}$$

$$= -\mu_i \left[C_v\right]_{(j,i)} \left[\Sigma^{-1}\right]_{(i,i)} / \sqrt{[C_v]_{(j,j)}}$$

$$= -\mu_i / \sigma_i^2 \left[C_v\right]_{(j,i)} / \sqrt{[C_v]_{(j,j)}}.$$

For the expectation of standardized residual of the ith measurement, we get

$$E\left(w_i\right) = -\mu_i / \sigma_i^2 \sqrt{[C_v]_{(i,i)}}. \tag{5.9}$$

5.4   RAIM (Receiver Autonomous Integrity Monitoring)

RAIM is a two-phased problem. First the algorithm determines whether or not there is an error in navigation measurements or models. This is called error detection. If an error was detected, the second phase of the algorithm finds a combination from available measurements that does not contain the faulty equation (signal and model).

The second phase of RAIM can be accomplished either by identifying the faulty equation and removing it from the solution, or alternatively by seeking a combination of available equations which seems error-free, i.e. which does not cause error detection. The first alternative is called FDI, fault detection and identification. The other method is called FDE, fault detection and exclusion (from the solution). Commonly the goal of FDI is also FDE, that is, to identify erroneous equation and remove it from the solution, then the whole process is often called FDE.

If satellite positioning is only a supplemental navigation method, then just detecting error can be sufficient. Then the system uses another navigation method, for instance INS, if RAIM detects an error in the satellite positioning equations. If satellite positioning is the only navigation method in use, it is often preferred to isolate the error from the solution rather than having no location solution at all.

Use of RAIM methods requires redundancy. In satellite positioning at least four measurements are required to solve location, whereas RAIM requires at least five measurements for detecting error, and for error isolation at least six measurements.

There exist several RAIM methods that are seemingly different: a method based on studying least squares residuals, a so-called parity method, a method based on comparison of ranges, and maximum residual. In the GPS literature these different methods have been shown to be equivalent. In this section we go through the method based on least squares residuals.

5.4.1   Error detection

The measurement model is assumed the same as (5.1) and covariance of measurement errors $\Sigma = \sigma^2 I$. Then the weight matrix Σ^{-1} in the position solution (5.2) is the identity matrix, and

thus $\mathrm{K} = \left(\mathrm{H}^T \mathrm{H}\right)^{-1} \mathrm{H}^T$ and redundancy matrix $\mathrm{R} = \mathrm{I} - \mathrm{H}\left(\mathrm{H}^T \mathrm{H}\right)^{-1} \mathrm{H}^T$.

To identify the error the observable is the sum of squared residuals, abbreviated with SSE (Sum of the Squares of range residual Errors): $SSE = \mathbf{v}^T \mathbf{v}$.

If measurement errors are zero mean, then $SSE \sim \chi^2\,(n-4)$, where n is number of available measurements. If the mean of the measurement errors is non-zero, the distribution of SSE is non-central, $SSE \sim \chi^2\,(n-4,\lambda)$.

The test statistics is chosen to be

$$T = \sqrt{SSE/\left(n-4\right)} = \sqrt{\mathbf{v}^T \mathbf{v}/\left(n-4\right)},$$

where the square root and division by the number of degrees of freedom is justified by the argument that this reduces the dependence of the test statistics on satellite geometry. The critical value of the test statistics is chosen to be

$$T_{cr} = \sqrt{\frac{\sigma^2 \times \chi^2_{1-p_{fa}}\,(n-4)}{n-4}},$$

where $\chi^2_{1-p_{fa}}\,(n-4)$ is the inverse function of the χ^2 cumulative distribution function, with probability $1 - p_{fa}$ as argument and parameter (degrees of freedom) in brackets. Probability p_{fa} is the probability of false alarms, false alarm rate. In typical RAIM implementations p_{fa} is small, for instance in aviation with GPS as is the primary navigation method, $p_{\mathrm{fa}} = 0.333 \times 10^{-6}$ [19].

5.4.2 Effect of geometry

Effect of satellite geometry on RAIM's capability to detect an error is studied by comparing the effect of measurement error on the estimate and its effect on the residual. If the effect of measurement error on residual is small but its effect on the estimate is large, RAIM's possibilities to protect against large errors in location estimate is poor. In this kind of situation the user should at least be aware of the situation — poor geometry does not generally mean that there is an error, but if there happens to be an error, it is not detected.

As an example we next present ARP checking (approximate radial error protected), whose purpose is to find poor geometries that do not allow RAIM to detect errors. Other methods have also been proposed to determine whether the geometry is sufficient for RAIM. *ARP* gives a rough estimate of the magnitude of the positioning error which RAIM protects against. *ARP* is computed with formula

$$ARP = S_{max} T_{cr},$$

where S_{max} is the largest of coefficients S_i that describe dependence of the test statistics and the horizontal location error caused by error in measurement i:

$$S_i = \sqrt{\frac{\left([\mathrm{K}]^2_{(1,i)} + [\mathrm{K}]^2_{(2,i)}\right)(n-4)}{[\mathrm{R}]_{(i,i)}}}.$$

It is assumed that the geometry is represented in ENU-coordinates, then the elements of the two first rows of matrix K pick the effect of the measurement to the horizontal coordinates of location estimate. To eliminate poor geometries the computed *ARP* value is compared to a threshold value ARP_{cr}: if $ARP < ARP_{cr}$, then RAIM can be used, if $ARP \geq ARP_{cr}$, then RAIM cannot detect errors of such magnitude that the ARP threshold is designed to protect against.

5.4.3 Error identification / isolation

If error identification detected an error and at least 6 measurements are available, from the available *n* measurements *n* subsets are formed, where each subset omits one of the original satellites. The ARP value is computed for each of these subsets. If all ARP-values are acceptable, the value of the test variable is computed for the subsets. If there is one faulty measurement in the whole measurement set, it will be missing from one of these subsets, and the value of the computed test statistics for this error-free subset should be less than the critical value. In other subsets however this error still exists, then test statistics value exceeds critical value. If testing of the subsets finds only one combination whose test result does not imply an error, the missing measurement (or equation) from this combination is found faulty and a new, acceptable location solution can be computed by leaving out the equation that is identified to be faulty. If however the test statistics value of more than one subset is below the critical value or the ARP value of some subset is not acceptable, the error cannot be isolated. Then the obtained positioning solution cannot be improved, and it is reported to be unreliable.

5.5 Reliability testing using global and local test

The reliability testing method represented in this section originates from reliability testing of geodetic measurement networks [22], where concepts differ somewhat from the ones used in the RAIM context. Principles are however similar in both: using statistical reliability testing we determine whether measurements are in accordance with the assumptions. First, the whole measurement set is observed, and a global test is used to detect an error (outlier detection). After that, if an error was detected, local tests are done to identify faulty measurements (gross error localization). The theory of this method is based on the assumption of one faulty measurement, but practical applications have been developed from it for a case of many simultaneous errors. Examples on how this method is used and more references are presented in [23].

Null hypothesis in testing is "there are no bias errors", when measurement errors are zero mean. Alternative hypothesis is "at least one measurement contains bias error", when measurement errors are not zero mean. If there are no bias errors, the distribution of the measurement errors is assumed to be $H_0 : \boldsymbol{\varepsilon} \sim N(0, \Sigma)$, in alternative case $H_1 : \boldsymbol{\varepsilon} \sim N(\boldsymbol{\mu}, \Sigma)$.

5.5.1 Global test

The purpose of a global test is to determine whether or not there are faulty measurements in the measurement set. The quadratic form of residuals is used as test statistics:

$$T = \mathbf{v}^T \Sigma^{-1} \mathbf{v}.$$

Based on Section 5.3.1, it is observed that the distribution of T is central χ^2 distribution if H_0 holds, and a non-central χ^2 distribution if H_1 holds. The number of degrees of freedom, $n - m$, of the distribution is defined based on the number of available measurements n and the number of unknown states m. Defining the parameter λ of the non-central distribution is left as a part of test design.

Because test statistics following a central χ^2 distribution have smaller values than test statistics following non-central χ^2 distribution with the same number of degrees of freedom, a right-sided critical area is selected for the test. Then the hypotheses to be tested are

$$\begin{aligned} H_0: & \quad T \leq T_{cr} \\ H_1: & \quad T > T_{cr}. \end{aligned}$$

In order to define the critical value the significance level α of the test is decided, i.e. how large a probability for type I error is still acceptable. The critical value is chosen such that when $T \sim \chi^2 (n - p)$, then $\mathrm{P}(T > T_{kr}) = 1 - \alpha$, i.e.

$$T_{kr} = \chi^2_{1-\alpha}(n - m),$$

where $\chi^2_p(\cdot)$ is the inverse function of the cumulative function of the χ^2 distribution, with probability p as an argument and parameters of the distribution in brackets.

Performing a global test does not in itself require fixing the probability β and the parameter λ. The dependence between these two is however nice to know, and additionally the values are needed for defining the critical value of the local test.

In example 24 (on page 76) it was noted that probabilities α and β of type I and II errors depend on each other, and the magnitude of β is affected by how much the parameter assumed by the null hypothesis differs from the true value. This is used to select the λ value. First it is decided what probability β of faulty decision is acceptable. After that a value of λ is sought such that the cumulative function of distribution $\chi^2(n - m)$ has the value $1 - \alpha$ and the cumulative function of distribution $\chi^2(n - m, \lambda)$ gets value β for the same T value:

$$\chi^2_{1-\alpha}(n - m) = \chi^2_\beta(n - m, \lambda).$$

In Figure 5.4 parameter selection is visualized in a case where the number of degrees of freedom is 3. Parameter values $\alpha = 0.11$ and $\beta = 0.20$ have been chosen, because using these values the test statistics areas leading to faulty decisions are easily shown in the figure. In practice especially the value of α is chosen to be smaller, $\alpha = 0.001$ is a typical choice. Commonly used β values are 0.2, 0.1 and 0.05.
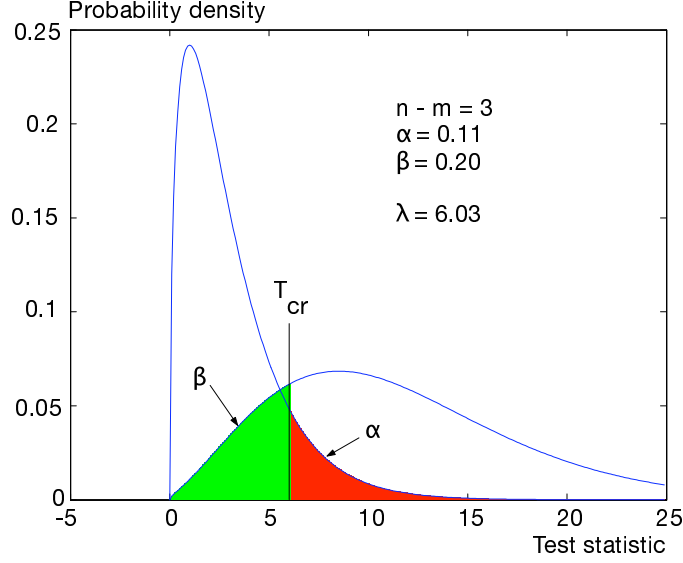
Figure 5.4: Defining parameter λ using α and β.

5.5.2 Local test

If the global test result is $T > T_{cr}$, it is inferrred that there is at least one erroneous measurement among the measurements. A local test is used to identify those measurements that are suspected to be faulty. According to H_0 the distribution of measurement errors $\boldsymbol{\varepsilon}$ is $\mathrm{N}(0, \Sigma)$, and then the distribution of residuals $\mathbf{v}$ is $\mathrm{N}\left(0, \mathrm{R}\Sigma\mathrm{R}^T\right)$. Standardized residuals are obtained from residuals according to formula (5.8) by scaling each residual by its own variance, so for these $w_i \sim \mathrm{N}(0,1)$.

As the test statistics is chosen to be a standardized residual the hypotheses to be tested concern normal distribution. In the global test, the whole residual vector is tested at once, whereas in a local test each residual is tested separately. Hypotheses concerning residual w_i are $H_0 : w_i \sim \mathrm{N}(0,1)$ and $H_1 : w_i \sim \mathrm{N}(\delta, 1)$, where $\delta \neq 0$. Tests corresponding to the hypotheses are

$$
\begin{aligned}
H_{0,i} : &\quad -w_{cr} \leq w_i \leq w_{cr} \\
H_{1,i} : &\quad w_i < -w_{cr} \text{ or } w_i > w_{cr}.
\end{aligned}
$$

Because the distribution is symmetrical, tests can be made right-sided by taking the absolute value of standardized residual as test statistics:

$$
\begin{aligned}
H_{0,i} : &\quad |w_i| \leq w_{kr} \\
H_{1,i} : &\quad |w_i| > w_{kr}.
\end{aligned}
$$

In the local test, the mean δ of the distribution in alternative hypothesis is non-zero, corresponding to the assumption in the global test that the non-centrality parameter λ is non-zero. These parameters depend on each other, and using this information the δ of the alternative hypothesis of a local test can be defined. Because in both tests the probability β and the deviation of null hypothesis's parameter (λ or δ) from the true value depend on each other, and since

84

on the other hand δ is defined from the parameter of the global test, it is natural to use the same probability β as in the global test for the selection of the critical value δ. Then the significance level α_0 of the local test is defined using fixed parameters δ and β.

In both global and local test the alternative hypothesis is based on the assumption that measurements are normally distributed, but their mean is non-zero. Assume that all other measurement errors are zero mean, but measurement i has bias: $E(\varepsilon_i) = \mu_i$. Let us compare the effect of bias of measurement i on parameter λ and on $E(w_i)$, the expectation of standardized residual. From formulas (5.7) and (5.9) we see that $\lambda = E(w_i)^2$, and from this it follows that $|E(w_i)| = \sqrt{\lambda}$. Because the absolute value of w_i is examined in the local test, those distributions of the alternative hypothesis whose mean is positive are interesting. Then

$$\delta = |E(w_i)| = \sqrt{\lambda}.$$

Because δ and β define the critical value location, the problem becomes defining what value of α_0 it corresponds to. From Figure 5.5 we can see the following equalities:

$$a = n_{1-\alpha_0/2}(0,1) = w_{kr}$$
$$b = n_\beta(\delta,1) = n_{1-\beta}(\delta,1) - \delta = n_{1-\beta}(0,1) + \delta - \delta$$
$$\delta = a + b = n_{1-\alpha_0/2}(0,1) + n_{1-\beta}(0,1),$$

where $n_p(\cdot)$ is the inverse function of the cumulative function of normal distribution with probability p as an argument and distribution parameters in brackets.
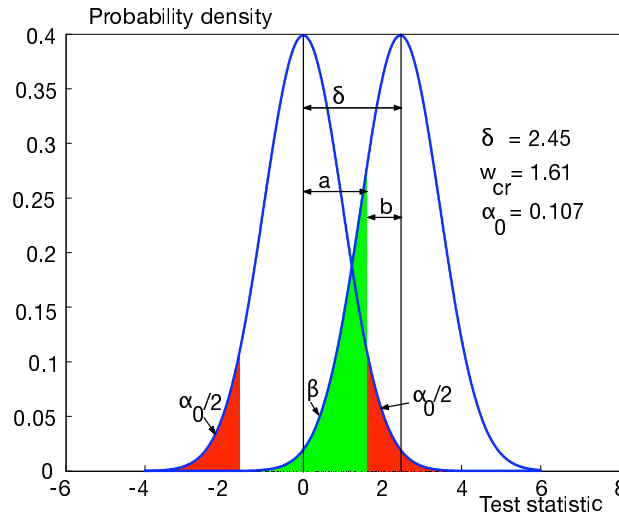
Figure 5.5: Defining parameter α_0 using δ and β.

5.5.3 Internal and external reliability

The smallest bias of the measurement, i.e. an outlier, that can still be detected by using statistical testing is called MDB, minimum detectable bias or marginally detectable bias). It is used as a

measure of internal reliability of positioning problem. *External reliability* on the other hand indicates the effect of the smallest detectable bias on the estimate.

Expectation δ is the smallest bias in standardized residuals that can be detected by using the test with probability $1 - \beta$ and with confidence level $1 - \alpha$. The smallest bias of measurement i that can be detected using the test is

$$m_i = \frac{\delta \left[\Sigma \right]_{(i,i)}}{\sqrt{\left[C_v \right]_{(i,i)}}}. \tag{5.10}$$

External reliability describes the effect of the largest undetected bias on the estimate. Let's assume that in measurement i there is bias with magnitude m_i, and that in other measurements there are no errors, i.e. the measurement error vector is $\Delta \mathbf{y_i} = \begin{bmatrix} 0 \ldots 0 & m_i & 0 \ldots 0 \end{bmatrix}^T$. Therefore, the external reliability is obtained from

$$\mathbf{e_i} = K \Delta \mathbf{y_i} = \left[K \right]_{(1:n,i)} m_i = \left[\left(H \Sigma^{-1} H^T \right) H^T \Sigma^{-1} \right]_{(1:n,i)} m_i. \tag{5.11}$$

External reliability tells how large estimation errors the testing can protect against — smaller errors will not be detected in the test. If we consider a typical satellite positioning problem and the three first elements of the state vector are location coordinates, then from external reliability we can compute the radius of a sphere that contains the error of a location estimate caused by a measurement error that is not detected in the test:

$$r_{PPL_i} = \sqrt{e_{i_1}^2 + e_{i_2}^2 + e_{i_3}^2}.$$

External reliability is also called the protection level (PPL, position protection level). If positioning computations have been done in ENU-coordinates or $\mathbf{e_i}$ is transformed to ENU, we can compute the horizontal protection level (HPL)

$$r_{HPL_i} = \sqrt{e_{i_1}^2 + e_{i_2}^2},$$

which is the radius of a circle that contains horizontal errors caused by a measurement that is not detected in the test.

Protection levels r_{PPL_i} and r_{HPL_i} are not equal for different measurements, so a conservative assessment of the general protection level of testing is obtained by computing protection levels for all measurements and selecting the largest. Here a larger number describing protection level means that larger errors go through the testing undetected.

Thus, reliability testing does not detect the smallest biases, and thus their identification cannot be done. On the other hand very large errors also affect the estimate to which the iterative linearization-based solution of navigation equations converges. Then the linearization point drifts too far from the true location, and then also the geometry matrix will contain large errors. These kinds of errors usually increase almost all residuals, which is usually seen in the test statistics of the global test and the error is detected. However, error identification can be difficult, because the geometry matrix that is used in error analysis is also erroneous.

5.5.4 FDE

The practical FDE process is iterative. The global test aims to detect whether there are errors in the measurement set. If an error is detected, a local test is done for each standardized residual w_i. Because there can be more than one error, more than one measurement may be found whose w_i does not pass the test.

A residual v_i with large absolute value can be caused by an error in some other measurement than y_i. Often the redundancy matrix R, which describes measurement errors as residuals, contains items outside its diagonal that have the properties

$$\left| [R]_{(j,i)} \right| > [R]_{(i,i)} \ \text{ or } \ \left| [R]_{(i,j)} \right| > [R]_{(i,i)}.$$

In the first case the error of measurement i affects residual j more strongly than its "own" residual. In the second case the effect of measurement j shows more strongly in residual i than the effect of measurement i. This makes error detection more difficult or impossible. The possibility of such a problem can be determined by studying the elements of R.

For the reasons mentioned earlier, a test variable w_i that does not pass the local test cannot be directly considered an indication that the corresponding measurement y_i contains error, it only gives a reason to suspect that measurement y_i is faulty. That is why the global test is done again after removing the suspected measurement. If the global test further gives reason to suspect that even the modified measurement set contains error, testing is continued. When finally a measurement set is found that passes the global test, it is often reasonable to try whether some removed measurement could be added to the set without the global test giving an error again. This is done because it is possible that some large error that has been removed later would have strongly appeared in some residual, and the measurement corresponding to it was removed in an earlier state of testing.

Exercises

5.1. (a) Show that the redundancy matrix R is idempotent.

(b) With what kind of weight matrix Σ^{-1} is matrix R symmetric?

5.2. (a) Show that residual covariance satisfies $C_v = \Sigma - H\left(H^T\Sigma^{-1}H\right)^{-1}H^T$ (5.4).

(b) Show that $R = C_v\Sigma^{-1}$ (5.5).

5.3. Show that $\text{rank}(R^T\Sigma^{-1}R) = n - m$. Hint: Show that

(a) $x \in \mathcal{N}(R^T\Sigma^{-1}R)$ *if and only if* $x \in \mathcal{N}(R)$

(b) $x \in \mathcal{N}(R)$ *if and only if* $x \in \mathcal{R}(H)$

(c) $\dim\left(\mathcal{R}(H)\right) = m$

Index

Bibliography

Most, if not all, of the references should be available either electronically or from the TUT library.

[1] ALANEN, K., WIROLA, L., KÄPPI, J., AND SYRJÄRINNE, J. Mobile RTK: Using low-cost GPS and internet-enabled wireless phones. *Inside GNSS 1* (May/June 2006), 32–39.

[2] ALLAN, D. W., ASHBY, N., AND HODGE, C. C. The science of timekeeping. Application note 1289, Hewlett Packard, 1997. http://www.allanstime.com/Publications/DWA/Science_Timekeeping/TheScienceOfTimekeeping.pdf.

[3] ASH, R. B., AND GARDNER, M. F. *Topics in Stochastic Processes*, vol. 27 of *Probability and Mathematical Statistics*. Academic Press, 1975.

[4] BANCROFT, S. An algebraic solution of the GPS equations. *IEEE Transactions on Aerospace and Electronic Systems 21*, 7 (1986), 56–59. http://ieeexplore.ieee.org/.

[5] BAR-SHALOM, Y., LI, R. X., AND KIRUBARAJAN, T. *Estimation with Applications to Tracking and Navigation, Theory Algorithms and Software*. John Wiley & Sons, 2001.

[6] BROWN, R. G. *Introduction to Random Signal Analysis and Kalman Filtering*. John Wiley & Sons, 1983.

[7] COLLIN, J. Investigations of self-contained sensors for personal navigation. Phd thesis, Tampere University of Technology, October 2006.

[8] DAM, E. B., KOCH, M., AND LILLHOLM, M. Quaternions, interpolation and animation. Tech. rep., Department of Computer Science, University of Copenhagen, 1998. http://www.itu.dk/people/erikdam/DOWNLOAD/98-5.pdf.

[9] FANG, B. T. Simple solutions for hyperbolic and related position fixes. *IEEE Transactions on Aerospace and Electronic Systems 26*, 5 (1990), 748–753. http://ieeexplore.ieee.org/.

[10] FARRELL, J. A., AND BARTH, M. *The Global Positioning System & Inertial Navigation*. McGraw-Hill, 1999.

[11] HATCH, R. Instantaneous ambiguity resolution. In *Proceedings of IAG International Symposium No. 107 on Kinematic Systems in Geodesy, Surveying and Remote Sensing* (September 1990), Springer Verlag, pp. 299–308.

[12] JAZWINSKI, A. H. *Stochastic Processes and Filtering Theory*, vol. 64 of *Mathematics in Science and Engineering*. Academic Press, 1970.

[13] JHS. Public government recommendation JHS 153: ETRS89 coordinates in finland, 2006. http://docs.jhs-suositukset.fi/jhs-suositukset/JHS153/JHS153.pdf.

[14] KALEVA, O. Matemaattinen tilastotiede. Hand-out, TUT, Institute of Mathematics, November 2005. http://www.tut.fi/~kaleva/.

[15] KALEVA, O. Matemaattinen optimointi 1. Hand-out, TUT, Institute of Mathematics, January 2006. http://www.tut.fi/~kaleva/.

[16] KALEVA, O. Stokastiset prosessit. Hand-out, TUT, Institute of Mathematics, January 2007. http://www.tut.fi/~kaleva/.

[17] KALMAN, R. E. A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, 82 (1960), 35–45.

[18] KANJI, G. K. *100 Statistical Tests*. SAGE Publications Ltd, 1999.

[19] KAPLAN, E. D., Ed. *Understanding GPS: principles and applications*. Artech House, Norwood, 2005.

[20] KINCAID, D., AND CHENEY, W. *Numerical analysis*, second ed. Brooks/Cole Publishing Company, Pacific Grove, California, 1996.

[21] KOCH, K.-R. *Parameter Estimation and Hypothesis Testing in Linear Models*, 2nd ed. Springer-Verlag Berlin Heidelberg, 1999.

[22] KUANG, S. *Geodetic Network Analysis and Optimal Design: concepts and Applications*. Ann Arbor Press Inc., 1996.

[23] KUUSNIEMI, H. User-level reliability and quality monitoring in satellite-based personal navigation. Phd thesis, Tampere University of Technology, September 2005. http://webhotel.tut.fi/library/tutdiss/pdf/kuusniemi.pdf.

[24] LEVA, J. L. An alternative closed-form solution to the GPS pseudo-range equations. *IEEE Transactions on Aerospace and Electronic Systems 32*, 4 (1996), 1430–1439. http://ieeexplore.ieee.org/.

[25] MAYBECK, P. S. *Stochastic Models, Estimation, and Control*, vol. 141 of *Mathematics in Science and Engineering*. Academic Press, 1979.

[26] MISRA, P., AND ENGE, P. *Global Positioning System: Signals, Measurements, and Performance*, 2nd ed. Ganga-Jamuna Press, 2006.

[27] PARKINSON, B., AND SPILKER, J., Eds. *Global Positioning System: Theory and Applications Volume I.* Charles Stark Draper Laboratory, Inc., Cambridge, 1996.

[28] PESONEN, H. Numerical integration in Bayesian positioning. M.Sc. thesis, Tampere University of Technology, June 2006. http://math.tut.fi/posgroup/pesonen_mscth.pdf.

[29] PHATAK, M., CHANSARKAR, M., AND KOHLI, S. Position fix from three GPS satellites and altitude: a direct method. *IEEE Transactions on Aerospace and Electronic Systems 36*, 1 (1999), 350–354. http://ieeexplore.ieee.org/.

[30] POHJAVIRTA, A., AND RUOHONEN, K. Laaja tilastomatematiikka. Hand-out, TUT, Institute of Mathematics, 2005. http://math.tut.fi/~ruohonen/LTM.pdf.

[31] POHJOLAINEN, S. Matrix algebra 1. Lecture Notes, TUT, Institute of Mathematics, 2004. http://matpc41.ee.tut.fi/PNF:byName:/ML02/.

[32] POOLE, D. *Linear Algebra*, second ed. Thomsom Brooks/Cole, 2006.

[33] PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A., AND VETTERLING, W. T. *Numerical recipes: the art of scientific computing*. Cambridge University Press, Cambridge, 1986. http://www.numerical-recipes.com/.

[34] RISTIC, B., ARULAMPALAM, S., AND GORDON, N. *Beyond the Kalman Filter, Particle Filters for Tracking Applications*. Artech House, Boston, London, 2004.

[35] SAVAGE, P. Strapdown inertial navigation integration algorithm design. *Journal of Guidance, Control and Dynamics 21*, 1-2 (1998).

[36] SAVAGE, P. G. Strapdown inertial navigation. Lecture Notes, 1997.

[37] STARK, H., AND WOODS, J. W. *Probability, random processes, and estimation theory for engineers*. Prentice Hall, Englewood Cliffs, 1994.

[38] SYRJÄRINNE, J. Studies of modern techniques for personal positioning. Phd thesis, Tampere University of Technology, March 2001.

[39] TEUNISSEN, P. A new method for fast carrier phase ambiguity estimation. In *IEEE Position Location and Navigation Symposium* (April 1994), pp. 562–573. http://ieeexplore.ieee.org/.

[40] TITTERTON, D. H. *Strapdown Inertial Navigation Technology*. IEE, 2004.

[41] VAUHKONEN, M., KARJALAINEN, P., AND TOSSAVAINEN, O.-P. Estimointiteoria. Lecture notes, University of Kuopio, 2006. http://venda.uku.fi/studies/kurssit/ETE/.

[42] WALPOLE, R. E., AND MYERS, R. H. *Probability and Statistics for Engineers and Scientists*, 4th ed. Macmillan Publishing Company, 1980.

[43] WGS84. Department of defence world geodetic system 1984. Tech. rep., National Imagery and Mapping Agency, 2000. ftp://164.214.2.65/pub/gig/tr8350.2/wgs84fin.pdf.

[44] WILLIAMS, D. *Probability with Martingales*, sixth ed. Cambridge University Press, Cambridge University, 1991.