



UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

Numerical integration for probabilistic reasoning skripsie

by

Jacobus Martin Louw



Bachelor of Engineering (Electrical and Electronic)

Final Report

Study leader: Dr CE van Daalen

2018

Declaration

By submitting this report electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date:

Copyright © 2018 Stellenbosch University
All rights reserved.

Contents

List of Figures	iii
List of Acronyms	iv
Abstract	v
1 Introduction	1
2 Gaussian Random Variables	2
2.1 Covariance form	2
2.1.1 Univariate Gaussians	2
2.1.2 Multivariate Gaussians	3
2.1.3 Error ellipse	4
2.2 Canonical form	5
2.2.1 Operations using the canonical form	6
2.2.2 Using the canonical form the present conditional distributions (linear transformations)	8
3 Localisation using Kalman Filters	10
3.1 Kalman Filter	10
3.2 Motion models	10
3.3 Measurement models	10
3.4 Extended Kalman Filter	10
3.5 Unscented Kalman Filter	10
4 Probabilistic Graphical Models	11
5 Localisation using Probabilistic Graphical Models	12
5.1 Linear localisation	12
5.2 Non-linear localisation	12
5.2.1 Taylor series expansion	12
5.2.2 Unscented transform	12
5.2.3 Monte Carlo integration	12
List of References	13

List of Figures

2.1	Univariate Gaussian Distribution PDF	3
2.2	Error Ellipse	5

List of Acronyms

RV - random variable

PGM - probabilistic graphical model

PDF - probability density function

KF - Kalman Filter

EKF - Extended Kalman Filter

UKF - Unscented Kalman Filter

Abstract

Text in default language ...

Chapter 1

Introduction

For modern mobile vehicles and robots it is important for them to have the ability to navigate their environment. It is critical for these devices to avoid collisions or dangerous environments. Some robots must move very precisely and therefore should have an accurate reading of their location.

Localisation is essential for robot navigation, but is used in various other applications such as missile tracking. For applications like this, it is crucial to have accurate and instantaneous information of the missile's location. Measurements of any object's location will always have some noise, the measured location is thus never 100% accurate. Therefore, one should rather approach the localisation problem in a probabilistic manner and the object's location can be described in terms of a PDF.

For systems with continuous random variables, most of the operations used in probabilistic reasoning use integration. These integrals can be solved analytically in the case of a problem with linear movement. Most systems are nonlinear. The integrals in nonlinear systems cannot be computed analytically and one has to use numerical methods. Commonly-used techniques such as the extended or unscented Kalman filters use primitive numerical integration that are very inaccurate, especially when measurements are also nonlinear. There are several numerical techniques available that are more accurate.

The end goal of this project is to investigate different numerical techniques to solve the nonlinear localisation problem. To reach the end goal, one should have a good understanding of Gaussian random variables, motion and measurement models, and traditional techniques such as the extended - and unscented kalman filters. Also, modeling the problem with a Probabilistic Graphical Model has advantages and is therefore investigated.

A relevant problem has been simulated in Python. Different techniques were implemented and compared in terms of accuracy and efficiency.

Chapter 2

Gaussian Random Variables

The Gaussian or normal distribution is commonly used in probability theory, as it is easy to work with and are remarkably close to real-world distributions. Gaussians make severe assumptions such as exponential decay of the distribution away from its mean and linear interactions between variables. Gaussians are in many cases a good estimation for real world distributions, such as noise, even though they make invalid assumptions. The Gaussian random variable (RV) is a very important concept in this paper as all probability distributions are approximated as Gaussian distributions. Key concepts and features of the Gaussian RV are discussed in this chapter. The canonical form and conditional distributions are very important concepts and will be used in following chapters. The following chapter is based on the work of Peebles and Shi [1] and Koller and Friedman [2].

2.1 Covariance form

2.1.1 Univariate Gaussians

An univariate Gaussian random variable X with mean μ and variance σ^2 is described by

$$p(x) = \mathcal{N}(\mu; \sigma) = \eta \exp \left[\frac{-(x - \mu)^2}{2\sigma^2} \right] \quad (2.1)$$

with normalisation coefficient

$$\eta = \frac{1}{\sqrt{2\pi\sigma^2}} \quad (2.2)$$

The mean parameter μ describes the location of the peak of the distribution and the variance parameter σ^2 describes the tempo which the distribution decays. The standard deviation is denoted by σ . The probability of x decays exponentially as x moves farther from the mean. Figure 2.1 indicates the mean μ and standard deviation σ of an univariate Gaussian PDF.

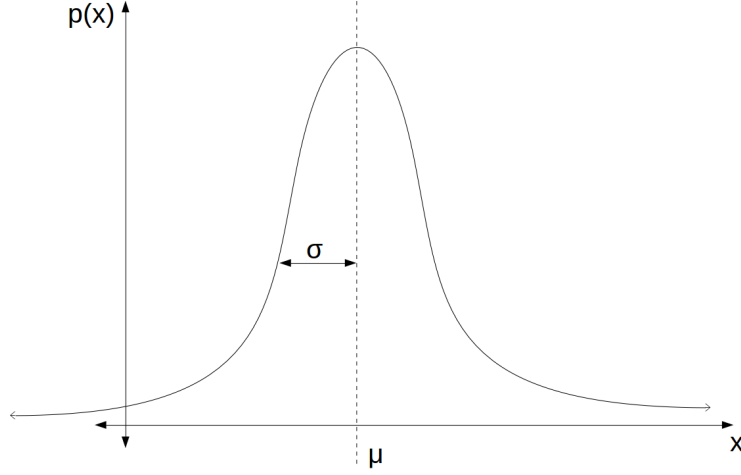


Figure 2.1: Univariate Gaussian Distribution PDF

2.1.2 Multivariate Gaussians

The multivariate Gaussian distribution is described by an n -dimensional mean vector $\boldsymbol{\mu}$ and an $n \times n$ covariance matrix Σ . The multivariate Gaussian distribution is commonly described by the following function

$$p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}; \Sigma) = \eta \exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right] \quad (2.3)$$

with normalisation coefficient

$$\eta = \frac{1}{(2\pi^{n/2})|\Sigma|^{\frac{1}{2}}} \quad (2.4)$$

$|\Sigma|$ is the determinant of Σ . It is important that the determinant of Σ is nonzero for the equation above to be valid. It is a requirement that Σ is positive definite and therefore also non-singular, which guarantees a determinant that is nonzero. In this paper we focus on positive definite covariance matrices, as it is invertible, consequentially one can use the alternative canonical or information parameterization. The canonical form is discussed in the following section.

The mean and covariance can sequentially be found by calculating the first and second moments of the Gaussian distribution.

The mean is equal to the first moment:

$$\boldsymbol{\mu} = \mathbf{E}[\mathbf{X}] \quad (2.5)$$

The covariance is equal to the second moment:

$$\Sigma = \mathbf{E}[\mathbf{X}\mathbf{X}^T] - \mathbf{E}[\mathbf{X}]\mathbf{E}[\mathbf{X}]^T \quad (2.6)$$

2.1.3 Error ellipse

Multivariate Gaussian distributions can be visualized as a series of ellipsoidal contours around the mean vector $\boldsymbol{\mu}$. The contours are parallel to each other and the distance between contours suggests the steepness of the density function. The steepness of the density function is determined by the covariance matrix Σ .

In the following chapters it is important to show the uncertainty of PDFs, a plot of an error ellipses is an effective way to illustrate this.

Eigenvalue decomposition

The following section is based on an article from H Abdi [3].

In order to plot an error ellipse, the directions the covariance varies the most and the variances in those directions are needed.

We can use eigenvalue decomposition to factorize the covariance matrix Σ in terms of its eigenvalues Λ and its eigenvectors Q .

$$\Sigma = Q\Lambda Q^{-1} \quad (2.7)$$

Each column of the eigenvector matrix Q contains an eigenvector \mathbf{v} . The eigenvectors indicate the direction which the covariance varies the most. Each eigenvector has a corresponding eigenvalue λ which specifies the variance in the direction of the eigenvector.

Plotting a 2D error ellipse

The following section is based on a webpage from "Computer vision for dummies" [4].

Error ellipses are specified in terms of the confidence or probability that a random data point will fall inside the ellipse.

To plot an error ellipse, the major - and minor axes' lengths are sequentially specified as $2\sqrt{s\lambda_1}$ and $2\sqrt{s\lambda_2}$, where λ_1 and λ_2 represent the eigenvalues of the covariance matrix ($\lambda_1 \geq \lambda_2$). For a 95% confident ellipse s is set to 5.991. If we are interested in other confidence intervals, the Chi-Square distribution can be used to find an s that less than or equal to a specific value which correlates with the probability of the confidence interval.

To obtain the orientation α of the error ellipse, we calculate the angle of the largest eigenvector relative to the x-axis.

$$\alpha = \arctan \left(\frac{\mathbf{v}_1(y)}{\mathbf{v}_1(x)} \right) \quad (2.8)$$

In the case of an uncorrelated covariance, the eigenvalues are equal to the variances of the covariance matrix and the eigenvectors are in the same direction as the x-axis and y-axis. Figure 2.2 shows α and indicates the major and minor axes of the error ellipse.

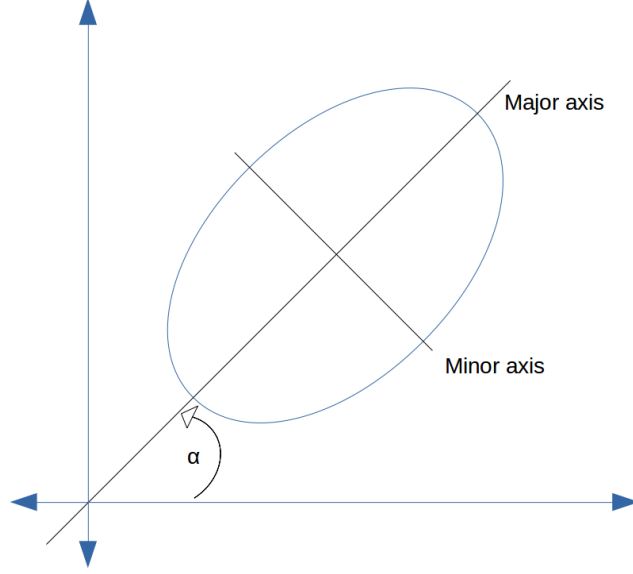


Figure 2.2: Error Ellipse

2.2 Canonical form

Linear Gaussian CPDs (conditional probability densities) are generally not Gaussians. It will be handy to find general representation that accommodates both Gaussian distributions and linear Gaussian models. The canonical or information form is a viable option. It is also much easier to perform certain operations in the canonical form. The following section is based on the work of Koller and Friedman [2] and JC Schoeman [5].

$$\mathcal{C}(\mathbf{X}; K, \mathbf{h}, g) = \exp \left(-\frac{1}{2} \mathbf{X}^T K \mathbf{X} + \mathbf{h}^T \mathbf{X} + g \right) \quad (2.9)$$

It is possible to represent every Gaussian as a canonical form. Equation 2.3 can be rewritten:

$$\begin{aligned} \eta \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \\ = \exp \left(-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} + \boldsymbol{\mu}^T \Sigma^{-1} \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} + \ln \eta \right) \end{aligned} \quad (2.10)$$

$\mathcal{N}(\boldsymbol{\mu}; \Sigma) = \mathcal{C}(K, \mathbf{h}, g)$ by comparing 2.10 with 2.9

$$K = \Sigma^{-1} \quad (2.11)$$

$$\mathbf{h} = \Sigma^{-1} \boldsymbol{\mu} \quad (2.12)$$

$$g = -\frac{1}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} + \ln \eta \quad (2.13)$$

\mathbf{h} is called the potential vector and K is called the information matrix.

The covariance parameters can again be recovered

$$\Sigma = K^{-1} \quad (2.14)$$

$$\boldsymbol{\mu} = \Sigma \mathbf{h} \quad (2.15)$$

The covariance form is not defined when K is not invertible, even though the distribution can still be represented in the canonical form. The canonical form is only a valid Gaussian density if and only if K (information matrix) is symmetric and positive definite. The canonical form is therefore more general than the covariance form. The canonical form is very useful to present linear Gaussian CPDs. From 2.13 it can be seen that only K and \mathbf{h} are necessary to calculate g . Thus, g can be omitted when working in the canonical form.

2.2.1 Operations using the canonical form

The main advantage of using the canonical form is that it is very easy to perform various operations on distributions. The variable g is included in the following operations for completeness sake, but can be omitted as stated above. Extending and rearranging scopes of canonical forms may be necessary before doing operations, as it is important that scopes of canonical forms are identical when doing operations.

Extending the scope of a canonical form:

If additional variables are needed, the scope of a canonical form can be extended by adding zero entries to K , \mathbf{h} and g .

$$\begin{aligned} \mathcal{C} \left(\begin{bmatrix} x \\ y \end{bmatrix} : \begin{bmatrix} k_{xx} & k_{xy} \\ k_{yx} & k_{yy} \end{bmatrix}, \begin{bmatrix} h_x \\ h_y \end{bmatrix}, \begin{bmatrix} g_x \\ g_y \end{bmatrix} \right) \\ = \mathcal{C} \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} : \begin{bmatrix} k_{xx} & k_{xy} & 0 \\ k_{yx} & k_{yy} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} h_x \\ h_y \\ 0 \end{bmatrix}, \begin{bmatrix} g_x \\ g_y \\ 0 \end{bmatrix} \right) \end{aligned} \quad (2.16)$$

Rearranging the scope of a canonical form:

The order of the scope can be rearranged by rearranging the rows and columns of K and rearranging the entries of \mathbf{h} and g .

$$\begin{aligned} \mathcal{C} \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} : \begin{bmatrix} k_{xx} & k_{xy} & k_{xz} \\ k_{yx} & k_{yy} & k_{yz} \\ k_{zx} & k_{zy} & k_{zz} \end{bmatrix}, \begin{bmatrix} h_x \\ h_y \\ h_z \end{bmatrix}, \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} \right) \\ = \mathcal{C} \left(\begin{bmatrix} y \\ x \\ z \end{bmatrix} : \begin{bmatrix} k_{yy} & k_{yx} & k_{yz} \\ k_{xy} & k_{xx} & k_{xz} \\ k_{zy} & k_{zx} & k_{zz} \end{bmatrix}, \begin{bmatrix} h_y \\ h_x \\ h_z \end{bmatrix}, \begin{bmatrix} g_y \\ g_x \\ g_z \end{bmatrix} \right) \end{aligned} \quad (2.17)$$

Multiplication of canonical forms:

When multiplying two canonical forms, it is important that their scopes are identical. Multiplying canonical forms with the same scope is simply:

$$\mathcal{C}(K_1, \mathbf{h}_1, g_1) \times \mathcal{C}(K_2, \mathbf{h}_2, g_2) = \mathcal{C}(K_1 + K_2, \mathbf{h}_1 + \mathbf{h}_2, g_1 + g_2) \quad (2.18)$$

Division of canonical forms:

Again, it important that the scopes of the distributions are identical.

$$\frac{\mathcal{C}(K_1, \mathbf{h}_1, g_1)}{\mathcal{C}(K_2, \mathbf{h}_2, g_2)} = \mathcal{C}(K_1 - K_2, \mathbf{h}_1 - \mathbf{h}_2, g_1 - g_2) \quad (2.19)$$

Marginalization of a canonical form:

A marginal distribution can be found by integrating over a subset of variables. For example the marginal distribution over \mathbf{X} can be found by integrating over \mathbf{Y} .

Let $\mathcal{C}(\mathbf{X}, \mathbf{Y}; K, \mathbf{h}, g)$ be a canonical form with subsets \mathbf{X} and \mathbf{Y} where

$$K = \begin{bmatrix} K_{\mathbf{X}\mathbf{X}} & K_{\mathbf{X}\mathbf{Y}} \\ K_{\mathbf{Y}\mathbf{X}} & K_{\mathbf{Y}\mathbf{Y}} \end{bmatrix}; \mathbf{h} = \begin{pmatrix} h_{\mathbf{X}} \\ h_{\mathbf{Y}} \end{pmatrix} \quad (2.20)$$

To obtain the marginal distribution over \mathbf{X} , we have to find the integral over \mathbf{Y} . Therefore

$$\int \mathcal{C}(\mathbf{X}, \mathbf{Y}; K, \mathbf{h}, g) d\mathbf{Y} = \mathcal{C}(\mathbf{X}; K', \mathbf{h}', g') \quad (2.21)$$

where

$$K' = K_{\mathbf{X}\mathbf{X}} - K_{\mathbf{X}\mathbf{Y}} K_{\mathbf{Y}\mathbf{Y}}^{-1} K_{\mathbf{Y}\mathbf{X}} \quad (2.22)$$

$$\mathbf{h}' = \mathbf{h}_{\mathbf{X}} - K_{\mathbf{X}\mathbf{X}} K_{\mathbf{Y}\mathbf{Y}}^{-1} \mathbf{h}_{\mathbf{Y}} \quad (2.23)$$

$$g' = g - \frac{1}{2} (\ln |2\pi K_{\mathbf{Y}\mathbf{Y}}^{-1}| + \mathbf{h}_{\mathbf{Y}}^T K_{\mathbf{Y}\mathbf{Y}}^{-1} \mathbf{h}_{\mathbf{Y}}) \quad (2.24)$$

It is important to that $K_{\mathbf{Y}\mathbf{Y}}$ is positive-definite for the result to be finite.

Reduction with evidence:

Let $\mathcal{C}(\mathbf{X}, \mathbf{Y}; K, \mathbf{h}, g)$ be a canonical form with subsets \mathbf{X} and \mathbf{Y} . If subset \mathbf{Y} is known ($\mathbf{Y} = \mathbf{y}$), then the canonical form can be reduced to $\mathcal{C}(\mathbf{X}; K', \mathbf{h}', g')$, where

$$K' = K_{\mathbf{X}\mathbf{X}} \quad (2.25)$$

$$\mathbf{h}' = \mathbf{h}_{\mathbf{X}} - K_{\mathbf{X}\mathbf{Y}}\mathbf{y} \quad (2.26)$$

$$g' = g + \mathbf{h}_{\mathbf{Y}}^T \mathbf{y} - \frac{1}{2} \mathbf{y}^T K_{\mathbf{Y}\mathbf{Y}} \mathbf{y} \quad (2.27)$$

2.2.2 Using the canonical form the present conditional distributions (linear transformations)

The following section is based on a skripsie from JC Schoeman [5] and notes from Dr CE van Daalen. One of the advantages of the canonical form is that it is possible to present conditional probability distributions (CPDs).

If a CPD such as $p(\mathbf{y}|\mathbf{x})$ can be described as a linear transform:

$$\mathbf{y} = F\mathbf{x} + \mathbf{g} + \mathbf{n} \quad (2.28)$$

where

$$\mathbf{n} = \mathcal{N}(\mathbf{0}, \Sigma_n) \quad (2.29)$$

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \mathcal{N}(F\mathbf{x} + \mathbf{g}, \Sigma) \\ &= \eta \exp \left[-\frac{1}{2} (\mathbf{y} - (F\mathbf{x} + \mathbf{g}))^T \Sigma^{-1} (\mathbf{y} - (F\mathbf{x} + \mathbf{g})) \right] \end{aligned} \quad (2.30)$$

Rewriting

$$\begin{aligned} \mathbf{y} - (F\mathbf{x} + \mathbf{g}) &= \begin{bmatrix} I & -F & -I \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix} \\ &= \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix}^T \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix} \end{aligned} \quad (2.31)$$

Equation 2.30 becomes

$$p(\mathbf{y}|\mathbf{x}) = \eta \exp \left[-\frac{1}{2} \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix}^T \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix} \Sigma^{-1} \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix}^T \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix} \right] \quad (2.32)$$

Defining a combined random vector

$$\mathbf{w} = \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix} \quad (2.33)$$

and information matrix

$$K' = \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix} \Sigma^{-1} \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix}^T \quad (2.34)$$

Equation 2.32 can be rewritten

$$p(\mathbf{y}|\mathbf{x}) = \eta \exp \left[-\frac{1}{2} \mathbf{w}^T K' \mathbf{w} \right] \quad (2.35)$$

This closely relates with the canonical form in equation 2.9

$$\begin{aligned} p(y|x) &= \mathcal{C}(\mathbf{w} : K', \mathbf{0}) \\ &= \mathcal{C} \left(\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix} ; \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix} \Sigma_n^{-1} \begin{bmatrix} I & -F & -I \end{bmatrix}, \mathbf{0}, - \right) \\ &= \mathcal{C} \left(\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} ; \begin{bmatrix} I \\ -F^T \end{bmatrix} \Sigma_n^{-1} \begin{bmatrix} I & -F \end{bmatrix}, \begin{bmatrix} I \\ -F^T \end{bmatrix} \Sigma_n^{-1} \mathbf{g}, - \right) \\ &= \mathcal{C} \left(\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} ; \begin{bmatrix} \Sigma_n^{-1} & -\Sigma_n^{-1} F \\ -F^T \Sigma_n^{-1} & F^T \Sigma_n^{-1} F \end{bmatrix}, \begin{bmatrix} -\Sigma_n^{-1} \mathbf{g} \\ -F^T \Sigma_n^{-1} \mathbf{g} \end{bmatrix}, - \right) \end{aligned} \quad (2.36)$$

Chapter 3

Localisation using Kalman Filters

3.1 Kalman Filter

3.2 Motion models

3.3 Measurement models

3.4 Extended Kalman Filter

3.5 Unscented Kalman Filter

Chapter 4

Probabilistic Graphical Models

Chapter 5

Localisation using Probabilistic Graphical Models

5.1 Linear localisation

5.2 Non-linear localisation

5.2.1 Taylor series expansion

5.2.2 Unscented transform

5.2.3 Monte Carlo integration

List of References

- [1] Peebles, P.Z.: *Probability, random variables, and random signal principles*, vol. 3. McGraw-Hill New York, NY, USA:, 2001.
- [2] Koller, D., Friedman, N. and Bach, F.: *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [3] Abdi, H.: The eigen-decomposition: Eigenvalues and eigenvectors. *Encyclopedia of measurement and statistics*, pp. 304–308, 2007.
- [4] Visiondummy, how to draw a covariance error ellipse. <http://www.visiondummy.com/2014/04/draw-error-ellipse-representing-covariance-matrix/>. Accessed: 2018-10-01.
- [5] Schoeman, J.: *Simultaneous Localisation and Mapping of a Robotic Vehicle using a Probabilistic Graphical Model*. Stellenbosch University, 2016.
- [6] Choset, H.M., Hutchinson, S., Lynch, K.M., Kantor, G., Burgard, W., Kavraki, L.E. and Thrun, S.: *Principles of robot motion: theory, algorithms, and implementation*. MIT press, 2005.