

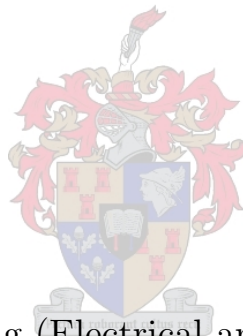


UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

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	iv
List of Acronyms	v
Abstract	vi
1 Introduction	1
2 Gaussian Random Variables	2
2.1 Covariance form	2
2.1.1 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)	7
3 Motion models	9
3.1 Basic concepts	9
3.2 Linear Motion	9
3.3 Nonlinear motion	9
3.3.1 Velocity motion model	9
4 Localisation using Kalman Filters	10
4.1 Motion models	10
4.2 Measurement models	10
4.3 Kalman Filter	10
4.4 Extended Kalman Filter	10
4.5 Unscented Kalman Filter	10
5 Probabilistic Graphical Models	11
5.1 Overview	11
5.2 Basic concepts of graphical models	11
5.3 Bayesian PGMs	12
5.4 Cluster graphs	13
5.5 Message Passing	15
5.6 Summary	16
6 Localisation using Probabilistic Graphical Models	17
6.1 Linear localisation	17
6.2 Non-linear localisation	17

6.2.1	Taylor series expansion	17
6.2.2	Unscented transform	17
6.2.3	Monte Carlo integration	17
List of References		18

List of Figures

2.1	Univariate Gaussian PDF	3
2.2	Error Ellipse	5
5.1	12
5.2	Bayesian PGM	13
5.3	Bayesian PGM with cluster boundaries	14
5.4	Cluster graph	14

List of Acronyms

RV - random variable
PGM - probabilistic graphical model
PDF - probability density function
CPD - conditional probability distribution
KF - Kalman Filter
EKF - Extended Kalman Filter
UKF - Unscented Kalman Filter

Abstract

Text in default language ...

Chapter 1

Introduction

Robot localisation is the process of estimating where a robot is located relative to its environment. It is essential for autonomous robots to have precise knowledge of their location in order to navigate their surroundings.

When localising a robot, the control input to the robot, and a map of the environment is usually available. Further the robot is generally fitted with sensors which, together with beacons, measure where the robot is located relative to the map. Unfortunately the information of the robot's movements and location always have noise to some degree.

The position and orientation of the robot should rather be estimated in a probabilistic manner from the information gathered from the sensors. The localisation techniques must therefore be able to deal with noise and describe the robot's location with a measure of uncertainty.

Probabilistic reasoning of systems with continuous random variables (RV) use integration for various operations. In most nonlinear systems these integrals cannot be calculated analytically and one must use numerical methods. Kalman filters have been used extensively to localise robots and other objects. Techniques such as the extended or unscented Kalman filters use primitive numerical integration that is very inaccurate, especially when measurements are also nonlinear. However, there are several alternative numerical techniques available that are more accurate.

A probabilistic graphical model (PGM) is a technique that is be used to model systems with uncertainty in an organised manner. It describes the relationship between random variables (RV) and allows to reason about their likelihood based on knowledge that is available to the system. Reasoning about RVs in a PGM consists out of concrete steps and one can easily identify where integration is used.

The end goal of this project is investigate different numerical techniques of integration and to implement them in a PGM to localise a simulated robot. The accuracy and efficiency of different techniques of numerical integration when used to inference a nonlinear system will be compared using a simulated environment.

Chapter 2

Gaussian Random Variables

The Gaussian or normal distribution is commonly used in probability theory, as it is very easy to work with. Although Gaussian distributions make severe assumptions such as exponential decay of the distribution away from its mean and linear interactions between variables, they are often surprisingly good estimations for real world distributions, such as noise. The Gaussian distribution is a very important concept in this report as all probability distributions are approximated as Gaussian distributions. Key concepts and features of the Gaussian distributions 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

The likelihood of the values of a continuous Gaussian RV can be described with a Gaussian distribution. The univariate Gaussian distribution is probability density function (PDF) of a single Gaussian RV. The covariance form fully describes a univariate Gaussian distribution by a mean μ and variance σ^2 . The univariate Gaussian distribution of a RV X , denoted

$$X \sim \mathcal{N}(\mu, \sigma), \quad (2.1)$$

has a PDF that is described by

$$p(x) = \eta \exp \left[\frac{-(x - \mu)^2}{2\sigma^2} \right], \quad (2.2)$$

with a normalisation coefficient

$$\eta = \frac{1}{\sqrt{2\pi\sigma^2}}. \quad (2.3)$$

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 probability of RV X having a value x decays exponentially as x moves farther from the mean. The standard deviation of the Gaussian distribution is denoted as σ .

If X is a continuous RV, then the mean or expectation of X is

$$\mu = E[X] = \int x \cdot p(x) dx. \quad (2.4)$$

The variance of X can be calculated as

$$\sigma^2 = E[X^2] - (E[X])^2 \quad (2.5)$$

Figure 2.1 indicates the mean μ and standard deviation σ of an univariate Gaussian PDF.

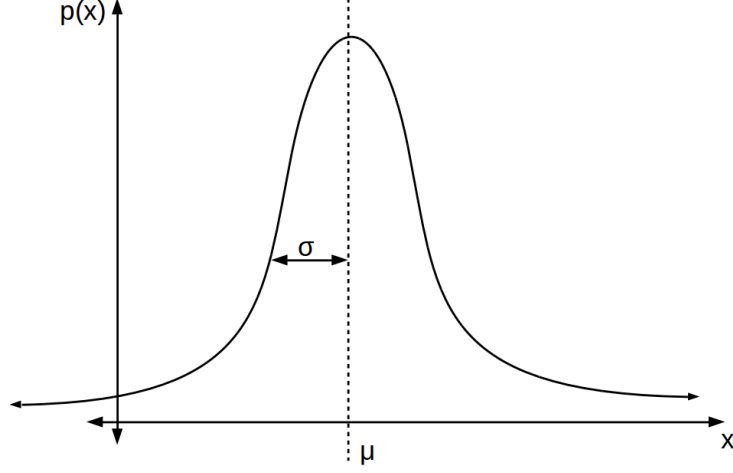


Figure 2.1: Univariate Gaussian PDF

The definition of the Gaussian distribution can be extended to describe the PDF of N random variables, this is called the multivariate Gaussian distribution. 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 of a RV vector \mathbf{x} with length of N , $\mathbf{X} = [X_1, \dots, X_N]$, denoted

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \quad (2.6)$$

has a PDF that is described by

$$p(\mathbf{x}) = \eta \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad (2.7)$$

with a normalisation coefficient

$$\eta = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}}. \quad (2.8)$$

$|\Sigma|$ is the the determinant of Σ .

For this report it is required that Σ is positive definite and thus also non-singular, which guarantees a determinant that is nonzero. Positive definite covariance matrices are also invertible, thus the alternative canonical or information parameterization, which requires Σ^{-1} , can be used. The canonical form is discussed in the following section.

The mean vector of RV vector \mathbf{X} is equal to the expectation of \mathbf{X} and is defined as

$$\boldsymbol{\mu} = E[\mathbf{X}]. \quad (2.9)$$

The covariance matrix Σ specifies the correlation between RVs and is equal to

$$\Sigma = E[\mathbf{X}\mathbf{X}^T] - E[\mathbf{X}]E[\mathbf{X}]^T. \quad (2.10)$$

2.1.1 Error ellipse

The following section is based on an article by H. Abdi [3] and on a webpage from "Computer vision for dummies" [4].

A multivariate Gaussian distribution with RV vector

$$\mathbf{X} = [X_1, X_2] \quad (2.11)$$

can be visualised as a series of ellipsoidal contours around the mean vector $\boldsymbol{\mu}$. The contours are an indication of the correlation between X_1 and X_2 and also show the uncertainty of the PDF. Contours close to each other suggest a steep incline and contours farther apart suggest a gradual incline in the PDF. Hence, small and concentrated contours represent certain Gaussian distributions and contours that are larger and further apart represent uncertain Gaussian distributions. Error ellipses are an effective way to indicate the mean, uncertainty and correlation of 2D Gaussian distributions.

An ellipse has a major axis and a minor axis as shown in Figure 2.2. The major axis of the error ellipse is always aligned in the direction the Gaussian distribution varies most. This direction can be found by determining the eigenvectors of the distribution. Each eigenvector has a corresponding eigenvalue which indicates the spread of the distribution in the direction of the eigenvector.

We can use eigenvalue decomposition to factorise the covariance matrix Σ ,

$$\Sigma = Q\Lambda Q^{-1}. \quad (2.12)$$

Each column of the eigenvector matrix Q contains an eigenvector \mathbf{v} . Q is defined as

$$Q = [\mathbf{v}_1 \ \mathbf{v}_2] = \begin{bmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{bmatrix}. \quad (2.13)$$

Λ is a diagonal matrix and each of its diagonal entries contains an eigenvalue λ . Λ is defined as

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}. \quad (2.14)$$

An error ellipse of a Gaussian PDF can be specified in terms of confidence levels which is the probability that a value drawn from the distribution will fall inside the ellipse. Thus, differently sized ellipses can be plotted for different confidence levels. The lengths of the major and minor axes are sequentially specified as $2\sqrt{s\lambda_1}$ and $2\sqrt{s\lambda_2}$, where $\lambda_1 \geq \lambda_2$. The value of s is determined by the confidence level of the error ellipse. In the case of an error ellipse with a confidence level of 95%, s is set to 5.991. The Chi-Square distribution can be used to find s for other confidence levels, but it is beyond the scope of this document.

The orientation α of the error ellipse is shown in Figure 2.2. To obtain α we calculate the angle, relative to the x-axis, of the eigenvector with the largest spread. The angle α is defined as

$$\alpha = \arctan\left(\frac{v_{11}}{v_{12}}\right), \quad (2.15)$$

where $\lambda_1 \geq \lambda_2$.

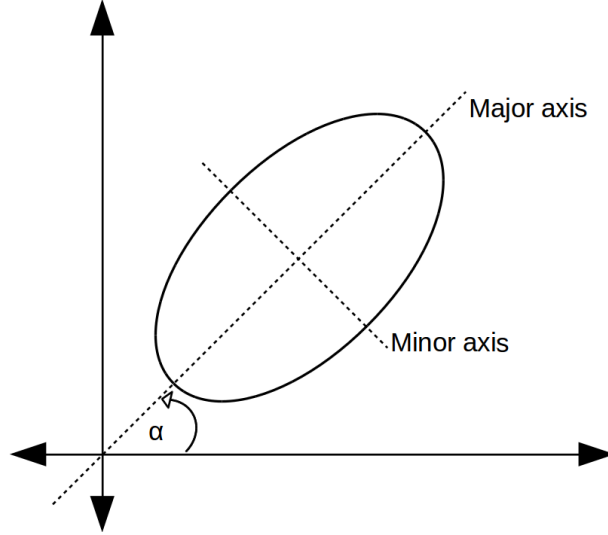


Figure 2.2: Error Ellipse

2.2 Canonical form

The canonical or information form is an alternative way to represent Gaussian distributions. A big advantage of using the canonical form is that it is much easier to perform certain operations. It can also be used to represent Gaussian conditional probability densities (CPDs), which will be discussed later in this section. The following section is based on the work of Koller and Friedman [2] and Schoeman [5].

Equation 2.7, which is the definition of a Gaussian PDF, can be rewritten

$$\begin{aligned} p(\mathbf{x}) &= \eta \exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right] \\ &= \exp \left(-\frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{x} + \boldsymbol{\mu}^T \Sigma^{-1} \mathbf{x} - \frac{1}{2}\boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} + \ln \eta \right), \end{aligned} \quad (2.16)$$

with the information matrix

$$K = \Sigma^{-1}, \quad (2.17)$$

the potential vector

$$\mathbf{h} = \Sigma^{-1} \boldsymbol{\mu}, \quad (2.18)$$

and

$$g = -\frac{1}{2}\boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} + \ln \eta. \quad (2.19)$$

Thus the definition of the canonical form is

$$\mathcal{C}(\mathbf{x}; K, \mathbf{h}, g) = \exp \left(-\frac{1}{2}\mathbf{x}^T K \mathbf{x} + \mathbf{h}^T \mathbf{x} + g \right). \quad (2.20)$$

As one can always determine g from \mathbf{h} and K , thus a more general notation for a Gaussian distribution in canonical form is $\mathcal{C}(\mathbf{x}; K, \mathbf{h})$.

It is easy to recover the covariance parameters:

$$\Sigma = K^{-1} \quad (2.21)$$

$$\boldsymbol{\mu} = \boldsymbol{\Sigma} \mathbf{h} \quad (2.22)$$

Sometimes K is not invertible and the covariance can not be recovered, the canonical form is therefore more general than the covariance form.

It can be seen from the above that it is very easy to switch between the canonical and covariance form.

2.2.1 Operations using the canonical form

A big advantage of using the canonical form is that it makes it easy to perform various operations on Gaussian distributions. This section will cover useful operations using the canonical form.

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:

The scope of a canonical form can be extended by adding zero entries to K and \mathbf{h} :

$$\mathcal{C} \left(\begin{bmatrix} x \\ y \end{bmatrix}; \begin{bmatrix} k_{xx} & k_{xy} \\ k_{yx} & k_{yy} \end{bmatrix}, \begin{bmatrix} h_x \\ h_y \end{bmatrix} \right) = \mathcal{C} \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}; \begin{bmatrix} k_{xx} & k_{xy} & 0 \\ k_{yx} & k_{yy} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} h_x \\ h_y \\ 0 \end{bmatrix} \right) \quad (2.23)$$

Rearranging the scope of a canonical form:

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

$$\mathcal{C} \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}; \begin{bmatrix} k_{xx} & k_{xy} & k_{xz} \\ k_{yx} & k_{yy} & k_{yz} \\ k_{zx} & k_{zy} & k_{zz} \end{bmatrix}, \begin{bmatrix} h_x \\ h_y \\ h_z \end{bmatrix} \right) = \mathcal{C} \left(\begin{bmatrix} y \\ x \\ z \end{bmatrix}; \begin{bmatrix} k_{yy} & k_{yx} & k_{yz} \\ k_{xy} & k_{xx} & k_{xz} \\ k_{zy} & k_{zx} & k_{zz} \end{bmatrix}, \begin{bmatrix} h_y \\ h_x \\ h_z \end{bmatrix} \right) \quad (2.24)$$

Multiplication of canonical forms:

If the scopes of two canonical forms are identical, one can multiply them by simply summing the K and \mathbf{h} parameters:

$$\mathcal{C}(K_1, \mathbf{h}_1, g_1) \times \mathcal{C}(K_2, \mathbf{h}_2, g_2) = \mathcal{C}(K_1 + K_2, \mathbf{h}_1 + \mathbf{h}_2, g_1 + g_2) \quad (2.25)$$

Marginalisation 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

$$K = \begin{bmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{bmatrix}; \quad \mathbf{h} = \begin{pmatrix} h_x \\ h_y \end{pmatrix}. \quad (2.26)$$

To obtain the marginal distribution over \mathbf{x} , we have to find the integral over \mathbf{y} . Therefore the marginal distribution over \mathbf{x} is

$$\int \mathcal{C}(\mathbf{X}, \mathbf{Y}; K, \mathbf{h}, g) d\mathbf{Y} = \mathcal{C}(\mathbf{X}; K', \mathbf{h}', g'), \quad (2.27)$$

where

$$K' = K_{xx} - K_{xy}K_{yy}^{-1}K_{yx}, \quad (2.28)$$

$$h' = h_x - K_{xx}K_{yy}^{-1}h_y, \quad (2.29)$$

and

$$g' = g - \frac{1}{2} (\ln |2\pi K_{yy}^{-1}| + h_y^T K_{yy}^{-1} h_y). \quad (2.30)$$

It is important to that K_{yy} 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

$$K' = K_{xx} \quad (2.31)$$

$$h' = h_x - K_{xy}y \quad (2.32)$$

$$g' = g + h_y^T y - \frac{1}{2} y^T K_{yy} y \quad (2.33)$$

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:

$$\mathbf{y} = F\mathbf{x} + \mathbf{g} + \mathbf{n} \quad (2.34)$$

where

$$\mathbf{n} = \mathcal{N}(\mathbf{0}, \Sigma_n) \quad (2.35)$$

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \mathcal{N}(F\mathbf{x} + \mathbf{g}, \Sigma) \\ &= \eta \exp \left[-\frac{1}{2} (\mathbf{y} - (F\mathbf{x} + \mathbf{g}))^T \Sigma^{-1} (\mathbf{y} - (F\mathbf{x} + \mathbf{g})) \right] \end{aligned} \quad (2.36)$$

Rewriting

$$\begin{aligned} \mathbf{y} - (F\mathbf{x} + \mathbf{g}) &= \begin{bmatrix} I & -F & -I \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix} \\ &= \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix}^T \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix} \end{aligned} \quad (2.37)$$

Equation 2.36 becomes

$$p(\mathbf{y}|\mathbf{x}) = \eta \exp \left[-\frac{1}{2} \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix}^T \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix} \Sigma^{-1} \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix}^T \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix} \right] \quad (2.38)$$

Defining a combined random vector

$$\mathbf{w} = \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix} \quad (2.39)$$

and information matrix

$$K' = \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix} \Sigma^{-1} \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix}^T \quad (2.40)$$

Equation 2.38 can be rewritten

$$p(\mathbf{y}|\mathbf{x}) = \eta \exp \left[-\frac{1}{2} \mathbf{w}^T K' \mathbf{w} \right] \quad (2.41)$$

This closely relates with the canonical form in equation 2.20

$$\begin{aligned} p(y|x) &= \mathcal{C}(\mathbf{w} : K', \mathbf{0}) \\ &= \mathcal{C} \left(\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \mathbf{g} \end{bmatrix} ; \begin{bmatrix} I \\ -F^T \\ -I \end{bmatrix} \Sigma_n^{-1} \begin{bmatrix} I & -F & -I \end{bmatrix}, \mathbf{0}, - \right) \\ &= \mathcal{C} \left(\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} ; \begin{bmatrix} I \\ -F^T \end{bmatrix} \Sigma_n^{-1} \begin{bmatrix} I & -F \end{bmatrix}, \begin{bmatrix} I \\ -F^T \end{bmatrix} \Sigma_n^{-1} \mathbf{g}, - \right) \\ &= \mathcal{C} \left(\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} ; \begin{bmatrix} \Sigma_n^{-1} & -\Sigma_n^{-1} F \\ -F^T \Sigma_n^{-1} & F^T \Sigma_n^{-1} F \end{bmatrix}, \begin{bmatrix} -\Sigma_n^{-1} \mathbf{g} \\ -F^T \Sigma_n^{-1} \mathbf{g} \end{bmatrix}, - \right) \end{aligned} \quad (2.42)$$

Chapter 3

Motion models

Before implementing a localisation algorithm to localise an object, it is important to have a model which describes the movement of the object. A simple linear motion model will be covered which will later be used to illustrate basic principles when localising linear moving objects. A more sophisticated nonlinear motion model will be explored which will be essential when investigating numerical integration techniques for localising nonlinear moving objects. This chapter is based on the work of S. Thrun, W. Burgard and D. Fox [6];

3.1 Basic concepts

The motion model describes $p(x_t|u_t, x_{t-1})$, which is the transition probability and it is vital in the prediction step of the Bayes filter. The material will cover kinematics in a stochastic form. Robot motion will be limited to planar movement as it is easier to convey concepts or ideas to the reader. Note that the theory discussed is not limited to planar movement.

3.2 Linear Motion

3.3 Nonlinear motion

3.3.1 Velocity motion model

Chapter 4

Localisation using Kalman Filters

This chapter will focus on localisation using traditional techniques such as the Kalman filter, extended Kalman filter and Unscented Kalman filter. This is important to investigate and understand as there is a strong correspondance to the PGM technique that will be discussed in a following section. The motion an measurement models will also be discussed

4.1 Motion models

4.2 Measurement models

4.3 Kalman Filter

4.4 Extended Kalman Filter

4.5 Unscented Kalman Filter

Chapter 5

Probabilistic Graphical Models

This chapter gives a brief introduction to Probabilistic Graphical Models (PGMs). PGMs are a very vast field which is beyond the scope of this document, therefore only the theory relevant to this paper will be covered. This chapter is based on the work of Koller and Friedman [2], Barber [7] and Schoeman [5].

5.1 Overview

Reasoning about systems with uncertainty can become very complex and sometimes it is completely unmanageable. The PGM is a graphical technique which allows to model probabilistic systems in a logical and compact manner which makes problems tractable. PGMs are thus used to describe the relationship between RVs and allows to reason about them. This action of querying a system or reasoning about a variable is called inference. The relationships in a PGM are usually in the form of conditional probability densities (CPDs). It is a very popular technique, as it is intuitive, transparent and easy to manipulate. Hence, PGMs have countless applications from making medical diagnosis to localising robots. Although it is a powerful technique, it has its limitations such as image recognition, where neural networks thrive.

PGMs can be divided in to two categories. The first is Markov networks, which are used for non-causal systems, and the second is Bayesian networks, which are used for causal systems. As the localisation problem is in most cases causal, the focus of this chapter is on Bayesian networks.

The theory is also discussed in terms of continuous RVs, because problems addressed in following chapters make use of continuous RVs. Note that all of the theory can also be applied to discrete RVs.

5.2 Basic concepts of graphical models

A graph consists of nodes and edges. Edges are the link between nodes and can be directed or undirected. If all edges are undirected then the graph is classified as undirected. If all edges are directed, then the graph is classified as directed. Bayesian networks are directed and Markov networks are undirected.

In the case of directed graphs, nodes can be classified as ancestors, parents, children and descendants. If there exists a directed path from $X_1 \rightarrow X_k$, then X_1 is an ancestor of X_k , and X_k is a descendant of X_1 . In Figure 5.1(b), a is an ancestor of c and c is a descendant

of a . A parent is an ancestor with only one edge between the ancestor and descendant. A child is a descendant with only one edge between the descendant and the ancestor. In Figure 5.1 (b), then a is a parent of b and b is a child of a . [7]

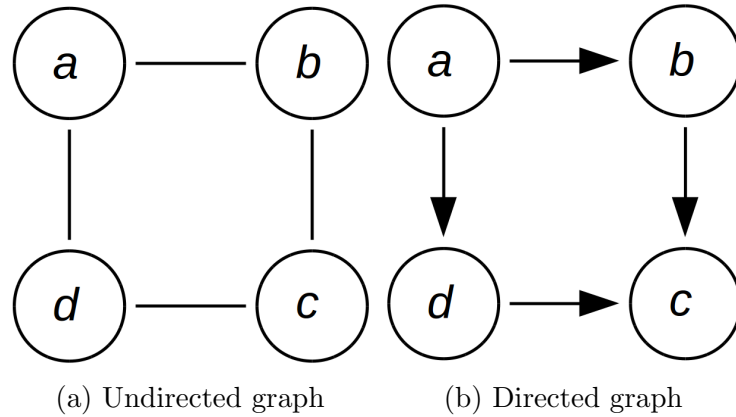


Figure 5.1

5.3 Bayesian PGMs

A system with uncertainty can be modeled with a Bayesian network. Bayesian networks consists out of a set of nodes and directed edges. Nodes represent RVs(X_1, \dots, X_n). Edges indicate the relationship between nodes.

Each node is associated with a conditional probability distribution.

$$X_i \sim p(X_i | Par(X_i)) \quad (5.1)$$

where $Par(X_i)$ indicates the parent of node X_i .

The CPD associated with each node can be written as a factor. A factor is a function that takes a number of random variables as arguments.

$$\phi_i(X_i, Par(X_i)) = p(X_i | Par(X_i)) \quad (5.2)$$

Where the scope of a factor is its' arguments

$$Scope\{\phi_i\} = \{X_i, Par(X_i)\} \quad (5.3)$$

Figure 5.2 is an example of a Bayesian network with seven nodes labeled from a to c . Each node is associated with a CPD. Directed edges between nodes are indicated with arrows.

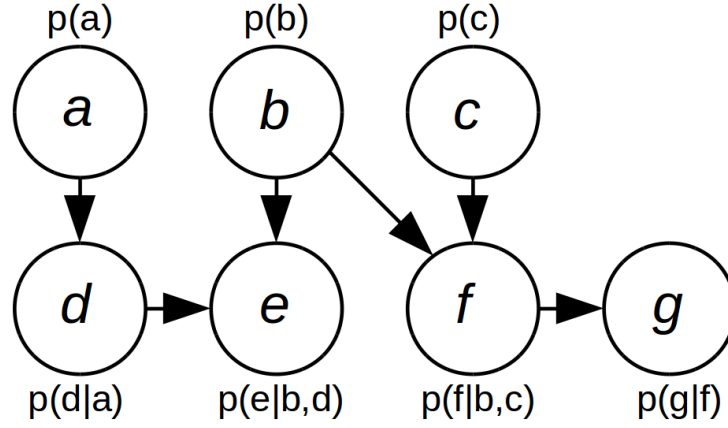


Figure 5.2: Bayesian PGM

Chain Rule for Bayesian Networks

The joint probability density distribution of all the variables in a Bayesian network can be found by finding the product of all the CPDs associated with nodes. (Koller and Friedman [2])

$$p(X_1, \dots, X_n) = \prod_i p(X_i | \text{Par}(X_i)) \quad (5.4)$$

Applying the Chain Rule to the Bayesian network in Figure 5.2

$$\begin{aligned} p(a, b, c, d, e, f, g) &= p(a)p(b)p(c)p(d|a)p(e|b, d)p(f|b, c)p(g|f) \\ &= \phi_a(a)\phi_b(b)\phi_c(c)\phi_d(d, a)\phi_e(e, b, d)\phi_f(f, b, c)\phi_g(g, f) \end{aligned} \quad (5.5)$$

A marginal PDF of a subset of variables can be found by integrating over the remaining variables.

$$p(a, b, c) = \int \int \int \int p(a, b, c, d, e, f, g) dd de df dg \quad (5.6)$$

5.4 Cluster graphs

After a Bayesian network has been constructed, it must be possible to reason about RVs in the network. Various methods can be used to inference a Bayesian network, one method is to construct a cluster graph. The cluster graph consists of clusters and undirected edges. Clusters are subsets of variables and can be formed by grouping factors together.

$$\mathbf{C}_i \subseteq \{X_1, \dots, X_n\} \quad (5.7)$$

The undirected edges between clusters are called sepsets and are responsible for passing messages between clusters. A sepset between two clusters contains information about variables that are common to both clusters.

$$\mathbf{S}_{i,j} \subseteq \mathbf{C}_i \cap \mathbf{C}_j \quad (5.8)$$

There are multiple ways to construct clusters, but it should adhere to two requirements specified Koller and Friedman [2].

1. Family Preservation

Every factor ϕ_k in a set of factors Φ should be accommodated by a cluster.

$$\text{Scope}[\phi_k] \subseteq \mathbf{C}_i \quad (5.9)$$

2. Running Intersection Property

There exists an unique path connecting a pair of clusters containing the same variable X , and every cluster and sepset along the path also contain X . This path allows clusters to share their beliefs of X . In other words, for any variable X , the set of sepsets and clusters containing X form a tree [2]. This prevents feedback loops and thus counters the phenomena where cluster's reinforce their own beliefs of variables.

Figure 5.2 is updated in Figure 5.3 where CPDs is replaced with factors and possible cluster boundaries are indicated with dashed rectangles. Note that there are other ways to construct the cluster graph.

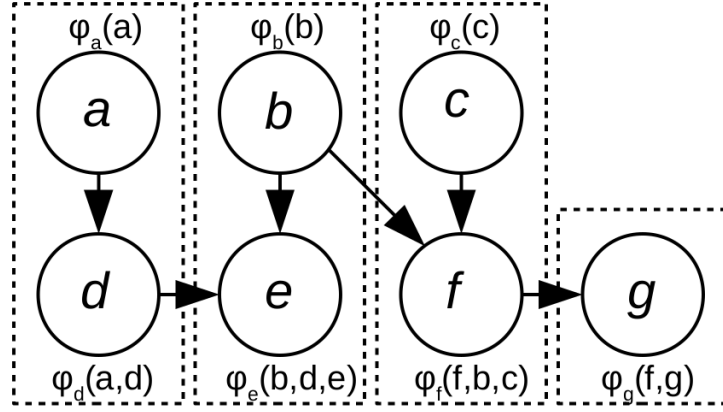


Figure 5.3: Bayesian PGM with cluster boundaries

The initial belief of a cluster can be calculated by finding the product of all the factors inside the cluster. [2]

$$\psi_i(C_i) = \prod_k \phi_k \quad (5.10)$$

Figure 5.3 is updated in Figure 5.4. Clusters are indicated by ellipsis. Initial beliefs are calculated and shown inside the ellipsis. Sepsets are indicated by rectangles.

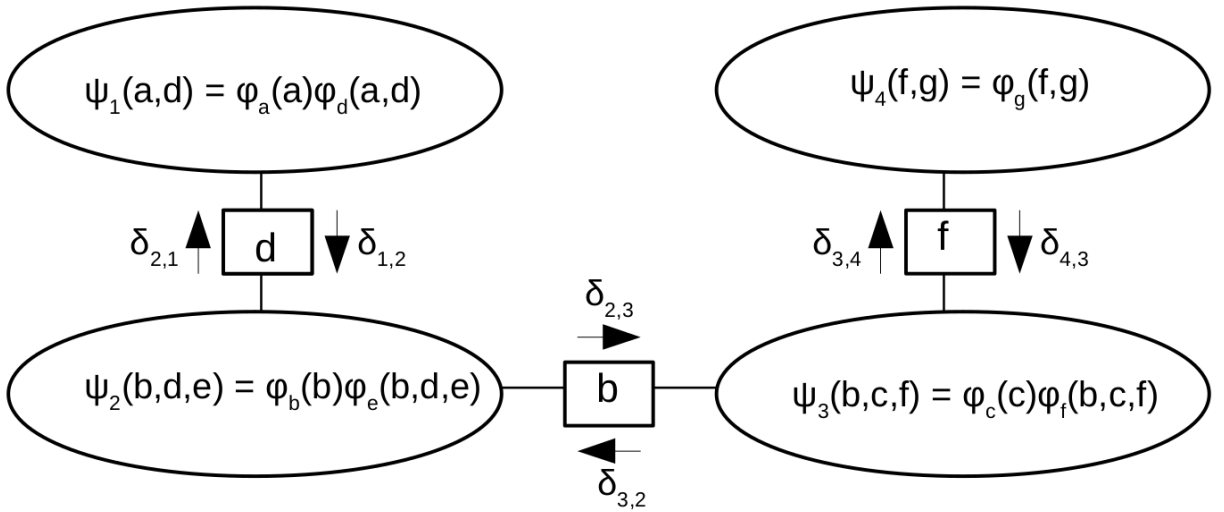


Figure 5.4: Cluster graph

5.5 Message Passing

Clusters can share beliefs of common variables with messages passing. Messages are associated with sepsets. An outgoing message of a cluster can be calculated by multiplying the cluster with all the incoming messages and then integrating over the variables that are not in the sepset.

Thus, messages can be calculated accordingly [2]

$$\delta_{i \rightarrow j}(\mathbf{S}_{i,j}) = \int_{\mathbf{C}_i - \mathbf{S}_{i,j}} \psi_i \times \prod_{k \neq j} \delta_{k \rightarrow i}(\mathbf{S}_{k,i}) \quad (5.11)$$

Applying to the cluster graph in Figure 5.4

$$\delta_{1,2}(d) = \int_a \psi_1(a, d) da \quad (5.12)$$

$$\delta_{2,3}(b) = \int_d \int_e \delta_{1,2}(d) \psi_2(b, d, e) dd \, de \quad (5.13)$$

$$\delta_{3,4}(f) = \int_b \int_c \delta_{2,3}(b) \psi_3(b, c, f) db \, dc \quad (5.14)$$

$$\delta_{4,3}(f) = \int_g \psi_4(f, g) dg \quad (5.15)$$

$$\delta_{3,2}(b) = \int_c \int_f \delta_{4,3}(f) \psi_3(b, c, f) dc \, df \quad (5.16)$$

$$\delta_{2,1}(d) = \int_b \int_e \delta_{3,2}(b) \psi_2(b, d, e) db \, de \quad (5.17)$$

Sometimes the value of certain variables in a PGM are known. This is known as evidence of a RV. If there is evidence of a RV, the evidence is available for entire the PGM and the RV itself doesn't appear in the PGM anymore. Evidence is indicated by a capital letter, thus X is evidence of the RV x . Evidence is used to reduce PDFs, before calculating outgoing messages. Therefore RVs with evidence is left out of the integral when calculating messages.

In equation a message, containing RV y , is calculated. Evidence of x is available and is indicated as X . The evidence of x is used to reduce the cluster ψ_i . After ψ_i has been multiplied with all other incoming messages, the result is integrated, but only over z .

$$\delta_{i \rightarrow j}(y) = \int_z \psi_i(x = X, y, z) \times \prod_{k \neq j} \delta_{k \rightarrow i}(\mathbf{S}_{k,i}) \quad (5.18)$$

After all the incoming messages of a cluster have been determined, the belief of a cluster can be calculated by multiplying the cluster's initial belief with all of the incoming messages, and normalizing the end result. Koller and Friedman [2] specifies this as

$$\beta_i(\mathbf{C}_i) \propto \psi_i \times \prod_k \delta_{k \rightarrow i}(\mathbf{S}_{k,i}) \quad (5.19)$$

5.6 Summary

Bayesian networks consists of factors and directed edges.

Cluster graphs can be constructed by grouping factors.

Clusters in a cluster graph are connected by sepsets.

Adjacent clusters pass information to each other about variables in sepset.

Outgoing messages can be calculated with equation 4.11.

Chapter 6

Localisation using Probabilistic Graphical Models

6.1 Linear localisation

6.2 Non-linear localisation

6.2.1 Taylor series expansion

6.2.2 Unscented transform

6.2.3 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.
- [7] Barber, D.: *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.