Introduction to
Simultaneous Localization and Mapping I

First Edition, 2023

Hyeonwoo Yu
Department of Intelligent Robotics & School of Mechanical Engineering,
Sungkyunkwan University

To Yunjeong

Contents

1

Introduction

“There is no spoon”

– Neo, *The Matrix (1999)*

Suppose you are standing alone in the environment, and only have eyesights. How do you define where you at? The location is relatve, therefore you can define your location by determining the locations of the surroundings exising in the environment. You may look around, see what exist and determine their locations related to your location. Here, you can simply set your present location as the origin of this environment for now. If you want to explore the environment, things become more challening. When you move, apparently your location has been changed, and you can naturaly feel the changes through your observations. Even further, you can also determine your changed location related to the locations of the surroundings. Here the problem occurs. In the viewpoimt of yourself, when you move, the locations of the sorroundings have also been changed because their locations are determined related to your location. Your location is determined by the locations of the surroundings observed by yourself, thus the locations of the surroundings should be known; the locations of the surroundings are determined by your location, thus your location should be known. In the real life, unfortunately both are unknown and you only have your own observations which are the relation between your location and the surroundings’ locations.

Exactly same issue happens in Robotics. Suppose you have a mobile robot explores in the environment, and obtain sequential observations. Since the robot’s location and the surroundings’ locations are both unknown, the robot should perform simultaneouse robot localization and the surroundings’ localization from sequential observations. The localization of the surrounding environment can be seen as drawing a map, called mapping of the environment. So in robotics simultaneous localization and mapping, SLAM is the basic essence of the mobile robot.

The localization of the Various sensors equiped to a mobile robot work as your eyesight. Similar to your observations of the surroundings, a mobile robot gets sensor information of the surroundings, called landmark. Any form of the sensor information can be used as landmarks, such as raw sensor values or geometric features like lines or edges computed from the sensor information, or more high-level semantic features.

When the environment is previously explored and known, the prior map can be given, and the global robot location also can be given from the external system. However, even in this well-known and well-set environment SLAM should be performed to 1) obtain the local map for place matching between the global map and the robot observations, 2) overcome the uncertainty of the global localization system and obtain the precise local localization, and 3) handle the dynamic environment.

In this textbook, the main concept of SLAM will be discussed.

2

Probability and Estimation

2.1 Basic Concepts in Probability

In this section, we briefly discuss the probabilistic essences for the robotics. Humans can experience the world by feeling various senses. Similarly, in the real-world scenarios, the mobile robots equipped sensors measure various quantities in real-time in order to experience the world. These quantities can be anything that can represent the robots such as the locations of the robots or the surrounding features. Unfortunately, every sensors do have noise and always have the uncertainties on their measurements, thus all the measurements should be considered as the random variable which follows the certain probability distributions according to the probabilistic modeling of sensor data. Probabilistic inference is therefore necessary in robotics in order to estimate the values of quantities from the noisy measurements, under the specific laws of probability.

Let $\mathbf{x}$ be the random variable of any quantities such as the robot location. Suppose $\mathbf{x}$ follows a probability distribution $p\left(\mathrm{x}\right)$. Let $\boldsymbol{x}$ be the specific state or values might take on, such as the specific states of the robot turn on/off, or specific coordinates of the current robot location. In the case of the discrete random variable like turn on/off of the robot, we have the probability

$$p\left(\boldsymbol{x}\right) = p\left(\mathbf{x} = \boldsymbol{x}\right) \geq 0 \tag{2.1}$$

to represent that the random variable $\mathbf{x}$ has the value of $\boldsymbol{x}$. From the axiom of probability, the sum of the probability over the all possible values is one as the following:

$$\sum_{\boldsymbol{x}} p\left(\mathbf{x} = \boldsymbol{x}\right) = 1. \tag{2.2}$$

For simplicity, from now we use $p\left(\boldsymbol{x}\right)$ instead of writing $p\left(\mathbf{x} = \boldsymbol{x}\right)$.

The major quantities we will handle in this book, such as the measurement of the robot location, are represented in continuous random various, because they exist in continuous spaces. These continuous random variables follow the specific probability density function (PDF), defined by their own probabilistic models which reflect the characteristics of the variables. In the case of the measurement of the sensor, for example, the measurement value always poses a noise due to the nature of the sensor. This noise model is used as the measurement model, which defines the probabilistic model and the PDF of the measurement. A common PDF is Gaussian distribution with mean μ and variance σ^2 given as the following:

$$\begin{aligned} p\left(\boldsymbol{x}\right) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{\left(\boldsymbol{x}-\mu\right)^2}{\sigma^2}\right) \\ &\stackrel{\text{def}}{=} \mathcal{N}\left(\mu, \sigma^2\right). \end{aligned} \tag{2.3}$$

In the aspect of the measurement example, μ can be regarded as the true quantity measured by the sensor, σ as the sensor uncertainty, and $\boldsymbol{x}$ as the value obtained from the measurement.

Since the quantities such as the sensor measurement contain multiple values in usual, the random variable $\boldsymbol{x}$ can take multi-dimensional variables. In the case of the multi-dimensional case, multivariate Gaussian distribution is given as the following:

$$\begin{aligned} p\left(\boldsymbol{x}\right) &= \frac{1}{\det\left(2\pi\boldsymbol{\Sigma}\right)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right) \\ &\stackrel{\text{def}}{=} \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), \end{aligned} \tag{2.4}$$

where $\boldsymbol{x}$ is the multivariate Gaussian random variable, $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ a symmetric, positive semidefinite matrix called covariance matrix.

Same as the discrete probability distribution, a PDF of continuous random variable integrates to 1:

$$\int_{\boldsymbol{x}} p\left(\boldsymbol{x}\right) d\boldsymbol{x} = 1. \tag{2.5}$$

Similar to the case of sensor measurement, in this book we assume all the continuous quantities can be represented as the continuous random variable; that is, all the continuous quantities are measurable, and possess probability densities.

Joint Distribution and Marginalization

Suppose we have two random variables $\mathbf{x}$ and $\mathbf{y}$. These random variables can be jointly defined for the measurements from multiple sensors of one quantity such as a robot location, or two different quantities such as each components of a robot location on two-dimensional coordinates. Here the *joint distribution* of $\mathbf{x}$ and $\mathbf{y}$ is given as:

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\mathrm{x} = \boldsymbol{x} \text{ and } \mathrm{y} = \boldsymbol{y}). \tag{2.6}$$

By integrating the joint distribution over the one random variable, say over $\boldsymbol{x}$, we have:

$$\int_{\boldsymbol{x}} p(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{x} = p(\boldsymbol{y}). \tag{2.7}$$

Here we call this integration as *marginalization*, and $p(\boldsymbol{y})$ is the *marginal distribution.*

Independence
If $\mathbf{x}$ and $\mathbf{y}$ are independent, the joint distribution can be expressed as:

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x})\, p(\boldsymbol{y}). \tag{2.8}$$

In usual, random variables involve information of other random variables. Suppose we observe that one of the sensor value $\mathbf{y}$ is determined as $\boldsymbol{y}$, and we want to figure out the probability that the other sensor value $\mathbf{x}$ is detected as $\boldsymbol{x}$ conditioned on our previous observation. In this case we denote the *conditional probability* as:

$$p(\boldsymbol{x}|\boldsymbol{y}) = p(\mathrm{x} = \boldsymbol{x}|\mathrm{y} = \boldsymbol{y}). \tag{2.9}$$

Using the marginal distribution and the joint distribution, the conditional probability is defined as:

$$p(\boldsymbol{x}|\boldsymbol{y}) \stackrel{\text{def}}{=} \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{y})} \tag{2.10}$$

When $p(\boldsymbol{y})$ is nonzero and $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent, substituting Eqn. (2.8) into Eqn. (2.10) we have:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{x})\, p(\boldsymbol{y})}{p(\boldsymbol{y})} = p(\boldsymbol{x}). \tag{2.11}$$

Therefore, when the random variables are independent their conditional distributions are not *conditioned* to other random variables, thereby the conditional distribution is equal to the marginal distribution.

Expectation
The random variable $\boldsymbol{x}$ is generated by its probability $p(\boldsymbol{x})$. Therefore, when we compute the mean of the variable $\boldsymbol{x}$ it is necessary to consider its probability, not by simple summation and division by the number of variables,

but by weighting the value $\boldsymbol{x}$. This weighted summation can be achieved as the following:

$$\mathbb{E}(\boldsymbol{x}) = \int_{\boldsymbol{x}} \boldsymbol{x} \cdot p(\boldsymbol{x})\, d\boldsymbol{x} \tag{2.12}$$

where $\mathbb{E}(\boldsymbol{x})$ is the *expectation* (mean). In general, when we want to compute the expectation of the value derived from the function f, then the expectation can be expressed as:

$$\mathbb{E}(\boldsymbol{x}) = \int_{\boldsymbol{x}} f(\boldsymbol{x})\, p(\boldsymbol{x})\, d\boldsymbol{x}. \tag{2.13}$$

2.2 Multivariate Gaussian and Conditional Probability

In the realm of Simultaneous Localization and Mapping (SLAM), the synergy between multivariate Gaussian distributions and conditional probability plays a pivotal role in enhancing our understanding of the environment. At the core of this integration lies the recognition that valuable information about the robot's position and surroundings is not solely confined to the direct measurements obtained from a single sensor. Instead, the intricate nature of the environment often necessitates the consideration of additional data sources, such as observations from other sensors.
Conditional probability becomes a necessity in this context as it enables us to quantify the likelihood of a certain measurement given the information gleaned from other sensors. Essentially, it allows us to refine our estimation by incorporating the influence of related observations, providing a more comprehensive and accurate representation of the true state of the environment. This chapter delves into the fundamental principles of multivariate Gaussian distributions and conditional probability, unraveling their significance in harnessing diverse sources of information for a more robust and nuanced SLAM framework.

Suppose we have two Gaussian random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ that are jointly Gaussian, where $\boldsymbol{x}$ is $k \times 1$ and $\boldsymbol{y}$ is $l \times 1$, with mean vector $\left[\mathbb{E}(\boldsymbol{x})^T\, \mathbb{E}(\boldsymbol{y})^T\right]^T$ and partitioned covariance matrix

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_{xx} & \boldsymbol{C}_{xy} \\ \boldsymbol{C}_{yx} & \boldsymbol{C}_{yy} \end{bmatrix} = \begin{bmatrix} k \times k & k \times l \\ l \times k & l \times l \end{bmatrix} \tag{2.14}$$

thus we have

$$p(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{\sqrt{(2\pi)^{k+l} \det \boldsymbol{C}}} \exp\left(-\frac{1}{2} \begin{bmatrix} \boldsymbol{x} - \mathbb{E}(\boldsymbol{x}) \\ \boldsymbol{y} - \mathbb{E}(\boldsymbol{y}) \end{bmatrix}^T \boldsymbol{C}^{-1} \begin{bmatrix} \boldsymbol{x} - \mathbb{E}(\boldsymbol{x}) \\ \boldsymbol{y} - \mathbb{E}(\boldsymbol{y}) \end{bmatrix} \right), \tag{2.15}$$

then the conditional PDF $p\left(\boldsymbol{y}|\boldsymbol{x}\right)$ is also Gaussian represented as:

$$\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}\left(\mathbb{E}\left(\boldsymbol{y}|\boldsymbol{x}\right), \boldsymbol{C}_{y|x}\right) \tag{2.16}$$

where

$$\mathbb{E}\left(\boldsymbol{y}|\boldsymbol{x}\right) = \mathbb{E}\left(\boldsymbol{y}\right) \boldsymbol{C}_{yx} \boldsymbol{C}_{xx}^{-1} \left(\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right)\right) \tag{2.17}$$

$$\boldsymbol{C}_{y|x} = \boldsymbol{C}_{yx} \boldsymbol{C}_{xx}^{-1} \boldsymbol{C}_{xy}. \tag{2.18}$$

In the case of Bivariate Gaussian random variable $[x, y]^T$, the covariance $\boldsymbol{C}$ can be simplified to:

$$\boldsymbol{C} = \begin{bmatrix} var\left(x\right) & cov\left(x, y\right) \\ cov\left(y, x\right) & var\left(y\right) \end{bmatrix} \tag{2.19}$$

so that

$$\mathbb{E}\left(y|x\right) = \mathbb{E}\left(y\right) + \frac{cov\left(x, y\right)}{var\left(x\right)} \left(x - \mathbb{E}\left(x\right)\right) \tag{2.20}$$

$$var\left(y|x\right) = var\left(y\right) - \frac{cov^2\left(x, y\right)}{var\left(x\right)} \tag{2.21}$$

$$= var\left(y\right)\left(1 - \frac{cov^2\left(x, y\right)}{var\left(x\right) var\left(y\right)}\right) \tag{2.22}$$

$$= var\left(y\right)\left(1 - \rho\right), \tag{2.23}$$

where ρ is Correlation coefficient.

Proof of Conditional Gaussian

We can easily compute the conditional mean and the conditional covariance by directly applying the definition of the conditional probability. Since we have the joint PDF $p\left(\boldsymbol{x}, \boldsymbol{y}\right)$, the conditional PDF $p\left(\boldsymbol{y}|\boldsymbol{x}\right)$ can be represented as:

$$\begin{aligned} p\left(\boldsymbol{y}|\boldsymbol{x}\right) &= \frac{p\left(\boldsymbol{x}, \boldsymbol{y}\right)}{p\left(\boldsymbol{x}\right)} \\ &= \frac{\frac{1}{\sqrt{(2\pi)^{k+l} \det \boldsymbol{C}}} \exp\left(-\frac{1}{2} \begin{bmatrix} \boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right) \\ \boldsymbol{y} - \mathbb{E}\left(\boldsymbol{y}\right) \end{bmatrix}^T \boldsymbol{C}^{-1} \begin{bmatrix} \boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right) \\ \boldsymbol{y} - \mathbb{E}\left(\boldsymbol{y}\right) \end{bmatrix}\right)}{\frac{1}{\sqrt{(2\pi)^{k} \det \boldsymbol{C}_{xx}}} \exp\left(-\frac{1}{2}\left(\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right)\right)^T \boldsymbol{C}_{xx}^{-1}\left(\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right)\right)\right)}. \end{aligned} \tag{2.24}$$

To calculate the term $\frac{\det \boldsymbol{C}}{\det \boldsymbol{C}_{xx}}$, we use the Schur complement. Suppose the matrices $\boldsymbol{A}_{11}$ and $\boldsymbol{A}_{22}$ are invertible and let $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$. Then the

Schur complement is defined as follows:

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{A}_{11}\boldsymbol{A}_{12}^{-1} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \end{bmatrix}. \tag{2.25}$$

Taking determinent to both side we have:

$$\det\left(\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}\right) = \det \boldsymbol{A}_{11} \det\left(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\right). \tag{2.26}$$

Substitute $\boldsymbol{A}$ with $\boldsymbol{C}$, we have:

$$\det \boldsymbol{C} = \det \boldsymbol{C}_{xx} \det\left(\boldsymbol{C}_{yy} - \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{yx}\right) \tag{2.27}$$

so that

$$\frac{\det \boldsymbol{C}}{\det \boldsymbol{C}_{xx}} = \det\left(\boldsymbol{C}_{yy} - \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{yx}\right). \tag{2.28}$$

Substituting Eqn. (2.28) to Eqn. (2.24) we have:

$$p\left(\boldsymbol{y}|\boldsymbol{x}\right) = \frac{1}{\sqrt{(2\pi)^l \det\left(\boldsymbol{C}_{yy} - \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{yx}\right)}} \exp\left(-\frac{1}{2}Q\right), \tag{2.29}$$

where

$$Q = \begin{bmatrix} \boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right) \\ \boldsymbol{y} - \mathbb{E}\left(\boldsymbol{y}\right) \end{bmatrix}^T \boldsymbol{C}^{-1} \begin{bmatrix} \boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right) \\ \boldsymbol{y} - \mathbb{E}\left(\boldsymbol{y}\right) \end{bmatrix} - \left(\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right)\right)^T \boldsymbol{C}_{xx}^{-1}\left(\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right)\right). \tag{2.30}$$

From Eqn. (2.28), we can also have $\boldsymbol{C}^{-1}$ as the following:

$$\boldsymbol{C}^{-1} = \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{C}_{xx}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1} & \boldsymbol{I} \end{bmatrix} \tag{2.31}$$

where $\boldsymbol{B} = \boldsymbol{C}_{yy} - \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{yx}$. Let $\bar{x} = \boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right)$ and $\bar{y} = \boldsymbol{y} - \mathbb{E}\left(\boldsymbol{y}\right)$. Substituting Eqn. (2.31) into Eqn. (2.30) we have:

$$\begin{aligned} Q &= \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{C}_{xx}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} - \bar{x}^T\boldsymbol{C}_{xx}^{-1}\bar{x} \\ &= \left(\bar{y} - \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\bar{x}\right)^T \boldsymbol{B}^{-1}\left(\bar{y} - \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\bar{x}\right) \\ &= \left[\boldsymbol{y} - \left(\mathbb{E}\left(\boldsymbol{y}\right) + \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\left(\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right)\right)\right)\right]^T \\ &\quad \left[\boldsymbol{C}_{yy} - \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{yx}\right]^{-1}\left[\boldsymbol{y} - \left(\mathbb{E}\left(\boldsymbol{y}\right) + \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\left(\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right)\right)\right)\right]. \end{aligned} \tag{2.32}$$

Substituting Eqn. (2.32) into Eqn. (2.29), we finally have:

Conditional Probability of Gaussian:

$$\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}\left(\mathbb{E}\left(\boldsymbol{y}|\boldsymbol{x}\right), \boldsymbol{C}_{\boldsymbol{y}|\boldsymbol{x}}\right) \tag{2.33}$$

where

$$\mathbb{E}\left(\boldsymbol{y}|\boldsymbol{x}\right) = \mathbb{E}\left(\boldsymbol{y}\right) + \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\left(\boldsymbol{x} - \mathbb{E}\left(\boldsymbol{x}\right)\right) \tag{2.34}$$

$$\boldsymbol{C}_{\boldsymbol{y}|\boldsymbol{x}} = \boldsymbol{C}_{yy} - \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{yx}. \tag{2.35}$$

The most important thing here is that the conditional distribution of two Gaussian random variables is still represented as a Gaussian distribution. We will utilize this fact for the remaining chapters.

3

Bayesian Estimation and Measurement

Consider the robot is about to explore the environment. In order to understand the surrounding environment, first the robot should obtain the sensor information about the environment, for example distances between the robot and the surrounding structures. The sensor information, in this case, is a measurement of the true distance. The problem lies here. In the realm of Robotics, our sensors serve as imperfect observers, providing us with measurements that are, in essence, true essence of the environment involving noises. This "true essence" represents the genuine state of the surrounding environments or events, unblemished by the inherent inaccuracies of the sensors. Our goal is to uncover the genuine nature concealed within the noisy sensor measurements, and achieve the true essence of the surroundings for the robot perception and further robot applications. Simply, our goal is to estimate the true value from the measurements.

In order to discern and achieve this authentic landscape, we turn to the probabilistic estimation theory. Apparently, it is impossible to know the correct noise model of the certain sensors. However this is not the end. We empirically know that the various noise models in nature approximately follow the known distributions such as Gaussian distribution, thereby it is possible to approximate and make sense of the true reality that lies beneath the surface of our sensor measurements. Here Bayesian estimation becomes our guiding light from now.

3.1 Maximum Likelihood Estimation

Suppose the robot is equipped with the time-of-flight (ToF) sensor and is in a stationary state. Consider that the robot is measuring the distance to the structure placed in front. Here, let $x[n]$ be the measurement obtained from the sensor in time n, and let A be the true distance between the sensor and the structure. Since the measurement of the sensor inevitably involves noise, we cannot say that the measurement $x[n]$ is the true distance A, but we can approximately find the value of A in probabilistic manner.

For this estimation, first we need to assume the noise model (or measurement model) of the sensor with known distributions. In usual, Gaussian distribution plays a main role as the measurement model. Assume that $w[n]$ is the random variable representing the noise occurred in the n^{th} measurement. Let us say the robot obtains N times measurements, then we can express the relation between $x[n]$ and A as the following:

$$x[n] = A + w[n]\,,\ \ n = 0, 1, ..., N-1 \tag{3.1}$$

where

$$w[n] \sim \mathcal{N}\left(0, \sigma^2\right). \tag{3.2}$$

Here σ^2 reflects the uncertainty of the sensor measurement. In this example let us assume that σ^2 is known, which can be obtained empirically in practice.

Since $w[n]$ is a Gaussian random variable and A is a constant, $x[n]$ is also a Gaussian random variable with the following probability distribution:

$$p(x[n]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x[n]-A)^2}{2\sigma^2}\right], \tag{3.3}$$

which is the *likelihood* of the random variable $x[n]$. In this equation, apparently A is a constant hence not a probabilistic random variable. However, in usual we express A as if a random variable for convenience, because when the value of A is determined, we can say the random variable $x[n]$ follows the above probability distribution which is determined by A. In other words, the distribution of $x[n]$ is conditionally determined by A. In this context, we can also rewrite this likelihood as if the conditional probability:

$$p(x[n]; A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x[n]-A)^2}{2\sigma^2}\right]. \tag{3.4}$$

Apparently, in order to represent the conditional probability the left-hand side of the above equation should be $p(x[n]\,|A)$. However, we need to note that the likelihood is not a conditional probability in usual because A in our case is a constant, not a random variable. Therefore we simply use a semicolon to express the relation between $x[n]$ and A in the likelihood.

The likelihood involves the noise model of the measurement, thus we can utilize the likelihood as the measurement model (or observation model). Since the robot measures the distance N times, under the assumption that all the measurements are independent to each other we can represent the likelihood of N measurements as the following:

$$
\begin{aligned}
p(\boldsymbol{x}; A) &= p(x[0], ..., x[N-1]; A) \\
&= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[n]-A)^2\right] \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right] \qquad (3.5)
\end{aligned}
$$

where $\boldsymbol{x} = \{x[0], ..., x[N-1]\}$. To find the optimal parameter of A, the first approach we can consider is Maximum Likelihood Estimation (MLE). Suppose we have optimal parameter $\hat{A}$ for the true distance. Since we assume that the measurements $\boldsymbol{x}$ is sampled from the likelihood, this likelihood has the maximum probability with the optimal parameter $\hat{A}$ thus we have the following expression:

$$
\hat{A} = \underset{A}{\operatorname{argmax}}\, p(\boldsymbol{x}; A) \qquad (3.6)
$$

Therefore, the optimal parameter $\hat{A}$ can be calculated by taking the derivation of the likelihood. For convenience, in usual we use log-likelihood for Gaussian distribution, since Gaussian involves quadratic form which is convex and the logarithm is monotonically increasing function so that the optimal point remains the same while in logarithm form. Therefore we have:

$$
\hat{A} = \underset{A}{\operatorname{argmax}}\, \log p(\boldsymbol{x}; A) \qquad (3.7)
$$

where

$$
\begin{aligned}
\log p(\boldsymbol{x}; A) &= -\frac{N}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2 \\
&= const - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2. \qquad (3.8)
\end{aligned}
$$

Note that in the case of Gaussian, we have quadratic form inside of the exponential form when taking logarithm of Gaussian. This will play an important role in further. By taking the partial derivative of the log-likelihood, we have the following:

$$
\frac{\partial \log p(\boldsymbol{x}; A)}{\partial A} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A). \qquad (3.9)
$$

Set to zero we finally have:

$$\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]\,. \tag{3.10}$$

Therefore, for the measurements of the constant quantities such as the distance between the static robot and the surrounding structure, we can conclude that the optimal parameters can be determined by taking average of the multiple measurement. Note that the noise model (or observation model) here is assumed to be Gaussian.

3.2 Bayesian Mean Square Error

Estimating the true value of the target quantities always poses a great challenge, due to the measurement noise or model uncertainty. In the previous section we handled this uncertainty by assuming the measurement model as Gaussian and by applying MLE. In this section we handle this noise by simply minimizing the error and compare the result to the MLE case. Assume that the robot at a standstill measures the distance to the static object and obtain the measurement $x[n]$ in time n, and let A be the true distance. Due to the noises from the measurement model, the measurement $x[n]$ and the true distance A always pose a difference, which is the error that we need to minimize. To compute the error, we define Least square error criterion J as the cost function as the following:

$$J(A) = \sum_{n=0}^{N-1} \left(x[n] - A\right)^2\,. \tag{3.11}$$

Apparently, the optimal solution $\hat{A}$ of this quadratic form can be given as:

$$\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]\,, \tag{3.12}$$

which is the same result of the optimal solution from MLE.

Although the least squares method enables us to attain the same solution in Gaussian scenarios, it falls short in providing insights beyond measurement models expressed solely in terms of likelihood.

In robotics, there are instances where various information can be anticipated in advance through pre-existing knowledge of the system model. For example, having prior information about the surrounding environment allows for a rough anticipation of the positions of observed objects. Additionally, knowing the control input provided to the robot enables the prediction of the robot's anticipated location before movement. From this, predictions can be made regarding how far an observed object will be from the robot compared to its position before movement.

To illustrate, consider having prior knowledge about the approximate distance of an object placed in front of a robot. Let the true distance from the robot to the object be denoted as A, and the observed distance from the measurement be denoted as x. Since likelihood only represents the measurement model, given A, it mimics the probability density function (PDF) for the observed value x. Therefore, there is no avenue to incorporate our existing knowledge about the object's position. Here, A is simply a constant parameter.

However, with prior knowledge about the approximate range of distances to the object, we now possess probabilistic information about A, treating it as a random variable. For instance, if we have prior knowledge that A lies between a and b, the PDF for A can be expressed as follows:

$$p(A) = \begin{cases} b - a, \text{if } a < A < b \\ 0 \text{ else} \end{cases} . \tag{3.13}$$

Considering such prior knowledge allows for compensating noisy observations or incorrect system models. For instance, when x is observed to be greater than or equal to b, since the probability of the true distance A being greater than or equal to b is zero, measures such as ignoring the least squares error in this case can be taken.

Now, let's apply the prior knowledge we have about the parameter A to define the least squares. As A is also a random variable, we can consider not just the likelihood but the joint probability density function (PDF) of $\boldsymbol{x}$ and A, denoted as $p(\boldsymbol{x}, A)$. Let's denote the optimal value for the distance as $\hat{A}$. The Bayesian mean square error ($Bmse$) is given by the following expression:

$$\begin{aligned} Bmse\left(\hat{A}\right) &= \mathbb{E}\left(A - \hat{A}\right)^2 \\ &= \int\int \left(A - \hat{A}\right)^2 p(\boldsymbol{x}, A)\, d\boldsymbol{x} dA. \end{aligned} \tag{3.14}$$

Note that this $Bmse$ can be considered as the weighted error summation, different from the traditional Least square error criterion J. Here each error $\left(A - \hat{A}\right)^2$ has its corresponding weight $p(\boldsymbol{x}, A)$ so that the probability of the target error now can be reflected.

Since $p(\boldsymbol{x}, A) = p(A|\boldsymbol{x})\, p(\boldsymbol{x})$, we have:

$$Bmse\left(\hat{A}\right) = \int \left(\int \left(A - \hat{A}\right)^2 p(A|\boldsymbol{x})\, dA\right) p(\boldsymbol{x})\, d\boldsymbol{x}. \tag{3.15}$$

Our goal is to have the optimal value $\hat{A}$ from the redefined error criterion *Bmse*. From the axiom of probability $p(\boldsymbol{x}) \geq 0$, therefore $\hat{A}$ can be obtained by minimizing the integral of A. Taking the derivation we have:

$$\begin{aligned}\frac{\partial}{\partial A}\int\left(A-\hat{A}\right)^2 p(A|\boldsymbol{x})\,dA &= \int\frac{\partial}{\partial A}\left(A-\hat{A}\right)^2 p(A|\boldsymbol{x})\,dA \\ &= \int -2\left(A-\hat{A}\right)p(A|\boldsymbol{x})\,dA \\ &= -2\int Ap(A|\boldsymbol{x})\,dA + 2\hat{A}\underbrace{\int p(A|\boldsymbol{x})\,dA}_{=1}.\end{aligned} \tag{3.16}$$

Set to zero we finally have:

$$\begin{aligned}\hat{A} &= \int Ap(A|\boldsymbol{x})\,dA \\ &= \mathbb{E}(A|\boldsymbol{x}).\end{aligned} \tag{3.17}$$

In the end, the mean becomes the optimal solution for Bmse. That is, for arbitrary parameter A, the minimum mean square error estimator considering the joint PDF is given by the mean of $p(A|\boldsymbol{x})$, denoted as $\mu_{A|\boldsymbol{x}}$. Then what is $p(A|\boldsymbol{x})$? How do we find $p(A|\boldsymbol{x})$?

Bayes' Rule

$p(A|\boldsymbol{x})$ in Eqn. (3.17) is the posterior, the probability of A given $\boldsymbol{x}$, which can be represented using the *likelihood* and the *prior* by the following Bayes' rule:

$$\begin{aligned}p(A|\boldsymbol{x}) &= \frac{p(\boldsymbol{x}|A)\,p(A)}{p(\boldsymbol{x})} && \text{(Bayes' Rule)} \\ &= \frac{p(\boldsymbol{x}|A)\,p(A)}{\int p(\boldsymbol{x},A)\,dA} && \text{(marginalization)} \\ &= \frac{p(\boldsymbol{x}|A)\,p(A)}{\int p(\boldsymbol{x}|A)\,p(A)\,dA} && \text{(Def. of Conditional Dist.)} \\ &\propto p(\boldsymbol{x}|A)\,p(A) && (p(\boldsymbol{x}) \text{ is } \mathit{Constant}).\end{aligned} \tag{3.18}$$

Note that $\boldsymbol{x}$ are measurements. In other words, $\boldsymbol{x}$ denote the observations of the true value A, which is the parameter of our probability. At this point of time, the observations $\boldsymbol{x}$ are the determined variables observed in the past, even if those variables are sampled from the probability distribution. Therefore $\boldsymbol{x}$ are the known variables thus we say $p(\boldsymbol{x})$ in Eqn. (3.18) becomes constant. In Bayesian estimation, $\boldsymbol{x}$ are also called as *evidence*, since the observations $\boldsymbol{x}$ play a key role of estimating the desired parameter A.

In general, directly obtaining, assuming or formulating the posterior is challenging, compared to the cases of the likelihood (measurement model), or the prior (range of the parameter). Similarly, considering $p(\boldsymbol{x})$ is more challenging, because this distribution is the *generativemodel*. Therefore, in order to computing the posterior, simply from the likelihood and the prior, we will utilize Eqn. (3.18) in the rest of the chapters in this book.

Consequently, finding the optimal solution of Bayesian mean square error is equal to the solution of MAP, and the solution is determined as the mean of the posterior distribution. In summary, for Bayesian estimation we follow:

- Find the likelihood, mostly representing the observation model
- Assume the reasonable prior
- Compute the posterior from the likelihood and the prior by Bayes' rule
- Use the mean of the posterior as the optimal solution of the Bayesian estimation