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:

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

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:

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

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:

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

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:

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

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:

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

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 $p(\boldsymbol{y}|\boldsymbol{x})$ is also Gaussian represented as:

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

where

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

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

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

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

so that

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

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

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

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

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 $p(\boldsymbol{x}, \boldsymbol{y})$, the conditional PDF $p(\boldsymbol{y}|\boldsymbol{x})$ can be represented as:

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

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

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.