1

Introduction

Robotics inherently deals with things that move in the world. We live in an era of rovers on Mars, drones surveying the Earth, and, soon, self-driving cars. And, although specific robots have their subtleties, there are also some common issues we must face in all applications, particularly *state estimation* and *control*.

The *state* of a robot is a set of quantities, such as position, orientation, and velocity, that, if known, fully describe that robot's motion over time. Here we focus entirely on the problem of estimating the state of a robot, putting aside the notion of control. Yes, control is essential, as we would like to make our robots behave in a certain way. But, the first step in doing so is often the process of determining the state. Moreover, the difficulty of state estimation is often underestimated for real-world problems, and thus it is important to put it on an equal footing with control.

In this book, we introduce the classic estimation results for linear systems corrupted by Gaussian measurement noise. We then examine some of the extensions to nonlinear systems with non-Gaussian noise. In a departure from typical estimation texts, we take a detailed look at how to tailor general estimation results to robots operating in three-dimensional space, advocating a particular approach to handling rotations.

The rest of this introduction provides a little history of estimation, discusses types of sensors and measurements, and introduces the problem of state estimation. It concludes with a breakdown of the contents of the book and provides some other suggested reading.

1.1 A Little History

About 4,000 years ago, the early seafarers were faced with a vehicular state estimation problem: how to determine a ship's position while at sea. Primitive charts and observations of the sun allowed local navigation along coastlines. Early instruments also helped with navigation. The astrolabe was a handheld model of the universe that allowed various astronomical problems to be solved; it could be used as an inclinometer to determine latitude, for example. Its origins can be traced to the Hellenistic civilization around 200 BC and was greatly advanced in the Islamic world starting in the eighth century by mathematician Muhammad al-Fazārī and astronomer Abū al-Battănī (aka, Albatenius). Also around 100 BC in ancient Greece, the so-called Antikythera mechanism was the world's first analogue computer capable of predicting astronomical positions and eclipses decades into the future.

Figure 1.1
Quadrant. A tool used to measure angles.

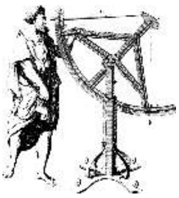


Figure 1.2
Harrison's H4. The first clock able to keep accurate time at sea, enabling determination of longitude.

Carl Friedrich Gauss (1777–1855) was a German mathematician who contributed significantly to many fields including statistics and estimation. Much of this book his based on his work.

Rudolf Emil Kálmán (1930–2016) was a Hungarian-born American electrical engineer, mathematician, and inventor. He is famous for the *Kalman filter* and introducing the notions of *controllability* and *observability* in systems theory.

Despite these early capabilities, it was not until the fifteenth century that global navigation on the open sea became widespread with the advent of additional key technologies and tools. The mariner's compass, an early form of the magnetic compass, allowed crude measurements of direction to be made. Together with coarse nautical charts, the compass made it possible to sail along rhumb lines between key destinations (i.e., following a compass bearing). A series of instruments was then gradually invented that made it possible to measure the angle between distant points (i.e., cross-staff, astrolabe, quadrant, sextant, theodolite) with increasing accuracy.

These instruments allowed latitude to be determined at sea fairly readily using celestial navigation. For example, in the Northern Hemisphere, the angle between the North Star, Polaris, and the horizon provides the latitude. Longitude, however, was a much more difficult problem. It was known early on that an accurate timepiece was the missing piece of the puzzle for the determination of longitude. The behaviours of key celestial bodies appear differently at different locations on the Earth. Knowing the time of day therefore allows longitude to be inferred. In 1764, British clockmaker John Harrison built the first accurate portable timepiece that effectively solved the longitude problem; a ship's longitude could be determined to within about 10 nautical miles.

Estimation theory also finds its roots in astronomy. The method of least squares was pioneered by Gauss,[1] who developed the technique to minimize the impact of measurement error in the prediction of orbits. Gauss reportedly used least squares to predict the position of the dwarf planet Ceres after it passed behind the Sun, accurate to within half a degree (about nine months after it was last seen). The year was 1801, and Gauss was 23. Later, in 1809, he proved that the least-squares method is optimal under the assumption of normally distributed errors (Gauss, 1809) and later still he removed this assumption (Gauss, 1821, 1823). Most of the classic estimation techniques in use today can be directly related to Gauss' least-squares method.

The idea of fitting models to minimize the impact of measurement error carried forward, but it was not until the middle of the twentieth century that estimation really took off. This was likely correlated with the dawn of the computer age. In 1960, Kalman published two landmark papers that have defined much of what has followed in the field of state estimation. First, he introduced the notion of *observability* (Kalman, 1960a), which tells us when a state can be inferred from a set of measurements in a dynamic system. Second, he introduced an optimal framework for estimating a system's state in the presence of measurement noise (Kalman, 1960b); this classic technique for linear systems (whose measurements are corrupted by Gaussian noise) is famously known as the *Kalman filter*, and has been the workhorse of estimation for the more than 60 years since its inception. Although used in many fields, it has been widely adopted in aerospace

[1] There is some debate as to whether Adrien Marie Legendre might have come up with least squares before Gauss.

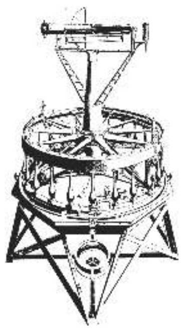
applications. Researchers at the *National Aeronautics and Space Administration (NASA)* were the first to employ the Kalman filter to aid in the estimation of spacecraft trajectories on the Ranger, Mariner, and Apollo programs. In particular, the on-board computer on the Apollo 11 Lunar Module, the first manned spacecraft to land on the surface of the Moon, employed a Kalman filter to estimate the module's position above the lunar surface based on noisy inertial and radar measurements.

Many incremental improvements have been made to the field of state estimation since these early milestones. Faster and cheaper computers have allowed much more computationally complex techniques to be implemented in practical systems. Today, exciting new sensing technologies are coming along (e.g., digital cameras, laser imaging, the Global Positioning System) that pose new challenges to this old field.

1.2 Sensors, Measurements and Problem Definition

To understand the need for state estimation is to understand the nature of sensors. All sensors have a limited precision. Therefore, all measurements derived from real sensors have associated uncertainty. Some sensors are better at measuring specific quantities than others, but even the best sensors still have a degree of imprecision. When we combine various sensor measurements into a state estimate, it is important to keep track of all the uncertainties involved and therefore (it is hoped) know how confident we can be in our estimate.

Figure 1.3 Theodolite. A better tool to measure angles.



In a way, state estimation is about doing the best we can with the sensors we have. This, however, does not prevent us from, in parallel, improving the quality of our sensors. A good example is the *theodolite* sensor that was developed in 1787 to allow triangulation across the English Channel. It was much more precise than its predecessors and helped show that much of England was poorly mapped by tying measurements to well-mapped France.

It is useful to put sensors into two categories: *interoceptive*[2] and *exteroceptive*. These are actually terms borrowed from human physiology, but they have become somewhat common in engineering. Some definitions follow:[3]

in·tero·cep·tive [int-ə-rō-'sep-tiv], *adjective*: of, relating to, or being stimuli arising within the body.

ex·tero·cep·tive [ek-stə-rō-'sep-tiv], *adjective*: relating to, being, or activated by stimuli received by an organism from outside.

Typical interoceptive sensors are the accelerometer (measures translational acceleration), gyroscope (measures angular rate), and wheel odometer (measures angular rate). Typical exteroceptive sensors are the camera (measures range/bearing to a landmark or landmarks) and time-of-flight transmitter/receiver (e.g., laser rangefinder, pseudolites, *Global Positioning System (GPS)* transmitter/receiver). Roughly speaking, we can think of exteroceptive measurements as being of the

[2] Sometimes *proprioceptive* is used synonomously.
[3] *Merriam-Webster's Dictionary*.

position and orientation of a vehicle, whereas interoceptive ones are of a vehicle's velocity or acceleration. In most cases, the best state estimation concepts make use of both interoceptive and exteroceptive measurements. For example, the combination of a GPS receiver (exteroceptive) and an *inertial measurement unit (IMU)* (three linear accelerometers and three rate gyros; interoceptive) is a popular means of estimating a vehicle's position/velocity on Earth. And, the combination of a Sun/star sensor (exteroceptive) and three rate gyros (interoceptive) is commonly used to carry out pose determination on satellites.

Now that we understand a little bit about sensors, we are prepared to define the problem that will be investigated in this book:

Estimation *is the problem of reconstructing the underlying state of a system given a sequence of measurements as well as a prior model of the system.*

EARLY ESTIMATION MILESTONES

1654	Pascal and Fermat lay foundations of probability theory
1764	Bayes' rule
1801	Gauss uses least-squares to estimate the orbit of the planetoid Ceres
1805	Legendre publishes 'least-squares'
1913	Markov chains
1933	(Chapman)–Kolmogorov equations
1949	Wiener filter
1960	Kalman (Bucy) filter
1965	Rauch–Tung–Striebel smoother
1970	Jazwinski coins 'Bayes filter'

There are many specific versions of this problem and just as many solutions. The goal is to understand which methods work well in which situations, in order to pick the best tool for the job.

1.3 How This Book Is Organized

The book is broken into three main parts:

I. Estimation Machinery
II. Three-Dimensional Machinery
III. Applications

The first part, *Estimation Machinery*, presents classic and state-of-the-art estimation tools, without the complication of dealing with things that live in three-dimensional space (and therefore translate and rotate); the state to be estimated is assumed to be a generic vector. For those not interested in the details of working in three-dimensional space, this first part can be read in a stand-alone manner. It covers both recursive state estimation techniques and batch methods (less common in classic estimation books). As is commonplace in robotics and machine learning today, we adopt a *Bayesian* approach to estimation in this book. We contrast (full) Bayesian methods with *maximum a posteriori (MAP)* methods, and attempt to make clear the difference between these when faced with nonlinear problems. The book also connects continuous-time estimation with Gaussian process regression from the machine-learning world. Finally, it touches on some practical issues, such as determining how well an estimator is performing, and handling outliers and biases.

The second part, *Three-Dimensional Machinery*, provides a basic primer on three-dimensional geometry and gives a detailed but accessible introduction to matrix Lie groups. To represent an object in three-dimensional space, we need to talk about that object's translation and rotation. The rotational part turns out to be a problem for our estimation tools because rotations are not *vectors* in the usual sense and so we cannot naively apply the methods from Part I to three-dimensional robotics problems involving rotations. Part II, therefore, examines the geometry, kinematics, and probability/statistics of rotations and poses (translation plus rotation).

Finally, in the third part, *Applications*, the first two parts of the book are brought together. We look at a number of classic three-dimensional estimation problems involving objects translating and rotating in three-dimensional space. We show how to adapt the methods from Part I based on the knowledge gained in Part II. The result is a suite of easy-to-implement methods for three-dimensional state estimation. The spirit of these examples can also, we hope, be adapted to create other novel techniques moving forward.

Appendix A provides a summary of matrix algebra and calculus that can serve as a primer or reference while reading this book.

1.4 Relationship to Other Books

There are many other books on state estimation and robotics, but very few cover both topics simultaneously. We briefly describe a few works that do cover these topics and their relationships to this book.

Probabilistic Robotics by Thrun et al. (2006) is a great introduction to mobile robotics, with a large focus on state estimation in relation to mapping and localization. It covers the probabilistic paradigm that is dominant in much of robotics today. It mainly describes robots operating in the two-dimensional, horizontal plane. The probabilistic methods described are not necessarily limited to the two-dimensional case, but the details of extending to three dimensions are not provided.

Computational Principles of Mobile Robotics by Dudek and Jenkin (2010) is a great overview book on mobile robotics that touches on state estimation, again in relation to localization and mapping methods. It does not work out the details of performing state estimation in 3D.

Mobile Robotics: Mathematics, Models, and Methods by Kelly (2013) is another excellent book on mobile robotics and covers state estimation extensively. Three-dimensional situations are covered, particularly in relation to satellite-based and inertial navigation. As the book covers all aspects of robotics, it does not delve deeply into how to handle rotational variables within three-dimensional state estimation.

Robotics, Vision, and Control by Corke (2011) is another great and comprehensive book that covers state estimation for robotics, including in three dimensions. Similarly to the previously mentioned book, the breadth of Corke's book necessitates that it not delve too deeply into the specific aspects of state estimation treated herein.

Bayesian Filtering and Smoothing by Särkkä (2013) is a super book focused on recursive Bayesian methods. It covers the recursive methods in far more depth than this book, but does not cover batch methods nor focus on the details of carrying out estimation in three dimensions.

Stochastic Models, Information Theory, and Lie Groups: Classical Results and Geometric Methods by Chirikjian (2009), an excellent two-volume work, is perhaps the closest in content to the current book. It explicitly investigates the consequences of carrying out state estimation on matrix Lie groups (and hence rotational variables). It is quite theoretical in nature and goes beyond the current book in this sense, covering applications beyond robotics.

Engineering Applications of Noncommutative Harmonic Analysis: With Emphasis on Rotation and Motion Groups by Chirikjian and Kyatkin (2001) and the recent update, *Harmonic Analysis for Engineers and Applied Scientists: Updated and Expanded Edition* (Chirikjian and Kyatkin, 2016), also provide key insights to representing probability globally on Lie groups. In the current book, we limit ourselves to approximate methods that are appropriate to the situation where rotational uncertainty is not too high.

Although they are not estimation books per se, it is worth mentioning *Optimization on Matrix Manifolds* by Absil et al. (2009) and *An Introduction to Optimization on Smooth Manifolds* by Boumal (2022), which discuss optimization problems where the quantity being optimized is not necessarily a vector, a concept that is quite relevant to robotics because rotations do not behave like vectors (they form a Lie group).

The current book is somewhat unique in focusing only on state estimation and working out the details of common three-dimensional robotics problems in enough detail to be easily implemented for many practical situations.

CAMBRIDGE

Cambridge University Press & Assessment
978-1-009-29989-3 — State Estimation for Robotics
Timothy D. Barfoot
Excerpt
More Information

Part I

Estimation Machinery

Cambridge University Press & Assessment
978-1-009-29989-3 — State Estimation for Robotics
Timothy D. Barfoot
Excerpt
More Information

2

Primer on Probability Theory

In what is to follow, we will be using a number of basic concepts from probability and statistics. This chapter serves to provide a review of these concepts. For a classic book on probability and random processes, see Papoulis (1965). For a light read on the history of probability theory, Devlin (2008) provides a wonderful introduction; this book also helps to understand the difference between the *frequentist* and *Bayesian* views of probability. We will primarily adopt the latter in our approach to estimation, although this chapter mentions some basic frequentist statistical concepts in passing. We begin by discussing general *probability density functions (PDFs)* and then focus on the specific case of Gaussian PDFs. The chapter concludes by introducing Gaussian processes, the continuous-time version of Gaussian random variables.

2.1 Probability Density Functions

2.1.1 Definitions

We say that a *random variable*, x, is distributed according to a particular PDF. Let $p(x)$ be a PDF for the random variable, x, over the interval $[a, b]$. This is a nonnegative function that satisfies

$$\int_a^b p(x)\, dx = 1. \tag{2.1}$$

That is, it satisfies the *axiom of total probability*. Note that this is *probability density*, not *probability*.

Probability is given by the area under the density function. For example, the probability that x lies between c and d, $\Pr(c \leq x \leq d)$, is given by

$$\Pr(c \leq x \leq d) = \int_c^d p(x)\, dx. \tag{2.2}$$

We will also make use of the *cumulative distribution function (CDF)*[1] on occasion, which is given by

[1] The classical treatment of probability theory starts with CDFs, Kolmogorov's three axioms, and works out the details of probability densities as a consequence of being the derivative of CDFs. As is common in robotics, we will work directly with densities in a Bayesian framework, and therefore we will skip these formalities and present primarily the results we need using densities. We shall be careful to use the term *density*, not *distribution*, as we are working with continuous variables throughout this book.

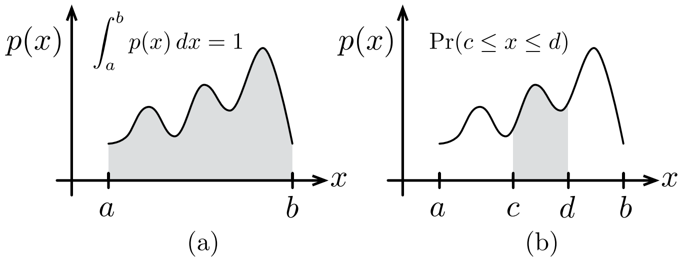


Figure 2.1 Probability density over a finite interval (a). Probability of being within a sub-interval (b).

$$P(x) = \Pr(x' \le x) = \int_{-\infty}^{x} p(x')\,dx', \tag{2.3}$$

the probability that a random variable is less than or equal to x. We have that $P(x)$ is nondecreasing, right-continuous, and $0 \le P(x) \le 1$ with $\lim_{x\to-\infty} P(x) = 0$ and $\lim_{x\to\infty} P(x) = 1$.

Figure 2.1 depicts a general PDF over a finite interval as well as the probability of being within a sub-interval. We will use PDFs to represent the *likelihood* of x being in all possible states in the interval, $[a, b]$, given some evidence in the form of data.

We can also introduce a conditioning variable to PDFs. Let $p(x|y)$ be a PDF over $x \in [a, b]$ conditioned on $y \in [r, s]$ such that

$$(\forall y) \qquad \int_a^b p(x|y)\,dx = 1. \tag{2.4}$$

We may also denote *joint probability densities* for N-dimensional continuous variables in our framework as $p(\mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_N)$ with $x_i \in [a_i, b_i]$. Note that we can also use the notation

$$p(x_1, x_2, \ldots, x_N) \tag{2.5}$$

in place of $p(\mathbf{x})$. Sometimes we even mix and match the two and write

$$p(\mathbf{x}, \mathbf{y}) \tag{2.6}$$

for the joint density of $\mathbf{x}$ and $\mathbf{y}$. In the N-dimensional case, the axiom of total probability requires

$$\int_{\mathbf{a}}^{\mathbf{b}} p(\mathbf{x})\,d\mathbf{x} = \int_{a_N}^{b_N} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} p\left(x_1, x_2, \ldots, x_N\right)\,dx_1\,dx_2 \cdots dx_N = 1, \tag{2.7}$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_N)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_N)$. In what follows, we will sometimes simplify notation by leaving out the integration limits, $\mathbf{a}$ and $\mathbf{b}$.

2.1.2 Marginalization, Bayes' Rule, Inference

We can always factor a joint probability density into a conditional and a unconditional factor:[2]

[2] In the specific case that $\mathbf{x}$ and $\mathbf{y}$ are *statistically independent*, we can factor the joint density as $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$.

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

This one statement has important ramifications.

First, the process of integrating[3] out one or more variables from a joint density, $p(\mathbf{x}, \mathbf{y})$, is called *marginalization*. For example, integrating the joint density over $\mathbf{x}$ reveals

$$\int p(\mathbf{x}, \mathbf{y})\, d\mathbf{x} = \int p(\mathbf{x}|\mathbf{y})p(\mathbf{y})\, d\mathbf{x} = \underbrace{\int p(\mathbf{x}|\mathbf{y})\, d\mathbf{x}}_{1}\, p(\mathbf{y}) = p(\mathbf{y}). \tag{2.9}$$

The result, $p(\mathbf{y})$, is the *marginal* of the joint density for $\mathbf{y}$. Clearly then, the marginal for $\mathbf{x}$ is $p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{y})\, d\mathbf{y}$.

Second, rearranging (2.8) gives *Bayes' rule* (aka Bayes' theorem):

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \tag{2.10}$$

Thomas Bayes (1701–1761) was an English statistician, philosopher and Presbyterian minister, known for having formulated a specific case of the theorem that bears his name. Bayes never published what would eventually become his most famous accomplishment; his notes were edited and published after his death by Richard Price (Bayes, 1764).

We can use this to *infer* the *posterior* or likelihood of the state given some measurements, $p(\mathbf{x}|\mathbf{y})$, if we have a *prior* PDF over the state, $p(\mathbf{x})$, and the sensor model, $p(\mathbf{y}|\mathbf{x})$. We do this by expanding the denominator so that

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

We compute the denominator, $p(\mathbf{y})$, by marginalization as follows:

$$p(\mathbf{y}) = \int p(\mathbf{x}, \mathbf{y})\, d\mathbf{x} = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\, d\mathbf{x}, \tag{2.12}$$

which can be quite expensive to calculate in the general case. In Bayesian inference, $p(\mathbf{x})$ is known as the *prior* density, while $p(\mathbf{x}|\mathbf{y})$ is known as the *posterior* density. Thus, all a priori information is encapsulated in $p(\mathbf{x})$, while $p(\mathbf{x}|\mathbf{y})$ contains the a posteriori information.

2.1.3 Expectations and Moments

The *expectation operator*, $E[\cdot]$, is an important tool when working with probabilities. It allows us to work out the 'average' value of a function of a random variable, $f(\mathbf{x})$, and is defined as

$$E[f(\mathbf{x})] = \int f(\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}, \tag{2.13}$$

where $p(\mathbf{x})$ is the PDF for the random variable, $\mathbf{x}$, and the integration is assumed to be over the domain of $\mathbf{x}$. This is sometimes referred to as the *law of the unconscious statistician (LOTUS)*.[4]

[3] When integration limits are not stated, they are assumed to be over the entire allowable domain of the variable; e.g., $\mathbf{x}$ from $\mathbf{a}$ to $\mathbf{b}$.

[4] LOTUS is named so due to the fact that many practitioners apply (2.13) as a definition without realizing that it requires a rigorous proof.