2 Preliminaries

This chapter provides a succinct review of key concepts from probability and stochastic processes, set theory (with a focus on sets in Euclidean space), and linear dynamical systems that appear throughout the book. The treatment of each subject is far from being complete and is mostly included as a reference for the developments in later chapters. The reader is expected to have certain mathematical maturity and already be familiar with the material as the majority of the results are presented without deriving them. The material presented builds on several concepts from matrix and vector analysis, all of which are briefly reviewed in Appendix A.

2.1 Probability and Stochastic Processes

2.1.1 Probability Spaces

An experiment whose outcome is uncertain can be mathematically described by a so-called probability space, whose components are:

P1. the sample space, denoted by Ω, which is the set of all possible outcomes of the experiment;

P2. an event algebra, denoted by $\mathcal{F}$, which is a set whose elements, referred to as events, are subsets of the sample space chosen so that (i) $\Omega \in \mathcal{F}$, (ii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ (clearly $\emptyset \in \mathcal{F}$ because $\Omega \in \mathcal{F}$), and (iii) if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_i A_i \in \mathcal{F}$; and

P3. a probability measure, which assigns each event $A \in \mathcal{F}$ a number $\Pr(A)$, referred to as the probability of event A, so that: (i) $\Pr(A) \geq 0$ for any $A \in \mathcal{F}$, (ii) $\Pr\left(\bigcup_i A_i\right) = \sum_i \Pr(A_i)$, for any events $A_1, A_2, \ldots$, such that $A_i \cap A_j = \emptyset,\ i \neq j$, and (iii) $\Pr(\Omega) = 1$.

Conditional Probability and Independence

We say an event A has occurred if the outcome of the experiment is within the set A. Given events A, B, the conditional probability of event A given that event B has occurred is defined as follows:

$$\Pr(A \,|\, B) = \begin{cases} \frac{\Pr(A \cap B)}{\Pr(B)}, & \text{if } \Pr(B) > 0, \\ \text{undefined}, & \text{if } \Pr(B) = 0. \end{cases} \tag{2.1}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

A similar expression can be obtained for $\Pr(B \mid A)$, i.e., the conditional probability of event B given event A has occurred.

Events A, B are said to be independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$; thus, if $\Pr(B) > 0$, we have that $\Pr(A \mid B) = \Pr(A)$, i.e., knowing event B has occurred does not yield additional information about event A. Similarly, if events A and B are independent and $\Pr(A) > 0$, we have that $\Pr(B \mid A) = \Pr(B)$, which is to say that knowing event A has occurred does not give us any additional information about event B. Events A, B, C are said to be pairwise independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$, $\Pr(A \cap C) = \Pr(A)\Pr(C)$, and $\Pr(B \cap C) = \Pr(B)\Pr(C)$. Events A, B, C are said to be independent if they are pairwise independent and

$$\Pr(A \cap B \cap C) = \Pr(A)\Pr(B)\Pr(C).$$

Bayes' Formula and the Law of Total Probability

Given events A, B, if both $\Pr(A) > 0$ and $\Pr(B) > 0$, then $\Pr(A \cap B) = \Pr(A \mid B)\ \Pr(B) = \Pr(B \mid A)\Pr(A)$; this leads to the following expression

$$\Pr(B \mid A) = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)},$$

which is known as Bayes' formula.

Events $E_1, E_2, \ldots, E_n$ are said to be mutually exclusive if $E_i \cap E_j = \emptyset,\ i \neq j$. Events $E_1, E_2, \ldots, E_n$ are said to form a partition of Ω if they are mutually exclusive and $\bigcup_{i=1}^{n} E_i = \Omega$. For any event A and any partition of the sample space, $E_1, E_2, \ldots E_n$, satisfying $\Pr(E_i) > 0,\ i = 1, 2, \ldots, n$, we have that

$$\Pr(A) = \Pr(A \mid E_1)\Pr(E_1) + \Pr(A \mid E_2)\Pr(E_2) + \cdots + \Pr(A \mid E_n)\Pr(E_n);$$

this is known as the law of total probability.

By using Bayes' formula and the law of total probability one can check that, for any event A and any partition of the sample space, $E_1, E_2, \ldots, E_n$, satisfying $\Pr(E_i) > 0, i = 1, 2, \ldots, n$, the following holds true:

$$\Pr(E_i \mid A) = \frac{\Pr(A \mid E_i)\Pr(E_i)}{\Pr(A \mid E_1)\Pr(E_1) + \Pr(A \mid E_2)\Pr(E_2) + \cdots + \Pr(A \mid E_n)\Pr(E_n)},$$

$i = 1, 2, \ldots, n$.

2.1.2 Random Variables

Given a probability space $(\Omega, \mathcal{F}, \Pr)$, a random variable is a function $X(\cdot)$ that maps each outcome $\omega \in \Omega$ to a real number, while satisfying that

$$\{\omega \in \Omega \colon X(\omega) \leq x\} \in \mathcal{F}$$

for every $x \in \mathbb{R}$. The values that $X(\cdot)$ takes are referred to as the realizations of the random variable. In order to simplify notation, we write $\{X \leq x\}$ as a shorthand for the event $\{\omega \in \Omega \colon X(\omega) \leq x\}$, and $\Pr(X \leq x)$ as a shorthand for its probability, $\Pr\big(\{\omega \in \Omega \colon X(\omega) \leq x\}\big)$.

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Since $\{\omega \in \Omega\colon X(\omega) \leq x\}$ is an event for every $x \in \mathbb{R}$, we can define the following function:

$$F_X(x) = \Pr(X \leq x), \quad x \in \mathbb{R},$$

which is referred to as the cumulative distribution function (cdf) of X. The cdf of a random variable X, $F_X(\cdot)$, always satisfies the following properties:

C1. It is always nondecreasing.
C2. $\lim_{x\to-\infty} F_X(x) = 0$, and $\lim_{x\to\infty} F_X(x) = 1$.
C3. It is right continuous, i.e., $\lim\limits_{\substack{\epsilon\to 0\\ \epsilon>0}} F_X(x+\epsilon) = F_X(x)$.

Given events $\{X \leq a\}$, $\{X \leq b\}$, $a < b$, we have that

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\},$$

with events $\{X \leq a\}$ and $\{a < X \leq b\}$ being mutually exclusive; thus, we can write $\Pr(X \leq b) = \Pr(X \leq a) + \Pr(a < X \leq b)$, and by using the definition of cdf, we obtain that

$$\Pr(a < X \leq b) = F_X(b) - F_X(a).$$

Discrete Random Variables

A random variable X is said to be discrete if it takes values in a finite or countable set $\mathcal{X} = \{x_1, x_2, \ldots\}$, $x_i \in \mathbb{R}$, (i.e., there is one-to-one correspondence between each element in $\mathcal{X}$ and a natural number) such that

$$\sum_{x_i \in \mathcal{X}} \Pr(X = x_i) = 1. \tag{2.2}$$

For each $x_i \in \mathcal{X}$, let $p_X(x_i)$ denote the probability that X takes value x_i; $p_X(\cdot)$ is referred to as the probability mass function (pmf) of X and in light of (2.2), it must satisfy

$$\sum_{x_i \in \mathcal{X}} p_X(x_i) = 1.$$

The expectation, or first moment, of a discrete random variable X with pmf $p_X(x_i)$, $x_i \in \mathcal{X}$, which we denote by $\mathrm{E}[X]$ and sometimes by μ_X, is defined as

$$\mathrm{E}[X] = \sum_{x_i \in \mathcal{X}} x_i p_X(x_i). \tag{2.3}$$

Given a discrete random variable X with its pmf taking values $p_X(x_i)$, $x_i \in \mathcal{X}$, and a function $g\colon \mathbb{R} \to \mathbb{R}$, we have that

$$\mathrm{E}\big[g(X)\big] = \sum_{x_i \in \mathcal{X}} g(x_i) p_X(x_i). \tag{2.4}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

The variance of a discrete random variable X with its pmf taking values $p_X(x_i), x_i \in \mathcal{X}$, which we denote by σ_X^2, is defined as

$$\begin{aligned}\sigma_X^2 &= \mathrm{E}\big[(X-\mu_X)^2\big] \\ &= \sum_{x_i \in \mathcal{X}} (x_i - \mu_X)^2 p_X(x_i). \end{aligned} \tag{2.5}$$

By using simple manipulations, one can check that

$$\sigma_X^2 = \mathrm{E}\big[(X-\mu_X)^2\big] = \mathrm{E}\big[X^2\big] - \mu_X^2.$$

The term $\mathrm{E}\big[X^2\big]$ is referred to as the second moment of X. More generally, $\mathrm{E}[X^n], n = 1, 2, \ldots$, is referred to as the n^{th} moment of X. The square root of the variance, σ_X, is referred to as the standard deviation of X.

Example 2.1 (Bernoulli distribution) A discrete random variable X with pmf $p_X(\cdot)$ defined as follows

$$p_X(x) = \begin{cases} p, & \text{if } x = 1, \\ 1-p, & \text{if } x = 0, \end{cases}$$

where $0 \le p \le 1$, is said to have a Bernoulli distribution with parameter p. By using (2.3) and (2.5), one can check that $\mu_X = p$ and $\sigma_X^2 = p(1-p)$.

Example 2.2 (Binomial distribution) A discrete random variable X with pmf $p_X(\cdot)$ defined as follows

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n,$$

where $0 \le p \le 1$, is said to have a Binomial distribution with parameters n and p. By using (2.3) and (2.5), one can check that $\mu_X = np$ and $\sigma_X^2 = np(1-p)$.

Example 2.3 (Geometric distribution) A discrete random variable X with pmf $p_X(\cdot)$ defined as follows

$$p_X(x) = (1-p)^{x-1} p, \quad x = 1, 2, \ldots,$$

where $0 \le p \le 1$, is said to have a Geometric distribution with parameter p. By using (2.3) and (2.5), one can check that $\mu_X = \frac{1}{p}$ and $\sigma_X^2 = \frac{1-p}{p^2}$.

Example 2.4 (Poisson distribution) A discrete random variable X with pmf $p_X(\cdot)$ defined as follows

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, \ldots,$$

where $\lambda \ge 0$, is said to have a Poisson distribution with parameter λ. By using (2.3) and (2.5), one can check that $\mu_X = \lambda$ and $\sigma_X^2 = \lambda$.

---

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Continuous Random Variables

A random variable X is said to be continuous if there exists some function $f_X : \mathbb{R} \to [0, \infty)$, referred to as the probability density function (pdf) of X, such that

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy;$$

thus, because of Property C2 above, it follows that

$$\int_{-\infty}^{\infty} f_X(x)dx = 1.$$

If the pdf of a continuous random variable X, $f_X(\cdot)$, is continuous, then its cdf, $F_X(\cdot)$, is continuously differentiable and

$$f_X(x) = \frac{dF_X(x)}{dx}. \tag{2.6}$$

In this remainder, we only consider continuous random variables whose pdf is piecewise continuous with a finite or countable number of discontinuity points; thus, except for those points at which the pdf is discontinuous, (2.6) holds.

The expectation, or first moment, of a continuous random variable with pdf $f_X(\cdot)$ is defined as

$$\mu_X = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx. \tag{2.7}$$

Given a continuous random variable X with pdf $f_X(x)$, we have that

$$\mathrm{E}\big[g(X)\big] = \int_{-\infty}^{\infty} g(x) f_X(x)dx; \tag{2.8}$$

this result is analogous to that in (2.4).

The variance of a continuous random variable with pdf $f_X(\cdot)$ is defined as

$$\begin{aligned}\sigma_X^2 &= \mathrm{E}[(X - \mu_X)^2] \\ &= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)dx.\end{aligned} \tag{2.9}$$

As in the discrete case, we have that

$$\sigma_X^2 = \mathrm{E}[(X - \mu_X)^2] = \mathrm{E}[X^2] - \mu_X^2.$$

Example 2.5 (Uniform distribution) A continuous random variable X with pdf

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a < x \le b, \\ 0, & \text{otherwise,} \end{cases}$$

$a, b \in \mathbb{R}$, $a < b$, is said to have a Uniform distribution on the interval $[a, b]$. One can check by using (2.7) and (2.9) that $\mu_X = \frac{a+b}{2}$ and $\sigma_X^2 = \frac{(b-a)^2}{12}$.

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Example 2.6 (Exponential distribution) A continuous random variable X with pdf

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

$\lambda > 0$, is said to have an Exponential distribution with parameter λ. One can check by using (2.7) and (2.9) that $\mu_X = \frac{1}{\lambda}$ and $\sigma_X^2 = \lambda^2$.

Example 2.7 (Gaussian distribution) A continuous random variable X with pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}, \quad -\infty \leq x \leq \infty,$$

$\mu \in \mathbb{R}$, $\sigma > 0$, is said to have a Gaussian (or Normal) distribution with parameters μ_X and σ_X^2. One can check by using (2.7) and (2.9) that indeed μ_X and σ_X^2 are respectively the mean and variance of X.

2.1.3 Jointly Distributed Random Variables

Given two random variables X and Y defined on the same probability space $(\Omega, \mathcal{F}, \Pr)$ and any $x, y \in \mathbb{R}$, we can define an event of the form

$$\{\omega \in \Omega \colon X(\omega) \leq x,\ Y(\omega) \leq y\},$$

and as before, we will write $\Pr(X \leq x, Y \leq y)$ as a shorthand for its probability. Then, we can define the function

$$F_{X,Y}(x, y) = \Pr(X \leq x, Y \leq y),$$

$x, y \in \mathbb{R}$, which we refer to as the joint cdf of X and Y. In the remainder of this section, we mostly focus on continuous random variables; however, there are counterpart notions for discrete random variables to the majority of concepts introduced.

Joint and Marginal pdfs

Given two continuous random variables X, Y, there exists a function $f_{X,Y}(\cdot, \cdot)$, referred to as the joint probability density function of X and Y, such that

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) du dv,$$

and, similar to the single continuous random variable case, we have that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1.$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Given two continuous random variables with joint pdf $f_{X,Y}(\cdot,\cdot)$, the marginal pdf $f_X(\cdot)$ is the pdf of X when considered by itself, and can be obtained as follows:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

One can similarly obtain an expression for the marginal pdf $f_Y(\cdot)$.

Example 2.8 Consider two continuous random variables X, Y with joint pdf as follows:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2}, & \text{if } |x|+|y| \le 1, \\ 0, & \text{otherwise;} \end{cases} \tag{2.10}$$

see Fig. 2.1 (left) for a graphical representation of the support of $f_{X,Y}(\cdot,\cdot)$. The marginal pdf $f_X(x)$ can be computed as follows. For $0 \le x \le 1$, we have that $f_{X,Y}(x,y) = \frac{1}{2}$ if $-1+x \le y \le 1-x$ and $f_{X,Y}(x,y) = 0$ otherwise; thus

$$\begin{aligned} f_X(x) &= \int_{-1+x}^{1-x} \frac{1}{2} dy \\ &= 1-x, \quad 0 \le x \le 1. \end{aligned}$$

A similar argument yields

$$f_X(x) = 1+x, \qquad -1 \le x < 0.$$

Finally, since $f_{X,Y}(x,y) = 0$ when $x > 1$ or $x < -1$, we have that $f_X(x) = 0$ for $x > 1$ or $x < -1$. Putting together the results above yields

$$f_X(x) = \begin{cases} 1-|x|, & \text{if } -1 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.11}$$

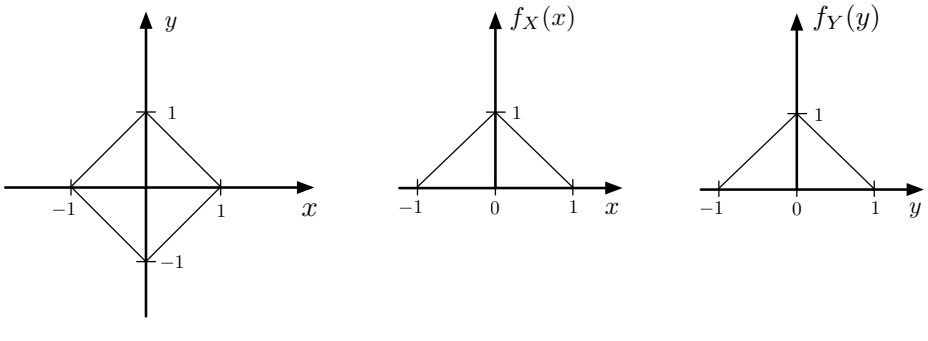

Figure 2.1 Example 2.8: support of joint pdf $f_{X,Y}(\cdot,\cdot)$ (left), marginal pdf $f_X(\cdot)$ (center), and marginal pdf $f_Y(\cdot)$ (right).

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Because of symmetry, it is easy to see that

$$f_Y(y) = \begin{cases} 1 - |y|, & \text{if } -1 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.12}$$

The graphs of $f_X(\cdot)$ and $f_Y(\cdot)$ are displayed in Fig. 2.1 (center) and Fig. 2.1 (right), respectively.

Conditional pdfs and Independence

As we did with events, if X and Y are continuous random variables, we can define the conditional pdf of X given $Y = y$ such that $f_Y(y) > 0$, which we denote by $f_{X\,|\,Y}(\cdot\,|\,y)$, as follows:

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}. \tag{2.13}$$

One can also write a similar expression for the conditional pdf of Y given $X = x$.

Two jointly distributed continuous random variables X and Y are said to be independent if we can factor their joint pdf as the product of the marginal pdfs, i.e.,

$$f_{X,Y}(x,y) = f_X(x) f_Y(y), \tag{2.14}$$

for all x and y. Then, by plugging (2.14) into (2.13), we obtain

$$f_{X\,|\,Y}(x\,|\,y) = f_X(x),$$

for all x and all y such that $f_Y(y) > 0$.

Example 2.9 We continue with Example 2.8 and compute the conditional pdf of X given Y. Recall from (2.12) that $f_Y(y) = 1 - |y| > 0$ if $-1 < y < 1$, and $f_Y(y) = 0$ otherwise; thus, $f_{X\,|\,Y}(x\,|\,y)$ is only defined if $y \in (-1, 1)$. Also, recall from (2.10) that for any $y \in (-1, 1)$, we have that $f_{X,Y}(x,y) = 1/2$ if $|x| \le 1 - |y|$, and $f_{X,Y}(x,y) = 0$ otherwise; thus,

$$f_{X\,|\,Y}(x\,|\,y) = \begin{cases} \frac{1}{2(1-|y|)}, & \text{if } -1 + |y| \le x \le 1 - |y|, \\ 0, & \text{otherwise,} \end{cases} \tag{2.15}$$

for any $y \in (-1, 1)$, i.e., given $Y = y$, $y \in (-1, 1)$, the distribution of X on the interval $\big[-1+|y|, 1-|y|\big]$ is uniform. The graph of $f_{X\,|\,Y}(\cdot\,|\,y_0)$ for some $y_0 \in (0, 1)$ is displayed in Fig. 2.2, where one can visualize the relation between the support of $f_{X\,|\,Y}(\cdot\,|\,y_0)$ and the shape of the support of the joint pdf $f_{X,Y}(x,y)$. By inspection of (2.10), (2.11), and (2.12), it is clear that $f_{X,Y}(x,y) \neq f_X(x) f_Y(y)$; thus, the random variables X and Y are not independent.

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

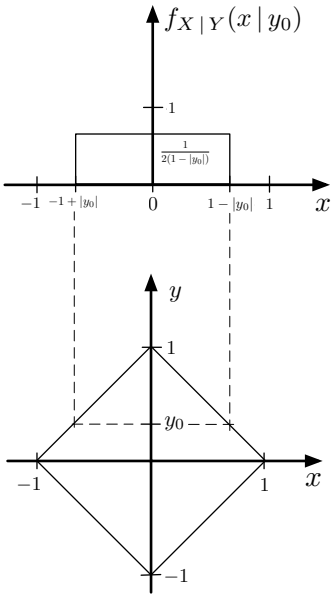

Figure 2.2 Example 2.9: conditional pdf $f_{X\,|\,Y}(\cdot\,|\,y_0)$ for some $y_0 \in (0,1)$ (top), and support of joint pdf $f_{X,Y}(\cdot,\cdot)$ (bottom).

Expectation, Covariance, and Correlation

Given two jointly distributed continuous random variables with joint pdf $f_{X,Y}(x,y)$, and a function $g\colon \mathbb{R}\times\mathbb{R}\to\mathbb{R}$, we have that

$$\mathrm{E}\big[g(X,Y)\big] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy; \tag{2.16}$$

this result is similar in spirit to the one in (2.8) for a single random variable and a real-valued function.

If X and Y are independent, then we have that

$$\mathrm{E}\big[XY\big] = \mathrm{E}\big[X\big]\mathrm{E}\big[Y\big].$$

To see this, we use (2.16) and (2.14) and proceed as follows:

$$\begin{aligned}
\mathrm{E}\big[XY\big] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y)\,dx\,dy \\
&= \underbrace{\int_{-\infty}^{\infty} x f_X(x)\,dx}_{\mathrm{E}[X]} \underbrace{\int_{-\infty}^{\infty} y f_Y(y)\,dy}_{\mathrm{E}[Y]}.
\end{aligned}$$

Given two jointly distributed continuous random variables X, Y, the conditional expectation of X given $Y = y$ is as follows:

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

$$\mathrm{E}\big[X \,|\, Y = y\big] = \int_{-\infty}^{\infty} x f_{X\,|\,Y}(x\,|\,y)dx, \tag{2.17}$$

which depends on y; thus, there exists some real-valued function $h(\cdot)$ such that

$$\begin{aligned} h(y) &:= \mathrm{E}\big[X \,|\, Y = y\big] \\ &= \int_{-\infty}^{\infty} x f_{X\,|\,Y}(x\,|\,y)dx. \end{aligned}$$

Then, $h(Y)$ is another random variable, denoted by $\mathrm{E}\big[X \,|\, Y\big]$, and referred to as the conditional expectation of X given Y. Now, by using (2.8), we have that

$$\begin{aligned} \mathrm{E}\Big[\mathrm{E}\big[X \,|\, Y\big]\Big] &= \mathrm{E}\big[h(Y)\big] \\ &= \int_{-\infty}^{\infty} h(y) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X\,|\,Y}(x\,|y)dx \right) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} \underbrace{f_{X\,|\,Y}(x\,|y) f_Y(y)}_{= f_{X,Y}(x,y)} dy \right) dx \\ &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= \mathrm{E}[X]. \end{aligned} \tag{2.18}$$

The covariance of two jointly distributed continuous random variables X, Y, which we denote by $c_{X,Y}$, is defined as follows:

$$\begin{aligned} c_{X,Y} &= \mathrm{E}\big[(X - \mu_X)(Y - \mu_Y)\big] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f_{X,Y}(x,y)dxdy, \end{aligned}$$

and similarly to the variance of a single random variable, one can check that

$$\begin{aligned} c_{X,Y} &= \mathrm{E}[XY] - \mu_X \mu_Y \\ &= \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x,y)dxdy}_{=:\, r_{X,Y}} - \mu_X \mu_Y. \end{aligned}$$

The term $r_{X,Y} := \mathrm{E}[XY]$ is referred to as the correlation of X and Y. Random variables X and Y are said to be uncorrelated when $c_{X,Y} = 0$. If this is the case, we have that

$$r_{X,Y} = \mu_X \mu_Y;$$

the converse is also true, i.e., if $r_{X,Y} = \mu_X \mu_Y$, then X and Y are uncorrelated. Recall that if two random variables X and Y are independent, then

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

$\mathrm{E}[XY] = \mu_X \mu_Y$; thus, if X and Y are independent, they are also uncorrelated. The converse is not true in general, i.e., if X and Y are uncorrelated, it does not imply that they are independent unless both X and Y are jointly distributed Gaussian random variables.

Example 2.10 We continue with Example 2.9. By plugging (2.15) into (2.17), we obtain

$$\begin{aligned}\mathrm{E}\big[X \,|\, Y = y\big] &= \int_{-1+|y|}^{1-|y|} x \frac{1}{2(1-|y|)} dx \\ &= 0 \\ &=: h(y), \quad -1 \le y \le 1. \end{aligned} \tag{2.19}$$

Then, by plugging (2.12) and (2.19) into (2.18), we obtain

$$\begin{aligned}\mathrm{E}\big[X\big] &= \mathrm{E}\Big[\mathrm{E}\big[X \,|\, Y\big]\Big] \\ &= \mathrm{E}\big[h(Y)\big] \\ &= 0, \end{aligned} \tag{2.20}$$

which matches the value of $\mathrm{E}[X]$ computed by using (2.11):

$$\begin{aligned}\mathrm{E}\big[X\big] &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_{-1}^{1} x\big(1-|x|\big) dx \\ &= \int_{-1}^{0} x(1+x) dx + \int_{0}^{1} x(1-x) dx \\ &= 0, \end{aligned} \tag{2.21}$$

as expected. Similar calculation yields $\mathrm{E}[Y] = 0$. Thus,

$$\begin{aligned} c_{X,Y} &= r_{X,Y} \\ &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} y f_{X,Y}(x,y) dy \right) dx \\ &= \frac{1}{2} \int_{-1}^{0} x \left(\int_{-(1+x)}^{1+x} y dy \right) dx + \frac{1}{2} \int_{0}^{1} x \left(\int_{-(1-x)}^{1-x} y dy \right) dx \\ &= 0, \end{aligned} \tag{2.22}$$

therefore, X and Y are uncorrelated; however, they are not independent as we established in Example 2.9.

---

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

2.1.4 Random Vectors

An n-dimensional random vector X is an n-tuple whose components, denoted by $X_1, X_2, \ldots, X_n$, are random variables all defined on the same probability space, $(\Omega, \mathcal{F}, \Pr)$. Because the X_i's take values in $\mathbb{R}$, the values that X takes, denoted by x, are real vectors in $\mathbb{R}^n$. (Unless otherwise stated, in this book we adopt the convention that the components of a vector are arranged in a column format.)

Now, for any $x = [x_1, x_2, \ldots, x_n]^\top \in \mathbb{R}^n$, we can define events of the form:

$$\{\omega \in \Omega \colon X_1(\omega) \leq x_1,\ X_2(\omega) \leq x_2, \ldots, X_n(\omega) \leq x_n\}, \tag{2.23}$$

and as before, we write $\Pr(X_1 \leq x_1,\ X_2 \leq x_2, \ldots, X_n \leq x_n)$ as a shorthand for their probability. Then, we can define a function $F_X \colon \mathbb{R}^n \to [0,1]$ as follows:

$$F_X(x_1, x_2, \ldots, x_n) = \Pr(X_1 \leq x_1,\ X_2 \leq x_2, \ldots, X_n \leq x_n);$$

this function is referred to as the joint cdf of the components of X.

If the components of an n-dimensional random vector X are jointly distributed discrete random variables, then X will take values in a finite or countable set $\mathcal{X} \subset \mathbb{R}^n$, and we refer to X as a discrete random vector. Thus, as in the case of a single random variable, the pmf of the random vector X, denoted by $p_X(\cdot, \cdot, \ldots, \cdot)$, is defined as follows:

$$p_X(x_1, x_2, \ldots, x_n) = \Pr(X_1 = x_1,\ X_2 = x_2, \ldots, X_n = x_n),$$

$x = [x_1, x_2, \ldots, x_n]^\top \in \mathcal{X}$, which satisfies $\sum_{x \in \mathcal{X}} p_X(x_1, x_2, \ldots, x_n) = 1$.

We can write the set $\mathcal{X} \subset \mathbb{R}^n$ containing the values taken by a discrete random vector $X = [X_1, X_2, \ldots, X_n]^\top$ as follows:

$$\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n \subset \mathbb{R}^n,$$

where $\mathcal{X}_i \subset \mathbb{R},\ i = 1, 2, \ldots, n$, is a countable set containing the values that X_i can take. Then, we can define the marginal pmf of X_i, denoted by $p_{X_i}(\cdot)$, as follows:

$$p_{X_i}(x_i) = \sum_{x_j \in \mathcal{X}_j, j \neq i} p_X(x_1, x_2, \ldots, x_i, \ldots, x_n).$$

For the case when the components of X are jointly distributed continuous random variables, in which case we refer to X as a continuous random vector, there exists some function $f_X \colon \mathbb{R}^n \to [0, \infty)$, referred to as the joint pdf of the components of X, such that

$$F_X(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f_X(y_1, y_2, \ldots, y_n) dy_1 dy_2 \ldots dy_n,$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_X(x_1, x_2, \ldots, x_n)dx_1dx_2\ldots dx_n = 1.$$

Given an n-dimensional continuous random vector X with pdf $f_X(\cdot)$, the marginal pdf of its i^{th} component, denoted by $f_{X_i}(\cdot)$, is defined as follows:

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_X(x_1, x_2, \ldots, x_i, \ldots, x_n)dx_1dx_2\ldots dx_{i-1}dx_{i+1}\ldots dx_n.$$

Conditional pdfs and Independence

Given two random vectors $X = [X_1, X_2, \ldots, X_n]^\top$ and $Y = [Y_1, Y_2, \ldots, Y_m]^\top$ defined on the same probability space, $(\Omega, \mathcal{F}, \Pr)$, we can define their joint cdf, which we denote by $F_{X,Y}(\cdot,\cdot)$, by using a similar approach to the one above used for defining the joint pdf of the components of a random vector X. Similarly, whether X and Y are discrete or continuous random vectors, we can also define their joint pmf and pdf, which we denote by $p_{X,Y}(\cdot,\cdot)$ and $f_{X,Y}(\cdot,\cdot)$, respectively, and the marginal pmf and pdf of X (or Y), which we denote by $p_X(x)$ $\big($or $p_Y(y)\big)$ and $f_X(x)$ $\big($or $f_Y(y)\big)$, respectively.

As we did with events, if X and Y are discrete random vectors, we can define the conditional pmf of X given Y, denoted by $p_{X\,|\,Y}(\cdot\,|\,\cdot)$, as follows:

$$p_{X\,|\,Y}(x\,|\,y) = \frac{p_{X,Y}(x,y)}{p_Y(x)},$$

and if

$$p_{X,Y}(x,y) = p_X(x)p_Y(y),$$

we say they are independent. Similarly, if X and Y are continuous random vectors, we can define the conditional pdf of X given Y, denoted by $f_{X\,|\,Y}(\cdot\,|\,\cdot)$, as follows:

$$f_{X\,|\,Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(x)},$$

and if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y),$$

we say they are independent.

Expectation, Covariance, and Correlation

The mean of a random vector $X = [X_1, X_2, \ldots, X_n]^\top$, denoted by $\mathrm{E}[X]$ or m_X, is a vector in $\mathbb{R}^n$ whose components are the expectations of the components of X, i.e.,

$$\mathrm{E}[X] = \Big[\mathrm{E}[X_1], \mathrm{E}[X_2], \ldots, \mathrm{E}[X_n]\Big]^\top. \tag{2.24}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

The covariance matrix of a random vector $X = [X_1, X_2, \ldots, X_n]^\top$, denoted by $\mathrm{E}\big[(X - m_X)(X - m_X)^\top\big]$ or Σ_X, is defined as:

$$\big[\Sigma_X\big]_{i,j} = \mathrm{E}\big[(X_i - m_{X_i})(X_j - m_{X_j})\big],$$

$i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$. Similarly, the correlation matrix of a random vector $X = [X_1, X_2, \ldots, X_n]^\top$, denoted by $\mathrm{E}\big[XX^\top\big]$ or S_X, is defined as:

$$\big[S_X\big]_{i,j} = \mathrm{E}\big[X_i X_j\big],$$

$i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$. One can check that the relation between the covariance and correlation matrices of X is given by:

$$\Sigma_X = S_X - m_X m_X^\top.$$

Given two random vectors, $X = [X_1, X_2, \ldots, X_n]^\top$ and $Y = [Y_1, Y_2, \ldots, Y_n]^\top$, defined on the same probability space, we can define their covariance matrix, which we denote by $\mathrm{E}\big[(X - m_X)(Y - m_Y)^\top\big]$ or $C_{X,Y}$, as follows:

$$\big[C_{X,Y}\big]_{i,j} = \mathrm{E}\Big[\big(X_i - \mathrm{E}[X_i]\big)\big(Y_j - \mathrm{E}[Y_j]\big)\Big], \tag{2.25}$$

$i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$. Similarly, we can define their correlation matrix, which we denote by $\mathrm{E}\big[XY^\top\big]$ or $R_{X,Y}$, as follows:

$$\big[R_{X,Y}\big]_{i,j} = \mathrm{E}\big[X_i Y_j\big], \tag{2.26}$$

$i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$. For the case when $Y = X$, we clearly have $C_{X,X} = \Sigma_X$ and $R_{X,X} = S_X$. Given two random vectors X and Y, one can check that the relation between their covariance and correlation matrices is given by:

$$C_{X,Y} = R_{X,Y} - m_X m_Y^\top.$$

2.1.5 Stochastic Processes

An n-dimensional stochastic process X is a collection of n-dimensional random vectors indexed in some set $\mathcal{T}$, all of which are defined on the same probability space $(\Omega, \mathcal{F}, \Pr)$. When $\mathcal{T} = \{0, 1, 2, \ldots\}$, X is called a vector-valued discrete-time stochastic process, and we typically write $X = \big\{X_k \colon k \in \mathcal{T}\big\}$, $\mathcal{T} = \{0, 1, 2, \ldots\}$, or $X = \{X_k : k \geq 0\}$ to represent it. When $\mathcal{T} = [0, \infty)$, X is called a vector-valued continuous-time stochastic process, and we typically write $X = \big\{X(t) : t \in \mathcal{T}\big\}, \mathcal{T} = [0, \infty)$, or $X = \big\{X(t) : t \geq 0\big\}$ to represent it.

To completely characterize a vector-valued discrete-time stochastic process, $X = \{X_k \colon k \geq 0\}$, it is necessary to obtain the joint cdf of the random vectors $X_{k_1}, X_{k_2}, \ldots, X_{k_\ell}$,

$$F_{X_{k_1}, X_{k_2}, \ldots, X_{k_\ell}}(x_1, x_2, \ldots, x_\ell),$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

for all ℓ and any $k_1, k_2, \ldots, k_\ell$. Similarly, to fully characterize a vector-valued continuous-time stochastic process, $X = \{X(t) : t \geq 0\}$, it is necessary to obtain the joint cdf of the random vectors $X(t_1), X(t_2), \ldots, X(t_\ell)$,

$$F_{X(t_1),X(t_2),\ldots,X(t_\ell)}(x_1, x_2, \ldots, x_\ell),$$

for all ℓ and all $t_1, t_2, \ldots, t_\ell$.

Example 2.11 (Bernoulli process) Consider a discrete-time stochastic process $X = \{X_k : k \geq 0\}$, where the X_k's are Bernoulli independent and identically distributed (i.i.d.) random variables with parameter p, i.e., $\Pr(X_k = 1) = p$ and $\Pr(X_k = 0) = 1 - p$; such process is referred to as a Bernoulli process. In this case, instead of characterizing

$$F_{X_{k_1},X_{k_2},\ldots,X_{k_\ell}}(x_1, x_2, \ldots, x_\ell),$$

for all ℓ and any $k_1, k_2, \ldots, k_\ell$, we can equivalently provide a complete description of the stochastic process by providing the joint pmf of $X_{k_1}, X_{k_2}, \ldots, X_{k_\ell}$, $p_{X_{k_1},X_{k_2},\ldots,X_{k_\ell}}(x_1, x_2, \ldots, x_\ell)$, for all ℓ and any $k_1, k_2, \ldots, k_\ell$. First note that

$$\Pr(X_k = x_k) = p^{x_k}(1-p)^{1-x_k},$$

with $x_k \in \{0, 1\}$; then, we have that

$$\begin{aligned} p_{X_{k_1},X_{k_2},\ldots,X_{k_\ell}}(x_1, x_2, \ldots, x_\ell) &= \Pr\left(X_{k_1} = x_1, X_{k_2} = x_2, \ldots, X_{k_\ell} = x_\ell\right) \\ &= \Pr\left(X_{k_1} = x_1\right) \Pr\left(X_{k_2} = x_2\right) \cdots \Pr\left(X_{k_\ell} = x_\ell\right) \\ &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \ldots p^{x_\ell}(1-p)^{1-x_\ell} \\ &= p^{\sum_{k=1}^{\ell} x_k}(1-p)^{\ell - \sum_{k=1}^{\ell} x_k}, \end{aligned} \tag{2.27}$$

with $x_i \in \{0, 1\},\ i = 1, 2, \ldots, \ell$.

Mean, Covariance, and Correlation Functions

Except for specific cases (as in the example above), it is hard in general to obtain a full characterization of a stochastic process. However, in many applications it suffices to characterize the first and second moments of the stochastic process, which we define next.

The first moment, or mean function, of an n-dimensional discrete-time stochastic process X, which we denote by $m_X[\cdot]$, is defined as follows:

$$m_X[k] = \mathrm{E}\left[X_k\right] \in \mathbb{R}^n, \quad k \geq 0. \tag{2.28}$$

Similarly, if X is an n-dimensional continuous-time stochastic process, its mean function, which we denote by $m_X(\cdot)$, is defined as follows:

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

$$m_X(t) = \mathrm{E}\big[X(t)\big] \in \mathbb{R}^n, \quad t \geq 0. \tag{2.29}$$

The covariance function of an n-dimensional discrete-time stochastic process, which we denote by $C_X[\cdot,\cdot]$, is defined as follows:

$$\begin{aligned} C_X[k_1,k_2] &:= C_{X_{k_1},X_{k_2}} \\ &= \mathrm{E}\Big[\big(X_{k_1} - m_X[k_1]\big)\big(X_{k_2} - m_X[k_2]\big)^\top\Big] \in \mathbb{R}^{n\times n}, \quad k_1,k_2 \geq 0. \end{aligned} \tag{2.30}$$

Similarly, if X is an n-dimensional continuous-time stochastic process, its covariance function, which we denote by $C_X(\cdot,\cdot)$, is defined as follows:

$$\begin{aligned} C_X(t_1,t_2) &:= C_{X(t_1),X(t_2)} \\ &= \mathrm{E}\Big[\big(X(t_1) - m_X(t_1)\big)\big(X(t_2) - m_X(t_2)\big)^\top\Big] \in \mathbb{R}^{n\times n}, \quad t_1,t_2 \geq 0. \end{aligned} \tag{2.31}$$

In words, given $k_1,k_2 \geq 0$ $(t_1,t_2 \geq 0)$, the value of the covariance function of X, $C_X[k_1,k_2]$ $\big(C_X(t_1,t_2)\big)$, is equal to the covariance of the random vectors X_{k_1} and X_{k_2} $\big(X(t_1)$ and $X(t_2)\big)$.

The correlation function of an n-dimensional discrete-time stochastic process, which we denote by $R_X[\cdot,\cdot]$, is defined as follows:

$$\begin{aligned} R_X[k_1,k_2] &:= R_{X_{k_1},X_{k_2}} \\ &= \mathrm{E}\big[X_{k_1}X_{k_2}^\top\big] \in \mathbb{R}^{n\times n}, \quad k_1,k_2 \geq 0. \end{aligned} \tag{2.32}$$

Similarly, if X is an n-dimensional continuous-time stochastic process, its covariance function, which we denote by $R_X(\cdot,\cdot)$, is defined as follows:

$$\begin{aligned} R_X(t_1,t_2) &:= R_{X(t_1),X(t_2)} \\ &= \mathrm{E}\big[X(t_1)X^\top(t_2)\big] \in \mathbb{R}^{n\times n}, \quad t_1,t_2 \geq 0. \end{aligned} \tag{2.33}$$

In words, given $k_1,k_2 \geq 0$ $(t_1,t_2 \geq 0)$, the value of the correlation function of X, $R_X[k_1,k_2]$ $\big(R_X(t_1,t_2)\big)$, is equal to the correlation of the random vectors X_{k_1} and X_{k_2} $\big(X(t_1)$ and $X(t_2)\big)$.

The correlation and covariance functions of a stochastic process X are referred to as the second moments of the process, and they are related as follows:

$$C_X[k_1,k_2] = R_X[k_1,k_2] - m_X[k_1]m_X^\top[k_2], \quad k_1,k_2 \geq 0, \tag{2.34}$$

if X is a discrete-time stochastic process, and

$$C_X(t_1,t_2) = R_X(t_1,t_2) - m_X(t_1)m_X^\top(t_2), \quad t_1,t_2 \geq 0, \tag{2.35}$$

if X is a continuous-time stochastic process.

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Example 2.12 (Bernoulli process moments) Consider again the Bernoulli process in Example 2.11. In this case, the mean, covariance, and correlation functions of the process are:

$$\begin{aligned} m_X[k] &= \mathrm{E}\big[X_k\big] \\ &= p, \quad k = 0, 1, 2, \ldots, \end{aligned} \tag{2.36}$$

$$\begin{aligned} C_X[k_1, k_2] &= C_{X_{k_1}, X_{k_2}} \\ &= \mathrm{E}\Big[\big(X_{k_1} - m_X[k_1]\big)\big(X_{k_2} - m_X[k_2]\big)\Big] \\ &= \begin{cases} p(1-p), & \text{if } k_1 = k_2, \\ 0, & \text{if } k_1 \neq k_2, \end{cases} \end{aligned} \tag{2.37}$$

$$\begin{aligned} R_X[k_1, k_2] &= C_X[k_1, k_2] + m_X[k_1] m_X[k_2] \\ &= \begin{cases} p, & \text{if } k_1 = k_2, \\ p^2, & \text{if } k_1 \neq k_2. \end{cases} \end{aligned} \tag{2.38}$$

Stationarity

An n-dimensional discrete-time stochastic process $X = \{X_k : k \geq 0\}$ is said to be strict-sense stationary if

$$F_{X_{k_1}, X_{k_2}, \ldots, X_{k_\ell}}(x_1, x_2, \ldots, x_\ell) = F_{X_{k_1+k'}, X_{k_2+k'}, \ldots, X_{k_\ell+k'}}(x_1, x_2, \ldots, x_\ell)$$

for all ℓ and all $k_1, k_2, \ldots, k_\ell, k' \geq 0$. Similarly, an n-dimensional continuous-time stochastic process $X = \big\{X(t) : t \in [0, \infty)\big\}$ is said to be strict-sense stationary if

$$F_{X(t_1), X(t_2), \ldots, X(t_\ell)}(x_1, x_2, \ldots, x_\ell) = F_{X(t_1+s), X(t_2+s), \ldots, X(t_\ell+s)}(x_1, x_2, \ldots, x_\ell)$$

for all ℓ and all $t_1, t_2, \ldots, t_\ell, s \in [0, \infty)$.

A discrete-time stochastic process $X = \{X_k : k \geq 0\}$ is said to be wide-sense stationary if

$$m_X[k] = m_X[k + k'], \tag{2.39}$$

$$C_X[k_1, k_2] = C_X[k_1 + k', k_2 + k'], \tag{2.40}$$

for all $k, k', k_1, k_2 \geq 0$, i.e., its mean function does not depend on time, and the value of the covariance function for any k_1, k_2, $C_X(k_1, k_2)$, only depends on $k_1 - k_2$. Similarly, a continuous-time stochastic process $X = \{X(t) : t \in [0, \infty)\}$ is said to be wide-sense stationary if

$$m_X(t) = m_X(t + s), \tag{2.41}$$

$$C_X(t_1, t_2) = C_X(t_1 + s, t_2 + s), \tag{2.42}$$

for all $t, s, t_1, t_2 \geq 0$. Any strict-sense stationary process is also wide-sense stationary; the converse is not true in general.

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Example 2.13 Consider again the Bernoulli process of Examples 2.11, 2.12. For any $k, k' \geq 0$, we have that

$$\begin{aligned} m_X[k+k'] &= \mathrm{E}[X_{k+k'}] \\ &= p; \end{aligned} \tag{2.43}$$

which matches the expression for $m_X[k]$ in (2.36). Similarly, for any $k_1, k_2, k' \geq 0$, direct calculation of $\mathrm{E}\big[\big(X_{k_1+k'} - m_X[k_1+k']\big)\big(X_{k_2+k'} - m_X[k_2+k']\big)\big]$ yields

$$C_X[k_1+k', k_2+k'] = \begin{cases} p(1-p), & \text{if } k_1 = k_2, \\ 0, & \text{if } k_1 \neq k_2, \end{cases} \tag{2.44}$$

which matches the expression for $C_X[k_1, k_2]$ in (2.37). Thus, we conclude that the Bernoulli process is wide-sense stationary.

By using a similar procedure to the one used in (2.27), we obtain that

$$p_{X_{k_1+k'}, X_{k_2+k'}, \ldots, X_{k_\ell+k'}}(x_1, x_2, \ldots, x_\ell) = p^{\sum_{k=1}^{\ell} x_k}(1-p)^{\ell - \sum_{k=1}^{\ell} x_k}, \tag{2.45}$$

for any $k_1, k_2, \ldots, k_\ell, k' \geq 0$, which matches the expression for

$$p_{X_{k_1}, X_{k_2}, \ldots, X_{k_\ell}}(x_1, x_2, \ldots, x_\ell)$$

in (2.27); thus, the Bernoulli process is also strict-sense stationary.

The Wiener Process

The Wiener process is a continuous-time stochastic process that will be heavily featured in Chapter 6. It can be used to formally describe the random motion of a tiny particle suspended in water, a phenomenon first observed by the botanist Robert Brown in 1827, and for this reason, a Wiener process is also referred to as Brownian motion process, or Brownian motion. However, it was not until the 1920s that the mathematician Norbert Wiener provided a rigorous mathematical description of the phenomenon (Albert Einstein also studied the problem in 1905). The formal definition of the Wiener process is as follows:

DEFINITION 2.1 Let $W = \big\{W(t) : t \in \mathcal{T} = [0, \infty)\big\}$ denote a real-valued stochastic process. We say W is a *standard Wiener process* if it satisfies the following properties:

B1. $W(0) = 0$.

B2. W has independent increments, i.e., the distribution of $W(t) - W(s)$ depends on $t - s$ alone, and the variables $W(t_j) - W(s_j)$, $j = 1, 2, \ldots, n$, are independent whenever the intervals $(s_j, t_j]$ are disjoint.

B3. $W(s+t) - W(s)$ is normally distributed with zero mean and variance t for all $s, t \geq 0$.

B4. The sample paths of W are continuous.

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Mean, Covariance, and Correlation Function. A standard Wiener process W is a zero-mean process, i.e.,

$$m_W(t) = 0, \tag{2.46}$$

for all $t \geq 0$; this can be easily established as follows. First note that $\mathrm{E}\big[W(0)\big] = 0$ by Property B1; then we have

$$\begin{aligned} m_W(t) &= \mathrm{E}\big[W(t)\big] \\ &= \mathrm{E}\big[W(t)\big] - \mathrm{E}\big[W(0)\big] \\ &= \mathrm{E}\big[W(t) - W(0)\big] \\ &= 0, \end{aligned} \tag{2.47}$$

where the last equality follows from Property B3. In addition, the values taken by the covariance and correlation functions of a Wiener process W, $C_W(t,s),\ t,s \geq 0$, and $R(t,s),\ t,s \geq 0$, respectively, are equal and given by

$$\begin{aligned} C_W(t,s) &= R_W(t,s) \\ &= \min\{t,s\}; \end{aligned} \tag{2.48}$$

this formula can be established as follows. Clearly $C_W(t,s) = R_W(t,s)$ since $m_W(t) = 0$ for all $t \geq 0$. Now, assume $t \geq s$, then by noting Property B1 (i.e., $W(0) = 0$), we have that

$$\begin{aligned} \mathrm{E}\big[W(t)W(s)\big] &= \mathrm{E}\Big[\big(W(t) - W(0)\big)\big(W(s) - W(0)\big)\Big] \\ &= \mathrm{E}\Big[\big(W(s) - W(0)\big)^2 + \big(W(t) - W(s)\big)\big(W(s) - W(0)\big)\Big] \\ &= \mathrm{E}\Big[\big(W(s) - W(0)\big)^2\Big] + \underbrace{\mathrm{E}\Big[\big(W(t) - W(s)\big)\big(W(s) - W(0)\big)\Big]}_{=\,0 \text{ by Properties B2 and B3}} \\ &= \mathrm{E}\Big[\big(W(s) - W(0)\big)^2\Big] \\ &= s, \end{aligned} \tag{2.49}$$

where the last equality follows from Property B3. A similar derivation for the case when $t \leq s$ results in $\mathrm{E}\big[W(t)W(s)\big] = t$; thus, $C_W(t,s) = R_W(t,s) = \min\{t,s\}$.

Probability Distribution. By Property B3, $\Delta W(t) = W(t) - W(0)$ is normally distributed with zero mean and variance t; thus, its pdf, which we denote by $f_{\Delta W}(t,\cdot)$, is given by

$$f_{\Delta W}(t,\Delta w) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{\Delta w^2}{2t}}, \tag{2.50}$$

and since $W(0) = 0$ by Property B1, we have that the pdf of

$$W(t) = W(0) + \Delta W(t),$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

denoted by $f_W(t,\cdot)$, is given by

$$f_W(t,w) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{w^2}{2t}}. \tag{2.51}$$

One can check that (2.51) satisfies the following partial differential equation

$$\frac{\partial f_W(t,w)}{\partial t} = \frac{1}{2}\frac{\partial^2 f_W(t,w)}{\partial w^2}, \tag{2.52}$$

with $f_W(0,w) = \delta(w)$.

Assume that $W(s) = x$, $s \geq 0$ and $x \in \mathbb{R}$; then, conditioned on this, we have that $W(t)$, $t \geq s$, is normally distributed with mean x and variance $t - s$, i.e., the pdf of $W(t)$ conditioned on $W(s) = x$, which we denote by $f_W(t,\cdot \,|\, s,x)$, is given by

$$f_W(t,w \,|\, s,x) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(w-x)^2}{2(t-s)}}. \tag{2.53}$$

One can check that $f_W(t,w \,|\, s,x)$ satisfies the following partial differential equation

$$\frac{\partial f_W(t,w \,|\, s,x)}{\partial t} = \frac{1}{2}\frac{\partial^2 f_W(t,w \,|\, s,x)}{\partial w^2}, \tag{2.54}$$

which has the same form as that in (2.52) governing the evolution of the unconditional pdf.

2.2 Set Theory

2.2.1 Basic Notions and Notation

A set $\mathcal{X}$ is a collection of distinct objects, referred to as the elements of $\mathcal{X}$. We write $x \in \mathcal{X}$ to denote that x is an element of $\mathcal{X}$ and $x \notin \mathcal{X}$ to denote that x is not an element of $\mathcal{X}$. The cardinality of a set $\mathcal{X}$, denoted by $|\mathcal{X}|$, is the number of elements in $\mathcal{X}$. A set $\mathcal{X}$ is called a singleton if it contains a single element, i.e., $|\mathcal{X}| = 1$. The empty set, denoted by $\emptyset$, is a set containing no elements; thus, $|\emptyset| = 0$. The sets of natural numbers (including 0), integer numbers, real numbers, and complex numbers are denoted by $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$, respectively.

A set can be described in terms of some relations that its elements satisfy. For example, let $\mathcal{X}$ denote a set whose elements, denoted by x, are real numbers satisfying the following relation:

$$x^2 - 1 \leq 0;$$

then we write

$$\mathcal{X} = \{x \in \mathbb{R} \colon x^2 - 1 \leq 0\}.$$

We say two sets $\mathcal{X}$ and $\mathcal{Y}$ are equal, and denote it by $\mathcal{X} = \mathcal{Y}$, if every element in $\mathcal{X}$ is also an element of $\mathcal{Y}$ and vice versa. We say the set $\mathcal{X}$ is a subset of

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

another set $\mathcal{Y}$, and denote it by $\mathcal{X} \subseteq \mathcal{Y}$, if every element in $\mathcal{X}$ is also an element of $\mathcal{Y}$. We say the set $\mathcal{X}$ is a strict subset of another set $\mathcal{Y}$, and denote it by $\mathcal{X} \subset \mathcal{Y}$ if $\mathcal{X} \subseteq \mathcal{Y}$ and $\mathcal{X}$ and $\mathcal{Y}$ are not equal, i.e., $\mathcal{X} \neq \mathcal{Y}$. We say two sets $\mathcal{X}$ and $\mathcal{Y}$ are disjoint if none of the elements in the set $\mathcal{X}$ are contained in the set $\mathcal{Y}$ and vice versa.

The union of two sets $\mathcal{X}$ and $\mathcal{Y}$, denoted by $\mathcal{X} \cup \mathcal{Y}$, is another set containing all the elements that are in either $\mathcal{X}$ or $\mathcal{Y}$; this can be written as

$$\mathcal{X} \cup \mathcal{Y} = \{x \colon x \in \mathcal{X} \text{ or } x \in \mathcal{Y}\}.$$

The intersection of two sets $\mathcal{X}$ and $\mathcal{Y}$, denoted by $\mathcal{X} \cap \mathcal{Y}$, is another set containing all the elements that are in both $\mathcal{X}$ and $\mathcal{Y}$; this can be written as

$$\mathcal{X} \cap \mathcal{Y} = \{x \colon x \in \mathcal{X} \text{ and } x \in \mathcal{Y}\}.$$

The difference of two sets $\mathcal{X}$ and $\mathcal{Y}$, denoted by $\mathcal{X} \setminus \mathcal{Y}$, is another set containing the elements in $\mathcal{X}$ that are not in $\mathcal{Y}$; this can be written as

$$\mathcal{X} \setminus \mathcal{Y} = \{x \colon x \in \mathcal{X} \text{ and } x \notin \mathcal{Y}\}.$$

Let $\mathcal{X} \subseteq \mathcal{U}$, where $\mathcal{U}$ denotes the universal set, i.e., the set that contains all considered elements; then we can define the complement of $\mathcal{X}$, which we denote by $\overline{\mathcal{X}}$ or $\mathcal{X}^c$, as $\overline{\mathcal{X}} = \mathcal{U} \setminus \mathcal{X}$.

The Cartesian product of two sets $\mathcal{X}$ and $\mathcal{Y}$, denoted by $\mathcal{X} \times \mathcal{Y}$, is another set containing ordered pairs of the form (x, y); this can be written as:

$$\mathcal{X} \times \mathcal{Y} = \{(x, y) \colon x \in \mathcal{X},\ y \in \mathcal{Y}\}.$$

2.2.2 Sets in Euclidean Space

We write $\mathcal{X} \subseteq \mathbb{R}^n$ to denote a set whose elements are vectors in the n-dimensional Euclidean space. Given a set $\mathcal{X} \subseteq \mathbb{R}^n$, we say an element $x \in \mathcal{X}$ is an interior point of $\mathcal{X}$ if there exists some $\epsilon > 0$ so that the set

$$\{y \in \mathbb{R}^n \colon (y - x)^\top (y - x) \leq \epsilon\}$$

is a subset of $\mathcal{X}$. The set of all interior points of $\mathcal{X}$ is called the interior of $\mathcal{X}$ and is denoted by $\mathbf{int}(\mathcal{X})$. A set $\mathcal{X} \subseteq \mathbb{R}^n$ is said to be open if $\mathbf{int}(\mathcal{X}) = \mathcal{X}$. A set $\mathcal{X} \subseteq \mathbb{R}^n$ is said to be closed if its complement,

$$\overline{\mathcal{X}} = \{x \in \mathbb{R}^n \colon x \notin \mathcal{X}\},$$

is open. The closure of a set $\mathcal{X} \subseteq \mathbb{R}^n$, denoted by $\mathbf{cl}(\mathcal{X})$, is defined as

$$\mathbf{cl}(\mathcal{X}) = \mathbb{R}^n \setminus \mathbf{int}(\mathbb{R}^n \setminus \mathcal{X}).$$

The boundary of a set $\mathcal{X} \subseteq \mathbb{R}^n$, denoted by $\mathbf{bd}(\mathcal{X})$ or $\partial\mathcal{X}$, is defined as

$$\mathbf{bd}(\mathcal{X}) = \mathbf{cl}(\mathcal{X}) \setminus \mathbf{int}(\mathcal{X}).$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Operations with Sets in $\mathbb{R}^n$

The (geometric) Minkowski sum of two sets $\mathcal{X} \in \mathbb{R}^n$ and $\mathcal{Y} \in \mathbb{R}^n$, denoted by $\mathcal{X} + \mathcal{Y}$, is another set whose elements result from adding each element in $\mathcal{X}$ to each element in $\mathcal{Y}$; this can be written as

$$\mathcal{X} + \mathcal{Y} = \big\{z \in \mathbb{R}^n \colon z = x + y,\ x \in \mathcal{X},\ y \in \mathcal{Y}\big\}.$$

The (geometric) Minkowski difference of two sets, $\mathcal{X} \in \mathbb{R}^n$ and $\mathcal{Y} \in \mathbb{R}^n$, denoted by $\mathcal{X} - \mathcal{Y}$, is another set defined as follows:

$$\mathcal{X} - \mathcal{Y} = \big\{x \in \mathbb{R}^n \colon \{x\} + \mathcal{Y} \subseteq \mathcal{X}\big\}.$$

For a given $H \in \mathbb{R}^{p \times n}$, the linear transformation of a set $\mathcal{X} \subseteq \mathbb{R}^n$ is another set $H\mathcal{X} \subseteq \mathbb{R}^p$ defined as follows:

$$H\mathcal{X} = \big\{y \in \mathbb{R}^p \colon y = Hx,\ x \in \mathcal{X}\big\}.$$

The Support Function

Given a closed set $\mathcal{X} \subseteq \mathbb{R}^n$, its support function, denoted by $S_{\mathcal{X}}(\cdot)$, is defined as

$$S_{\mathcal{X}}(\eta) = \max_{x \in \mathcal{X}} \eta^\top x, \quad \eta \in \mathbb{R}^n. \tag{2.55}$$

Let

$$x^*(\eta) = \arg \underbrace{\max_{x \in \mathcal{X}} \eta^\top x}_{S_{\mathcal{X}}(\eta)}; \tag{2.56}$$

then, clearly $S_{\mathcal{X}}(\eta) = \eta^\top x^*(\eta)$ and $x^*(\eta) \in \mathbf{bd}(\mathcal{X})$; to see this, assume that $S_{\mathcal{X}}(\eta) > 0$. Then, if $x^*(\eta) \notin \mathbf{bd}(\mathcal{X})$, there exists some $\tilde{x}^* \in \mathcal{X}$ satisfying

$$\tilde{x}^*(\eta) = x^*(\eta) + \varepsilon\eta$$

for some $\varepsilon > 0$. Then we have that

$$\begin{aligned} \eta^\top \tilde{x}^* &= \eta^\top \big(x^*(\eta) + \varepsilon\eta\big) \\ &= S_{\mathcal{X}}(\eta) + \varepsilon\eta^\top\eta \\ &> S_{\mathcal{X}}(\eta), \end{aligned}$$

but this contradicts the fact that $S_{\mathcal{X}}(\eta) = \max_{x \in \mathcal{X}} \eta^\top x$. A similar argument can be made when $S_{\mathcal{X}}(\eta) < 0$.

The support function of the Minkowski sum of two sets $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^n$, denoted by $S_{\mathcal{X}+\mathcal{Y}}(\cdot)$, is given by

$$S_{\mathcal{X}+\mathcal{Y}}(\eta) = S_{\mathcal{X}}(\eta) + S_{\mathcal{Y}}(\eta), \quad \eta \in \mathbb{R}^n. \tag{2.57}$$

To see this, define

$$\mathcal{Z} := \mathcal{X} + \mathcal{Y} = \big\{z \in \mathbb{R}^n \colon z = x + y,\ x \in \mathcal{X},\ y \in \mathcal{Y}\big\};$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

then, by using (2.55), we have

$$\begin{aligned} S_{\mathcal{X}+\mathcal{Y}}(\eta) &= S_{\mathcal{Z}}(\eta) \\ &= \max_{z\in\mathcal{Z}} \eta^\top z \\ &= \max_{x\in\mathcal{X},\ y\in\mathcal{Y}} \eta^\top (x+y) \\ &= \max_{x\in\mathcal{X}} \eta^\top x + \max_{y\in\mathcal{Y}} \eta^\top y \\ &= S_{\mathcal{X}}(\eta) + S_{\mathcal{Y}}(\eta), \end{aligned} \tag{2.58}$$

as claimed in (2.57).

The support function of $H\mathcal{X}$, where $H \in \mathbb{R}^{p\times n}$ and $\mathcal{X} \subseteq \mathbb{R}^n$, is given by

$$S_{H\mathcal{X}}(\eta) = S_{\mathcal{X}}(H^\top \eta); \tag{2.59}$$

this can be established as follows. Let $\mathcal{Y} := H\mathcal{X}$; then, by using (2.55), we have

$$\begin{aligned} S_{\mathcal{H}\mathcal{X}}(\eta) &= S_{\mathcal{Y}}(\eta) \\ &= \max_{y\in\mathcal{Y}} \eta^\top y \\ &= \max_{x\in\mathcal{X}} \eta^\top (Hx) \\ &= \max_{x\in\mathcal{X}} \left(H^\top \eta\right)^\top x \\ &= S_{\mathcal{X}}(H^\top \eta), \end{aligned} \tag{2.60}$$

as claimed in (2.59).

Convex Sets

A set $\mathcal{X} \in \mathbb{R}^n$ is convex if the line segment between any two points in $\mathcal{X}$ is contained in $\mathcal{X}$, i.e., for any $x_1, x_2 \in \mathcal{X}$, and any $\theta \in [0,1]$, we have that

$$\theta x_1 + (1-\theta)x_2 \in \mathcal{X}.$$

Let $\mathcal{B} \subseteq \mathbb{R}^n$ denote the unit ball, i.e.,

$$\mathcal{B} = \left\{\eta \in \mathbb{R}^n \colon \eta^\top \eta = 1\right\}.$$

A closed convex set $\mathcal{X} \subseteq \mathbb{R}^n$ can then be described via its support function as follows:

$$\mathcal{X} = \left\{x \in \mathbb{R}^n \colon \eta^\top x \le S_{\mathcal{X}}(\eta),\ \eta \in \mathcal{B}\right\}. \tag{2.61}$$

To establish this, we need the following result.

THEOREM 2.2 (Supporting Hyperplane Theorem) *If $\mathcal{X} \subseteq \mathbb{R}^n$ is a nonempty convex set, then for any $\overline{x}$ in the boundary of $\mathcal{X}$, there exists a vector $a \in \mathbb{R}^n$ such that*

$$a^\top x \le a^\top \overline{x} \tag{2.62}$$

for all $x \in \mathcal{X}$.

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

The geometric interpretation of the result in the theorem above is that the hyperplane,

$$\mathcal{H}(a) = \{x \in \mathbb{R}^n : a^\top x = a^\top \overline{x}\},$$

is tangent to $\mathcal{X}$ at $\overline{x}$; thus, $\mathcal{H}(a)$ is called a supporting hyperplane to $\mathcal{X}$ at $\overline{x}$, hence the name of the theorem.

From the supporting hyperplane theorem, associated to each $\overline{x} \in \mathbf{bd}(\mathcal{X})$, there exists some a so that $a^\top x \leq a^\top \overline{x}$ for all $x \in \mathcal{X}$, which is equivalent to saying that

$$\eta^\top x \leq \eta^\top \overline{x}, \quad \eta = \frac{a}{\|a\|},$$

for all $x \in \mathcal{X}$. Now, recall from (2.56) that $S_{\mathcal{X}}(\eta) = \eta^\top x^*(\eta)$, with

$$x^*(\eta) = \arg\max_{x \in \mathcal{X}} \eta^\top x$$

in the boundary of $\mathcal{X}$; thus,

$$\eta^\top x \leq \eta^\top x^*(\eta) \tag{2.63}$$

for all $x \in \mathcal{X}$. Therefore, each $\overline{x} \in \mathbf{bd}(\mathcal{X})$ can be written as

$$\overline{x} = x^*(\eta), \quad \eta \in \mathcal{B},$$

thus,

$$\mathcal{X} = \{x \in \mathbb{R}^n : \eta^\top x \leq S_{\mathcal{X}}(\eta),\ \eta \in \mathcal{B}\}, \tag{2.64}$$

with

$$\mathcal{H}(\eta) = \{x \in \mathbb{R}^n : \eta^\top x = S_{\mathcal{X}}(\eta)\} \tag{2.65}$$

defining a supporting hyperplane to $\mathcal{X}$ at $x^*(\eta)$, i.e., $\mathcal{H}(\eta)$ is tangent to $\mathcal{X}$ at $x^*(\eta)$.

Consider two closed and convex sets $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^n$. Then, as a consequence of the supporting hyperplane theorem and the result in (2.65), if $S_{\mathcal{X}}(\eta) = S_{\mathcal{Y}}(\eta)$ for some $\eta \in \mathbb{R}^n$, then the boundaries of $\mathcal{X}$ and $\mathcal{Y}$ touch each other at the following point:

$$\begin{aligned} x^*(\eta) &= \arg \underbrace{\max_{x \in \mathcal{X}} \eta^\top x}_{= S_{\mathcal{X}}(\eta)} \\ &= \arg \underbrace{\max_{x \in \mathcal{Y}} \eta^\top x}_{= S_{\mathcal{Y}}(\eta)}. \end{aligned} \tag{2.66}$$

Let $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^n$ be closed and convex, then, $\mathcal{X} \subseteq \mathcal{Y}$ if and only if $S_{\mathcal{X}}(\eta) \leq S_{\mathcal{Y}}(\eta)$ for all $\eta \in \mathbb{R}^n$; this can be established as follows. Since $\mathcal{X}$ and $\mathcal{Y}$ are convex, we can describe them using (2.61) as follows:

$$\mathcal{X} = \{x \in \mathbb{R}^n : \eta^\top x \leq S_{\mathcal{X}}(\eta),\ \eta \in \mathcal{B}\}, \tag{2.67}$$

$$\mathcal{Y} = \{y \in \mathbb{R}^n : \eta^\top y \leq S_{\mathcal{Y}}(\eta),\ \eta \in \mathcal{B}\}. \tag{2.68}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

If $S_{\mathcal{X}}(\eta) \leq S_{\mathcal{Y}}(\eta)$ for all $\eta \in \mathbb{R}^n$, then, by using (2.67), we have that every $x \in \mathcal{X}$ satisfies $\eta^\top x \leq S_{\mathcal{X}}(\eta) \leq S_{\mathcal{Y}}(\eta),\ \eta \in \mathcal{B}$; thus, every $x \in \mathcal{X}$ is contained in $\mathcal{Y}$ by virtue of (2.68), therefore, $\mathcal{X} \subseteq \mathcal{Y}$. If $\mathcal{X} \subseteq \mathcal{Y}$, then for every $\eta \in \mathcal{B}$, and because $\mathcal{X}$ is closed, there exists some $x' \in \mathcal{X}$ such that $\eta^\top x' = S_{\mathcal{X}}(\eta)$ by virtue of (2.67), which also satisfies $\eta^\top x' \leq S_{\mathcal{Y}}(\eta)$ by virtue of (2.68); thus, $S_{\mathcal{X}}(\eta) \leq S_{\mathcal{Y}}(\eta)$ for all $\eta \in \mathcal{B}$. Now, if we multiply η by any $\alpha \in \mathbb{R}$, it follows from the definition of support function that $S_{\mathcal{X}}(\alpha\eta) = \alpha S_{\mathcal{X}}(\eta)$ and $S_{\mathcal{Y}}(\alpha\eta) = \alpha S_{\mathcal{Y}}(\eta)$, thus $S_{\mathcal{X}}(\alpha\eta) \leq S_{\mathcal{Y}}(\alpha\eta)$ for any $\alpha \in \mathbb{R}$ and all $\eta \in \mathcal{B}$, therefore, $S_{\mathcal{X}}(\eta) \leq S_{\mathcal{Y}}(\eta)$ for all $\eta \in \mathbb{R}^n$.

The Minkowski sum of two closed convex sets $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^n$ is also a closed convex set. To see this, let w and z denote two elements of $\mathcal{X} + \mathcal{Y}$. Then, there exist $w_1, z_1 \in \mathcal{X}$ and $w_2, z_2 \in \mathcal{Y}$ so that $w = w_1 + w_2$ and $z = z_1 + z_2$. Now, for $\theta \in [0, 1]$, we have that

$$\theta w + (1-\theta)z = \underbrace{\theta w_1 + (1-\theta)z_1}_{=:\, x} + \underbrace{\theta w_2 + (1-\theta)z_2}_{=:\, y},$$

and by convexity of $\mathcal{X}$ and $\mathcal{Y}$, we have that $x \in \mathcal{X}$ and $y \in \mathcal{Y}$; thus, we conclude that $\theta w + (1-\theta)z \in \mathcal{X} + \mathcal{Y}$, therefore $\mathcal{X} + \mathcal{Y}$ is convex.

Ellipsoids. An ellipsoid is a closed convex set $\mathcal{E} \subseteq \mathbb{R}^n$ defined as follows:

$$\mathcal{E} = \left\{x \in \mathbb{R}^n \colon (x - x_0)^\top E^{-1}(x - x_0) \leq 1\right\}, \tag{2.69}$$

where $x_0 \in \mathbb{R}^n$ is the center of the ellipsoid, and $E \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix referred to as the shape matrix. The directions of the the semi-axes of $\mathcal{E}$ are defined by the eigenvectors of the matrix E, whereas its eigenvalues are the squares of the semi-axis lengths.

Support function: Consider an ellipsoid $\mathcal{E} \subseteq \mathbb{R}^n$ with center $x_0 \in \mathbb{R}^n$ and shape matrix $E \in \mathbb{R}^{n \times n}$. Then, its support function, $S_{\mathcal{E}}(\cdot)$, is given by

$$S_{\mathcal{E}}(\eta) = \eta^\top x_0 + \sqrt{\eta^\top E \eta}, \quad \eta \in \mathbb{R}^n, \tag{2.70}$$

and

$$\mathcal{H}(\eta) = \left\{ x \in \mathbb{R}^n \colon \eta^\top x = \underbrace{\eta^\top x_0 + \sqrt{\eta^\top E \eta}}_{= S_{\mathcal{E}}(\eta)} \right\}$$

is a supporting hyperplane to $\mathcal{E}$ at

$$x^*(\eta) = x_0 + \frac{1}{\sqrt{\eta^\top E \eta}} E\eta.$$

To see this, recall that $S_{\mathcal{E}}(\eta) = \eta^\top x^*(\eta)$, where $x^*(\eta) = \arg\max_{x \in \mathcal{E}} \eta^\top x$. Since $x^*(\eta) \in \mathbf{bd}(\mathcal{E})$, we can obtain $x^*(\eta)$ from local extremum points of the following optimization problem:

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & \eta^\top x \\ \text{subject to} \quad & (x - x_0)^\top E^{-1}(x - x_0) = 1. \end{aligned} \tag{2.71}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

By introducing a Lagrange multiplier, λ, we can reformulate (2.71) as an unconstrained optimization problem as follows:

$$\underset{x,\lambda}{\text{maximize}}\ f(x,\lambda), \tag{2.72}$$

where $f(x,\lambda) = \eta^\top x - \lambda\big((x-x_0)^\top E^{-1}(x-x_0)-1\big)$. The values that maximize $f(x,\lambda)$ can be obtained by computing the values of x and λ for which the gradient of $f(\cdot,\cdot)$ is equal to zero. Thus, we have that

$$\begin{aligned}\frac{\partial f(x,\lambda)}{\partial x} &= \eta^\top - 2\lambda(x-x_0)^\top E^{-1}\\ &= 0, \qquad (2.73)\\ \frac{\partial f(x,\lambda)}{\partial \lambda} &= (x-x_0)^\top E^{-1}(x-x_0) - 1\\ &= 0. \qquad (2.74)\end{aligned}$$

Then, by using (2.73), we obtain

$$x = x_0 + \frac{1}{2\lambda}E\eta, \tag{2.75}$$

and by plugging this expression into (2.74), and solving for λ, we obtain

$$\lambda = \pm\frac{1}{2}\sqrt{\eta^\top E\eta}. \tag{2.76}$$

Now, one can easily see that

$$x = x_0 + \frac{1}{\sqrt{\eta^\top E\eta}}E\eta, \qquad \lambda = \frac{1}{2}\sqrt{\eta^\top E\eta},$$

maximize the value of $f(x,\lambda)$, while

$$x = x_0 - \frac{1}{\sqrt{\eta^\top E\eta}}E\eta, \qquad \lambda = -\frac{1}{2}\sqrt{\eta^\top E\eta},$$

minimize it. Thus,

$$x^*(\eta) = x_0 + \frac{1}{\sqrt{\eta^\top E\eta}}E\eta,$$

which, by plugging into $S_{\mathcal{E}}(\eta) = \eta^\top x^*(\eta)$, yields

$$S_{\mathcal{E}}(\eta) = \eta^\top x_0 + \sqrt{\eta^\top E\eta},$$

as claimed in (2.70). The fact that

$$\mathcal{H}(\eta) = \left\{x \in \mathbb{R}^n \colon \eta^\top x = \eta^\top x_0 + \sqrt{\eta^\top E\eta}\right\}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

is a supporting hyperplane to $\mathcal{E}$ at

$$x^*(\eta) = x_0 + \frac{1}{\sqrt{\eta^\top E\eta}} E\eta$$

follows directly from (2.65).

Minkowski sum: Consider two ellipsoids $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^n$ defined as follows:

$$\mathcal{X} = \left\{x \in \mathbb{R}^n \colon (x - x_0)^\top X^{-1}(x - x_0) \leq 1\right\},$$
$$\mathcal{Y} = \left\{y \in \mathbb{R}^n \colon (y - y_0)^\top Y^{-1}(y - y_0) \leq 1\right\},$$

where X and Y are positive definite matrices; thus, their support functions are

$$S_{\mathcal{X}}(\eta) = \eta^\top x_0 + \sqrt{\eta^\top X\eta},$$
$$S_{\mathcal{Y}}(\eta) = \eta^\top y_0 + \sqrt{\eta^\top Y\eta},$$

respectively. Now, let $\mathcal{Z} \subseteq \mathbb{R}^n$ denote the set that results from the Minkowski sum of $\mathcal{X}$ and $\mathcal{Y}$. Then, by using (2.57), we have that

$$\begin{aligned} S_{\mathcal{Z}}(\eta) &= S_{\mathcal{X}}(\eta) + S_{\mathcal{Y}}(\eta) \\ &= \eta^\top (x_0 + y_0) + \sqrt{\eta^\top X\eta} + \sqrt{\eta^\top Y\eta}; \end{aligned} \tag{2.77}$$

thus, in general, $\mathcal{Z}$ is not an ellipsoid because, except for specific cases, we can not find some positive definite matrix Z such that $\sqrt{\eta^\top Z\eta} = \sqrt{\eta^\top X\eta} + \sqrt{\eta^\top Y\eta}$.

Volume: Consider an ellipsoid $\mathcal{E} \subseteq R^n$ with center $x_0 \in \mathbb{R}^n$ and shape matrix $E \in \mathbb{R}^{n\times n}$. Then its volume, denoted by $\mathbf{vol}(\mathcal{E})$, is defined as follows:

$$\mathbf{vol}(\mathcal{E}) = \frac{\pi^{n/2}\sqrt{\mathbf{det}(E)}}{\Gamma\left(\frac{n}{2} + 1\right)}, \tag{2.78}$$

where $\Gamma(\cdot)$ is Euler's gamma function, which is defined as follows:

$$\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du. \tag{2.79}$$

For $n = 2$, we have that

$$\begin{aligned} \Gamma(2) &= \int_0^\infty ue^{-u}du \\ &= 1; \end{aligned} \tag{2.80}$$

thus, in this case, $\mathbf{vol}(\mathcal{X}) = \pi\sqrt{\mathbf{det}(E)}$. Now, if $E = r^2 I_2$, the resulting ellipsoid is just a disc of radius r, and since $\mathbf{det}(E) = r^4$, we have that $\mathbf{vol}(\mathcal{X}) = \pi r^2$, which is the familiar formula for the surface area of a disc of radius r. For $n = 3$, we have that

$$\begin{aligned} \Gamma(2.5) &= \int_0^\infty u^{1.5}e^{-u}du \\ &= \frac{3\sqrt{\pi}}{4}; \end{aligned} \tag{2.81}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

thus, in this case, $\mathbf{vol}(\mathcal{X}) = \frac{4\pi}{3}\sqrt{\mathbf{det}(E)}$. Now, if $E = r^2 I_3$, the resulting ellipsoid is just a sphere of radius r, and since $\mathbf{det}(E) = r^6$, we have that $\mathbf{vol}(\mathcal{X}) = \frac{4\pi r^3}{3}$, which is the familiar formula for the volume of a sphere of radius r.

Zonotopes. A zonotope is a closed convex set $\mathcal{Z} \subseteq \mathbb{R}^n$ defined as follows:

$$\mathcal{Z} = \left\{ x \in \mathbb{R}^n \colon x = x_0 + \sum_{i=1}^{s} \alpha_i e_i, \ \ -1 \leq \alpha_i \leq 1 \right\}, \tag{2.82}$$

where $x_0 \in \mathbb{R}^n$ is the center of the zonotope, and $e_1, e_2, \ldots, e_s$ are vectors in $\mathbb{R}^n$ referred to as the generators of the zonotope.

Support function: Consider a zonotope $\mathcal{Z} \subseteq \mathbb{R}^n$ with center $x_0 \in \mathbb{R}^n$ and generators $e_1, e_2, \ldots, e_s$. Then, its support function, $S_{\mathcal{Z}}(\cdot)$, is defined as follows:

$$S_{\mathcal{Z}}(\eta) = \eta^\top x_0 + \sum_{i=1}^{s} |\eta^\top e_i|, \quad \eta \in \mathbb{R}^n. \tag{2.83}$$

To see this, define $\mathcal{X}_0 = \{x_0\}$ and

$$\mathcal{L}_i = \{ x \in \mathbb{R}^n \colon x = \alpha_i e_i, \ \ -1 \leq \alpha_i \leq 1 \}, \quad i = 1, 2, \ldots, s;$$

then, clearly $\mathcal{Z}$ can be written as the Minkowski sum of $\mathcal{X}_0,\ \mathcal{L}_i,\ i = 1, 2, \ldots, s$; thus,

$$S_{\mathcal{Z}}(\eta) = S_{\mathcal{X}_0}(\eta) + \sum_{i=1}^{s} S_{\mathcal{L}_i}(\eta).$$

Now, from the definition of support function in (2.55), we have

$$\begin{aligned} S_{\mathcal{X}_0}(\eta) &= \max_{x \in \mathcal{X}_0} \eta^\top x \\ &= \eta^\top x_0, \end{aligned} \tag{2.84}$$

and

$$\begin{aligned} S_{\mathcal{L}_i}(\eta) &= \max_{x \in \mathcal{L}_i} \eta^\top x \\ &= \max_{-1 \leq \alpha_i \leq 1} \alpha_i \eta^\top e_i \\ &= |\eta^\top e_i|; \end{aligned} \tag{2.85}$$

thus,

$$S_{\mathcal{Z}}(\eta) = \eta^\top x_0 + \sum_{i=1}^{s} |\eta^\top e_i|,$$

as claimed in (2.83).

Minkowski sum: Consider two zonotopes $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^n$ defined as follows:

$$\mathcal{X} = \left\{ x \in \mathbb{R}^n \colon x = x_0 + \sum_{i=1}^{p} \alpha_i e_i, \ \ -1 \leq \alpha_i \leq 1 \right\},$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

$$\mathcal{Y} = \left\{ y \in \mathbb{R}^n \colon y = y_0 + \sum_{i=1}^{q} \beta_i f_i, \;\; -1 \leq \beta_i \leq 1 \right\}; \tag{2.86}$$

thus, their support functions are

$$\begin{aligned} S_{\mathcal{X}}(\eta) &= \eta^\top x_0 + \sum_{i=1}^{p} |\eta^\top e_i|, \\ S_{\mathcal{Y}}(\eta) &= \eta^\top y_0 + \sum_{i=1}^{q} |\eta^\top f_i|, \end{aligned} \tag{2.87}$$

respectively. Now, let $\mathcal{Z} \subseteq \mathbb{R}^n$ denote the set that results from the Minkowski sum of $\mathcal{X}$ and $\mathcal{Y}$. Then, by using (2.57), we have that

$$\begin{aligned} S_{\mathcal{Z}}(\eta) &= S_{\mathcal{X}}(\eta) + S_{\mathcal{Y}}(\eta) \\ &= \eta^\top (x_0 + y_0) + \sum_{i=1}^{p} |\eta^\top e_i| + \sum_{i=1}^{q} |\eta^\top f_i|; \end{aligned} \tag{2.88}$$

thus, $\mathcal{Z}$ is clearly a zonotope with center $z_0 = x_0 + y_0$ and generators

$$e_1, e_2, \ldots, e_p, f_1, f_2, \ldots, f_q.$$

Volume: Consider a zonotope $\mathcal{Z} \subseteq \mathbb{R}^n$ with center $x_0 \in \mathbb{R}^n$ and generators $e_1, e_2, \ldots, e_s$. Then, its volume, denoted by $\mathbf{vol}(\mathcal{Z})$, is defined as follows. For $n > s$, we have that

$$\mathbf{vol}(\mathcal{Z}) = 0. \tag{2.89}$$

For $n \leq s$, consider all n-combinations of the set of generators,

$$\{e_1, e_2, \ldots, e_s\},$$

i.e., all sets formed by taking n distinct elements of the set $\{e_1, e_2, \ldots, e_s\}$; there are $N = \binom{s}{n} = \frac{s!}{n!(s-n)!}$ such sets, which we denote by $\mathcal{E}_i, \; i = 1, 2, \ldots, N$. Let E_i denote the $(n \times n)$-dimensional matrix formed by horizontal concatenation of the elements in $\mathcal{E}_i, \; i = 1, 2, \ldots, N$, then,

$$\mathbf{vol}(\mathcal{Z}) = 2^n \sum_{i=1}^{N} \left| \mathbf{det}(E_i) \right|. \tag{2.90}$$

Consider the case when $\mathcal{Z} \subseteq R^3$ is a zonotope with center $x_0 = [0, 0, 0]^\top$ and generators $e_1 = [l/2, 0, 0]^\top$, $e_2 = [0, l/2, 0]^\top$, and $e_3 = [0, 0, l/2]^\top$; thus $\mathcal{Z}$ is a cube with sides of length l. In this case, $N = 1$ and $E_1 = \frac{l}{2} I_3$; thus,

$$\mathbf{vol}(\mathcal{Z}) = 2^3 \mathbf{det}\left(\frac{l}{2} I_3 \right) = l^3,$$

which is the familiar formula for the volume of a cube of length l.

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

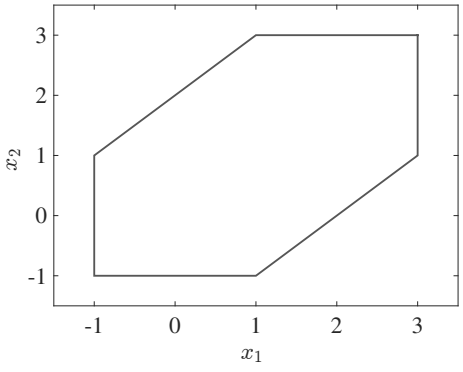

Figure 2.3 Zonotope in $\mathbb{R}^2$ with center $x_0 = [1, 1]^\top$ and generators $e_1 = [1, 0]^\top$, $e_2 = [0, 1]^\top$, and $e_3 = [1, 1]^\top$.

Example 2.14 (Zonotope in $\mathbb{R}^2$) Consider a zonotope $\mathcal{Z} \subseteq \mathbb{R}^2$ with center

$$x_0 = [1, 1]^\top,$$

and generators

$$e_1 = [1, 0]^\top, \quad e_2 = [0, 1]^\top, \quad e_3 = [1, 1]^\top; \tag{2.91}$$

thus,

$$\mathcal{Z} = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 + \alpha_1 + \alpha_3 \\ 1 + \alpha_2 + \alpha_3 \end{bmatrix} \right.$$
$$\left. -1 \leq \alpha_1 \leq 1, -1 \leq \alpha_2 \leq 1, -1 \leq \alpha_3 \leq 1 \right\}; \tag{2.92}$$

see Fig. 2.3 for a graphical depiction.

Now, by tailoring (2.90) to the setting here, we have that

$$E_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \tag{2.93}$$

from where it follows that

$$\begin{aligned} \mathbf{vol}(\mathcal{Z}) &= 2^2 \Big(\big|\mathbf{det}(E_1)\big| + \big|\mathbf{det}(E_2)\big| + \big|\mathbf{det}(E_3)\big| \Big) \\ &= 12. \end{aligned} \tag{2.94}$$

By inspection of Fig. 2.3, one can verify that the area enclosed by the boundary of the zonotope is 12, which is consistent with the result in (2.94).

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

2.3 Linear Dynamical Systems

In this section, we study discrete-time and continuous-time linear dynamical systems described by state-space models. We first introduce the notion of the state-transition matrix, which we subsequently use for trajectory characterization and stability analysis.

2.3.1 Discrete-Time Systems

Here we study linear time-varying dynamical systems described by a difference equation of the form:

$$x_{k+1} = G_k x_k + H_k w_k, \quad k = 0, 1, \ldots, \tag{2.95}$$

where $x_k \in \mathbb{R}^n$ is referred to as the system state, $w_k \in \mathbb{R}^m$ is referred to as the input, $G_k \in \mathbb{R}^{n\times n}$, and $H_k \in \mathbb{R}^{n\times m}$. Given x_0, and a sequence of inputs, $w_0, w_1, \ldots$, we would like to find an expression describing the trajectory followed by the state, $x_k, k \geq 0$. In addition, we would like to provide a formal characterization of the stability of such systems.

Trajectory Characterization. Consider first the case when $w_k = \mathbf{0}_m$ for all $k \geq 0$; then, (2.95) reduces to

$$x_{k+1} = G_k x_k, \quad k \geq 0, \tag{2.96}$$

from where it follows that

$$x_k = G_{k-1} G_{k-2} \ldots G_1 G_0 x_0, \quad k \geq 0. \tag{2.97}$$

More generally, we can relate the value the system state takes at time instant k to the value taken at an earlier time instant ℓ as follows, By defining the following matrix,

$$\Phi_{k,\ell} = \begin{cases} G_{k-1} G_{k-2} \cdots G_{\ell+1} G_\ell, & \text{if } k > \ell \geq 0, \\ I_n, & \text{if } k = \ell, \end{cases} \tag{2.98}$$

referred to as the discrete-time state-transition matrix, we can write

$$x_k = \Phi_{k,\ell} x_\ell, \quad k \geq \ell \geq 0, \tag{2.99}$$

and in particular

$$x_k = \Phi_{k,0} x_0, \quad k \geq 0. \tag{2.100}$$

From (2.98), it is clear that discrete-time state-transition matrix, $\Phi_{k,\ell}$, can also be described recursively as follows:

$$\Phi_{k+1,\ell} = G_k \Phi_{k,\ell}, \quad k \geq \ell \geq 0. \tag{2.101}$$

If G_k is invertible for all k, we can relate the value that the state takes at time instant k to the value it would take at a later time instant ℓ as follows:

$$x_k = G_k^{-1} G_{k+1}^{-1} \ldots G_{\ell-2}^{-1} G_{\ell-1}^{-1} x_\ell; \tag{2.102}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

one can see this by noting that

$$x_\ell = G_{\ell-1}G_{\ell-2}\dots G_{k+1}G_k x_k;$$

thus,

$$\begin{aligned} x_k &= (G_{\ell-1}G_{\ell-2}\dots G_{k+1}G_k)^{-1}x_\ell \\ &= G_k^{-1}G_{k+1}^{-1}\dots G_{\ell-2}^{-1}G_{\ell-1}^{-1}x_\ell. \end{aligned} \tag{2.103}$$

Thus, for such class of systems, we could generalize the definition of the discrete-time state-transition matrix in (2.98) as follows:

$$\Phi_{k,\ell} = \begin{cases} G_{k-1}G_{k-2}\cdots G_{\ell+1}G_\ell, & \text{if } k > \ell \geq 0, \\ I_n, & \text{if } k = \ell, \\ G_k^{-1}G_{k+1}^{-1}\dots G_{\ell-2}^{-1}G_{\ell-1}^{-1}, & \text{if } 0 \leq k < \ell, \end{cases} \tag{2.104}$$

which would allow us to generalize (2.99) as

$$x_k = \Phi_{k,\ell}x_\ell, \quad k,\ell \geq 0, \tag{2.105}$$

and (2.101) as

$$\Phi_{k+1,\ell} = G_k\Phi_{k,\ell}, \quad k,\ell \geq 0, \tag{2.106}$$

with $\Phi_{\ell,\ell} = I_n$. Unless otherwise stated, we will assume $\Phi_{k,\ell}$ is only defined for $k \geq \ell \geq 0$ as in (2.98).

Now, consider the general case when w_k can take any values in $\mathbb{R}^m$; then, by using (2.95) for $k = 0, 1$, we have

$$\begin{aligned} x_1 &= G_0x_0 + H_0w_0, \\ x_2 &= G_1x_1 + H_1w_1 \\ &= G_1\big(G_0x_0 + H_0w_0\big) + H_1w_1 \\ &= G_1G_0x_0 + G_1H_0w_0 + H_1w_1 \\ &= \Phi_{2,0}x_0 + \Phi_{2,1}H_0w_0 + \Phi_{2,2}H_1w_1, \end{aligned} \tag{2.107}$$

and more generally one can check that

$$x_k = \Phi_{k,0}x_0 + \sum_{\ell=0}^{k-1}\Phi_{k,\ell+1}H_\ell w_\ell, \quad k \geq 0. \tag{2.108}$$

When the system is time-invariant, i.e., $G_k = G$ and $H_k = H$ for all k, where G and H are some constant matrices, we have that

$$\Phi_{k,\ell} = G^{k-\ell}, \ k \geq \ell \geq 0;$$

thus, (2.108) reduces to

$$x_k = G^kx_0 + \sum_{\ell=0}^{k-1}G^{k-\ell-1}Hw_\ell, \quad k \geq 0. \tag{2.109}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Stability Characterization. Consider the homogeneous part of the system in (2.95), i.e.,

$$x_{k+1} = G_k x_k, \tag{2.110}$$

with $x_0 \in \mathbb{R}^n$ given. Recall the expression in (2.99):

$$x_k = \Phi_{k,\ell} x_\ell, \tag{2.111}$$

with $\Phi_{k,\ell}$ as defined in (2.98). If $x_\ell = \mathbf{0}_n$, it follows that $x_k = \mathbf{0}_n,\ k \geq \ell$; thus, we say $x^\circ = \mathbf{0}_n$ is an equilibrium point of the system in (2.110). Next, we introduce two important notions for characterizing the stability of the system in (2.110) around x°.

The system in (2.110) is said to be *stable in the sense of Lyapunov* around x° if the following property is satisfied:

S1. For a given $\epsilon > 0$, there exists some $\delta_1 > 0$ such that if $\|x_0\|_2 \leq \delta_1$, then $\|x_k\|_2 < \epsilon$ for all $k \geq 0$.

Furthermore, the system in (2.110) is said to be *asymptotically stable* around x° if it is stable in the sense of Lyapunov and the following property is additionally satisfied:

S2. There exists some $\delta_2 > 0$ such that if $\|x_0\|_2 \leq \delta_2$, then $\lim_{k\to\infty} x_k = \mathbf{0}_n$.

Since $x_k = \Phi_{k,0} x_0,\ k \geq 0$, the matrix $\Phi_{k,0}$ completely determines whether or not Properties S1 and S2 are satisfied.

For homogeneous linear time-invariant systems, i.e., $x_{k+1} = Gx_k$, we have that $\Phi_{k,0} = G^k$. Then, Property S1 is satisfied if and only if (i) the magnitude of all the eigenvalues of the matrix G is smaller than or equal to one, and (ii) for each eigenvalue whose magnitude is equal to one, the associated algebraic and geometric multiplicities must be identical. For Property S2 to be satisfied, the magnitude of all the eigenvalues of the matrix G must be strictly smaller than one. Finally, consider time-invariant systems of the form

$$x_{k+1} = Gx_k + Hw_k,$$

where $w_k,\ k \geq 0$, is bounded, i.e., there exists some positive constant, K_w, such that $\|w_k\|_2 \leq K_w$ for all k. Then, the system state will also remain bounded, i.e., there exists some positive constant, K_x, such that $\|x_k\|_2 \leq K_x$ for all k, if the magnitude of all the eigenvalues of the matrix G is strictly smaller than one.

2.3.2 Continuous-Time Systems

Consider the continuous-time counterpart of the system in (2.95), which is described by a differential equation as follows:

$$\frac{d}{dt}x(t) = A(t)x(t) + B(t)w(t), \tag{2.112}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

where $x(t) \in \mathbb{R}^n$ is the system state, $w(t) \in \mathbb{R}^m$ is the input, $A(t) \in \mathbb{R}^{n\times n}$, and $B(t) \in \mathbb{R}^{n\times m}$. Given $x(0)$, and $w(\tau),\ 0 \le \tau \le t$, where $w(\cdot)$ is some integrable function, and assuming the entries of $A(t)$ and $B(t)$ are sufficiently well behaved, we would like to find an expression describing the trajectory followed by the state, $x(t),\ t \ge 0$.

In the discrete-time case discussed earlier, we saw that the solution can be expressed as a function of the state-transition matrix $\Phi_{k,\ell}$, which is completely characterized by the recursion in (2.101); here, we will follow a similar approach and describe the solution in terms of some matrix $\Phi(t,s),\ t,s \ge 0$, referred to as the continuous-time state-transition matrix. To this end, we first consider a matrix $F(t) \in \mathbb{R}^{n\times n}$ that satisfies the following matrix differential equation:

$$\frac{d}{dt}F(t) = A(t)F(t), \quad t \ge 0, \tag{2.113}$$

with $F(0)$ given such that $\mathbf{det}\big(F(0)\big) \neq 0$; we refer to $F(t)$ as a fundamental matrix of the system in (2.112) for the special case when $B(t)w(t) = 0$ for all $t \ge 0$. The matrix $F(t)$ is invertible for all $t \ge 0$; one can see this by using Jacobi's formula (see (A.11)) as follows:

$$\begin{aligned}
\frac{d}{dt}\mathbf{det}\Big(F(t)\Big) &= \mathbf{tr}\left(\mathbf{adj}\big(F(t)\big)\frac{d}{dt}F(t)\right)\\
&= \mathbf{tr}\Big(\mathbf{adj}\big(F(t)\big)A(t)F(t)\Big)\\
&= \mathbf{tr}\Big(\underbrace{F(t)\mathbf{adj}\big(F(t)\big)}_{=\mathbf{det}\left(F(t)\right)I_n}A(t)\Big)\\
&= \mathbf{det}\big(F(t)\big)\mathbf{tr}\big(A(t)\big),
\end{aligned} \tag{2.114}$$

which by integrating results in

$$\mathbf{det}\big(F(t)\big) = \mathbf{det}\big(F(0)\big)e^{\int_0^t \mathbf{tr}\left(A(\tau)\right)d\tau}; \tag{2.115}$$

thus, clearly $\mathbf{det}\big(F(t)\big) \neq 0$, therefore $F(t)$ is invertible. Now, we use $F(t)$ to define the continuous-time state-transition matrix $\Phi(t,\tau),\ t,\tau \ge 0$, as follows:

$$\Phi(t,\tau) := F(t)F^{-1}(\tau), \quad t,\tau \ge 0. \tag{2.116}$$

The continuous-time state-transition matrix satisfies the following properties:

F1. $\Phi(t,t) = I_n$ for any $t \ge 0$. One can see this as follows:

$$\begin{aligned}
\Phi(t,t) &= F(t)F^{-1}(t)\\
&= I_n.
\end{aligned} \tag{2.117}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

F2. $\Phi^{-1}(t,s) = \Phi(s,t)$ for any $t, s \geq 0$. One can see this as follows:

$$\begin{aligned}\Phi^{-1}(t,s) &= \big(F(t)F^{-1}(s)\big)^{-1} \\ &= F(s)F^{-1}(t) \\ &= \Phi(s,t). \end{aligned} \tag{2.118}$$

F3. $\Phi(t,s) = \Phi(t,\tau)\Phi(\tau,s)$ for any $t, \tau, s \geq 0$. One can see this as follows:

$$\begin{aligned}\Phi(t,s) &= F(t)F^{-1}(s) \\ &= F(t)F^{-1}(\tau)F(\tau)F^{-1}(s) \\ &= \Phi(t,\tau)\Phi(\tau,s). \end{aligned} \tag{2.119}$$

F4. $\frac{\partial}{\partial t}\Phi(t,\tau) = A(t)\Phi(t,\tau)$, for any $t, \tau \geq 0$. One can see this as follows:

$$\begin{aligned}\frac{\partial}{\partial t}\Phi(t,\tau) &= \frac{\partial}{\partial t}\big(F(t)F^{-1}(\tau)\big) \\ &= \frac{d}{dt}F(t)F^{-1}(\tau) = A(t)F(t)F^{-1}(\tau) \\ &= A(t)\Phi(t,\tau). \end{aligned} \tag{2.120}$$

Now, we use $\Phi(t,\tau)$ to construct the solution of (2.112). For the special case when $w(\tau) = 0$ for all $\tau \in [0,t]$, the expression in (2.112) reduces to

$$\frac{d}{dt}x(t) = A(t)x(t), \tag{2.121}$$

with $x(0)$ given, the solution of which is given by

$$x(t) = \Phi(t,0)x(0); \tag{2.122}$$

this can be seen by differentiating both sides of the expression above with respect to t and using Property F4:

$$\begin{aligned}\frac{d}{dt}x(t) &= \frac{d}{dt}\Phi(t,0)x(0) \\ &= A(t)\Phi(t,0)x(0) \\ &= A(t)x(t), \end{aligned} \tag{2.123}$$

which matches the expression in (2.121). More generally, we have

$$x(t) = \Phi(t,s)x(s), \tag{2.124}$$

for any $t, s \geq 0$. (This is unlike the discrete time case, where, unless the matrix G_k is invertible for all k, the state-transition matrix, $\Phi_{k,\ell}$, was only defined for $k \geq \ell \geq 0$.) To see this, note that

$$\begin{aligned}x(t) &= \Phi(t,0)x(0), \\ x(s) &= \Phi(s,0)x(0). \end{aligned}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Now, by using Property F2, we have that

$$\begin{aligned} x(0) &= \Phi^{-1}(s,0)x(s) \\ &= \Phi(0,s)x(s), \end{aligned}$$

thus,

$$\begin{aligned} x(t) &= \Phi(t,0)x(0) \\ &= \Phi(t,0)\Phi(0,s)x(s) \\ &= \Phi(t,s)x(s), \end{aligned}$$

where the last equality follows from Property F3.

For the general case when $w(\tau)$ is not equal to zero for all $\tau \in [0,t]$, we have that the solution of (2.112) is given by

$$x(t) = \Phi(t,0)x(0) + \int_0^t \Phi(t,\tau)B(\tau)w(\tau)d\tau. \tag{2.125}$$

To see this, we differentiate both sides of (2.125) with respect to t and check that the result matches (2.112). To this end, recall Leibniz's rule for differentiating an integral:

$$\begin{aligned} \frac{d}{dx}\left(\int_{a(x)}^{b(x)} f(x,y)dy\right) &= f\big(x,b(x)\big)\frac{d}{dx}b(x) - f\big(x,a(x)\big)\frac{d}{dx}a(x) \\ &\quad + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,y)dy; \end{aligned} \tag{2.126}$$

then, it follows that

$$\begin{aligned} \frac{d}{dt}x(t) &= \frac{d}{dt}\Phi(t,0)x(0) + \frac{d}{dt}\int_0^t \Phi(t,\tau)B(\tau)w(\tau)d\tau \\ &= \frac{d}{dt}\Phi(t,0)x(0) + \int_0^t \left(\frac{\partial}{\partial t}\Phi(t,\tau)\right)B(\tau)w(\tau)d\tau + \underbrace{\Phi(t,t)}_{I_n}B(t)w(t) \\ &= A(t)\Phi(t,0)x(0) + \int_0^t A(t)\Phi(t,\tau)B(\tau)w(\tau)d\tau + B(t)w(t) \\ &= A(t)\underbrace{\left[\Phi(t,0)x(0) + \int_0^t \Phi(t,\tau)B(\tau)w(\tau)d\tau\right]}_{=\,x(t)\text{ by (2.125)}} + B(t)w(t) \\ &= A(t)x(t) + B(t)w(t), \end{aligned} \tag{2.127}$$

which matches the expression in (2.112).

When the system is time-invariant, i.e., $A(t) = A$ and $B(t) = B$ for all t, where A and B are some constant matrices, it follows from (2.113) that

$$\frac{d}{dt}\Phi(t,0) = A\Phi(t,0), \quad t \geq 0, \tag{2.128}$$

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

with $\Phi(0,0) = I_n$, the solution of which is

$$\begin{aligned}\Phi(t,0) &= e^{tA} \\ &:= \sum_{k=0}^{\infty} \frac{1}{k!} t^k A^k; \end{aligned} \tag{2.129}$$

one can check this by noting that

$$\begin{aligned}\frac{d}{dt}\Phi(t,0) &= \frac{d}{dt}\left(\sum_{k=0}^{\infty} \frac{1}{k!} t^k A^k\right) \\ &= \sum_{k=1}^{\infty} \frac{1}{(k-1)!} t^{k-1} A^k \\ &= A \sum_{k=1}^{\infty} \frac{1}{(k-1)!} t^{k-1} A^{k-1} \\ &= A \sum_{m=0}^{\infty} \frac{1}{m!} t^m A^m \\ &= A\Phi(t,0). \end{aligned} \tag{2.130}$$

Then, by plugging $\Phi(t,0) = e^{tA}$ into (2.125), we obtain

$$x(t) = e^{tA} + \int_0^t e^{(t-\tau)A} B w(\tau) d\tau. \tag{2.131}$$

Stability Characterization. Consider the homogeneous part of the system in (2.112), i.e.,

$$\frac{d}{dt} x(t) = A(t)x(t), \tag{2.132}$$

with $x(0) \in \mathbb{R}^n$. As in the discrete-time case, $x^\circ = \mathbf{0}_n$ is an equilibrium point of the system in (2.132). Similar to the discrete-time case, the system in (2.132) is said to be *stable in the sense of Lyapunov* around x°, if, for a given $\epsilon > 0$, there exists some $\delta_1 > 0$ such that if $\|x(0)\|_2 \leq \delta_1$; then $\|x(t)\|_2 < \epsilon$ for all $t \geq 0$. Furthermore, the system in (2.132) is said to be *asymptotically stable* around x°, if it is stable in the sense of Lyapunov and there exists some $\delta_2 > 0$ such that if $\|x(0)\|_2 \leq \delta_2$, then $\lim_{t\to\infty} x(t) = \mathbf{0}_n$.

For the time-invariant case, i.e., $\frac{d}{dt}x(t) = Ax(t)$, stability in the sense of Lyapunov around x° is guaranteed if and only if (i) the real part of all the eigenvalues of A is smaller than or equal to zero, and (ii) for those eigenvalues whose real part is equal to zero, the associated algebraic and geometric multiplicities are identical. Furthermore, asymptotic stability around x° is guaranteed if and only if each eigenvalue of the matrix A has real part strictly smaller than zero. Finally, consider the non-homogeneous case $\frac{d}{dt}x(t) = Ax(t) + Bw(t)$, where $w(t)$, $t \geq 0$, is bounded, i.e., $\|w(t)\|_2 \leq K_w$, where K_w is some positive constant.

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press

Then, the $x(t),\ t \geq 0$, will be bounded, i.e., $\|x(t)\|_2 \leq K_x$, where K_x is some positive constant if the real part of all the eigenvalues of A is strictly negative.

2.4 Notes and References

The material on probability theory is standard and follows the developments in [11, 12, 13, 14]. The basic material on stochastic processes can be found in [15]. The material on the Wiener process follows the developments in [11]. Basic set-theoretic notions are covered in [16]. The material on sets in Euclidean space can be found in [10, 17, 18, 19, 20]. The derivation of the formula for the support function of an ellipsoid follows ideas from [10]. The formula for the volume of an ellipsoid given in (2.78) can be obtained from that given in [21], where an n-dimensional ellipsoid is referred to as a hyperellipsoid and the term "content" is used instead of "volume." The formula for the volume of a zonotope given in (2.90) follows from that given in [22], where a zonotope $\mathcal{Z} \subseteq \mathbb{R}^n$ is parametrized as follows:

$$\mathcal{Z} = \left\{ x \in \mathbb{R}^n \colon x = \sum_{i=1}^{s} \alpha_i' e_i',\ 0 \leq \alpha_i' \leq 1 \right\}. \tag{2.133}$$

Let $\mathcal{E}_i',\ i = 1, 2, \ldots, N$, where $N = \binom{s}{n} = \frac{(s)!}{n!(s-n)!}$, denote all n-combinations of the set $\{e_1', e_2', \ldots, e_s'\}$. Then, the formula for the volume of $\mathcal{Z}$ given in [22] is as follows:

$$\mathbf{vol}(\mathcal{Z}) = \sum_{i=1}^{N} \left| \mathbf{det}(E_i') \right|, \tag{2.134}$$

where E_i' denotes the $(n \times n)$-dimensional matrix formed by horizontal concatenation of the elements in $\mathcal{E}_i'$. Note that $\mathcal{Z}$ in (2.133) can be equivalently written as

$$\mathcal{Z} = \left\{ x = \mathbb{R}^n \colon x = x_0 + \sum_{i=1}^{s} \alpha_i e_i,\ -1 \leq \alpha_i \leq 1 \right\}, \tag{2.135}$$

where $e_i = \frac{1}{2} e_i'$ and $\alpha_i = -1 + 2\alpha_i'$ for all $i = 1, 2, \ldots, s$, and $x_0 = \frac{1}{2} \sum_{i=1}^{s} e_i'$ (this is the parametrization given in (2.82)). As above, consider all n-combinations of the set $\{e_1, e_2, \ldots, e_s\}$, which we denote by $\mathcal{E}_i,\ i = 1, 2, \ldots, N$, and associated to each $\mathcal{E}_i$, define a matrix E_i formed by horizontal concatenation of the elements in $\mathcal{E}_i$. Clearly $E_i' = 2E_i,\ i = 1, 2, \ldots, s$; thus

$$\begin{aligned} \mathbf{vol}(\mathcal{Z}) &= \sum_{i=1}^{N} \left| \mathbf{det}(E_i') \right| \\ &= 2^n \sum_{i=1}^{N} \left| \mathbf{det}(E_i) \right|, \end{aligned} \tag{2.136}$$

which coincides with the formula of a zonotope given in (2.90).

https://doi.org/10.1017/9781108123853.004 Published online by Cambridge University Press