Optimization-Based Control

Richard M. Murray
Control and Dynamical Systems
California Institute of Technology

DRAFT v2.1a, January 3, 2010

Chapter 4
Stochastic Systems

In this chapter we present a focused overview of stochastic systems, oriented toward the material that is required in Chapters 5 and 6. After a brief review of random variables, we define discrete-time and continuous-time random processes, including the expectation, (co-)variance and correlation functions for a random process. These definitions are used to describe linear stochastic systems (in continuous time) and the stochastic response of a linear system to a random process (e.g., noise). We initially derive the relevant quantities in the state space, followed by a presentation of the equivalent frequency domain concepts.

Prerequisites. Readers should be familiar with basic concepts in probability, including random variables and standard distributions. We do not assume any prior familiarity with random processes.

Caveats. This chapter is written to provide a brief introduction to stochastic processes that can be used to derive the results in the following chapters. In order to keep the presentation compact, we gloss over several mathematical details that are required for rigorous presentation of the results. A more detailed (and mathematically precise) derivation of this material is available in the book by Åström [Åst06].

4.1 Brief Review of Random Variables

To help fix the notation that we will use, we briefly review the key concepts of random variables. A more complete exposition is available in standard books on probability, such as Hoel, Port and Stone [HPS71].

A *(real-valued) random variable* X is a variable that can take on any value according to a *probability distribution* P:

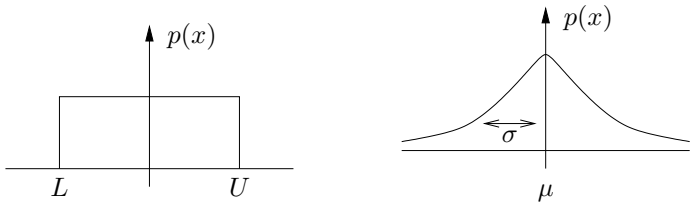
$$P(x_l \leq X \leq x_u) = \text{probability that } x \text{ takes on a value in the range } x_l,\, x_u.$$

More generally, we write $P(A)$ as the probability that an event A will occur (e.g., $A = \{x_l \leq X \leq x_u\}$). It follows from the definition that if X is a random variable in the range $[L, U]$ then $P(L \leq X \leq U) = 1$. Similarly, if $Y \in [L, U]$ then $P(L \leq X \leq Y) = 1 - P(Y \leq X \leq U)$.

We characterize a random variable in terms of the *probability density function* (pdf) $p(x)$:

$$P(x_l \leq X \leq x_u) = \int_{x_l}^{x_u} p(x)dx. \tag{4.1}$$

This can be taken as the definition of the pdf, but it is also possible to compute



(a) Uniform distribution (b) Gaussian distribution

Figure 4.1: Probability density function (pdf) for uniform and Gaussian distributions.

$p(x)$ given the distribution P as long as the distribution is suitably smooth:

$$p(x) = \left.\frac{\partial P(x_l \le x \le x_u)}{\partial x_u}\right|_{\substack{x_l \text{ fixed},\\ x_u = x,}} \qquad x > x_l.$$

We will sometimes write $p_X(x)$ when we wish to make explicit that the pdf is associated with the random variable X. Note that we use capital letters to refer to a random variable and lower case letters to refer to a specific value.

Probability distributions provide a general way to describe stochastic phenomena. Some standard probability distributions include a *uniform distribution*,

$$p(x) = \frac{1}{U - L}, \tag{4.2}$$

and a *Gaussian distribution* (also called a *normal distribution*),

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{4.3}$$

In the Gaussian distribution, the parameter μ is called the *mean* of the distribution and σ is called the *standard deviation* of the distribution. Figure 4.1 gives a graphical representation of uniform and Gaussian pdfs. There many other distributions that arise in applications, but for the purpose of these notes we focus on uniform distributions and Gaussian distributions.

If two random variables are related, we can talk about their *joint probability*: $P_{X,Y}(A, B)$ is the probability that both event A occurs for X and B occurs for Y. This is sometimes written as $P(A \cap B)$. For continuous random variables, these can be characterized in terms of a *joint probability density function*

$$P(x_l \le X \le x_u,\, y_l \le Y \le y_u) = \int_{y_l}^{y_u} \int_{x_l}^{x_u} p(x, y)\, dxdy. \tag{4.4}$$

The joint pdf thus describes the relationship between X and Y, and for sufficiently smooth distributions we have

$$p(x, y) = \left.\frac{\partial^2 P(x_l \le X \le x_u,\, y_l \le Y \le y_u)}{\partial x_u \partial y_u}\right|_{\substack{x_l, y_l \text{ fixed},\\ x_u = x,\, y_u = y,}} \qquad \begin{array}{l} x > x_l, \\ y > y_l. \end{array}$$

We say that X and Y are *independent* if $p(x,y) = p(x)p(y)$, which implies that $P_{X,Y}(A,B) = P_X(A)P_Y(B)$ for events A associated with X and B associated with Y. Equivalently, $P(A \cap B) = P(A)P(B)$ if A and B are independent.

The *conditional probability* for an event A given that an event B has occurred, written as $P(A|B)$, is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{4.5}$$

If the events A and B are independent, then $P(A|B) = P(A)$. Note that the individual, joint and conditional probability distributions are all different, so we should really write $P_{X,Y}(A \cap B)$, $P_{X|Y}(A|B)$ and $P_Y(B)$.

If X is dependent on Y then Y is also dependent on X. *Bayes' theorem* relates the conditional and individual probabilities:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \qquad P(B) \neq 0. \tag{4.6}$$

Bayes' theorem gives the conditional probability of event A on event B given the inverse relationship (B given A). It can be used in situations in which we wish to evaluate a hypothesis H given data D when we have some model for how likely the data is given the hypothesis, along with the unconditioned probabilities for both the hypothesis and the data. As we shall see, Bayes' theorem can be used to construct estimates of a system's state given measurements and a model.

The analog of the probability density function for conditional probability is the *conditional probability density function* $p(x|y)$

$$p(x|y) = \begin{cases} \dfrac{p(x,y)}{p(y)} & 0 < p(y) < \infty \\ 0 & \text{otherwise.} \end{cases} \tag{4.7}$$

It follows that

$$p(x,y) = p(x|y)p(y) \tag{4.8}$$

and

$$\begin{aligned} P(x_l \leq X \leq x_u|y) &:= P(x_l \leq X \leq x_u|Y = y) \\ &= \int_{x_l}^{x_u} p(x|y)dx = \frac{\int_{x_l}^{x_u} p(x,y)dx}{p(y)}. \end{aligned} \tag{4.9}$$

If X and Y are independent than $p(x|y) = p(x)$ and $p(y|x) = p(y)$. Note that $p(x,y)$ and $p(x|y)$ are different density functions, though they are related through equation (4.8). If X and Y are related with joint probability density function $p(x,y)$ and conditional probability density function $p(x|y)$ then

$$p(x) = \int_{-\infty}^{\infty} p(x,y)dy = \int_{-\infty}^{\infty} p(x|y)p(y)dy.$$

Example 4.1 Conditional probability for sum
Consider three random variables X, Y and Z related by the expression

$$Z = X + Y.$$

In other words, the value of the random variable Z is given by choosing values from two random variables X and Y and adding them. We assume that X and Y are independent Gaussian random variables with mean μ_1 and μ_2 and standard deviation $\sigma = 1$ (the same for both variables).

Clearly the random variable Z is not independent of X (or Y) since if we know the values of X then it provides information about the likely value of Z. To see this, we compute the joint probability between Z and X. Let

$$A = \{x_l \le x \le x_u\}, \qquad B = \{z_l \le z \le z_u\}.$$

The joint probability of both events A and B occurring is given by

$$\begin{aligned} P_{X,Z}(A \cap B) &= P(x_l \le x \le x_u,\ z_l \le x + y \le z_u) \\ &= P(x_l \le x \le x_u,\ z_l - x \le y \le z_u - x). \end{aligned}$$

We can compute this probability by using the probability density functions for X and Y:

$$\begin{aligned} P(A \cap B) &= \int_{x_l}^{x_u} \Big(\int_{z_l - x}^{z_u - x} p_Y(y) dy \Big) p_X(x) dx \\ &= \int_{x_l}^{x_u} \int_{z_l}^{z_u} p_Y(z - x) p_X(x) dz dx =: \int_{z_l}^{z_u} \int_{x_l}^{x_u} p_{Z,X}(z, x) dx dz. \end{aligned}$$

Using Gaussians for X and Y we have

$$\begin{aligned} p_{Z,X}(z, x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z - x - \mu_Y)^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x - \mu_X)^2} \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}\left((z - x - \mu_Y)^2 + (x - \mu_X)^2\right)}. \end{aligned}$$

A similar expression holds for $p_{Z,Y}$. ∇

Given a random variable X, we can define various standard measures of the distribution. The *expectation* or *mean* of a random variable is defined as

$$\mathbb{E}[X] = \langle X \rangle = \int_{-\infty}^{\infty} x\, p(x)\, dx,$$

and the *mean square* of a random variable is

$$\mathbb{E}[X^2] = \langle X^2 \rangle = \int_{-\infty}^{\infty} x^2\, p(x)\, dx.$$

If we let μ represent the expectation (or mean) of X then we define the *variance* of X as

$$\mathbb{E}[(X - \mu)^2] = \langle (X - \langle X \rangle)^2 \rangle = \int_{-\infty}^{\infty} (x - \mu)^2\, p(x)\, dx.$$

We will often write the variance as σ^2. As the notation indicates, if we have a Gaussian random variable with mean μ and (stationary) standard deviation σ, then the expectation and variance as computed above return μ and σ^2.

Several useful properties follow from the definitions.

Proposition 4.1 (Properties of random variables)**.**

1. *The expected value preserves linearity:* $\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y]$
2. *If X is a Gaussian random variable with mean μ and variance σ^2, then αX is Gaussian with mean αX and variance $\alpha^2\sigma^2$.*
3. *If X and Y are Gaussian random variables with means μ_X, μ_Y and variances σ_X^2, σ_Y^2,*

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}, \qquad p(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2},$$

then $X + Y$ is a Gaussian random variable with mean $\mu_Z = \mu_X + \mu_Y$ and variance $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$,

$$p(x+y) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{1}{2}\left(\frac{x+y-\mu_Z}{\sigma_Z}\right)^2}.$$

Proof. The first item follows directly from the definition of expectation. The second statement is proved using the definitions:

$$\begin{aligned} P(x_l \le \alpha X \le x_u) &= P(\frac{x_l}{\alpha} \le X \le \frac{x_u}{\alpha}) \\ &= \int_{\frac{x_l}{\alpha}}^{\frac{x_u}{\alpha}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \int_{x_l}^{x_u} \frac{1}{\alpha\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y/\alpha-\mu}{\sigma}\right)^2} dy \\ &= \int_{x_l}^{x_u} \frac{1}{\sqrt{2\pi\alpha^2\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\alpha\mu}{\alpha\sigma}\right)^2} dy \quad = \int_{x_l}^{x_u} p(y)\, dy \end{aligned}$$

The third item is left as an exercise. □

4.2 Introduction to Random Processes

In this section we generalize the concept of a random variable to that of a random process. We first focus on the scalar state, discrete-time case and try to build up some intuition for the basic concepts that can later be extended to the continuous-time, multi-state case.

A *discrete-time random process* is a stochastic system characterized by the *evolution* of a sequence of random variables $X[k]$. As an example, consider a discrete-time linear system with dynamics

$$x[k+1] = Ax[k] + Bu[k] + Fw[k], \qquad y[k] \quad = Cx[k] + v[k]. \tag{4.10}$$

As in ÅM08, $x \in \mathbb{R}^n$ represents the state of the system, $u \in \mathbb{R}^m$ is the vector of inputs and $y \in \mathbb{R}^p$ is the vector of outputs. The (possibly vector-valued) signal w represents disturbances to the process dynamics and v represents noise in the

measurements. To try to fix the basic ideas, we will take $u = 0$, $n = 1$ (single state) and $F = 1$, then generalize in a later chapter.

We wish to describe the evolution of the dynamics when the disturbances and noise are not given as deterministic signals, but rather are chosen from some probability distribution. Thus we will let $W[k]$ be a collection of random variables where the values at each instant k are chosen from the probability distribution $P_{W,k}$. As the notation indicates, the distributions might depend on the time instant k, although the most common case is to have a *stationary* distribution in which the distributions are independent of k (defined more formally below).

In addition to stationarity, we will often also assume that distribution of values of W at time k is independent of the values of W at time l if $k \neq l$. In other words, $W[k]$ and $W[l]$ are two separate random variables that are independent of each other. We say that the corresponding random process is *uncorrelated* (also defined more formally below). As a consequence of our independence assumption, we have that

$$\mathbb{E}[W[k]W[l]] = \mathbb{E}[W^2[k]]\delta(k-l) = \begin{cases} \mathbb{E}[W^2[k]] & k = l \\ 0 & k \neq l. \end{cases}$$

In the case that $W[k]$ is a Gaussian with mean zero and (stationary) standard deviation σ, then $\mathbb{E}[W[k]W[l]] = \sigma^2\,\delta(k-l)$.

We next wish to describe the evolution of the state x in equation (4.10) in the case when W is a random variable. In order to do this, we describe the state x as a sequence of random variables $X[k]$, $k = 1, \cdots, N$. Looking back at equation (4.10), we see that even if $W[k]$ is an uncorrelated sequence of random variables, then the states $X[k]$ are not uncorrelated since

$$X[k+1] = AX[k] + FW[k],$$

and hence the probability distribution for X at time $k+1$ depends on the value of X at time k (as well as the value of W at time k), similar to the situation in Example 4.1.

Since each $X[k]$ is a random variable, we can define the mean and variance as $\mu[k]$ and $\sigma^2[k]$ using the previous definitions at each time k:

$$\mu[k] := \mathbb{E}[X[k]] = \int_{-\infty}^{\infty} x\,p(x,k)\,dx,$$
$$\sigma^2[k] := \mathbb{E}[(X[k]-\mu[k])^2] = \int_{-\infty}^{\infty} (x-\mu[k])^2\,p(x,k)\,dx.$$

To capture the relationship between the current state and the future state, we define the *correlation function* for a random process as

$$\rho(k_1,k_2) := \mathbb{E}[X[k_1]X[k_2]] = \int_{-\infty}^{\infty} x_1x_2\,p(x_1,x_2;k_1,k_2)\,dx_1dx_2$$

The function $p(x_i,x_j;k_1,k_2)$ is the *joint probability density function*, which depends on the times k_1 and k_2. A process is *stationary* if $p(x,k+d) = p(x,d)$ for all k, $p(x_i,x_j;k_1+d,k_2+d) = p(x_i,x_j;k_1,k_2)$, etc. In this case we can write $p(x_i,x_j;d)$ for the joint probability distribution. We will almost always restrict to this case. Similarly, we will write $p(k_1,k_2)$ as $p(d) = p(k,k+d)$.

We can compute the correlation function by explicitly computing the joint pdf (see Example 4.1) or by directly computing the expectation. Suppose that we take a random process of the form (4.10) with $x[0] = 0$ and W having zero mean and standard deviation σ. The correlation function is given by

$$\begin{aligned}\mathbb{E}[X[k_1]X[k_2]] &= E\Big\{\big(\sum_{i=0}^{k_1-1} A^{k_1-i}BW[i]\big)\big(\sum_{j=0}^{k_2-1} A^{k_2-j}BW[j]\big)\Big\} \\ &= E\Big\{\sum_{i=0}^{k_1-1}\sum_{j=0}^{k_2-1} A^{k_1-i}BW[i]W[j]BA^{k_2-j}\Big\}.\end{aligned}$$

We can now use the linearity of the expectation operator to pull this inside the summations:

$$\begin{aligned}\mathbb{E}[X[k_1]X[k_2]] &= \sum_{i=0}^{k_1-1}\sum_{j=0}^{k_2-1} A^{k_1-i}B\mathbb{E}[W[i]W[j]]BA^{k_2-j} \\ &= \sum_{i=0}^{k_1-1}\sum_{j=0}^{k_2-1} A^{k_1-i}B\sigma^2\delta(i-j)BA^{k_2-j} \\ &= \sum_{i=0}^{k_1-1} A^{k_1-i}B\sigma^2BA^{k_2-i}.\end{aligned}$$

Note that the correlation function depends on k_1 and k_2.

We can see the dependence of the correlation function on the time more clearly by letting $d = k_2 - k_1$ and writing

$$\begin{aligned}\rho(k, k+d) = \mathbb{E}[X[k]X[k+d]] &= \sum_{i=0}^{k_1-1} A^{k-i}B\sigma^2BA^{d+k-i} \\ &= \sum_{j=1}^{k} A^jB\sigma^2BA^{j+d} = \Big(\sum_{j=1}^{k} A^jB\sigma^2BA^j\Big)A^d.\end{aligned}$$

In particular, if the discrete time system is stable then $|A| < 1$ and the correlation function decays as we take points that are further departed in time (d large). Furthermore, if we let $k \to \infty$ (i.e., look at the steady state solution) then the correlation function only depends on d (assuming the sum converges) and hence the steady state random process is stationary.

4.3 Continuous-Time, Vector-Valued Random Processes

A *continuous-time random process* is a stochastic system characterized by the evolution of a random variable $X(t)$, $t \in [0, T]$. As in the case of the a discrete-time random process, we are interested in understanding how the (random) state of the system is related at separate times. The process is defined in terms of the "correlation" of $X(t_1)$ with $X(t_2)$.

We call $X(t) \in \mathbb{R}^n$ the *state* of the random process. For the case $n > 1$, we have a vector of random processes:

$$X(t) = \begin{bmatrix} X_1(t) \\ \vdots \\ X_n(t) \end{bmatrix}$$

We can characterize the state in terms of a (vector-valued) time-varying pdf,

$$P(x_l \leq X_i(t) \leq x_u) = \int_{x_l}^{x_u} p_{X_i}(x;t)dx.$$

Note that the state of a random process is not enough to determine the next state (otherwise it would be a deterministic process). We typically omit indexing of the individual states unless the meaning is not clear from context.

We can characterize the dynamics of a random process by its statistical characteristics, written in terms of *joint probability* density functions:

$$P(x_{1l} \leq X_i(t_1) \leq x_{1u}, x_{2l} \leq X_j(t_2) \leq x_{2u}) \\ = \int_{x_{2l}}^{x_{2u}} \int_{x_{1l}}^{x_{1u}} p_{X_i,Y_i}(x_1, x_2; t_1, t_2)\, dx_1 dx_2$$

The function $p(x_i, x_j; t_1, t_2)$ is called a *joint probability density function* and depends both on the individual states that are being compared and the time instants over which they are compared. Note that if $i = j$, then p_{X_i,X_i} describes how X_i at time t_1 is related to X_i at time t_2.

Typically we will *assume* a certain pdf (or class of pdfs) as a model and then do our calculations across this class. One of the most common classes of random variables are Gaussian distributions and, as we shall see, one can often compute closed form solutions in this case. In practice, other distributions can arise and these are often handled using computational techniques. We will focus here on Gaussian pdfs, but some remarks on more general formulations are given in Chapter 6.

In general, the distributions used to describe a random process depend on the specific time or times that we evaluate the random variables. However, in some cases the relationship only depends on the difference in time and not the absolute times (similar to the notion of time invariance in deterministic systems, as described in ÅM08). A process is *stationary* if $p(x, t + \tau) = p(x, t)$ for all τ, $p(x_i, x_j; t_1 + \tau, t_2 + \tau) = p(x_i, x_j; t_1, t_2)$, etc. In this case we can write $p(x_i, x_j; \tau)$ for the joint probability distribution. Stationary distributions roughly correspond to the steady state properties of a random process and we will often restrict our attention to this case.

In practice we do not usually specify random processes via the joint probability distribution $p(x_i, x_j; t_1, t_2)$ but instead describe them in terms of their means, covariances and correlation functions. The previous definitions for mean, variance and correlation can be extended to the continuous time, vector-valued case by indexing

the individual states:

$$\mathbb{E}[X(t)] = \begin{bmatrix} \mathbb{E}[X_1(t)] \\ \vdots \\ \mathbb{E}[X_n(t)] \end{bmatrix} =: \mu(t)$$

$$\mathbb{E}[X(t)X^T(s)] = \begin{bmatrix} \mathbb{E}[X_1(t)X_1(s)] & \dots & \mathbb{E}[X_1(t)X_n(s)] \\ & \ddots & \vdots \\ & & \mathbb{E}[X_n(t)X_n(s)] \end{bmatrix} =: R(t,s)$$

As in the discrete time case, the random variables and their statistical properties are all indexed by the time t (and s). The matrix $R(t,s)$ is called the *correlation matrix* for $X(t) \in \mathbb{R}^n$. If $t = s$ then $R(t,t)$ describes how the elements of x are correlated at time t (with each other) and is called the *covariance matrix*. Note that the elements on the diagonal of $R(t,t)$ are the variances of the corresponding scalar variables. A random process is uncorrelated if $R(t,s) = 0$ for all $t \neq s$. This implies that $X(t)$ and $X(s)$ are independent random events and is equivalent to $p_{X,Y}(x,y) = p_X(x)p_Y(y)$.

If a random process is stationary, then it can be shown that $R(t+\tau, s+\tau) = R(t,s)$ and it follows that the correlation matrix depends only on $t-s$. In this case we will often write $R(t,s) = R(s-t)$ or simple $R(\tau)$ where τ is the correlation time. The correlation matrix in this case is simply $R(0)$.

In the case where X is also scalar random process, the correlation matrix is also a scalar and we will write $\rho(\tau)$, which we refer to as the (scalar) correlation function. Furthermore, for stationary scalar random processes, the correlation function depends only on the absolute value of the correlation function, so $\rho(tau) = \rho(-\tau) = \rho(|\tau|)$. This property also holds for the diagonal entries of the correlation matrix since $R_{ii}(s,t) = R_{ii}(t,s)$ from the definition.

Example 4.2 Ornstein-Uhlenbeck process
Consider a scalar random process defined by a Gaussian pdf with $\mu = 0$,

$$p(x,t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}},$$

and a correlation function given by

$$\rho(t_1,t_2) = \frac{Q}{2\omega_0} e^{-\omega_0|t_2-t_1|}.$$

The correlation function is illustrated in Figure 4.2. This process is also known as an *Ornstein-Uhlenbeck process*, a term that is commonly used in the scientific literature. This is a stationary process. ∇

The terminology and notation for covariance and correlation varies between disciplines. In some communities (e.g., statistics), the term "cross-covariance" is used to refer to the covariance between two random vectors X and Y, to distinguish this from the covariance of the elements of X with each other. The term "cross-correlation" is sometimes also used. `MATLAB` has a number of functions to implement covariance and correlation, which mostly match the terminology here:

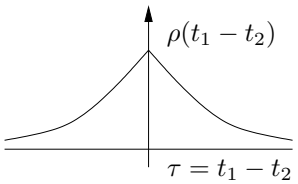


Figure 4.2: Correlation function for a first-order Markov process.

- `cov(X)` - this returns the variance of the vector `X` that represents samples of a given random variable or the covariance of the columns of a matrix X where the rows represent observations.
- `cov(X, Y)` - equivalen to `cov([X(:), Y(:)])`. Computes the covariance between the columns of X and Y, where the row are observations.
- `xcorr(X, Y)` - the "cross-correlation" between two random sequences. If these sequences came from a random process, this is basically the correlation function.
- `xcov(X, Y)` - this returns the "cross-covariance", which `MATLAB` defines as the "mean-removed cross-correlation".

The `MATLAB` help pages give the exact formulas used for each, so the main point here is to be careful to make sure you know what you really want.

We will also make use of a special type of random process referred to as "white noise". A *white noise process* $X(t)$ satisfies $\mathbb{E}[X(t)] = 0$ and $R(t, s) = W\delta(s - t)$, where $\delta(\tau)$ is the impulse function and W is called the *noise intensity*. White noise is an idealized process, similar to the impulse function or Heaviside (step) function in deterministic systems. In particular, we note that $\rho(0) = \mathbb{E}[X^2(t)] = \infty$, so the covariance is infinite and we never see this signal in practice. However, like the step function, it is very useful for characterizing the responds of a linear system, as described in the following proposition. A more formal definition of white noise, which gives better insight into the terminology, is given in Section 4.5.

4.4 Linear Stochastic Systems

We now consider the problem of how to compute the response of a linear system to a random process. We assume we have a linear system described in state space as

$$\dot{X} = AX + FW, \qquad Y = CX \tag{4.11}$$

Given an "input" W, which is itself a random process with mean $\mu(t)$, variance $\sigma^2(t)$ and correlation $\rho(t, t + \tau)$, what is the description of the random process Y?

Let W be a white noise process, with zero mean and noise intensity Q:

$$\rho(\tau) = Q\delta(\tau).$$

We can write the output of the system in terms of the convolution integral

$$Y(t) = \int_0^t h(t - \tau)W(\tau)\, d\tau,$$

where $h(t-\tau)$ is the impulse response for the system

$$h(t-\tau) = Ce^{A(t-\tau)}B + D\delta(t-\tau).$$

We now compute the statistics of the output, starting with the mean:

$$\begin{aligned}\mathbb{E}[Y(t)] &= \mathbb{E}[\int_0^t h(t-\eta)W(\eta)\,d\eta] \\ &= \int_0^t h(t-\eta)\mathbb{E}[W(\eta)]\,d\eta = 0.\end{aligned}$$

Note here that we have relied on the linearity of the convolution integral to pull the expectation inside the integral.

We can compute the covariance of the output by computing the correlation $\rho(\tau)$ and setting $\sigma^2 = \rho(0)$. The correlation function for y is

$$\begin{aligned}\rho_Y(t,s) &= \mathbb{E}[Y(t)Y(s)] = \mathbb{E}[\int_0^t h(t-\eta)W(\eta)\,d\eta \cdot \int_0^s h(s-\xi)W(\xi)\,d\xi] \\ &= \mathbb{E}[\int_0^t \int_0^s h(t-\eta)W(\eta)W(\xi)h(s-\xi)\,d\eta d\xi]\end{aligned}$$

Once again linearity allows us to exchange expectation and integration

$$\begin{aligned}\rho_Y(t,s) &= \int_0^t \int_0^s h(t-\eta)\mathbb{E}[W(\eta)W(\xi)]h(s-\xi)\,d\eta d\xi \\ &= \int_0^t \int_0^s h(t-\eta)Q\delta(\eta-\xi)h(s-\xi)\,d\eta d\xi \\ &= \int_0^t h(t-\eta)Qh(s-\eta)\,d\eta\end{aligned}$$

Now let $\tau = s - t$ and write

$$\begin{aligned}\rho_Y(\tau) = \rho_Y(t,t+\tau) &= \int_0^t h(t-\eta)Qh(t+\tau-\eta)\,d\eta \\ &= \int_0^t h(\xi)Qh(\xi+\tau)\,d\xi \qquad (\text{setting } \xi = t-\eta)\end{aligned}$$

Finally, we let $t \to \infty$ (steady state)

$$\lim_{t\to\infty} \rho_Y(t,t+\tau) = \bar{\rho}_Y(\tau) = \int_0^\infty h(\xi)Qh(\xi+\tau)d\xi \tag{4.12}$$

If this integral exists, then we can compute the second order statistics for the output Y.

We can provide a more explicit formula for the correlation function ρ in terms of the matrices A, F and C by expanding equation (4.12). We will consider the general case where $W \in \mathbb{R}^m$ and $Y \in \mathbb{R}^p$ and use the correlation matrix $R(t,s)$ instead of the correlation function $\rho(t,s)$. Define the *state transition matrix* $\Phi(t,t_0) = e^{A(t-t_0)}$ so that the solution of system (4.11) is given by

$$x(t) = \Phi(t,t_0)x(t_0) + \int_{t_0}^t \Phi(t,\lambda)Fw(\lambda)d\lambda$$

Proposition 4.2 (Stochastic response to white noise)**.** *Let* $\mathbb{E}[X(t_0)X^T(t_0)] = P(t_0)$ *and* W *be white noise with* $\mathbb{E}[W(\lambda)W^T(\xi)] = R_W\delta(\lambda - \xi)$*. Then the correlation matrix for* X *is given by*

$$R_X(t,s) = P(t)\Phi^T(s,t)$$

where $P(t)$ *satisfies the linear matrix differential equation*

$$\dot{P}(t) = AP + PA^T + FR_WF, \qquad P(0) \quad = P_0.$$

Proof. Using the definition of the correlation matrix, we have

$$\begin{aligned}
\mathbb{E}[X(t)X^T(s)] &= E\Big\{\Phi(t,0)X(0)X^T(0)\Phi^T(t,0) + \text{cross terms} \\
&\qquad + \int_0^t \Phi(t,\xi)FW(\xi)\,d\xi \int_0^s W^t(\lambda)F^T\Phi(s,\lambda)\,d\lambda\Big\} \\
&= \Phi(t,0)\mathbb{E}[X(0)X^T(0)]\Phi(s,0) \\
&\qquad + \int_0^t\int_0^s \Phi(t,\xi)F\mathbb{E}[W(\xi)W^T(\lambda)]F^T\Phi(s,\lambda)\,d\xi\,d\lambda \\
&= \Phi(t,0)P(0)\phi^T(s,0) + \int_0^t \Phi(t,\lambda)FR_W(\lambda)F^T\Phi(s,\lambda)\,d\lambda.
\end{aligned}$$

Now use the fact that $\Phi(s,0) = \Phi(s,t)\Phi(t,0)$ (and similar relations) to obtain

$$R_X(t,s) = P(t)\Phi^T(s,t)$$

where

$$P(t) = \Phi(t,0)P(0)\Phi^T(t,0) + \int_0^T \Phi(t,\lambda)FR_WF^T(\lambda)\Phi^T(t,\lambda)d\lambda$$

Finally, differentiate to obtain

$$\dot{P}(t) = AP + PA^T + FR_WF, \qquad P(0) \quad = P_0$$

(see Friedland for details). □

The correlation matrix for the output Y can be computing using the fact that $Y = CX$ and hence $R_Y = C^TR_XC$. We will often be interested in the steady state properties of the output, which given by the following proposition.

Proposition 4.3 (Steady state response to white noise)**.** *For a time-invariant linear system driven by white noise, the correlation matrices for the state and output converge in steady state to*

$$R_X(\tau) = R_X(t,t+\tau) = Pe^{A^T\tau}, \qquad R_Y(\tau) = CR_X(\tau)C^T$$

where P *satisfies the algebraic equation*

$$AP + PA^T + FR_WF^T = 0 \qquad P > 0. \tag{4.13}$$

Equation (4.13) is called the *Lyapunov equation* and can be solved in MATLAB using the function `lyap`.

Example 4.3 First-order system
Consider a scalar linear process

$$\dot{X} = -aX + W, \qquad Y = cX,$$

where W is a white, Gaussian random process with noise intensity σ^2. Using the results of Proposition 4.2, the correlation function for X is given by

$$R_X(t, t+\tau) = p(t)e^{-a\tau}$$

where $p(t) > 0$ satisfies

$$p(t) = -2ap + \sigma^2.$$

We can solve explicitly for $p(t)$ since it is a (non-homogeneous) linear differential equation:

$$p(t) = e^{-2at}p(0) + (1 - e^{-2at})\frac{\sigma^2}{2a}.$$

Finally, making use of the fact that $Y = cX$ we have

$$\rho(t, t+\tau) = c^2(e^{-2at}p(0) + (1 - e^{-2at})\frac{\sigma^2}{2a})e^{-a\tau}.$$

In steady state, the correlation function for the output becomes

$$\rho(\tau) = \frac{c^2\sigma^2}{2a}e^{-a\tau}.$$

Note correlation function has the same form as the Ornstein-Uhlenbeck process in Example 4.2 (with $Q = c^2\sigma^2$). ∇

4.5 Random Processes in the Frequency Domain

As in the case of deterministic linear systems, we can analyze a stochastic linear system either in the state space or the frequency domain. The frequency domain approach provides a very rich set of tools for modeling and analysis of interconnected systems, relying on the frequency response and transfer functions to represent the flow of signals around the system.

Given a random process $X(t)$, we can look at the frequency content of the properties of the response. In particular, if we let $\rho(\tau)$ be the correlation function for a (scalar) random process, then we define the *power spectral density function* as the Fourier transform of ρ:

$$S(\omega) = \int_{-\infty}^{\infty} \rho(\tau)e^{-j\omega\tau}\, d\tau, \qquad \rho(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{j\omega\tau}\, d\tau.$$

The power spectral density provides an indication of how quickly the values of a random process can change through the frequency content: if there is high frequency content in the power spectral density, the values of the random variable can change quickly in time.

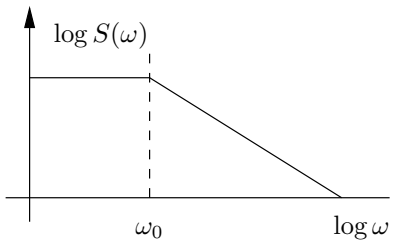


Figure 4.3: Spectral power density for a first-order Markov process.

.

Example 4.4 First-order Markov process
To illustrate the use of these measures, consider a first-order Markov process as defined in Example 4.2. The correlation function is

$$\rho(\tau) = \frac{Q}{2\omega_0} e^{-\omega_0(\tau)}.$$

The power spectral density becomes

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} \frac{Q}{2\omega_0} e^{-\omega|\tau|} e^{-j\omega\tau}\, d\tau \\ &= \int_{-\infty}^{0} \frac{Q}{2\omega_0} e^{(\omega - j\omega)\tau}\, d\tau + \int_{0}^{\infty} \frac{Q}{2\omega_0} e^{(-\omega - j\omega)\tau}\, d\tau = \frac{Q}{\omega^2 + \omega_0^2}. \end{aligned}$$

We see that the power spectral density is similar to a transfer function and we can plot $S(\omega)$ as a function of ω in a manner similar to a Bode plot, as shown in Figure 4.3. Note that although $S(\omega)$ has a form similar to a transfer function, it is a real-valued function and is not defined for complex s. ∇

Using the power spectral density, we can more formally define "white noise": a *white noise process* is a zero-mean, random process with power spectral density $S(\omega) = W = \text{constant}$ for all ω. If $X(t) \in \mathbb{R}^n$ (a random vector), then $W \in \mathbb{R}^{n \times n}$. We see that a random process is white if all frequencies are equally represented in its power spectral density; this spectral property is the reason for the terminology "white". The following proposition verifies that this formal definition agrees with our previous (time domain) definition.

Proposition 4.4. *For a white noise process,*

$$\rho(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{j\omega\tau}\, d\tau = W\delta(\tau),$$

where $\delta(\tau)$ is the unit impulse function.

Proof. If $\tau \neq 0$ then

$$\rho(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(\cos(\omega\tau) + j\sin(\omega\tau)\, d\tau = 0$$

If $\tau = 0$ then $\rho(\tau) = \infty$. Can show that

$$\rho(0) = \lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} \int_{-\infty}^{\infty} (\cdots)\, d\omega d\tau = W\delta(0)$$

□

Given a linear system

$$\dot{X} = AX + FW, \qquad Y = CX,$$

with W given by white noise, we can compute the spectral density function corresponding to the output Y. We start by computing the Fourier transform of the steady state correlation function (4.12):

$$\begin{aligned} S_Y(\omega) &= \int_{-\infty}^{\infty} \left[\int_0^{\infty} h(\xi) Q h(\xi + \tau) d\xi \right] e^{-j\omega\tau}\, d\tau \\ &= \int_0^{\infty} h(\xi) Q \left[\int_{-\infty}^{\infty} h(\xi + \tau) e^{-j\omega\tau}\, d\tau \right] d\xi \\ &= \int_0^{\infty} h(\xi) Q \left[\int_0^{\infty} h(\lambda) e^{-j\omega(\lambda - \xi)}\, d\lambda \right] d\xi \\ &= \int_0^{\infty} h(\xi) e^{j\omega\xi}\, d\xi \cdot QH(j\omega) = H(-j\omega) Q_u H(j\omega) \end{aligned}$$

This is then the (steady state) response of a linear system to white noise.

As with transfer functions, one of the advantages of computations in the frequency domain is that the composition of two linear systems can be represented by multiplication. In the case of the power spectral density, if we pass white noise through a system with transfer function $H_1(s)$ followed by transfer function $H_2(s)$, the resulting power spectral density of the output is given by

$$S_Y(\omega) = H_1(-j\omega) H_2(-j\omega) Q_u H_2(j\omega) H_1(j\omega).$$

As stated earlier, white noise is an idealized signal that is not seen in practice. One of the ways to produced more realistic models of noise and disturbances it to apply a filter to white noise that matches a measured power spectral density function. Thus, we wish to find a covariance W and filter $H(s)$ such that we match the statistics $S(\omega)$ of a measured noise or disturbance signal. In other words, given $S(\omega)$, find $W > 0$ and $H(s)$ such that $S(\omega) = H(-j\omega) W H(j\omega)$. This problem is know as the *spectral factorization problem.*

Figure 4.4 summarizes the relationship between the time and frequency domains.

4.6 Further Reading

There are several excellent books on stochastic systems that cover the results in this chapter in much more detail. For discrete-time systems, the textbook by Kumar and Varaiya [KV86] provides an derivation of the key results. Results for continuous-time systems can be found in the textbook by Friedland [Fri04]. Åström [Åst06] gives a very elegant derivation in a unified framework that integrates discrete-time and continuous-time systems.

$$p(w) = \frac{1}{\sqrt{2\pi R_W}} e^{-\frac{w^2}{2R_W}} \qquad W \longrightarrow \boxed{\quad H \quad} \longrightarrow Y \qquad p(y) = \frac{1}{\sqrt{2\pi R_Y}} e^{-\frac{y^2}{2R_Y}}$$
$$S_W(\omega) = R_W \qquad\qquad S_Y(\omega) = H(-j\omega) R_W H(j\omega)$$

$$\rho_W(\tau) = R_W \delta(\tau) \qquad \dot{X} = AX + FW \qquad \rho_Y(\tau) = R_Y(\tau) = CPe^{-A|\tau|}C^T$$
$$Y = CX \qquad AP + PA^T + FR_W F^T = 0$$

Figure 4.4: Summary of steady state stochastic response.

Exercises

4.1 Let Z be a random random variable that is the sum of two independent normally (Gaussian) distributed random variables X_1 and X_2 having means m_1, m_2 and variances σ_1^2, σ_2^2 respectively. Show that the probability density function for Z is

$$p(z) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left\{ -\frac{(z-x-m_1)^2}{2\sigma_1^2} - \frac{(x-m_2)^2}{2\sigma_2^2} \right\} dx$$

and confirm that this is normal (Gaussian) with mean m_1+m_2 and variance $\sigma_1^2+\sigma_2^2$. (Hint: Use the fact that $p(z|x_2) = p_{X_1}(x_1) = p_{X_1}(z-x_2)$.)

4.2 (ÅM08, Exercise 7.13) Consider the motion of a particle that is undergoing a random walk in one dimension (i.e., along a line). We model the position of the particle as

$$x[k+1] = x[k] + u[k],$$

where x is the position of the particle and u is a white noise processes with $E\{u[i]\} = 0$ and $E\{u[i]\,u[j]\}R_u\delta(i-j)$. We assume that we can measure x subject to additive, zero-mean, Gaussian white noise with covariance 1. Show that the expected value of the particle as a function of k is given by

$$E\{x[k]\} = E\{x[0]\} + \sum_{i=0}^{k-1} E\{u[i]\} = E\{x[0]\} =: \mu_x$$

and the covariance $E\{(x[k]-\mu_x)^2\}$ is given by

$$E\{(x[k]-\mu_x)^2\} = \sum_{i=0}^{k-1} E\{u^2[i]\} = kR_u$$

4.3 Consider a second order system with dynamics

$$\begin{bmatrix} \dot{X}_1 \\ \dot{X}_2 \end{bmatrix} = \begin{bmatrix} -a & 0 \\ 0 & -b \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} v, \qquad Y = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

that is forced by Gaussian white noise with zero mean and variance σ^2. Assume $a, b > 0$.

(a) Compute the correlation function $\rho(\tau)$ for the output of the system. Your answer should be an explicit formula in terms of a, b and σ.

(b) Assuming that the input transients have died out, compute the mean and variance of the output.

4.4 Find a constant matrix A and vectors F and C such that for

$$\dot{X} = AX + FW, \qquad Y = CX$$

the power spectrum of Y is given by

$$S(\omega) = \frac{1 + \omega^2}{(1 - 7\omega^2)^2 + 1}$$

Describe the sense in which your answer is unique.

Bibliography

[Åst06] K. J. Åström. *Introduction to Stochastic Control Theory*. Dover, New York, 2006. Originally published by Academic Press, New York, 1970.

[Fri04] B. Friedland. *Control System Design: An Introduction to State Space Methods*. Dover, New York, 2004.

[KV86] P. R. Kumar and P. Varaiya. *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Prentice Hall, Inc., 1986.