Optimization-Based Control

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

DRAFT v1.6b, 13 February 2008

Chapter 4
Stochastic Systems

In this chapter we present a focused review 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.

4.1 Review of Random Variables

A *(continuous) 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. 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.

Some standard probability distributions include a *uniform distribution*,

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

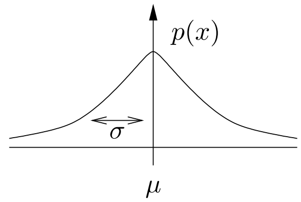


Figure 4.1: Probability density function (pdf) for a Gaussian distribution.

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 normal 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 a Gaussian pdf. 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 thust describes the relationship between X and Y. 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.

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

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{4.6}$$

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

$$P(x_l \le X \le x_u|y) = \int_{x_l}^{x_u} p(x|y)dx = \frac{\int_{x_l}^{x_u} p(x,y)dx}{p(y)}. \tag{4.9}$$

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 conditional probability distribution $p(x|y)$ 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

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

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

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

Using Gaussians for X and Y we have

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

∇

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

$$E\{X\} = \int_{-\infty}^{\infty} x\, p(x)\, dx,$$

and the *mean square* of a random variable is

$$E\{X^2\} = \int_{-\infty}^{\infty} x^2\, p(x)\, dx.$$

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

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

We will often write the variance as σ^2. As the notation indicates, if we have a Gaussian random variable with mean μ and variance σ^2, then the expectation and variance as computed above return precisely those quantities.

The following properties follow from the definitions.

Proposition 4.1.

1. $E\{\alpha X + \beta Y\} = \alpha E\{X\} + \beta E\{Y\}$
2. *If X and Y are Gaussian random processes/variables with*

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \qquad p(y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}$$

   then $X+Y$ is a Gaussian random process/variable with

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

   where

$$\mu_z = \mu_x + \mu_y \qquad \sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

3. *If X is a Gaussian random variable with means μ and variance σ^2, then αX is Gaussian with mean αX and variance $\alpha^2\sigma^2$.*

Proof. The first item follows directly from the definition of expection. The second item is left as an exercise. The third statement is proved using the definitions:

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

□

4.2 Introduction to Random Processes

In this section we generalize the concept of a random variable to that of a random process. We 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

$$x[k+1] = Ax[k] + Bu[k] + Fv[k]. \tag{4.10}$$

As in ÅM08, $x \in \mathbb{R}^n$ represents the state of the system, $u \in \mathbb{R}^p$ is the vector of inputs and $y \in \mathbb{R}^q$ is the vector of outputs. The signal v represents the process disturbances. The (possibly vector-valued) signal v represents disturbances to the process dynamics and w represents noise in the measurements. For simplicity, we will take $u = 0$, $n = 1$ (single state) and $F = 1$.

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 $V[k]$ be random variables where the values at each instant k are chosen from the probability distribution $P_{V,k}$. 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.

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

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

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

We next wish to describe the evolution of the state x in equation (4.10) in the case when V 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 $V[k]$ is an uncorrelated sequence of random variables, then the states $X[k]$ are not uncorrelated since

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

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 V 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:

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

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

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

$$\rho(k_1,k_2) = E\{X[k_1]X[k_2]\} = \int_{-\infty}^{\infty} x_1 x_2\, p(x_1,x_2;k_1,k_2)\, dx_1 dx_2$$

The function $p(x_1,x_2;k_1,k_2)$ is the joint density function, which depends on the times k_1 and k_2. A process is *stationary* if $p(x,k+\kappa) = p(x,\kappa)$ for all k, $p(x_1,x_2;k_1+\kappa,k_2+\kappa) = p(x_1,x_2;k_1,k_2)$, etc. In this case we can write $p(x_1,x_2;\kappa)$ for the joint probability distribution. We will almost always restrict to this case. Similarly, we will write $p(k_1,k_2)$ as $p(\kappa) = p(\kappa, k+\kappa)$.

We can compute the 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 V having zero mean and standard deviation σ. The correlation function is given by

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

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

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

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

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

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 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:

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

The function $p(x_1, x_2; t_1, t_2)$ 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. In practice, pdf's are not available for most random processes, so this formulation is mainly useful for analytical derivations. 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 for solutions in this case.

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 different 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, $p(x_1, x_2; t_1 + \tau, t_2 + \tau) = p(x_1, x_2; t_1, t_2)$, etc. In this case we can write $p(x_1, x_2; \tau)$ for the joint probability distribution and we write $R(\tau) := R(t, t + \tau)$. 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 don't 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 mean, covariance and correlation function. The previous definitions for mean, variance and correlation can be extended to the continuous time,

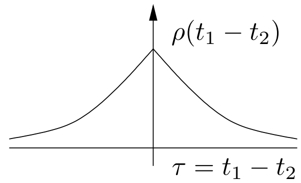


Figure 4.2: Correlation function for a first-order Markov process.

vector-valued case by indexing the individual state:

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

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

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.

Example 4.2 First-order Markov process
Consider a first order *Markov 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 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:

- `cov(X)` - this returns the variance of the vector `X` that represents samples of a given random variable.

- `cov(X, Y)` - I'm not sure what this means yet.
- `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.

4.4 Linear Stochastic Systems

We now consider the problem of how to compute the response of a linear system to a random variable. We assume we have a linear system described either in state space or as a transfer function:

$$\begin{aligned}\dot{X} &= AX + BU \\ Y &= CX\end{aligned} \qquad H(s) = C(sI - A)^{-1}B.$$

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

Let U be an uncorrelated, Gaussian random process, with zero mean and covariance Q:

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

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

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

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

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

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

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

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

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

Once again linearity allows us to exchange expectation and integration

$$\begin{aligned}\rho_y(t,s) &= \int_0^t \int_0^s h(t-\eta)E\{U(\eta)U(\xi)\}h(s-\xi)\,d\eta d\xi \\ &= \int_0^t \int_0^s h(t-\eta)Q\delta(\eta-\xi)h(s-\xi)\,d\eta d\xi \\ &= \int_0^t h(t-\eta)Qh(s-\eta)\,d\eta\end{aligned}$$

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)

$$\lim_{t\to\infty} \rho_y(t,t+\tau) = \bar{\rho}_y(\tau) = \int_0^\infty h(\xi)Qh(\xi+\tau)d\xi \tag{4.11}$$

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 (4.11). We will consider the general case where $V \in \mathbb{R}^m$ 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 (**??**) is given by

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

Proposition 4.2 (Full stochastic reponse)**.** *Let $E\{X(t_0)X^T(t_0)\} = P(t_0)$ and V be white noise with $E\{V(\lambda)V^T(\xi)\} = Q_v\delta(\lambda-\xi)$. 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

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

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

$$\begin{aligned}
E\{X(t)X^T(s)\} &= E\Big\{\Phi(t,0)X(0)X^T(0)\Phi^T(t,0) + \text{cross terms} \\
&\quad + \int_0^t \Phi(t,\xi)FV(\xi)d\xi \int_0^s V^t(\lambda)F^T\Phi(s,\lambda)d\lambda\Big\} \\
&= \Phi(t,0)E\{X(0)X^T(0)\}\Phi(s,0) \\
&\quad + \int_0^t \int_0^s \Phi(t,\xi)FE\{V(\xi)V^T(\lambda)\}F^T\Phi(s,\lambda)d\xi d\lambda \\
&= \Phi(t,0)P(0)\phi^T(s,0) \\
&\quad + \int_0^t \Phi(t,\lambda)FQ_V(\lambda)F^T\Phi(s,\lambda)d\lambda
\end{aligned}$$

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

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

Finally, differentiate to obtain

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

(see Friedland for details). □

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

Proposition 4.3 (Steady-state stochastic response)**.** *For a time-invariant linear system driven by white noise, the correlation matrices for the state and output are given by*

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

where P satisfies the algebraic equation

$$AP + PA^T + FQ_VF^T = 0 \qquad P > 0. \tag{4.12}$$

Equation (4.12) 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

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

where V is a white, Gaussian random process with variance 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 first-order Markov process in Example 4.2 (with $Q = c^2\sigma 2$). ∇

4.5 Random Processes in the Frequency Domain

Power Spectral Density

We can also characterize the spectrum of a random process. Let $\rho(\tau)$ be the correlation function for a random process. We define the *power spectral density function* as

$$S(\omega) = \int_{-\infty}^{\infty} \rho(\tau)e^{-j\omega\tau}\,d\tau \qquad \text{Fourier transform}$$

$$\rho(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{j\omega\tau}\,d\tau \quad \text{Inverse Fourier transform}$$

Definition 4.1. A process is *white noise* if $E\{X(t)\} = 0$ and $S(\omega) = W =$ constant for all ω. If $X(t) \in \mathbb{R}^n$ (a random vector), then $W \in \mathbb{R}^{n\times n}$.

Properties:

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

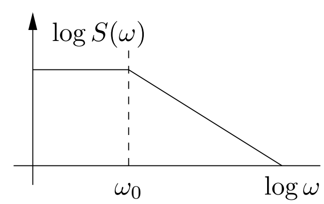
$$\int_{-\epsilon}^{\epsilon} \int_{-\infty}^{\infty} (\cdots)\, d\omega d\tau = W$$

2. $\rho(0) = E\{x^2(t)\} = \infty \implies$ idealization; never see this in practice.
3. Typically we use white noise as an idealized input to a system and characterize the output of the system (similar to the impulse response in deterministic systems).

Example. First order Markov process

$$\rho(\tau) = \frac{Q}{2\omega_0} e^{-\omega_0(\tau)} \qquad p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

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



Be careful: $S(\omega)$ is *not* a transfer function. $S(\omega)$ is always real.

Spectral response

We now compute the spectral density function corresponding to Y

$$\begin{aligned} S_y(\omega) &= \int_{-\infty}^{\infty} \left[\int_{0}^{\infty} h(\xi) Q h(\xi + \tau) d\xi \right] e^{-j\omega\tau}\, d\tau \\ &= \int_{0}^{\infty} h(\xi) Q \left[\int_{-\infty}^{\infty} h(\xi + \tau) e^{-j\omega\tau}\, d\tau \right] d\xi \\ &= \int_{0}^{\infty} h(\xi) Q \left[\int_{0}^{\infty} h(\lambda) e^{-j\omega(\lambda - \xi)}\, d\lambda \right] d\xi \\ &= \int_{0}^{\infty} h(\xi) e^{j\omega\xi}\, d\xi \cdot Q H(j\omega) = H(-j\omega) Q_u H(j\omega) \end{aligned}$$

This is then the response of a linear system to white noise.

Composition:

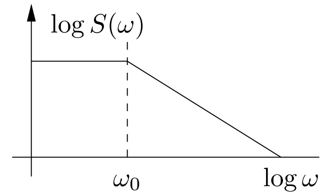
$$S_y(\omega) = H_2(-j\omega) H_1(-j\omega) Q_u H_1(j\omega) H_2(j\omega)$$

Note that in the frequency domain, we get *multiplication* of frequency responses (same as composition transfer functions)

Example: first order response

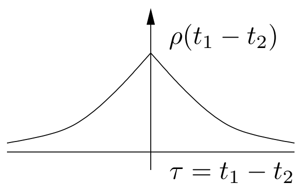
$$H(j\omega) = \frac{1}{s+\omega_0} \qquad \Longrightarrow$$

$$S_y(\omega) = \frac{1}{-j\omega+\omega_0} \cdot Q \cdot \frac{1}{j\omega+\omega_0} = \frac{Q}{\omega^2+\omega_0^2}$$



We can also compute the correlation function for a random process Y

$$\bar\rho_Y(\tau) = \int_0^\infty H(\xi) Q H(\xi+\tau) d\xi = \frac{Q}{2\omega_0} e^{-}\omega|\tau|$$



- Shows that $Y(t)$ is correlated to $Y(t+\tau)$ even though input is white (uncorrelated)
- Correlation drops off as τ increases

Spectral factorization

We often want to find a Q and H such that we match the statistics of measured noise. Eg, given $S(\omega)$, find $Q > 0$ and $H(s)$ such that $S(\omega) = H(-j\omega) Q H(j\omega)$.

4.6 Further Reading

Exercises

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

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

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 $p(z|y) = p_x(x) = p_x(z-y)$.)

4.2 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.

4.3 (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 coveriance $E\{(x[k] - \mu_x)^2\}$ is given by

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

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

$$\dot{x} = Ax + Fw, \; y = Cx$$

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.