Chapter 7

Random processes and noise

7.1 Introduction

Chapter 6 discussed modulation and demodulation, but replaced any detailed discussion of the noise by the assumption that a minimal separation is required between each pair of signal points. This chapter develops the underlying principles needed to understand noise, and the next chapter shows how to use these principles in detecting signals in the presence of noise.

Noise is usually the fundamental limitation for communication over physical channels. This can be seen intuitively by accepting for the moment that different possible transmitted waveforms must have a difference of some minimum energy to overcome the noise. This difference reflects back to a required distance between signal points, which along with a transmitted power constraint, limits the number of bits per signal that can be transmitted.

The transmission rate in bits per second is then limited by the product of the number of bits per signal times the number of signals per second, *i.e.*, the number of degrees of freedom per second that signals can occupy. This intuitive view is substantially correct, but must be understood at a deeper level which will come from a probabilistic model of the noise.

This chapter and the next will adopt the assumption that the channel output waveform has the form $y(t) = x(t) + z(t)$ where $x(t)$ is the channel input and $z(t)$ is the noise. The channel input $x(t)$ depends on the random choice of binary source digits, and thus $x(t)$ has to be viewed as a particular selection out of an ensemble of possible channel inputs. Similarly, $z(t)$ is a particular selection out of an ensemble of possible noise waveforms.

The assumption that $y(t) = x(t) + z(t)$ implies that the channel attenuation is known and removed by scaling the received signal and noise. It also implies that the input is not filtered or distorted by the channel. Finally it implies that the delay and carrier phase between input and output is known and removed at the receiver.

The noise should be modeled probabilistically. This is partly because the noise is a priori unknown, but can be expected to behave in statistically predictable ways. It is also because encoders and decoders are designed to operate successfully on a variety of different channels, all of which are subject to different noise waveforms. The noise is usually modeled as zero mean, since a mean can be trivially removed.

Modeling the waveforms $x(t)$ and $z(t)$ probabilistically will take considerable care. If $x(t)$ and $z(t)$ were defined only at discrete values of time, such as $\{t = kT; k \in \mathbb{Z}\}$, then they could

be modeled as sample values of sequences of random variables (rv's). These sequences of rv's could then be denoted as $X(t) = \{X(kT);\, k \in \mathbb{Z}\}$ and $Z(t) = \{Z(kT);\, k \in \mathbb{Z}\}$. The case of interest here, however, is where $x(t)$ and $z(t)$ are defined over the continuum of values of t, and thus a continuum of rv's is required. Such a probabilistic model is known as a *random process* or, synonomously, a *stochastic process*. These models behave somewhat similarly to random sequences, but they behave differently in a myriad of small but important ways.

7.2 Random processes

A *random process* $\{Z(t);\, t \in \mathbb{R}\}$ is a collection[1] of rv's, one for each $t \in \mathbb{R}$. The parameter t usually models time, and any given instant in time is often referred to as an *epoch*. Thus there is one rv for each epoch. Sometimes the range of t is restricted to some finite interval, $[a,\, b]$, and then the process is denoted as $\{Z(t);\, t \in [a, b]\}$.

There must be an underlying sample space Ω over which these rv's are defined. That is, for each epoch $t \in \mathbb{R}$ (or $t \in [a,\, b]$), the rv $Z(t)$ is a function $\{Z(t, \omega);\, \omega{\in}\Omega\}$ mapping sample points $\omega \in \Omega$ to real numbers.

A given sample point $\omega \in \Omega$ within the underlying sample space determines the sample values of $Z(t)$ for each epoch t. The collection of all these sample values for a given sample point ω, *i.e.*, $\{Z(t, \omega);\, t \in \mathbb{R}\}$ is called a *sample function* $\{z(t) : \mathbb{R} \to \mathbb{R}\}$ of the process.

Thus $Z(t, \omega)$ can be viewed as a function of ω for fixed t, in which case it is the rv $Z(t)$, or it can be viewed as a function of t for fixed ω, in which case it is the sample function $\{z(t) : \mathbb{R} \to \mathbb{R}\} = \{Z(t, \omega);\, t \in \mathbb{R}\}$ corresponding to the given ω. Viewed as a function of both t and ω, $\{Z(t, \omega); t \in \mathbb{R},\, \omega \in \Omega\}$ is the random process itself; the sample point ω is usually suppressed, denoting the process as $\{Z(t);\, t \in \mathbb{R}\}$

Suppose a random process $\{Z(t);\, t \in \mathbb{R}\}$ models the channel noise and $\{z(t) : \mathbb{R} \to \mathbb{R}\}$ is a sample function of this process. At first this seems inconsistent with the traditional elementary view that a random process or set of rv's models an experimental situation *a priori* (before performing the experiment) and the sample function models the result *a posteriori* (after performing the experiment). The trouble here is that the experiment might run from $t = -\infty$ to $t = \infty$, so there can be no "before" for the experiment and "after" for the result.

There are two ways out of this perceived inconsistency. First, the notion of 'before and after' in the elementary view is inessential; the only important thing is the view that a multiplicity of sample functions might occur, but only one actually occurs. This point of view is appropriate in designing a cellular telephone for manufacture. Each individual phone that is sold experiences its own noise waveform, but the device must be manufactured to work over the multiplicity of such waveforms.

Second, whether we view a function of time as going from $-\infty$ to $+\infty$ or going from some large negative to large positive time is a matter of mathematical convenience. We often model waveforms as persisting from $-\infty$ to $+\infty$, but this simply indicates a situation in which the starting time and ending time are sufficiently distant to be irrelevant.

[1]Since a random variable is a mapping from Ω to $\mathbb{R}$, the sample values of a rv are real and thus the sample functions of a random process are real. It is often important to define objects called complex random variables that map Ω to $\mathbb{C}$. One can then define a complex random process as a process that maps each $t \in \mathbb{R}$ into a complex random variable. These complex random processes will be important in studying noise waveforms at baseband.

In order to specify a random process $\{Z(t); t \in \mathbb{R}\}$, some kind of rule is required from which joint distribution functions can, at least in principle, be calculated. That is, for all positive integers n, and all choices of n epochs $t_1, t_2, \ldots, t_n$, it must be possible (in principle) to find the joint distribution function,

$$F_{Z(t_1),\ldots,Z(t_n)}(z_1, \ldots, z_n) = \Pr\{Z(t_1) \leq z_1, \ldots, Z(t_n) \leq z_n\}, \tag{7.1}$$

for all choices of the real numbers $z_1, \ldots, z_n$. Equivalently, if densities exist, it must be possible (in principle) to find the joint density,

$$f_{Z(t_1),\ldots,Z(t_n)}(z_1, \ldots, z_n) = \frac{\partial^n F_{Z(t_1),\ldots,Z(t_n)}(z_1, \ldots, z_n)}{\partial z_1 \cdots \partial z_n}, \tag{7.2}$$

for all real $z_1, \ldots, z_n$. Since n can be arbitrarily large in (7.1) and (7.2), it might seem difficult for a simple rule to specify all these quantities, but a number of simple rules are given in the following examples that specify all these quantities.

7.2.1 Examples of random processes

The following generic example will turn out to be both useful and quite general. We saw earlier that we could specify waveforms by the sequence of coefficients in an orthonormal expansion. In the following example, a random process is similarly specified by a sequence of rv's used as coefficients in an orthonormal expansion.

Example 7.2.1. Let $Z_1, Z_2, \ldots$, be a sequence of rv's defined on some sample space Ω and let $\{\phi_1(t)\}, \{\phi_2(t)\}, \ldots$, be a sequence of orthogonal (or orthonormal) real functions. For each $t \in \mathbb{R}$, let the rv $Z(t)$ be defined as $Z(t) = \sum_k Z_k \phi_k(t)$. The corresponding random process is then $\{Z(t); t \in \mathbb{R}\}$. For each t, $Z(t)$ is simply a sum of rv's, so we could, in principle, find its distribution function. Similarly, for each n-tuple, $t_1, \ldots, t_n$ of epochs, $Z(t_1), \ldots, Z(t_n)$ is an n-tuple of rv's whose joint distribution could in principle be found. Since $Z(t)$ is a countably infinite sum of rv's, $\sum_{k=1}^{\infty} Z_k \phi_k(t)$, there might be some mathematical intricacies in finding, or even defining, its distribution function. Fortunately, as will be seen, such intricacies do not arise in the processes of most interest here.

It is clear that random processes can be defined as in the above example, but it is less clear that this will provide a mechanism for constructing reasonable models of actual physical noise processes. For the case of Gaussian processes, which will be defined shortly, this class of models will be shown to be broad enough to provide a flexible set of noise models.

The next few examples specialize the above example in various ways.

Example 7.2.2. Consider binary PAM, but view the input signals as independent identically distributed (iid) rv's $U_1, U_2, \ldots$, which take on the values ± 1 with probability $1/2$ each. Assume that the modulation pulse is $\text{sinc}(\frac{t}{T})$ so the baseband random process is

$$U(t) = \sum_k U_k \,\text{sinc}\left(\frac{t - kT}{T}\right).$$

At each sampling epoch kT, the rv $U(kT)$ is simply the binary rv U_k. At epochs between the sampling epochs, however, $U(t)$ is a countably infinite sum of binary rv's whose variance will later be shown to be 1, but whose distribution function is quite ugly and not of great interest.

Example 7.2.3. A random variable is said to be zero-mean Gaussian if it has the probability density

$$f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[\frac{-z^2}{2\sigma^2}], \tag{7.3}$$

where σ^2 is the variance of Z. A common model for a noise process $\{Z(t); t \in \mathbb{R}\}$ arises by letting

$$Z(t) = \sum_k Z_k \operatorname{sinc}\left(\frac{t - kT}{T}\right), \tag{7.4}$$

where $\ldots, Z_{-1}, Z_0, Z_1, \ldots$, is a sequence of iid zero-mean Gaussian rv's of variance σ^2. At each sampling epoch kT, the rv $Z(kT)$ is the zero-mean Gaussian rv Z_k. At epochs between the sampling epochs, $Z(t)$ is a countably infinite sum of independent zero-mean Gaussian rv's, which turns out to be itself zero-mean Gaussian of variance σ^2. The next section considers sums of Gaussian rv's and their inter-relations in detail. The sample functions of this random process are simply sinc expansions and are limited to the baseband $[-1/(2T), 1/(2T)]$. This example, as well as the previous example, brings out the following mathematical issue: the expected energy in $\{Z(t); t \in \mathbb{R}\}$ turns out to be infinite. As discussed later, this energy can be made finite either by truncating $Z(t)$ to some finite interval much larger than any time of interest or by similarly truncating the sequence $\{Z_k; k \in \mathbb{Z}\}$.

Another slightly disturbing aspect of this example is that this process cannot be 'generated' by a sequence of Gaussian rv's entering a generating device that multiplies them by T-spaced sinc functions and adds them. The problem is the same as the problem with sinc functions in the previous chapter - they extend forever and thus the process cannot be generated with finite delay. This is not of concern here, since we are not trying to generate random processes, only to show that interesting processes can be defined. The approach here will be to define and analyze a wide variety of random processes, and then to see which are useful in modeling physical noise processes.

Example 7.2.4. Let $\{Z(t); t \in [-1, 1]\}$ be defined by $Z(t) = tZ$ for all $t \in [-1, 1]$ where Z is a zero-mean Gaussian rv of variance 1. This example shows that random processes can be very degenerate; a sample function of this process is fully specified by the sample value $z(t)$ at $t = 1$. The sample functions are simply straight lines through the origin with random slope. This illustrates that the sample functions of a random process do not necessarily "look" random.

7.2.2 The mean and covariance of a random process

Often the first thing of interest about a random process is the mean at each epoch t and the covariance between any two epochs t, τ. The mean, $\mathsf{E}[Z(t)] = \overline{Z}(t)$ is simply a real valued function of t and can be found directly from the distribution function $F_{Z(t)}(z)$ or density $f_{Z(t)}(z)$. It can be verified that $\overline{Z}(t)$ is 0 for all t for Examples 7.2.2, 7.2.3, and 7.2.4 above. For Example 7.2.1, the mean can not be specified without specifying more about the random sequence and the orthogonal functions.

The covariance[2] is a real-valued function of the epochs t and τ. It is denoted by $\mathsf{K}_Z(t, \tau)$ and

[2]This is often called the *autocovariance* to distinguish it from the covariance between two processes; we will not need to refer to this latter type of covariance.

defined by

$$\mathsf{K}_Z(t,\tau) = \mathsf{E}\left[[Z(t)-\overline{Z}(t)][Z(\tau)-\overline{Z}(\tau)]\right]. \tag{7.5}$$

This can be calculated (in principle) from the joint distribution function $F_{Z(t),Z(\tau)}(z_1,z_2)$ or from the density $f_{Z(t),Z(\tau)}(z_1,z_2)$. To make the covariance function look a little simpler, we usually split each random variable $Z(t)$ into its mean, $\overline{Z}(t)$, and its fluctuation, $\widetilde{Z}(t) = Z(t)-\overline{Z}(t)$. The covariance function is then

$$\mathsf{K}_Z(t,\tau) = \mathsf{E}\left[\widetilde{Z}(t)\widetilde{Z}(\tau)\right]. \tag{7.6}$$

The random processes of most interest to us are used to model noise waveforms and usually have zero mean, in which case $Z(t) = \widetilde{Z}(t)$. In other cases, it often aids intuition to separate the process into its mean (which is simply an ordinary function) and its fluctuation, which is by definition zero mean.

The covariance function for the generic random process in Example 7.2.1 above can be written as

$$\mathsf{K}_Z(t,\tau) = \mathsf{E}\left[\sum_k \widetilde{Z}_k\phi_k(t)\sum_m \widetilde{Z}_m\phi_m(\tau)\right]. \tag{7.7}$$

If we assume that the rv's $Z_1, Z_2, \ldots$ are iid with variance σ^2, then $\mathsf{E}[\widetilde{Z}_k\widetilde{Z}_m] = 0$ for $k \neq m$ and $\mathsf{E}[\widetilde{Z}_k\widetilde{Z}_m] = \sigma^2$ for $k = m$. Thus, ignoring convergence questions, (7.7) simplifies to

$$\mathsf{K}_Z(t,\tau) = \sigma^2\sum_k \phi_k(t)\phi_k(\tau). \tag{7.8}$$

For the sampling expansion, where $\phi_k(t) = \mathrm{sinc}(\frac{t}{T}-k)$, it can be shown (see (7.48)) that the sum in (7.8) is simply $\mathrm{sinc}(\frac{t-\tau}{T})$. Thus for Examples 7.2.2 and 7.2.3, the covariance is given by

$$\mathsf{K}_Z(t,\tau) = \sigma^2\mathrm{sinc}\left(\frac{t-\tau}{T}\right)$$

where $\sigma^2 = 1$ for the binary PAM case of Example 7.2.2. Note that this covariance depends only on $t-\tau$ and not on the relationship between t or τ and the sampling points kT. These sampling processes are considered in more detail later.

7.2.3 Additive noise channels

The communication channels of greatest interest to us are known as *additive noise channels*. Both the channel input and the noise are modeled as random processes, $\{X(t);\, t\in\mathbb{R}\}$ and $\{Z(t);\, t\in\mathbb{R}\}$, both on the same underlying sample space Ω. The channel output is another random process $\{Y(t);\, t\in\mathbb{R}\}$ and $Y(t) = X(t)+Z(t)$. This means that for each epoch t the random variable $Y(t)$ is equal to $X(t)+Z(t)$.

Note that one could always define the noise on a channel as the difference $Y(t)-X(t)$ between output and input. The notion of *additive noise* inherently also includes the assumption that the processes $\{X(t);\, t\in\mathbb{R}\}$ and $\{Z(t);\, t\in\mathbb{R}\}$ are statistically independent.[3]

[3]More specifically, this means that for all $k > 0$, all epochs $t_1,\ldots,t_k$, and all epochs $\tau_1,\ldots,\tau_k$, the rvs $X(t_1),\ldots,X(t_k)$ are statistically independent of $Z(\tau_1),\ldots,Z(\tau_k)$.

As discussed earlier, the additive noise model $Y(t) = X(t) + Z(t)$ implicitly assumes that the channel attenuation, propagation delay, and carrier frequency and phase are perfectly known and compensated for. It also assumes that the input waveform is not changed by any disturbances other than the noise, $Z(t)$.

Additive noise is most frequently modeled as a Gaussian process, as discussed in the next section. Even when the noise is not modeled as Gaussian, it is often modeled as some modification of a Gaussian process. Many rules of thumb in engineering and statistics about noise are stated without any mention of Gaussian processes, but are often valid only for Gaussian processes.

7.3 Gaussian random variables, vectors, and processes

This section first defines Gaussian random variables (rv's), then jointly-Gaussian random vectors (r**v**'s), and finally Gaussian random processes. The covariance function and joint density function for Gaussian random vectors are then derived. Finally several equivalent conditions for r**v**'s to be jointly Gaussian are derived.

A rv W is a *normalized Gaussian* rv, or more briefly a *normal*[4] rv, if it has the probability density

$$f_W(w) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-w^2}{2}\right].$$

This density is symmetric around 0 and thus the mean of W is zero. The variance is 1, which is probably familiar from elementary probability and is demonstrated in Exercise 7.1. A random variable Z is a *Gaussian* rv if it is a scaled and shifted version of a normal rv, *i.e.*, if $Z = \sigma W + \bar{Z}$ for a normal rv W. It can be seen that $\bar{Z}$ is the mean of Z and σ^2 is the variance[5]. The density of Z (for $\sigma^2 > 0$) is

$$f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(z-\bar{Z})^2}{(2\sigma^2)}\right]. \tag{7.9}$$

A Gaussian rv Z of mean $\bar{Z}$ and variance σ^2 is denoted as $Z \sim \mathcal{N}(\bar{Z}, \sigma^2)$. The Gaussian rv's used to represent noise are almost invariably zero-mean. Such rv's have the density $f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[\frac{-z^2}{2\sigma^2}]$ and are denoted by $Z \sim \mathcal{N}(0, \sigma^2)$.

Zero-mean Gaussian rv's are important in modeling noise and other random phenomena for the following reasons:

- They serve as good approximations to the sum of many independent zero-mean rv's (recall the central limit theorem).
- They have a number of extremal properties; as discussed later, they are, in several senses, the most random rv's for a given variance.
- They are easy to manipulate analytically, given a few simple properties.
- They serve as common channel noise models, and in fact the literature often assumes that noise is modeled by zero-mean Gaussian rv's without explicitly stating it.

[4]Some people use normal rv as a synonym for Gaussian rv.

[5]It is convenient to denote Z as Gaussian even in the deterministic case where $\sigma = 0$, but (7.9) is invalid then.

Definition 7.3.1. A set of n of random variables, $Z_1, \ldots, Z_n$ is *zero-mean jointly Gaussian* if there is a set of iid normal rv's $W_1, \ldots, W_\ell$ such that each Z_k, $1 \le k \le n$, can be expressed as

$$Z_k = \sum_{m=1}^{\ell} a_{km} W_m; \qquad 1 \le k \le n, \tag{7.10}$$

where $\{a_{km};\, 1{\le}k{\le}n,\, 1{\le}m{\le}\ell\}$ is an array of real numbers. $Z'_1, \ldots, Z'_n$ is *jointly Gaussian* if $Z'_k = Z_k + \bar{Z}'_k$ where the set $Z_1, \ldots, Z_n$ is zero-mean jointly Gaussian and $\bar{Z}'_1, \ldots, \bar{Z}'_n$ is a set of real numbers.

It is convenient notationally to refer to a set of n random variables, $Z_1, \ldots, Z_n$ as a random vector[6] (rv) $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\mathsf{T}}$. Letting A be the n by ℓ real matrix with elements $\{a_{km};\, 1{\le}k{\le}n,\, 1{\le}m{\le}\ell\}$, (7.10) can then be represented more compactly as

$$\boldsymbol{Z} = \mathsf{A}\,\boldsymbol{W}. \tag{7.11}$$

Similarly the jointly-Gaussian random vector $\boldsymbol{Z}'$ above can be represented as $\boldsymbol{Z}' = \mathsf{A}\boldsymbol{Z} + \bar{\boldsymbol{Z}}'$ where $\bar{\boldsymbol{Z}}'$ is an n-vector of real numbers.

In the remainder of this chapter, all random variables, random vectors, and random processes are assumed to be zero-mean unless explicitly designated otherwise. Viewed differently, only the fluctuations are analyzed with the means added at the end[7].

It is shown in Exercise 7.2 that any sum $\sum_m a_{km} W_m$ of iid normal rv's $W_1, \ldots, W_n$ is a Gaussian rv, so that each Z_k in (7.10) is Gaussian. Jointly Gaussian means much more than this, however. The random variables $Z_1, \ldots, Z_n$ must also be related as linear combinations of the same set of iid normal variables. Exercises 7.3 and 7.4 illustrate some examples of pairs of random variables which are individually Gaussian but not jointly Gaussian. These examples are slightly artificial, but illustrate clearly that the joint density of jointly-Gaussian rv's is much more constrained than the possible joint densities arising from constraining marginal distributions to be Gaussian.

The above definition of jointly Gaussian looks a little contrived at first, but is in fact very natural. Gaussian rv's often make excellent models for physical noise processes because noise is often the summation of many small effects. The central limit theorem is a mathematically precise way of saying that the sum of a very large number of independent small zero-mean random variables is approximately zero-mean Gaussian. Even when different sums are statistically dependent on each other, they are different linear combinations of a common set of independent small random variables. Thus the jointly-Gaussian assumption is closely linked to the assumption that the noise is the sum of a large number of small, essentially independent, random disturbances. Assuming that the underlying variables are Gaussian simply makes the model analytically clean and tractable.

An important property of any jointly-Gaussian n-dimensional rv $\boldsymbol{Z}$ is the following: for any real m by n real matrix B, the rv $\boldsymbol{Y} = \mathsf{B}\boldsymbol{Z}$ is also jointly Gaussian. To see this, let $\boldsymbol{Z} = \mathsf{A}\,\boldsymbol{W}$ where $\boldsymbol{W}$ is a normal rv. Then

$$\boldsymbol{Y} = \mathsf{B}\boldsymbol{Z} = \mathsf{B}(\mathsf{A}\,\boldsymbol{W}) = (\mathsf{BA})\,\boldsymbol{W}. \tag{7.12}$$

[6]The class of random vectors for a given n over a given sample space satisfies the axioms of a vector space, but here the vector notation is used simpy as a notational convenience.

[7]When studying estimation and conditional probabilities, means become an integral part of many arguments, but these arguments will not be central here.

Since BA is a real matrix, $\boldsymbol{Y}$ is jointly Gaussian. A useful application of this property arises when A is diagonal, so $\boldsymbol{Z}$ has arbitrary independent Gaussian components. This implies that $\boldsymbol{Y} = \mathsf{B}\boldsymbol{Z}$ is jointly Gaussian whenever a rv $\boldsymbol{Z}$ has independent Gaussian components.

Another important application is where B is a 1 by n matrix and Y is a random variable. Thus every linear combination $\sum_{k=1}^{n} b_k Z_k$ of a jointly-Gaussian rv $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\mathsf{T}$ is Gaussian. It will be shown later in this section that this is an if and only if property; that is, if every linear combination of a rv $\boldsymbol{Z}$ is Gaussian, then $\boldsymbol{Z}$ is jointly Gaussian.

We now have the machinery to define zero-mean Gaussian processes.

Definition 7.3.2. $\{Z(t); t \in \mathbb{R}\}$ is a *zero-mean Gaussian process* if, for all positive integers n and all finite sets of epochs $t_1, \ldots, t_n$, the set of random variables $Z(t_1), \ldots, Z(t_n)$ is a (zero-mean) jointly-Gaussian set of random variables.

If the covariance, $\mathsf{K}_Z(t, \tau) = \mathsf{E}[Z(t)Z(\tau)]$, is known for each pair of epochs t, τ, then for any finite set of epochs $t_1, \ldots, t_n$, $\mathsf{E}[Z(t_k)Z(t_m)]$ is known for each pair (t_k, t_m) in that set. The next two subsections will show that the joint probability density for any such set of (zero-mean) jointly-Gaussian rv's depends only on the covariances of those variables. This will show that a zero-mean Gaussian process is specified by its covariance function. A nonzero-mean Gaussian process is similarly specified by its covariance function and its mean.

7.3.1 The covariance matrix of a jointly-Gaussian random vector

Let an n-tuple of (zero-mean) random variables (rv's) $Z_1, \ldots, Z_n$ be represented as a random vector (rv) $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\mathsf{T}$. As defined in the previous section, $\boldsymbol{Z}$ is jointly Gaussian if $\boldsymbol{Z} = \mathsf{A}\boldsymbol{W}$ where $\boldsymbol{W} = (W_1, W_2, \ldots, W_\ell)^\mathsf{T}$ is a vector of iid normal rv's and A is an n by ℓ real matrix. Each rv Z_k, and all linear combinations of $Z_1, \ldots, Z_n$, are Gaussian.

The covariance of two (zero-mean) rv's Z_1, Z_2 is $\mathsf{E}[Z_1 Z_2]$. For a rv $\boldsymbol{Z} = (Z_1, \ldots Z_n)^\mathsf{T}$ the covariance between all pairs of random variables is very conveniently represented by the n by n covariance matrix,

$$\mathsf{K}_{\boldsymbol{Z}} = \mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^\mathsf{T}].$$

Appendix 7A.1 develops a number of properties of covariance matrices (including the fact that they are identical to the class of nonnegative definite matrices). For a vector $\boldsymbol{W} = W_1, \ldots, W_\ell$ of independent normalized Gaussian rv's, $\mathsf{E}[W_j W_m] = 0$ for $j \neq m$ and 1 for $j = m$. Thus

$$\mathsf{K}_{\boldsymbol{W}} = \mathsf{E}[\boldsymbol{W}\boldsymbol{W}^\mathsf{T}] = \mathsf{I}_\ell,$$

where I_ℓ is the ℓ by ℓ identity matrix. For a zero-mean jointly-Gaussian vector $\boldsymbol{Z} = \mathsf{A}\boldsymbol{W}$, the covariance matrix is thus

$$\mathsf{K}_{\boldsymbol{Z}} = \mathsf{E}[\mathsf{A}\boldsymbol{W}\boldsymbol{W}^\mathsf{T}\mathsf{A}^\mathsf{T}] = \mathsf{A}\mathsf{E}[\boldsymbol{W}\boldsymbol{W}^\mathsf{T}]\mathsf{A}^\mathsf{T} = \mathsf{A}\mathsf{A}^\mathsf{T}. \tag{7.13}$$

The *probability density*, $f_{\boldsymbol{Z}}(\boldsymbol{z})$, of a rv $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)^\mathsf{T}$ is the joint probability density of the components $Z_1, \ldots, Z_n$. An important example is the iid rv $\boldsymbol{W}$ where the components W_k, $1 \leq k \leq n$, are iid and normal, $W_k \sim \mathcal{N}(0, 1)$. By taking the product of the n densities of the individual random variables, the density of $\boldsymbol{W} = (W_1, W_2, \ldots, W_n)^\mathsf{T}$ is

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{1}{(2\pi)^{n/2}} \exp\left(\frac{-w_1^2 - w_2^2 - \cdots - w_n^2}{2}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(\frac{-\|\boldsymbol{w}\|^2}{2}\right). \tag{7.14}$$

This shows that the density of $\boldsymbol{W}$ at a sample value $\boldsymbol{w}$ depends only on the squared distance $\|\boldsymbol{w}\|^2$ of the sample value from the origin. That is, $f_{\boldsymbol{W}}(\boldsymbol{w})$ is spherically symmetric around the origin, and points of equal probability density lie on concentric spheres around the origin.

7.3.2 The probability density of a jointly-Gaussian random vector

Consider the transformation $\boldsymbol{Z} = \mathsf{A}\boldsymbol{W}$ where $\boldsymbol{Z}$ and $\boldsymbol{W}$ each have n components and A is n by n. If we let $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ be the n columns of A, then this means that $\boldsymbol{Z} = \sum_m \boldsymbol{a}_m W_m$. That is, for any sample values $w_1, \ldots w_n$ for $\boldsymbol{W}$, the corresponding sample value for $\boldsymbol{Z}$ is $\boldsymbol{z} = \sum_m \boldsymbol{a}_m w_m$. Similarly, if we let $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ be the rows of A, then $Z_k = \boldsymbol{b}_k \boldsymbol{W}$.

Let $\mathcal{B}_\delta$ be a cube, δ on a side, of the sample values of $\boldsymbol{W}$ defined by $\mathcal{B}_\delta = \{\boldsymbol{w} : 0 \leq w_k \leq \delta;\ 1 \leq k \leq n\}$ (see Figure 7.1). The set $\mathcal{B}'_\delta$ of vectors $\boldsymbol{z} = \mathsf{A}\boldsymbol{w}$ for $\boldsymbol{w} \in \mathcal{B}_\delta$ is a parallepiped whose sides are the vectors $\delta\boldsymbol{a}_1, \ldots, \delta\boldsymbol{a}_n$. The determinant, $\det(\mathsf{A})$, of A has the remarkable geometric property that its magnitude, $|\det(\mathsf{A})|$, is equal to the volume of the parallelepiped with sides $\boldsymbol{a}_k;\ 1 \leq k \leq n$. Thus the unit cube $\mathcal{B}_\delta$ above, with volume δ^n, is mapped by A into a parallelepiped of volume $|\det \mathsf{A}|\delta^n$.

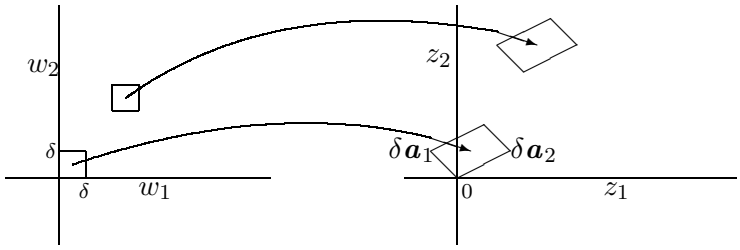


Figure 7.1: Example illustrating how $\boldsymbol{Z} = \mathsf{A}\boldsymbol{W}$ maps cubes into parallelepipeds. Let $Z_1 = -W_1 + 2W_2$ and $Z_2 = W_1 + W_2$. The figure shows the set of sample pairs z_1, z_2 corresponding to $0 \leq w_1 \leq \delta$ and $0 \leq w_2 \leq \delta$. It also shows a translation of the same cube mapping into a translation of the same parallelepiped.

Assume that the columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are linearly independent. This means that the columns must form a basis for $\mathbb{R}^n$, and thus that every vector $\boldsymbol{z}$ is some linear combination of these columns, *i.e.*, that $\boldsymbol{z} = \mathsf{A}\boldsymbol{w}$ for some vector $\boldsymbol{w}$. The matrix A must then be invertible, *i.e.*, there is a matrix A^{-1} such that $\mathsf{A}\mathsf{A}^{-1} = \mathsf{A}^{-1}\mathsf{A} = \mathsf{I}_n$ where I_n is the n by n identity matrix. The matrix A maps the unit vectors of $\mathbb{R}^n$ into the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ and the matrix A^{-1} maps $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ back into the unit vectors.

If the columns of A are not linearly independent, *i.e.*, A is not invertible, then A maps the unit cube in $\mathbb{R}^n$ into a subspace of dimension less than n. In terms of Fig. 7.1, the unit cube would be mapped into a straight line segment. The area, in 2 dimensional space, of a straight line segment is 0, and more generally, the volume in n-space of a lower dimensional set of points is 0. In terms of the determinant, $\det \mathsf{A} = 0$ for any noninvertible matrix A.

Assuming again that A is invertible, let $\boldsymbol{z}$ be a sample value of $\boldsymbol{Z}$, and let $\boldsymbol{w} = \mathsf{A}^{-1}\boldsymbol{z}$ be the corresponding sample value of $\boldsymbol{W}$. Consider the incremental cube $\boldsymbol{w} + \mathcal{B}_\delta$ cornered at $\boldsymbol{w}$. For δ very small, the probability $P_\delta(\boldsymbol{w})$ that $\boldsymbol{W}$ lies in this cube is $f_{\boldsymbol{W}}(\boldsymbol{w})\delta^n$ plus terms that go to zero faster than δ^n as $\delta \to 0$. This cube around $\boldsymbol{w}$ maps into a parallelepiped of volume $\delta^n|\det(\mathsf{A})|$ around $\boldsymbol{z}$, and no other sample value of $\boldsymbol{W}$ maps into this parallelepiped. Thus $P_\delta(\boldsymbol{w})$ is also

equal to $f_{\boldsymbol{Z}}(\boldsymbol{z})\delta^n|\det(\mathsf{A})|$ plus negligible terms. Going to the limit $\delta \to 0$, we have

$$f_{\boldsymbol{Z}}(\boldsymbol{z})|\det(\mathsf{A})| = \lim_{\delta\to 0} \frac{P_\delta(\boldsymbol{w})}{\delta^n} = f_{\boldsymbol{W}}(\boldsymbol{w}). \tag{7.15}$$

Since $\boldsymbol{w} = \mathsf{A}^{-1}\boldsymbol{z}$, we get the explicit formula

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{f_{\boldsymbol{W}}(\mathsf{A}^{-1}\boldsymbol{z})}{|\det(\mathsf{A})|}. \tag{7.16}$$

This formula is valid for any random vector $\boldsymbol{W}$ with a density, but we are interested in the vector $\boldsymbol{W}$ of iid Gaussian random variables, $\mathcal{N}(0,1)$. Substituting (7.14) into (7.16),

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}|\det(\mathsf{A})|} \exp\left(\frac{-\|\mathsf{A}^{-1}\boldsymbol{z}\|^2}{2}\right) \tag{7.17}$$

$$= \frac{1}{(2\pi)^{n/2}|\det(\mathsf{A})|} \exp\left[-\frac{1}{2}\boldsymbol{z}^{\mathsf{T}}(\mathsf{A}^{-1})^{\mathsf{T}}\mathsf{A}^{-1}\boldsymbol{z}\right] \tag{7.18}$$

We can simplify this somewhat by recalling from (7.13) that the covariance matrix of $\boldsymbol{Z}$ is given by $\mathsf{K}_{\boldsymbol{Z}} = \mathsf{A}\mathsf{A}^{\mathsf{T}}$. Thus $\mathsf{K}_{\boldsymbol{Z}}^{-1} = (\mathsf{A}^{-1})^{\mathsf{T}}\mathsf{A}^{-1}$.

Substituting this into (7.18) and noting that $\det(\mathsf{K}_{\boldsymbol{Z}}) = |\det(\mathsf{A})|^2$,

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\mathsf{K}_{\boldsymbol{Z}})}} \exp\left[-\frac{1}{2}\boldsymbol{z}^{\mathsf{T}}\mathsf{K}_{\boldsymbol{Z}}^{-1}\boldsymbol{z}\right]. \tag{7.19}$$

Note that this probability density depends only on the covariance matrix of $\boldsymbol{Z}$ and not directly on the matrix A.

The above density relies on A being nonsingular. If A is singular, then at least one of its rows is a linear combination of the other rows, and thus, for some m, $1 \le m \le n$, Z_m is a linear combination of the other Z_k. The random vector $\boldsymbol{Z}$ is still jointly Gaussian, but the joint probability density does not exist (unless one wishes to view the density of Z_m as a unit impulse at a point specified by the sample values of the other variables). It is possible to write out the distribution function for this case, using step functions for the dependent rv's, but it is not worth the notational mess. It is more straightforward to face the problem and find the density of a maximal set of linearly independent rv's, and specify the others as deterministic linear combinations.

It is important to understand that there is a large difference between rv's being *statistically dependent* and *linearly dependent*. If they are linearly dependent, then one or more are deterministic functions of the others, whereas statistical dependence simply implies a probabilistic relationship.

These results are summarized in the following theorem:

Theorem 7.3.1 (Density for jointly-Gaussian rv's). *Let $\boldsymbol{Z}$ be a (zero-mean) jointly-Gaussian rv with a nonsingular covariance matrix $\mathsf{K}_{\boldsymbol{Z}}$. Then the probability density $f_{\boldsymbol{Z}}(\boldsymbol{z})$ is given by (7.19). If $\mathsf{K}_{\boldsymbol{Z}}$ is singular, then $f_{\boldsymbol{Z}}(\boldsymbol{z})$ does not exist but the density in (7.19) can be applied to any set of linearly independent rv's out of $Z_1, \ldots, Z_n$.*

For a zero-mean Gaussian process $Z(t)$, the covariance function $\mathsf{K}_Z(t,\tau)$ specifies $\mathsf{E}[Z(t_k)Z(t_m)]$ for arbitrary epochs t_k and t_m and thus specifies the covariance matrix for any finite set of epochs

$t_1, \ldots, t_n$. From the above theorem, this also specifies the joint probability distribution for that set of epochs. Thus the covariance function specifies all joint probability distributions for all finite sets of epochs, and thus specifies the process in the sense[8] of Section 7.2. In summary, we have the following important theorem.

Theorem 7.3.2 (Gaussian process). *A zero-mean Gaussian process is specified by its covariance function $\mathsf{K}(t,\tau)$.*

7.3.3 Special case of a 2-dimensional zero-mean Gaussian random vector

The probability density in (7.19) is now written out in detail for the 2-dimensional case. Let $\mathsf{E}[Z_1^2] = \sigma_1^2$, $\mathsf{E}[Z_2^2] = \sigma_2^2$ and $\mathsf{E}[Z_1 Z_2] = \kappa_{12}$. Thus

$$\mathsf{K}_{\boldsymbol{Z}} = \begin{bmatrix} \sigma_1^2 & \kappa_{12} \\ \kappa_{12} & \sigma_2^2 \end{bmatrix}.$$

Let ρ be the *normalized covariance* $\rho = \kappa_{12}/(\sigma_1\sigma_2)$. Then $\det(\mathsf{K}_{\boldsymbol{Z}}) = \sigma_1^2\sigma_2^2 - \kappa_{12}^2 = \sigma_1^2\sigma_2^2(1-\rho^2)$. Note that ρ must satisfy $|\rho| \le 1$, and $|\rho| < 1$ for $\mathsf{K}_{\boldsymbol{Z}}$ to be nonsingular.

$$\mathsf{K}_{\boldsymbol{Z}}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2 - \kappa_{12}^2}\begin{bmatrix} \sigma_2^2 & -\kappa_{12} \\ -\kappa_{12} & \sigma_1^2 \end{bmatrix} = \frac{1}{1-\rho^2}\begin{bmatrix} 1/\sigma_1^2 & -\rho/(\sigma_1\sigma_2) \\ -\rho/(\sigma_1\sigma_2) & 1/\sigma_2^2 \end{bmatrix}.$$

$$\begin{aligned} f_{\boldsymbol{Z}}(\boldsymbol{z}) &= \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2 - \kappa_{12}^2}} \exp\left(\frac{-z_1^2\sigma_2^2 + 2z_1z_2\kappa_{12} - z_2^2\sigma_1^2}{2(\sigma_1^2\sigma_2^2 - \kappa_{12}^2)}\right) \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{-(z_1/\sigma_1)^2 + 2\rho(z_1/\sigma_1)(z_2/\sigma_2) - (z_2/\sigma_2)^2}{2(1-\rho^2)}\right). \end{aligned} \tag{7.20}$$

Curves of equal probability density in the plane correspond to points where the argument of the exponential function in (7.20) is constant. This argument is quadratic and thus points of equal probability density form an ellipse centered on the origin. The ellipses corresponding to different values of probability density are concentric, with larger ellipses corresponding to smaller densities.

If the normalized covariance ρ is 0, the axes of the ellipse are the horizontal and vertical axes of the plane; if $\sigma_1 = \sigma_2$, the ellipse reduces to a circle, and otherwise the ellipse is elongated in the direction of the larger standard deviation. If $\rho > 0$, the density in the first and third quadrants is increased at the expense of the second and fourth, and thus the ellipses are elongated in the first and third quadrants. This is reversed, of course, for $\rho < 0$.

The main point to be learned from this example, however, is that the detailed expression for 2 dimensions in (7.20) is messy. The messiness gets far worse in higher dimensions. Vector notation is almost essential. One should reason directly from the vector equations and use standard computer programs for calculations.

[8] As will be discussed later, focusing on the pointwise behavior of a random process at all finite sets of epochs has some of the same problems as specifying a function pointwise rather than in terms of $\mathcal{L}_2$ equivalence. This can be ignored for the present.

7.3.4 $\boldsymbol{Z = \mathsf{A}\,W}$ where A is orthogonal

An n by n real matrix A for which $\mathsf{AA}^{\mathsf{T}} = \mathsf{I}_n$ is called an *orthogonal matrix* or *orthonormal matrix* (orthonormal is more appropriate, but orthogonal is more common). For $\boldsymbol{Z} = \mathsf{A}\,\boldsymbol{W}$, where $\boldsymbol{W}$ is iid normal and A is orthogonal, $\mathsf{K}_{\boldsymbol{Z}} = \mathsf{AA}^{\mathsf{T}} = \mathsf{I}_n$. Thus $\mathsf{K}_{\boldsymbol{Z}}^{-1} = \mathsf{I}_n$ also and (7.19) becomes

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{\exp\left[-\frac{1}{2}\boldsymbol{z}^{\mathsf{T}}\boldsymbol{z}\right]}{(2\pi)^{n/2}} = \prod_{k=1}^{n} \frac{\exp[-z_k^2/2]}{\sqrt{2\pi}}. \tag{7.21}$$

This means that A transforms $\boldsymbol{W}$ into a random vector $\boldsymbol{Z}$ with the same probability density, and thus the components of $\boldsymbol{Z}$ are still normal and iid. To understand this better, note that $\mathsf{AA}^{\mathsf{T}} = \mathsf{I}_n$ means that A^{T} is the inverse of A and thus that $\mathsf{A}^{\mathsf{T}}\mathsf{A} = \mathsf{I}_n$. Letting $\boldsymbol{a}_m$ be the m^{th} column of A, the equation $\mathsf{A}^{\mathsf{T}}\mathsf{A} = \mathsf{I}_n$ means that $\boldsymbol{a}_m^{\mathsf{T}}\boldsymbol{a}_j = \delta_{mj}$ for each m, j, $1 \leq m, j \leq n$, i.e., that the columns of A are orthonormal. Thus, for the two dimensional example, the unit vectors $\boldsymbol{e}_1, \boldsymbol{e}_2$ are mapped into orthonormal vectors $\boldsymbol{a}_1, \boldsymbol{a}_2$, so that the transformation simply rotates the points in the plane. Although it is difficult to visualize such a transformation in higher dimensional space, it is still called a rotation, and has the property that $||\mathsf{A}\boldsymbol{w}||^2 = \boldsymbol{w}^{\mathsf{T}}\mathsf{A}^{\mathsf{T}}\mathsf{A}\boldsymbol{w}$, which is just $\boldsymbol{w}^{\mathsf{T}}\boldsymbol{w} = ||\boldsymbol{w}||^2$. Thus, each point $\boldsymbol{w}$ maps into a point $\mathsf{A}\boldsymbol{w}$ at the same distance from the origin as itself.

Not only the columns of an orthogonal matrix are orthonormal, but the rows, say $\{\boldsymbol{b}_k;\, 1 \leq k \leq n\}$ are also orthonormal (as is seen directly from $\mathsf{AA}^{\mathsf{T}} = \mathsf{I}_n$). Since $Z_k = \boldsymbol{b}_k\,\boldsymbol{W}$, this means that, for any set of orthonormal vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$, the random variables $Z_k = \boldsymbol{b}_k\,\boldsymbol{W}$ are normal and iid for $1 \leq k \leq n$.

7.3.5 Probability density for Gaussian vectors in terms of principal axes

This subsection describes what is often a more convenient representation for the probability density of an n-dimensional (zero-mean) Gaussian r**v** $\boldsymbol{Z}$ with a nonsingular covariance matrix $\mathsf{K}_{\boldsymbol{Z}}$. As shown in Appendix 7A.1, the matrix $\mathsf{K}_{\boldsymbol{Z}}$ has n real orthonormal eigenvectors, $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$, with corresponding nonnegative (but not necessarily distinct[9]) real eigenvalues, $\lambda_1, \ldots, \lambda_n$. Also, for any vector $\boldsymbol{z}$, it is shown that $\boldsymbol{z}^{\mathsf{T}}\mathsf{K}_{\boldsymbol{Z}}^{-1}\boldsymbol{z}$ can be expressed as $\sum_k \lambda_k^{-1}|\langle \boldsymbol{z}, \boldsymbol{q}_k\rangle|^2$. Substituting this in (7.19), we have

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\mathsf{K}_{\boldsymbol{Z}})}} \exp\left[-\sum_k \frac{|\langle \boldsymbol{z}, \boldsymbol{q}_k\rangle|^2}{2\lambda_k}\right]. \tag{7.22}$$

Note that $\langle \boldsymbol{z}, \boldsymbol{q}_k\rangle$ is the projection of $\boldsymbol{z}$ on the kth of n orthonormal directions. The determinant of an n by n real matrix can be expressed in terms of the n eigenvalues (see Appendix 7A.1) as $\det(\mathsf{K}_{\boldsymbol{Z}}) = \prod_{k=1}^{n} \lambda_k$. Thus (7.22) becomes

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\lambda_k}} \exp\left[\frac{-|\langle \boldsymbol{z}, \boldsymbol{q}_k\rangle|^2}{2\lambda_k}\right]. \tag{7.23}$$

[9] If an eigenvalue λ has multiplicity m, it means that there is an m dimensional subspace of vectors $\boldsymbol{q}$ satisfying $\mathsf{K}_{\boldsymbol{Z}}\boldsymbol{q} = \lambda\boldsymbol{q}$; in this case any orthonormal set of m such vectors can be chosen as the m eigenvectors corresponding to that eigenvalue.

This is the product of n Gaussian densities. It can be interpreted as saying that the Gaussian random variables $\{\langle \boldsymbol{Z}, \boldsymbol{q}_k\rangle; 1 \le k \le n\}$ are statistically independent with variances $\{\lambda_k; 1 \le k \le n\}$. In other words, if we represent the r**v** $\boldsymbol{Z}$ using $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ as a basis, then the components of $\boldsymbol{Z}$ in that coordinate system are independent random variables. The orthonormal eigenvectors are called *principal axes* for $\boldsymbol{Z}$.

This result can be viewed in terms of the contours of equal probability density for $\boldsymbol{Z}$ (see Figure 7.2). Each such contour satisfies

$$c = \sum_k \frac{|\langle \boldsymbol{z}, \boldsymbol{q}_k\rangle|^2}{2\lambda_k}$$

where c is proportional to the log probability density for that contour. This is the equation of an ellipsoid centered on the origin, where $\boldsymbol{q}_k$ is the kth axis of the ellipsoid and $\sqrt{2c\lambda_k}$ is the length of that axis.

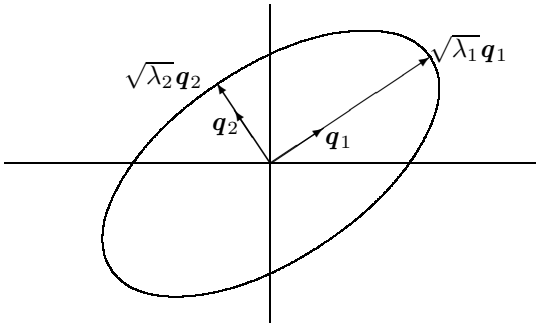


Figure 7.2: Contours of equal probability density. Points $\boldsymbol{z}$ on the $\boldsymbol{q}_1$ axis are points for which $\langle \boldsymbol{z}, \boldsymbol{q}_2\rangle = 0$ and points on the $\boldsymbol{q}_2$ axis satisfy $\langle \boldsymbol{z}, \boldsymbol{q}_1\rangle = 0$. Points on the illustrated ellipse satisfy $\boldsymbol{z}^{\mathsf{T}}\mathsf{K}_{\boldsymbol{Z}}^{-1}\boldsymbol{z} = 1$.

The probability density formulas in (7.19) and (7.23) suggest that for every covariance matrix K, there is a jointly Gaussian r**v** that has that covariance, and thus has that probability density. This is in fact true, but to verify it, we must demonstrate that for every covariance matrix K, there is a matrix A (and thus a r**v** $\boldsymbol{Z} = \mathsf{A}\boldsymbol{W}$) such that $\mathsf{K} = \mathsf{A}\mathsf{A}^{\mathsf{T}}$. There are many such matrices for any given K, but a particularly convenient one is given in (7.88). As a function of the eigenvectors and eigenvalues of K, it is $\mathsf{A} = \sum_k \sqrt{\lambda_k}\boldsymbol{q}_k\boldsymbol{q}_k^{\mathsf{T}}$. Thus, for every nonsingular covariance matrix, K, there is a jointly Gaussian r**v** whose density satisfies (7.19) and (7.23)

7.3.6 Fourier transforms for joint densities

As suggested in Exercise 7.2, Fourier transforms of probability densities are useful for finding the probability density of sums of independent random variables. More generally, for an n-dimensional r**v**, $\boldsymbol{Z}$, we can define the n-dimensional Fourier transform of $f_{\boldsymbol{Z}}(\boldsymbol{z})$ as

$$\hat{f}_{\boldsymbol{Z}}(\boldsymbol{s}) = \int \cdots \int f_{\boldsymbol{Z}}(\boldsymbol{z}) \exp(-2\pi i \boldsymbol{s}^{\mathsf{T}}\boldsymbol{z})\, dz_1 \cdots dz_n = \mathsf{E}[\exp(-2\pi i \boldsymbol{s}^{\mathsf{T}}\boldsymbol{Z})]. \tag{7.24}$$

If $\boldsymbol{Z}$ is jointly Gaussian, this is easy to calculate. For any given $\boldsymbol{s} \neq \boldsymbol{0}$, let $X = \boldsymbol{s}^{\mathsf{T}}\boldsymbol{Z} = \sum_k s_k Z_k$. Thus X is Gaussian with variance $\mathsf{E}[\boldsymbol{s}^{\mathsf{T}}\boldsymbol{Z}\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{s}] = \boldsymbol{s}^{\mathsf{T}}\mathsf{K}_{\boldsymbol{Z}}\boldsymbol{s}$. From Exercise 7.2,

$$\hat{f}_X(\theta) = \mathsf{E}[\exp(-2\pi i\theta \boldsymbol{s}^{\mathsf{T}}\boldsymbol{Z})] = \exp\left[-\frac{(2\pi\theta)^2 \boldsymbol{s}^{\mathsf{T}}\mathsf{K}_{\boldsymbol{Z}}\boldsymbol{s}}{2}\right]. \tag{7.25}$$

Comparing (7.25) for $\theta = 1$ with (7.24), we see that

$$\hat{f}_{\boldsymbol{Z}}(\boldsymbol{s}) = \exp\left[-\frac{(2\pi)^2 \boldsymbol{s}^{\mathsf{T}}\mathsf{K}_{\boldsymbol{Z}}\boldsymbol{s}}{2}\right]. \tag{7.26}$$

The above derivation also demonstrates that $\hat{f}_{\boldsymbol{Z}}(\boldsymbol{s})$ is determined by the Fourier transform of each linear combination of the elements of $\boldsymbol{Z}$. In other words, if an arbitrary r**v** $\boldsymbol{Z}$ has covariance $\mathsf{K}_{\boldsymbol{Z}}$ and has the property that all linear combinations of $\boldsymbol{Z}$ are Gaussian, then the Fourier transform of its density is given by (7.26). Thus, assuming that the Fourier transform of the density uniquely specifies the density, $\boldsymbol{Z}$ must be jointly Gaussian if all linear combinations of $\boldsymbol{Z}$ are Gaussian.

A number of equivalent conditions have now been derived under which a (zero-mean) random vector $\boldsymbol{Z}$ is jointly Gaussian. In summary, each of the following are necessary and sufficient conditions for a r**v** $\boldsymbol{Z}$ with a nonsingular covariance $\mathsf{K}_{\boldsymbol{Z}}$ to be jointly Gaussian.

- $\boldsymbol{Z} = \mathsf{A}\boldsymbol{W}$ where the components of $\boldsymbol{W}$ are iid normal and $\mathsf{K}_{\boldsymbol{Z}} = \mathsf{A}\mathsf{A}^{\mathsf{T}}$;
- $\boldsymbol{Z}$ has the joint probability density given in (7.19);
- $\boldsymbol{Z}$ has the joint probability density given in (7.23);
- All linear combinations of $\boldsymbol{Z}$ are Gaussian random variables.

For the case where $\mathsf{K}_{\boldsymbol{Z}}$ is singular, the above conditions are necessary and sufficient for any linearly independent subset of the components of $\boldsymbol{Z}$.

This section has considered only zero-mean random variables, vectors, and processes. The results here apply directly to the fluctuation of arbitrary random variables, vectors, and processes. In particular the probability density for a jointly Gaussian r**v** $\boldsymbol{Z}$ with a nonsingular covariance matrix $\mathsf{K}_{\boldsymbol{Z}}$ and mean vector $\overline{\boldsymbol{Z}}$ is

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\mathsf{K}_{\boldsymbol{Z}})}} \exp\left[-\frac{1}{2}(\boldsymbol{z}-\overline{\boldsymbol{Z}})^{\mathsf{T}}\mathsf{K}_{\boldsymbol{Z}}^{-1}(\boldsymbol{z}-\overline{\boldsymbol{Z}})\right]. \tag{7.27}$$

7.4 Linear functionals and filters for random processes

This section defines the important concept of linear functionals on arbitrary random processes $\{Z(t); t \in \mathbb{R}\}$ and then specializes to Gaussian random processes, where the results of the previous section can be used. Assume that the sample functions $Z(t,\omega)$ of $Z(t)$ are real $\mathcal{L}_2$ waveforms. These sample functions can be viewed as vectors over $\mathbb{R}$ in the $\mathcal{L}_2$ space of real waveforms. For any given real $\mathcal{L}_2$ waveform $g(t)$, there is an inner product,

$$\langle Z(t,\omega), g(t)\rangle = \int_{-\infty}^{\infty} Z(t,\omega)g(t)\,dt.$$

By the Schwarz inequality, the magnitude of this inner product in the space of real $\mathcal{L}_2$ functions is upper bounded by $\|Z(t,\omega)\|\|g(t)\|$ and is thus a finite real value for each ω. This then maps

sample points ω into real numbers and is thus a random variable,[10] denoted $V = \int_{-\infty}^{\infty} Z(t)g(t)\,dt$. This random variable V is called a *linear functional* of the process $\{Z(t); t \in \mathbb{R}\}$.

As an example of the importance of linear functionals, recall that the demodulator for both PAM and QAM contains a filter $q(t)$ followed by a sampler. The output of the filter at a sampling time kT for an input $u(t)$ is $\int u(t)q(kT-t)\,dt$. If the filter input also contains additive noise $Z(t)$, then the output at time kT also contains the linear functional $\int Z(t)q(kT-t)\,dt$.

Similarly, for any random process $\{Z(t); t \in \mathbb{R}\}$ (again assuming $\mathcal{L}_2$ sample functions) and any real $\mathcal{L}_2$ function $h(t)$, consider the result of passing $Z(t)$ through the filter with impulse response $h(t)$. For any $\mathcal{L}_2$ sample function $Z(t,\omega)$, the filter output at any given time τ is the inner product

$$\langle Z(t,\omega), h(\tau - t)\rangle = \int_{-\infty}^{\infty} Z(t,\omega)h(\tau - t)\,dt.$$

For each real τ, this maps sample points ω into real numbers and thus (aside from measure theoretic issues),

$$V(\tau) = \int Z(t)h(\tau - t)\,dt \tag{7.28}$$

is a rv for each τ. This means that $\{V(\tau); \tau \in \mathbb{R}\}$ is a random process. This is called the *filtered process* resulting from passing $Z(t)$ through the filter $h(t)$. Not much can be said about this general problem without developing a great deal of mathematics, so instead we restrict ourselves to Gaussian processes and other relatively simple examples.

For a Gaussian process, we would hope that a linear functional is a Gaussian random variable. The following examples show that some restrictions are needed even on the class of Gaussian processes.

Example 7.4.1. Let $Z(t) = tX$ for all $t \in \mathbb{R}$ where $X \sim \mathcal{N}(0,1)$. The sample functions of this Gaussian process have infinite energy with probability 1. The output of the filter also has infinite energy except except for very special choices of $h(t)$.

Example 7.4.2. For each $t \in [0,1]$, let $W(t)$ be a Gaussian rv, $W(t) \sim \mathcal{N}(0,1)$. Assume also that $\mathsf{E}[W(t)W(\tau)] = 0$ for each $t \neq \tau \in [0,1]$. The sample functions of this process are discontinuous everywhere[11]. We do not have the machinery to decide whether the sample functions are integrable, let alone whether the linear functionals above exist; we come back later to further discuss this example.

In order to avoid the mathematical issues in Example 7.4.2 above, along with a host of other mathematical issues, we start with Gaussian processes defined in terms of orthonormal expansions.

[10] One should use measure theory over the sample space Ω to interpret these mappings carefully, but this is unnecessary for the simple types of situations here and would take us too far afield.

[11] Even worse, the sample functions are not measurable. This process would not even be called a random process in a measure theoretic formulation, but it provides an interesting example of the occasional need for a measure theoretic formulation.

7.4.1 Gaussian processes defined over orthonormal expansions

Let $\{\phi_k(t);\, k \geq 1\}$ be a countable set of real orthonormal functions and let $\{Z_k;\, k \geq 1\}$ be a sequence of independent Gaussian random variables, $\{\mathcal{N}(0, \sigma_k^2)\}$. Consider the Gaussian process defined by

$$Z(t) = \sum_{k=1}^{\infty} Z_k \phi_k(t). \tag{7.29}$$

Essentially all zero-mean Gaussian processes of interest can be defined this way, although we will not prove this. Clearly a mean can be added if desired, but zero-mean processes are assumed in what follows. First consider the simple case in which σ_k^2 is nonzero for only finitely many values of k, say $1 \leq k \leq n$. In this case, $Z(t)$, for each $t \in \mathbb{R}$, is a finite sum,

$$Z(t) = \sum_{k=1}^{n} Z_k \phi_k(t), \tag{7.30}$$

of independent Gaussian rv's and thus is Gaussian. It is also clear that $Z(t_1), Z(t_2), \ldots Z(t_\ell)$ are jointly Gaussian for all $\ell, t_1, \ldots, t_\ell$, so $\{Z(t); t \in \mathbb{R}\}$ is in fact a Gaussian random process. The energy in any sample function, $z(t) = \sum_k z_k \phi_k(t)$ is $\sum_{k=1}^n z_k^2$. This is finite (since the sample values are real and thus finite), so every sample function is $\mathcal{L}_2$. The covariance function is then easily calculated to be

$$\mathsf{K}_{\boldsymbol{Z}}(t, \tau) = \sum_{k,m} \mathsf{E}[Z_k Z_m] \phi_k(t) \phi_m(\tau) = \sum_{k=1}^{n} \sigma_k^2\, \phi_k(t) \phi_k(\tau). \tag{7.31}$$

Next consider the linear functional $\int Z(t) g(t)\, dt$ where $g(t)$ is a real $\mathcal{L}_2$ function,

$$V = \int_{-\infty}^{\infty} Z(t) g(t)\, dt = \sum_{k=1}^{n} Z_k \int_{-\infty}^{\infty} \phi_k(t) g(t)\, dt. \tag{7.32}$$

Since V is a weighted sum of the zero-mean independent Gaussian rv's $Z_1, \ldots, Z_n$, V is also Gaussian with variance

$$\sigma_V^2 = \mathsf{E}[V^2] = \sum_{k=1}^{n} \sigma_k^2 |\langle \boldsymbol{\phi}_k, \boldsymbol{g} \rangle|^2 . \tag{7.33}$$

Next consider the case where n is infinite but $\sum_k \sigma_k^2 < \infty$. The sample functions are still $\mathcal{L}_2$ (at least with probability 1). Equations (7.29), (7.30), (7.31), (7.32) and (7.33) are still valid, and Z is still a Gaussian rv. We do not have the machinery to easily prove this, although Exercise 7.7 provides quite a bit of insight into why these results are true.

Finally, consider a finite set of $\mathcal{L}_2$ waveforms $\{g_m(t);\ 1 \leq m \leq \ell\}$. Let $V_m = \int_{-\infty}^{\infty} Z(t) g_m(t)\, dt$. By the same argument as above, V_m is a Gaussian rv for each m. Furthermore, since each linear combination of these variables is also a linear functional, it is also Gaussian, so $\{V_1, \ldots, V_\ell\}$ is jointly Gaussian.

7.4.2 Linear filtering of Gaussian processes

We can use the same argument as in the previous subsection to look at the output of a linear filter for which the input is a Gaussian process $\{Z(t); t \in \mathbb{R}\}$. In particular, assume that $Z(t) = \sum_k Z_k \phi_k(t)$ where $Z_1, Z_2, \ldots$ is an independent sequence $\{Z_k \sim \mathcal{N}(0, \sigma_k^2\}$ satisfying $\sum_k \sigma_k^2 < \infty$ and where $\phi_1(t), \phi_2(t), \ldots$, is a sequence of orthonormal functions.

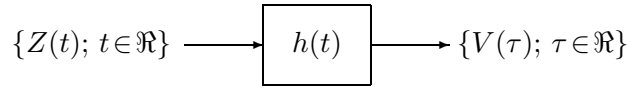


Figure 7.3: Filtered random Process

Assume that the impulse response $h(t)$ of the filter is a real $\mathcal{L}_2$ waveform. Then for any given sample function $Z(t, \omega) = \sum_k Z_k(\omega)\phi_k(t)$ of the input, the filter output at any epoch τ is given by

$$V(\tau, \omega) = \int_{-\infty}^{\infty} Z(t, \omega)h(\tau - t)\, dt = \sum_k Z_k(\omega) \int_{-\infty}^{\infty} \phi_k(t)h(\tau - t)\, dt. \tag{7.34}$$

Each integral on the right side of (7.34) is an $\mathcal{L}_2$ function of τ whose energy is upper bounded by $\|\boldsymbol{h}\|^2$ (see Exercise 7.5). It follows from this (see Exercise 7.7) that $\int_{-\infty}^{\infty} Z(t, \omega)h(\tau - t)\, dt$ is an $\mathcal{L}_2$ waveform with probability 1. For any given epoch τ, (7.34) maps sample points ω to real values and thus $V(\tau, \omega)$ is a sample value of a random variable $V(\tau)$.

$$V(\tau) = \int_{-\infty}^{\infty} Z(t)h(\tau - t)\, dt = \sum_k Z_k \int_{-\infty}^{\infty} \phi_k(t)h(\tau - t)\, dt. \tag{7.35}$$

This is a Gaussian rv for each epoch τ. For any set of epochs, $\tau_1, \ldots, \tau_\ell$, we see that $V(\tau_1), \ldots, V(\tau_\ell)$ are jointly Gaussian. Thus $\{V(\tau); \tau \in \mathbb{R}\}$ is a Gaussian random process.

We summarize the last two subsections in the following theorem.

Theorem 7.4.1. *Let $\{Z(t); t \in \mathbb{R}\}$ be a Gaussian process, $Z(t) = \sum_k Z_k \phi_k(t)$, where $\{Z_k; k \geq 1\}$ is a sequence of independent Gaussian rv's $\mathcal{N}(0, \sigma_k^2)$ where $\sum \sigma_k^2 < \infty$ and $\{\phi_k(t);\ k \geq 1\}$ is an orthonormal set. Then*

- *For any set of $\mathcal{L}_2$ waveforms $g_1(t), \ldots, g_\ell(t)$, the linear functionals $\{Z_m;\ 1 \leq m \leq \ell\}$ given by $Z_m = \int_{-\infty}^{\infty} Z(t)g_m(t)\, dt$ are zero-mean jointly Gaussian.*
- *For any filter with real $\mathcal{L}_2$ impulse response $h(t)$, the filter output $\{V(\tau); \tau \in \mathbb{R}\}$ given by (7.35) is a zero-mean Gaussian process.*

These are important results. The first, concerning sets of linear functionals, is important when we represent the input to the channel in terms of an orthonormal expansion; the noise can then often be expanded in the same orthonormal expansion. The second, concerning linear filtering, shows that when the received signal and noise are passed through a linear filter, the noise at the filter output is simply another zero-mean Gaussian process. This theorem is often summarized by saying that linear operations preserve Gaussianity.

7.4.3 Covariance for linear functionals and filters

Assume that $\{Z(t); t \in \mathbb{R}\}$ is a random process and that $g_1(t), \ldots, g_\ell(t)$ are real $\mathcal{L}_2$ waveforms. We have seen that if $\{Z(t); t \in \mathbb{R}\}$ is Gaussian, then the linear functionals $V_1, \ldots, V_\ell$ given by

$V_m = \int_{-\infty}^{\infty} Z(t)g_m(t)\,dt$ are jointly Gaussian for $1 \le m \le \ell$. We now want to find the covariance for each pair V_j, V_m of these random variables. The result does not depend on the process $Z(t)$ being Gaussian. The computation is quite simple, although we omit questions of limits, interchanges of order of expectation and integration, etc. A more careful derivation could be made by returning to the sampling theorem arguments before, but this would somewhat obscure the ideas. Assuming that the process $Z(t)$ is zero mean,

$$\begin{aligned}
\mathsf{E}[V_jV_m] &= \mathsf{E}\left[\int_{-\infty}^{\infty} Z(t)g_j(t)\,dt \int_{-\infty}^{\infty} Z(\tau)g_m(\tau)\,d\tau\right] && (7.36)\\
&= \int_{t=-\infty}^{\infty}\int_{\tau=-\infty}^{\infty} g_j(t)\mathsf{E}[Z(t)Z(\tau)]g_m(\tau)\,dt\,d\tau && (7.37)\\
&= \int_{t=-\infty}^{\infty}\int_{\tau=-\infty}^{\infty} g_j(t)\mathsf{K}_{\boldsymbol{Z}}(t,\tau)g_m(\tau)\,dt\,d\tau. && (7.38)
\end{aligned}$$

Each covariance term (including $\mathsf{E}[V_m^2]$ for each m) then depends only on the covariance function of the process and the set of waveforms $\{\boldsymbol{g}_m; 1 \le m \le \ell\}$.

The convolution $V(r) = \int Z(t)h(r-t)\,dt$ is a linear functional at each time r, so the covariance for the filtered output of $\{Z(t); t \in \mathbb{R}\}$ follows in the same way as the results above. The output $\{V(r)\}$ for a filter with a real $\mathcal{L}_2$ impulse response $\boldsymbol{h}$ is given by (7.35), so the covariance of the output can be found as

$$\begin{aligned}
\mathsf{K}_{\boldsymbol{V}}(r,s) &= \mathsf{E}[V(r)V(s)]\\
&= \mathsf{E}\left[\int_{-\infty}^{\infty} Z(t)h(r-t)dt \int_{-\infty}^{\infty} Z(\tau)h(s-\tau)d\tau\right]\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(r-t)\mathsf{K}_{\boldsymbol{Z}}(t,\,\tau)h(s-\tau)dtd\tau. && (7.39)
\end{aligned}$$

7.5 Stationarity and related concepts

Many of the most useful random processes have the property that the location of the time origin is irrelevant, *i.e.*, they "behave" the same way at one time as at any other time. This property is called *stationarity* and such a process is called a *stationary process*.

Since the location of the time origin must be irrelevant for stationarity, random processes that are defined over any interval other than $(-\infty, \infty)$ cannot be stationary. Thus assume a process that is defined over $(-\infty, \infty)$.

The next requirement for a random process $\{Z(t); t \in \mathbb{R}\}$ to be stationary is that $Z(t)$ must be identically distributed for all epochs $t \in \mathbb{R}$. This means that, for any epochs t and $t+\tau$, and for any real number x, $\Pr\{Z(t) \le x\} = \Pr\{Z(t+\tau) \le x\}$. This does not mean that $Z(t)$ and $Z(t+\tau)$ are the same random variables; for a given sample outcome ω of the experiment, $Z(t,\omega)$ is typically unequal to $Z(t+\tau,\omega)$. It simply means that $Z(t)$ and $Z(t+\tau)$ have the same distribution function, *i.e.*,

$$F_{Z(t)}(x) = F_{Z(t+\tau)}(x) \qquad \text{for all } x. \tag{7.40}$$

This is still not enough for stationarity, however. The joint distributions over any set of epochs must remain the same if all those epochs are shifted to new epochs by an arbitrary shift τ. This includes the previous requirement as a special case, so we have the definition:

Definition 7.5.1. A random process $\{Z(t);\, t\in\mathbb{R}\}$ is *stationary* if, for all positive integers ℓ, for all sets of epochs $t_1,\ldots,t_\ell \in \mathbb{R}$, for all amplitudes $z_1,\ldots,z_\ell$, and for all shifts $\tau \in \mathbb{R}$,

$$F_{Z(t_1),\ldots,Z(t_\ell)}(z_1\ldots,z_\ell) = F_{Z(t_1+\tau),\ldots,Z(t_\ell+\tau)}(z_1\ldots,z_\ell). \tag{7.41}$$

For the typical case where densities exist, this can be rewritten as

$$f_{Z(t_1),\ldots,Z(t_\ell)}(z_1\ldots,z_\ell) = f_{Z(t_1+\tau),\ldots,Z(t_\ell+\tau)}(z_1\ldots,z_\ell) \tag{7.42}$$

for all $z_1,\ldots,z_\ell \in \mathbb{R}$.

For a (zero-mean) Gaussian process, the joint distribution of $Z(t_1),\ldots,Z(t_\ell)$ depends only on the covariance of those variables. Thus, this distribution will be the same as that of $Z(t_1{+}\tau)$, $\ldots, Z(t_\ell{+}\tau)$ if $\mathsf{K}_Z(t_m,t_j) = \mathsf{K}_Z(t_m{+}\tau, t_j{+}\tau)$ for $1\le m,j\le \ell$. This condition will be satisfied for all τ, all ℓ, and all $t_1,\ldots,t_\ell$ (verifying that $\{Z(t)\}$ is stationary) if $\mathsf{K}_Z(t_1,t_2) = \mathsf{K}_Z(t_1{+}\tau, t_2{+}\tau)$ for all τ and all t_1,t_2. This latter condition will be satisfied if $\mathsf{K}_{\boldsymbol{Z}}(t_1,t_2) = \mathsf{K}_{\boldsymbol{Z}}(t_1{-}t_2,0)$ for all t_1,t_2. We have thus shown that a zero-mean Gaussian process is stationary if

$$\mathsf{K}_{\boldsymbol{Z}}(t_1,t_2) = \mathsf{K}_{\boldsymbol{Z}}(t_1{-}t_2,0) \qquad \text{for all } t_1,t_2\in\mathbb{R}. \tag{7.43}$$

Conversely, if (7.43) is not satisfied for some choice of t_1,t_2, then the joint distribution of $Z(t_1),Z(t_2)$ must be different from that of $Z(t_1{-}t_2),Z(0)$, and the process is not stationary. The following theorem summarizes this.

Theorem 7.5.1. *A zero-mean Gaussian process $\{Z(t);\, t\in\mathbb{R}\}$ is stationary if and only if (7.43) is satisfied.*

An obvious consequence of this is that a Gaussian process with a nonzero mean is stationary if and only if its mean is constant and its fluctuation satisfies (7.43).

7.5.1 Wide-sense stationary (WSS) random processes

There are many results in probability theory that depend only on the covariances of the random variables of interest (and also the mean if nonzero). For random processes, a number of these classical results are simplified for stationary processes, and these simplifications depend only on the mean and covariance of the process rather than full stationarity. This leads to the following definition:

Definition 7.5.2. A random process $\{Z(t);\, t\in\Re\}$ is *wide-sense stationary (WSS)* if $\mathsf{E}[Z(t_1)] = \mathsf{E}[Z(0)]$ and $\mathsf{K}_Z(t_1,t_2) = \mathsf{K}_Z(t_1{-}t_2,0)$ for all $t_1,t_2\in\mathbb{R}$.

Since the covariance function $\mathsf{K}_Z(t{+}\tau,t)$ of a WSS process is a function of only one variable τ, we will often write the covariance function as a function of one variable, namely $\tilde{\mathsf{K}}_Z(\tau)$ in place of $\mathsf{K}_Z(t{+}\tau,t)$. In other words, the single variable in the single argument form represents the difference between the two arguments in two argument form. Thus for a WSS process, $\mathsf{K}_Z(t,\tau) = \mathsf{K}_Z(t{-}\tau,0) = \tilde{\mathsf{K}}_Z(t-\tau)$.

The random processes defined as expansions of T-spaced sinc functions have been discussed several times. In particular let

$$V(t) = \sum_k V_k \,\mathrm{sinc}\left(\frac{t-kT}{T}\right), \tag{7.44}$$

where $\{\dots, V_{-1}, V_0, V_1, \dots\}$ is a sequence of (zero-mean) iid rv's. As shown in 7.8, the covariance function for this random process is

$$\mathsf{K}_{\boldsymbol{V}}(t,\tau) = \sigma_V^2 \sum_k \text{sinc}\left(\frac{t-kT}{T}\right)\text{sinc}\left(\frac{\tau-kT}{T}\right), \tag{7.45}$$

where σ_V^2 is the variance of each V_k. The sum in (7.45), as shown below, is a function only of $t-\tau$, leading to the theorem:

Theorem 7.5.2 (Sinc expansion). *The random process in (7.44) is WSS. In addition, if the rv's $\{V_k; k \in \mathbb{Z}\}$ are iid Gaussian, the process is stationary. The covariance function is given by*

$$\tilde{K}_{\boldsymbol{V}}(t-\tau) = \sigma_V^2 \,\text{sinc}\left(\frac{t-\tau}{T}\right). \tag{7.46}$$

Proof: From the sampling theorem, any $\mathcal{L}_2$ function $u(t)$, baseband limited to $1/(2T)$, can be expanded as

$$u(t) = \sum_k u(kT)\text{sinc}\left(\frac{t-kT}{T}\right). \tag{7.47}$$

For any given τ, take $u(t)$ to be $\text{sinc}(\frac{t-\tau}{T})$. Substituting this in (7.47),

$$\text{sinc}\left(\frac{t-\tau}{T}\right) = \sum_k \text{sinc}\left(\frac{kT-\tau}{T}\right)\text{sinc}\left(\frac{t-kT}{T}\right) = \sum_k \text{sinc}\left(\frac{\tau-kT}{T}\right)\text{sinc}\left(\frac{t-kT}{T}\right). \tag{7.48}$$

Substituting this in (7.45) shows that the process is WSS with the stated covariance. As shown in subsection 7.4.1, $\{V(t); t \in \mathbb{R}\}$ is Gaussian if the rv's $\{V_k\}$ are Gaussian. From Theorem 7.5.1, this Gaussian process is stationary since it is WSS. □

Next consider another special case of the sinc expansion in which each V_k is binary, taking values ± 1 with equal probability. This corresponds to a simple form of a PAM transmitted waveform. In this case, $V(kT)$ must be ± 1, whereas for values of t between the sample points, $V(t)$ can take on a wide range of values. Thus this process is WSS but cannot be stationary. Similarly, any discrete distribution for each V_k creates a process that is WSS but not stationary.

There are not many important models of *noise* processes that are WSS but not stationary[12], despite the above example and the widespread usage of the term WSS. Rather, the notion of wide-sense stationarity is used to make clear, for some results, that they depend only on the mean and covariance, thus perhaps making it easier to understand them.

The Gaussian sinc expansion brings out an interesting theoretical nonsequitur. Assuming that $\sigma_V^2 > 0$, *i.e.*, that the process is not the trivial process for which $V(t) = 0$ with probability 1 for all t, the expected energy in the process (taken over all time) is infinite. It is not difficult to convince oneself that the sample functions of this process have infinite energy with probability 1. Thus stationary noise models are simple to work with, but the sample functions of these processes don't fit into the $\mathcal{L}_2$ theory of waveforms that has been developed. Even more important than the issue of infinite energy, stationary noise models make unwarranted assumptions about the

[12] An important exception is interference from other users, which as the above sinc expansion with binary samples shows, can be WSS but not stationary. Even in this case, if the interference is modeled as just part of the noise (rather than specifically as interference), the nonstationarity is usually ignored.

very distant past and future. The extent to which these assumptions affect the results about the present is an important question that must be asked.

The problem here is not with the peculiarities of the Gaussian sinc expansion. Rather it is that stationary processes have constant power over all time, and thus have infinite energy. One practical solution[13] to this is simple and familiar. The random process is simply truncated in any convenient way. Thus, when we say that noise is stationary, we mean that it is stationary within a much longer time interval than the interval of interest for communication. This is not very precise, and the notion of *effective stationarity* is now developed to formalize this notion of a truncated stationary process.

7.5.2 Effectively stationary and effectively WSS random processes

Definition 7.5.3. A (zero-mean) random process is *effectively stationary within* $[-\frac{T_0}{2}, \frac{T_0}{2}]$ if the joint probability assignment for $t_1, \ldots, t_n$ is the same as that for $t_1+\tau, t_2+\tau, \ldots, t_n+\tau$ whenever $t_1, \ldots, t_n$ and $t_1+\tau, t_2+\tau, \ldots, t_n+\tau$ are all contained in the interval $[-\frac{T_0}{2}, \frac{T_0}{2}]$. It is *effectively WSS within* $[-\frac{T_0}{2}, \frac{T_0}{2}]$ if $\mathsf{K}_{\boldsymbol{Z}}(t,\tau)$ is a function only of $t-\tau$ for $t, \tau \in [-\frac{T_0}{2}, \frac{T_0}{2}]$. A random process with nonzero mean is effectively stationary (effectively WSS) if its mean is constant within $[-\frac{T_0}{2}, \frac{T_0}{2}]$ and its fluctuation is effectively stationary (WSS) within $[-\frac{T_0}{2}, \frac{T_0}{2}]$.

One way to view a stationary (WSS) random process is in the limiting sense of a process that is effectively stationary (WSS) for all intervals $[-\frac{T_0}{2}, \frac{T_0}{2}]$. For operations such as linear functionals and filtering, the nature of this limit as T_0 becomes large is quite simple and natural, whereas for frequency domain results, the effect of finite T_0 is quite subtle.

For an effectively WSS process within $[-\frac{T_0}{2}, \frac{T_0}{2}]$, the covariance within $[-\frac{T_0}{2}, \frac{T_0}{2}]$ is a function of a single parameter, $\mathsf{K}_{\boldsymbol{Z}}(t,\tau) = \tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau)$ for $t, \tau \in [-\frac{T_0}{2}, \frac{T_0}{2}]$. Note however that $t-\tau$ can range from $-T_0$ (for $t{=}-\frac{T_0}{2}$, $\tau{=}\frac{T_0}{2}$) to T_0 (for $t{=}\frac{T_0}{2}$, $\tau{=}-\frac{T_0}{2}$).

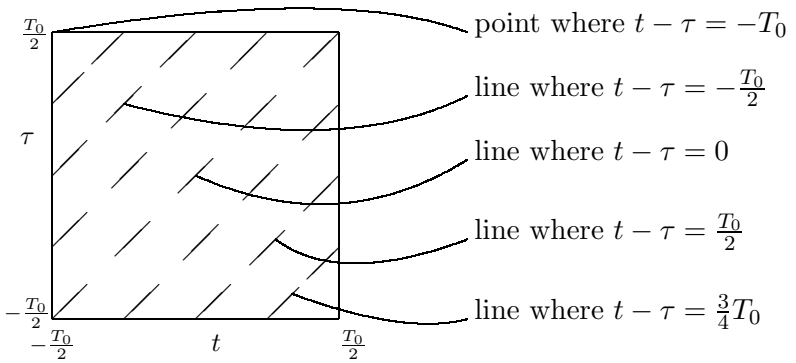


Figure 7.4: The relationship of the two argument covariance function $\mathsf{K}_{\boldsymbol{Z}}(t,\tau)$ and the one argument function $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau)$ for an effectively WSS process. $\mathsf{K}_{\boldsymbol{Z}}(t,\tau)$ is constant on each dashed line above. Note that, for example, the line for which $t-\tau = \frac{3}{4}T_0$ applies only for pairs (t,τ) where $t \geq T_0/2$ and $\tau \leq -T_0/2$. Thus $\tilde{\mathsf{K}}_Z(\frac{3}{4}T_0)$ is not necessarily equal to $\mathsf{K}_{\boldsymbol{Z}}(\frac{3}{4}T_0, 0)$. It can be easily verified, however, that $\tilde{\mathsf{K}}_Z(\alpha T_0) = \mathsf{K}_{\boldsymbol{Z}}(\alpha T_0, 0)$ for all $\alpha \leq 1/2$.

[13] There is another popular solution to this problem. For any $\mathcal{L}_2$ function $g(t)$, the energy in $g(t)$ outside of $[-\frac{T_0}{2}, \frac{T_0}{2}]$ vanishes as $T_0 \to \infty$, so intuitively the effect of these tails on the linear functional $\int g(t)Z(t)\,dt$ vanishes as $T_0 \to 0$. This provides a nice intuitive basis for ignoring the problem, but it fails, both intuitively and mathematically, in the frequency domain.

Since a Gaussian process is determined by its covariance function and mean, it is effectively stationary within $[-\frac{T_0}{2}, \frac{T_0}{2}]$ if it is effectively WSS.

Note that the difference between a stationary and effectively stationary random process for large T_0 is primarily a difference in the model and not in the situation being modeled. If two models have a significantly different behavior over the time intervals of interest, or more concretely, if noise in the distant past or future has a significant effect, then the entire modeling issue should be rethought.

7.5.3 Linear functionals for effectively WSS random processes

The covariance matrix for a set of linear functionals and the covariance function for the output of a linear filter take on simpler forms for WSS or effectively WSS processes than the corresponding forms for general processes derived in Subsection 7.4.3.

Let $Z(t)$ be a zero-mean WSS random process with covariance function $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau)$ for $t, \tau \in [-\frac{T_0}{2}, \frac{T_0}{2}]$ and let $g_1(t), g_2(t), \ldots, g_\ell(t)$ be a set of $\mathcal{L}_2$ functions nonzero only within $[-\frac{T_0}{2}, \frac{T_0}{2}]$. For the conventional WSS case, we can take $T_0 = \infty$. Let the linear functional V_m be given by $\int_{-T_0/2}^{T_0/2} Z(t) g_m(t)\, dt$ for $1 \le m \le \ell$. The covariance $\mathsf{E}[V_m V_j]$ is then given by

$$\begin{aligned}\mathsf{E}[V_m V_j] &= \mathsf{E}\left[\int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} Z(t) g_m(t)\, dt \int_{-\infty}^{\infty} Z(\tau) g_j(\tau)\, d\tau\right] \\ &= \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} g_m(t) \tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau) g_j(\tau)\, d\tau\, dt. \end{aligned} \tag{7.49}$$

Note that this depends only on the covariance where $t, \tau \in [-\frac{T_0}{2}, \frac{T_0}{2}]$, *i.e.*, where $\{Z(t)\}$ is effectively WSS. This is not surprising, since we would not expect V_m to depend on the behavior of the process outside of where $g_m(t)$ is nonzero.

7.5.4 Linear filters for effectively WSS random processes

Next consider passing a random process $\{Z(t); t \in \mathbb{R}\}$ through a linear time-invariant filter whose impulse response $h(t)$ is $\mathcal{L}_2$. As pointed out in 7.28, the output of the filter is a random process $\{V(\tau); \tau \in \mathbb{R}\}$ given by

$$V(\tau) = \int_{-\infty}^{\infty} Z(t_1) h(\tau - t_1)\, dt_1.$$

Note that $V(\tau)$ is a linear functional for each choice of τ. The covariance function evaluated at t, τ is the covariance of the linear functionals $V(t)$ and $V(\tau)$. Ignoring questions of orders of integration and convergence,

$$\mathsf{K}_{\boldsymbol{V}}(t,\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t-t_1) \mathsf{K}_{\boldsymbol{Z}}(t_1, t_2) h(\tau - t_2) dt_1 dt_2. \tag{7.50}$$

First assume that $\{Z(t);\, t \in \mathbb{R}\}$ is WSS in the conventional sense. Then $\mathsf{K}_{\boldsymbol{Z}}(t_1, t_2)$ can be replaced by $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t_1 - t_2)$. Replacing $t_1 - t_2$ by s (*i.e.*, t_1 by $t_2 + s$),

$$\mathsf{K}_{\boldsymbol{V}}(t,\tau) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(t - t_2 - s) \tilde{\mathsf{K}}_{\boldsymbol{Z}}(s)\, ds\right] h(\tau - t_2)\, dt_2.$$

Replacing t_2 by $\tau+\mu$,

$$\mathsf{K}_{\boldsymbol{V}}(t,\tau) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(t-\tau-\mu-s)\tilde{\mathsf{K}}_{\boldsymbol{Z}}(s)\, ds \right] h(-\mu)\, d\mu. \tag{7.51}$$

Thus $\mathsf{K}_{\boldsymbol{V}}(t,\tau)$ is a function only of $t-\tau$. This means that $\{V(t); t \in \mathbb{R}\}$ is WSS. This is not surprising; passing a WSS random process through a linear time-invariant filter results in another WSS random process.

If $\{Z(t); t \in \mathbb{R}\}$ is a Gaussian process, then, from Theorem 7.4.1, $\{V(t); t \in \mathbb{R}\}$ is also a Gaussian process. Since a Gaussian process is determined by its covariance function, it follows that if $Z(t)$ is a stationary Gaussian process, then $V(t)$ is also a stationary Gaussian process.

We do not have the mathematical machinery to carry out the above operations carefully over the infinite time interval[14]. Rather, it is now assumed that $\{Z(t); t \in \mathbb{R}\}$ is effectively WSS within $[-\frac{T_0}{2}, \frac{T_0}{2}]$. It will also be assumed that the impulse response $h(t)$ above is time-limited in the sense that for some finite A, $h(t) = 0$ for $|t| > A$.

Theorem 7.5.3. *Let $\{Z(t); t \in \mathbb{R}\}$ be effectively WSS within $[-\frac{T_0}{2}, \frac{T_0}{2}]$ and have sample functions that are $\mathcal{L}_2$ within $[-\frac{T_0}{2}, \frac{T_0}{2}]$ with probability 1. Let $Z(t)$ be the input to a filter with an $\mathcal{L}_2$ time-limited impulse response $\{h(t);\, [-A,\, A] \to \mathbb{R}\}$. Then for $\frac{T_0}{2} > A$, the output random process $\{V(t); t \in \mathbb{R}\}$ is WSS within $[-\frac{T_0}{2}+A, \frac{T_0}{2}-A]$ and its sample functions within $[-\frac{T_0}{2}+A, \frac{T_0}{2}-A]$ are $\mathcal{L}_2$ with probability 1.*

Proof: Let $z(t)$ be a sample function of $Z(t)$ and assume $z(t)$ is $\mathcal{L}_2$ within $[-\frac{T_0}{2}, \frac{T_0}{2}]$. Let $v(\tau) = \int z(t)h(\tau - t)\, dt$ be the corresponding filter output. For each $\tau \in [-\frac{T_0}{2}+A, \frac{T_0}{2}-A]$, $v(\tau)$ is determined by $z(t)$ in the range $t \in [-\frac{T_0}{2}, \frac{T_0}{2}]$. Thus, if we replace $z(t)$ by $z_0(t) = z(t)\text{rect}[T_0]$, the filter output, say $v_0(\tau)$ will equal $v(\tau)$ for $\tau \in [-\frac{T_0}{2}+A, \frac{T_0}{2}-A]$. The time-limited function $z_0(t)$ is $\mathcal{L}_1$ as well as $\mathcal{L}_2$. This implies that the Fourier transform $\hat{z}_0(f)$ is bounded, say by $\hat{z}_0(f) \le B$, for each f. Since $\hat{v}_0(f) = \hat{z}_0(f)\hat{h}(f)$, we see that

$$\int |\hat{v}_0(f)|^2\, df = \int |\hat{z}_0(f)|^2 |\hat{h}(f)|^2\, df \le B^2 \int |\hat{h}(f)|^2\, df < \infty$$

This means that $\hat{v}_0(f)$, and thus also $v_0(t)$, is $\mathcal{L}_2$. Now $v_0(t)$, when truncated to $[-\frac{T_0}{2}+A, \frac{T_0}{2}-A]$ is equal to $v(t)$ truncated to $[-\frac{T_0}{2}+A, \frac{T_0}{2}-A]$, so the truncated version of $v(t)$ is $\mathcal{L}_2$. Thus the sample functions of $\{V(t)\}$, truncated to $[-\frac{T_0}{2}+A, \frac{T_0}{2}-A]$, are $\mathcal{L}_2$ with probability 1.

Finally, since $\{Z(t); t \in \mathbb{R}\}$ can be truncated to $[-\frac{T_0}{2}, \frac{T_0}{2}]$ with no lack of generality, it follows that $\mathsf{K}_{\boldsymbol{Z}}(t_1, t_2)$ can be truncated to $t_1, t_2 \in [-\frac{T_0}{2}, \frac{T_0}{2}]$. Thus, for $t, \tau \in [-\frac{T_0}{2}+A, \frac{T_0}{2}-A]$, (7.50) becomes

$$\mathsf{K}_{\boldsymbol{V}}(t,\tau) = \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} h(t-t_1)\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t_1-t_2)h(\tau-t_2)dt_1 dt_2. \tag{7.52}$$

The argument in (7.50, 7.51) shows that $V(t)$ is effectively WSS within $[-\frac{T_0}{2}+A, \frac{T_0}{2}-A]$. □

The above theorem, along with the effective WSS result about linear functionals, shows us that results about WSS processes can be used within finite intervals. The result in the theorem about

[14] More important, we have no justification for modeling a process over the infinite time interval. Later, however, after building up some intuition about the relationship of an infinite interval to a very large interval, we can use the simpler equations corresponding to infinite intervals.

the interval of effective stationarity being reduced by filtering should not be too surprising. If we truncate a process, and then pass it through a filter, the filter spreads out the effect of the truncation. For a finite duration filter, however, as assumed here, this spreading is limited.

The notion of stationarity (or effective stationarity) makes sense as a modeling tool where T_0 above is very much larger than other durations of interest, and in fact where there is no need for explicit concern about how long the process is going to be stationary.

The above theorem essentially tells us that we can have our cake and eat it too. That is, transmitted waveforms and noise processes can be truncated, thus making use of both common sense and $\mathcal{L}_2$ theory, but at the same time insights about stationarity can still be relied upon. More specifically, random processes can be modeled as stationary, without specifying a specific interval $[-\frac{T_0}{2}, \frac{T_0}{2}]$ of effective stationarity, because stationary processes can now be viewed as asymptotic versions of finite duration processes.

Appendices 7A.2 and 7A.3 provide a deeper analysis of WSS processes truncated to an interval. The truncated process is represented as a Fourier series with random variables as coefficients. This gives a clean interpretation of what happens as the interval size is increased without bound, and also gives a clean interpretation of the effect of time-truncation in the frequency domain. Another approach to a truncated process is the Karhunen-Loeve expansion, which is discussed in 7A.4.

7.6 Stationary and WSS processes in the Frequency Domain

Stationary and WSS zero-mean processes, and particularly Gaussian processes, are often viewed more insightfully in the frequency domain than in the time domain. An effectively WSS process over $[-\frac{T_0}{2}, \frac{T_0}{2}]$ has a single variable covariance function $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(\tau)$ defined over $[T_0, T_0]$. A WSS process can be viewed as a process that is effectively WSS for each T_0. The energy in such a process, truncated to $[-\frac{T_0}{2}, \frac{T_0}{2}]$, is linearly increasing in T_0, but the covariance simply becomes defined over a larger and larger interval as $T_0 \to \infty$. Assume in what follows that this limiting covariance is $\mathcal{L}_2$. This does not appear to rule out any but the most pathological processes.

First we look at linear functionals and linear filters, ignoring limiting questions and convergence issues and assuming that T_0 is 'large enough'. We will refer to the random processes as stationary, while still assuming $\mathcal{L}_2$ sample functions.

For a zero-mean WSS process $\{Z(t); t \in \mathbb{R}\}$ and a real $\mathcal{L}_2$ function $g(t)$, consider the linear functional $V = \int g(t)Z(t)\,dt$. From (7.49),

$$\mathsf{E}[V^2] = \int_{-\infty}^{\infty} g(t) \left[\int_{-\infty}^{\infty} \tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau) g(\tau)\, d\tau \right] dt \tag{7.53}$$

$$= \int_{-\infty}^{\infty} g(t) \left[\tilde{\mathsf{K}}_{\boldsymbol{Z}} * \boldsymbol{g} \right](t)\, dt. \tag{7.54}$$

where $\tilde{\mathsf{K}}_{\boldsymbol{Z}} * \boldsymbol{g}$ denotes the convolution of the waveforms $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t)$ and $g(t)$. Let $S_{\boldsymbol{Z}}(f)$ be the Fourier transform of $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t)$. The function $S_{\boldsymbol{Z}}(f)$ is called the *spectral density* of the stationary process $\{Z(t); t \in \mathbb{R}\}$. Since $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t)$ is $\mathcal{L}_2$, real, and symmetric, its Fourier transform is also $\mathcal{L}_2$, real, and symmetric, and, as shown later, $S_{\boldsymbol{Z}}(f) \geq 0$. It is also shown later that $S_{\boldsymbol{Z}}(f)$ at each frequency f can be interpreted as the power per unit frequency at f.

Let $\theta(t) = [\tilde{\mathsf{K}}_{\boldsymbol{Z}} * \boldsymbol{g}](t)$ be the convolution of $\tilde{\mathsf{K}}_{\boldsymbol{Z}}$ and $\boldsymbol{g}$. Since $\boldsymbol{g}$ and $\mathsf{K}_{\boldsymbol{Z}}$ are real, $\theta(t)$ is also real

so $\theta(t) = \theta^*(t)$. Using Parseval's theorem for Fourier transforms,

$$\mathsf{E}[V^2] = \int_{-\infty}^{\infty} g(t)\theta^*(t)\,dt = \int_{-\infty}^{\infty} \hat{g}(f)\hat{\theta}^*(f)\,df.$$

Since $\theta(t)$ is the convolution of $\mathsf{K}_{\boldsymbol{Z}}$ and $\boldsymbol{g}$, we see that $\hat{\theta}(f) = S_{\boldsymbol{Z}}(f)\hat{g}(f)$. Thus,

$$\mathsf{E}[V^2] = \int_{-\infty}^{\infty} \hat{g}(f)S_{\boldsymbol{Z}}(f)\hat{g}^*(f)\,df = \int_{-\infty}^{\infty} |\hat{g}(f)|^2\, S_{\boldsymbol{Z}}(f)\,df. \tag{7.55}$$

Note that $\mathsf{E}[V^2] \geq 0$ and that this holds for all real $\mathcal{L}_2$ functions $g(t)$. The fact that $g(t)$ is real constrains the transform $\hat{g}(f)$ to satisfy $\hat{g}(f) = \hat{g}^*(-f)$, and thus $|\hat{g}(f)| = |\hat{g}(-f)|$ for all f. Subject to this constraint and the constraint that $|\hat{g}(f)|$ be $\mathcal{L}_2$, $|\hat{g}(f)|$ can be chosen as any $\mathcal{L}_2$ function. Stated another way, $\hat{g}(f)$ can be chosen arbitrarily for $f \geq 0$ subject to being $\mathcal{L}_2$.

Since $S_{\boldsymbol{Z}}(f) = S_{\boldsymbol{Z}}(-f)$, (7.55) can be rewritten as

$$\mathsf{E}[V^2] = \int_{0}^{\infty} 2\,|\hat{g}(f)|^2\, S_{\boldsymbol{Z}}(f)\,df.$$

Since $\mathsf{E}[V^2] \geq 0$ and $|\hat{g}(f)|$ is arbitrary, it follows that $S_{\boldsymbol{Z}}(f) \geq 0$ for all $f \in \mathbb{R}$.

The conclusion here is that the spectral density of any WSS random process must be nonnegative. Since $S_{\boldsymbol{Z}}(f)$ is also the Fourier transform of $\tilde{\mathsf{K}}(t)$, this means that a necessary property of any single variable covariance function is that it have a nonnegative Fourier transform.

Next, let $V_m = \int g_m(t)Z(t)\,dt$ where the function $g_m(t)$ is real and $\mathcal{L}_2$ for $m = 1, 2$. From (7.49),

$$\mathsf{E}[V_1V_2] = \int_{-\infty}^{\infty} g_1(t)\left[\int_{-\infty}^{\infty} \tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau)g_2(\tau)\,d\tau\right] dt \tag{7.56}$$

$$= \int_{-\infty}^{\infty} g_1(t)\left[\tilde{\mathsf{K}} * \boldsymbol{g}_2\right](t)\,dt. \tag{7.57}$$

Let $\hat{g}_m(f)$ be the Fourier transform of $g_m(t)$ for $m = 1, 2$, and let $\theta(t) = [\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t) * \boldsymbol{g}_2](t)$ be the convolution of $\tilde{\mathsf{K}}_{\boldsymbol{Z}}$ and $\boldsymbol{g}_2$. Let $\hat{\theta}(f) = S_{\boldsymbol{Z}}(f)\hat{g}_2(f)$ be its Fourier transform. As before, we have

$$\mathsf{E}[V_1V_2] = \int \hat{g}_1(f)\hat{\theta}^*(f)\,df = \int \hat{g}_1(f)S_{\boldsymbol{Z}}(f)\hat{g}_2^*(f)\,df. \tag{7.58}$$

There is a remarkable feature in the above expression. If $\hat{g}_1(f)$ and $\hat{g}_2(f)$ have no overlap in frequency, then $\mathsf{E}[V_1V_2] = 0$. In other words, for any stationary process, two linear functionals over different frequency ranges must be uncorrelated. If the process is Gaussian, then the linear functionals are independent. This means in essence that Gaussian noise in different frequency bands must be independent. That this is true simply because of stationarity is surprising. Appendix 7A.3 helps to explain this puzzling phenomenon, especially with respect to effective stationarity.

Next, let $\{\phi_m(t);\, m \in \mathbb{Z}\}$ be a set of real orthonormal functions and let $\{\hat{\phi}_m(f)\}$ be the corresponding set of Fourier transforms. Letting $V_m = \int Z(t)\phi_m(t)\,dt$, (7.58) becomes

$$\mathsf{E}[V_mV_j] = \int \hat{\phi}_m(f)S_{\boldsymbol{Z}}(f)\hat{\phi}_j^*(f)\,df. \tag{7.59}$$

If the set of orthonormal functions $\{\phi_m(t); m \in \mathbb{Z}\}$ is limited to some frequency band, and if $S_{\boldsymbol{Z}}(f)$ is constant, say with value $N_0/2$ in that band, then

$$\mathsf{E}[V_m V_j] = N_0/2 \int \hat{\phi}_m(f)\hat{\phi}_j^*(f)\, df. \tag{7.60}$$

By Parseval's theorem for Fourier transforms, we have $\int \hat{\phi}_m(f)\hat{\phi}_j^*(f)\, df = \delta_{mj}$, and thus

$$\mathsf{E}[V_m V_j] = \frac{N_0}{2}\delta_{mj}. \tag{7.61}$$

The rather peculiar looking constant $N_0/2$ is explained in the next section. For now, however, it is possible to interpret the meaning of the spectral density of a noise process. Suppose that $S_{\boldsymbol{Z}}(f)$ is continuous and approximately constant with value $S_{\boldsymbol{Z}}(f_c)$ over some narrow band of frequencies around f_c and suppose that $\phi_1(t)$ is constrained to that narrow band. Then the variance of the linear functional $\int_{-\infty}^{\infty} Z(t)\phi_1(t)\, dt$ is approximately $S_{\boldsymbol{Z}}(f_c)$. In other words, $S_{\boldsymbol{Z}}(f_c)$ in some fundamental sense describes the energy in the noise per degree of freedom at the frequency f_c. The next section interprets this further.

7.7 White Gaussian noise

Physical noise processes are very often reasonably modeled as zero mean, stationary, and Gaussian. There is one further simplification that is often reasonable. This is that the covariance between the noise at two epochs dies out very rapidly as the interval between those epochs increases. The interval over which this covariance is significantly nonzero is often very small relative to the intervals over which the signal varies appreciably. This means that the covariance function $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(\tau)$ looks like a short-duration pulse around $\tau = 0$.

We know from linear system theory that $\int \tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau)g(\tau)d\tau$ is equal to $g(t)$ if $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t)$ is a unit impulse. Also, this integral is approximately equal to $g(t)$ if $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t)$ has unit area and is a narrow pulse relative to changes in $g(t)$. It follows that under the same circumstances, (7.56) becomes

$$\mathsf{E}[V_1 V_2^*] = \int_t \int_\tau g_1(t)\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau)g_2(\tau)\, d\tau\, dt \approx \int g_1(t)g_2(t)\, dt. \tag{7.62}$$

This means that if the covariance function is very narrow relative to the functions of interest, then its behavior relative to those functions is specified by its area. In other words, the covariance function can be viewed as an impulse of a given magnitude. We refer to a zero-mean WSS Gaussian random process with such a narrow covariance function as *White Gaussian Noise (WGN)*. The area under the covariance function is called the *intensity* or the *spectral density* of the WGN and is denoted by the symbol $N_0/2$. Thus, for $\mathcal{L}_2$ functions $g_1(t), g_2(t), \ldots$ in the range of interest, and for WGN (denoted by $\{W(t); t \in \mathbb{R}\}$) of intensity $N_0/2$, the random variable $V_m = \int W(t)g_m(t)\, dt$ has the variance

$$\mathsf{E}[V_m^2] = (N_0/2) \int g_m^2(t)\, dt. \tag{7.63}$$

Similarly, the random variables V_j and V_m have the covariance

$$\mathsf{E}[V_j V_m] = (N_0/2) \int g_j(t)g_m(t)\, dt. \tag{7.64}$$

Also $V_1, V_2, \ldots$ are jointly Gaussian.

The most important special case of (7.63) and (7.64) is to let $\phi_j(t)$ be a set of orthonormal functions and let $W(t)$ be WGN of intensity $N_0/2$. Let $V_m = \int \phi_m(t)W(t)\,dt$. Then, from (7.63) and (7.64),

$$\mathsf{E}[V_j V_m] = (N_0/2)\delta_{jm}. \tag{7.65}$$

This is an important equation. It says that if the noise can be modeled as WGN, then when the noise is represented in terms of *any* orthonormal expansion, the resulting random variables are iid. Thus, we can represent signals in terms of an arbitrary orthonormal expansion, and represent WGN in terms of the same expansion, and the result is iid Gaussian random variables.

Since the coefficients of a WGN process in any orthonormal expansion are iid Gaussian, it is common to also refer to a random vector of iid Gaussian rv's as WGN.

If $\mathsf{K}_{\boldsymbol{W}}(t)$ is approximated by $(N_0/2)\delta(t)$, then the spectral density is approximated by $S_{\boldsymbol{W}}(f) = N_0/2$. If we are concerned with a particular band of frequencies, then we are interested in $S_{\boldsymbol{W}}(f)$ being constant within that band, and in this case, $\{W(t); t \in \mathbb{R}\}$ can be represented as white noise within that band. If this is the only band of interest, we can model[15] $S_{\boldsymbol{W}}(f)$ as equal to $N_0/2$ everywhere, in which case the corresponding model for the covariance function is $(N_0/2)\delta(t)$.

The careful reader will observe that WGN has not really been defined. What has been said, in essence, is that if a stationary zero-mean Gaussian process has a covariance function that is very narrow relative to the variation of all functions of interest, or a spectral density that is constant within the frequency band of interest, then we can pretend that the covariance function is an impulse times $N_0/2$, where $N_0/2$ is the value of $S_{\boldsymbol{W}}(f)$ within the band of interest. Unfortunately, according to the definition of random process, there cannot be any Gaussian random process $W(t)$ whose covariance function is $\tilde{\mathsf{K}}(t) = (N_0/2)\delta(t)$. The reason for this dilemma is that $\mathsf{E}[W^2(t)] = \mathsf{K}_{\boldsymbol{W}}(0)$. We could interpret $\mathsf{K}_{\boldsymbol{W}}(0)$ to be either undefined or ∞, but either way, $W(t)$ cannot be a random variable (although we could think of it taking on only the values plus or minus ∞).

Mathematicians view WGN as a generalized random process, in the same sense as the unit impulse $\delta(t)$ is viewed as a generalized function. That is, the impulse function $\delta(t)$ is not viewed as an ordinary function taking the value 0 for $t \neq 0$ and the value ∞ at $t = 0$. Rather, it is viewed in terms of its effect on other, better behaved, functions $g(t)$, where $\int_{-\infty}^{\infty} g(t)\delta(t)\,dt = g(0)$. In the same way, WGN is not viewed in terms of random variables at each epoch of time. Rather it is viewed as a generalized zero-mean random process for which linear functionals are jointly Gaussian, for which variances and covariances are given by (7.63) and (7.64), and for which the covariance is formally taken to be $(N_0/2)\delta(t)$.

Engineers should view WGN within the context of an overall bandwidth and time interval of interest, where the process is effectively stationary within the time interval and has a constant spectral density over the band of interest. Within that context, the spectral density can be viewed as constant, the covariance can be viewed as an impulse, and (7.63) and (7.64) can be used.

The difference between the engineering view and the mathematical view is that the engineering view is based on a context of given time interval and bandwidth of interest, whereas the math-

[15] This is not at obvious as it sounds, and will be further discussed in terms of the theorem of irrelevance in the next chapter.

ematical view is based on a very careful set of definitions and limiting operations within which theorems can be stated without explicitly defining the context. Although the ability to prove theorems without stating the context is valuable, any application must be based on the context.

7.7.1 The sinc expansion as an approximation to WGN

Theorem 7.5.2 treated the process $Z(t) = \sum_k Z_k \text{sinc}\left(\frac{t-kT}{T}\right)$ where each rv $\{Z_k; k \in \mathbb{Z}\}$ is iid and $\mathcal{N}(0, \sigma^2)$. We found that the process is zero-mean Gaussian and stationary with covariance function $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau) = \sigma^2 \text{sinc}(\frac{t-\tau}{T})$. The spectral density for this process is then given by

$$S_{\boldsymbol{Z}}(f) = \sigma^2 T \,\text{rect}(fT). \tag{7.66}$$

This process has a constant spectral density over the baseband bandwidth $\mathsf{W} = 1/(2T)$, so by making T sufficiently small, the spectral density is constant over a band sufficiently large to include all frequencies of interest. Thus this process can be viewed as WGN of spectral density $\frac{N_0}{2} = \sigma^2 T$ for any desired range of frequencies $\mathsf{W} = 1/(2T)$ by making T sufficiently small. Note, however, that to approximate WGN of spectral density $N_0/2$, the noise power, *i.e.*, the variance of $Z(t)$ is $\sigma^2 = \mathsf{W}N_0$. In other words, σ^2 must increase with increasing W. This also says that N_0 is the noise power per unit *positive frequency*. The spectral density, $N_0/2$, is defined over both positive and negative frequencies, and so becomes N_0 when positive and negative frequencies are combined as in the standard definition of bandwidth[16].

If a sinc process is passed through a linear filter with an arbitrary impulse response $h(t)$, the output is a stationary Gaussian process with spectral density $|\hat{h}(f)|^2 \sigma^2 T \,\text{rect}(fT)$. Thus, by using a sinc process plus a linear filter, a stationary Gaussian process with any desired non-negative spectral density within any desired finite bandwith can be generated. In other words, stationary Gaussian processes with arbitrary covariances (subject to $S(f) \geq 0$ can be generated from orthonormal expansions of Gaussian variables.

Since the sinc process is stationary, it has sample waveforms of infinite energy. As explained in subsection 7.5.2, this process may be truncated to achieve an effectively stationary process with $\mathcal{L}_2$ sample waveforms. Appendix 7A.3 provides some insight about how an effectively stationary Gaussian process over an interval T_0 approaches stationarity as $T_0 \to \infty$.

The sinc process can also be used to understand the strange, everywhere uncorrelated, process in Example 7.4.2. Holding $\sigma^2 = 1$ in the sinc expansion as T approaches 0, we get a process whose limiting covariance function is 1 for $t-\tau = 0$ and 0 elsewhere. The corresponding limiting spectral density is 0 everywhere. What is happening is that the power in the process (*i.e.*, $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(0)$) is 1, but that power is being spread over a wider and wider band as $T \to 0$, so the power per unit frequency goes to 0.

To explain this in another way, note that any measurement of this noise process must involve filtering over some very small but nonzero interval. The output of this filter will have zero variance. Mathematically, of course, the limiting covariance is $\mathcal{L}_2$-equivalent to 0, so again the mathematics[17] corresponds to engineering reality.

[16] One would think that this field would have found a way to be consistent about counting only positive frequencies or positive and negative frequencies. However, the word bandwidth is so widely used among the mathophobic, and Fourier analysis is so necessary for engineers, that one must simply live with such minor confusions.

[17] This process also can not be satisfactorily defined in a measure theoretic way.

7.7.2 Poisson process noise

The sinc process of the last subsection is very convenient for generating noise processes that approximate WGN in an easily used formulation. On the other hand, this process is not very believable[18] as a physical process. A model that corresponds better to physical phenomena, particularly for optical channels, is a sequence of very narrow pulses which arrive according to a Poisson distribution in time.

The Poisson distribution, for our purposes, can be simply viewed as a limit of a discrete time process where the time axis is segmented into intervals of duration Δ and a pulse of width Δ arrives in each interval with probability $\Delta\rho$, independent of every other interval. When such a process is passed through a linear filter, the fluctuation of the output at each instant of time is approximately Gaussian if the filter is of sufficiently small bandwidth to integrate over a very large number of pulses. One can similarly argue that linear combinations of filter outputs tend to be approximately Gaussian, making the process an approximation of a Gaussian process.

We do not analyze this carefully, since our point of view is that WGN, over limited bandwidths, is a reasonable and canonic approximation to a large number of physical noise processes. After understanding how this affects various communication systems, one can go back and see whether the model is appropriate for the given physical noise process. When we study wireless communication, we will find that the major problem is not that the noise is poorly approximated by WGN, but rather that the channel itself is randomly varying.

7.8 Adding noise to modulated communication

Consider the QAM communication problem again. A complex $\mathcal{L}_2$ baseband waveform $u(t)$ is generated and modulated up to passband as a real waveform $x(t) = 2\Re[u(t)e^{2\pi i f_c t}]$. A sample function $w(t)$ of a random noise process $W(t)$ is then added to $x(t)$ to produce the output $y(t) = x(t)+w(t)$, which is then demodulated back to baseband as the received complex baseband waveform $v(t)$.

Generalizing QAM somewhat, assume that $u(t)$ is given by $u(t) = \sum_k u_k\theta_k(t)$ where the functions $\theta_k(t)$ are complex orthonormal functions and the sequence of symbols $\{u_k; k \in \mathbb{Z}\}$ are complex numbers drawn from the symbol alphabet and carrying the information to be transmitted. For each symbol u_k, $\Re(u_k)$ and $\Im(u_k)$ should be viewed as sample values of the random variables $\Re(U_k)$ and $\Im(U_k)$. The joint probability distributions of these random variables is determined by the incoming random binary digits and how they are mapped into symbols. The *complex random variable*[19] $\Re(U_k) + i\Im(U_k)$ is then denoted by U_k.

In the same way, $\Re(\sum_k U_k\theta_k(t))$ and $\Im(\sum_k U_k\theta_k(t))$ are random processes denoted respec-

[18] To many people, defining these sinc processes with their easily analyzed properties but no physical justification, is more troublesome than our earlier use of discrete memoryless sources in studying source coding. Actually, the approach to modeling is the same in each case: first understand a class of easy-to-analyze but perhaps impractical processes, then build on that understanding to understand practical cases. Actually, sinc processes have an advantage here: the band limited statationary Gaussian random processes defined this way (although not the method of generation) are widely used as practical noise models, whereas there are virtually no uses of discrete memoryless sources as practical source models.

[19] Recall that a random variable (rv) is a mapping from sample points to real numbers, so that a complex rv is a mapping from sample points to complex numbers. Sometimes in discussions involving both rv's and complex rv's, it helps to refer to rv's as real rv's, but the modifier 'real' is superflous.

tively by $\Re(U(t))$ and $\Im(U(t))$. We then call $U(t) = \Re(U(t)) + i\Im(U(t))$ for $t \in \mathbb{R}$ a *complex random process*. A complex random process $U(t)$ is defined by the joint distribution of $U(t_1), U(t_2), \ldots, U(t_n)$ for all choices of $n, t_1, \ldots, t_n$. This is equivalent to defining both $\Re(U(t))$ and $\Im(U(t))$ as joint processes.

Recall from the discussion of the Nyquist criterion that if the QAM transmit pulse $p(t)$ is chosen to be square-root of Nyquist, then $p(t)$ and its T-spaced shifts are orthogonal and can be normalized to be orthonormal. Thus a particularly natural choice here is $\theta_k(t) = p(t - kT)$ for such a $\boldsymbol{p}$. Note that this is a generalization of the previous chapter in the sense that $\{U_k;\, k \in \mathbb{Z}\}$ is a sequence of complex rv's using random choices from the signal constellation rather than some given sample function of that random sequence. The transmitted passband (random) waveform is then

$$X(t) = \sum_k 2\Re\left\{U_k \theta_k(t) \exp[2\pi i f_c t]\right\}. \tag{7.67}$$

Recall that the transmitted waveform has twice the power of the baseband waveform. Now define

$$\begin{aligned} \psi_{k,1}(t) &= \Re\left\{2\theta_k(t) \exp[2\pi i f_c t]\right\}; \\ \psi_{k,2}(t) &= \Im\left\{-2\theta_k(t) \exp[2\pi i f_c t]\right\}. \end{aligned}$$

Also, let $U_{k,1} = \Re(U_k)$ and $U_{k,2} = \Im(U_k)$. Then

$$X(t) = \sum_k [U_{k,1}\psi_{k,1}(t) + U_{k,2}\psi_{k,2}(t)].$$

As shown in Theorem 6.6.1, the set of bandpass functions $\{\psi_{k,\ell}; k \in \mathbb{Z}, \ell \in \{1,2\}\}$ are orthogonal and each have energy equal to 2. This again assumes that the carrier frequency f_c is greater than all frequencies in each baseband function $\theta_k(t)$.

In order for $u(t)$ to be $\mathcal{L}_2$, assume that the number of orthogonal waveforms $\theta_k(t)$ is arbitrarily large but finite, say $\theta_1(t), \ldots, \theta_n(t)$. Thus $\{\psi_{k,\ell}\}$ is also limited to $1 \le k \le n$.

Assume that the noise $\{W(t); t \in \mathbb{R}\}$ is white over the band of interest and effectively stationary over the time interval of interest, but has $\mathcal{L}_2$ sample functions[20]. Since $\{\psi_{k,l}; 1 \le k \le n, \ell = 1,2\}$ is a finite real orthogonal set, the projection theorem can be used to express each sample noise waveform $\{w(t); t \in \mathbb{R}\}$ as

$$w(t) = \sum_{k=1}^{n} [z_{k,1}\psi_{k,1}(t) + z_{k,2}\psi_{k,2}(t)] + w_\perp(t), \tag{7.68}$$

where $w_\perp(t)$ is the component of the sample noise waveform perpendicular to the space spanned by $\{\psi_{k,l}; 1 \le k \le n, \ell = 1,2\}$. Let $Z_{k,\ell}$ be the rv with sample value $z_{k,\ell}$. Then each rv $Z_{k,\ell}$ is a linear functional on $W(t)$. Since $\{\psi_{k,l}; 1 \le k \le n, \ell = 1,2\}$ is an orthogonal set, the rv's $Z_{k,\ell}$ are iid Gaussian rv's. Let $W_\perp(t)$ be the random process corresponding to the sample function $w_\perp(t)$ above. Expanding $\{W_\perp(t); t \in \mathbb{R}\}$ in an orthonormal expansion orthogonal to $\{\psi_{k,l}; 1 \le k \le n, \ell = 1,2\}$, the coefficients are assumed to be independent of the $Z_{k,\ell}$, at least

[20]Since the set of orthogonal waveforms $\theta_k(t)$ are not necessarily time or frequency limited, the assumption here is that the noise is white over a much larger time and frequency interval than the nominal bandwidth and time interval used for communication. This assumption is discussed further in the next chapter.

over the time and frequency band of interest. What happens to these coefficients outside of the region of interest is of no concern, other than assuming that $W_\perp(t)$ is independent of $U_{k,\ell}$ and $Z_{k,\ell}$ for $1 \le k \le n$ and $\ell = \{1, 2\}$. The received waveform $Y(t) = X(t) + W(t)$ is then

$$Y(t) = \sum_{k=1}^{n} [(U_{k,1}+Z_{k,1})\,\psi_{k,1}(t) + (U_{k,2}+Z_{k,2})\,\psi_{k,2}(t)] + W_\perp(t).$$

When this is demodulated,[21] the baseband waveform is represented as the complex waveform

$$V(t) = \sum_k (U_k + Z_k)\theta_k(t) + Z_\perp(t). \tag{7.69}$$

where each Z_k is a complex rv given by $Z_k = Z_{k,1} + iZ_{k,2}$ and the baseband residual noise $Z_\perp(t)$ is independent of $\{U_k, Z_k;\, 1 \le k \le n\}$. The variance of each real rv $Z_{k,1}$ and $Z_{k,2}$ is taken by convention to be $N_0/2$. We follow this convention because we are measuring the input power at baseband; as mentioned many times, the power at passband is scaled to be twice that at baseband. The point here is that N_0 is not a physical constant - rather it is the noise power per unit positive frequency *in the units used to represent the signal power.*

7.8.1 Complex Gaussian random variables and vectors

Noise waveforms, after demodulation to baseband, are usually complex and are thus represented, as in (7.69), by a sequence of complex random variables, best regarded as a complex random vector (r**v**). It is possible to view any such n dimensional complex r**v** $\boldsymbol{Z} = \boldsymbol{Z}_{\rm re} + i\boldsymbol{Z}_{\rm im}$ as a $2n$ dimensional real r**v** $\begin{bmatrix} \boldsymbol{Z}_{\rm re} \\ \boldsymbol{Z}_{\rm im} \end{bmatrix}$ where $\boldsymbol{Z}_{\rm re} = \Re(\boldsymbol{Z})$ and $\boldsymbol{Z}_{\rm im} = \Im(\boldsymbol{Z})$.

For many of the same reasons that it is desirable to work directly with a complex baseband waveform rather than a pair of real passband waveforms, it is often beneficial to work directly with complex r**v**'s.

Definition 7.8.1. A complex random variable $Z = Z_{\rm re} + iZ_{\rm im}$ is *Gaussian* if $Z_{\rm re}$ and $Z_{\rm im}$ are jointly Gaussian; Z is *circularly-symmetric Gaussian*[22] if it is Gaussian and $Z_{\rm re}$ and $Z_{\rm im}$ are zero mean and iid.

The amplitude of a circularly-symmetric Gaussian rv is Rayleigh distributed and the phase is uniform, *i.e.*, it has circular symmetry. A circularly-symmetric Gaussian rv Z is fully described by its variance $\sigma^2 = \mathsf{E}[ZZ^*]$ and is denoted as $Z \sim \mathcal{CN}(0, \sigma^2)$. Note that the real and imaginary parts of Z are then iid with variance $\sigma^2/2$ each.

Definition 7.8.2. A complex random vector (r**v**) $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\mathsf{T}}$ is *jointly Gaussian* if the $2n$ real and imaginary components of $\boldsymbol{Z}$ are jointly Gaussian. It is *circularly symmetric* if the distribution of $\boldsymbol{Z}$ (*i.e.*, the joint distribution of the real and imaginary parts) is the same as that of $e^{i\theta}\boldsymbol{Z}$ for all phase angles θ. It is *circularly -symmetric Gaussian* if it is jointly-Gaussian and circularly symmetric.

[21] Some filtering is necessary before demodulation to remove the residual noise that is far out of band, but we do not want to analyze that here.

[22] This is sometimes referred to as complex proper Gaussian.

Example 7.8.1. An important example of a circularly-symmetric Gaussian r**v** is $\boldsymbol{W} = (W_1, \ldots, W_n)^{\mathsf{T}}$ where the components $W_k, 1 \leq k \leq n$ are statistically independent and each is $\mathcal{CN}(0,1)$. Since each W_k, is $\mathcal{CN}(0,1)$, it can be seen that $e^{i\theta}W_k$ has the same distribution as W_k. Using the independence, it can be seen that $e^{i\theta}\boldsymbol{W}$ then has the same distribution as $\boldsymbol{W}$. The $2n$ real and imaginary components of $\boldsymbol{W}$ are iid and $\mathcal{N}(0,1/2)$ so that the probability density is

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{1}{(\pi)^n} \exp\left[\sum_{k=1}^{n} -|w_k|^2\right], \tag{7.70}$$

where we have used the fact that $|w_k|^2 = \Re(w_k)^2 + \Im(w_k)^2$ for each k to replace a sum over $2n$ terms with a sum over n terms.

Definition 7.8.3. The *covariance matrix* $\mathsf{K}_{\boldsymbol{Z}}$ and the *pseudo-covariance matrix* $\mathsf{M}_{\boldsymbol{Z}}$ of a zero-mean complex r**v** $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\mathsf{T}}$ are the n by n matrices given respectively by

$$\mathsf{K}_{\boldsymbol{Z}} = \mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^{\dagger}] \qquad \mathsf{M}_{\boldsymbol{Z}} = \mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^{\mathsf{T}}], \tag{7.71}$$

where $\boldsymbol{Z}^{\dagger}$ is the the conjugate of the transpose, $\boldsymbol{Z}^{\mathsf{T}*}$.

For *real* zero-mean random vectors, the covariance matrix specifies all the second moments, and thus in the jointly-Gaussian case, specifies the distribution. For *complex* r**v**'s, both $\mathsf{K}_{\boldsymbol{Z}}$ and $\mathsf{M}_{\boldsymbol{Z}}$ combine to specify all the second moments. Specifically, a little calculation shows that

$$\mathsf{E}[\Re(Z_k)\Re(Z_j)] = \tfrac{1}{2}\Re[\mathsf{K}_{\boldsymbol{Z}}(k,j) + \mathsf{M}_{\boldsymbol{Z}}(k,j)] \qquad \mathsf{E}[\Im(Z_k)\Im(Z_j)] = \tfrac{1}{2}\Re[\mathsf{K}_{\boldsymbol{Z}}(k,j) - \mathsf{M}_{\boldsymbol{Z}}(k,j)]$$

$$\mathsf{E}[\Re(Z_k)\Im(Z_j)] = \tfrac{1}{2}\Im[\text{-}K_{\boldsymbol{Z}}(k,j) + \mathsf{M}_{\boldsymbol{Z}}(k,j)] \qquad \mathsf{E}[\Im(Z_k)\Re(Z_j)] = \tfrac{1}{2}\Im[\mathsf{K}_{\boldsymbol{Z}}(k,j) + \mathsf{M}_{\boldsymbol{Z}}(k,j)]$$

When $\boldsymbol{Z}$ is a zero-mean, complex jointly-Gaussian r**v** then $\mathsf{K}_{\boldsymbol{Z}}$ and $\mathsf{M}_{\boldsymbol{Z}}$ specify the distribution of $\boldsymbol{Z}$, and thus $\boldsymbol{Z}$ is circularly-symmetric Gaussian if and only if $\mathsf{K}_{\boldsymbol{Z}} = \mathsf{K}_{e^{i\theta}\boldsymbol{Z}}$ and $\mathsf{M}_{\boldsymbol{Z}} = \mathsf{M}_{e^{i\theta}\boldsymbol{Z}}$ for all phases θ. Calculating these matrices for an arbitrary r**v**,

$$\mathsf{K}_{e^{i\theta}\boldsymbol{Z}} = \mathsf{E}[e^{i\theta}\boldsymbol{Z} \cdot e^{-i\theta}\boldsymbol{Z}^{\dagger}] = \mathsf{K}_{\boldsymbol{Z}}; \qquad \mathsf{M}_{e^{i\theta}\boldsymbol{Z}} = \mathsf{E}[e^{i\theta}\boldsymbol{Z} \cdot e^{i\theta}\boldsymbol{Z}^{\mathsf{T}}] = e^{2i\theta}\mathsf{M}_{\boldsymbol{Z}}$$

Thus, $\mathsf{K}_{e^{i\theta}\boldsymbol{Z}}$ is always equal to $\mathsf{K}_{\boldsymbol{Z}}$ but $\mathsf{M}_{e^{i\theta}\boldsymbol{Z}}$ is equal to $\mathsf{M}_{\boldsymbol{Z}}$ for all real θ if and only if $\mathsf{M}_{\boldsymbol{Z}}$ is the zero matrix. We have proven the following theorem.

Theorem 7.8.1. *A zero-mean, complex jointly-Gaussian r***v** *is circularly-symmetric Gaussian if and only if the pseudo-covariance matrix* $\mathsf{M}_{\boldsymbol{Z}}$ *is 0.*

Since $\mathsf{M}_{\boldsymbol{Z}}$ is zero for any circularly-symmetric Gaussian r**v** $\boldsymbol{Z}$, the distribution of $\boldsymbol{Z}$ is determined solely by $\mathsf{K}_{\boldsymbol{Z}}$ and is denoted as $\boldsymbol{Z} \sim \mathcal{CN}(0, \mathsf{K}_{\boldsymbol{Z}})$ where $\mathcal{C}$ denotes that $\boldsymbol{Z}$ is both complex and circularly symmetric. The complex normalized iid r**v** of Example 7.8.1 is thus denoted as $\boldsymbol{W} \sim \mathcal{CN}(0, \mathsf{I}_n)$.

The following two examples illustrate some subtleties in Theorem 7.8.1.

Example 7.8.2. Let $\boldsymbol{Z} = (Z_1, Z_2)^{\mathsf{T}}$ where $Z_1 \sim \mathcal{CN}(0,1)$ and $Z_2 = UZ_1$ where U is statistically independent of Z_1 and has possible values ± 1 with probability 1/2 each. It is easy to see that $Z_2 \sim \mathcal{CN}(0,1)$, but the real and imaginary parts of Z_1 and Z_2 together are not jointly Gaussian. In fact, the joint distribution of $\Re(Z_1)$ and $\Re(Z_2)$ is concentrated on the two diagonal axes and

$\Im(Z_1)$ and $\Im(Z_2)$ is similarly distributed. Thus, $\boldsymbol{Z}$ is not jointly Gaussian, and the theorem doesn't apply. Even though Z_1 and Z_2 are individually circularly-symmetric Gaussian, $\boldsymbol{Z}$ is not circularly-symmetric Gaussian. In this example, it turns out that $\boldsymbol{Z}$ *is* circularly symmetric and $\mathsf{M}_{\boldsymbol{Z}} = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$. The example could be changed slightly, changing the definition of Z_2 to $\Re(Z_2) = U\Re(Z_1)$ and $\Im(Z_2) \sim \mathcal{N}(0, 1/2)$, where $\Im(Z_2)$ is statistically independent of all the other variables. Then $\mathsf{M}_{\boldsymbol{Z}}$ is still 0, but $\boldsymbol{Z}$ is not circularly symmetric. Thus, without the jointly-Gaussian property, the pseudo-covariance matrix does not specify whether $\boldsymbol{Z}$ is circularly symmetric.

Example 7.8.3. Consider a vector $\boldsymbol{Z} = (Z_1, Z_2)^{\mathsf{T}}$ where $Z_1 \sim \mathcal{CN}(0, 1)$ and $Z_2 = Z_1^*$. Since $\Re(Z_2) = \Re(Z_1)$ and $\Im(Z_2) = -\Im(Z_1)$, we see that the four real and imaginary components of $\boldsymbol{Z}$ are jointly Gaussian, so $\boldsymbol{Z}$ is complex jointly Gaussian and the theorem applies. We see that $\mathsf{M}_{\boldsymbol{Z}} = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, and thus $\boldsymbol{Z}$ is jointly Gaussian but not circularly symmetric. This makes sense, since when Z_1 is real (or approximately real), $Z_2 = Z_1$ (or $Z_2 \approx Z_1$) and when Z_1 is pure imaginary (or close to pure imaginary), Z_2 is the negative of Z_1 (or $Z_2 \approx -Z_1$). Thus the relationship of Z_2 to Z_1 is certainly not phase invariant.

What makes this example interesting is that both $Z_1 \sim \mathcal{CN}(0, 1)$ and $Z_2 \sim \mathcal{CN}(0, 1)$. Thus, as in Example 7.8.2, it is the relationship between Z_1 and Z_2 that breaks up the circularly-symmetric Gaussian property. Here it is the circular symmetry that causes the problem, whereas in Example 7.8.2 it was the lack of a jointly-Gaussian distribution.

In Section 7.3, we found that an excellent approach to real jointly-Gaussian r$\mathbf{v}$'s was to view them as linear transformations of a r$\mathbf{v}$ with iid components, each $\mathcal{N}(0, 1)$. We will find here that the same approach applies to circularly-symmetric Gaussian vectors. Thus let A be an arbitrary complex m by n matrix and let the complex r$\mathbf{v}$ $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\mathsf{T}}$ be defined by

$$\boldsymbol{Z} = \mathsf{A}\,\boldsymbol{W}, \tag{7.72}$$

where $\boldsymbol{W} \sim \mathcal{CN}(0, \mathsf{I}_m)$. The complex r$\mathbf{v}$ defined in this way has jointly Gaussian real and imaginary parts. To see this, represent (7.72) as the following real linear transformation of $2n$ real space:

$$\begin{bmatrix} \boldsymbol{Z}_{\rm re} \\ \boldsymbol{Z}_{\rm im} \end{bmatrix} = \begin{bmatrix} \mathsf{A}_{\rm re} & -\mathsf{A}_{\rm im} \\ \mathsf{A}_{\rm im} & \mathsf{A}_{\rm re} \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_{\rm re} \\ \boldsymbol{W}_{\rm im} \end{bmatrix}, \tag{7.73}$$

where $\boldsymbol{Z}_{\rm re} = \Re(\boldsymbol{Z})$, $\boldsymbol{Z}_{\rm im} = \Im(\boldsymbol{Z})$, $\mathsf{A}_{\rm re} = \Re(\mathsf{A})$, and $\mathsf{A}_{\rm im} = \Im(\mathsf{A})$.

The r$\mathbf{v}$ $\boldsymbol{Z}$ is also circularly symmetric.[23] To see this, note that

$$\mathsf{K}_{\boldsymbol{Z}} = \mathsf{E}[\mathsf{A}\,\boldsymbol{W}\,\boldsymbol{W}^\dagger \mathsf{A}^\dagger] = \mathsf{A}\mathsf{A}^\dagger \qquad \mathsf{M}_{\boldsymbol{Z}} = \mathsf{E}[\mathsf{A}\,\boldsymbol{W}\,\boldsymbol{W}^{\mathsf{T}} \mathsf{A}^{\mathsf{T}}] = 0 \tag{7.74}$$

Thus from Theorem 7.8.1, $\boldsymbol{Z}$ is circularly-symmetric Gaussian and $\boldsymbol{Z} \sim \mathcal{CN}(0, \mathsf{A}\mathsf{A}^\dagger)$.

This proves the *if* part of the following theorem.

Theorem 7.8.2. *A complex r$\mathbf{v}$ $\boldsymbol{Z}$ is circularly-symmetric Gaussian if and only if it can be expressed as $\boldsymbol{Z} = \mathsf{A}\boldsymbol{W}$ for a complex matrix A and an iid circularly-symmetric Gaussian r$\mathbf{v}$ $\boldsymbol{W} \sim \mathcal{CN}(0, \mathsf{I}_m)$.*

[23] Conversely, as we will see later, all circularly symmetric jointly-Gaussian r$\mathbf{v}$'s can be defined this way.

Proof: Let $\boldsymbol{Z} \sim \mathsf{K}_{\boldsymbol{Z}}$ be an arbitrary circularly-symmetric Gaussian r**v**. From Appendix 7A.1, $\mathsf{K}_{\boldsymbol{Z}}$ can be expressed as

$$\mathsf{K}_{\boldsymbol{Z}} = \mathsf{Q}\Lambda\mathsf{Q}^{-1}, \tag{7.75}$$

where Q is unitary and its columns are orthonormal eignevectors of $\mathsf{K}_{\boldsymbol{Z}}$ The matrix Λ is diagonal and its entries are the eignevalues of $\mathsf{K}_{\boldsymbol{Z}}$, all of which are nonnegative. We can then express $\boldsymbol{Z}$ as $\boldsymbol{Z} = \mathsf{R}\,\boldsymbol{W}$ where $\mathsf{R} = \mathsf{Q}\sqrt{\Lambda}\mathsf{Q}^{-1}$ and $\mathsf{W} \sim \mathcal{CN}(0, \mathsf{I})$. □

Next note that any linear functional, say $V = \boldsymbol{b}^\dagger \boldsymbol{Z}$ of a circularly-symmetric Gaussian r**v** $\boldsymbol{Z}$ can be expressed as $V = (\boldsymbol{b}^\dagger A)\,\boldsymbol{W}$ and is thus a circularly symmetric random variable. In particular, for each orthonormal eigenfunction $\boldsymbol{q}_k$ of $\mathsf{K}_{\boldsymbol{Z}}$, we see that $\boldsymbol{q}_k^\dagger \boldsymbol{Z} = \langle \boldsymbol{Z}, \boldsymbol{q}_k \rangle$ is a circularly-symmetric rv. Furthermore, using (7.75), it is easy to show that these variables are uncorrelated, and in particular,

$$\mathsf{E}[\langle \boldsymbol{Z}, \boldsymbol{q}_k \rangle \langle \boldsymbol{Z}, \boldsymbol{q}_j \rangle^*] = \lambda_k \delta_{k,j}$$

Since these rv's are jointly Gaussian, this also means that they are statistically independent. From the projection theorem, any sample value $\boldsymbol{z}$ of the r**v** $\boldsymbol{Z}$ can be represented as $\boldsymbol{z} = \sum_j \langle \boldsymbol{z}, \boldsymbol{q}_j \rangle \boldsymbol{z}$, so we also have

$$\boldsymbol{Z} = \sum_j \langle \boldsymbol{Z}, \boldsymbol{q}_j \rangle \boldsymbol{q}_j \tag{7.76}$$

This represents $\boldsymbol{Z}$ as an orthonormal expansion whose coefficients, $\langle \boldsymbol{Z}, \boldsymbol{q}_j \rangle$ are independent circularly-symmetric Gaussian rv's. The probability density of $\boldsymbol{Z}$ is then simply the probability density of the sequence of coefficients.[24] Remembering that each circularly-symmetric Gaussian rv $\langle \boldsymbol{Z}, \boldsymbol{q}_k \rangle$ corresponds to two independent real rv's with variance $\lambda_k/2$, the resulting density, assuming that all eigenvalues are positive is

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \prod_{j=1}^{n} \frac{1}{\pi \lambda_j} \exp\left(-|\langle \boldsymbol{z}, \boldsymbol{q}_j \rangle|^2 \lambda_j^{-1}\right) \tag{7.77}$$

This is the density of n independent circularly-symmetric Gaussian random variables, $(\langle \boldsymbol{Z}, \boldsymbol{q}_1 \rangle, \ldots, \langle \boldsymbol{Z}, \boldsymbol{q}_n \rangle)$ with variances $\lambda_1, \ldots, \lambda_n$ respectively. This is the same as the analogous result for jointly-Gaussian real random vectors which says that there is always an orthonormal basis in which the variables are Gaussian and independent. This analogy forms the simplest way to (sort of) visualize circularly-symmetric Gaussian vectors – they have the same kind of elliptical symmetry as the real case, except that here, each complex random variable is also circularly symmetric.

It is often more convenient to express $f_{\boldsymbol{Z}}$ for $\boldsymbol{Z} \sim \mathcal{CN}(0, \mathsf{K}_{\boldsymbol{Z}}$ directly in terms of $\mathsf{K}_{\boldsymbol{Z}}$. Recognizing that $\mathsf{K}_{\boldsymbol{Z}}^{-1} = \mathsf{Q}\Lambda^{-1}\mathsf{Q}^{-1}$, (7.77) becomes

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{\pi^n \det(\mathsf{K}_{\boldsymbol{Z}})} \exp(-\boldsymbol{z}^\dagger \mathsf{K}_{\boldsymbol{Z}}^{-1} \boldsymbol{z}). \tag{7.78}$$

It should be clear that (7.77) or (7.78) are also if-and-only-if conditions for circularly-symmetric jointly-Gaussian random vectors with a positive-definite covariance matrix.

[24] This relies on the 'obvious' fact that incremental volume is the same in any orthonormal basis. The sceptical reader, with some labor, can work out the probability density in $\mathbb{R}^{2n}$ and then transform to $\mathbb{C}^n$.

7.9 Signal to noise ratio

There are a number of different measures of signal power, noise power, energy per symbol, energy per bit, and so forth, which are defined here. These measures are explained in terms of QAM and PAM, but they also apply more generally. In the previous section, a fairly general set of orthonormal functions was used, and here a specific set is assumed. Consider the orthonormal functions $p_k(t) = p(t - kT)$ as used in QAM, and use a nominal passband bandwidth $\mathsf{W} = 1/T$. Each QAM symbol U_k can be assumed to be iid with energy $E_s = \mathsf{E}[|U_k|^2]$. This is the signal energy per real component plus imaginary component. The noise energy per real plus imaginary component is defined to be N_0. Thus the signal to noise ratio is defined to be

$$\text{SNR} = \frac{E_s}{N_0} \qquad \text{for QAM.} \tag{7.79}$$

For baseband PAM, using real orthonormal functions satisfying $p_k(t) = p(t - kT)$, the signal energy per symbol is $E_s = \mathsf{E}[|U_k|^2]$. Since the symbol is one dimensional, *i.e.*, real, the noise energy in this single dimension is defined to be $N_0/2$. Thus SNR is defined to be

$$\text{SNR} = \frac{2E_s}{N_0} \qquad \text{for PAM.} \tag{7.80}$$

For QAM there are W complex degrees of freedom per second, so the signal power is given by $P = E_s\mathsf{W}$. For PAM at baseband, there are $2\mathsf{W}$ degrees of freedom per second, so the signal power is $P = 2E_s\mathsf{W}$. Thus in each case, the SNR becomes

$$\text{SNR} = \frac{P}{N_0\mathsf{W}} \qquad \text{for QAM and PAM.} \tag{7.81}$$

We can interpret the denominator here as the overall noise power in the bandwidth W, so SNR is also viewed as the signal power divided by the noise power in the nominal band. For those who like to minimize the number of formulas they remember, all of these equations for SNR follow from a basic definition as the signal energy per degree of freedom divided by the noise energy per degree of freedom.

PAM and QAM each use the same signal energy for each degree of freedom (or at least for each complex pair of degrees of freedom), whereas other systems might use the available degrees of freedom differently. For example, PAM with baseband bandwidth W occupies bandwidth $2\mathsf{W}$ if modulated to passband, and uses only half the available degrees of freedom. For these situations, SNR can be defined in several different ways depending on the context. As another example, frequency hopping is a technique used both in wireless and in secure communication. It is the same as QAM, except that the carrier frequency f_c changes pseudo-randomly at intervals long relative to the symbol interval. Here the bandwidth W might be taken as the bandwidth of the underlying QAM system, or might be taken as the overall bandwidth within which f_c hops. The SNR in (7.81) is quite different in the two cases.

The appearance of W in the denominator of the expression for SNR in (7.81) is rather surprising and disturbing at first. It says that if more bandwidth is allocated to a communication system with the same available power, then SNR *decreases*. This is best interpreted by viewing SNR in terms of signal to noise energy per degree of freedom. As the number of degrees of freedom per second increases, the SNR decreases, but the available number of degrees of freedom increases. We will later see that the net gain is positive.

Another important parameter is the rate R; this is the number of transmitted bits per second, which is the number of bits per symbol, $\log_2 |\mathcal{A}|$, times the number of symbols per second. Thus

$$R = \mathsf{W} \log_2 |\mathcal{A}|, \quad \text{for QAM;} \qquad R = 2\mathsf{W} \log_2 |\mathcal{A}|, \quad \text{for PAM.} \tag{7.82}$$

An important parameter is the *spectral efficiency* of the system, which is defined as $\rho = R/\mathsf{W}$. This is the transmitted number of bits/sec in each unit frequency interval. For QAM and PAM, ρ is given by (7.82) to be

$$\rho = \log_2 |\mathcal{A}|, \quad \text{for QAM;} \qquad \rho = 2 \log_2 |\mathcal{A}|, \quad \text{for PAM.} \tag{7.83}$$

More generally the spectral efficiency ρ can be defined as the number of transmitted bits per degree of freedom. From (7.83), achieving a large value of spectral efficiency requires making the symbol alphabet large; Note that ρ increases only logarithmically with $|\mathcal{A}|$.

Yet another parameter is the energy per bit E_b. Since each symbol contains $\log_2 \mathcal{A}$ bits, E_b is given for both QAM and PAM by

$$E_b = \frac{E_s}{\log_2 |\mathcal{A}|}. \tag{7.84}$$

One of the most fundamental quantities in communication is the ratio E_b/N_0. Both E_b and N_0 are measured in the same way, so the ratio is dimensionless, and it is the ratio that is important rather than either alone. Finding ways to reduce E_b/N_0 is important, particularly where transmitters use batteries. For QAM, we substitute (7.79) and (7.83) into (7.84), getting

$$\frac{E_b}{N_0} = \frac{\mathrm{SNR}}{\rho}. \tag{7.85}$$

The same equation is seen to be valid for PAM. This says that achieving a small value for E_b/N_0 requires a small ratio of SNR to ρ. We look at this next in terms of channel capacity.

One of Shannon's most famous results was to develop the concept of the capacity C of an additive WGN communication channel. This is defined as the supremum of the number of bits per second that can be transmitted and received with arbitrarily small error probability. For the WGN channel with a constraint W on the bandwidth and a constraint P on the received signal power, he showed that

$$C = \mathsf{W} \log_2 \left(1 + \frac{P}{\mathsf{W} N_0}\right). \tag{7.86}$$

He showed that any rate $R < C$ could be achieved with arbitrarily small error probability by using channel coding of arbitrarily large constraint length. He also showed, and later results strengthened, the fact that larger rates would lead to larger error probabilities. This result will be demonstrated in the next chapter. This result is widely used as a benchmark for comparison with particular systems. Figure 7.5 shows a sketch of C as a function of W. Note that C increases monotonically with W, reaching a limit of $(P/N_0) \log_2 e$ as $\mathsf{W} \to \infty$. This is known as the ultimate Shannon limit on achievable rate. Note also that when $\mathsf{W} = P/N_0$, *i.e.*, when the bandwidth is large enough for the SNR to reach 1, then C is within $1/\log_2 e$, which is 69%, of the ultimate Shannon limit.

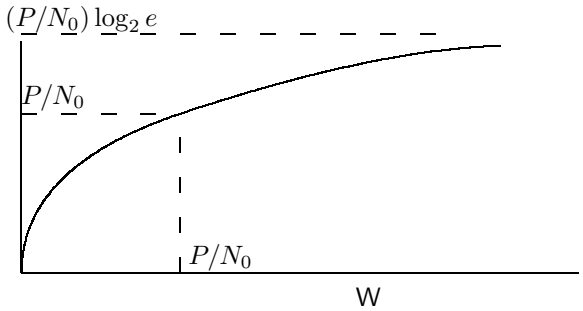


Figure 7.5: Capacity as a function of bandwidth W for fixed P/N_0.

For any achievable rate, R, *i.e.*, any rate at which the error probability can be made arbitrarily small by coding and other clever strategems, the theorem above says that $R < C$. If we rewrite (7.86), substituting SNR for $P/(\mathsf{W}N_0)$ and substituting ρ for R/W, we get

$$\rho < \log_2(1 + \text{SNR}). \tag{7.87}$$

If we substitute this into (7.85), we get

$$\frac{E_b}{N_0} > \frac{\text{SNR}}{\log_2(1 + \text{SNR})}.$$

This is a monotonic increasing function of the single variable SNR, which in turn is decreasing in W. Thus $(E_b/N_0)_{\min}$ is monotonic decreasing in W. As $\mathsf{W} \to \infty$ it reaches the limit $\ln 2 = 0.693$, *i.e.*, -1.59 dB. As W decreases, it grows, reaching 0 dB at SNR =1, and increasing without bound for yet smaller W. The limiting spectral efficiency, however, is C/W. This is also monotonic decreasing in W, going to 0 as $\mathsf{W} \to \infty$. In other words, there is a trade-off between E_b/N_0 (which we would like to be small) and spectral efficiency (which we would like to be large). This is further discussed in the next chapter.

7.10 Summary of Random Processes

The additive noise in physical communication systems is usually best modeled as a random process, *i.e.*, a collection of random variables, one at each real-valued instant of time. A random process can be specified by its joint probability distribution over all finite sets of epochs, but additive noise is most often modeled by the assumption that the random variables are all zero-mean Gaussian and their joint distribution is jointly Gaussian.

These assumptions were motivated partly by the central limit theorem, partly by the simplicity of working with Gaussian processes, partly by custom, and partly by various extremal properties. We found that jointly Gaussian means a great deal more than individually Gaussian, and that the resulting joint densities are determined by the covariance matrix. These densities have ellipsoidal equiprobability contours whose axes are the eigenfunctions of the covariance matrix.

A sample function, say $Z(t, \omega)$ of a random process $Z(t)$ can be viewed as a waveform and interpreted as an $\mathcal{L}_2$ vector. For any fixed $\mathcal{L}_2$ function $g(t)$, the inner product $\langle g(t), Z(t, \omega)\rangle$ maps ω into a real number and thus can be viewed over Ω as a random variable. This rv is called a linear function of $Z(t)$ and is denoted by $\int g(t)Z(t)\,dt$.

These linear functionals arise when expanding a random process into an orthonormal expansion and also at each epoch when a random process is passed through a linear filter. For simplicity these linear functionals and the underlying random processes are not viewed in a measure theoretic form, although the $\mathcal{L}_2$ development in Chapter 4 provides some insight about the mathematical subtleties involved.

Noise processes are usually viewed as being stationary, which effectively means that their statistics do not change in time. This generates two problems - first that the sample functions have infinite energy and second that there is no clear way to see whether results are highly sensitive to time-regions far outside the region of interest. Both of these problems are treated by defining effective stationarity (or effective wide-sense stationarity) in terms of the behavior of the process over a finite interval. This analysis shows, for example, that Gaussian linear functionals depend only on effective stationarity over the region of interest. From a practical standpoint, this means that the simple results arising from the assumption of stationarity can be used without concern for the process statistics outside the time-range of interest.

The spectral density of a stationary process can also be used without concern for the process outside the time-range of interest. If a process is effectively WSS, it has a single variable covariance function corresponding to the interval of interest, and this has a Fourier transform which operates as the spectral density over the region of interest. How these results change as the region of interest approaches ∞ is explained in Appendix 7A.3.

7A Appendix: Supplementary topics

7A.1 Properties of covariance matrices

This appendix summarizes some properties of covariance matrices that are often useful but not absolutely critical to our treatment of random processes. Rather than repeat everything twice, we combine the treatment for real and complex r**v** together. On a first reading, however, one might assume everything to be real. Most of the results are the same in each case, although the complex-conjugate signs can be removed in the real case. It is important to realize that the properties developed here apply to nonGaussian as well as Gaussian r**v**'s. All rv's and r**v**'s here are assumed to be zero-mean.

A square matrix K is a *covariance matrix* if a (real or complex) r**v** $\boldsymbol{Z}$ exists such that $\mathsf{K} = \mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^{\mathsf{T}*}]$. The complex conjugate of the transpose, $\boldsymbol{Z}^{\mathsf{T}*}$, is called the *Hermitian transpose* and denoted by $\boldsymbol{Z}^\dagger$. If $\boldsymbol{Z}$ is real, of course, $\boldsymbol{Z}^\dagger = \boldsymbol{Z}^{\mathsf{T}}$. Similarly, for a matrix K, the Hermitian conjugate, denoted $\mathsf{K}^\dagger$, is $\mathsf{K}^{\mathsf{T}*}$. A matrix is *Hermitian* if $\mathsf{K} = \mathsf{K}^\dagger$. Thus a real Hermitian matrix (a Hermitian matrix containing all real terms) is a symmetric matrix.

An n by n square matrix K with real or complex terms is *nonnegative definite* if it is Hermitian and if, for all $\boldsymbol{b} \in \mathbb{C}^n$, $\boldsymbol{b}^\dagger \mathsf{K} \boldsymbol{b}$ is real and nonnegative. It is *positive definite* if, in addition, $\boldsymbol{b}^\dagger \mathsf{K} \boldsymbol{b} > 0$ for $\boldsymbol{b} \neq 0$. We now list some of the important relationships between nonnegative definite, positive definite, and covariance matrices and state some other useful properties of covariance matrices.

1. Every covariance matrix K is nonnegative definite. To see this, let $\boldsymbol{Z}$ be a r**v** such that $\mathsf{K} = \mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^\dagger]$. K is Hermitian since $\mathsf{E}[Z_k Z_m^*] = \mathsf{E}[Z_m^* Z_k]$ for all k, m. For any $\boldsymbol{b} \in \mathbb{C}^n$, let $X = \boldsymbol{b}^\dagger \boldsymbol{Z}$. Then $0 \leq \mathsf{E}[|X|^2] = \mathsf{E}\left[(\boldsymbol{b}^\dagger \boldsymbol{Z})(\boldsymbol{b}^\dagger \boldsymbol{Z})^*\right] = \mathsf{E}\left[\boldsymbol{b}^\dagger \boldsymbol{Z}\boldsymbol{Z}^\dagger \boldsymbol{b}\right] = \boldsymbol{b}^\dagger \mathsf{K} \boldsymbol{b}$.

2. For any complex n by n matrix A, the matrix $\mathsf{K} = \mathsf{A}\mathsf{A}^\dagger$ is a covariance matrix. In fact, let $\boldsymbol{Z}$ have n independent unit-variance elements so that $\mathsf{K}_{\boldsymbol{Z}}$ is the identity matrix I_n. Then $\boldsymbol{Y} = \mathsf{A}\boldsymbol{Z}$ has the covariance matrix $\mathsf{K}_{\boldsymbol{Y}} = \mathsf{E}[(\mathsf{A}\boldsymbol{Z})(\mathsf{A}\boldsymbol{Z})^\dagger] = \mathsf{E}[\mathsf{A}\boldsymbol{Z}\boldsymbol{Z}^\dagger\mathsf{A}^\dagger] = \mathsf{A}\mathsf{A}^\dagger$. Note that if A is real and $\boldsymbol{Z}$ is real, then $\boldsymbol{Y}$ is real and, of course, $\mathsf{K}_{\boldsymbol{Y}}$ is real. It is also possible for A to be real and $\boldsymbol{Z}$ complex, and in this case $\mathsf{K}_{\boldsymbol{Y}}$ is still real but $\boldsymbol{Y}$ is complex.

3. A covariance matrix K is positive definite if and only if K is nonsingular. To see this, let $\mathsf{K} = \mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^\dagger]$ and note that if $\boldsymbol{b}^\dagger\mathsf{K}\boldsymbol{b} = 0$ for some $\boldsymbol{b} \neq 0$, then $X = \boldsymbol{b}^\dagger\boldsymbol{Z}$ has zero variance, and therefore is zero with probability 1. Thus $\mathsf{E}[X\boldsymbol{Z}^\dagger] = 0$, so $\boldsymbol{b}^\dagger\mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^\dagger] = 0$. Since $\boldsymbol{b} \neq 0$ and $\boldsymbol{b}^\dagger\mathsf{K} = 0$, K must be singular. Conversely, if K is singular, there is some $\boldsymbol{b}$ such that $\mathsf{K}\boldsymbol{b} = 0$, so $\boldsymbol{b}^\dagger\mathsf{K}\boldsymbol{b}$ is also 0.

4. A complex number λ is an *eigenvalue* of a square matrix K if $\mathsf{K}\boldsymbol{q} = \lambda\boldsymbol{q}$ for some nonzero vector $\boldsymbol{q}$; the corresponding $\boldsymbol{q}$ is an *eigenvector* of K. The following results about the eigenvalues and eigenvectors of positive (nonnegative) definite matrices K are standard linear algebra results (see for example, Strang, section 5.5):

   All eigenvalues of K are positive (nonnegative). If K is real, the eigenvectors can be taken to be real. Eigenvectors of different eigenvalues are orthogonal, and the eigenvectors of any one eigenvalue form a subspace whose dimension is called the *multiplicity* of that eigenvalue. If K is n by n, then n orthonormal eigenvectors, $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ can be chosen. The corresponding list of eigenvalues, $\lambda_1, \ldots, \lambda_n$ need not be distinct; specifically, the number of repetitions of each eigenvalue equals the multiplicity of that eigenvalue. Finally $\det(\mathsf{K}) = \prod_{k=1}^n \lambda_k$.

5. If K is nonnegative definite, let Q be the matrix with the orthonormal columns, $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ defined above. Then Q satisfies $\mathsf{K}\mathsf{Q} = \mathsf{Q}\Lambda$ where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. This is simply the vector version of the eigenvector/eigenvalue relationship above. Since $\boldsymbol{q}_k^\dagger\boldsymbol{q}_m = \delta_{nm}$, Q also satisfies $\mathsf{Q}^\dagger\mathsf{Q} = \mathsf{I}_n$. We then also have $\mathsf{Q}^{-1} = \mathsf{Q}^\dagger$ and thus $\mathsf{Q}\mathsf{Q}^\dagger = \mathsf{I}_n$; this says that the rows of Q are also orthonormal. Finally, by post-multiplying $\mathsf{K}\mathsf{Q} = \mathsf{Q}\Lambda$ by $\mathsf{Q}^\dagger$, we see that $\mathsf{K} = \mathsf{Q}\Lambda\mathsf{Q}^\dagger$. The matrix Q is called *unitary* if complex, and *orthogonal* if real.

6. If K is positive definite, then $\mathsf{K}\boldsymbol{b} \neq 0$ for $\boldsymbol{b} \neq 0$. Thus K can have no zero eigenvalues and Λ is nonsingular. It follows that K can be inverted as $\mathsf{K}^{-1} = \mathsf{Q}\Lambda^{-1}\mathsf{Q}^\dagger$. For any n-vector $\boldsymbol{b}$,

$$\boldsymbol{b}^\dagger\mathsf{K}^{-1}\boldsymbol{b} = \sum_k \lambda_k^{-1}|\langle \boldsymbol{b}, \boldsymbol{q}_k\rangle|^2.$$

   To see this, note that $\boldsymbol{b}^\dagger\mathsf{K}^{-1}\boldsymbol{b} = \boldsymbol{b}^\dagger Q\Lambda^{-1}\mathsf{Q}^\dagger\boldsymbol{b}$. Letting $\boldsymbol{v} = \mathsf{Q}^\dagger\boldsymbol{b}$ and using the fact that the rows of $\mathsf{Q}^\intercal$ are the orthonormal vectors $\boldsymbol{q}_k$, we see that $\langle \boldsymbol{b}, \boldsymbol{q}_k\rangle$ is the kth component of $\boldsymbol{v}$. We then have $\boldsymbol{v}^\dagger\Lambda^{-1}\boldsymbol{v} = \sum_k \lambda_k^{-1}|v_k|^2$, which is equivalent to the desired result. Note that $\langle \boldsymbol{b}, \boldsymbol{q}_k\rangle$ is the projection of $\boldsymbol{b}$ in the direction of $\boldsymbol{q}_k$.

7. $\det \mathsf{K} = \prod_{k=1}^n \lambda_k$ where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of K repeated according to their multiplicity. Thus if K is positive definite, $\det \mathsf{K} > 0$ and if K is nonnegative definite, $\det \mathsf{K} \geq 0$.

8. If K is a positive definite (semi-definite) matrix, then there is a unique positive definite (semi-definite) square root matrix R satisfying $\mathsf{R}^2 = \mathsf{K}$. In particular, R is given by

$$\mathsf{R} = \mathsf{Q}\Lambda^{1/2}\mathsf{Q}^\dagger \text{ where } \Lambda^{1/2} = \mathrm{diag}\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n}\right). \tag{7.88}$$

9. If K is nonnegative definite, then K is a covariance matrix. In particular, K is the covariance matrix of $\boldsymbol{Y} = \mathsf{R}\boldsymbol{V}$ where R is the square root matrix in (7.88) and $\mathsf{K}_{\boldsymbol{V}} = \mathsf{I}_m$

This shows that zero-mean jointly-Gaussian r**v**'s exist with any desired covariance matrix; the definition of jointly Gaussian here as a linear combination of normal rv's does not limit the possible set of covariance matrices.

For any given covariance matrix K, there are usually many choices for A satisfying $\mathsf{K} = \mathsf{AA}^{\mathsf{T}}$. The square root matrix R above is simply a convenient choice. Some of the results in this section are summarized in the following theorem:

Theorem 7A.1. An n by n matrix K is a covariance matrix if and only if it is nonnegative definite. Also it is a covariance matrix if and only if $\mathsf{K} = \mathsf{AA}^{\dagger}$ for an n by n matrix A. One choice for A is the square root matrix R in (7.88).

7A.2 The Fourier series expansion of a truncated random process

Consider a (real zero-mean) random process that is effectively WSS over some interval $[-\frac{T_0}{2}, \frac{T_0}{2}]$ where T_0 is viewed intuitively as being very large. Let $\{Z(t); |t| \leq \frac{T_0}{2}\}$ be this process truncated to the interval $[-\frac{T_0}{2}, \frac{T_0}{2}]$. The objective of this and the next appendix is to view this truncated process in the frequency domain and discover its relation to the spectral density of an untruncated WSS process. A second objective is to interpret the statistical independence between different frequencies for stationary Gaussian processes in terms of a truncated process.

Initially assume that $\{Z(t); |t| \leq \frac{T_0}{2}\}$ is arbitrary; the effective WSS assumption will be added later. Assume the sample functions of the truncated process are $\mathcal{L}_2$ real functions with probability 1. Each $\mathcal{L}_2$ sample function, say $\{Z(t,\omega); |t| \leq \frac{T_0}{2}\}$ can then be expanded in a Fourier series,

$$Z(t,\omega) = \sum_{m=-\infty}^{\infty} \hat{Z}_k(\omega) e^{2\pi i k t/T_0}, \qquad |t| \leq \frac{T_0}{2}. \tag{7.89}$$

The orthogonal functions here are complex and the coefficients $\hat{Z}_k(\omega)$ can be similarly complex. Since the sample functions $\{Z(t,\omega);\ |t| \leq \frac{T_0}{2}\}$ are real, $\hat{Z}_k(\omega) = \hat{Z}^*_{-k}(\omega)$ for each k. This also implies that $\hat{Z}_0(\omega)$ is real. The inverse Fourier series is given by

$$\hat{Z}_k(\omega) = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} Z(t,\omega) e^{-2\pi i k t/T_0}\, dt. \tag{7.90}$$

For each sample point ω, $\hat{Z}_k(\omega)$ is a complex number, so $\hat{Z}_k$ is a complex random variable, *i.e.*, $\Re(\hat{Z}_k)$ and $\Im(\hat{Z}_k)$ are each rv's. Also, $\Re(\hat{Z}_k) = \Re(\hat{Z}_{-k})$ and $\Im(\hat{Z}_k) = -\Im(\hat{Z}_{-k})$ for each k. It follows that the truncated process $\{Z(t); |t| \leq \frac{T_0}{2}\}$ defined by

$$Z(t) = \sum_{k=-\infty}^{\infty} \hat{Z}_k e^{2\pi i k t/T_0}, \qquad -\frac{T_0}{2} \leq t \leq \frac{T_0}{2}. \tag{7.91}$$

is a (real) random process and the complex random variables $\hat{Z}_k$ are complex linear functionals of $Z(t)$ given by

$$\hat{Z}_k = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} Z(t) e^{-2\pi i k t/T_0}\, dt. \tag{7.92}$$

Thus (7.91) and (7.92) are a Fourier series pair between a random process and a sequence of complex rv's. The sample functions satisfy

$$\frac{1}{T_0}\int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} Z^2(t,\omega)\,dt = \sum_{k\in\mathbb{Z}} |\hat{Z}_k(\omega))|^2,$$

so that

$$\frac{1}{T_0}\mathsf{E}\left[\int_{t=-\frac{T_0}{2}}^{\frac{T_0}{2}} Z^2(t)\,dt\right] = \sum_{k\in\mathbb{Z}} \mathsf{E}\left[|\hat{Z}_k|^2\right]. \tag{7.93}$$

The assumption that the sample functions are $\mathcal{L}_2$ with probability 1 can be seen to be equivalent to the assumption that

$$\sum_{k\in\mathbb{Z}} S_k < \infty \qquad \text{where } S_k = \mathsf{E}[|\hat{Z}_k|^2]. \tag{7.94}$$

This is summarized in the following theorem.

Theorem 7A.2. *If a zero-mean (real) random process is truncated to $[-\frac{T_0}{2}, \frac{T_0}{2}]$ and the truncated sample functions are $\mathcal{L}_2$ with probability 1, then the truncated process is specified by the joint distribution of the complex Fourier-coefficient random variables $\{\hat{Z}_k\}$. Furthermore, any joint distribution of $\{\hat{Z}_k; k\in\mathbb{Z}\}$ that satisfies (7.94) specifies such a truncated process.*

The covariance function of a truncated process can be calculated from (7.91) as follows:

$$\begin{aligned}\mathsf{K}_{\boldsymbol{Z}}(t,\tau) &= \mathsf{E}[Z(t)Z^*(\tau)] = \mathsf{E}\left[\sum_k \hat{Z}_k e^{2\pi ikt/T_0}\sum_m \hat{Z}_m^* e^{-2\pi im\tau/T_0}\right]\\ &= \sum_{k,m}\mathsf{E}[\hat{Z}_k\hat{Z}_m^*]e^{2\pi ikt/T_0}e^{-2\pi im\tau/T_0}, \qquad \text{for } -\frac{T_0}{2}\le t,\tau\le\frac{T_0}{2}.\end{aligned} \tag{7.95}$$

Note that if the function on the right of (7.95) is extended over all $t,\tau\in\mathbb{R}$, it becomes periodic in t with period T_0 for each τ, and periodic in τ with period T_0 for each t.

Theorem 7A.2 suggests that virtually any truncated process can be represented as a Fourier series. Such a representation becomes far more insightful and useful, however, if the Fourier coefficients are uncorrelated. The next two subsections look at this case and then specialize to Gaussian processes, where uncorrelated implies independent.

7A.3 Uncorrelated coefficients in a Fourier series

Consider the covariance function in (7.95) under the additional assumption that the Fourier coefficients $\{\tilde{Z}_k; k\in\mathbb{Z}\}$ are uncorrelated, *i.e.*, that $\mathsf{E}[\hat{Z}_k\hat{Z}_m^*]=0$ for all k,m such that $k\neq m$. This assumption also holds for $m=-k$, and, since $Z_k = Z_{-k}^*$ for all k, implies both that $\mathsf{E}[(\Re(Z_k))^2] = \mathsf{E}[(\Im(Z_k))^2]$ and $\mathsf{E}[\Re(Z_k)\Im(Z_k)]=0$ (see Exercise 7.10). Since $\mathsf{E}[\hat{Z}_k\hat{Z}_m^*]=0$ for $k\neq m$, (7.95) simplifies to

$$\mathsf{K}_{\boldsymbol{Z}}(t,\tau) = \sum_{k\in\mathbb{Z}} S_k e^{2\pi ik(t-\tau)/T_0}, \qquad \text{for } -\frac{T_0}{2}\le t,\tau\le\frac{T_0}{2}. \tag{7.96}$$

This says that $\mathsf{K}_{\boldsymbol{Z}}(t,\tau)$ is a function only of $t-\tau$ over $-\frac{T_0}{2} \le t,\tau \le \frac{T_0}{2}$, *i.e.*, that $\mathsf{K}_{\boldsymbol{Z}}(t,\tau)$ is effectively WSS over $[\frac{T_0}{2}, \frac{T_0}{2}]\}$. Thus $\mathsf{K}_{\boldsymbol{Z}}(t,\tau)$ can be denoted as $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau)$ in this region, and

$$\tilde{\mathsf{K}}_{\boldsymbol{Z}}(\tau) = \sum_k S_k e^{2\pi ik\tau/T_0}. \tag{7.97}$$

This means that the variances S_k of the sinusoids making up this process are the Fourier series coefficients of the covariance function $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(r)$.

In summary, the assumption that a truncated (real) random process has uncorrelated Fourier series coefficients over $[-\frac{T_0}{2}, \frac{T_0}{2}]$ implies that the process is WSS over $[-\frac{T_0}{2}, \frac{T_0}{2}]$ and that the variances of those coefficients are the Fourier coefficients of the single variable covariance. This is intuitively plausible since the sine and cosine components of each of the corresponding sinusoids are uncorrelated and have equal variance.

Note that $\mathsf{K}_{\boldsymbol{Z}}(t,\tau)$ in the above example is defined for all $t,\tau \in [-\frac{T_0}{2}, \frac{T_0}{2}]$ and thus $t-\tau$ ranges from $-T_0$ to T_0 and $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(r)$ must satisfy (7.97) for $-T_0 \le r \le T_0$. From (7.97), $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(r)$ is also periodic with period T_0, so the interval $[-T_0, T_0]$ constitutes 2 periods of $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(r)$. This means, for example, that $\mathsf{E}[Z(-\varepsilon)Z^*(\varepsilon)] = \mathsf{E}[Z(\frac{T_0}{2}-\varepsilon)Z^*(-\frac{T_0}{2}+\varepsilon)]$. More generally, the periodicity of $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(r)$ is reflected in $\mathsf{K}_{\boldsymbol{Z}}(t,\tau)$ as illustrated in figure 7.6.

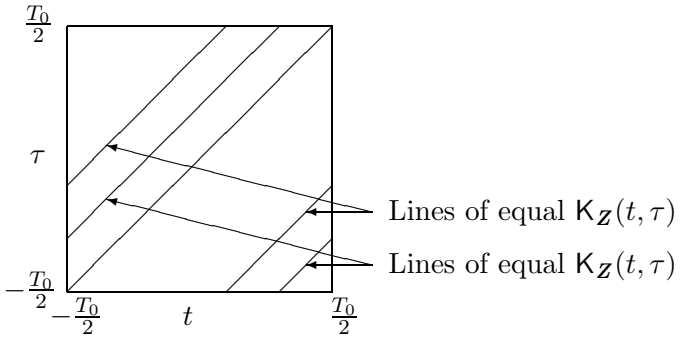


Figure 7.6: Constraint on $\mathsf{K}_{\boldsymbol{Z}}(t,\tau)$ imposed by periodicity of $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau)$.

We have seen that essentially any random process, when truncated to $[-\frac{T_0}{2}, \frac{T_0}{2}]$, has a Fourier series representation, and that if the Fourier series coefficients are uncorrelated, then the truncated process is WSS over $[-\frac{T_0}{2}, \frac{T_0}{2}]$ and has a covariance function which is periodic with period T_0. This proves the first half of the following theorem:

Theorem 7A.3. *Let $\{Z(t);\, t\in[-\frac{T_0}{2}, \frac{T_0}{2}]\}$ be a finite-energy zero-mean (real) random process over $[-\frac{T_0}{2}, \frac{T_0}{2}]$ and let $\{\hat{Z}_k;\, k\in\mathbb{Z}\}$ be the Fourier series rv's of (7.91) and (7.92).*

- *If $\mathsf{E}[Z_k Z_m^*] = S_k\delta_{k,m}$ for all $k,m \in \mathbb{Z}$, then $\{Z(t); t \in [-\frac{T_0}{2}, \frac{T_0}{2}]\}$ is effectively WSS within $[-\frac{T_0}{2}, \frac{T_0}{2}]$ and satisfies (7.97).*
- *If $\{Z(t);\, t\in[-\frac{T_0}{2}, \frac{T_0}{2}]\}$ is effectively WSS within $[-\frac{T_0}{2}, \frac{T_0}{2}]$ and if $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau)$ is periodic with period T_0 over $[-T_0, T_0]$, then $\mathsf{E}[Z_k Z_m^*] = S_k\delta_{k,m}$ for some choice of $S_k \ge 0$ and for all $k,m \in \mathbb{Z}$.*

Proof: To prove the second part of the theorem, note from (7.92) that

$$\mathsf{E}[\hat{Z}_k \hat{Z}_m^*] = \frac{1}{T_0^2}\int_{-\frac{T_0}{2}}^{\frac{T_0}{2}}\int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} \mathsf{K}_{\boldsymbol{Z}}(t,\tau)e^{-2\pi ikt/T_0}e^{2\pi im\tau/T_0}\, dt\, d\tau. \tag{7.98}$$

By assumption, $\mathsf{K}_{\boldsymbol{Z}}(t,\tau) = \tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau)$ for $t,\tau \in [-\frac{T_0}{2}, \frac{T_0}{2}]$ and $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(t-\tau)$ is periodic with period T_0. Substituting $s = t-\tau$ for t as a variable of integration, (7.98) becomes

$$\mathsf{E}[Z_k Z_m^*] = \frac{1}{T_0^2} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} \left(\int_{-\frac{T_0}{2}-\tau}^{\frac{T_0}{2}-\tau} \tilde{\mathsf{K}}_{\boldsymbol{Z}}(s) e^{-2\pi i k s/T_0}\, ds \right) e^{-2\pi i k\tau/T_0} e^{2\pi i m\tau/T_0}\, d\tau. \tag{7.99}$$

The integration over s does not depend on τ because the interval of integration is one period and $\tilde{\mathsf{K}}_{\boldsymbol{Z}}$ is periodic. Thus this integral is only a function of k, which we denote by $T_0 S_k$. Thus

$$\mathsf{E}[Z_k Z_m^*] = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} S_k e^{-2\pi i(k-m)\tau/T_0}\, d\tau = \begin{cases} S_k & \text{for } m = k \\ 0 & \text{otherwise} \end{cases} \tag{7.100}$$

This shows that the Z_k are uncorrelated, completing the proof. □

The next issue is to find the relationship between these processes and processes that are WSS over all time. This can be done most cleanly for the case of Gaussian processes. Consider a WSS (and therefore stationary) zero-mean Gaussian random process[25] $\{Z'(t); t \in \mathbb{R}\}$ with covariance function $\tilde{\mathsf{K}}_{\boldsymbol{Z}'}(\tau)$ and assume a limited region of nonzero covariance *i.e.*,

$$\tilde{\mathsf{K}}_{\boldsymbol{Z}'}(\tau) = 0 \qquad \text{for} \quad |\tau| > \frac{T_1}{2}.$$

Let $S_{Z'}(f) \geq 0$ be the spectral density of Z' and let T_0 satisfy $T_0 > T_1$. The Fourier series coefficients of $\tilde{\mathsf{K}}_{\boldsymbol{Z}'}(\tau)$ over the interval $[-\frac{T_0}{2}, \frac{T_0}{2}]$ are then given by $S_k = \frac{S_{Z'}(k/T_0)}{T_0}$. Suppose this process is approximated over the interval $[-\frac{T_0}{2}, \frac{T_0}{2}]$ by a truncated Gaussian process $\{Z(t);\, t{\in}[-\frac{T_0}{2}, \frac{T_0}{2}]\}$ composed of independent Fourier coefficients $\hat{Z}_k$, *i.e.*

$$Z(t) = \sum_k \hat{Z}_k e^{2\pi i k t/T_0}, \qquad -\frac{T_0}{2} \leq t \leq \frac{T_0}{2},$$

where

$$\mathsf{E}[\hat{Z}_k \hat{Z}_m^*] = S_k \delta_{k,m} \qquad \text{for all } k, m \in \mathbb{Z}.$$

By Theorem 7A.3, the covariance function of $Z(t)$ is $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(\tau) = \sum_k S_k e^{2\pi i k t/T_0}$. This is periodic with period T_0 and for $|\tau| \leq \frac{T_0}{2}$, $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(\tau) = \tilde{\mathsf{K}}_{\boldsymbol{Z}'}(\tau)$. The original process $Z'(t)$ and the approximation $Z(t)$ thus have the same covariance for $|\tau| \leq \frac{T_0}{2}$. For $|\tau| > \frac{T_0}{2}$, $\tilde{\mathsf{K}}_{\boldsymbol{Z}'}(\tau) = 0$ whereas $\tilde{\mathsf{K}}_{\boldsymbol{Z}}(\tau)$ is periodic over all τ. Also, of course, Z' is stationary, whereas Z is effectively stationary within its domain $[-\frac{T_0}{2}, \frac{T_0}{2}]$. The difference between Z and Z' becomes more clear in terms of the two-variable covariance function, illustrated in Figure 7.7.

It is evident from the figure that if Z' is modeled as a Fourier series over $[-\frac{T_0}{2}, \frac{T_0}{2}]$ using independent complex circularly symmetric Gaussian coefficients, then $\mathsf{K}_{Z'}(t,\tau) = \mathsf{K}_Z(t,\tau)$ for $|t|, |\tau| \leq \frac{T_0 - T_1}{2}$. Since zero-mean Gaussian processes are completely specified by their covariance functions, this means that Z' and Z are statistically identical over this interval.

In summary, a stationary Gaussian process Z' can not be perfectly modeled over an interval $[-\frac{T_0}{2}, \frac{T_0}{2}]$ by using a Fourier series over that interval. The anomalous behavior is avoided,

[25] Equivalently, one can assume that Z' is effectively WSS over some interval much larger than the intervals of interest here.

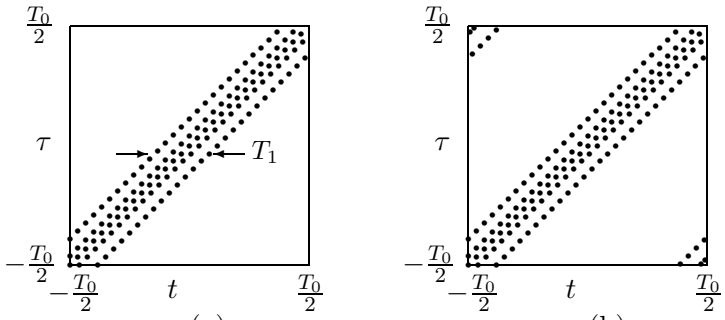


Figure 7.7: Part (a) illustrates $\mathsf{K}_{Z'}(t,\tau)$ over the region $-\frac{T_0}{2} \le t,\tau \le \frac{T_0}{2}$ for a stationary process Z' satisfying $\tilde{\mathsf{K}}_{\boldsymbol{Z}'}(\tau) = 0$ for $|\tau| > T_1/2$. Part (b) illustrates the approximating process Z comprised of independent sinusoids, spaced by $1/T_0$ and with uniformly distribuited phase. Note that the covariance functions are identical except for the anomalous behavior at the corners where t is close to $T_0/2$ and τ is close to $-T_0/2$ or vice versa.

however, by using a Fourier series over a larger interval, large enough to include the interval of interest plus the interval over which $\tilde{\mathsf{K}}_{\boldsymbol{Z}'}(\tau) \neq 0$. If this latter interval is unbounded, then the Fourier series model can only be used as an approximation. The following theorem has been established.

Theorem 7A.4. *Let $Z'(t)$ be a zero-mean stationary Gaussian random process with spectral density $S(f)$ and covariance $\tilde{\mathsf{K}}_{\boldsymbol{Z}'}(\tau) = 0$ for $|\tau| \ge T_1/2$. Then for $T_0 > T_1$, the truncated process $Z(t) = \sum_k Z_k e^{2\pi i k t/T_0}$ for $|t| \le \frac{T_0}{2}$, where the Z_k are independent and $Z_k \sim \mathcal{CN}(\frac{S(k/T_0)}{T_0})$ for all $k \in \mathbb{Z}$ is statistically identical to $Z'(t)$ over $[-\frac{T_0-T_1}{2}, \frac{T_0-T_1}{2}]$.*

The above theorem is primarily of conceptual use, rather than as a problem solving tool. It shows that, aside from the anomalous behavior discussed above, stationarity can be used over the region of interest without concern for how the process behaves outside far outside the interval of interest. Also, since T_0 can be arbitrarily large, and thus the sinusoids arbitrarily closely spaced, we see that the relationship between stationarity of a Gaussian process and independence of frequency bands is quite robust and more than something valid only in a limiting sense.

7A.4 The Karhunen-Loeve expansion

There is another approach, called the Karhunen-Loeve expansion for representing a random process that is truncated to some interval $[-\frac{T_0}{2}, \frac{T_0}{2}]$ by an orthonormal expansion. The objective is to choose a set of orthonormal functions such that the coefficients in the expansion are uncorrelated.

We start with the covariance function $\mathsf{K}(t,\tau)$ defined for $t,\tau \in [-\frac{T_0}{2}, \frac{T_0}{2}]$. The basic facts about these time-limited covariance functions are virtually the same as the facts about covariance matrices in Appendix 7A.1. $\mathsf{K}(t,\tau)$ is nonnegative definite in the sense that for all $\mathcal{L}_2$ functions $g(t)$,

$$\int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} g(t)\mathsf{K}_{\boldsymbol{Z}}(t,\tau)g(\tau)\,dt\,d\tau \ge 0$$

$\mathsf{K}_{\boldsymbol{Z}}$ also has real valued orthonormal eigenvectors defined over $[-\frac{T_0}{2}, \frac{T_0}{2}]$ and nonnegative eigen-

values. That is

$$\int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} \mathsf{K}_{\boldsymbol{Z}}(t,\tau)\phi_m(\tau)\,d\tau = \lambda_m\phi_m(t); \quad t \in \left[-\frac{T_0}{2},\, \frac{T_0}{2}\right] \qquad \text{where } \langle \boldsymbol{\phi}_m, \boldsymbol{\phi}_k\rangle = \delta_{m,k}$$

These eigenvectors span the $\mathcal{L}_2$ space of real functions over $[-\frac{T_0}{2},\, \frac{T_0}{2}]$. By using these eigenvectors as the orthonormal functions of $Z(t) = \sum_m Z_m\phi_m(t)$, it is easy to show that $\mathsf{E}[Z_mZ_k] = \lambda_m\delta_{m,k}$. In other words, given an arbitrary covariance function over the truncated interval $[-\frac{T_0}{2},\, \frac{T_0}{2}]$, we can find a particular set of orthonormal functions so that $Z(t) = \sum_m Z_m\phi_m(t)$ and $\mathsf{E}[Z_mZ_k] = \lambda_m\delta_{m,k}$. This is called the Karhunen-Loeve expansion.

These equations for the eigenvectors and eigenvalues are well-known integral equations and can be calculated by computer. Unfortunately they do not provide a great deal of insight into the frequency domain.

7.E Exercises

7.1. (a) Let X, Y be iid rv's, each with density $f_X(x) = \alpha \exp(-x^2/2)$. In part (b), we show that α must be $1/\sqrt{2\pi}$ in order for $f_X(x)$ to integrate to 1, but in this part, we leave α undetermined. Let $S = X^2 + Y^2$. Find the probability density of S in terms of α. Hint: Sketch the contours of equal probability density in the X, Y plane.

(b) Prove from part (a) that α must be $1/\sqrt{2\pi}$ in order for S, and thus X and Y, to be random variables. Show that $\mathsf{E}[X] = 0$ and that $\mathsf{E}[X^2] = 1$.

(c) Find the probability density of $R = \sqrt{S}$. R is called a *Rayleigh* rv.

7.2. (a) Let $X \sim \mathcal{N}(0, \sigma_X^2)$ and $Y \sim \mathcal{N}(0, \sigma_Y^2)$ be independent zero-mean Gaussian rv's. By convolving their densities, find the density of $X+Y$. Hint: In performing the integration for the convolution, you should do something called "completing the square" in the exponent. This involves multiplying and dividing by $e^{\alpha y^2/2}$ for some α, and you can be guided in this by knowing what the answer is. This technique is invaluable in working with Gaussian rv's.

(b) The Fourier transform of a probability density $f_X(x)$ is $\hat{f}_X(\theta) = \int f_X(x) e^{-2\pi i x\theta}\, dx = \mathsf{E}[e^{-2\pi i X\theta}]$. By scaling the basic Gaussian transform in (4.48), show that for $X \sim \mathcal{N}(0, \sigma_X^2)$,

$$\hat{f}_X(\theta) = \exp\left[-\frac{(2\pi\theta)^2\sigma_X^2}{2}\right].$$

(b) Now find the density of $X + Y$ by using Fourier transforms of the densities.

(c) Using the same Fourier transform technique, find the density of $V = \sum_{k=1}^{n} \alpha_k W_k$ where $W_1, \ldots, W_k$ are independent normal rv's.

7.3. In this exercise you will construct two rv's that are individually Gaussian but not jointly Gaussian. Consider the nonnegative random variable X with the density

$$f_X(x) = \sqrt{\frac{2}{\pi}}\, \exp\left(\frac{-x^2}{2}\right) \qquad \text{for } x \geq 0.$$

Let U be binary, ± 1, with $p_U(1) = p_U(-1) = 1/2$.

(a) Find the probability density of $Y_1 = UX$. Sketch the density of Y_1 and find its mean and variance.

(b) Describe two normalized Gaussian rv's, say Y_1 and Y_2, such that the joint density of Y_1, Y_2 is *zero* in the second and fourth quadrants of the plane. It is nonzero in the first and third quadrants where it has the density $\frac{1}{\pi}\exp(\frac{-y_1^2-y_2^2}{2})$. Hint: Use part (a) for Y_1 and think about how to construct Y_2.

(c) Find the covariance $\mathsf{E}[Y_1 Y_2]$. Hint: First find the mean of the rv X above.

(d) Use a variation of the same idea to construct two normalized Gaussian rv's V_1, V_2 whose probability is concentrated on the diagonal axes $v_1 = v_2$ and $v_1 = -v_2$, *i.e.*, for which $\Pr(V_1 \neq V_2 \text{ and } V_1 \neq -V_2) = 0$.

7.4. Let $W_1 \sim \mathcal{N}(0,1)$ and $W_2 \sim \mathcal{N}(0,1)$ be independent normal rv's. Let $X = \max(W_1, W_2)$ and $Y = \min(W_1, W_2)$.

(a) Sketch the transformation from sample values of W_1, W_2 to sample values of X, Y. Which sample pairs w_1, w_2 of W_1, W_2 map into a given sample pair x, y of X, Y?

(b) Find the probability density $f_{XY}(x,y)$ of X,Y. Explain your argument briefly but work from your sketch rather than equations.

(c) Find $f_S(s)$ where $S = X + Y$.

(d) Find $f_D(d)$ where $D = X - Y$.

(e) Let U be a random variable taking the values ± 1 with probability 1/2 each and let U be statistically independent of W_1, W_2. Are S and UD jointly Gaussian?

7.5. Let $\phi(t)$ be an $\mathcal{L}_2$ function of energy 1 and let $h(t)$ be $\mathcal{L}_2$. Show that $\int_{-\infty}^{\infty} \phi(t)h(\tau - t)\,dt$ is an $\mathcal{L}_2$ function of τ with energy upper bounded by $\|\boldsymbol{h}\|^2$. Hint: Consider the Fourier transform of $\phi(t)$ and $h(t)$.

7.6. (a) Generalize the random process of (7.30) by assuming that the Z_k are arbitrarily correlated. Show that every sample function is still $\mathcal{L}_2$.

(b) For this same case, show that $\iint |\mathsf{K}_{\mathbf{Z}}(t,\tau)|^2\,dtd\tau < \infty$.

7.7. (a) Let $Z_1, Z_2, \ldots$, be a sequence of independent Gaussian rv's, $Z_k \sim \mathcal{N}(0, \sigma_k^2)$ and let $\{\phi_k(t) : \mathbb{R} \to \mathbb{R}\}$ be a sequence of orthonormal functions. Argue from fundamental definitions that for each t, $Z(t) = \sum_{k=1}^{n} Z_k\phi_k(t)$ is a Gaussian random variable. Find the variance of $Z(t)$ as a function of t.

(b) For any set of epochs, $t_1, \ldots, t_\ell$, let $Z(t_m) = \sum_{k=1}^{n} Z_k\phi_k(t_m)$ for $1 \le m \le \ell$. Explain carefully from the basic definitions why $\{Z(t_1), \ldots, Z(t_\ell)\}$ are jointly Gaussian and specify their covariance matrix. Explain why $\{Z(t); t \in \mathbb{R}\}$ is a Gaussian random process.

(c) Now let $n = \infty$ above and assume that $\sum_k \sigma_k^2 < \infty$. Also assume that the orthonormal functions are bounded for all k and t in the sense that for some constant A, $|\phi_k(t)| \le A$ for all k and t. Consider the linear combination of rv's

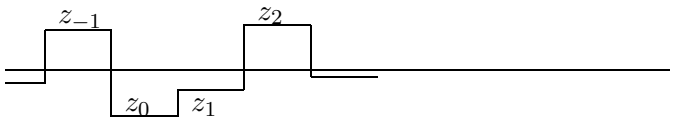
$$Z(t) = \sum_k Z_k\phi_k(t) = \lim_{n\to\infty} \sum_{k=1}^{n} Z_k\phi_k(t)$$

Let $Z^{(n)}(t) = \sum_{k=1}^{n} Z_k\phi_k(t)$. For any given t, find the variance of $Z^{(j)}(t) - Z^{(n)}(t)$ for $j > n$. Show that for all $j > n$, this variance approaches 0 as $n \to \infty$. Explain intuitively why this indicates that $Z(t)$ is a Gaussian rv. Note: $Z(t)$ is in fact a Gaussian rv, but proving this rigorously requires considerable background. $Z(t)$ is a limit of a sequence of rv's, and each rv is a function of a sample space - the issue here is the same as that of a sequence of functions going to a limit function, where we had to invoke the Riesz-Fischer theorem.

(d) For the above Gaussian random process $\{Z(t); t \in \mathbb{R}\}$, let $z(t)$ be a sample function of $Z(t)$ and find its energy, *i.e.*, $\|\boldsymbol{z}\|^2$ in terms of the sample values $z_1, z_2, \ldots$ of $Z_1, Z_2, \ldots$. Find the expected energy in the process, $\mathsf{E}[\|\{Z(t); t \in \mathbb{R}\}\|^2]$.

(e) Find an upper bound on $\Pr\{\|\{Z(t); t \in \mathbb{R}\}\|^2 > \alpha\}$ that goes to zero as $\alpha \to \infty$. Hint: You might find the Markov inequality useful. This says that for a nonnegative rv Y, $\Pr\{Y \ge \alpha\} \le \frac{\mathsf{E}[Y]}{\alpha}$. Explain why this shows that the sample functions of $\{Z(t)\}$ are $\mathcal{L}_2$ with probability 1.

7.8. Consider a stochastic process $\{Z(t); t \in \mathbb{R}\}$ for which each sample function is a sequence of rectangular pulses as in the figure below.



Analytically, $Z(t) = \sum_{k=-\infty}^{\infty} Z_k \text{rect}(t-k)$ where $\ldots Z_{-1}, Z_0, Z_1, \ldots$ is a sequence of iid normal variables, $Z_k \sim \mathcal{N}(0,1)$..

(a) Is $\{Z(t);\, t \in \mathbb{R}\}$ a Gaussian random process? Explain why or why not carefully.

(b) Find the covariance function of $\{Z(t);\, t \in \mathbb{R}\}$.

(c) Is $\{Z(t);\, t \in \mathbb{R}\}$ a stationary random process? Explain carefully.

(d) Now suppose the stochastic process is modified by introducing a random time shift Φ which is uniformly distributed between 0 and 1. Thus, the new process, $\{V(t);\, t \in \mathbb{R}\}$ is defined by $V(t) = \sum_{k=-\infty}^{\infty} Z_k \text{rect}(t-k-\Phi)$. Find the conditional distribution of $V(0.5)$ conditional on $V(0) = v$.

(e) Is $\{V(t);\, t \in \mathbb{R}\}$ a Gaussian random process? Explain why or why not carefully.

(f) Find the covariance function of $\{V(t);\, t \in \mathbb{R}\}$.

(g) Is $\{V(t);\, t \in \mathbb{R}\}$ a stationary random process? It is easier to explain this than to write a lot of equations.

7.9. Consider the Gaussian sinc process, $V(t) = \sum_k V_k \,\text{sinc}\left(\frac{t-kT}{T}\right)$ where $\{\ldots, V_{-1}, V_0, V_1, \ldots,\}$ is a sequence of iid rv's, $V_k \sim \mathcal{N}(0, \sigma^2)$.

(a) Find the probability density for the linear functional $\int V(t)\text{sinc}(\frac{t}{T})\, dt$.

(b) Find the probability density for the linear functional $\int V(t)\text{sinc}(\frac{\alpha t}{T})\, dt$ for $\alpha > 1$.

(c) Consider a linear filter with impulse response $h(t) = \text{sinc}(\frac{\alpha t}{T})$ where $\alpha > 1$. Let $\{Y(t)\}$ be the output of this filter when $V(t)$ above is the input. Find the covariance function of the process $\{Y(t)\}$. Explain why the process is Gaussian and why it is stationary.

(d) Find the probability density for the linear functional $Y(\tau) = \int V(t)\text{sinc}(\frac{\alpha(t-\tau)}{T})\, dt$ for $\alpha \geq 1$ and arbitrary τ.

(e) Find the spectral density of $\{Y(t); t \in \mathbb{R}\}$.

(f) Show that $\{Y(t); t \in \mathbb{R}\}$ can be represented as $Y(t) = \sum_k Y_k \text{sinc}\left(\frac{t-kT}{T}\right)$ and characterize the rv's $\{Y_k; k \in Z\}$.

(g) Repeat parts (c), (d), and (e) for $\alpha < 1$.

(h) Show that $\{Y(t)\}$ in the $\alpha < 1$ case can be represented as a Gaussian sinc process (like $\{V(t)\}$ but with an appropriately modified value of T).

(i) Show that if any given process $\{Z(t); t \in \mathbb{R}\}$ is stationary, then so is the process $\{Y(t); t \in \mathbb{R}\}$ where $Y(t) = Z^2(t)$ for all $t \in \mathbb{R}$.

7.10. (Complex random variables)(a) Suppose the zero-mean complex random variables X_k and X_{-k} satisfy $X^*_{-k} = X_k$ for all k. Show that if $\mathsf{E}[X_k X^*_{-k}] = 0$ then $\mathsf{E}[(\Re(X_k))^2] = \mathsf{E}[(\Im(X_k))^2]$ and $\mathsf{E}[\Re(X_k)\Im(X_{-k})] = 0$.

(b) Use this to show that if $\mathsf{E}[X_k X^*_m] = 0$ then $\mathsf{E}[\Re(X_k)\Re(X_m)] = 0$, $\mathsf{E}[\Re(X_k)\Im(X_m)] = 0$, and $\mathsf{E}[\Im(X_k)\Im(X_m)] = 0$ for all m not equal to either k or $-k$.

7.11. Explain why the integral in (7.58) must be real for $g_1(t)$ and $g_2(t)$ real, but the integrand $\hat{g}_1(f)S_{\boldsymbol{Z}}(f)\hat{g}^*_2(f)$ need not be real.

7.12. (Filtered white noise) Let $\{Z(t)\}$ be a White Gaussian noise process of spectral density $N_0/2$.

(a) Let $Y = \int_0^T Z(t)\,dt$. Find the probability density of Y.

(b) Let $Y(t)$ be the result of passing $Z(t)$ through an ideal baseband filter of bandwidth W whose gain is adjusted so that its impulse response has unit energy. Find the joint distribution of $Y(0)$ and $Y(\frac{1}{4\mathsf{W}})$.

(c) Find the probability density of

$$V = \int_0^\infty e^{-t} Z(t)\,dt.$$

7.13. (Power spectral density) (a) Let $\{\phi_k(t)\}$ be any set of real orthonormal $\mathcal{L}_2$ waveforms whose transforms are limited to a band B, and let $\{W(t)\}$ be white Gaussian noise with respect to B with power spectral density $S_W(f) = N_0/2$ for $f \in B$. Let the orthonormal expansion of $W(t)$ with respect to the set $\{\phi_k(t)\}$ be defined by

$$\tilde{W}(t) = \sum_k W_k \phi_k(t),$$

where $W_k = \langle W(t), \phi_k(t)\rangle$. Show that $\{W_k\}$ is an iid Gaussian sequence, and give the probability distribution of each W_k.

(b) Let the band B be $B = [-1/2T, 1/2T]$, and let $\phi_k(t) = (1/\sqrt{T})\mathrm{sinc}(\frac{t-kT}{T}), k \in \mathbb{Z}$. Interpret the result of part (a) in this case.

7.14. (Complex Gaussian vectors) (a) Give an example of a 2 dimensional complex r**v** $\boldsymbol{Z} = (Z_1, Z_2)$ where $Z_k \sim \mathcal{CN}(0,1)$ for $k = 1, 2$ and where $\boldsymbol{Z}$ has the same joint probability distribution as $e^{i\phi}\boldsymbol{Z}$ for all $\phi \in [0,\, 2\pi]$ but where $\boldsymbol{Z}$ is not jointly Gaussian and thus not circularly symmetric. Hint: Extend the idea in part (d) of Exercise 7.3.

(b) Suppose a complex random variable $Z = Z_{\mathrm{re}} + iZ_{\mathrm{im}}$ has the properties that Z_{re} and Z_{im} are individually Gaussian and that Z has the same probability density as $e^{i\phi}Z$ for all $\phi \in [0,\, 2\pi]$. Show that Z is complex circularly symmetric Gaussian.