Chapter 3

GAUSSIAN RANDOM VECTORS AND PROCESSES

3.1 Introduction

Poisson processes and Gaussian processes are similar in terms of their simplicity and beauty. When we first look at a new problem involving stochastic processes, we often start with insights from Poisson and/or Gaussian processes. Problems where queueing is a major factor tend to rely heavily on an understanding of Poisson processes, and those where noise is a major factor tend to rely heavily on Gaussian processes.

Poisson and Gaussian processes share the characteristic that the results arising from them are so simple, well known, and powerful that people often forget how much the results depend on assumptions that are rarely satisfied perfectly in practice. At the same time, these assumptions are often approximately satisfied, so the results, if used with insight and care, are often useful.

This chapter is aimed primarily at Gaussian processes, but starts with a study of Gaussian (normal[1]) random variables and vectors, These initial topics are both important in their own right and also essential to an understanding of Gaussian processes. The material here is essentially independent of that on Poisson processes in Chapter 2.

3.2 Gaussian random variables

A random variable (rv) W is defined to be a *normalized Gaussian rv* if it has the density

$$\mathsf{f}_W(w) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-w^2}{2}\right); \qquad \text{for all } w \in \mathbb{R}. \tag{3.1}$$

[1]Gaussian rv's are often called normal rv's. I prefer Gaussian, first because the corresponding processes are usually called Gaussian, second because Gaussian rv's (which have arbitrary means and variances) are often *normalized* to zero mean and unit variance, and third, because calling them normal gives the false impression that other rv's are abnormal.

Exercise 3.1 shows that $\mathsf{f}_W(w)$ integrates to 1 (i.e., it is a probability density), and that W has mean 0 and variance 1.

If we scale a normalized Gaussian rv W by a positive constant σ, *i.e.*, if we consider the rv $Z = \sigma W$, then the distribution functions of Z and W are related by $\mathsf{F}_Z(\sigma w) = \mathsf{F}_W(w)$. This means that the probability densities are related by $\sigma \mathsf{f}_Z(\sigma w) = \mathsf{f}_W(w)$. Thus the PDF of Z is given by

$$\mathsf{f}_Z(z) \;=\; \frac{1}{\sigma}\mathsf{f}_W\Big(\frac{z}{\sigma}\Big) \;=\; \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(\frac{-z^2}{2\sigma^2}\right). \tag{3.2}$$

Thus the PDF for Z is scaled horizontally by the factor σ, and then scaled vertically by $1/\sigma$ (see Figure 3.1). This scaling leaves the integral of the density unchanged with value 1 and scales the variance by σ^2. If we let σ approach 0, this density approaches an impulse, i.e., Z becomes the atomic rv for which $\Pr\{Z{=}0\} = 1$. For convenience in what follows, we use (3.2) as the density for Z for all $\sigma \geq 0$, with the above understanding about the $\sigma = 0$ case. A rv with the density in (3.2), for any $\sigma \geq 0$, is defined to be a *zero-mean Gaussian rv*. The values $\Pr\{|Z| > \sigma\} = .318$, $\Pr\{|Z| > 3\sigma\} = .0027$, and $\Pr\{|Z| > 5\sigma\} = 2.2 \cdot 10^{-12}$ give us a sense of how small the tails of the Gaussian distribution are.

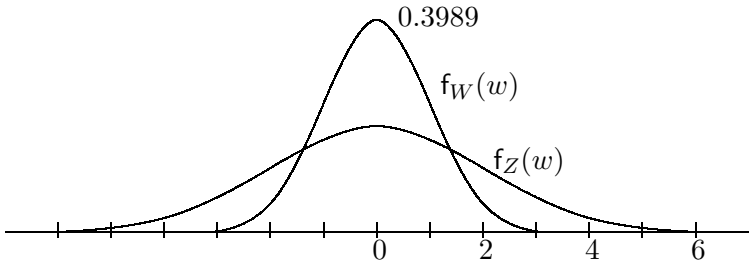


Figure 3.1: Graph of the PDF of a normalized Gaussian rv W (the taller curve) and of a zero-mean Gaussian rv Z with standard deviation 2 (the flatter curve).

If we shift Z by an arbitrary $\mu \in \mathbb{R}$ to $U = Z + \mu$, then the density shifts so as to be centered at $\mathsf{E}[U] = \mu$, and the density satisfies $\mathsf{f}_U(u) = \mathsf{f}_Z(u - \mu)$. Thus

$$\mathsf{f}_U(u) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(\frac{-(u-\mu)^2}{2\sigma^2}\right). \tag{3.3}$$

A random variable U with this density, for arbitrary μ and $\sigma \geq 0$, is defined to be a *Gaussian random variable* and is denoted $U \sim \mathcal{N}(\mu, \sigma^2)$.

The added generality of a mean often obscures formulas; we usually assume zero-mean rv's and random vectors (r**v**'s) and add means later if necessary. Recall that any rv U with a mean μ can be regarded as a constant μ plus the fluctuation, $U - \mu$, of U.

The moment generating function, $\mathsf{g}_Z(r)$, of a Gaussian rv $Z \sim \mathcal{N}(0, \sigma^2)$, can be calculated

as follows:

$$\begin{aligned}
\mathsf{g}_Z(r) &= \mathsf{E}\left[\exp(rZ)\right] = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}\exp(rz)\exp\left[\frac{-z^2}{2\sigma^2}\right]dz \\
&= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}\exp\left[\frac{-z^2+2\sigma^2rz-r^2\sigma^4}{2\sigma^2}+\frac{r^2\sigma^2}{2}\right]dz \qquad (3.4)\\
&= \exp\left[\frac{r^2\sigma^2}{2}\right]\left\{\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}\exp\left[\frac{-(z-r\sigma)^2}{2\sigma^2}\right]dz\right\} \qquad (3.5)\\
&= \exp\left[\frac{r^2\sigma^2}{2}\right]. \qquad (3.6)
\end{aligned}$$

We completed the square in the exponent in (3.4). We then recognized that the term in braces in (3.5) is the integral of a probability density and thus equal to 1.

Note that $\mathsf{g}_Z(r)$ exists for all real r, although it increases rapidly with $|r|$. As shown in Exercise 3.2, the moments for $Z \sim \mathcal{N}(0,\sigma^2)$, can be calculated from the MGF to be

$$\mathsf{E}\left[Z^{2k}\right] = \frac{(2k)!\sigma^{2k}}{k!\,2^k} = (2k-1)(2k-3)(2k-5)\ldots(3)(1)\sigma^{2k}. \qquad (3.7)$$

Thus, $\mathsf{E}\left[Z^4\right]=3\sigma^4$, $\mathsf{E}\left[Z^6\right]=15\sigma^6$, etc. The odd moments of Z are all zero since z^{2k+1} is an odd function of z and the Gaussian density is even.

For an arbitrary Gaussian rv $U \sim \mathcal{N}(\mu,\sigma^2)$, let $Z = U-\mu$, Then $Z \sim \mathcal{N}(0,\sigma^2)$ and $\mathsf{g}_U(r)$ is given by

$$\mathsf{g}_U(r) = \mathsf{E}\left[\exp(r(\mu+Z))\right] = e^{r\mu}\mathsf{E}\left[e^{rZ}\right] = \exp(r\mu + r^2\sigma^2/2). \qquad (3.8)$$

The characteristic function, $\mathsf{g}_Z(i\theta) = \mathsf{E}\left[e^{i\theta Z}\right]$ for $Z \sim \mathcal{N}(0,\sigma^2)$ and $i\theta$ imaginary can be shown to be (e.g., see Chap. 2.12 in [27]).

$$\mathsf{g}_Z(i\theta) = \exp\left[\frac{-\theta^2\sigma^2}{2}\right], \qquad (3.9)$$

The argument in (3.4) to (3.6) does not show this since the term in braces in (3.5) is not a probability density for r imaginary. As explained in Section 1.5.5, the characteristic function is useful first because it exists for all rv's and second because an inversion formula (essentially the Fourier transform) exists to uniquely find the distribution of a rv from its characteristic function.

3.3 Gaussian random vectors

An $n \times \ell$ *matrix* $[A]$ is an array of $n\ell$ elements arranged in n rows and ℓ columns; A_{jk} denotes the k^{th} element in the j^{th} row. Unless specified to the contrary, the elements are real numbers. The *transpose* $[A^{\mathsf{T}}]$ of an $n\times\ell$ matrix $[A]$ is an $\ell\times n$ matrix $[B]$ with $B_{kj} = A_{jk}$

for all j,k. A matrix is *square* if $n=\ell$ and a square matrix $[A]$ is *symmetric* if $[A]=[A]^{\mathsf{T}}$. If $[A]$ and $[B]$ are each $n\times\ell$ matrices, $[A]+[B]$ is an $n\times\ell$ matrix $[C]$ with $C_{jk}=A_{jk}+B_{jk}$ for all j,k. If $[A]$ is $n\times\ell$ and $[B]$ is $\ell\times r$, the matrix $[A][B]$ is an $n\times r$ matrix $[C]$ with elements $C_{jk}=\sum_i A_{ji}B_{ik}$. A *vector* (or *column vector*) of dimension n is an n by 1 matrix and a *row vector* of dimension n is a 1 by n matrix. Since the transpose of a vector is a row vector, we denote a vector $\boldsymbol{a}$ as $(a_1,\ldots,a_n)^{\mathsf{T}}$. Note that if $\boldsymbol{a}$ is a (column) vector of dimension n, then $\boldsymbol{a}\boldsymbol{a}^{\mathsf{T}}$ is an $n\times n$ matrix whereas $\boldsymbol{a}^{\mathsf{T}}\boldsymbol{a}$ is a number. The reader is expected to be familiar with these vector and matrix manipulations.

The covariance matrix, $[K]$ (if it exists) of an arbitrary zero-mean n-r**v** $\boldsymbol{Z}=(Z_1,\ldots,Z_n)^{\mathsf{T}}$ is the matrix whose components are $K_{jk}=\mathsf{E}[Z_jZ_k]$. For a non-zero-mean n-r**v** $\boldsymbol{U}$, let $\boldsymbol{U}=\boldsymbol{m}+\boldsymbol{Z}$ where $\boldsymbol{m}=\mathsf{E}[\boldsymbol{U}]$ and $\boldsymbol{Z}=\boldsymbol{U}-\boldsymbol{m}$ is the fluctuation of $\boldsymbol{U}$. The covariance matrix $[K]$ of $\boldsymbol{U}$ is defined to be the same as the covariance matrix of the fluctuation $\boldsymbol{Z}$, *i.e.*, $K_{jk}=\mathsf{E}[Z_jZ_k]=\mathsf{E}[(U_j-m_j)(U_k-m_k)]$. It can be seen that if an $n\times n$ covariance matrix $[K]$ exists, it must be symmetric, *i.e.*, it must satisfy $K_{jk}=K_{kj}$ for $1\le j,k\le n$.

3.3.1 Generating functions of Gaussian random vectors

The *moment generating function* (MGF) of an n-r**v** $\boldsymbol{Z}$ is defined as $\mathsf{g}_{\boldsymbol{Z}}(\boldsymbol{r})=\mathsf{E}[\exp(\boldsymbol{r}^{\mathsf{T}}\boldsymbol{Z})]$ where $\boldsymbol{r}=(r_1,\ldots,r_n)^{\mathsf{T}}$ is an n-dimensional real vector. The n-dimensional MGF might not exist for all $\boldsymbol{r}$ (just as the one-dimensional MGF discussed in Section 1.5.5 need not exist everywhere). As we will soon see, however, the MGF exists everywhere for Gaussian n-r**v**'s.

The characteristic function, $\mathsf{g}_{\boldsymbol{Z}}(i\boldsymbol{\theta})=\mathsf{E}\left[e^{i\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Z}}\right]$, of an n-r**v** $\boldsymbol{Z}$, where $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_n)^{\mathsf{T}}$ is a real n-vector, is equally important. As in the one-dimensional case, the characteristic function always exists for all real $\boldsymbol{\theta}$ and all n-r**v** $\boldsymbol{Z}$. In addition, there is a uniqueness theorem[2] stating that the characteristic function of an n-r**v** $\boldsymbol{Z}$ uniquely specifies the joint distribution of $\boldsymbol{Z}$.

If the components of an n-r**v** are independent and identically distributed (IID), we call the vector an IID n-r**v**.

3.3.2 IID normalized Gaussian random vectors

An example that will become familiar is that of an IID n-r**v** $\boldsymbol{W}$ where each component W_j, $1\le j\le n$, is normalized Gaussian, $W_j\sim\mathcal{N}(0,1)$. By taking the product of n densities as given in (3.1), the joint density of $\boldsymbol{W}=(W_1,W_2,\ldots,W_n)^{\mathsf{T}}$ is

$$\mathsf{f}_{\boldsymbol{W}}(\boldsymbol{w})=\frac{1}{(2\pi)^{n/2}}\exp\left(\frac{-w_1^2-w_2^2-\cdots-w_n^2}{2}\right)=\frac{1}{(2\pi)^{n/2}}\exp\left(\frac{-\boldsymbol{w}^{\mathsf{T}}\boldsymbol{w}}{2}\right). \tag{3.10}$$

[2]See Shiryaev, [27], for a proof in the one-dimensional case and an exercise providing the extension to the n-dimensional case. It appears that the exercise is a relatively straightforward extension of the proof for one dimension, but the one-dimensional proof is measure theoretic and by no means trivial. The reader can get an engineering understanding of this uniqueness theorem by viewing the characteristic function and joint probability density essentially as n-dimensional Fourier transforms of each other.

The joint density of $\boldsymbol{W}$ at a sample value $\boldsymbol{w}$ depends only on the squared distance $\boldsymbol{w}^\mathsf{T}\boldsymbol{w}$ of the sample value $\boldsymbol{w}$ from the origin. That is, $\mathsf{f}_{\boldsymbol{W}}(\boldsymbol{w})$ is spherically symmetric around the origin, and points of equal probability density lie on concentric spheres around the origin (see Figure 3.2).

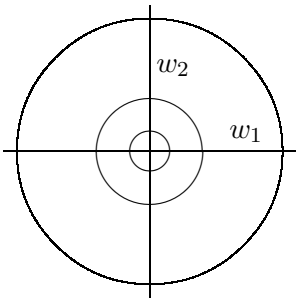


Figure 3.2: Equi-probability contours for an IID Gaussian 2-r$\mathbf{v}$.

The moment generating function of $\boldsymbol{W}$ is easily calculated as follows:

$$\begin{aligned}\mathsf{g}_{\boldsymbol{W}}(\boldsymbol{r}) &= \mathsf{E}\left[\exp \boldsymbol{r}^\mathsf{T}\boldsymbol{W}\right] = \mathsf{E}\left[\exp(r_1W_1 + \cdots + r_nW_n\right] = \mathsf{E}\left[\prod_j \exp(r_jW_j)\right] \\ &= \prod_j \mathsf{E}\left[\exp(r_jW_j)\right] = \prod_j \exp\left(\frac{r_j^2}{2}\right) = \exp\left[\frac{\boldsymbol{r}^\mathsf{T}\boldsymbol{r}}{2}\right]. \qquad (3.11)\end{aligned}$$

The interchange of the expectation with the product above is justified because, first, the rv's W_j (and thus the rv's $\exp(r_jW_j)$) are independent, and, second, the expectation of a product of independent rv's is equal to the product of the expected values. The MGF of each W_j then follows from (3.6). The characteristic function of $\boldsymbol{W}$ is similarly calculated using (3.9),

$$\mathsf{g}_{\boldsymbol{W}}(i\boldsymbol{\theta}) = \exp\left[\frac{-\boldsymbol{\theta}^\mathsf{T}\boldsymbol{\theta}}{2}\right], \qquad (3.12)$$

Next consider rv's that are linear combinations of $W_1, \ldots, W_n$, *i.e.*, rv's of the form $Z = \boldsymbol{a}^\mathsf{T}\boldsymbol{W} = a_1W_1 + \cdots + a_nW_n$. By convolving the densities of the components a_jW_j, it is shown in Exercise 3.4 that Z is Gaussian, $Z \sim \mathcal{N}(0, \sigma^2)$ where $\sigma^2 = \sum_{j=1}^n a_j^2$, *i.e.*, $Z \sim \mathcal{N}(0, \sum_j a_j^2)$.

3.3.3 Jointly-Gaussian random vectors

We now go on to define the general class of zero-mean jointly-Gaussian n-r$\mathbf{v}$'s.

Definition 3.3.1. *$\{Z_1, Z_2, \ldots, Z_n\}$ is a set of jointly-Gaussian zero-mean rv's, and $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\mathsf{T}$ is a Gaussian zero-mean n-r$\mathbf{v}$, if, for some finite set of IID $\mathcal{N}(0,1)$ rv's,*

$W_1, \ldots, W_m$, each Z_j can be expressed as

$$Z_j = \sum_{\ell=1}^{m} a_{j\ell} W_\ell \qquad \text{i.e., } \boldsymbol{Z} = [A]\,\boldsymbol{W} \tag{3.13}$$

where $\{a_{j\ell}, 1 \le j \le n, 1 \le \ell \le m,\}$ is a given array of real numbers. More generally, $\boldsymbol{U} = (U_1, \ldots, U_n)^\mathsf{T}$ is a Gaussian n-r$\mathbf{v}$ if $\boldsymbol{U} = \boldsymbol{Z} + \boldsymbol{\mu}$, where $\boldsymbol{Z}$ is a zero-mean Gaussian n-r$\mathbf{v}$ and $\boldsymbol{\mu}$ is a real n vector.

We already saw that each linear combination of IID $\mathcal{N}(0,1)$ rv's is Gaussian. This definition defines $Z_1, \ldots, Z_n$ to be jointly Gaussian if all of them are linear combinations of a common set of IID normalized Gaussian rv's. This definition might not appear to restrict jointly-Gaussian rv's far beyond being individually Gaussian, but several examples later show that being jointly Gaussian in fact implies a great deal more than being individually Gaussian. We will also see that the remarkable properties of jointly Gaussian rv's depend very heavily on this linearity property.

Note from the definition that a Gaussian n-r$\mathbf{v}$ is a vector whose components are *jointly* Gaussian rather than only individually Gaussian. When we define Gaussian processes later, the requirement that the components be jointly Gaussian will again be present.

The intuition behind jointly-Gaussian rv's is that in many physical situations there are multiple rv's each of which is a linear combination of a common large set of small essentially independent rv's. The central limit theorem indicates that each such sum can be approximated by a Gaussian rv, and, more to the point here, linear combinations of those sums are also approximately Gaussian. For example, when a broadband noise waveform is passed through a narrowband linear filter, the output at any given time is usually well approximated as the sum of a large set of essentially independent rv's. The outputs at different times are different linear combinations of the same set of underlying small, essentially independent, rv's. Thus we would expect a set of outputs at different times to be jointly Gaussian according to the above definition.

The following simple theorem begins the process of specifying the properties of jointly-Gaussian rv's. These results are given for zero-mean rv's since the extension to non-zero mean is obvious.

Theorem 3.3.1. *Let $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\mathsf{T}$ be a zero-mean Gaussian n-r$\mathbf{v}$. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_k)^\mathsf{T}$ be a k-r$\mathbf{v}$ satisfying $\boldsymbol{Y} = [B]\boldsymbol{Z}$. Then $\boldsymbol{Y}$ is a zero-mean Gaussian k-r$\mathbf{v}$.*

Proof: Since $\boldsymbol{Z}$ is a zero-mean Gaussian n-r$\mathbf{v}$, it can be represented as $\boldsymbol{Z} = [A]\,\boldsymbol{W}$ where the components of $\boldsymbol{W}$ are IID and $\mathcal{N}(0,1)$. Thus $\boldsymbol{Y} = [B][A]\,\boldsymbol{W}$. Since $[B][A]$ is a matrix, $\boldsymbol{Y}$ is a zero-mean Gaussian k-r$\mathbf{v}$. □

For $k = 1$, this becomes the trivial but important corollary:

Corollary 3.3.1. *Let $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\mathsf{T}$ be a zero-mean Gaussian n-r$\mathbf{v}$. Then for any real n-vector $\boldsymbol{a} = (a_1, \ldots, a_n)^\mathsf{T}$, the linear combination $\boldsymbol{a}^\mathsf{T}\boldsymbol{Z}$ is a zero-mean Gaussian rv.*

We next give an example of two rv's, Z_1, Z_2 that are each zero-mean Gaussian but for which $Z_1 + Z_2$ is not Gaussian. From the theorem, then, Z_1 and Z_2 are not jointly Gaussian and the 2-r$\boldsymbol{v}$ $\boldsymbol{Z} = (Z_1, Z_2)^\mathsf{T}$ is not a Gaussian vector. This is the first of a number of later examples of rv's that are marginally Gaussian but not jointly Gaussian.

Example 3.3.1. Let $Z_1 \sim \mathcal{N}(0,1)$, and let X be independent of Z_1 and take equiprobable values ± 1. Let $Z_2 = Z_1 X_1$. Then $Z_2 \sim \mathcal{N}(0,1)$ and $\mathsf{E}[Z_1 Z_2] = 0$. The joint probability density, $\mathsf{f}_{Z_1 Z_2}(z_1, z_2)$, however, is impulsive on the diagonals where $z_2 = \pm z_1$ and is zero elsewhere. Then $Z_1 + Z_2$ can not be Gaussian, since it takes on the value 0 with probability one half.

This example also shows the falseness of the frequently heard statement that uncorrelated Gaussian rv's are independent. The correct statement, as we see later, is that uncorrelated *jointly* Gaussian rv's are independent.

The next theorem specifies the moment generating function (MGF) of an arbitrary zero-mean Gaussian n-r$\mathbf{v}$ $\boldsymbol{Z}$. The important feature is that the MGF depends only on the covariance function $[K]$. Essentially, as developed later, Z is characterized by a probability density that depends only on $[K]$.

Theorem 3.3.2. *Let $\boldsymbol{Z}$ be a zero-mean Gaussian n-r$\mathbf{v}$ with covariance matrix $[K]$. Then the MGF, $\mathsf{g}_{\boldsymbol{Z}}(\boldsymbol{r}) = \mathsf{E}[\exp(\boldsymbol{r}^T \boldsymbol{Z})]$ and the characteristic function $\mathsf{g}_{\boldsymbol{Z}}(i\boldsymbol{\theta}) = \mathsf{E}[\exp(i\boldsymbol{\theta}^T \boldsymbol{Z})]$ are given by*

$$\mathsf{g}_{\boldsymbol{Z}}(\boldsymbol{r}) = \exp\left[\frac{\boldsymbol{r}^T[K]\,\boldsymbol{r}}{2}\right]; \qquad \mathsf{g}_{\boldsymbol{Z}}(i\boldsymbol{\theta}) = \exp\left[\frac{-\boldsymbol{\theta}^T[K]\,\boldsymbol{\theta}}{2}\right]. \tag{3.14}$$

Proof: For any given real n-vector $\boldsymbol{r} = (r_1, \ldots, r_n)^\mathsf{T}$, let $X = \boldsymbol{r}^\mathsf{T}\boldsymbol{Z}$. Then from Corollary 3.3.1, X is zero-mean Gaussian and from (3.6),

$$\mathsf{g}_X(s) = \mathsf{E}[\exp(sX)] = \exp(\sigma_X^2 s^2/2). \tag{3.15}$$

Thus for the given $\boldsymbol{r}$,

$$\mathsf{g}_{\boldsymbol{Z}}(r) = \mathsf{E}[\exp(\boldsymbol{r}^\mathsf{T}\boldsymbol{Z})] = \mathsf{E}[\exp(X)] = \exp(\sigma_X^2/2), \tag{3.16}$$

where the last step uses (3.15) with $s = 1$. Finally, since $X = \boldsymbol{r}^\mathsf{T}\boldsymbol{Z}$, we have

$$\sigma_X^2 = \mathsf{E}\left[|\boldsymbol{r}^\mathsf{T}\boldsymbol{Z}|^2\right] = \mathsf{E}[\boldsymbol{r}^\mathsf{T}\boldsymbol{Z}\boldsymbol{Z}^\mathsf{T}\boldsymbol{r}] = \boldsymbol{r}^\mathsf{T}\mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^\mathsf{T}]\,\boldsymbol{r} = \boldsymbol{r}^\mathsf{T}[K]\boldsymbol{r}. \tag{3.17}$$

Substituting (3.17) into (3.16), yields (3.14). The proof is the same for the characteristic function except (3.9) is used in place of (3.6). □

Since the characteristic function of an n-r$\mathbf{v}$ uniquely specifies the CDF, this theorem also shows that the joint CDF of a zero-mean Gaussian n-r$\mathbf{v}$ is completely determined by the covariance function. To make this story complete, we will show later that for any possible covariance function for any n-r$\mathbf{v}$, there is a corresponding zero-mean Gaussian n-r$\mathbf{v}$ with that covariance.

As a slight generaliization of (3.14), let $\boldsymbol{U}$ be a Gaussian n-r$\mathbf{v}$ with an arbitrary mean, *i.e.*, $\boldsymbol{U} = \boldsymbol{m} + \boldsymbol{Z}$ where the n-vector $\boldsymbol{m}$ is the mean of $\boldsymbol{U}$ and the zero-mean Gaussian n-r$\mathbf{v}$ $\boldsymbol{Z}$ is the fluctuation of $\boldsymbol{U}$. Note that the covariance matrix $[K]$ of $\boldsymbol{U}$ is the same as that for $\boldsymbol{Z}$, yielding

$$\mathsf{g}_U(\boldsymbol{r}) = \exp\left(\boldsymbol{r}^\mathsf{T}\boldsymbol{m} + \frac{\boldsymbol{r}^\mathsf{T}[K]\,\boldsymbol{r}}{2}\right); \qquad \mathsf{g}_U(i\boldsymbol{\theta}) = \exp\left[i\boldsymbol{\theta}^\mathsf{T}\boldsymbol{m} - \frac{\boldsymbol{\theta}^\mathsf{T}[K]\,\boldsymbol{\theta}}{2}\right]. \tag{3.18}$$

We denote a Gaussian n-r$\mathbf{v}$ $\boldsymbol{U}$ of mean $\boldsymbol{m}$ and covariance $[K]$ as $\boldsymbol{U} \sim \mathcal{N}(\boldsymbol{m}, [K])$.

3.3.4 Joint probability density for Gaussian n-rv's (special case)

A zero-mean Gaussian n-r$\mathbf{v}$, by definition, has the form $\boldsymbol{Z} = [A]\,\boldsymbol{W}$ where $\boldsymbol{W}$ is $\mathcal{N}(0, [I_n])$. In this section we look at the special case where $[A]$ is $n \times n$ and non-singular. The covariance matrix of $\boldsymbol{Z}$ is then

$$\begin{aligned}[K] &= \mathsf{E}\,[\boldsymbol{Z}\boldsymbol{Z}^\mathsf{T}] = \mathsf{E}\,[[A]\,\boldsymbol{W}\,\boldsymbol{W}^\mathsf{T}[A]^\mathsf{T}] \\ &= [A]\mathsf{E}\,[\boldsymbol{W}\,\boldsymbol{W}^\mathsf{T}]\,[A]^\mathsf{T} = [A][A]^\mathsf{T}\end{aligned} \tag{3.19}$$

since $\mathsf{E}\,[WW^\mathsf{T}]$ is the identity matrix, $[I_n]$.

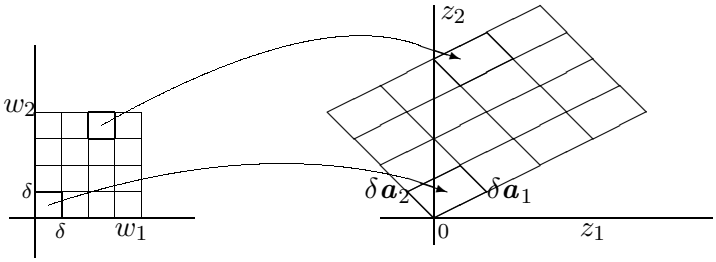
To find $\mathsf{f}_{\boldsymbol{Z}}(\boldsymbol{z})$ in this case, we first consider the transformation of real-valued vectors, $\boldsymbol{z} = [A]\boldsymbol{w}$. Let $\boldsymbol{e}_j$ be the jth unit vector (i.e., the vector whose jth component is 1 and whose other components are 0). Then $[A]\boldsymbol{e}_j = \boldsymbol{a}_j$, where $\boldsymbol{a}_j$ is the jth column of $[A]$. Thus, $\boldsymbol{z} = [A]\boldsymbol{w}$ transforms each unit vector $\boldsymbol{e}_j$ into the column $\boldsymbol{a}_j$ of [A]. For $n = 2$, Figure 3.3 shows how this transformation carries each vector $\boldsymbol{w}$ into the vector $\boldsymbol{z} = [A]\boldsymbol{w}$. Note that an incremental square, δ on a side is carried into an parallelogram with corners $\boldsymbol{0}$, $\boldsymbol{a}_1\delta$, $\boldsymbol{a}_2\delta$, and $(\boldsymbol{a}_1 + \boldsymbol{a}_2)\delta$.

For an arbitrary number of dimensions, the unit cube in the $\boldsymbol{w}$ space is the set of points $\{\boldsymbol{w} : 0 \leq w_j \leq 1; 1 \leq j \leq n\}$ There are 2^n corners of the unit cube, and each is some 0/1 combination of the unit vectors, i.e., each has the form $\boldsymbol{e}_{j_1} + \boldsymbol{e}_{j_2} + \cdots + \boldsymbol{e}_{j_k}$. The transformation $[A]\boldsymbol{w}$ carries the unit cube into a parallelepiped, where each corner of the cube, $\boldsymbol{e}_{j_1} + \boldsymbol{e}_{j_2} + \cdots + \boldsymbol{e}_{j_k}$, is carried into a corresponding corner $\boldsymbol{a}_{j_1} + \boldsymbol{a}_{j_2} + \cdots + \boldsymbol{a}_{j_n}$ of the parallelepiped. One of the most interesting and geometrically meaningful properties of the determinant, $\det[A]$, of a square real matrix $[A]$ is that the *magnitude* of that determinant, $|\det[A]|$, is equal to the volume of that parallelepiped (see Strang, [28]). If $\det[A] = 0$, i.e., if $[A]$ is singular, then the n-dimensional unit cube in the $\boldsymbol{w}$ space is transformed into a lower-dimensional parallelepiped whose volume (as a region of n-dimensional space) is 0. This case is considered in Section 3.4.4.

Now let $\boldsymbol{z}$ be a sample value of $\boldsymbol{Z}$, and let $\boldsymbol{w} = [A]^{-1}\boldsymbol{z}$ be the corresponding sample value of $\boldsymbol{W}$. The joint density at $\boldsymbol{z}$ must satisfy

$$\mathsf{f}_{\boldsymbol{Z}}(z)|d\boldsymbol{z}| = \mathsf{f}_{\boldsymbol{W}}(\boldsymbol{w})|d\boldsymbol{w}|, \tag{3.20}$$

where $|d\boldsymbol{w}|$ is the volume of an incremental cube with dimension $\delta = dw_j$ on each side, and $|d\boldsymbol{z}|$ is the volume of that incremental cube transformed by $[A]$. Thus $|d\boldsymbol{w}| = \delta^n$ and



$$[A] = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$$

Figure 3.3: Example illustrating how $\boldsymbol{z} = [A]\boldsymbol{w}$ maps cubes into parallelepipeds. Let $z_1 = 2w_1 - w_2$ and $z_2 = w_1 + w_2$. Thus $\boldsymbol{w} = (1,0)^\mathsf{T}$ transforms to $\boldsymbol{a}_1 = (2,1)^\mathsf{T}$ and $\boldsymbol{w} = (0,1)^\mathsf{T}$ transforms to $\boldsymbol{a}_2 = (-1,1)^\mathsf{T}$. The lower left square in the first figure is the set $\{(w_1, w_2) : 0 \le w_1 \le \delta;\ 0 \le w_2 \le \delta\}$. This square is transformed into the parallelogram with sides $\delta\boldsymbol{a}_1$ and $\delta\boldsymbol{a}_2$. The figure also shows how the w_1, w_2 space can be quantized into adjoining squares, which map into corresponding adjoining parallelograms in the z_1, z_2 space.

$|d\boldsymbol{z}| = \delta^n |\det[A]|$ so that $|d\boldsymbol{z}|/|d\boldsymbol{w}| = |\det[A]|$. Using this in (3.20), and using (3.10) for $\mathsf{f}_{\boldsymbol{W}}(\boldsymbol{w})$, we see that the density of a jointly-Gaussian r**v** $\boldsymbol{Z} = [A]\,\boldsymbol{W}$ is

$$\mathsf{f}_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{\exp\left(-\frac{1}{2}\boldsymbol{z}^\mathsf{T}[A^{-1}]^\mathsf{T}[A^{-1}]\boldsymbol{z}\right)}{(2\pi)^{n/2}|\det[A]|}. \tag{3.21}$$

From (3.19), we have $[K] = [AA^\mathsf{T}]$, so $[K^{-1}] = [A^{-1}]^\mathsf{T}[A^{-1}]$. Also, for arbitrary square real matrices $[A]$ and $[B]$, $\det[AB] = \det[A]\det[B]$ and $\det[A] = \det[A^\mathsf{T}]$. Thus $\det[K] = \det[A]\det[A^\mathsf{T}] = \big(\det[A]\big)^2 > 0$ and (3.21) becomes

$$\mathsf{f}_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{\exp\left(-\frac{1}{2}\boldsymbol{z}^\mathsf{T}[K^{-1}]\boldsymbol{z}\right)}{(2\pi)^{n/2}\sqrt{\det[K]}}. \tag{3.22}$$

Note that this density depends only on $[K]$, so the density depends on $[A]$ only through $[A][A^\mathsf{T}] = [K]$. This is not surprising, since we saw that the characteristic function of $\boldsymbol{Z}$ also depended only on the covariance matrix of $\boldsymbol{Z}$.

The expression in (3.22) is quite beautiful. It arises, first, because the density of $\boldsymbol{W}$ is spherically symmetric, and second, because $\boldsymbol{Z}$ is a linear transformation of $\boldsymbol{W}$. We show later that this density applies to any zero-mean Gaussian n-r**v** for which the covariance is a non-singular matrix $[K]$.

Example 3.3.2. Consider (3.22) for the 2-dimensional case. Let $\mathsf{E}\left[Z_1^2\right] = \sigma_1^2$, $\mathsf{E}\left[Z_2^2\right] = \sigma_2^2$ and $\mathsf{E}\left[Z_1 Z_2\right] = k_{12}$. Define the *normalized covariance*, ρ, as $k_{12}/(\sigma_1\sigma_2)$. Then $\det[K] = \sigma_1^2\sigma_2^2 - k_{12}^2 = \sigma_1^2\sigma_2^2(1-\rho^2)$. For $[A]$ to be non-singular, we need $\det[K] = \big(\det[A]\big)^2 > 0$, so we need $|\rho| < 1$. We then have

$$[K]^{-1} = \frac{1}{\sigma_1^2\sigma_2^2 - k_{12}^2}\begin{bmatrix} \sigma_2^2 & -k_{12} \\ -k_{12} & \sigma_1^2 \end{bmatrix} = \frac{1}{1-\rho^2}\begin{bmatrix} 1/\sigma_1^2 & -\rho/(\sigma_1\sigma_2) \\ -\rho/(\sigma_1\sigma_2) & 1/\sigma_2^2 \end{bmatrix}.$$

$$\begin{aligned} \mathsf{f}_{\boldsymbol{Z}}(\boldsymbol{z}) &= \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2 - k_{12}^2}} \exp\left(\frac{-z_1^2\sigma_2^2 + 2z_1z_2k_{12} - z_2^2\sigma_1^2}{2(\sigma_1^2\sigma_2^2 - k_{12}^2)}\right) \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{-\frac{z_1^2}{\sigma_1^2} + \frac{2\rho z_1 z_2}{\sigma_1\sigma_2} - \frac{z_2^2}{\sigma_2^2}}{2(1-\rho^2)}\right). \end{aligned} \tag{3.23}$$

The exponent in (3.23) is a quadratic in z_1, z_2 and from this it can be deduced that the equiprobability contours for $\boldsymbol{Z}$ are concentric ellipses. This will become clearer (both for $n = 2$ and $n > 2$) in Section 3.4.4.

Perhaps the more important lesson from (3.23), however, is that vector notation simplifies such equations considerably even for $n = 2$. We must learn to reason directly from the vector equations and use standard computer programs for required calculations.

For completeness, let $\boldsymbol{U} = \boldsymbol{\mu} + \boldsymbol{Z}$ where $\boldsymbol{\mu} = \mathsf{E}[U]$ and $\boldsymbol{Z}$ is a zero-mean Gaussian n-r**v** with the density in (3.21). Then the density of $\boldsymbol{U}$ is given by

$$\mathsf{f}_{U}(\boldsymbol{u}) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{\mu})^{\mathsf{T}}[K^{-1}](\boldsymbol{u}-\boldsymbol{\mu})\right)}{(2\pi)^{n/2}\sqrt{\det[K]}}, \tag{3.24}$$

where $[K]$ is the covariance matrix of both $\boldsymbol{U}$ and $\boldsymbol{Z}$.

3.4 Properties of covariance matrices

In this section, we summarize some simple properties of covariance matrices that will be used frequently in what follows. We start with symmetric matrices before considering covariance matrices.

3.4.1 Symmetric matrices

A number λ is said to be an eigenvalue of an $n \times n$ matrix, $[B]$, if there is a non-zero n-vector $\boldsymbol{q}$ such that $[B]\boldsymbol{q} = \lambda\boldsymbol{q}$, *i.e.*, such that $[B - \lambda I]\boldsymbol{q} = 0$. In other words, λ is an eigenvalue of $[B]$ if $[B - \lambda I]$ is singular. We are interested only in real matrices here, but the eigenvalues and eigenvectors might be complex. The values of λ that are eigenvalues of $[B]$ are the solutions to the characteristic equation, $\det[B - \lambda I] = 0$, *i.e.*, they are the roots of $\det[B - \lambda I]$. As a function of λ, $\det[B - \lambda I]$ is a polynomial of degree n. From the fundamental theorem of algebra, it therefore has n roots (possibly complex and not necessarily distinct).

If $[B]$ is symmetric, then the eigenvalues are all real.[3] Also, the eigenvectors can all be chosen to be real. In addition, eigenvectors of distinct eigenvalues must be orthogonal, and if an eigenvalue λ has multiplicity ℓ (*i.e.*, $\det[B - \lambda I]$ as a polynomial in λ has an ℓth order root at λ), then ℓ orthogonal eigenvectors can be chosen for that λ.

[3]See Strang [28] or other linear algebra texts for a derivation of these standard results.

What this means is that we can list the eigenvalues as $\lambda_1, \lambda_2, \ldots, \lambda_n$ (where each distinct eigenvalue is repeated according to its multiplicity). To each eigenvalue λ_j, we can associate an eigenvector $\boldsymbol{q}_j$ where $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are orthogonal. Finally, each eigenvector can be normalized so that $\boldsymbol{q}_j{}^\mathsf{T}\boldsymbol{q}_k = \delta_{jk}$ where $\delta_{jk} = 1$ for $j = k$ and $\delta_{jk} = 0$ otherwise; the set $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$ is then called orthonormal.

If we take the resulting n equations, $[B]\boldsymbol{q}_j = \lambda_j \boldsymbol{q}_j$ and combine them into a matrix equation, we get

$$[BQ] = [Q\Lambda], \tag{3.25}$$

where $[Q]$ is the $n \times n$ matrix whose columns are the orthonormal vectors $\boldsymbol{q}_1, \ldots \boldsymbol{q}_n$ and where $[\Lambda]$ is the $n \times n$ diagonal matrix whose diagonal elements are $\lambda_1, \ldots, \lambda_n$.

The matrix $[Q]$ is called an orthonormal or orthogonal matrix and, as we have seen, has the property that its columns are orthonormal. The matrix $[Q]^\mathsf{T}$ then has the rows $\boldsymbol{q}_j^\mathsf{T}$ for $1 \le j \le n$. If we multiply $[Q]^\mathsf{T}$ by $[Q]$, we see that the j,k element of the product is $\boldsymbol{q}_j{}^\mathsf{T}\boldsymbol{q}_k = \delta_{jk}$. Thus $[Q^\mathsf{T}Q] = [I]$ and $[Q^\mathsf{T}]$ is the inverse, $[Q^{-1}]$, of $[Q]$. Finally, since $[QQ^{-1}] = [I] = [QQ^\mathsf{T}]$, we see that the rows of Q are also orthonormal. This can be summarized in the following theorem:

Theorem 3.4.1. *Let $[B]$ be a real symmetric matrix and let $[\Lambda]$ be the diagonal matrix whose diagonal elements $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $[B]$, repeated according to multiplicity.. Then a set of orthonormal eigenvectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ can be chosen so that $[B]\boldsymbol{q}_j = \lambda_j \boldsymbol{q}_j$ for $1 \le j \le n$. The matrix $[Q]$ with orthonormal columns $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ satisfies (3.25). Also $[Q^\mathsf{T}] = [Q^{-1}]$ and the rows of $[Q]$ are orthonormal. Finally $[B]$ and $[Q]$ satisfy*

$$[B] = [Q\Lambda Q^{-1}]; \qquad [Q^{-1}] = [Q^\mathsf{T}] \tag{3.26}$$

Proof: The only new statement is the initial part of (3.26), which follows from (3.25) by post-multiplying both sides by $[Q^{-1}]$. □

3.4.2 Positive definite matrices and covariance matrices

Definition 3.4.1. *A real $n \times n$ matrix $[K]$ is positive definite if it is symmetric and if $\boldsymbol{b}^\mathsf{T}[K]\boldsymbol{b} > 0$ for all real n-vectors $\boldsymbol{b} \neq 0$. It is positive semi-definite[4] if $\boldsymbol{b}^\mathsf{T}[K]\boldsymbol{b} \ge 0$. It is a covariance matrix if there is a zero-mean n-rv $\boldsymbol{Z}$ such that $[K] = \mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^\mathsf{T}]$.*

We will see shortly that the class of positive semi-definite matrices is the same as the class of covariance matrices and that the class of positive definite matrices is the same as the class of non-singular covariance matrices. First we develop some useful properties of positive (semi-) definite matrices.

[4]Positive semi-definite is sometimes referred to as nonnegative definite, which is more transparent but less common.

Theorem 3.4.2. *A symmetric matrix $[K]$ is positive definite[5] if and only if each eigenvalue of $[K]$ is positive. It is positive semi-definite if and only if each eigenvalue is nonnegative.*

Proof: Assume that $[K]$ is positive definite. It is symmetric by the definition of positive definiteness, so for each eigenvalue λ_j of $[K]$, we can select a real normalized eigenvector $\boldsymbol{q}_j$ as a vector $\boldsymbol{b}$ in Definition 3.4.1. Then

$$0 < \boldsymbol{q}_j^{\mathsf{T}}[K]\boldsymbol{q}_j = \lambda_j \boldsymbol{q}_j^{\mathsf{T}}\boldsymbol{q}_j = \lambda_j,$$

so each eigenvalue is positive. To go the other way, assume that each $\lambda_j > 0$ and use the expansion of (3.26) with $[Q^{-1}] = [Q^{\mathsf{T}}]$. Then for any real $\boldsymbol{b} \neq 0$,

$$\boldsymbol{b}^{\mathsf{T}}[K]\boldsymbol{b} = \boldsymbol{b}^{\mathsf{T}}[Q\Lambda Q^{\mathsf{T}}]\boldsymbol{b} = \boldsymbol{c}^{\mathsf{T}}[\Lambda]\boldsymbol{c} \qquad \text{where } \boldsymbol{c} = [Q^{\mathsf{T}}]\boldsymbol{b}.$$

Now $[\Lambda]\boldsymbol{c}$ is a vector with components $\lambda_j c_j$. Thus $\boldsymbol{c}^{\mathsf{T}}[\Lambda]\boldsymbol{c} = \sum_j \lambda_j c_j^2$. Since each c_j is real, $c_j^2 \geq 0$ and thus $c_j^2\lambda_j \geq 0$. Since $\boldsymbol{c} \neq 0$, $c_j \neq 0$ for at least one j and thus $\lambda_j c_j^2 > 0$ for at least one j, so $\boldsymbol{c}^{\mathsf{T}}[\Lambda]\boldsymbol{c} > 0$. The proof for the positive semi-definite case follows by replacing the strict inequalitites above with non-strict inequalities. □

Theorem 3.4.3. *If $[K] = [AA^{\mathsf{T}}]$ for some real $n \times n$ matrix $[A]$, then $[K]$ is positive semi-definite. If $[A]$ is also non-singular, then $[K]$ is positive definite.*

Proof: For the hypothesized $[A]$ and any real n-vector $\boldsymbol{b}$,

$$\boldsymbol{b}^{\mathsf{T}}[K]\boldsymbol{b} = \boldsymbol{b}^{\mathsf{T}}[AA^{\mathsf{T}}]\boldsymbol{b} = \boldsymbol{c}^{\mathsf{T}}\boldsymbol{c} \geq 0 \qquad \text{where } \boldsymbol{c} = [A^{\mathsf{T}}]\boldsymbol{b}.$$

Thus $[K]$ is positive semi-definite. If $[A]$ is non-singular, then $\boldsymbol{c} \neq 0$ if $\boldsymbol{b} \neq 0$. Thus $\boldsymbol{c}^{\mathsf{T}}\boldsymbol{c} > 0$ for $\boldsymbol{b} \neq 0$ and $[K]$ is positive definite. □

A converse can be established showing that if $[K]$ is positive (semi-)definite, then an $[A]$ exists such that $[K] = [A][A^{\mathsf{T}}]$. It seems more productive, however, to actually specify a matrix with this property.

From (3.26) and Theorem 3.4.2, we have

$$[K] = [Q\Lambda Q^{-1}]$$

where, for $[K]$ positive semi-definite, each element λ_j on the diagonal matrix $[\Lambda]$ is nonnegative. Now define $[\Lambda^{1/2}]$ as the diagonal matrix with the elements $\sqrt{\lambda_j}$. We then have

$$[K] = [Q\Lambda^{1/2}\Lambda^{1/2}Q^{-1}] = [Q\Lambda^{1/2}Q^{-1}][Q\Lambda^{1/2}Q^{-1}]. \tag{3.27}$$

Define the square-root matrix $[R]$ for $[K]$ as

$$[R] = [Q\Lambda^{1/2}Q^{-1}]. \tag{3.28}$$

[5]Do not confuse the positive definite and positive semi-definite matrices here with the positive and nonnegative matrices we soon study as the stochastic matrices of Markov chains. The terms positive definite and semi-definite relate to the eigenvalues of symmetric matrices, whereas the terms positive and nonnegative matrices relate to the elements of typically non-symmetric matrices.

Comparing (3.27) with (3.28), we see that $[K] = [R\,R]$. However, since $[Q^{-1}] = [Q^{\mathsf{T}}]$, we see that $[R]$ is symmetric and consequently $[R] = [R^{\mathsf{T}}]$. Thus

$$[K] = [RR^{\mathsf{T}}], \tag{3.29}$$

and $[R]$ is one choice for the desired matrix $[A]$. If $[K]$ is positive definite, then each $\lambda_j > 0$ so each $\sqrt{\lambda_j} > 0$ and $[R]$ is non-singular. This then provides a converse to Theorem 3.4.3, using the square-root matrix for $[A]$. We can also use the square root matrix in the following simple theorem:

Theorem 3.4.4. *Let $[K]$ be an $n \times n$ semi-definite matrix and let $[R]$ be its square-root matrix. Then $[K]$ is the covariance matrix of the Gaussian zero-mean n-rv $\boldsymbol{Y} = [R]\,\boldsymbol{W}$ where $\boldsymbol{W} \sim \mathcal{N}(0, [I_n])$.*

Proof:

$$\mathsf{E}\,[\boldsymbol{Y}\,\boldsymbol{Y}^{\mathsf{T}}] = [R]\mathsf{E}\,[\boldsymbol{W}\,\boldsymbol{W}^{\mathsf{T}}]\,[R^{\mathsf{T}}] = [R\,R^{\mathsf{T}}] = [K]. \qquad \square$$

We can now finally relate covariance matrices to positive (semi-) definite matrices.

Theorem 3.4.5. *An $n \times n$ real matrix $[K]$ is a covariance matrix if and only if it is positive semi-definite. It is a non-singular covariance matrix if and only if it is positive definite.*

Proof: First assume $[K]$ is a covariance matrix, *i.e.*, assume there is a zero-mean n-rv $\boldsymbol{Z}$ such that $[K] = \mathsf{E}\,[\boldsymbol{Z}\,\boldsymbol{Z}^{\mathsf{T}}]$. For any given real n-vector $\boldsymbol{b}$, let the zero-mean rv X satisfy $X = \boldsymbol{b}^{\mathsf{T}}\boldsymbol{Z}$. Then

$$0 \leq \mathsf{E}\left[X^2\right] = \mathsf{E}\,[\boldsymbol{b}^{\mathsf{T}}\boldsymbol{Z}\,\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{b}] = \boldsymbol{b}^{\mathsf{T}}\mathsf{E}\,[\boldsymbol{Z}\,\boldsymbol{Z}^{\mathsf{T}}]\,\boldsymbol{b} = \boldsymbol{b}^{\mathsf{T}}[K]\boldsymbol{b}.$$

Since $\boldsymbol{b}$ is arbitrary, this shows that $[K]$ is positive semi-definite. If in addition, $[K]$ is non-singular, then it's eigenvalues are all non-zero and thus positive. Consequently $[K]$ is positive definite.

Conversely, if $[K]$ is positive semi-definite, Theorem 3.4.4 shows that $[K]$ is a covariance matrix. If, in addition, $[K]$ is positive definite, then $[K]$ is non-singular and $[K]$ is then a non-singular covariance matrix. $\square$

3.4.3 Joint probability density for Gaussian n-rv's (general case)

Recall that the joint probability density for a Gaussian n-rv $\boldsymbol{Z}$ was derived in Section 3.3.4 only for the special case where $\boldsymbol{Z} = [A]\,\boldsymbol{W}$ where the $n \times n$ matrix $[A]$ is non-singular and $\boldsymbol{W} \sim \mathcal{N}(0, [I_n])$. The above theorem lets us generalize this as follows:

Theorem 3.4.6. *Let a Gaussian zero-mean n-rv $\boldsymbol{Z}$ have a non-singular covariance matrix $[K]$. Then the probability density of $\boldsymbol{Z}$ is given by (3.22).*

Proof: Let $[R]$ be the square root matrix of $[K]$ as given in (3.28). From Theorem 3.4.4, the Gaussian vector $\boldsymbol{Y} = [R]\,\boldsymbol{W}$ has covariance $[K]$. Also $[K]$ is positive definite, so from Theorem 3.4.3 $[R]$ is non-singular. Thus $\boldsymbol{Y}$ satisfies the conditions under which (3.22) was derived, so $\boldsymbol{Y}$ has the probability density in (3.22). Since $\boldsymbol{Y}$ and $\boldsymbol{Z}$ have the same covariance and are both Gaussian zero-mean n-r**v**'s, they have the same characteristic function, and thus the same distribution. □

The question still remains about the distribution of a zero-mean Gaussian n-r**v** $\boldsymbol{Z}$ with a singular covariance matrix $[K]$. In this case $[K^{-1}]$ does not exist and thus the density in (3.22) has no meaning. From Theorem 3.4.4, $\boldsymbol{Y} = [R]\,\boldsymbol{W}$ has covariance $[K]$ but $[R]$ is singular. This means that the individual sample vectors $\boldsymbol{w}$ are mapped into a proper linear subspace of $\mathbb{R}^n$. The n-r**v** $\boldsymbol{Z}$ has zero probability outside of that subspace and, viewed as an n-dimensional density, is impulsive within that subspace.

In this case $[R]$ has one or more linearly dependent combinations of rows. As a result, one or more components Z_j of $\boldsymbol{Z}$ can be expressed as a linear combination of the other components. Very messy notation can then be avoided by viewing a maximal linearly-independent set of components of $\boldsymbol{Z}$ as a vector $\boldsymbol{Z}'$. All other components of $\boldsymbol{Z}$ are linear combinations of $\boldsymbol{Z}'$. Thus $\boldsymbol{Z}'$ has a non-singular covariance matrix and its probability density is given by (3.22).

Jointly-Gaussian rv's are often defined as rv's all of whose linear combinations are Gaussian. The next theorem shows that this definition is equivalent to the one we have given.

Theorem 3.4.7. *Let $Z_1, \ldots, Z_n$ be zero-mean rv's. These rv's are jointly Gaussian if and only if $\sum_{j=1}^{n} a_j Z_j$ is zero-mean Gaussian for all real $a_1, \ldots, a_n$.*

Proof: First assume that $Z_1, \ldots, Z_n$ are zero-mean jointly Gaussian, *i.e.*, $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\mathsf{T}}$ is a zero-mean Gaussian n-r**v**. Corollary 3.3.1 then says that $\boldsymbol{a}^{\mathsf{T}}\boldsymbol{Z}$ is zero-mean Gaussian for all real $\boldsymbol{a} = (a_1, \ldots, a_n)^{\mathsf{T}}$.

Second assume that for all real vectors $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^{\mathsf{T}}$, $\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Z}$ is zero-mean Gaussian. For any given $\boldsymbol{\theta}$, let $X = \boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Z}$, from which it follows that $\sigma_X^2 = \boldsymbol{\theta}^{\mathsf{T}}[K]\boldsymbol{\theta}$, where $[K]$ is the covariance matrix of $\boldsymbol{Z}$. By assumption, X is zero-mean Gaussian, so from (3.9), the characteristic function, $\mathsf{g}_X(i\phi) = \mathsf{E}\,[\exp(i\phi X]$, of X is

$$\mathsf{g}_X(i\phi) \;=\; \exp\left(\frac{-\phi^2\sigma_X^2}{2}\right) \;=\; \exp\left(\frac{-\phi^2\boldsymbol{\theta}^{\mathsf{T}}[K]\boldsymbol{\theta}}{2}\right) \tag{3.30}$$

Setting $\phi = 1$, we see that

$$\mathsf{g}_X(i) = \mathsf{E}\,[\exp(iX)] = \mathsf{E}\,[\exp(i\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Z})]\,.$$

In other words, the characteristic function of $X = \boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Z}$, evaluated at $\phi = 1$, is the characteristic function of $\boldsymbol{Z}$ evaluated at the given $\boldsymbol{\theta}$. Since this applies for all choices of $\boldsymbol{\theta}$,

$$\mathsf{g}_{\boldsymbol{Z}}(i\boldsymbol{\theta}) \;=\; \exp\left(\frac{-\boldsymbol{\theta}^{\mathsf{T}}[K]\boldsymbol{\theta}}{2}\right) \tag{3.31}$$

From (3.14), this is the characteristic function of an arbitrary $\boldsymbol{Z} \sim \mathcal{N}(0, [K])$. Since the characteristic function uniquely specifies the distribution of $\boldsymbol{Z}$, we have shown that $\boldsymbol{Z}$ is a zero-mean Gaussian n-r**v**. □

The following theorem summarizes the conditions under which a set of zero-mean rv's are jointly Gaussian

Theorem 3.4.8. *The following four sets of conditions are each necessary and sufficient for a zero-mean n-r$\mathbf{v}$ Z to be a zero-mean Gaussian n-r$\mathbf{v}$,* i.e.*, for the components $Z_1, \ldots, Z_n$ of $\boldsymbol{Z}$ to be jointly Gaussian:*

- *$\boldsymbol{Z}$ can be expressed as $\boldsymbol{Z} = [A]\,\boldsymbol{W}$ where $[A]$ is real and $\boldsymbol{W}$ is $\mathcal{N}(0, [I])$.*
- *For all real n-vectors $\boldsymbol{a}$, the rv $\boldsymbol{a}^{\mathsf{T}}\boldsymbol{Z}$ is zero-mean Gaussian.*
- *The linearly independent components of $\boldsymbol{Z}$ have the probability density in (3.22).*
- *The characteristic function of $\boldsymbol{Z}$ is given by (3.9).*

We emphasize once more that the distribution of a zero-mean Gaussian n-r$\mathbf{v}$ depends only on the covariance, and for every covariance matrix, zero-mean Gaussian n-r$\mathbf{v}$'s exist with that covariance. If that covariance matrix is diagonal (*i.e.*, the components of the Gaussian n-r$\mathbf{v}$ are uncorrelated), then the components are also independent. As we have seen from several examples, this depends on the definition of a Gaussian n-r$\mathbf{v}$ as having jointly-Gaussian components.

3.4.4 Geometry and principal axes for Gaussian densities

The purpose of this section is to explain the geometry of the probability density contours of a zero-mean Gaussian n-r$\mathbf{v}$ with a non-singular covariance matrix $[K]$. From (3.22), the density is constant over the region of vectors $\boldsymbol{z}$ for which $\boldsymbol{z}^{\mathsf{T}}[K^{-1}]\boldsymbol{z} = c$ for any given $c > 0$. We shall see that this region is an ellipsoid centered on 0 and that the ellipsoids for different c are concentric and expanding with increasing c.

First consider a simple special case where $Z_1, \ldots, Z_n$ are independent with different variances, *i.e.*, $Z_j \sim \mathcal{N}(0, \lambda_j)$ where $\lambda_j = \mathsf{E}\left[Z_j^2\right]$. Then $[K]$ is diagonal with elements $\lambda_1, \ldots, \lambda_n$ and $[K^{-1}]$ is diagonal with elements $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$. Then the contour for a given c is

$$\boldsymbol{z}^{\mathsf{T}}[K^{-1}]\boldsymbol{z} \;=\; \sum_{j=1}^{n} z_j^2 \lambda_j^{-1} \;=\; c. \tag{3.32}$$

This is the equation of an ellipsoid which is centered at the origin and has axes lined up with the coordinate axes. We can view this ellipsoid as a deformed n-dimensional sphere where the original sphere has been expanded or contracted along each coordinate axis j by a linear factor of $\sqrt{\lambda_j}$. An example is given in Figure 3.4.

For the general case with $\boldsymbol{Z} \sim \mathcal{N}(0, [K])$, the equiprobability contours are similar, except that the axes of the ellipsoid become the eigenvectors of $[K]$. To see this, we represent $[K]$ as $[Q\Lambda Q^{\mathsf{T}}]$ where the orthonormal columns of $[Q]$ are the eigenvectors of $[K]$ and $[\Lambda]$ is the

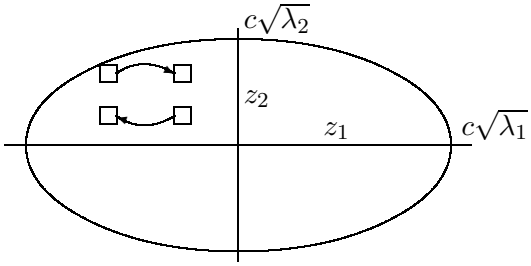


Figure 3.4: A contour of equal probability density for 2 dimensions with diagonal $[K]$. The figure assumes that $\lambda_1 = 4\lambda_2$. The figure also shows how the joint probability density can be changed without changing the Gaussian marginal probability densitities. For any rectangle aligned with the coordinate axes, incremental squares can be placed at the vertices of the rectangle and ϵ probability can be transferred from left to right on top and right to left on bottom with no change in the marginals. This transfer can be done simultaneously for any number of rectangles, and by reversing the direction of the transfers appropriately, zero covariance can be maintained. Thus the elliptical contour property depends critically on the variables being jointly Gaussian rather than merely individually Gaussian.

diagonal matrix of eigenvalues, all of which are positive. Thus we want to find the set of vectors $\boldsymbol{z}$ for which

$$\boldsymbol{z}^{\mathsf{T}}[K^{-1}]\boldsymbol{z} \;=\; \boldsymbol{z}^{\mathsf{T}}[Q\Lambda^{-1}Q^{\mathsf{T}}]\boldsymbol{z} \;=\; c. \tag{3.33}$$

Since the eigenvectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are orthonormal, they span $\mathbb{R}^n$ and any vector $\boldsymbol{z} \in \mathbb{R}^n$ can be represented as a linear combination, say $\sum_j v_j \boldsymbol{q}_j$ of $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$. In vector terms this is $\boldsymbol{z} = [Q]\boldsymbol{v}$. Thus $\boldsymbol{v}$ represents $\boldsymbol{z}$ in the coordinate basis in which the axes are the eigenvectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$. Substituting $\boldsymbol{z} = [Q]\boldsymbol{v}$ in (3.33),

$$\boldsymbol{z}^{\mathsf{T}}[K^{-1}]\boldsymbol{z} \;=\; \boldsymbol{v}^{\mathsf{T}}[\Lambda^{-1}]\boldsymbol{v} \;=\; \sum_{j=1}^{n} v_j^2 \lambda_j^{-1} \;=\; c. \tag{3.34}$$

This is the same as (3.32) except that here the ellipsoid is defined in terms of the representation $v_j = \boldsymbol{q}_j^{\mathsf{T}}\boldsymbol{z}$ for $1 \le j \le n$. Thus the equiprobability contours are ellipsoids whose axes are the eigenfunctions of $[K]$. (see Figure 3.5). We can also substitute this into (3.22) to obtain what is often a more convenient expression for the probability density of $\boldsymbol{Z}$.

$$\mathsf{f}_{\boldsymbol{Z}}(\boldsymbol{z}) \;=\; \frac{\exp\left(-\frac{1}{2}\sum_{j=1}^{n} v_j^2 \lambda_j^{-1}\right)}{(2\pi)^{n/2}\sqrt{\det[K]}} \tag{3.35}$$

$$\;=\; \prod_{j=1}^{n} \frac{\exp(-v_j^2/(2\lambda_j))}{\sqrt{2\pi\lambda_j}}, \tag{3.36}$$

where $v_j = \boldsymbol{q}_j^{\mathsf{T}}\boldsymbol{z}$ and we have used the fact that $\det[K] = \prod_j \lambda_j$.

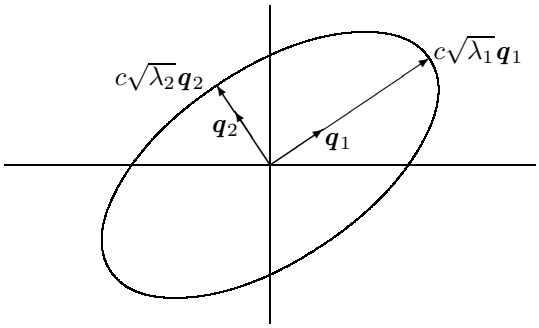


Figure 3.5: Contours of equal probability density. Points $\boldsymbol{z}$ on the $\boldsymbol{q}_j$ axis are points for which $v_k = 0$ for all $k \neq j$. Points on the illustrated ellipse satisfy $\boldsymbol{z}^{\mathsf{T}}[K^{-1}]\boldsymbol{z} = c$.

3.5 Conditional PDF's for Gaussian random vectors

Next consider the conditional probability $\mathsf{f}_{X|Y}(x|y)$ for two zero-mean jointly-Gaussian random vectors X and Y with a non-singular covariance matrix. From (3.23),

$$\mathsf{f}_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[\frac{-(x/\sigma_X)^2 + 2\rho(x/\sigma_X)(y/\sigma_Y) - (y/\sigma_Y)^2}{2(1-\rho^2)}\right],$$

where $\rho = \mathsf{E}\left[XY\right]/(\sigma_X\sigma_Y)$. Since $\mathsf{f}_Y(y) = (2\pi\sigma_Y^2)^{-1/2}\exp(-y^2/2\sigma_Y^2)$, we have

$$\mathsf{f}_{X|Y}(x|y) = \frac{1}{\sigma_X\sqrt{2\pi(1-\rho^2)}} \exp\left[\frac{-(x/\sigma_X)^2 + 2\rho(x/\sigma_X)(y/\sigma_Y) - \rho^2(y/\sigma_Y)^2}{2(1-\rho^2)}\right].$$

The numerator of the exponent is the negative of the square $(x/\sigma_x - \rho y/\sigma_y)^2$. Thus

$$\mathsf{f}_{X|Y}(x|y) = \frac{1}{\sigma_X\sqrt{2\pi(1-\rho^2)}} \exp\left[\frac{-\left[x - \rho(\sigma_X/\sigma_Y)y\right]^2}{2\sigma_X^2(1-\rho^2)}\right]. \tag{3.37}$$

This says that, given any particular sample value y for the rv Y, the conditional density of X is Gaussian with variance $\sigma_X^2(1-\rho^2)$ and mean $\rho(\sigma_X/\sigma_Y)y$. Given $Y=y$, we can view X as a random variable in the restricted sample space where $Y=y$. In that restricted sample space, X is $\mathcal{N}\left(\rho(\sigma_X/\sigma_Y)y,\ \sigma_X^2(1-\rho^2)\right)$.

We see that the variance of X, given $Y = y$, has been reduced by a factor of $1-\rho^2$ from the variance before the observation. It is not surprising that this reduction is large when $|\rho|$ is close to 1 and negligible when ρ is close to 0. It is surprising that this conditional variance is the same for all values of y. It is also surprising that the conditional mean of X is linear in y and that the conditional distribution is Gaussian with a variance constant in y.

Another way to interpret this conditional distribution of X conditional on Y is to use the above observation that the conditional fluctuation of X, conditional on $Y = y$, does not

depend on y. This fluctuation can then be denoted as a rv V that is independent of Y. Thus we can represent X as $X = \rho(\sigma_X/\sigma_Y)Y + V$ where $V \sim \mathcal{N}(0, (1-\rho^2)\sigma_X^2)$ and V is independent of Y.

As will be seen in Chapter 10, this simple form for the conditional distribution leads to important simplifications in estimating X from Y. We now go on to show that this same kind of simplification occurs when we study the conditional density of one Gaussian random vector conditional on another Gaussian random vector, assuming that all the variables are jointly Gaussian.

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\mathsf{T}$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^\mathsf{T}$ be zero-mean jointly Gaussian r$\mathbf{v}$'s of length n and m (*i.e.*, $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ are jointly Gaussian). Let their covariance matrices be $[K_X]$ and $[K_Y]$ respectively. Let $[K]$ be the covariance matrix of the $(n+m)$-r$\mathbf{v}$ $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)^\mathsf{T}$.

The $(n+m) \times (n+m)$ covariance matrix $[K]$ can be partitioned into n rows on top and m rows on bottom, and then further partitioned into n and m columns, yielding:

$$[K] = \begin{bmatrix} [K_{\boldsymbol{X}}] & [K_{\boldsymbol{X}\cdot\boldsymbol{Y}}] \\ [K_{\boldsymbol{X}\cdot\boldsymbol{Y}}^\mathsf{T}] & [K_{\boldsymbol{Y}}] \end{bmatrix}. \tag{3.38}$$

Here $[K_{\boldsymbol{X}}] = \mathsf{E}[\boldsymbol{X}\,\boldsymbol{X}^\mathsf{T}]$, $[K_{\boldsymbol{X}\cdot\boldsymbol{Y}}] = \mathsf{E}[\boldsymbol{X}\,\boldsymbol{Y}^\mathsf{T}]$, and $[K_{\boldsymbol{Y}}] = \mathsf{E}[\boldsymbol{Y}\,\boldsymbol{Y}^\mathsf{T}]$. Note that if $\boldsymbol{X}$ and $\boldsymbol{Y}$ have means, then $[K_{\boldsymbol{X}}] = \mathsf{E}\left[(\boldsymbol{X} - \overline{\boldsymbol{X}})(\boldsymbol{X} - \overline{\boldsymbol{X}})^\mathsf{T}\right]$, $[K_{\boldsymbol{X}\cdot\boldsymbol{Y}}] = \mathsf{E}\left[(\boldsymbol{X} - \overline{\boldsymbol{X}})(\boldsymbol{Y} - \overline{\boldsymbol{Y}})^\mathsf{T}\right]$, etc.

In what follows, assume that $[K]$ is non-singular. We then say that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are *jointly non-singular*, which implies that none of the rv's $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ can be expressed as a linear combination of the others. The inverse of $[K]$ then exists and can be denoted in block form as

$$[K^{-1}] = \begin{bmatrix} [B] & [C] \\ [C^\mathsf{T}] & [D] \end{bmatrix}. \tag{3.39}$$

The blocks $[B], [C], [D]$ can be calculated directly from $[KK^{-1}] = [I]$ (see Exercise 3.16), but for now we simply use them to find $\mathsf{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$.

We shall find that for any given $\boldsymbol{y}$, $\mathsf{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$ is a jointly-Gaussian density with a conditional covariance matrix equal to $[B^{-1}]$ (Exercise 3.11 shows that $[B]$ is non-singular). As in (3.37), where X and Y are one-dimensional, this covariance does not depend on $\boldsymbol{y}$. Also, the conditional mean of $\boldsymbol{X}$, given $\boldsymbol{Y} = \boldsymbol{y}$, will turn out to be $-[B^{-1}C]\,\boldsymbol{y}$. More precisely, we have the following theorem:

Theorem 3.5.1. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be zero-mean, jointly Gaussian, jointly non-singular r$\mathbf{v}$'s. Then $\boldsymbol{X}$, conditional on $\boldsymbol{Y} = \boldsymbol{y}$, is $\mathcal{N}\big(-[B^{-1}C]\,\boldsymbol{y},\, [B^{-1}]\big)$, i.e.,*

$$\mathsf{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) = \frac{\exp\left\{-\frac{1}{2}\Big(\boldsymbol{x} + [B^{-1}C]\,\boldsymbol{y}^\mathsf{T}\Big)\;[B]\;\Big(\boldsymbol{x} + [B^{-1}C]\,\boldsymbol{y}\Big)\right\}}{(2\pi)^{n/2}\sqrt{\det[B^{-1}]}}. \tag{3.40}$$

Proof: Express $\mathsf{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$ as $\mathsf{f}_{\boldsymbol{X}\,\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})/\mathsf{f}_{\boldsymbol{Y}}(\boldsymbol{y})$. From (3.22),

$$\begin{aligned}\mathsf{f}_{\boldsymbol{X}\,\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) &= \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{x}^{\mathsf{T}},\boldsymbol{y}^{\mathsf{T}})[K^{-1}](\boldsymbol{x}^{\mathsf{T}},\boldsymbol{y}^{\mathsf{T}})^{\mathsf{T}}\right\}}{(2\pi)^{(n+m)/2}\sqrt{\det[K^{-1}]}} \\ &= \frac{\exp\left\{-\frac{1}{2}\left(\boldsymbol{x}^{\mathsf{T}}[B]\boldsymbol{x}+\boldsymbol{x}^{\mathsf{T}}[C]\boldsymbol{y}+\boldsymbol{y}^{\mathsf{T}}[C^{\mathsf{T}}]\boldsymbol{x}+\boldsymbol{y}^{\mathsf{T}}[D]\boldsymbol{y}\right)\right\}}{(2\pi)^{(n+m)/2}\sqrt{\det[K^{-1}]}}.\end{aligned}$$

Note that $\boldsymbol{x}$ appears only in the first three terms of the exponent above, and that $\boldsymbol{x}$ does not appear at all in $\mathsf{f}_{\boldsymbol{Y}}(\boldsymbol{y})$. Thus we can express the dependence on $\boldsymbol{x}$ in $\mathsf{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$ by

$$\mathsf{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\mid\boldsymbol{y}) = \phi(\boldsymbol{y})\exp\left\{-\frac{1}{2}\Big[\boldsymbol{x}^{\mathsf{T}}[B]\boldsymbol{x}+\boldsymbol{x}^{\mathsf{T}}[C]\boldsymbol{y}+\boldsymbol{y}^{\mathsf{T}}[C^{\mathsf{T}}]\boldsymbol{x}\Big]\right\}, \tag{3.41}$$

where $\phi(\boldsymbol{y})$ is some function of $\boldsymbol{y}$. We now complete the square around $[B]$ in the exponent above, getting

$$\mathsf{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\mid\boldsymbol{y}) = \phi(\boldsymbol{y})\exp\left\{-\frac{1}{2}\Big[(\boldsymbol{x}+[B^{-1}C]\,\boldsymbol{y})^{\mathsf{T}}[B]\,(\boldsymbol{x}+[B^{-1}C]\,\boldsymbol{y}) + \boldsymbol{y}^{\mathsf{T}}[C^{\mathsf{T}}B^{-1}C]\,\boldsymbol{y}\Big]\right\}.$$

Since the last term in the exponent does not depend on $\boldsymbol{x}$, we can absorb it into $\phi(\boldsymbol{y})$. The remaining expression has the form of the density of a Gaussian n-r**v** with non-zero mean as given in (3.24). Comparison with (3.24) also shows that $\phi(\boldsymbol{y})$ must be $(2\pi)^{-n/2}(\det[B^{-1})^{-1/2}]$. With this substituted for $\phi(\boldsymbol{y})$, we have (3.40). □

To interpret (3.40), note that for any sample value $\boldsymbol{y}$ for $\boldsymbol{Y}$, the conditional distribution of $\boldsymbol{X}$ has a mean given by $-[B^{-1}C]\boldsymbol{y}$ and a Gaussian fluctuation around the mean of variance $[B^{-1}]$. This fluctuation has the same distribution for all $\boldsymbol{y}$ and thus can be represented as a r**v** $\boldsymbol{V}$ that is independent of $\boldsymbol{Y}$. Thus we can represent $\boldsymbol{X}$ as

$$\boldsymbol{X} = [G]\,\boldsymbol{Y} + \boldsymbol{V}\,; \quad \boldsymbol{Y},\,\boldsymbol{V} \text{ independent}, \tag{3.42}$$

where

$$[G] = -[B^{-1}C] \quad \text{and} \quad \boldsymbol{V} \sim \mathcal{N}(0,[B^{-1}]). \tag{3.43}$$

We often call $\boldsymbol{V}$ an *innovation*, because it is the part of $\boldsymbol{X}$ that is independent of $\boldsymbol{Y}$. It is also called a *noise term* for the same reason. We will call $[K_{\boldsymbol{V}}] = [B^{-1}]$ the *conditional covariance* of $\boldsymbol{X}$ given a sample value $\boldsymbol{y}$ for $\boldsymbol{Y}$. In summary, the unconditional covariance, $[K_{\boldsymbol{X}}]$, of $\boldsymbol{X}$ is given by the upper left block of $[K]$ in (3.38), while the conditional covariance $[K_{\boldsymbol{V}}]$ is the inverse of the upper left block, $[B]$, of the inverse of $[K]$.

The following theorem expresses (3.42) and (3.43) directly in terms of the covariances of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Theorem 3.5.2. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be zero-mean, jointly Gaussian, and jointly non-singular. Then $\boldsymbol{X}$ can be expressed as $\boldsymbol{X} = [G]\,\boldsymbol{Y} + \boldsymbol{V}$ where $\boldsymbol{V}$ is statistically independent of $\boldsymbol{Y}$ and*

$$G = [K_{X\cdot Y}K_Y^{-1}] \tag{3.44}$$

$$[K_{\boldsymbol{V}}] = [K_{\boldsymbol{X}}] - [K_{\boldsymbol{X}\cdot\boldsymbol{Y}}K_{\boldsymbol{Y}}^{-1}K_{\boldsymbol{X}\cdot\boldsymbol{Y}}^{T}] \tag{3.45}$$

Proof: From (3.42), we know that $\boldsymbol{X}$ can be represented as $[G]\,\boldsymbol{Y} + \boldsymbol{V}$ with $\boldsymbol{Y}$ and $\boldsymbol{V}$ independent, so we simply have to evaluate $[G]$ and $[K_{\boldsymbol{V}}]$. Using (3.42), the covariance of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is given by

$$[K_{\boldsymbol{X}\cdot\boldsymbol{Y}}] \;=\; \mathsf{E}\,[\boldsymbol{X}\,\boldsymbol{Y}^{\mathsf{T}}] \;=\; \mathsf{E}\,[[G]\,\boldsymbol{Y}\,\boldsymbol{Y}^{\mathsf{T}} + \boldsymbol{V}\,\boldsymbol{Y}^{\mathsf{T}}] \;=\; [GK_{\boldsymbol{Y}}],$$

where we used the fact that $\boldsymbol{V}$ and $\boldsymbol{Y}$ are independent. Post-multiplying both sides by $[K_{\boldsymbol{Y}}^{-1}]$ yields (3.44). To verify (3.45), we use (3.42) to express $[K_{\boldsymbol{X}}]$ as

$$\begin{aligned}
[K_{\boldsymbol{X}}] &= \mathsf{E}\,[\boldsymbol{X}\,\boldsymbol{X}^{\mathsf{T}}] \;=\; \mathsf{E}\,[([G]\,\boldsymbol{Y} + \boldsymbol{V})([G]\,\boldsymbol{Y} + \boldsymbol{V})^{\mathsf{T}}] \\
&= [GK_{\boldsymbol{Y}}G^{\mathsf{T}}] + [K_{\boldsymbol{V}}], \qquad \text{so} \\
[K_{\boldsymbol{V}}] &= [K_{\boldsymbol{X}}] - [GK_{\boldsymbol{Y}}G^{\mathsf{T}}].
\end{aligned}$$

This yields (3.45) when (3.44) is used for $[G]$. □

We have seen that $[K_{\boldsymbol{V}}]$ is the covariance of $\boldsymbol{X}$ conditional on $\boldsymbol{Y} = \boldsymbol{y}$ for each sample value $\boldsymbol{y}$. The expression in (3.45) provides some insight into how this covariance is reduced from $[K_{\boldsymbol{X}}]$. More particularly, for any n-vector $\boldsymbol{b}$,

$$\boldsymbol{b}^{\mathsf{T}}[K_{\boldsymbol{X}}]\boldsymbol{b} \geq \boldsymbol{b}^{\mathsf{T}}[K_{\boldsymbol{V}}]\boldsymbol{b},$$

i.e., the unconditional variance of $\boldsymbol{b}^{\mathsf{T}}\boldsymbol{X}$ is always greater than or equal to the variance of $\boldsymbol{b}^{\mathsf{T}}\boldsymbol{X}$ conditional on $\boldsymbol{Y} = \boldsymbol{y}$.

In the process of deriving these results, we have also implicity evaluated the matrices $[C]$ and $[B]$ in the inverse of $[K]$ in (3.39). Combining the second part of (3.43) with (3.45),

$$[B] \;=\; \Big([K_{\boldsymbol{X}}] - [K_{\boldsymbol{X}\cdot\boldsymbol{Y}}K_{\boldsymbol{Y}}^{-1}K_{\boldsymbol{X}\cdot\boldsymbol{Y}}^{\mathsf{T}}]\Big)^{-1} \tag{3.46}$$

Combining the first part of (3.43) with (3.44), we get

$$[C] \;=\; -[BK_{X\cdot Y}K_{Y}^{-1}] \tag{3.47}$$

Finally, reversing the roles of $\boldsymbol{X}$ and $\boldsymbol{Y}$, we can express D as

$$[D] = \Big([K_{\boldsymbol{Y}}] - [K_{\boldsymbol{Y}\cdot\boldsymbol{X}}K_{\boldsymbol{X}}^{-1}K_{\boldsymbol{Y}\cdot\boldsymbol{X}}^{\mathsf{T}}]\Big)^{-1} \tag{3.48}$$

Reversing the roles of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is even more important in another way, since Theorem 3.5.2 then also says that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are related by

$$\boldsymbol{Y} \;=\; [H]\boldsymbol{X} + \boldsymbol{Z}, \qquad \text{where } \boldsymbol{X} \text{ and } \boldsymbol{Z} \text{ are independent and} \tag{3.49}$$

$$[H] \;=\; [K_{\boldsymbol{Y}\cdot\boldsymbol{X}}K_{\boldsymbol{X}}^{-1}], \tag{3.50}$$

$$[K_{\boldsymbol{Z}}] \;=\; [K_{\boldsymbol{Y}}] - [K_{\boldsymbol{Y}\cdot\boldsymbol{X}}K_{\boldsymbol{X}}^{-1}K_{\boldsymbol{Y}\cdot\boldsymbol{X}}^{\mathsf{T}}]. \tag{3.51}$$

This gives us three ways of representing any pair $\boldsymbol{X}, \boldsymbol{Y}$ of zero-mean jointly Gaussian r**v**'s whose combined covariance is non-singular. First, they can be represented simply as

an overall r$\mathbf{v}$, $(X_1, \ldots, X_n Y_1, \ldots, Y_m)^\mathsf{T}$, second as $\boldsymbol{X} = [G]\,\boldsymbol{Y} + \boldsymbol{V}$ where $\boldsymbol{Y}$ and $\boldsymbol{V}$ are independent, and third as $\boldsymbol{Y} = [H]\boldsymbol{X} + \boldsymbol{Z}$ where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are independent.

Each of these formulations essentially implies the existence of the other two. If we start with formulation 3, for example, Exercise 3.17 shows simply that if $\boldsymbol{X}$ and $\boldsymbol{Z}$ are each zero-mean Gaussian r$\mathbf{v}$'s, the independence between them assures that they are jointly Gaussian, and thus that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are also jointly Gaussian. Similarly, if $[K_{\boldsymbol{X}}]$ and $[K_{\boldsymbol{Z}}]$ are nonsingular, the overall $[K]$ for $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)^\mathsf{T}$ must be non-singular. In Chapter 10, we will find that this provides a very simple and elegant solution to jointly Gaussian estimation problems.

3.6 Gaussian processes

Recall that a stochastic process (or random process) $\{X(t); t \in \mathcal{T}\}$ is a collection of rv's, one for each value of the parameter t in some parameter set $\mathcal{T}$. The parameter t usually denotes time, so there is one rv for each instant of time. For discrete-time processes, $\mathcal{T}$ is usually limited to the set of integers, $\mathbb{Z}$, and for continuous-time, $\mathcal{T}$ is usually limited to $\mathbb{R}$. In each case, t is sometimes additionally restricted to $t \geq 0$; this is denoted $\mathbb{Z}^+$ and $\mathbb{R}^+$ respectively. We use the word *epoch* to denote a value of t within $\mathcal{T}$.

Definition 3.6.1. *A Gaussian process $\{X(t); t \in \mathcal{T}\}$ is a stochastic process such that for all positive integers k and all choices of epochs $t_1, \ldots, t_k \in \mathcal{T}$, the set of rv's $X(t_1), \ldots, X(t_k)$ is a jointly-Gaussian set of rv's.*

The previous sections of this chapter should motivate both the simplicity and usefulness associated with this jointly-Gaussian requirement. In particular, the joint probability density of any k-r$\mathbf{v}$ $(X(t_1), \ldots, X(t_k))^\mathsf{T}$, is essentially specified by (3.24), using only the covariance matrix and the mean for each rv. If the rv's are individually Gaussian but not jointly Gaussian, none of this holds.

Definition 3.6.2. *The* **covariance function**, *$K_X(t, \tau)$, of a stochastic process $\{X(t);\, t{\in}\mathcal{T}\}$ is defined for all $t, \tau \in \mathcal{T}$ by*

$$K_X(t, \tau) = \mathsf{E}\left[(X(t) - \overline{X}(t))(X(\tau) - \overline{X}(\tau)\right] \tag{3.52}$$

Note that for each k-r$\mathbf{v}$ $(X(t_1), \ldots, X(t_k))^\mathsf{T}$, the (j, ℓ) element of the covariance matrix is simply $K_X(t_j, t_\ell)$. Thus the covariance function and the mean of a process specify the covariance matrix and mean of each k-r$\mathbf{v}$. This establishes the following simple but important result.

Theorem 3.6.1. *For a Gaussian process $\{X(t); t \in \mathcal{T}\}$, the covariance function $K_X(t, \tau)$ and the mean $\mathsf{E}[X(t)]$ for each $t, \tau \in \mathcal{T}$ specify the joint probability density for all k-r$\mathbf{v}$'s $(X(t_1), \ldots, X(t_k))^\mathsf{T}$ for all $k > 1$.*

We now give several examples of discrete-time Gaussian processes and their covariance functions. As usual, we look at the zero-mean case, since a mean can always be simply added later. Continuous-time Gaussian processes are a considerably more complicated and are considered in Section 3.6.3

Example 3.6.1 (Discrete time IID Gaussian process). Consider the stochastic process $\{W(n);\ n\in\mathbb{Z}\}$ where $\ldots, W(-1), W(0), W(1), \ldots$ is a sequence of IID Gaussian rv's, $W(n) \sim \mathcal{N}(0,\sigma^2)$. The mean is zero for all n and the covariance function is $K_W(n,k) = \sigma^2\delta_{nk}$. For any k epochs, $n_1, n_2, \ldots, n_k$, the joint density is

$$p_{W(n_1),\ldots,W(n_k)}(w_1,\ldots,w_k) = \frac{1}{(2\pi\sigma^2)^{k/2}} \exp\left(-\sum_{i=1}^{k}\frac{w_i^2}{2\sigma^2}\right). \tag{3.53}$$

Note that this process is very much like the IID Gaussian vectors we have studied. The only difference is that we now have an infinite number of dimensions (*i.e.*, an infinite number of IID rv's) for which all finite subsets are jointly Gaussian.

Example 3.6.2 (Discrete-time Gaussian sum process). Consider the stochastic process $\{S(n); n \geq 1\}$ which is defined from the discrete-time IID Gaussian process by $S(n) = W(1)+W(2)+\cdots+W(n)$. Viewing $(S_1,\ldots,S_n)^\mathsf{T}$ as a linear transformation of $(W_1,\ldots,W_n)^\mathsf{T}$, we see that $S_1,\ldots,S_n$ is a zero-mean jointly-Gaussian set of rv's. Since this is true for all $n \geq 1$, $\{S(n); n \geq 1\}$ is a zero-mean Gaussian process. For $n \leq k$, the covariance function is

$$K_X(n,k) = \mathsf{E}\left[\sum_{j=1}^{n} W_j \sum_{\ell=1}^{k} W_\ell\right] = \sum_{j=1}^{n} \mathsf{E}\left[W_j^2\right] = n\sigma^2.$$

Using a similar argument for $n > k$, the general result is

$$K_X(n,k) = \min(n,k)\sigma^2.$$

Example 3.6.3 (Discrete-time Gauss-Markov process). Let α be a real number, $|\alpha| < 1$ and consider a stochastic process $\{X(n);\ n \in \mathbb{Z}^+\}$ which is defined in terms of the previous example of an IID Gaussian process $\{W_n;\ n \in \mathbb{Z}\}$ by

$$X(n+1) = \alpha X(n) + W(n); \quad \text{for } n \in \mathbb{Z}^+; \quad X(0) = 0 \tag{3.54}$$

By applying (3.54) recursively,

$$X(n) = W(n-1) + \alpha W(n-2) + \alpha^2 W(n-3) + \cdots + \alpha^{n-1}W(0) \tag{3.55}$$

This is another example in which the new process $\{X(n);\ n \geq 1\}$ is a linear transformation of another process $\{W(n);\ n \geq 0\}$. Since $\{W(n);\ n \geq 0\}$ is a zero-mean Gaussian process, $\{X_n;\ n \geq 0\}$ is also. Thus $\{X(n);\ n \geq 0\}$ is specified by its covariance function, calculated in Exercise 3.22 to be

$$\mathsf{E}\left[X(n)X(n+k)\right] = \frac{\sigma^2(1-\alpha^{2n})\alpha^k}{1-\alpha^2} \tag{3.56}$$

Since $|\alpha| < 1$, the coefficients α^k in (3.55) are geometrically decreasing in k, and therefore, for large n it makes little difference whether the sum stops with the term $\alpha^{n-1}W(0)$ or whether terms $\alpha^n W(-1)$, $\alpha^{n+1}W_{-2}, \ldots$, are added.[6] Similarly, from (3.56), we see that

[6]One might ask whether the limit $\sum_{j=1}^{\infty}\alpha^{j-1}W(n-j)$ exists as a rv. As intuition almost demands, the answer is yes. We will show this in Section 9.9.2 as a consequence of the martingale convergence theorem.

$\lim_{n\to\infty} \mathsf{E}[X(n)X(n+k)] = \frac{\sigma^2\alpha^k}{(1-\alpha^2)}$. This suggests that the starting time of this process is irrelevant if it is far enough into the past, and thus suggests that we could define essentially the same process over all integer times n by

$$X(n+1) = \alpha X(n) + W(n); \quad \text{for all } n \in \mathbb{Z} \tag{3.57}$$

By applying (3.57) recursively, $X(n) = \sum_{j=1}^{\infty} \alpha^{j-1} W(n-j)$.

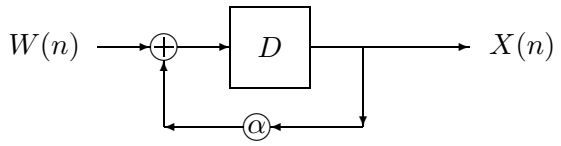


Figure 3.6: Schematic of the generation of $\{X(n);\, n \geq 1\}$ from $X(0) = 0$ and $\{W(n);\, n \geq 0\}$. The element D is a unit delay. It can be seen from the figure that X_{n+1} depends probabilistically on the past history $X_1, \ldots, X_n$ only through X_n. This is called a Gauss-Markov process, and the sample value x_n of X_n is called the *state* of the process at time n. This process differs from the Markov processes developed in Chapters 4, 6, and 7 in the sense that the state is an arbitrary real number rather than a discrete value.

3.6.1 Stationarity and related concepts:

Many of the most useful stochastic processes have the property that the location of the time origin is irrelevant, *i.e.*, that the process "behaves" 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*. A precise definition will be given shortly.

An obvious requirement for stationarity is that $X(t)$ must be identically distributed for all $t \in \mathcal{T}$. A more subtle requirement is that for every $k > 1$ and set of epochs $t_1, \ldots, t_k \in \mathcal{T}$, the joint distribution over these epochs should be the same as that over a shift in time of these epochs to, say, $t_1+\tau, \ldots, t_k+\tau \in \mathcal{T}$.

This shift requirement for stationarity becomes quite obscure and meaningless unless $\mathcal{T}$ is chosen so that a shift of a set of epochs in $\mathcal{T}$ is also in $\mathcal{T}$. This explains why the definition of $\mathcal{T}$ is restricted in the following definition.

Definition 3.6.3. *Let a stochastic process $\{X(t);\, t \in \mathcal{T}\}$ be defined over a set of epochs $\mathcal{T}$ where $\mathcal{T}$ is either $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{Z}^+$, or $\mathbb{R}^+$. The process is* **stationary** *if, for all positive integers k and all $\tau, t_1, \ldots, t_k$ in $\mathcal{T}$,*

$$\mathsf{F}_{X(t_1),\ldots,X(t_k)}(x_1 \ldots, x_k) = \mathsf{F}_{X(t_1+\tau),\ldots,X(t_k+\tau)}(x_1 \ldots, x_k) \tag{3.58}$$

Note that the restriction on $\mathcal{T}$ in the definition guarantees that if $X(t_1), \ldots, X(t_k) \in \mathcal{T}$, then $X(t_1+\tau), \ldots, X(t_k+\tau) \in \mathcal{T}$ also. In this chapter, $\mathcal{T}$ is usually $\mathbb{Z}$ or $\mathbb{R}$, whereas in Chapters 4, 6, and 7, $\mathcal{T}$ is usually restricted to $\mathbb{Z}^+$ or $\mathbb{R}^+$.

The discrete-time IID Gaussian process in Example 3.6.1 is stationary since all joint distributions of a given number of distinct variables from $\{W(n);\, n \in \mathbb{Z}\}$ are the same. More generally, for any Gaussian process, the joint distribution of $X(t_1), \ldots, X(t_k)$ depends only on the mean and covariance of those variables. In order for this distribution to be the same as that of $X(t_1+\tau), \ldots, X(t_k+\tau)$, it is necessary that $\mathsf{E}[X(t)] = \mathsf{E}[X(0)]$ for all epochs t and also that $K_X(t_1, t_2) = K_X(t_1{+}\tau,\, t_2{+}\tau)$ for all epochs t_1, t_2, and τ. This latter condition can be simplified to the statement that $K_X(t,\, t{+}u)$ is a function only of u and not of t. It can be seen that these conditions are also sufficient for a Gaussian process $\{X(t)\}$ to be stationary. We summarize this in the following theorem.

Theorem 3.6.2. *A Gaussian process $\{X(t);\, t \in \mathcal{T}\}$ (where $\mathcal{T}$ is $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{Z}^+$, or $\mathbb{R}^+$) is stationary if and only if $\mathsf{E}[X(t)] = \mathsf{E}[X(0)]$ and $K_X(t,\, t{+}u) = K_X(0, u)$ for all $t, u \in \mathcal{T}$.*

With this theorem, we see that the Gauss Markov process of Example 3.6.3, extended to the set of all integers, is a discrete-time stationary process. The Gaussian sum process of Example 3.6.2, however, is non-stationary.

For non-Gaussian processes, it is frequently difficult to calculate joint distributions in order to determine if the process is stationary. There are a number of results that depend only on the mean and the covariance function, and these make it convenient to have the following more relaxed definition:

Definition 3.6.4. *A stochastic process $\{X(t);\, t \in \mathcal{T}\}$ (where $\mathcal{T}$ is $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{Z}^+$, or $\mathbb{R}^+$) is* **wide sense stationary**[7] **(WSS)** *if $\mathsf{E}[X(t)] = \mathsf{E}[X(0)]$ and $K_X(t,\, t{+}u) = K_X(0, u)$ for all $t, u \in \mathcal{T}$.*

Since the covariance function $K_X(t,\, t{+}u)$ of a stationary or WSS process is a function of only one variable u, we will often write the covariance function of a WSS process as a function of one variable, namely $K_X(u)$ in place of $K_X(t,\, t{+}u)$. The single variable in the single-argument form represents the difference between the two arguments in the two-argument form. Thus, the covariance function $K_X(t, \tau)$ of a WSS process must be a function only of $t - \tau$ and is expressed in single-argument form as $K_X(t - \tau)$. Note also that since $K_X(t, \tau) = K_X(\tau, t)$, the covariance function of a WSS process must be symmetric, *i.e.*, $K_X(u) = K_X(-u)$,

The reader should not conclude from the frequent use of the term WSS in the literature that there are many important processes that are WSS but not stationary. Rather, the use of WSS in a result is used primarily to indicate that the result depends only on the mean and covariance.

3.6.2 Orthonormal expansions

The previous Gaussian process examples were discrete-time processes. The simplest way to generate a broad class of continuous-time Gaussian processes is to start with a discrete-time process (*i.e.*, a sequence of jointly-Gaussian rv's) and use these rv's as the coefficients in

[7] This is also called weakly stationary, covariance stationary, and second-order stationary.

an orthonormal expansion. We describe some of the properties of orthonormal expansions in this section and then describe how to use these expansions to generate continuous-time Gaussian processes in Section 3.6.3.

A set of functions $\{\phi_n(t);\, n \geq 1\}$ is defined to be orthonormal if

$$\int_{-\infty}^{\infty} \phi_n(t)\phi_k^*(t)\, dt = \delta_{nk} \qquad \text{for all integers } n, k. \tag{3.59}$$

These functions can be either complex or real functions of the real variable t; the complex case (using the reals as a special case) is most convenient.

The most familiar orthonormal set is that used in the Fourier series.

$$\phi_n(t) = \begin{cases} (1/\sqrt{T}) \exp[i2\pi nt/T] & \text{for } |t| \leq T/2 \\ 0 & \text{for } |t| > T/2 \end{cases}. \tag{3.60}$$

We can then take any square-integrable real or complex function $x(t)$ over $(-T/2, T/2)$ and essentially[8] represent it by

$$x(t) = \sum_n x_n \phi_n(t)\,; \quad \text{where } x_n = \int_{-T/2}^{T/2} x(t)\phi_n^*(t)dt \tag{3.61}$$

The complex exponential form of the Fourier series could be replaced by the sine/cosine form when expanding real functions (as here). This has the conceptual advantage of keeping everying real, but doesn't warrant the added analytical complexity.

Many features of the Fourier transform are due not to the special nature of sinusoids, but rather to the fact that the function is being represented as a series of orthonormal functions. To see this, let $\{\phi_n(t);\, n \in \mathbb{Z}\}$ be any set of orthonormal functions, and assume that a function $x(t)$ can be represented as

$$x(t) = \sum_n x_n \phi_n(t). \tag{3.62}$$

Multiplying both sides of (3.62) by $\phi_m^*(t)$ and integrating,

$$\int x(t)\phi_m^*(t)dt = \int \sum_n x_n \phi_n(t)\phi_m^*(t)dt.$$

Using (3.59) to see that only one term on the right is non-zero, we get

$$\int x(t)\phi_m^*(t)dt = x_m. \tag{3.63}$$

[8]More precisely, the difference between $x(t)$ and its Fourier series $\sum_n x_n\phi_n(t)$ has zero energy, *i.e.*, $\int \left|x(t) - \sum_n x_n\phi_n(t)\right|^2 dt = 0$. This allows $x(t)$ and $\sum_n x_n\phi_n(t)$ to differ at isolated values of t such as points of discontinuity in $x(t)$. Engineers view this as essential equality and mathematicians define it carefully and call it L_2 equivalence.

We don't have the mathematical tools to easily justify this interchange and it would take us too far afield to acquire those tools. Thus for the remainder of this section, we will concentrate on the results and ignore a number of mathematical fine points.

If a function can be represented by orthonormal functions as in (3.62), then the coefficients $\{x_n\}$ must be determined as in (3.63), which is the same pair of relations as in (3.61). We can also represent the energy in $x(t)$ in terms of the coefficients $\{x_n; n \in \mathbb{Z}\}$. Since $|x^2(t)| = (\sum_n x_n \phi_n(t))(\sum_m x_m^* \phi_m^*(t))$, we get

$$\int |x^2(t)|dt = \int \sum_n \sum_m x_n x_m^* \phi_n(t)\phi_m^*(t)dt = \sum_n |x_n|^2. \tag{3.64}$$

Next suppose $x(t)$ is any square-integrable function and $\{\phi_n(t); n \in \mathbb{Z}\}$ is an orthonormal set. Let $x_n = \int x(t)\phi_n^*(t)dt$. Let $\epsilon_k(t) = x(t) - \sum_{n=1}^k x_n\phi_n(t)$ be the error when $x(t)$ is represented by the first k of these orthonormal functions. First we show that $\epsilon_k(t)$ is orthogonal to $\phi_m(t)$ for $1 \le m \le k$.

$$\int \epsilon_k(t)\phi_m^*(t)dt = \int x(t)\phi_m^*(t)dt - \int \sum_{n=1}^k x_n\phi_n(t)\phi_m^*(t)dt = x_m - x_m = 0. \tag{3.65}$$

Viewing functions as vectors, $\epsilon_k(t)$ is the difference between $x(t)$ and its projection on the linear subspace spanned by $\{\phi_n(t); 1 \le n \le k\}$. The integral of the magnitude squared error is given by

$$\int |x^2(t)|dt = \int \left|\epsilon_k(t) + \sum_{n=1}^k x_n\phi_n(t)\right|^2 dt \tag{3.66}$$

$$= \int |\epsilon_k^2(t)|dt + \int \sum_{n=1}^k \sum_{n=1}^k x_n x_m^* \phi_n(t)\phi_m^*(t)dt \tag{3.67}$$

$$= \int |\epsilon_k^2(t)|dt + \sum_{n=1}^k |x_n^2|. \tag{3.68}$$

Since $|\epsilon_k^2(t)|dt \ge 0$, the following inequality, known as Bessel's inequality, follows.

$$\sum_{n=1}^k |x_n^2| \le \int |x^2(t)|dt. \tag{3.69}$$

We see from (3.68) that $\int |\epsilon_k^2(t)|^2 dt$ is non-increasing with k. Thus, in the limit $k \to \infty$, either the energy in $\epsilon_k(t)$ approaches 0 or it approaches some positive constant. A set of orthonormal functions is said to *span* a class $\mathcal{C}$ of functions if this error energy approaches 0 for all $x(t) \in \mathcal{C}$. For example, the Fourier series set of functions in (3.60) spans the set of functions that are square integrable and zero outside of $[-T/2, T/2]$. There are many other countable sets of functions that span this class of functions and many others that span the class of square-integrable functions over $(-\infty, \infty)$.

In the next subsection, we use a sequence of independent Gaussian rv's as coefficients in these orthonormal expansions to generate a broad class of continuous-time Gaussian processes.

3.6.3 Continuous-time Gaussian processes

Given an orthonormal set of real-valued functions, $\{\phi_n(t);\, n \in \mathbb{Z}\}$ and given a sequence $\{X_n;\, n \in \mathbb{Z}\}$ of independent rv's[9] with $X_n \sim \mathcal{N}(0, \sigma_n^2)$, consider the following expression:

$$X(t) = \lim_{\ell \to \infty} \sum_{n=-\ell}^{\ell} X_n \phi_n(t). \tag{3.70}$$

Note that for any given t and ℓ, the sum above is a Gaussian rv of variance $\sum_{n=-\ell}^{\ell} \sigma_n^2 \phi_n^2(t)$. If this variance increases without bound as $\ell \to \infty$, then it is not hard to convince oneself that there cannot be a limiting distribution, so there is no limiting rv. The more important case of bounded variance is covered in the following theorem. Note that the theorem does not require the functions $\phi_n(t)$ to be orthonormal.

Theorem 3.6.3. *Let $\{X_n;\, n \in \mathbb{Z}\}$ be a sequence of independent rv's, $X_n \sim \mathcal{N}(0, \sigma_n^2)$ and let $\{\phi_n(t);\, n \in \mathbb{Z}\}$ be a sequence of real-valued functions. Assume that $\sum_{n=-\ell}^{\ell} \sigma_n^2 \phi_n^2(t)$ converges to a finite value as $\ell \to \infty$ for each t. Then $\{X(t);\, t \in \mathbb{R}\}$ as given in (3.70) is a Gaussian process.*

Proof: The difficult part of the proof is showing that $X(t)$ is a rv for any given t under the conditions of the theorem. This means that, for a given t, the sequence of rv's $\{\sum_{n=-\ell}^{\ell} X_n \phi_n(t);\, \ell \geq 1\}$ must converge WP1 to a rv as $\ell \to \infty$. This is proven in Section 9.9.2 as a special case of the martingale convergence theorem, so we simply accept that result for now. Since this sequence converges WP1, it also converges in distribution, so, since each term in the sequence is Gaussian, the limit is also Gaussian. Thus $X(t)$ exists and is Gaussian for each t.

Next, we must show that for any k, any $t_1, \ldots, t_k$, and any $a_1, \ldots, a_k$, the sum $a_1 X(t_1) + \cdots + a_k X(t_k)$ is Gaussian. This sum, however, is just the limit

$$\lim_{\ell \to \infty} \sum_{n=-\ell}^{\ell} [a_1 X_n \phi_n(t_1) + \cdots + a_k X_n \phi_n(t_k)].$$

This limit exists and is Gaussian by the same argument as used above for $k = 1$. Thus the process is Gaussian. □

Example 3.6.4. First consider an almost trivial example. Let $\{\phi_n(t);\, n \in \mathbb{Z}\}$ be a sequence of unit pulses each of unit duration, *i.e.*, $\phi_n(t) = 1$ for $n \leq t < n+1$ and $\phi_n(t) = 0$ elsewhere. Then $X(t) = X_{\lfloor t \rfloor}$. In other words, we have converted the discrete-time process $\{X_n;\, n \in \mathbb{Z}\}$ into a continuous time process simply by maintaining the value of X_n as a constant over each unit interval.

Note that $\{X_n;\, n \in \mathbb{Z}\}$ is stationary as a discrete-time process, but the resulting continuous-time process is non-stationary because the covariance of two points within a unit interval differs from that between the two points shifted so that an integer lies between them.

[9] Previous sections have considered possibly complex orthonormal functions, but we restrict them here to be real. Using rv's (which are real by definition) with complex orthonormal functions is an almost trivial extension, but using complex rv's and complex functions is less trivial and is treated in Section 3.7.8.

Example 3.6.5 (The Fourier series expansion). Consider the real-valued orthonormal functions in the sine/cosine form of the Fourier series over an interval $[-T/2, T/2)$, *i.e.*,

$$\phi_n(t) = \begin{cases} \sqrt{2/T}\cos(2\pi nt/T) & \text{for } n > 0,\ |t| \le T/2 \\ \sqrt{1/T} & \text{for } n = 0,\ |t| \le T/2 \\ \sqrt{2/T}\sin(-2\pi nt/T) & \text{for } n < 0,\ |t| \le T/2 \\ 0 & \text{for } |t| > T/2 \end{cases}.$$

If we represent a real-valued function $x(t)$ over $(-T/2,\, T/2)$ as $x(t) = \sum_n x_n \phi_n(t)$, then the coefficients x_n and x_{-n} essentially represent how much of the frequency n/T is contained in $x(t)$. If an orchestra plays a particular chord during $(-T/2,\, T/2)$, then the corresponding coefficients of X_n will tend to be larger in magnitude than the coefficients of frequencies not in the chord. If there is considerable randomness in what the orchestra is playing then these coefficients might be modeled as rv's.

When we represent a zero-mean Gaussian process, $X(t) = \sum_n X_n \phi_n(t)$, by these orthonormal functions, then the variances σ_n^2 signify, in some sense that will be refined later, how the process is distributed between different frequencies. We assume for this example that the variances σ_n^2 of the X_n satisfy $\sum_n \sigma_n^2 < \infty$, since this is required to ensure that $\mathsf{E}\left[X^2(t)\right]$ is finite for each t. The only intention of this example is to show, first, that a Gaussian process can be defined in this way, second that joint probability densities over any finite set of epochs, $-T/2 < t_1 < t_2 < \cdots < t_n < T/2$ are in principle determined by $\{\sigma_n^2;\, n{\in}\mathbb{Z}\}$, and third, that these variances have some sort of relation to the frequency content of the Gaussian process.

The above example is very nice if we want to model noise over some finite time interval. As suggested in Section 3.6.1, however, we often want to model noise as being stationary over $(-\infty,\, \infty)$. Neither the interval $(-T/2,\, T/2)$ nor its limit as $T \to \infty$ turn out to be very productive in this case. The next example, based on the sampling theorem of linear systems, turns out to work much better.

3.6.4 Gaussian sinc processes

The sinc function is defined to be $\mathrm{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$ and is sketched in Figure 3.7.

The Fourier transform of $\mathrm{sinc}(t)$ is a square pulse that is 1 for $|f| \le 1/2$ and 0 elsewhere. This can be verified easily by taking the inverse transform of the square pulse. The most remarkable (and useful) feature of the sinc function is that it and its translates over integer intervals form an orthonormal set, *i.e.*,

$$\int \mathrm{sinc}(t-n)\mathrm{sinc}(t-k)\, dt = \delta_{nk} \qquad \text{for } n, k \in \mathbb{Z}. \tag{3.71}$$

This can be verified (with effort) by direct integration, but the following approach is more insightful: the Fourier transform of $\mathrm{sinc}(t-n)$ is $e^{-i2\pi nf}$ for $|f| \le 1/2$ and is 0 elsewhere.

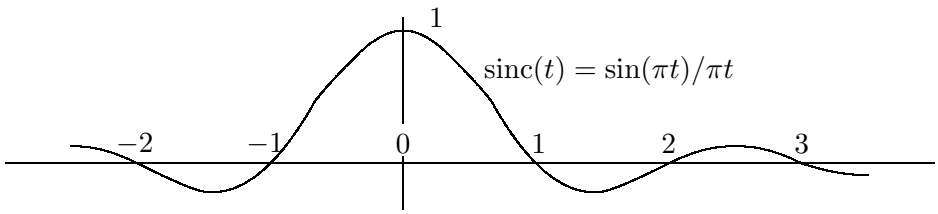


Figure 3.7: The function sinc(t) is 1 at $t = 0$ and 0 at every other integer t. The amplitude of its oscillations goes to 0 with increasing $|t|$ as $1/|t|$

Thus the Fourier transform of sinc$(t-n)$ is easily seen to be orthonormal to that of sinc$(t-k)$ for $n \neq k$. By Parseval's theorem, then, sinc$(t-n)$ and sinc$(t-k)$ are themselves orthonormal for $n \neq k$.

If we now think of representing any square-integrable function of *frequency*, say $v(f)$ over the frequency interval $(-1/2, 1/2)$ by a Fourier *series*, we see that $v(f) = \sum_n v_n e^{i2\pi nf}$ over $f \in (-1/2, 1/2)$, where $v_n = \int_{-1/2}^{1/2} v(f) e^{-i2\pi nf}\, df$. Taking the inverse Fourier *transform* we see that any function of time that is frequency limited to $(-1/2, 1/2)$ can be represented by the set $\{\text{sinc}(t-n);\, n \in \mathbb{Z}\}$. In other words, if $x(t)$ is a square-integrable continuous[10] function whose Fourier transform is limited to $f \in [-1/2, 1/2]$, then

$$x(t) = \sum_n x_n \text{sinc}(t-n) \qquad \text{where } x_n = \int x(t)\text{sinc}(t-n)\, dt \tag{3.72}$$

There is one further simplification that occurs here: for any integer value of t, say $t = k$, sinc$(t-n) = \delta_{kn}$, so $x(n) = x_n$. Thus for any square-integrable continuous function, limited in frequency to $[-1/2, 1/2]$,

$$x(t) = \sum_n x(n)\text{sinc}(t-n) \tag{3.73}$$

This sinc function expansion (better known as the sampling theorem expansion) is much more useful when it is linearly scaled in time, replacing the functions sinc$(t-n)$ with sinc$(2Bt-n)$ for some given bandwidth $B > 0$ (see Figure 3.8). The set of functions $\{\text{sinc}(2Bt-n);\, n \in \mathbb{Z}\}$ is still an orthogonal set, but the scaling in time by a factor of $(2B)^{-1}$ causes the squared integral to become $(2B)^{-1}$. Since the scaling by $(2B)^{-1}$ in time causes a scaling of $2B$ in frequency, these orthogonal function are now limited in frequency to $[-B, B]$. The argument above, applied to this scaled orthogonal set, leads to the well known sampling theorem:

[10]The reason for requiring continuity here is that a function can be altered at a finite (or even countable) number of points without changing its Fourier transform. The inverse transform of the Fourier transform of a bandlimited function, however, is continuous and is the function referred to. It is the same as the original function except at those originally altered points. The reader who wants a more complete development here is referred to [10].

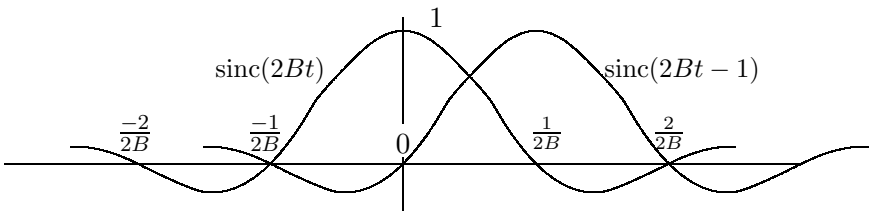


Figure 3.8: The function $\mathrm{sinc}(2Bt)$ is 1 at $t = 0$ and 0 at every other integer multiple of $(2B)^{-1}$. The function $\mathrm{sinc}(2Bt-1)$ is 1 at $t = (2B)^{-1}$ and 0 at every other integer multiple of $(2B)^{-1}$

Theorem 3.6.4. *Let $x(t)$ be a continuous square-integrable real or complex function of $t \in \mathbb{R}$ which is limited in frequency to $[-B, B]$ for any given $B > 0$. Then*

$$x(t) = \sum_n x\Big(\frac{n}{2B}\Big)\,\mathrm{sinc}(2Bt - n) \tag{3.74}$$

This theorem adds precision to the notion that any well-behaved function of a real variable can be approximated by its samples, saying that if the function is frequency limited, then sufficiently close samples represent the function perfectly when the points between the samples are filled in by this sinc expansion.

Now suppose that $\{X_n;\, n \in \mathbb{Z}\}$ is a sequence of IID Gaussian rv's and consider the following *Gaussian sinc process*,

$$X(t) = \sum_{-\infty}^{\infty} X_n \mathrm{sinc}(2Bt - n); \qquad \text{where } X_n \sim \mathcal{N}(0, \sigma^2) \tag{3.75}$$

The following theorem shows that the Gaussian sinc process of (3.75) is indeed a Gaussian process, calculates its covariance function, and shows that the process is stationary.

Theorem 3.6.5. *The Gaussian sinc process $\{X(t);\, t \in \mathbb{R}\}$ in (3.75) is a stationary Gaussian process with*

$$K_X(t) = \sigma^2 \mathrm{sinc}(2Bt). \tag{3.76}$$

Proof: From (3.75), we have

$$\begin{aligned}
K_X(t,\tau) &= \mathsf{E}\left[\left(\sum_n X_n \mathrm{sinc}(2Bt - n)\right)\left(\sum_k X_k \mathrm{sinc}(2B\tau - k)\right)\right] \\
&= \mathsf{E}\left[\sum_n X_n^2 \,\mathrm{sinc}(2Bt - n)\mathrm{sinc}(2B\tau - n)\right] && (3.77)\\
&= \sigma^2 \sum_n \mathrm{sinc}(2Bt - n)\mathrm{sinc}(2B\tau - n) && (3.78)\\
&= \sigma^2 \mathrm{sinc}(2B(t - \tau)), && (3.79)
\end{aligned}$$

where (3.77) results from $\mathsf{E}[X_n X_k] = 0$ for $k \neq n$ and (3.78) results from $\mathsf{E}\left[X_n^2\right] = \sigma^2$ for all n. To establish the identity between (3.78) and (3.79), let $y(t) = \text{sinc}(2B(t-\tau))$ for any given τ. The Fourier transform of $y(t)$ is $Y(f) = \sqrt{(2B)^{-1}}\exp(-i2\pi B\tau f)$ for $-B \leq f \leq B$ and 0 elsewhere. Thus $y(t)$ is frequency limited to $[-B, B]$ and therefore satisfies (3.74), which is the desired identity.

Now note that $K_X(t,t) = \sigma^2 = \sigma^2 \sum_n \text{sinc}^2(2Bt - n)$. Thus this series converges, and from Theorem 3.6.3, $\{X(t);\, t \in \mathbb{R}\}$ is a Gaussian process. Finally, since the covariance depends only on $t - \tau$, the process is stationary and the covariance in single variable form is $K_X(t) = \sigma^2 \text{sinc}(2Bt)$. □

3.6.5 Filtered Gaussian sinc processes

Many important applications of stochastic processes involve linear filters where the filter input is one stochastic process and the output is another. The filter might be some physical phenomenon, or it might be a filter being used to detect or estimate something from the input stochastic process. It might also be used simply to demonstrate the existence of a stochastic process with certain properties. In this section, we restrict our attention to the case where the input stochastic process is the Gaussian sinc process described in the previous section. We then show that the output is a stationary Gaussian process and find its covariance function. Figure 3.9 illustrates the situation.

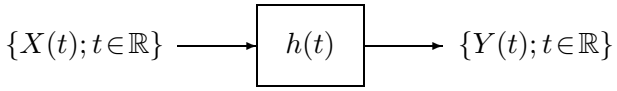


Figure 3.9: A stochastic process $\{X(t);\, t \in \mathbb{R}\}$ is the input to a linear time-invariant filter, and the output is another stochastic process. A WSS input leads to a WSS output and a Gaussian input leads to a Gaussian output.

A linear time-invariant filter with impulse response $h(t)$ creates a linear transformation from an input function $x(t)$ to an output function y(t) defined by $y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)\,d\tau$. In other words, the output at time t is a linear combination of the inputs over all time. The time invariance refers to the property that if the input function is translated by a given d, then the output function is translated by the same d.

In many situations, $h(t)$ is restricted to be *realizable*, meaning that $h(t) = 0$ for $t < 0$. This indicates that the output at a given t is a function only of the inputs up to and including t. In other situations, the filtering is done 'off-line,' meaning that the entire function $x(t)$ is available before performing the filtering. In some cases, the time reference at the filter output might have a delay of d relative to that at the input. This often occurs when a communication channel is subject to both filtering and propagation delay, and in these cases, $h(t)$ might be non-zero for all $t \geq -d$; this can still be regarded as a realizable filter, since only the time reference at the output has been altered.

In this section, we assume that $h(t)$ has a Fourier transform that is 0 for $|f| > B$, where B

is the bandwidth of the input Gaussian sinc process. We shall find later that this implies that the filter is non-realizable. This is of no concern here since our purpose is simply to characterize a family of Gaussian processes at the filter output.

Suppose a stochastic process $\{X(t);\, t \in \mathbb{R}\}$ is the input to a linear time-invariant (LTI) filter. Let Ω be the underlying sample space and let ω be a sample point of Ω. The corresponding sample function of the process $\{X(t);\, t \in \mathbb{R}\}$ is then $X(t,\omega)$. The output of the LTI filter with impulse response $h(t)$ and input $X(t,\omega)$ is given by

$$Y(t,\omega) = \int_{-\infty}^{\infty} X(\tau,\omega)h(t-\tau)d\tau.$$

If the integrals exist for each ω, this (in principle) defines a rv for each t and thus (in principle) defines a stochastic process $\{Y(t);\, t \in \mathbb{R}\}$. Developing a theory of integration for a continuum of rv's is quite difficult[11] and would take us too far afield. This is why we are concentrating on stochastic processes that can be represented as orthonormal expansions using a sequence of rv's as coefficients. The next section generalizes this to other input processes that can also be represented as orthogonal expansions.

If we express the input sinc process $X(t)$ as in (3.75), then the output process is given by

$$Y(t) = \int_{-\infty}^{\infty} \sum_n X_n \mathrm{sinc}(2B\tau - n)h(t-\tau)\, d\tau \quad \text{where } X_n \sim \mathcal{N}(0,\sigma^2). \tag{3.80}$$

Assuming that the integration and summation can be interchanged, we see that

$$\begin{aligned}\int_{-\infty}^{\infty} \mathrm{sinc}(2B\tau - n)\, h(t-\tau)\, d\tau &= \int_{-\infty}^{\infty} \mathrm{sinc}(2B\tau)\, h(t - \frac{n}{2B} - \tau)\, d\tau \\ &= \frac{1}{2B}\, h\left(t - \frac{n}{2B}\right), \end{aligned} \tag{3.81}$$

where we have viewed the convolution as a product in the frequency domain and used the fact that the transform of the sinc function is constant over $[-B,B]$ and that $H(f)$ is zero outside that range. Thus, substituting (3.81) into (3.80) we have

$$Y(t) = \sum_n \frac{X_n}{2B}\; h\left(t - \frac{n}{2B}\right). \tag{3.82}$$

From Theorem 3.6.3, if $\sum_n h^2(t-n/2B)$ is finite for each t, then $\{Y(t);\, t \in \mathbb{R}\}$ is a Gaussian process (and the previous interchange of integration and summation is justified). Exercise 3.22 shows that $\sum_n h^2(t-n/2B) = \int_{-\infty}^{\infty} h^2(\tau)\, d\tau$ for each t. This shows that if $h(t)$ is square integrable, then $Y(t)$ is a Gaussian process.

In the next section, we show that when a WSS stochastic process is filtered by an LTI filter, the output is also WSS. Thus our final result is that if the X_n are IID and $\mathcal{N}(0,\sigma^2)$ and if $h(t)$ is square integrable and bandlimited to $[-B,B]$, then $Y(t)$ in (3.82) is a stationary Gaussian process. We discuss this further in the next section.

[11] Readers who believe that stochastic processes are sort of like ordinary functions and can be integrated and understood in the same way should look at Example 3.6.6.

3.6.6 Filtered continuous-time stochastic processes

As discussed in the previous section, if a sample function $X(t,\omega)$ is the input to an LTI filter with impulse response $h(t)$, then the output sample function (assuming that the integral converges) is $Y(t,\omega) = \int X(\tau,\omega)h(t-\tau)\,d\tau$. If $\{X_n;\, n \in \mathbb{Z}\}$ is a sequence of IID rv's, $\{\phi_n(t);\, n \in \mathbb{Z}\}$ is a sequence of orthonormal functions, and $X(t) = \sum_n X_n\phi_n(t)$, then we can visualize $Y(t)$ as an output stochastic process described as

$$Y(t) = \int_{-\infty}^{\infty} X(\tau)h(t-\tau)d\tau = \int_{-\infty}^{\infty} \sum_n X_n\phi_n(\tau)h(t-\tau)d\tau \tag{3.83}$$

There are some mathematical issues about whether the infinite summation in the final expression converges and whether the integral converges, but we saw how to treat these questions for the case in which $X(t)$ is a Gaussian sinc process and $h(t)$ is bandlimited and square integrable. More broadly, if the X_n are IID Gaussian, we saw how to use Theorem 3.6.3 to show that the output process is also Gaussian if $Y(t)$ has a finite variance for all t.

We will ignore these convergence questions in more general cases for the time being, and also use the middle expression in (3.83), viewing it as shorthand for the final expression.

In what follows, we will find the covariance function $K_Y(t,\tau)$ of the output process in terms of the covariance function K_X of the input and the impulse response of the filter. We also introduce and interpret the spectral density if the process is WSS. We will simply assume that any needed limits exist, although a few examples will be given later where more care is needed.

Assume throughout that the input stochastic process, $\{X(t);\, t\in\mathbb{R}\}$, is real and zero mean and that $h(t)$ is real. It follows then that $Y(t)$ is real and zero-mean. If we rewrite $\int X(\tau)h(t-\tau)\,d\tau$ as $\int X(t-\tau)h(\tau)\,d\tau$, then the covariance function of $Y(t)$ can be expressed as

$$K_Y(t,u) = \mathsf{E}\left[\int_{-\infty}^{\infty} X(t-\tau)h(\tau)d\tau \int_{-\infty}^{\infty} X(u-s)h(s)\,ds\right] \tag{3.84}$$

Interchanging expectation and integration,

$$K_Y(t,u) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_X(t-\tau,\, u-s)h(\tau)h(s)\,d\tau ds \tag{3.85}$$

This equation is valid whether or not X is WSS. Assuming that X is WSS, we can rewrite $K_X(t-\tau,\, u-s)$ in the single argument form as $K_X(t-u-\tau+s)$,

$$K_Y(t,u) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_X(t-u-\tau+s)h(\tau)h(s)d\tau\, ds \tag{3.86}$$

This is a function only of $t-u$, showing that Y is WSS. Thus $K_Y(t,u)$ can be written in the single argument form $K_Y(t-u)$. Replacing $t-u$ by v, we have

$$K_Y(v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_X(v-\tau+s)h(\tau)h(s)d\tau\, ds \tag{3.87}$$

We now interpret the right hand side of (3.87) as the convolution of three functions. To do this, we first rewrite (3.87) as

$$K_Y(v) = \int_s h(s) \left[\int_\tau K_X(v+s-\tau)h(\tau)\, d\tau \right] ds \tag{3.88}$$

The term in brackets is the convolution of h and K_X evaluated at $v+s$, which we denote as $[h * K_X](v+s)$. Now define $h_b(s) = h(-s)$. That is, h_b is h reversed in time. Replacing s with $-s$, (3.88) becomes

$$K_Y(v) = \int_s h_b(s)[h * K_X](v-s)\, ds = [h_b * h * K_X](v) \tag{3.89}$$

One of the simplest and best known results of linear systems is that convolution in the time domain corresponds to multiplication in the Fourier transform domain. This leads us to define spectral density.

Definition 3.6.5. *The spectral density $S_Y(f)$ of a WSS stochastic process $\{Y(t);\, t \in \mathbb{R}\}$ is the Fourier transform of its covariance function $K_Y(t)$,* i.e.*,*

$$S_Y(f) = \int_{-\infty}^{\infty} K_Y(t) e^{-i2\pi ft}\, dt \tag{3.90}$$

We now express (3.89) in terms of spectral densities. Let $H(f)$ be the Fourier transform of the impulse response $h(t)$,

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-i2\pi ft}\, dt \tag{3.91}$$

The Fourier transform of the backward impulse response, $h_b(t)$ is then

$$H_b(f) = \int_{-\infty}^{\infty} h_b(t) e^{-i2\pi ft}\, dt = \int_{-\infty}^{\infty} h(\tau) e^{i2\pi f\tau}\, d\tau = H^*(f) \tag{3.92}$$

The transform of (3.89) is then

$$S_Y(f) = H^*(f)H(f)S_X(f) = |H(f)|^2 S_X(f) \tag{3.93}$$

Thus the covariance function of $Y(t)$ is best expressed as the inverse Fourier transform, $\mathcal{F}^{-1}[|H(f)|^2 S_X(f)]$.

3.6.7 Interpretation of spectral density and covariance

First note that the covariance function of a real WSS process must be real and symmetric around 0. Thus $S_X(f)$ and $S_Y(f)$ are real and symmetric around 0. Also, since $h(t)$ is real, $|H(f)|^2$ must be real and symmetric around 0, which is also implied by (3.93).

Now consider a very narrow band filter around some given frequency f_0. In particular, assume a filter with frequency response

$$H(f) = \begin{cases} 1; & \text{for } f_0 - \epsilon/2 \le |f| \le f_0 + \epsilon/2 \\ 0; & \text{elsewhere} \end{cases}.$$

If we pass a zero-mean WSS stochastic process $\{X(t)\}$ through this filter, then from (3.93),

$$S_Y(f) = \begin{cases} S_X(f); & \text{for } f_0 - \epsilon/2 \le |f| \le f_0 + \epsilon/2 \\ 0; & \text{elsewhere} \end{cases} . \tag{3.94}$$

The expected power out of this filter, i.e., $\mathsf{E}\left[Y^2(t)\right] = K_Y(0)$, is independent of t because Y is WSS. Since $K_Y(t)$ is the inverse Fourier transform of $S_Y(f)$, $K_Y(0) = \mathsf{E}\left[Y^2(t)\right]$ is given by

$$\mathsf{E}\left[Y^2(t)\right] = \int S_Y(f)\, df \approx 2\epsilon\, S_X(f_0). \tag{3.95}$$

where we assume that $S_X(f)$ is continuous and ϵ is so small that $S_X(f) = S_X(-f_0)$ does not vary appreciably from $f_0 - \epsilon/2$ to $f_0 + \epsilon/2$. This means that the expected output power from the filter is proportional to $S_X(f_0)$. This output power can then be interpreted as the input power over the range of frequencies $\pm(f_0 - \epsilon/2,\, f_0 + \epsilon/2)$. Since this is proportional to 2ϵ (the aggregate range of positive and negative frequencies in the filter passband), we interpret spectral density as the power per unit frequency in the WSS process. This also says (with a little care about points of discontinuity in $S_X(f)$) that $S_X(f) \ge 0$ for all f.

Now consider the class of filtered Gaussian sinc processes again. For a Gaussian sinc process bandlimited to B, we have $S_X(f) = \sigma^2/2B$ for $|f| \le B$ and $S_X(f) = 0$ for $|f| > B$. If this is filtered with frequency response $H(f)$, bandlimited to B, then from (3.93)), we have

$$S_Y(f) = \begin{cases} \frac{\sigma^2 |H(f)|^2}{2B} & \text{for } f \le B, \\ 0 & \text{for } f > B \end{cases}$$

There are several remarkable features about this. First, the covariance function $K_Y(t)$ is determined by this spectral density, and since $Y(t)$ is Gaussian, this determines all the joint distributions of $Y(t)$, *i.e.*, it determines the stochastic process, aside from possible interactions with other stochastic processes.

Second, the spectral density, and thus all joint CDF's depend on $H(f)$ only through $|H(f)|^2$. A given choice for $|H(f)|^2$ can leave significant freedom in choosing $H(f)$, but that freedom does not change the joint probability distributions of the process.

Third, since there are essentially no constraints on $|H(f)|$ other than being nonnegative and limited to B, any desired stationary Gaussian process bandlimited to B can be chosen in this way. Since B is arbitrary, this limitation does not at first appear significant, although this will be discussed shortly.

Fourth, we have seen that $S_Y(f) \ge 0$ for all f for all WSS stochastic processes. If we restrict our attention to WSS processes for which $K_Y(0) < \infty$ (and we really can't make much sense out of other processes), then $\int S_Y(f)\, df < \infty$, so it follows that we can approximate $S_Y(f)$ by a bandlimited choice with large enough B. Since any bandlimited choice for $S_Y(f) \ge 0$ is the spectral density of a filtered Gaussian sinc process, we see that the only constraint on $K_Y(t)$ to be a covariance function is that $K_Y(0) < \infty$ and $S_Y(f) \ge 0$ for all f. Furthermore

any such spectral density can be approximated (in some sense) by the spectral density of a filtered Gaussian sinc process.

The trouble with filtered Gaussian sinc processes is that realizable filters cannot be bandlimited. In fact, the Paley-Wiener theorem (see [19]) says says that a necessary and sufficient condition on the Fourier transform for a non-zero square-integrable function $h(t)$ to be 0 for all $t < 0$ is that $\int \frac{|\ln|H(f)||}{1+f^2}\, df < \infty$. This is more than an obscure mathematical issue, since it turns out that bandlimited stationary Gaussian processes have some peculiar properties even though their spectral densities closely approximate those of processes that are not bandlimited. We will not resolve these issues here, and readers are urged to exercise care when approximating non-bandlimited processes by bandlimited processes.

The purpose of this section has not been to imply that filtered Gaussian sinc processes provide a universally acceptable way to deal with continuous-time stationary Gaussian processes. Rather these banlimited processes provide a great deal of insight into more general stationary Gaussian processes and, with care, can be used to solve a large number of engineering problems concerning these processes.

3.6.8 White Gaussian noise

Physical noise processes are often well-modeled as stationary Gaussian processes, as we have pointed out earlier. Often they also have the characteristic that their spectral density is quite flat over the bandwidths of interest in a given situation. In this latter situation, we can simplify and idealize the model by assuming that the spectral density is constant over all frequencies. This idealization is called *white Gaussian noise*. Unfortunately, this simplification comes at a steep price — the power in such a process $\{W(t);\, t \in \mathbb{R}\}$ is

$$\mathsf{E}\left[|W(t)|^2\right] = K_W(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_W(f)df \;=\; \infty \tag{3.96}$$

Thus $W(t)$ at any given t can not even be approximated as a Gaussian rv. On the other hand, if a stationary Gaussian process has spectral density $S_X(f)$ and is filtered with frequency function $H(f)$, then, from (3.93) the output process $Y(t)$ has spectral density $S_Y(f) = |H(f)|^2 S_X(f)$. If $S_X(f)$ is flat over the range of f where $H(f)$ is non-zero, then the output Gaussian process $\{Y(t);\, t \in \mathbb{R}\}$ has the same joint distributions no matter what $S_X(f)$ is outside the bandwidth of $H(f)$.

If a Gaussian noise process is looked at only through the outputs of various linear filters, and if its spectral density is constant over the frequency response of each filter, then we might as well assume that the process is white and not concern ourselves with the spectral density outside of the range of interest. Since measurement devices generally involve filtering to some extent (even if flat over such a broad bandwidth that it can usually be ignored), this view of white Gaussian noise as an idealization is usually the only view that is physically meaningful.

In summary then, white Gaussian noise is an idealization of a Gaussian process with spectral density[12] $S_W(f) = N_0/2$ over all f such that $|f| \le B$ where B is larger than all frequencies

[12] $N_0/2$ is the standard term among engineers to denote the spectral density of white noise. Spectral

of interest. In the limit $B \to \infty$, the covariance function of white noise can be taken to be $(N_0/2)\delta(t)$ where $\delta(t)$ is the Dirac unit impulse. This is a generalized function, roughly defined by the property that for any well-behaved function $a(t)$, we have $\int a(t)\delta(t)\,dt = a(0)$. We can visualize $\delta(t)$ as an idealization of a narrow pulse of unit area, narrow relative to the smallest interval over which any $a(t)$ of interest can change. With a little thought, It can be seen that this is just a another way of saying, as before, that white Gaussian noise is an idealization of a stationary Gaussian noise process whose spectral density is constant over the frequency range of interest.

One of the nice features of white Gaussian noise is that we can view any zero-mean stationary Gaussian process as a filtered version of white Gaussian noise. That is, a zero-mean stationary Gaussian process $\{Y(t);\, t \in R\}$ with spectral density $S_Y(f)$ can be viewed as white noise of unit spectral density passed through a filter with frequency response $H(f)$ such that $|H(f)|^2 = S_X(f)$. Recall that this view was quite valuable in studying Gaussian vectors, and it is equally valuable here.

It almost appears that white Gaussian noise can be viewed as the limit of a sequence of Gaussian sinc processes where process ℓ has bandwidth B_ℓ and power $\mathsf{E}\left[X^2(t)\right] = B_\ell N_0$. Thus the spectral density for process ℓ is $N_0/2$ for $|f| \le B_\ell$. For any realizable filter with frequency response $H(f)$, we have observed from the Paley-Wiener theorem that $H(f)$ can only approach 0 at a limited rate as $f \to \infty$. Thus there is no B large enough that white noise filtered by $H(f)$ is quite the same as a filtered sinc Gaussian process, although it could certainly be approximated that way.

The following two examples show that we have slightly oversimplified matters in viewing zero-mean stationary Gaussian processes as being characterized by their spectral densities.

Example 3.6.6 (Pathological barely visible Gaussian noise). Consider a stationary Gaussian process $\{X(t)\,;\, t \in \mathbb{R}\}$ for which $X(t) \sim \mathcal{N}(0,1)$ for each $t \in \mathbb{R}$. Assume that $X(t)$ and $X(\tau)$ are independent for all t, τ with $\tau \neq t$. Thus $K_X(t,\tau)$ is 1 for $t = \tau$ and 0 otherwise. This process is Gaussian and stationary, and its single-variable covariance function $K_X(t)$ is 1 for $t = 0$ and 0 elsewhere. It follows that $S_X(f) = 0$ for all f. Also, if we express $X(t)$ in terms of any set of orthonormal functions, we see that $\int X(t)\phi_n(t)\,dt = 0$ WP1 for all n. In the same way, if $X(t)$ is passed through any square-integrable linear filter, the output process is 0 WP1 for all t. Thus in a very real sense, this Gaussian process is effectively 0. From a physical point of view, one could never observe such a process, because any physical measurement requires some type of averaging over a very small but non-zero interval of time. The sample-average measurement over any such interval would then be 0 WP1.

We can compare this pathological process to a sequence of Gaussian sinc processes with bandwidths $B_1 \le B_2, \ldots \to \infty$ as before. Here, however, we take the power in each process to be 1. Thus the spectral density of the ℓth process is $(2B_\ell)^{-1}$ for $|f| \le B_\ell$, so the spectral density at each f approaches 0 with increasing ℓ. As before, however, there isn't any decent kind of limit for the process,

density is the power per unit frequency in a process, counting positive and negative frequencies separately. Thus if we look at the power in a bandwidth B, i.e., a frequency interval of width B in positive frequencies and another B in negative frequencies, the noise power in that band is $N_0 B$.

There are a number of broad-band communication systems where the transmitted channel waveform can be roughly modeled as a Gaussian sinc process with large bandwidth and negligible spectral density. Such a process appears almost non-existant to other communication systems, but as will be seen in Chapter 8 on detection, the signals can still be detected, in contrast to the strange process here.

The strangeness of the pathological process in this example arises largely from the fact that the covariance function is not continuous. Exercise 3.25 shows that if a WSS process has a covariance function $K_X(t)$ that is continuous at $t = 0$, then it is continuous everywhere. A large part of the theory for constructing orthonormal expansions for continuous random processes depends on a continuous covariance function. From a more application oriented viewpoint, the properties arising from discontinuities in the covariance can not be observed (as in the example here). Thus a continuous covariance function is almost always assumed.

Example 3.6.7 (Pathological invisible Gaussian noise). Let $X \sim \mathcal{N}(0,1)$ and let Y be a uniform rv over $(0,1]$. Let $Z(t)$ be a stochastic process where $Z(Y) = X$ and $Z(t) = 0$ for all $t \neq Y$. Now for any given t, the probability that $Y = t$ is 0, so $Z(t) = 0$ WP1. Thus $Z(t)$ can be viewed as Gaussian with variance 0 and, of course, $K_Z(t) = 0$ for all t.

This is more pathological than the previous example, but is important in showing that specifying the covariance function of a Gaussian process does not fully specify the process. Here every sample function of this process is discontinuous, whereas the conventional zero function is continuous. For any finite set of epochs $t_1, \ldots, t_k$, we see that $X(t_1), \ldots, X(t_k)$ are all 0 WP1, so these joint distributions do not distingish this process from the all-zero process. The difference between this process and the all-zero process could never be measured, so essentially, the covariance function, and thus the set of finite joint distributions, specify a process. However, as shown here, more intricate tools (common sense for engineers and measure theory for mathematicians) are needed to make sense of continuity for these sample functions.

3.6.9 The Wiener process / Brownian motion

Recall that one of the major properties of the Poisson counting process (see Chapter 2) is that it has stationary and independent increments. These properties can be be defined for arbitrary stochastic processes as well as for counting processes. They are fundamental properties of the Wiener process, which is also known as Brownian motion.[13]

Definition 3.6.6. *Let a stochastic process $\{X(t);\, t \in \mathcal{T}\}$ be defined over a set of epochs $\mathcal{T}$ where $\mathcal{T}$ is either the nonnegative reals or nonnegative integers. Then $\{X(t);\, t \in \mathcal{T}\}$ has stationary increments if for any epochs $t_1 < t_2$, the increment $X(t_2) - X(t_1)$ has the same distribution as $X(t_2 - t_1) - X(0)$.*

Definition 3.6.7. *Let a stochastic process $\{X(t);\, t \in \mathcal{T}\}$ be defined over a set of epochs $\mathcal{T}$ where $\mathcal{T}$ is either the nonnegative reals or nonnegative integers. Then $\{X(t);\, t \geq 0\}$ has*

[13] Brownian motion is a standard model for the motion of small particles in a gas. Norbert Wiener substantially developed its mathematical analysis. We will discuss only the one-dimensional version of the process.

independent increments if for any sequence of epochs $t_1 < t_2 < t_3 < \cdots < t_k$, the random variables

$$\big[X(t_2) - X(t_1)\big],\ \big[X(t_3) - X(t_2)\big],\ \ldots,\big[X(t_k) - X(t_{k-1})\big]$$

are statistically independent.

Now consider an arbitrary process $\{X(t);\, t \geq 0\}$ with independent and stationary increments and with $X(0) = 0$. Let Δ be an arbitrary increment size and, for an arbitrary positive integer n, write $X(n\Delta)$ as

$$X(n\Delta) = \big[X(n\Delta) - X((n-1)\Delta)\big] + \big[X((n-1)\Delta) - X((n-2)\Delta)\big] + \cdots + \big[X(\Delta) - X(0)\big].$$

Because of this, we see that $\mathsf{E}[X(n\Delta)] = n\mathsf{E}[X(\Delta)]$ and $\mathsf{VAR}[X(n\Delta)] = n\,\mathsf{VAR}[X(\Delta)]$. Thus the mean and variance of $X(t)$ must each be linear in t. Because of the independent increments, we can also see that $K_X(t,\tau)$, for any $\tau \geq t$, is equal to $\mathsf{VAR}[X(t)]$. We summarize this in the following theorem.

Theorem 3.6.6. *Let $\{X(t);\, 0 \leq t\}$ have independent and stationary increments and let $X(0) = 0$. Then for any epochs t and $\tau > t$,*

$$\mathsf{E}[X(t)] = t\,\mathsf{E}[X(1)]\,; \qquad K_X(t,\tau) = t\,\mathsf{VAR}[X(1)] \tag{3.97}$$

One interesting consequence of this is that (except in the uninteresting case of zero variances) processes with independent and stationary increments cannot be stationary. That is, $\{X(t);\, t \geq 0\}$ has stationary increments if the *changes* $X(t) - X(t-\Delta)$ do not depend probabilistically on t, whereas (essentially) the process is stationary if the *process values themselves*, $X(t)$ do not depend probabilistically on t. Another consequence is that these processes are not meaningful over the entire time interval from $-\infty$ to $+\infty$. This is because the variance is growing linearly with t and must remain nonnegative for all epochs t.

The restriction that $X(0) = 0$ in the theorem is inessential, and the extension to the case where $X(0)$ is an arbitrary rv is contained in Exercise 3.20.

Definition 3.6.8 (The Wiener process / Brownian motion). *A Wiener process is a zero-mean Gaussian process $\{X(t);\, t \geq 0\}$ which has stationary and independent increments, satisfies $X(0) = 0$, and has continuous sample functions WP1.*

The continuity restriction rules out the addition of 'invisible' processes such as that in Example 3.6.7 to a continuous process with stationary and independent increments. See Feller [14] for a proof that Wiener processes exist. Given this existence, we see that $\mathsf{E}[X(t)] = 0$ for all $t \geq 0$ and also, from Theorem 3.6.6, $K_X(t,\tau) = \min(t,\tau)\sigma^2$ where $\sigma^2 = \mathsf{E}\big[X^2(1)\big]$. Since a zero-mean Gaussian process is essentially specified by its covariance function, we see that the Wiener process is essentially specified by the single parameter σ^2. Also, since the covariance function has been derived using only the stationary and incependent increments property, we see that the added assumption about continuity is not required for specifying all the joint CDF's of the process.

A type of continuity also follows directly from the independent and stationary increment property without the added requirement of continuity in the definition of a Wiener process. Consider the increment $X(t+\Delta) - X(t)$ for very small Δ. The increment has the variance $\Delta\sigma^2$, and by the Chebyshev inequality,

$$\Pr\{X(t+\Delta) - X(t) > \epsilon\} \leq \frac{\Delta\sigma^2}{\epsilon^2}.$$

This means that as $\Delta \to 0$, the probability that $X(t)$ changes by more than ϵ goes to zero. To understand the type of continuity implied by this, consider the sequence $X(t+\Delta), X(t+\Delta/2), X(t+\Delta/3), \ldots$ for some given t, Δ. This sequence approaches $X(t)$ in probability (see Section 1.7.2). This is weaker, of course, than the continuity of sample functions WP1 required in the definition of a Wiener process.

Strangely enough, although the sample functions of a Wiener process have these continuity properties, they are essentially not differentiable. To see this, note that $[X(t+\Delta)-X(t)]/\Delta$ has variance σ^2/Δ. This goes to ∞ as $\Delta \to 0$. Despite these strange properties, the Wiener process is widely used by engineers, and often provides sound insights into real issues.

The Poisson counting process and Wiener process are similar in the sense that both are, in a sense, modeling an "integral" of independent objects. In the Poisson case, we are interested in random point arrivals. If we view a sample function of these arrivals as a sequence of unit impulses, then the corresponding sample function of the counting process is the integral of that impulse chain. The Wiener process models an accumulation or integral of individually small but very dense independent disturbances (noise). One can envision the process being integrated as white Gaussian noise, although, as we have seen, the derivative of the Wiener process does not exist and white Gaussian noise does not exist except as a generalized form of stochastic process.

We now show that the Wiener process can be viewed as a limit of a sum of IID rv's if the limit uses the appropriate kind of scaling. Let $\{Y_n;\, n \geq 1\}$ be a sequence of zero-mean IID rv's each with finite variance σ^2. Consider a sequence of processes $\{X_\ell(t);\, t \geq 0\}$ where the ℓth process is defined in terms of $\{Y_n;\, n \geq 1\}$ by

$$X_\ell(t) = \sum_{k=1}^{\lfloor 2^\ell t \rfloor} 2^{-\ell/2} Y_k.$$

Then $\mathsf{E}[X_\ell(t)] = 0$ and $\mathsf{E}\left[X_\ell^2(t)\right] = \sigma^2 t$, where we are ignoring the difference between $\lfloor 2^\ell t \rfloor$ and $2^\ell t$.

Note that each unit increase in ℓ doubles the number of IID rv's added together in each unit of time. Note also that the magnitude of the IID rv's are scaled down by the square root of this rate doubling. Thus the variance of the scaled sum for a given t remains constant as ℓ increases. By the CLT, the distribution of this scaled sum approaches the Gaussian. It is easy to see that the covariance of $X_\ell(t)$ approaches that of the Wiener process (in fact, it is only the integer approximation $\lfloor 2^\ell t \rfloor \approx 2^\ell t$ that is involved in the covariance).

We don't want to address the issue of how a limit of a sequence of stochastic processes approaches another process. The important thing is that a sum of (finite variance zero-

mean) rv's can be modeled as a Wiener process with the appropriate scaling; this explains why the Wiener process appears in so many applications.

This completes our discussion of (real) Gaussian processes. The next section discusses the complex case.

3.7 Circularly-symmetric complex random vectors

Many of the (real-valued) waveforms used for communication and other purposes have the property that their Fourier transforms are 0 except in two relatively narrow bands of frequencies, one around a positive carrier frequency f_0, and the other around $-f_0$. Such waveforms are often represented as

$$x(t) = z_{\rm re}(t)\cos(2\pi f_0 t) + z_{\rm im}(t)\sin(2\pi f_0 t) = \Re\left[z(t)e^{-2\pi i f_0 t}\right], \qquad i=\sqrt{-1}.$$

Representing $x(t)$ in terms of a complex 'baseband waveform' $z(t) = z_{\rm re}(t) + iz_{\rm im}(t)$ or in terms of two real baseband waveforms, $z_{\rm re}(t)$ and $z_{\rm im}(t)$ is often convenient analytically, since if the bandwidth is small and f_0 is large, then $z(t)$ changes slowly relative to $x(t)$, while still specifying the waveform exactly for a given f_0.

The same relationship, $X(t) = \Re[Z(t)\exp[-2\pi i f_0 t]$, is equally convenient for a stochastic process rather than an individual waveform in a limited bandwidth. Note however that $\sin(2\pi f_0 t)$ is the same as $\cos(2\pi f_0 t)$ except for a small delay, $1/(4f_0)$. Normally, we would not expect the statistics of the noise to be sensitive to this small delay; in more graphic terms, we would not expect the noise to 'know' where our time reference $t = 0$ is. Thus we often model bandpass noise so that $Z_{\rm re}(t)$ and $Z_{\rm im}(t)$ are identically distributed. By extending this slightly, we often model bandpass noise so that $Z(t)$ and $Z(t)e^{-i\theta}$ are identically distributed for all phases θ. More specifically, we often model bandpass noise so that for each $t_1, t_2, \ldots, t_n$, the joint distribution of the complex random vector $(Z(t_1), \ldots, Z(t_n))^{\mathsf{T}}$ is the same as that of $(Z(t_1)e^{i\theta}, \ldots, Z(t_n)e^{i\theta})^{\mathsf{T}}$ for each real θ.

The purpose of the above argument is not to convince the reader that this joint distribution property is 'necessary' for band-pass noise, but simply to motivate why this kind of phase invariance, which is called circular symmetry, might be useful to understand. The results here are widely used in many branches of engineering, mathematics, and physics, but not widely accessible in a systematic form.

3.7.1 Circular symmetry and complex Gaussian rv's

Definition 3.7.1. *A complex rv $Z = Z_{\rm re} + iZ_{\rm im}$ is* Gaussian *if $Z_{\rm re}$ and $Z_{\rm im}$ are jointly Gaussian; Z is* circularly symmetric *if Z and $Ze^{i\theta}$ have the same distribution for all real θ.*

Note that if Z has a PDF and is circularly symmetric, then the PDF is constant on any circle centered on the origin. If Z is Gaussian, then its equal probability contours are ellipses; these are circular and centered on the origin if and only if $Z_{\rm re}$ and $Z_{\rm im}$ are IID zero-mean

Gaussian. The amplitude $|Z|$ of a circularly-symmetric Gaussian rv is Rayleigh-distributed and the phase is uniformly distributed.

If we multiply a circularly-symmetric rv Z by a complex constant c, then the amplitude of cZ is the product of the amplitudes of Z and c; the phase is the sum of the individual phases. It is intuitively clear (from the original uniform phase of Z) that such an addition of phases maintains the circular symmetry.

A circularly-symmetric Gaussian rv Z is fully described by its variance, $\sigma^2 = \mathsf{E}[ZZ^*] = \mathsf{E}\left[|Z|^2\right]$. The complex conjugate is necessary in the definition of variance, since $\mathsf{E}[ZZ^*] = \mathsf{E}\left[Z_{\rm re}^2\right] + \mathsf{E}\left[Z_{\rm im}^2\right]$ whereas $\mathsf{E}\left[Z^2\right] = \mathsf{E}\left[Z_{\rm re}^2\right] - \mathsf{E}\left[Z_{\rm im}^2\right]$.

Just as a Gaussian rv X of mean a and variance σ^2 is described as $X \sim \mathcal{N}(a, \sigma^2)$, a circularly-symmetric Gaussian rv Z of variance σ^2 is described 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. The terminology allows for a complex rv with a mean a and a fluctuation that is circularly-symmetric Gaussian to be referred to as $\mathcal{CN}(a, \sigma^2)$.

Definition 3.7.2. *An n-dimensional complex random vector (complex n-rv) $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$ is* 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 Gaussian and circularly symmetric.*

Example 3.7.1. An important example of a circularly-symmetric Gaussian r**v** is $\boldsymbol{W} = (W_1, \ldots, W_n)^{\mathsf{T}}$ where the components $W_k, 1 \le k \le n$ 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, $e^{i\theta}\boldsymbol{W}$ 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 (being careful about the factors of $1/2$) is

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{1}{\pi^n} \exp\left[\sum_{k=1}^{n} -|w_k|^2\right], \tag{3.98}$$

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.

3.7.2 Covariance and pseudo-covariance of complex n-rv's

We saw in Section 3.3.4 that the distribution of a real zero-mean Gaussian n-r**v** (*i.e.*, a vector with jointly-Gaussian components) is completely determined by its covariance matrix. Here we will find that the distribution of a *circularly-symmetric* Gaussian n-r**v** is also determined by its covariance matrix. *Without circular symmetry, the covariance matrix is not sufficient to determine the distribution.* In order to understand this, we first define both the covariance matrix and the pseudo-covariance matrix of a complex n-r**v**.

Definition 3.7.3. *The* covariance matrix $[K_{\boldsymbol{Z}}]$ *and the* pseudo-covariance matrix $[M_{\boldsymbol{Z}}]$ *of a zero-mean complex n-r**v** $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$ are the $n \times n$ matrices of complex components*

given respectively by

$$[K_{\boldsymbol{Z}}] = \mathsf{E}\left[\boldsymbol{Z}\boldsymbol{Z}^\dagger\right] \qquad [M_{\boldsymbol{Z}}] = \mathsf{E}\left[\boldsymbol{Z}\boldsymbol{Z}^T\right], \tag{3.99}$$

where $\boldsymbol{Z}^\dagger$ *is the the complex-conjugate of the transpose,* i.e., $\boldsymbol{Z}^\dagger = \boldsymbol{Z}^{T*}$.

As shown below, $[K_{\boldsymbol{Z}}]$ and $[M_{\boldsymbol{Z}}]$ determine the covariance matrix of the real $2n$-r$\mathbf{v}$ $\begin{bmatrix} \boldsymbol{Z}_{\rm re} \\ \boldsymbol{Z}_{\rm im} \end{bmatrix}$.

$$\begin{aligned}
\mathsf{E}\left[\Re(Z_k)\Re(Z_j)\right] &= \frac{1}{2}\Re\left([K_{\boldsymbol{Z}}]_{kj} + [M_{\boldsymbol{Z}}]_{kj}\right), \\
\mathsf{E}\left[\Im(Z_k)\Im(Z_j)\right] &= \frac{1}{2}\Re\left([K_{\boldsymbol{Z}}]_{kj} - [M_{\boldsymbol{Z}}]_{kj}\right), \\
\mathsf{E}\left[\Re(Z_k)\Im(Z_j)\right] &= \frac{1}{2}\Im\left(-[K_{\boldsymbol{Z}}]_{kj} + [M_{\boldsymbol{Z}}]_{kj}\right), \\
\mathsf{E}\left[\Im(Z_k)\Re(Z_j)\right] &= \frac{1}{2}\Im\left([K_{\boldsymbol{Z}}]_{kj} + [M_{\boldsymbol{Z}}]_{kj}\right)
\end{aligned} \tag{3.100}$$

If $\boldsymbol{Z}$ is also Gaussian, this shows that $[K_{\boldsymbol{Z}}]$ and $[M_{\boldsymbol{Z}}]$ together specify not only the covariance but also the distribution of $\begin{bmatrix} \boldsymbol{Z}_{\rm re} \\ \boldsymbol{Z}_{\rm im} \end{bmatrix}$, and thus specify the distribution of $\boldsymbol{Z}$. We next start to connect the notion of circular symmetry with the pseudo-covariance matrix.

Lemma 3.7.1. *Let* $\boldsymbol{Z}$ *be a circularly-symmetric complex* n*-r*$\mathbf{v}$*. Then the pseudo-covariance matrix satisfies* $[M_{\boldsymbol{Z}}] = 0$ *.*

Proof: Since $\boldsymbol{Z}$ and $e^{i\theta}\boldsymbol{Z}$ have the same joint distribution for any given θ, they have the same pseudo-covariance matrix, *i.e.*, $[M_{e^{i\theta}\boldsymbol{Z}}] = [M_{\boldsymbol{Z}}]$. Denote the j,ℓ component of $[M_{e^{i\theta}\boldsymbol{Z}}]$ as $[M_{e^{i\theta}\boldsymbol{Z}}]_{j,\ell}$. Then

$$[M_{e^{i\theta}\boldsymbol{Z}}]_{j,\ell} = \mathsf{E}\left[e^{i\theta}Z_j \cdot e^{i\theta}Z_\ell\right] = e^{i2\theta}[M_Z]_{j\ell}.$$

For $i = \pi/2$ then, $[M_{\boldsymbol{Z}}]_{j,\ell} = -[M_{\boldsymbol{Z}}]_{j,\ell}$. Thus $[M_{\boldsymbol{Z}}]_{j,\ell} = 0$ for all j,ℓ. □

In general, $[M_{\boldsymbol{Z}}] = 0$ is not enough to ensure that $\boldsymbol{Z}$ is circularly symmetric. For example, in the one dimensional case, if $Z_{\rm re}$ and $Z_{\rm im}$ are IID, binary equiprobable (1, -1), then $[M_Z] = 0$ but Z is obviously not circularly symmetric. The next theorem, however, shows that $[M_{\boldsymbol{Z}}] = 0$ is enough in the Gaussian case.

Theorem 3.7.1. *Let* $\boldsymbol{Z}$ *be a zero-mean complex Gaussian* n*–r*$\mathbf{v}$*. Then* $[M_{\boldsymbol{Z}}] = 0$ *if and only if* $\boldsymbol{Z}$ *is circularly-symmetric Gaussian.*

Proof: The lemma shows that $[M_{\boldsymbol{Z}}] = 0$ if $\boldsymbol{Z}$ is circularly-symmetric. For the only-if side, assume $[M_{\boldsymbol{Z}}] = 0$. Then $[M_{e^{i\theta}\boldsymbol{Z}}] = 0$ also, so $[M_{\boldsymbol{Z}}] = [M_{e^{i\theta}\boldsymbol{Z}}]$.

We must next consider $[K_{e^{i\theta}\boldsymbol{Z}}]$. The j,ℓ component of this matrix for any j,ℓ is

$$\mathsf{E}\left[e^{i\theta}Z_k \cdot e^{-i\theta}Z_\ell^*\right] = \mathsf{E}\left[Z_k \cdot Z_\ell^*\right] = [K_Z]_{j\ell}.$$

Thus, $[K_{e^{i\theta}\boldsymbol{Z}}] = [K_{\boldsymbol{Z}}]$, so $e^{i\theta}\boldsymbol{Z}$ has the same covariance and pseudo-covariance as $\boldsymbol{Z}$.

Since $e^{i\theta}\boldsymbol{Z}$ and $\boldsymbol{Z}$ are each zero-mean complex Gaussian, each distribution is specified by its covariance and pseudo-covariance. Since these are the same, $e^{i\theta}\boldsymbol{Z}$ and $\boldsymbol{Z}$ must have the same distribution. This holds for all real θ, so $\boldsymbol{Z}$ is circularly-symmetric Gaussian. □

Since $[M_{\boldsymbol{Z}}]$ is zero for any circularly-symmetric Gaussian n-r**v** $\boldsymbol{Z}$, the distribution of $\boldsymbol{Z}$ is determined solely by $[K_{\boldsymbol{Z}}]$ and is denoted as $\boldsymbol{Z} \sim \mathcal{CN}(0, [K_{\boldsymbol{Z}}])$ where $\mathcal{C}$ denotes that $\boldsymbol{Z}$ is both complex and circularly symmetric. The complex normalized IID r**v** of Example 3.7.1 is thus denoted as $\boldsymbol{W} \sim \mathcal{CN}(0, [I_n])$.

The following two examples illustrate some subtleties in Theorem 3.7.1.

Example 3.7.2. Let $\boldsymbol{Z} = (Z_1, Z_2)^{\mathsf{T}}$ where $Z_1 \sim \mathcal{CN}(0,1)$ and $Z_2 = XZ_1$ where X 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 the distribution of $\Im(Z_1)$ and $\Im(Z_2)$ is similarly concentrated. Thus, $\boldsymbol{Z}$ is not Gaussian. Even though Z_1 and Z_2 are individually circularly-symmetric Gaussian, $\boldsymbol{Z}$ is not circularly-symmetric Gaussian according to the definition. In this example, it turns out that $\boldsymbol{Z}$ *is* circularly symmetric and $[M_{\boldsymbol{Z}}] = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$. The example can be changed slightly, changing the definition of Z_2 to $\Re(Z_2) = X\Re(Z_1)$ and $\Im(Z_2) \sim \mathcal{N}(0, 1/2)$, where $\Im(Z_2)$ is statistically independent of all the other variables. Then $[M_{\boldsymbol{Z}}]$ is still 0, but $\boldsymbol{Z}$ is not circularly symmetric. Thus, without the jointly-Gaussian property, the relation between circular symmetry and $[M_{\boldsymbol{Z}}] = 0$ is not an if-and-only-if relation.

Example 3.7.3. 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 Gaussian and the theorem applies. We see that $[M_{\boldsymbol{Z}}] = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, and thus $\boldsymbol{Z}$ is 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 3.7.2, 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 3.7.2 it was the lack of a jointly-Gaussian distribution.

3.7.3 Covariance matrices of complex n-rv

The covariance matrix of a complex n-r**v** $\boldsymbol{Z}$ is $[K_{\boldsymbol{Z}}] = \mathsf{E}\left[\boldsymbol{Z}\boldsymbol{Z}^\dagger\right]$. The properties of these covariance matrices are quite similar to those for real n-r**v** except that $[K_{\boldsymbol{Z}}]$ is no longer symmetric ($K_{kj} = K_{jk}$), but rather is *Hermitian*, defined as a square matrix $[K]$ for which $K_{kj} = K_{jk}^*$ for all j, k. These matrices are analyzed in virtually the same way as the symmetric matrices considered in Section 3.4.1, so we simply summarize the results we need here.

If $[K]$ is Hermitian, then the eigenvalues are all real and the eigenvectors $\boldsymbol{q}_j$ and $\boldsymbol{q}_k$ of distinct eigenvalues are orthogonal in the sense that $\boldsymbol{q}_j^\dagger \boldsymbol{q}_k = 0$. Also if an eigenvalue has multiplicity ℓ, then ℓ orthogonal eigenvectors can be chosen for that eigenvalue.

The eigenvalues $\lambda_1, \ldots, \lambda_n$, repeating each distinct eigenvalue according to its multiplicity, can be used as the elements of a diagonal matrix $[\Lambda]$. To each λ_j, we can associate an eigenvector $\boldsymbol{q}_j$ where the eigenvectors are chosen to be orthonormal ($\boldsymbol{q}_j^\dagger \boldsymbol{q}_k = \delta_{jk}$). Letting $[Q]$ be the matrix with orthonormal columns[14] $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$, we have the relationship

$$[K] = [Q\Lambda Q^{-1}] \qquad [Q^\dagger] = [Q^{-1}] \qquad \text{for } [K] \text{ Hermitian.} \tag{3.101}$$

An $n \times n$ Hermitian matrix $[K]$ is positive semi-definite if, for all complex n-vectors $\boldsymbol{b}$, the equation $\boldsymbol{b}^\dagger [K] \boldsymbol{b} \geq 0$ holds. It is positive definite if $\boldsymbol{b}^\dagger [K] \boldsymbol{b} > 0$ for all $\boldsymbol{b} \neq 0$. By the same argments as in the real case, we have the following lemma:

Lemma 3.7.2. *If $\boldsymbol{Z}$ is a complex n-r$\mathbf{v}$ with covariance matrix $[K]$, then $[K]$ satisfies (3.101) and is positive semi-definite. It is positive definite if $[K]$ is non-singular. Also, for any complex $n \times n$ matrix $[A]$, the matrix $[AA^\dagger]$ is positive semi-definite and is positive definite if $[A]$ is non-singular. For any positive semi-definite $[K]$, there is a square-root matrix $[R] = [Q\sqrt{\Lambda}Q^{-1}]$ as given in (3.101) such that $\boldsymbol{Z} = [R]\,\boldsymbol{W}$ (where $\boldsymbol{W} \sim \mathcal{CN}(0, I)$) is circularly-symmetric Gaussian with $[K_{\boldsymbol{Z}}] = [K]$.*

We have seen that the major change in going from real n-r$\mathbf{v}$'s to complex n-r$\mathbf{v}$'s is a judicious conversion of transposes into complex-conjugate transposes.

3.7.4 Linear transformations of $\boldsymbol{W} \sim \mathcal{CN}(0, [I_\ell])$

One of the best ways to understand real Gaussian n-r$\mathbf{v}$'s is to view them as linear transformations of an ℓ-r$\mathbf{v}$ (for given ℓ) with IID components, each $\mathcal{N}(0,1)$. The same approach turns out to work equally well for circularly-symmetric Gaussian vectors. Thus let $[A]$ be an arbitrary complex $n \times \ell$ matrix and let the complex n-r$\mathbf{v}$ $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\mathsf{T}$ be defined by

$$\boldsymbol{Z} = [A]\,\boldsymbol{W} \qquad \text{where } \boldsymbol{W} \sim \mathcal{CN}(0, [I_\ell]) \tag{3.102}$$

The complex n-r$\mathbf{v}$ defined by this complex linear transformation has jointly Gaussian real and imaginary parts. To see this, represent the complex n-dimensional transformation in (3.102) by the following $2n$ dimensional real linear transformation:

$$\begin{bmatrix} \boldsymbol{Z}_{\rm re} \\ \boldsymbol{Z}_{\rm im} \end{bmatrix} = \begin{bmatrix} [A_{\rm re}] & -[A_{\rm im}] \\ [A_{\rm im}] & [A_{\rm re}] \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_{\rm re} \\ \boldsymbol{W}_{\rm im} \end{bmatrix}, \tag{3.103}$$

where $\boldsymbol{Z}_{\rm re} = \Re(\boldsymbol{Z})$, $\boldsymbol{Z}_{\rm im} = \Im(\boldsymbol{Z})$, $[A_{\rm re}] = \Re([A])$, and $[A]_{\rm im} = \Im([A])$. By definition, real linear transformations on real IID Gaussian r$\mathbf{v}$'s have jointly-Gaussian components. Thus $\boldsymbol{Z}_{\rm re}$ and $\boldsymbol{Z}_{\rm im}$ are jointly Gaussian and $\boldsymbol{Z}$ is a complex Gaussian n-r$\mathbf{v}$.

[14] A square complex matrix with orthonormal columns is said to be *unitary*. Viewed as a transformation, $Q\boldsymbol{z}$ has the same length as $\boldsymbol{z}$ where the length of $\boldsymbol{z}$ is $\sqrt{\boldsymbol{z}^\dagger \boldsymbol{z}}$.

The r**v** $\boldsymbol{Z}$ is also circularly symmetric.[15] To see this, note that

$$[K_{\boldsymbol{Z}}] = \mathsf{E}\left[[A]\,\boldsymbol{W}\,\boldsymbol{W}^\dagger[A^\dagger]\right] = [AA^\dagger] \qquad [M_{\boldsymbol{Z}}] = \mathsf{E}\left[[A]\,\boldsymbol{W}\,\boldsymbol{W}^\mathsf{T}[A^\mathsf{T}]\right] = 0 \tag{3.104}$$

Thus, from Theorem 3.7.1, $\boldsymbol{Z}$ is circularly-symmetric Gaussian and $\boldsymbol{Z} \sim \mathcal{CN}(0, [AA^\dagger])$.

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

Theorem 3.7.2. *A complex r***v** $\boldsymbol{Z}$ *is circularly-symmetric Gaussian if and only if it can be expressed as* $\boldsymbol{Z} = [A]\,\boldsymbol{W}$ *for a complex matrix* $[A]$ *and an IID circularly-symmetric Gaussian r***v** $\boldsymbol{W} \sim \mathcal{CN}(0, [I])$.

Proof: For the *only if* part, choose $[A]$ to be the square root matrix $[R]$ of Lemma 3.7.2. Then $\boldsymbol{Z} = [R]\,\boldsymbol{W}$ is circularly-symmetric Gaussian with $[K_{\boldsymbol{Z}}] = [RR^\dagger]$ □

We now have three equivalent characterizations for circularly-symmetric Gaussian n-r**v**'s. First, phase invariance, second, zero pseudo-covariance, and third, linear transformations of IID circularly symmetric Gaussian vectors. One advantage of the third characterization is that the jointly-Gaussian requirement is automatically met, whereas the other two depend on that as a separate requirement. Another advantage of the third characterization is that the usual motivation for modeling r**v**'s as circularly-symmetric Gaussian is that they are linear transformations of essentially IID circularly-symmetric Gaussian random vectors.

3.7.5 Linear transformations of $\boldsymbol{Z} \sim \mathcal{CN}(0, [K])$

Let $\boldsymbol{Z} \sim \mathcal{CN}(0, [K])$. If some other random vector $\boldsymbol{Y}$ can be expressed as $\boldsymbol{Y} = [B]\boldsymbol{Z}$, then $\boldsymbol{Y}$ is also a circularly-symmetric Gaussian random vector. To see this, represent $\boldsymbol{Z}$ as $\boldsymbol{Z} = [A]\,\boldsymbol{W}$ where $\boldsymbol{W} \sim \mathcal{CN}(0, [I])$. Then $\boldsymbol{Y} = [BA]\,\boldsymbol{W}$, so $\boldsymbol{Y} \sim \mathcal{CN}(0, [BKB^\dagger])$. This helps show why circular symmetry is important — it is invariant to linear transformations.

If $[B]$ is 1 by n (*i.e.*, if it is a row vector $\boldsymbol{b}^\mathsf{T}$) then $Y = \boldsymbol{b}^\mathsf{T}\boldsymbol{Z}$ is a complex rv. Thus all linear combinations of a circularly-symmetric Gaussian random vector are circularly-symmetric Gaussian rv's.

Conversely, we now want to show that if all linear combinations of a complex random vector $\boldsymbol{Z}$ are circularly-symmetric Gaussian, then $\boldsymbol{Z}$ must also be circularly-symmetric Gaussian. The question of being Gaussian can be separated from that of circular symmetry. Thus assume that for all complex n-vectors $\boldsymbol{b}$, the complex rv $\boldsymbol{b}^\mathsf{T}\boldsymbol{Z}$ is complex Gaussian. It follows that $\Re(\boldsymbol{b}^\mathsf{T}\boldsymbol{Z}) = \boldsymbol{b}_{\rm re}^\mathsf{T}\boldsymbol{Z}_{\rm re} - \boldsymbol{b}_{\rm im}^\mathsf{T}\boldsymbol{Z}_{\rm im}$ is a real Gaussian rv for all choices of $\boldsymbol{b}_{\rm re}$ and $\boldsymbol{b}_{\rm im}$. Thus from Theorem 3.4.7, the real $2n$-r$\boldsymbol{v}$ $\begin{bmatrix} \boldsymbol{Z}_{\rm re} \\ \boldsymbol{Z}_{\rm im} \end{bmatrix}$ is a Gaussian $2n$-r**v**. By defininition, then, $\boldsymbol{Z}$ is complex Gaussian.

We could now show that $\boldsymbol{Z}$ is also circularly-symmetric Gaussian if $\boldsymbol{b}^\mathsf{T}\boldsymbol{Z}$ is circularly-symmetric for all $\boldsymbol{b}$, but it is just as easy, and yields a slightly stronger result, to show that if $\boldsymbol{Z}$ is Gaussian and the pairwise linear combinations $Z_j + Z_k$ are circularly-symmetric for all j, k, then $\boldsymbol{Z} \sim \mathcal{CN}(0, [K_{\boldsymbol{Z}}])$. If $Z_j + Z_j$ is circularly symmetric for all j, then $\mathsf{E}\left[Z_j^2\right] = 0$,

[15] Conversely, as shown later, all circularly-symmetric Gaussian r**v**'s can be defined this way.

so that the main diagonal of $[M_{\boldsymbol{Z}}]$ is zero. If in addition, $Z_j + Z_k$ is circularly symmetric, then $\mathsf{E}\left[(Z_j + Z_k)^2\right] = 0$. But since $\mathsf{E}\left[Z_j^2\right] = \mathsf{E}\left[Z_k^2\right] = 0$, we must have $2\mathsf{E}\left[Z_j Z_k\right] = 0$. Thus the j,k element of $[M_{\boldsymbol{Z}}] = 0$. Thus if $Z_j + Z_k$ is circularly symmetric for all j,k, it follows that $[M_{\boldsymbol{Z}}] = 0$ and $\boldsymbol{Z}$ is circularly symmetric.[16] Summarizing,

Theorem 3.7.3. *A complex random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\mathsf{T}}$ is circularly-symmetric Gaussian if and only if all linear combinations of $\boldsymbol{Z}$ are complex Gaussian and $Z_j + Z_k$ is circularly symmetric for all j,k.*

3.7.6 The PDF of circularly-symmetric Gaussian n-rv's

Since the probability density of a complex random variable or vector is defined in terms of the real and imaginary parts of that variable or vector, we now pause to discuss these relationships. The major reason for using complex vector spaces and complex random vectors is to avoid all the detail of the real and imaginary parts, but our intuition comes from $\mathbb{R}^2$ and $\mathbb{R}^3$, and the major source of confusion in treating complex random vectors comes from assuming that $\mathbb{C}^n$ is roughly the same as $\mathbb{R}^n$. This assumption causes additional confusion when dealing with circular symmetry.

Assume that $\boldsymbol{Z} \sim \mathcal{CN}(0, [K_{\boldsymbol{Z}}])$, and let $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{Z}_{\rm re} \\ \boldsymbol{Z}_{\rm im} \end{bmatrix}$ be the corresponding real $2n$-r$\mathbf{v}$. Let $[K_{\boldsymbol{U}}]$ be the covariance of the real $2n$-r$\mathbf{v}$ $\boldsymbol{U}$. From (3.100), with $[M_Z] = 0$, we can express $[K_{\boldsymbol{U}}]$ as

$$[K_{\boldsymbol{U}}] = \begin{bmatrix} \frac{1}{2}[K_{\rm re}] & -\frac{1}{2}[K_{\rm im}] \\ \frac{1}{2}[K_{\rm im}] & \frac{1}{2}[K_{\rm re}] \end{bmatrix}, \tag{3.105}$$

where $[K_{\rm re}]$ is the $n \times n$ matrix whose components are the real parts of the components of $[K_{\boldsymbol{Z}}]$ and correspondingly $[K_{\rm im}]$ is the matrix of imaginary parts.

Now suppose that $(\lambda, \boldsymbol{q})$ is an eigenvalue, eigenvector pair for $[K_{\boldsymbol{Z}}]$. Separating $[K_{\boldsymbol{Z}}]\boldsymbol{q} = \lambda \boldsymbol{q}$ into real and imaginary parts,

$$[K_{\rm re}]\boldsymbol{q}_{\rm re} - [K_{\rm im}]\boldsymbol{q}_{\rm im} = \lambda \boldsymbol{q}_{\rm re}\ ; \qquad [K_{\rm im}]\boldsymbol{q}_{\rm re} + [K_{\rm re}]\boldsymbol{q}_{\rm im} = \lambda \boldsymbol{q}_{\rm im}.$$

Comparing this with $[K_{\boldsymbol{U}}]\begin{bmatrix} \boldsymbol{q}_{\rm re} \\ \boldsymbol{q}_{\rm im} \end{bmatrix}$, where $[K_{\boldsymbol{U}}]$ is given in (3.105), we see that $\lambda/2$ is an eigenvalue of $[K_{\boldsymbol{U}}]$ with eigenvector $\begin{bmatrix} \boldsymbol{q}_{\rm re} \\ \boldsymbol{q}_{\rm im} \end{bmatrix}$. Furthermore,assuming that $\boldsymbol{q}$ is normalized over complex n-space, $\begin{bmatrix} \boldsymbol{q}_{\rm re} \\ \boldsymbol{q}_{\rm im} \end{bmatrix}$ is normalized over real $2n$-space. As a complex n-vector, $i\boldsymbol{q}$ (where $i = \sqrt{-1}$) is a complex scalar times $\boldsymbol{q}$. It is an eigenvector of $[K_{\boldsymbol{Z}}]$ but not independent of $\boldsymbol{q}$. The corresponding real $2n$-vector $\begin{bmatrix} -\boldsymbol{q}_{im} \\ \boldsymbol{q}_{\rm re} \end{bmatrix}$, is orthonormal to $\begin{bmatrix} \boldsymbol{q}_{\rm re} \\ \boldsymbol{q}_{\rm im} \end{bmatrix}$ and is also an eigenvector of $[K_{\boldsymbol{U}}]$. In addition, for any orthonormal complex n-vectors, the corresponding real $2n$-vectors are orthonormal. This establishes the following lemma.

[16] Example 3.7.3 showed that if $\boldsymbol{Z}$ is Gaussian with individually circularly symmetric components, then $\boldsymbol{Z}$ is not necessarily circularly-symmetric Gaussian. This shows that the only additional requirement is for $Z_k + Z_j$ to be circularly-symmetric for all k,j.

Lemma 3.7.3. *Let $(\lambda_1, \boldsymbol{q}_1), \ldots, (\lambda_n, \boldsymbol{q}_n)$ denote the n pairs of eigenvalues and orthonormal eigenvectors of the covariance matrix $[K_{\boldsymbol{Z}}]$ of a circularly-symmetric n-r$\mathbf{v}$ $\boldsymbol{Z}$. Then the real $2n$-r$\mathbf{v}$ $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{Z}_{\rm re} \\ \boldsymbol{Z}_{\rm im} \end{bmatrix}$ has a covariance matrix $[K_{\boldsymbol{U}}]$ with the $2n$ eigenvalue, orthonormal eigenvector pairs*

$$(\frac{\lambda_1}{2}, \begin{bmatrix} \boldsymbol{q}_{1,\rm re} \\ \boldsymbol{q}_{1,\rm im} \end{bmatrix}), \ldots, (\frac{\lambda_n}{2}, \begin{bmatrix} \boldsymbol{q}_{n,\rm re} \\ \boldsymbol{q}_{n,\rm im} \end{bmatrix}) \ (\frac{\lambda_1}{2}, \begin{bmatrix} -\boldsymbol{q}_{1,\rm im} \\ \boldsymbol{q}_{1,\rm re} \end{bmatrix}), \ldots, (\frac{\lambda_n}{2}, \begin{bmatrix} -\boldsymbol{q}_{n,\rm im} \\ \boldsymbol{q}_{n,\rm re} \end{bmatrix}). \tag{3.106}$$

Since the determinant of a matrix is the product of the eigenvalues, we see that

$$\det[K_{\boldsymbol{U}}] \ = \ \prod_{j=1}^{n} \left(\frac{\lambda_j}{2}\right)^2 \ = \ 2^{-2n}(\det[K_{\boldsymbol{Z}}])^2 \tag{3.107}$$

Recall that the probability density of $\boldsymbol{Z}$ (if it exists) is the same as the probability density of $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{Z}_{\rm re} \\ \boldsymbol{Z}_{\rm im} \end{bmatrix}$, *i.e.*, it is the probability density taken over the real and imaginary components of $\boldsymbol{Z}$. This plus (3.107) makes it easy to find the probability density for $\boldsymbol{Z}$ assuming that $\boldsymbol{Z} \sim \mathcal{CN}(0, [K_{\boldsymbol{Z}}])$.

Theorem 3.7.4. *Assume that $\boldsymbol{Z} \sim \mathcal{CN}(0, [K_{\boldsymbol{Z}}])$ and assume that $[K_{\boldsymbol{Z}}]$ is non-singular. Then the probability density of $\boldsymbol{Z}$ exists everywhere and is given by*

$$\mathsf{f}_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{\pi^n \det[K_{\boldsymbol{Z}}]} \exp(-\boldsymbol{z}^\dagger [K_{\boldsymbol{Z}}^{-1}] \boldsymbol{z}). \tag{3.108}$$

Proof: Since $[K_{\boldsymbol{Z}}]$ is non-singular, its eigenvalues are all positive, so the eigenvalues of $[K_{\boldsymbol{U}}]$ are also positive and $[K_{\boldsymbol{U}}]$ is non-singular. Since $\boldsymbol{Z}$ is circularly-symmetric Gaussian, $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{Z}_{\rm re} \\ \boldsymbol{Z}_{\rm im} \end{bmatrix}$ must be zero-mean Gaussian. Since $\boldsymbol{U}$ is a zero-mean Gaussian $2n$-r$\mathbf{v}$, its PDF is given from (3.22) as

$$\mathsf{f}_{\boldsymbol{U}}(\boldsymbol{u}) = \prod_{j=1}^{2n} \frac{1}{\sqrt{2\pi\mu_j}} \exp(-v_j^2/(2\mu_j), \tag{3.109}$$

where μ_j is the jth eigenvalue of $[K_{\boldsymbol{U}}]$ and $v_j = \boldsymbol{a}_j^{\mathsf{T}} \boldsymbol{u}_j$, where $\boldsymbol{a}_j$ is the jth orthonormal eigenvector of $[K_{\boldsymbol{U}}]$. We have seen that the eigenvalues λ_j of $[K_{\boldsymbol{Z}}]$ are related to those of $[K_{\boldsymbol{U}}]$ by $\mu_j = \lambda_j/2$ and $\mu_{j+n} = \lambda_j/2$ for $1 \le j \le n$. Similarly the eigenvectors can be related by $\boldsymbol{a}_j^{\mathsf{T}} = (\boldsymbol{q}_{{\rm re},j}^{\mathsf{T}}, \boldsymbol{q}_{{\rm im},j}^{\mathsf{T}})$ and $\boldsymbol{a}_{j+n}^{\mathsf{T}} = (-\boldsymbol{q}_{{\rm im},j}^{\mathsf{T}}, \boldsymbol{q}_{{\rm re},j}^{\mathsf{T}})$. With a little calculation, we get

$$\begin{aligned} v_j^2 + v_{j+n}^2 \ &= \ (\boldsymbol{q}_{{\rm re},j}^{\mathsf{T}} \boldsymbol{z}_{\rm re} + \boldsymbol{q}_{{\rm im},j}^{\mathsf{T}} \boldsymbol{z}_{\rm im})^2 + (-\boldsymbol{q}_{{\rm im},j}^{\mathsf{T}} \boldsymbol{z}_{\rm re} + \boldsymbol{q}_{{\rm re},j}^{\mathsf{T}} \boldsymbol{z}_{\rm im})^2 \\ &= \ [\Re(\boldsymbol{q}_j^\dagger \boldsymbol{z})]^2 + [\Im(\boldsymbol{q}_j^\dagger \boldsymbol{z})]^2 \ = \ |\boldsymbol{q}_j^\dagger \boldsymbol{z}|^2 \end{aligned}$$

Substituting this into (3.109) and recognizing that the density is now given directly in terms of $\boldsymbol{Z}$,

$$\begin{aligned} \mathsf{f}_{\boldsymbol{Z}}(\boldsymbol{z}) \ &= \ \prod_{j=1}^{n} \frac{1}{\pi\lambda_j} \exp(-|\boldsymbol{q}_j^\dagger \boldsymbol{z}|^2/(\lambda_j). \\ &= \ \frac{1}{\pi^n \det[K_{\boldsymbol{Z}}]} \exp(\sum_{j=1}^{n} -|\boldsymbol{q}_j^\dagger \boldsymbol{z}|^2/(\lambda_j). \end{aligned} \tag{3.110}$$

Finally, recalling that $\boldsymbol{q}_j$ is the jth column of $[Q]$,

$$\sum_{j=1}^{n} |\boldsymbol{q}_j^\dagger \boldsymbol{z}|^2/\lambda_j = \boldsymbol{z}^\dagger [Q\Lambda^{-1}Q^{-1}]\boldsymbol{z} = \boldsymbol{z}^\dagger K_{\boldsymbol{Z}}^{-1}\boldsymbol{z}.$$

Substituting this into (3.110) completes the proof. □

Note that (3.110) is also a useful expression for the density of circularly-symmetric Gaussian n-r**v**'s. The geometric picture is not as easy to interpret as for real zero-mean Gaussian n-r**v**'s, but the regions of equal density are still ellipsoids. In this case, however, $e^{i\theta}\boldsymbol{z}$ is on the same ellipsoid for all phases θ.

The following theorem summarizes circularly-symmetric Gaussian n-r**v**.

Theorem 3.7.5. *A complex n-r**v** $\boldsymbol{Z}$ is circularly-symmetric Gaussian if and only if any one of the following conditions is satisfied.*

- *$\boldsymbol{Z}$ is a Gaussian n-r**v** and has the same distribution as $e^{i\theta}\boldsymbol{Z}$ for all real ϕ.*
- *$\boldsymbol{Z}$ is a zero-mean Gaussian n-r**v** and the pseudo-covariance matrix $[M_{\boldsymbol{Z}}]$ is zero.*
- *$\boldsymbol{Z}$ can be expressed as $\boldsymbol{Z} = [A]\,\boldsymbol{W}$ where $\boldsymbol{W} \sim \mathcal{CN}(0, [I_n])$.*
- *For non-singular $[K_{\boldsymbol{Z}}]$, the probability density of $\boldsymbol{Z}$ is given in (3.108). For singular $[K_{\boldsymbol{Z}}]$, (3.108) gives the density of $\boldsymbol{Z}$ after removal of the deterministically dependent components.*
- *All linear combinations of $\boldsymbol{Z}$ are complex Gaussian and $Z_j + Z_k$ is circularly symmetric for all j, k.*

Note that either all or none of these conditions are satisfied. The significance of the theorem is that any one of the conditions may be used to either establish the circularly-symmetric Gaussian property or to show that it does not hold. We have also seen (in Lemma 3.7.2) that if K is the covariance matrix for any complex n-r**v**, then it is also the covariance matrix of some circularly-symmetric Gaussian n-r**v**.

3.7.7 Conditional PDF's for circularly symmetric Gaussian rv's

It turns out that conditional PDF's for circularly-symmetric Gaussian r**v**'s are virtually the same as those for real-valued r**v**'s. Operationally, the only difference is that transposes must be replaced with Hermitian conjugates and the basic form for the unconditional real Gaussian PDF must be replaced with the basic form of the unconditional complex circularly symmetric Gaussian PDF. This is not obvious without going through all the calculations used to find conditional PDF's for real r**v**'s, but the calculations are virtually the same, so we won't repeat them here. We simply repeat and discuss Theorem 3.5.1, modified as above.

Theorem 3.7.6. *Let $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)^\mathsf{T}$ be circularly symmetric and jointly Gaussian with the non-singular covariance matrix $[K]$ partitioned into $n+m$ columns and rows as*

$$[K] = \begin{bmatrix} [K_{\boldsymbol{X}}] & [K_{\boldsymbol{X}\cdot\boldsymbol{Y}}] \\ [K^\dagger_{\boldsymbol{X}\cdot\boldsymbol{Y}}] & [K_{\boldsymbol{Y}}] \end{bmatrix} \qquad [K^{-1}] = \begin{bmatrix} [B] & [C] \\ [C^\dagger] & [D] \end{bmatrix}$$

Then the joint PDF of $\boldsymbol{X} = (X_1, \ldots, X_n)^\mathsf{T}$ conditional on $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^\mathsf{T}$ is given by

$$\mathsf{f}_{\boldsymbol{X}|\,\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) = \frac{\exp\left\{-\Big(\boldsymbol{x} + [B^{-1}C]\,\boldsymbol{y}\Big)^\dagger\,[B]\,\Big(\boldsymbol{x} + [B^{-1}C]\,\boldsymbol{y}\Big)\right\}}{(\pi)^n \det[B^{-1}]}, \tag{3.111}$$

The theorem can be interpreted as saying that $\boldsymbol{X}$ can be represented as $\boldsymbol{X} = [G]\,\boldsymbol{Y} + \boldsymbol{V}$, where $\boldsymbol{Y}$ and $\boldsymbol{V}$ are independent, circularly symmetric, and Gaussian. For a given $\boldsymbol{Y} = \boldsymbol{y}$, the conditional density of $\boldsymbol{X}$ can be denoted as $\mathcal{CN}(-[B^{-1}C]\,\boldsymbol{y}\,, [B^{-1}])$. The notation $\mathcal{CN}([A], [K])$ here means that the r$\mathbf{v}$ has a mean $[A]$ and a fluctuation around $[A]$ which is circularly symmetric Gaussian with covariance $[B]$. Thus, in this case, $\boldsymbol{X}$, conditional on $\boldsymbol{Y} = \boldsymbol{y}$, is cicularly symmetric with covariance $[B]^{-1}$ around the mean $[B^{-1}C]$.

The matrix $[G] = -[B^{-1}C]$ and the covariance matrix of $\boldsymbol{V}$ is $[B^{-1}]$. As in Theorem 3.5.2, the matrices $[G]$ and $[K_{\boldsymbol{V}}]$ can be expressed directly in terms of the joint covariances of $\boldsymbol{X}$ and $\boldsymbol{Y}$ as

$$G = [K_{X\cdot Y} K_Y^{-1}] \tag{3.112}$$

$$[K_{\boldsymbol{V}}] = [K_{\boldsymbol{X}}] - [K_{\boldsymbol{X}\cdot\boldsymbol{Y}} K_{\boldsymbol{Y}}^{-1} K^\dagger_{\boldsymbol{X}\cdot\boldsymbol{Y}}] \tag{3.113}$$

Conversely, if $\boldsymbol{X}$ can be expressed as $\boldsymbol{X} = G\,\boldsymbol{Y} + \boldsymbol{V}$ where $\boldsymbol{Y}$ and $\boldsymbol{V}$ are independent and each circularly-symmetric Gaussian, then it is easy to see that $(\boldsymbol{X}^\mathsf{T}, \boldsymbol{Y}^\mathsf{T})^\mathsf{T}$ must be circularly symmetric Gaussian. Using the resultant symmetry between $\boldsymbol{X}$ and $\boldsymbol{Y}$, we see that there must be a matrix $[H]$ and a r$\mathbf{v}$ $\boldsymbol{Z}$ so that $\boldsymbol{Y} = \boldsymbol{H}\,\boldsymbol{X} + \boldsymbol{Z}$ where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are independent. We will see how this is used for estimation in Section 10.5

3.7.8 Circularly-symmetric Gaussian processes

In this section, we modify Section 3.6 on continuous-time Gaussian processes to briefly outline the properties of circularly-symetric Gaussian processes.

Definition 3.7.4. *A circularly-symmetric Gaussian process $\{X(t); t \in \mathbb{R}\}$ is a complex stochastic process such that for all positive integers k and all choices of epochs $t_1, \ldots, t_k \in \mathbb{R}$, the complex n-r$\mathbf{v}$ with components $X(t_1), \ldots, X(t_k)$ is a circularly-symmetric Gaussian n-r$\mathbf{v}$.*

Now assume that $\{X(t);\, t \in \mathbb{R}\}$ is a circularly-symmetric Gaussian process. Since each n-r$\mathbf{v}$ $(X(t_1), \ldots, X(t_k))^\mathsf{T}$ is circularly symmetric, the corresponding pseudo-covariance matrix is 0 and the covariance matrix specifies the distribution of $(X(t_1), \ldots, X(t_k))^\mathsf{T}$. It

follows then that the pseudo-covariance function, $M_X(t,\tau) = \mathsf{E}[X(t)X(\tau)] = 0$ for all t,τ and the covariance function $K_X(t,\tau) = \mathsf{E}[X(t)X^*(\tau)]$ for all t,τ specifies all finite joint distributions.

A convenient way of generating a circularly-symmetric Gaussian process is to start with a sequence of (complex) orthonormal functions $\{\phi_n(t);\, n \in \mathbb{Z}\}$ and a sequence of independent circularly-symmetric Gaussian rv's $\{X_n \sim \mathcal{CN}(0,\sigma_n^2);\, n \in \mathbb{Z}\}$. Then if $\sum_n \sigma_n^2 \phi_n^2(t) < \infty$ for all t, it follows, as in Theorem 3.6.3 for ordinary Gaussian processes, that $X(t) = \sum_n X_n \phi_n(t)$ is a circularly-symmetric Gaussian process.

One convenient such orthonormal expansion is the set of functions $\phi_n(t) = e^{i2\pi nt/T}$ for $t \in (-T/2,\, T/2)$ used in the Fourier series over that time interval. The interpretation here is very much like that in Example 3.6.5, but here the functions are complex, the rv's are circularly symmetric Gaussian, and the arithmetic is over $\mathbb{C}$.

Another particularly convenient such expansion is the sinc-function expansion of Section 3.6.4. The sinc functions are real, but the expansion is now over the complex field using circularly symmetric rv's. It is intuitively plausible in this case that the process is circularly symmetric, since the real and imaginary parts of the process are identically distributed.

A complex stochastic process $\{X(t);\, t \in \mathbb{R}\}$ can be filtered by a complex filter with impulse response $h(t)$. The output is then the complex convolution $Y(\tau) = \int X(t)h(\tau - t)\, dt$. If $X(t)$ is a circularly symmetric Gaussian process expressed as an orthonormal expansion, then by looking at $Y(\tau)$ over say $\tau_1, \ldots, \tau_k$, we can argue as before that $\{Y(\tau); \tau \in \mathbb{R}\}$ is a circularly symmetric process if its power is finite at all τ. When circularly symmetric sinc processes are passed through filters, we have a broad class of circularly symmetric processes.

The definition of stationarity is the same for complex stochastic processes as for (real) stochastic processes, but the CDF over say $X(t_1), \ldots, X(t_k)$ is now over both the real and imaginary parts of those complex rv's. If $X(t_1), \ldots, X(t_k)$ are circularly symmetric Gaussian, however, then these distributions are determined by the covariance matrices. Thus, circularly symmetric Gaussian processes are stationary if and only if the covariance function satisfies $K_X(t, t+u)) = K_X(0,u)$.

For a stationary circularly-symmetric Gaussian process $\{X(t);\, t \in \mathbb{R}\}$, the covariance function can be expressed as a function of a single variable, $K_X(u)$. This function must be Hermition (*i.e.*, it must satisfy $K_X(t) = K_X^*(-t)$). The Fourier transform of a Hermitian function must be real, and by repeating the argument in Section 3.6.7, we see that this fourier transform must be nonnegative. This Fourier transform is called the spectral density of $\{X(t);\, t \in \mathbb{R}\}$.

The spectral density of a stationary circularly-symmetric Gaussian process has the same interpretation as the spectral density of a (real) stationary Gaussian process. White Gaussian noise is defined and interpreted the same way as in the real case, and can be approximated in the same way by Gaussian sinc processes.

It is important to understand that these very close analogies between real and complex Gaussian processes are actually between real and *circularly-symmetric* Gaussian processes. A complex Gaussian process that is not circularly symmetric does not have very nice properties and is perhaps better thought of as a pair of processes, one real and one pure imaginary.

3.8 Summary

The sum of sufficiently many rv's that are not too dependent tends toward the Gaussian distribution, and multiple such sums tend toward a jointly-Gaussian distribution. The requirements for a set of n rv's to be jointly Gaussian are far more stringent than the requirement that each be Gaussian alone, but fortunately, as above, the conditions that lead individual rv's to be Gaussian often also lead multiple rv's to be jointly Gaussian. Theorem 3.4.8 collects 4 sets of necessary and sufficient conditions for zero-mean rv's to be jointly Gaussian. Non-zero-mean rv's are jointly Gaussian if their fluctuations are jointly Gaussian. Finally, a random vector is defined to be Gaussian if its components are jointly Gaussian.

The distribution of a Gaussian vector $\boldsymbol{Z}$ is completely specified by its mean $\overline{\boldsymbol{Z}}$ and covariance matrix $[K_{\boldsymbol{Z}}]$. The distribution is denoted as $\mathcal{N}(\overline{\boldsymbol{Z}}, [K_{\boldsymbol{Z}}])$.

If $X_1, X_2, \ldots, X_n, Y_1, \ldots, Y_m$ are zero mean and jointly Gaussian with a non-singular covariance matrix, then the conditional density $\mathsf{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y})$ is jointly Gaussian for each $\boldsymbol{y}$. The covariance of this conditional distribution is $[K_{\boldsymbol{X}}] - [K_{\boldsymbol{X}\cdot\boldsymbol{Y}} K_{\boldsymbol{Y}}^{-1} K_{\boldsymbol{X}\cdot\boldsymbol{Y}}^{\mathsf{T}}]$, which does not depend on the particular sample value $\boldsymbol{y}$. The conditional mean, $[K_{\boldsymbol{X}\cdot\boldsymbol{Y}} K_{\boldsymbol{Y}}^{-1}]\boldsymbol{y}$, depends linearly on $\boldsymbol{y}$. This situation can be equivalently formulated as $\boldsymbol{X} = [G]\,\boldsymbol{Y} + \boldsymbol{V}$, where $\boldsymbol{V}$ is a zero-mean Gaussian n-r$\mathbf{v}$ independent of $\boldsymbol{Y}$. Using the symmetry between the roles of $\boldsymbol{X}$ and $\boldsymbol{Y}$, we have $\boldsymbol{Y} = [H]\boldsymbol{X} + \boldsymbol{Z}$ where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are independent.

A stochastic process $\{X(t);\, t \in \mathcal{T}\}$ is a Gaussian process if, for all finite sets $t_1, \ldots, t_k$ of epochs, the rv's $X(t_1), \ldots, X(t_k)$ are jointly Gaussian. If $\mathcal{T}$ is $\mathbb{R}$ or $\mathbb{R}^+$, then Gaussian processes can be easily generated as orthonormal expansions. When a Gaussian orthonormal expansion is used as the input to a linear filter, the output is essentially also a Gaussian process.

A stochastic process is stationary if all finite joint distributions are invariant to time shifts. It is WSS if the covariance function and mean are invariant to time shifts. A zero-mean Gaussian process is stationary if the covariance function is invariant to time shifts, *i.e.*, if $K_X(t, t+u) = K_X(0, u)$. Thus a stationary zero-mean Gaussian process is determined by its single-variable covariance function $K_X(u)$. A stationary zero-mean Gaussian process is also essentially determined by its spectral density, which is the Fourier transform of $K_X(u)$.

The spectral density $S_X(f)$ is interpreted as the process's power per unit frequency at frequency f. If a stationary zero-mean Gaussian process has a positive spectral density equal to a constant value, say $N_0/2$, over all frequencies of interest, it is called white Gaussian noise. Approximating $S_X(f)$ as constant over all f is often convenient, but implies infinite aggregate power, which is both mathematically and physically absurd. The Gaussian sinc process also models white Gaussian noise over an arbitrarily broad but finite band of frequencies.

Circularly-symmetric rv's are complex random variables for which the distribution over the real and imaginary plane is circularly symmetric. A random vector $\boldsymbol{Z}$ is circularly symmetric if $\boldsymbol{Z}$ and $e^{i\theta}\boldsymbol{Z}$ have the same distribution for all phases ϕ. Theorem 3.7.5 collects 5 sets of necessary and sufficient conditions for complex random vectors to be circularly-symmetric Gaussian.

Vectors and processes of circularly-symmetric Gaussian rv's have many analogies with ordinary Gaussian random vectors and processes, and many of the equations governing the real case can be converted to the circularly-symmetric case simply by replacing transposes by Hermitian transposes. This is true for conditional distributions also. Unfortunately, this extreme simplicity relating the equations sometimes hides more fundamental differences. Complex Gaussian random vectors and processes that are not circularly symmetric are usually best modeled as separate real and imaginary parts, since almost all of the insights that we might try to transfer from the real to complex case fail except when circular symmetry is present.

3.9 Exercises

Exercise 3.1. a) Let X, Y be IID rv's, each with density $\mathsf{f}_X(x) = \alpha \exp(-x^2/2)$. In part (b), we show that α must be $1/\sqrt{2\pi}$ in order for $\mathsf{f}_X(x)$ 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.

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 $\mathsf{E}\left[X^2\right] = 1$.

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

Exercise 3.2. a) By expanding in a power series in $(1/2)r^2\sigma^2$, show that

$$\exp\left(\frac{r^2\sigma^2}{2}\right) = 1 + \frac{r^2\sigma^2}{2} + \frac{r^4\sigma^4}{2(2^2)} + \cdots + \frac{r^{2k}\sigma^{2k}}{k!2^k} + \cdots .$$

b) By expanding e^{rZ} in a power series in rZ, show that

$$\mathsf{g}_Z(r) = \mathsf{E}\left[e^{rZ}\right] = 1 + r\mathsf{E}[Z] + \frac{r^2\mathsf{E}\left[Z^2\right]}{2} + \cdots + \frac{r^k\mathsf{E}\left[Z^i\right]}{(k)!} + \cdots .$$

c) By matching powers of r between parts (a) and (b), show that for all integer $k \geq 1$,

$$\mathsf{E}\left[Z^{2k}\right] = \frac{(2k)!\sigma^{2k}}{k!2^k} = (2k-1)(2k-3)\cdots(3)(1)\sigma^{2k} \quad ; \qquad \mathsf{E}\left[Z^{2k+1}\right] = 0.$$

Exercise 3.3. Let X and Z be IID normalized Gaussian random variables. Let $Y = |Z|\,\mathrm{Sgn}(X)$, where $\mathrm{Sgn}(X)$ is 1 if $X \geq 0$ and -1 otherwise. Show that X and Y are each Gaussian, but are not jointly Gaussian. Sketch the contours of equal joint probability density.

Exercise 3.4. a) Let $X_1 \sim \mathcal{N}(0, \sigma_1^2)$ and let $X_2 \sim \mathcal{N}(0, \sigma_2^2)$ be independent of X_1. Convolve the density of X_1 with that of X_2 to show that $X_1 + X_2$ is Gaussian, $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$.

b) Let W_1, W_2 be IID normalized Gaussian rv's . Show that $a_1W_1 + a_2W_2$ is Gaussian, $\mathcal{N}(0, a_1^2+a_2^2$. Hint: You could repeat all the equations of part a), but the insightful approach is to let $X_i = a_iW_i$ for $i = 1, 2$ and then use part a) directly.

c) Combine part (b) with induction to show that all linear combinations of IID normalized Gaussian rv's are Gaussian.

Exercise 3.5. a) Let $\boldsymbol{U}$ be an n-r**v** with mean $\boldsymbol{m}$ and covariance $[K]$ whose MGF is given by (3.18). Let $X = r^{\mathsf{T}}\boldsymbol{U}$ for an arbitrary real vector $\boldsymbol{r}$. Show that the MGF of X is given by $\mathsf{g}_X(r) = \exp\left[r\mathsf{E}[X] + r^2\sigma_X^2/2\right]$ and relate $\mathsf{E}[X]$ and σ_X^2 to $\boldsymbol{m}$ and $[K]$.

b) Show that $\boldsymbol{U}$ is a Gaussian r**v**.

Exercise 3.6. a) Let $\boldsymbol{Z} \sim \mathcal{N}(0, [K])$ be n-dimensional. By expanding in a power series in $(1/2)\boldsymbol{r}^{\mathsf{T}}[K]\,\boldsymbol{r}$, show that

$$\mathsf{g}_{\boldsymbol{Z}}(\boldsymbol{r}) = \exp\left[\frac{\boldsymbol{r}^{\mathsf{T}}[K]\boldsymbol{r}}{2}\right] = 1 + \frac{\sum_{j,k} r_j r_k K_{j,k}}{2} + \cdots + \frac{\left(\sum_{j,k} r_j r_k K_{j,k}\right)^m}{2^m m!} + \cdots .$$

b) By expanding $e^{r_j Z_j}$ in a power series in $r_j Z_j$ for each j, show that

$$\mathsf{g}_{\boldsymbol{Z}}(r) = \mathsf{E}\left[\exp\left(\sum_j r_j Z_j\right)\right] = \sum_{j_1=0}^{\infty} \cdots \sum_{j_n=0}^{\infty} \frac{r_1^{j_1}}{(j_1)!} \cdots \frac{r_n^{j_n}}{(j_n)!} \mathsf{E}\left[Z_1^{j_1} \ldots Z_n^{j_n}\right].$$

c) Let $D = \{j_1, j_2, \ldots, j_{2m}\}$ be a set of $2m$ distinct integers each between 1 and n. Consider the term $r_{j_1} r_{j_2} \cdots r_{j_{2m}} \mathsf{E}[Z_{j_1} Z_{j_2} \cdots Z_{j_{2m}}]$ in part (b). By comparing with the set of terms in part (a) containing the same product $r_{j_1} r_{j_2} \cdots r_{j_{2m}}$, show that

$$\mathsf{E}[Z_{j_1} Z_{j_2} \cdots Z_{j_{2m}}] = \frac{\sum_{j_1 j_2 \ldots j_{2m}} K_{j_1 j_2} K_{j_3 j_4} \cdots K_{j_{2m-1} j_{2m}}}{2^m m!},$$

where the sum is over all permutations $(j_1, j_2, \ldots, j_{2m})$ of the set D.

d) Find the number of permutations of D that contain the same set set of unordered pairs $(\{j_1, j_2\}, \ldots, \{j_{2m-1}, j_{2m}\})$. For example, $(\{1, 2\}, \{3, 4\})$ is the same set of unordered pairs as $(\{3, 4\}, \{2, 1\})$. Show that

$$\mathsf{E}[Z_{j_1} Z_{j_2} \cdots Z_{j j_{2m}}] = \sum_{j_1, j_2, \cdots, j_{2m}} K_{j_1 j_2} K_{j_3 j_4} \cdots K_{j_{2m-1} j_{2m}}, \tag{3.114}$$

where the sum is over distinct sets of unordered pairs of the set D. Note: another way to say the same thing is that the sum is over the set of all permutations of D for which $j_{2k-1} < j_{2k}$ for $1 \le k \le m$ and $j_{2k-1} < j_{2k+1}$ for $1 \le ki \le m - 1$.

e) To find $\mathsf{E}\left[Z_1^{j_1} \cdots Z_n^{j_n}\right]$, where $j_1 + j_2 + \cdots + j_n = 2m$, construct the random variables $U_1, \ldots, U_{2m}$, where $U_1, \ldots, U_{j_1}$ are all identically equal to Z_1, where $U_{j_1+1}, \ldots, U_{j_1+j_2}$ are identically equal to Z_2, etc., and use (i) to find $\mathsf{E}[U_1 U_2 \cdots U_{2m}]$. Use this formula to find $\mathsf{E}\left[Z_1^2 Z_2 Z_3\right]$, $\mathsf{E}\left[Z_1^2 Z_2^2\right]$, and $\mathsf{E}[Z_1]^4$.

Exercise 3.7. Let $[Q]$ be an orthonormal matrix. Show that the squared distance between any two vectors $\boldsymbol{z}$ and $\boldsymbol{y}$ is equal to the squared distance between $[Q]\boldsymbol{z}$ and $[Q]\boldsymbol{y}$.

Exercise 3.8. a) Let $[K] = \left[\begin{smallmatrix} .75 & .25 \\ .25 & .75 \end{smallmatrix}\right]$. Show that 1 and 1/2 are eigenvalues of $[K]$ and find the normalized eigenvectors. Express $[K]$ as $[Q\Lambda Q^{-1}]$ where $[\Lambda]$ is diagonal and $[Q]$ is orthonormal.

b) Let $[K'] = \alpha[K]$ for real $\alpha \neq 0$. Find the eigenvalues and eigenvectors of $[K']$. Don't use brute force—think!

c) Find the eigenvalues and eigenvectors of $[K^m]$, where $[K^m]$ is the m^{th} power of $[K]$.

Exercise 3.9. Let X and Y be jointly Gaussian with means m_X, m_Y, variances σ_X^2, σ_Y^2, and normalized covariance ρ. Find the conditional density $\mathsf{f}_{X|Y}(x \mid y)$.

Exercise 3.10. a) Let X and Y be zero-mean jointly Gaussian with variances σ_X^2, σ_Y^2, and normalized covariance ρ. Let $V = Y^3$. Find the conditional density $\mathsf{f}_{X|V}(x \mid v)$. Hint: This requires no computation.

b) Let $U = Y^2$ and find the conditional density of $\mathsf{f}_{X|U}(x \mid u)$. Hint: first understand why this is harder than part a).

Exercise 3.11. a) Let $(\boldsymbol{X}^\mathsf{T}, \boldsymbol{Y}^\mathsf{T})$ have a non-singular covariance matrix $[K]$. Show that $[K_{\boldsymbol{X}}]$ and $[K_{\boldsymbol{Y}}]$ are positive definite, and thus non-singular.

b) Show that the matrices $[B]$ and $[D]$ in (3.39) are also positive definite and thus non-singular.

Exercise 3.12. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be jointly-Gaussian r**v**'s with means $\boldsymbol{m}_{\boldsymbol{X}}$ and $\boldsymbol{m}_{\boldsymbol{Y}}$, covariance matrices $[K_{\boldsymbol{X}}]$ and $[K_{\boldsymbol{Y}}]$ and cross covariance matrix $[K_{\boldsymbol{X}\boldsymbol{Y}}]$. Find the conditional probability density $\mathsf{f}_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y})$. Assume that the covariance of $(\boldsymbol{X}^\mathsf{T}, \boldsymbol{Y}^\mathsf{T})$ is non-singular. Hint: think of the fluctuations of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Exercise 3.13. a) Let $\boldsymbol{W}$ be a normalized IID Gaussian n-r**v** and let $\boldsymbol{Y}$ be a Gaussian m-r**v**. Suppose we would like the joint covariance $\mathsf{E}[\boldsymbol{W}\boldsymbol{Y}^\mathsf{T}]$ to be some arbitrary real-valued $n \times m$ matrix $[K]$. Find the matrix $[A]$ such that $\boldsymbol{Y} = [A]\boldsymbol{W}$ achieves the desired joint covariance. Note: this shows that any real-valued $n \times m$ matrix is the joint covariance matrix for some choice of random vectors.

b) Let $\boldsymbol{Z}$ be a zero-mean Gaussian n-r**v** with non-singular covariance $[K_{\boldsymbol{Z}}]$, and let $\boldsymbol{Y}$ be a Gaussian m-r**v**. Suppose we would like the joint covariance $\mathsf{E}[\boldsymbol{Z}\boldsymbol{Y}^\mathsf{T}]$ to be some arbitrary $n \times m$ matrix $[K']$. Find the matrix $[B]$ such that $\boldsymbol{Y} = [B]\boldsymbol{Z}$ achieves the desired joint covariance. Note: this shows that any real valued $n \times m$ matrix is the joint covariance matrix for some choice of random vectors $\boldsymbol{Z}$ and $\boldsymbol{Y}$ where $[K_{\boldsymbol{Z}}]$ is given (and non-singular).

c) Now assume that $\boldsymbol{Z}$ has a singular covariance matrix in part b). Explain the constraints this places on possible choices for the joint covariance $\mathsf{E}[\boldsymbol{Z}\boldsymbol{Y}^\mathsf{T}]$. Hint: your solution should involve the eigenvectors of $[K_{\boldsymbol{Z}}]$.

Exercise 3.14. a) Let $\boldsymbol{W} = (W_1, W_2, \ldots, W_{2n})^\mathsf{T}$ be a $2n$ dimensional IID normalized Gaussian r**v**. Let $S_{2n} = W_1^2 + W_2^2 + \cdots + W_{2n}^2$. Show that S_{2n} is an nth order Erlang rv with parameter 1/2, i.e., that $\mathsf{f}_{S_{2n}}(s) = 2^{-n} s^{n-1} e^{-s/2}/(n-1)!$. Hint: look at S_2 from Exercise 3.1.

b) Let $R_{2n} = \sqrt{S_{2n}}$. Find the probability density of R_{2n}.

c) Let $v_{2n}(r)$ be the volume of a $2n$ dimensional sphere of radius r and let $b_{2n}(r)$ be the surface area of that sphere, i.e., $b_{2n}(r) = dv_{2n}(r)/dr$. The point of this exercise is to show how to calculate these quantities. By considering an infinitesimally thin spherical shell of thickness δ at radius r, show that

$$\mathsf{f}_{R_{2n}}(r) = b_{2n}(r) \mathsf{f}_{\boldsymbol{W}}(\boldsymbol{w}) \,|_{\boldsymbol{W}:\, \boldsymbol{W}^\mathsf{T}\boldsymbol{W}=r^2} \,.$$

d) Calculate $b_{2n}(r)$ and $v_{2n}(r)$. Note that for any fixed $\delta \ll r$, the volume within δ of the surface of a sphere of radius r to the total volume of the sphere approaches 1 with increasing n.

Exercise 3.15. a) Solve directly for $[B]$, $[C]$, and $[D]$ in (3.39) for the one dimensional case where $n = m = 1$. Show that (3.40) agrees with (3.37)

Exercise 3.16. a) Express $[B]$, $[C]$, and $[D]$, as defined in (3.39), in terms of $[K_{\boldsymbol{X}}]$, $[K_{\boldsymbol{Y}}]$ and $[K_{\boldsymbol{X}\cdot\boldsymbol{Y}}]$ by multiplying the block expression for $[K]$ by that for $[K]^{-1}$. You can check your solutions against those in (3.46) to (3.48). Hint: You can solve for $[B]$ and $[C]$ by looking at only two of the four block equations in $[KK^{-1}]$. You can use the symmetry between $\boldsymbol{X}$ and $\boldsymbol{Y}$ to solve for $[D]$.

b) Use your result in part a) for $[C]$ plus the symmetry between $\boldsymbol{X}$ and $\boldsymbol{Y}$ to show that

$$[BK_{\boldsymbol{X}\cdot\boldsymbol{Y}}K_{\boldsymbol{Y}}^{-1}] = [K_{\boldsymbol{X}}^{-1}K_{\boldsymbol{X}\cdot\boldsymbol{Y}}D]$$

c) For the formulations $\boldsymbol{X} = [G]\,\boldsymbol{Y} + \boldsymbol{V}$ and $\boldsymbol{Y} = [H]\boldsymbol{X} + \boldsymbol{Z}$ where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are zero-mean, jointly Gaussian and have a non-singular combined covariance matrix, show that

$$[K_{\boldsymbol{V}}^{-1}G] = [H^\mathsf{T}K_{\boldsymbol{Z}}^{-1}] \tag{3.115}$$

Hint: This is almost trivial from part b), (3.43), (3.44), and the symmetry.

Exercise 3.17. Let $\boldsymbol{X}$ and $\boldsymbol{Z}$ be statistically independent Gaussian r**v**'s of arbitrary dimension n and m respectively. Let $\boldsymbol{Y} = [H]\boldsymbol{X} + \boldsymbol{Z}$ where $[H]$ is an arbirary real $n \times m$ matrix.

a) Explain why $X_1, \ldots, X_n, Z_1, \ldots, Z_m$ must be jointly Gaussian rv's. Then explain why $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ must be jointly Gaussian.

b) Show that if $[K_{\boldsymbol{X}}]$ and $[K_{\boldsymbol{Z}}]$ are non-singular, then the combined covariance matrix $[K]$ for $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)^\mathsf{T}$ must be non-singular.

Exercise 3.18. a) Verify (3.56) for $k = 0$ by the use of induction on (3.54).

b Verify (3.56) for $k = 0$ directly from (3.55)

c) Verify (3.56) for $k > 0$ by using induction on k.

d) Verify that

$$\lim_{n\to\infty} \mathsf{E}\left[X(n)X(n+k)\right] = \frac{\sigma^2\alpha^k}{1-\alpha^2}.$$

Exercise 3.19. Let $\{X(t);\, t\in\Re\}$ be defined by $X(t) = tA$ for all $t \in \Re$ where $A \sim N(0,1)$. Show that this is a Gaussian process. Find its mean for each t and find its covariance function. Note: The purpose of this exercise is to show that Gaussian processes can be very degenerate and trivial.

Exercise 3.20. Let $\{X(t);\, t \geq 0\}$ be a stochastic process with independent and stationary increments and let $X(0)$ be an arbitrary random variable. Show that $\mathsf{E}\left[X(t)\right] = \mathsf{E}\left[X(0)\right] + t\,\mathsf{E}\left[X(1) - X(0)\right]$ and that

$$K_X(t,\tau) = \mathsf{VAR}\left[X(0)\right] + t\left[\mathsf{VAR}\left[X(1)\right] - \mathsf{VAR}\left[X(0)\right]\right].$$

Exercise 3.21. a) Let $X(t) = R\cos(2\pi ft + \theta)$ where R is a Rayleigh rv and the rv θ is independent of R and uniformly distributed over the interval 0 to 2π. Show that $\mathsf{E}\left[X(t)\right] = 0$.

b) Show that $\mathsf{E}\left[X(t)X(t+\tau)\right] = \frac{1}{2}\mathsf{E}\left[R^2\right]\cos(2\pi f\tau)$.

c) Show that $X(t); t\in\Re$ is a Gaussian process.

Exercise 3.22. Let $h(t)$ be a real square-integrable function whose Fourier transform is 0 for $|f| > B$ for some $B > 0$. Show that $\sum_n h^2(2Bt - n) = (1/2B)\int h^2(\tau)\,d\tau$ for all $t \in \mathbb{R}$. Hint: find the sampling theorem expansion for a time shifted sinc function.

Exercise 3.23. a) Let $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\mathsf{T}}$ be a circularly-symmetric n-r**v**. Show that Z_k is circularly symmetric for each k, $1 \leq k \leq n$. Hint: use the definition directly (you cannot assume that $\boldsymbol{Z}$ is also Gaussian).

b) Show that $Z_1 + Z_2$ is circularly symmetric. For any complex n-vector $\boldsymbol{c}$, show that $\boldsymbol{c}^{\mathsf{T}}\boldsymbol{Z}$ is a circularly symmetric rv.

Exercise 3.24. Let A be a complex Gaussian rv, i.e., $A = A_1 + iA_2$ where A_1 and A_2 are zero-mean jointly-Gaussian real rv's with variances σ_1^2 and σ_2^2 respectively.

a) Show that $\mathsf{E}\left[AA^*\right] = \sigma_1^2 + \sigma_2^2$.

b) Show that

$$\mathsf{E}\left[(AA^*)^2\right] = 3\sigma_1^4 + 3\sigma_2^4 + 2\sigma_1^2\sigma_2^2 + 4\left(\mathsf{E}\left[A_1A_2\right]\right)^2.$$

c) Show that $\mathsf{E}\left[(AA^*)^2\right] \geq 2\left(\mathsf{E}\left[AA^*\right]\right)^2$ with equality if and only if A_1 and A_2 are IID. Hint: Lower bound $\left(\mathsf{E}\left[A_1A_2\right]\right)^2$ by 0.

d Show that $\mathsf{VAR}\left[AA^*\right] \geq \left(\mathsf{E}\left[AA^*\right]\right)^2$.

Exercise 3.25. Let $K_X(t)$ be the covariance function of a WSS process $\{X(t); t \in \Re\}$. Show that if $K_X(t)$ is continuous at $t = 0$, then it is continuous everywhere. Hint: You must show that $\lim_{\delta \to 0} \mathsf{E}\left[X(0)(X(t+\delta) - X(t))\right] = 0$ for all t. Use the Schwarz inequality.